**2019 Urban Rate Survey – Fixed Broadband Service Analysis**

**Introduction**

Every year, the Wireline Competition Bureau (Bureau) conducts a survey of residential standalone Internet access service rates “to help ensure that universal service support recipients offering fixed voice and broadband services do so at reasonably comparable rates to those in urban areas.”[[1]](#footnote-2) The Bureau adopted the general methodology for surveying terrestrial fixed broadband providers in 2013. The 2019 Urban Rate Survey (URS) for fixed broadband services follows the same methodology as the 2018 survey. This document shows how the fixed broadband reasonable comparability benchmark was calculated for 2019, including differences in the data received in the survey and changes from the analysis used for the 2018 data.

The 2019 reasonable comparability benchmarks calculated for fixed broadband services are similar to the 2017 reasonable comparability benchmarks, which are lower than 2018 reasonable comparability benchmarks.[[2]](#footnote-3) The reason for the change is that the standard deviation in survey data varies over survey years. It appears that this variation in standard deviation over time is likely due to an insufficient sample size – that the standard deviation of the sample collected via the URS likely differs from the standard deviation of the population of all urban rates. The increase in the number of plans offered and survey responses (the rapid changes in the rate plans offered in the consumer market for fixed broadband services as described below) and the resulting large number of sample strata, as well as the changes in standard deviation in the survey responses from 2017-2019, demonstrate that the sample size is too small.

**Sample Design and Selection**

The sampling unit for the 2019 fixed broadband survey was a service provider census tract pair. The frame[[3]](#footnote-4) for the 2019 URS was the set of sampling units of providers offering terrestrial fixed broadband service to residential customers in urban census tracts based on FCC Form 477 December 2017 data. The frame consisted of 150,092 sampling units, encompassing 1,344 service providers and 58,162 census tracts. The likelihood of a service provider census tract pair receiving a survey is based on the number of potential subscribers for that provider in that census tract.

For each sampling unit other than the terrestrial fixed wireless providers, the number of potential subscribers is calculated as:

*Number of potential subscribers = Provider Presence Ratio x (Number of households in the sampling unit’s census tract)*

Provider Presence Ratio was calculated as the fraction of housing units in the census tract for which the provider reported service availability via Form 477.

We calculated the number of potential subscribers differently for the terrestrial fixed wireless providers because the number of potential customers for such services is limited by geographic and technological factors. Many terrestrial fixed wireless providers serve suburban areas that are of moderate population density. Also, the number of housing units in these areas is likely higher than fixed wireless providers have capacity to serve. Accordingly, for each sampling unit of these providers, the number of potential subscribers is calculated as:[[4]](#footnote-5)

*Number of potential subscribers = 2 x (Number of residential subscribers in the sampling unit’s census tract).*

The number of potential subscribers was not allowed to exceed the number of households in the sampling unit’s census tract.

The 2019 URS follows the stratification of the 2018 URS. The frame was divided into strata to account for the differing rate variability in each stratum.

The strata included in the 2019 URS are listed below. There are 28 strata: 13 strata for services with download bandwidth less than 500 Mbps, 13 strata for services with download bandwidth greater than or equal to 500 Mbps (high bandwidth strata),[[5]](#footnote-6) and two Alaska strata.

* Service download bandwidth < 500 Mbps
	+ AT&T (AT&T Services, Inc.)
	+ CenturyLink (CenturyLink, Inc., CenturyLink Communications, LLC)
	+ Charter (Charter Communications, Inc.)[[6]](#footnote-7)
	+ Comcast (Comcast Cable Communications, INC.)
	+ Cox (Cox Communications)
	+ CSC Holdings (CSC Holdings LLC)
	+ Frontier (Frontier Communications Corporation)
	+ Verizon (Verizon New York Inc., Verizon Pennsylvania LLC, Verizon New Jersey Inc., Verizon California Inc., Verizon New England Inc., Verizon Virginia LLC, Verizon Maryland LLC, Verizon Florida LLC, Verizon Delaware LLC, GTE Southwest Incorporated dba Verizon Southwest, Verizon Washington, DC Inc.)
	+ WideOpenWest (Knology, WideOpenWest, and Wiregrass Telcom)
	+ Windstream (service providers identifying Windstream as their holding company)
	+ Terrestrial fixed wireless providers
	+ Major[[7]](#footnote-8)
	+ Minor
* Service download bandwidth ≥500 Mbps (high bandwidth strata)
	+ AT&T
	+ CenturyLink
	+ Charter
	+ Comcast
	+ Cox
	+ CSC Holdings
	+ Frontier
	+ Verizon
	+ WideOpenWest
	+ Windstream
	+ Terrestrial Fixed Wireless
	+ Major
	+ Minor
* Alaska
	+ Fixed wireline services
	+ Terrestrial fixed wireless

The table below presents the sampling plan including the sample size for each stratum. Sampling units were selected randomly from each stratum, with unequal selection probability proportional to providers’ number of potential subscribers in a given tract.[[8]](#footnote-9) The sample sizes for each stratum are a reflection of the estimated number of potential subscribers in the stratum and the estimated variability of offered rates from last year’s URS.

|  |  |  |
| --- | --- | --- |
|   | Frame | Sample |
| Stratum | Units | Providers | Census Tracts | Offers | Units | Providers | Census Tracts | Offers |
| Overall |  150,092  |  1,344  |  58,162  |  197,431,411  |  550  |  176  |  548  |  871,621  |
| AT&T Services, Inc. |  15,163  |  1  |  15,163  |  20,471,517  |  10  |  1  |  10  |  15,889  |
| CenturyLink |  5,250  |  1  |  5,250  |  7,447,452  |  8  |  1  |  8  |  11,334  |
| Charter |  16,920  |  1  |  16,920  |  26,337,968  |  54  |  1  |  54  |  89,250  |
| Comcast Cable Communications, LLC |  4,892  |  1  |  4,892  |  6,902,581  |  5  |  1  |  5  |  9,705  |
| Cox |  3,682  |  1  |  3,682  |  5,113,724  |  5  |  1  |  5  |  10,153  |
| CSC Holdings LLC |  2,867  |  1  |  2,867  |  4,208,171  |  5  |  1  |  5  |  8,591  |
| Frontier |  6,295  |  1  |  6,295  |  8,827,913  |  10  |  1  |  10  |  16,540  |
| Verizon |  4,897  |  8  |  4,897  |  7,240,840  |  5  |  4  |  5  |  8,786  |
| WideOpenWest |  56  |  5  |  55  |  63,598  |  5  |  3  |  5  |  10,577  |
| Windstream |  1,436  |  38  |  1,319  |  977,876  |  5  |  5  |  5  |  8,472  |
| Terrestrial Fixed Wireless |  12,977  |  611  |  9,042  |  790,214  |  79  |  64  |  78  |  31,534  |
| Major |  3,859  |  321  |  3,156  |  6,822,805  |  22  |  17  |  22  |  45,270  |
| Minor |  7,395  |  583  |  6,474  |  2,364,082  |  21  |  15  |  21  |  12,704  |
| AT&T Services, Inc. (high bandwidth) |  12,291  |  1  |  12,291  |  21,146,314  |  56  |  1  |  56  |  108,864  |
| CenturyLink (high bandwidth) |  4,010  |  1  |  4,010  |  7,008,293  |  5  |  1  |  5  |  13,563  |
| Charter (high bandwidth) |  4,082  |  1  |  4,082  |  6,441,831  |  13  |  1  |  13  |  28,068  |
| Comcast Cable Communications, LLC (high bandwidth) |  21,111  |  1  |  21,111  |  32,915,763  |  82  |  1  |  82  |  154,451  |
| Cox (high bandwidth) |  1,586  |  1  |  1,586  |  2,556,035  |  5  |  1  |  5  |  9,362  |
| CSC Holdings LLC (high bandwidth) |  6  |  1  |  6  |  10,912  |  5  |  1  |  5  |  9,451  |
| Frontier (high bandwidth) |  407  |  1  |  407  |  756,035  |  5  |  1  |  5  |  8,467  |
| Verizon (high bandwidth) |  8,206  |  8  |  8,206  |  12,881,738  |  5  |  4  |  5  |  8,639  |
| WideOpenWest (high bandwidth) |  2,090  |  2  |  2,090  |  2,718,368  |  5  |  2  |  5  |  5,775  |
| Windstream (high bandwidth) |  413  |  23  |  411  |  700,467  |  5  |  5  |  5  |  14,777  |
| Terrestrial Fixed Wireless (high bandwidth) |  502  |  53  |  502  |  74,645  |  10  |  10  |  10  |  11,271  |
| Major (high bandwidth) |  5,628  |  197  |  5,348  |  10,376,467  |  93  |  47  |  93  |  198,172  |
| Minor (high bandwidth) |  3,845  |  237  |  3,606  |  1,956,500  |  17  |  16  |  17  |  13,188  |
| Alaska |  174  |  3  |  87  |  317,434  |  5  |  3  |  5  |  7,544  |
| Alaska TFW |  52  |  3  |  51  |  1,868  |  5  |  2  |  5  |  1,224  |

**Survey Response**

The table below presents the number of responses, the number of different service providers, and the number of different census tracts requested, received, and received with rates (service provided) in the 2019 URS for fixed broadband service.

|  |  |  |  |
| --- | --- | --- | --- |
| Survey Status | Responses | Service Providers | Census Tracts |
| Requested | 550 | 176 | 548 |
| Received | 511 | 145 | 509 |
| Service Provided | 506 | 141 | 504 |

The next table presents the number of responses, the number of different service providers, the number of different census tracts, and the number of rates for each technology among responses received with rates as of June 30 2018 for the 2019 benchmark.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Technology | Responses | Service Providers | Census Tracts | Rates |
| Cable | 267 | 42 | 267 | 1391 |
| DSL | 152 | 35 | 152 | 1203 |
| FTTH[[9]](#footnote-10) | 69 | 53 | 68 | 275 |
| Fixed Wireless | 153 | 57 | 153 | 743 |

A total of 3,612 rates were provided at a variety of service levels as of September 2018 for the 2019 benchmark. Several rates were excluded from the analysis for the reasons described in Appendix A, resulting in a total of 3,550 rates available for the analysis. The table below presents the number of responses, the number of different service providers, the number of different census tracts, and the number of rates for each technology among responses received with rates available for the analysis.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Technology | Responses | Service Providers | Census Tracts | Rates |
| Cable | 267 | 42 | 267 | 1381 |
| DSL | 152 | 35 | 152 | 1201 |
| FTTH | 69 | 53 | 68 | 253 |
| Fixed Wireless | 152 | 56 | 152 | 715 |
| All | 506 | 141 | 504 | 3550 |

**Monthly Rates and Rate Spreads**

Monthly rates were treated as unique for a combination of census tract, FCC Registration Number (FRN), service name, technology, download bandwidth, upload bandwidth, and capacity allowance. The following average monthly rate was used if the service provider offered multiple rates in the census tract for each unique combination:

* Minimum Rate = Minimum Monthly Charge + Minimum Other Mandatory Charge + Minimum Surcharge
* Maximum Rate = Maximum Monthly Charge + Maximum Other Mandatory Charge + Maximum Surcharge
* Average Rate = (Minimum Rate + Maximum Rate)/2
* Rate Spread = Maximum Rate - Minimum Rate

The following average monthly rate was used if the service provider did not offer multiple rates in the census tract:

* Average Rate = Minimum Monthly Charge + Minimum Other Mandatory Charge + Minimum Surcharge
* Rate Spread = 0

**Weights**

Weights are required to ensure the contributions of each response properly represent the offers that consumers possibly receive nationwide. Weights are also used to ensure that a service provider’s rates do not exert extra influence on the estimate only because the provider offers different services using multiple technologies.

The 2019 survey weight construction is consistent with the 2018 survey weight construction. Each rate was assigned a weight:

*Weight = Sampling Weight x Nonresponse Weight x Same Rate Weight x Service Level Weight x Number of Potential Subscribers*

*Sampling Weight* is the inverse of the selection probability for each sample unit. The selection probability is determined by the total number of units in each stratum, the sample size in each stratum, and the units’ number of potential subscribers described in the sample selection section earlier. Each sample is assigned a sampling weight to reflect its selection probability.

*Nonresponse Weight* is assigned to each stratum in order to compensate for unit nonresponse in each stratum. It is the total number of potential subscribers sampled over the total number of potential subscribers in the sampled census tracts of a given provider who has provided rate responses in each stratum.

*Same Rate Weight* is assigned to the respondents who provided i) multiple service levels or ii) equal service levels via different technologies for the same rate in the same census tract.[[10]](#footnote-11) In such cases, the rate was assigned a Same Rate Weight equal to 1/R, where R is the number of rate responses provided by a service provider at the same rate in the census tract.

*Service Level Weight* is assigned to the respondents who provided multiple rates for the same service level offered via different technologies and/or service names. Each rate was assigned a Service Level Weight equal to 1/L, where L is the number of responses with different rates provided by a service provider for the same service plan (same download bandwidth, upload bandwidth, and monthly capacity allowance) in the census tract.

*Number of Potential Subscribers* is the estimated number of potential customers to whom the providers advertise their service.

The final weight is the product of Sampling Weight, Nonresponse Weight, Same Rate Weight, Service Level Weight, and the Number of Potential Subscribers.

**Average Rate Model**

The 2019 URS shows that broadband rate is nonlinear in proportion to download bandwidth and upload bandwidth (see Appendix B). To estimate an average rate for every possible bandwidth tier combination, we applied a weighted Generalized Boosted Model (GBM),[[11]](#footnote-12) which is an algorithm allowing nonlinearity in our estimation,[[12]](#footnote-13) to all terrestrial fixed broadband services with download bandwidths between 2 and 1000 Mbps, inclusive.[[13]](#footnote-14)

This sub-sample of the data consisted of 3,495 rates from 506 responses encompassing 141 different providers in 504 different census tracts. The table below presents the number of responses, the number of different service providers, the number of different census tracts, and the number of rates for each technology used for constructing the average rate model.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Technology | Responses | Service Providers | Census Tracts | Rates |
| Cable | 267 | 42 | 267 | 1380 |
| DSL | 152 | 35 | 152 | 1163 |
| FTTH | 67 | 52 | 66 | 240 |
| Fixed Wireless | 152 | 56 | 152 | 712 |
| All | 506 | 141 | 504 | 3495 |

The rates in this sub-sample ranged from $9.95 to $1459.98 with a weighted standard deviation of $35.85. The rates vary widely across technologies. The following table shows the rate range, the weighted rate mean, the weighted rate standard deviation, and the weighted download bandwidth mean for different technologies in this sub-sample.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|   | min | max | Weighted rate mean | Weighted rate standard deviation | Weighted download bandwidth mean |
| Cable | 9.95 | 299.99 | 79.97 | 34.75 | 280.93 |
| DSL | 24.95 | 179.00 | 64.11 | 16.76 | 29.85 |
| Fixed wireless | 11.99 | 1459.98 | 67.33 | 75.04 | 18.62 |
| FTTH | 15.00 | 205.17 | 106.48 | 41.78 | 465.92 |

We undertook a weighted GBM[[14]](#footnote-15) based on the following form:[[15]](#footnote-16)

Average Monthly Rate ($) = Y = f(D , U , A, ST)

where D is download bandwidth in Mbps, U is upload bandwidth in Mbps, and A is the inverse of usage allowance in GB. ST includes 15 stratum groups: Alaska, Alaska TFW, AT&T Services, Inc. (all service download bandwidth), CenturyLink (all service download bandwidth), Charter (all service download bandwidth), Comcast Cable Communications, LLC (all service download bandwidth), Cox (all service download bandwidth), CSC Holdings LLC (all service download bandwidth), Frontier (all service download bandwidth), Major (all service download bandwidth), Minor (all service download bandwidth), Terrestrial Fixed Wireless (all service download bandwidth), Verizon (all service download bandwidth), WideOpenWest (all service download bandwidth), and Windstream (all service download bandwidth). The average monthly rate estimate is a function of D, U, A, and ST.

We estimated the U.S. average monthly rate as:

U.S. Average Monthly Rate ($) = $\sum\_{i=1}^{n}γ\_{i}E(Y | D, U, A, ST=ST\_{i})$

where n = 13, which represents 13 stratum groups in the continental U.S. E(Y| D, U, A, ST = STi) is the expected value conditioned on combinations of download bandwidth, upload bandwidth, and capacity allowance for a given stratum group. The $γ$i is the proportion of total continental U.S. potential subscribers in a given stratum group. As of December 2017, the proportion of total continental U.S. potential subscribers in a given stratum group is listed in the table below.

|  |  |
| --- | --- |
| Stratum Group | $$γ\_{i}$$ |
| AT&T Services, Inc. | 21.08% |
| CenturyLink | 7.32% |
| Charter | 16.60% |
| Comcast Cable Communications, LLC | 20.17% |
| Cox | 3.88% |
| CSC Holdings LLC | 2.14% |
| Frontier | 4.85% |
| Major | 8.71% |
| Minor | 2.19% |
| Terrestrial Fixed Wireless | 0.44% |
| Verizon | 10.19% |
| WideOpenWest | 1.41% |
| Windstream | 0.85% |

The plot below shows how the model fits the raw data. The closer the dots are to the 45 degree line, the better the fit. The size of the circles represents the weights of the sample rates.

 

**U.S. reasonable comparability benchmark**

Under the methodology previously adopted by the Bureau, the reasonable comparability benchmark is the estimated average monthly rate plus twice the standard deviation of rates for terrestrial fixed broadband service plans with download bandwidths of 10 Mbps or greater, upload bandwidths of 1 Mbps or greater, and meeting or exceeding the minimum monthly usage allowance. The root weighted mean squared residual (RWMSR) is an estimate of the standard deviation of rates for service plans meeting the reasonable comparability benchmark criteria.[[16]](#footnote-17)

The 2019 URS broadband average rate model approximates rate per download bandwidth and upload bandwidth closely. Therefore, the RWMSR of rates does not show a trend by download bandwidth and upload bandwidth. For the 2019 URS, we calculate the RWMSR by Alaska and continental U.S. The table below shows the RWMSR by Alaska and continental U.S.

|  |  |
| --- | --- |
|   | RWMSR |
| Continental U.S. | 13.53 |
| Alaska | 24.51 |

The calculation of the U.S. reasonable comparability benchmark is the following:

U.S. reasonable comparability benchmark ($) =

U.S. Average Monthly Rate + 2 (RWMSRContinentalUS) =

U.S. Average Monthly Rate + 27.06

The U.S. average monthly rate estimator is described in the previous section.

**Alaska reasonable comparability benchmark**

For the Alaska reasonable comparability benchmark, the average monthly rate model is defined as follows:

AK Average Monthly Rate ($) = $\sum\_{j=1}^{m}γ\_{j}E(Y | D, U, A, ST=ST\_{j})$

where m = 2, which represents 2 stratum groups in Alaska (Alaska and Alaska TFW). E(Y| D, U, A, ST = STj) is the expected value conditioned on combinations of download bandwidth, upload bandwidth, and capacity allowance for a given stratum group in Alaska. The $γ$j is the proportion of total Alaska potential subscribers in a given stratum group. As of December 2017, the proportion of total Alaska potential subscribers in a given stratum group is listed in the table below.

|  |  |
| --- | --- |
| Stratum Group | $$γ\_{j}$$ |
| Alaska | 99.41% |
| Alaska TFW | 0.59% |

The AK reasonable comparability benchmark is the Alaska average monthly rate plus two RWMSR as the following:

AK reasonable comparability benchmark ($) =

AK Average Monthly Rate + 2 (RWMSRAlaska) =

AK Average Monthly Rate + 49.02

**Reasonable comparability benchmark results**

The table directly below provides examples of reasonable comparability benchmarks (rounded up to the nearest cent) for several service plan levels. The estimates are available for a reasonable comparability benchmark for lower download bandwidths (greater than or equal to 4 Mbps) if needed and up to download bandwidths of 1000 Mbps.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Download Bandwidth (Mbps)** | **Upload Bandwidth (Mbps)** | **Capacity Allowance (GB)** | **2019 U.S.** | **2019 AK** |
| 4 | 1 | 200 | 66.12 | 113.19 |
| 4 | 1 | Unlimited | 70.76 | 119.06 |
| 10 | 1 | 200 | 72.31 | 121.54 |
| 10 | 1 | Unlimited | 77.30 | 127.75 |
| 25 | 3 | 200 | 77.65 | 129.52 |
| 25 | 3 | Unlimited | 82.66 | 135.75 |
| 25 | 5 | 200 | 78.49 | 129.78 |
| 25 | 5 | Unlimited | 83.50 | 136.01 |
| 50 | 5 | Unlimited | 100.85 | 153.64 |
| 100 | 10 | Unlimited | 106.23 | 161.16 |
| 250 | 25 | Unlimited | 128.69 | 203.67 |
| 500 | 50 | Unlimited | 148.35 | 223.87 |
| 1000 | 100 | Unlimited | 162.33 | 232.38 |

**Constraints**

The 2019 reasonable comparability benchmarks for fixed broadband services are similar to the 2017 reasonable comparability benchmarks, which are lower than 2018 reasonable comparability benchmarks. The reasonable comparability benchmark is the estimated average monthly rate plus twice the standard deviation of rates. While the estimated average monthly rate remains relatively unchanged over survey years, the reason for the increase in 2018 and the subsequent decrease in 2019 is mainly because the standard deviation of rates is inconsistent. In other words, the survey data contains differing rate variation over survey years. Wider rate variation on the market, insufficient sample size to capture all variations, and enlarged random noise due to insufficient sample size are main reasons that cause differing rate variation in survey data over years.

We have observed shifting in broadband service plans over time. A provider’s rate for a given service plan may change along with shifting in the service plans that this provider currently offers to the new customers at a given location. For example, a provider may offer a 100/10 service plan to new customers in a city with the same rate as for a 25/5 service plan which was offered previously but is no longer available for new customers. The changes do not happen uniformly, neither across providers nor across geographic areas. Therefore, the national average monthly rate is not heavily influenced, but the rate variation for service plans may increase or decrease substantially over time.

Additionally, the sample size for a given service plan may change dramatically depending on what providers currently offer. The changes in sample size for a given service plan also increase the year-to-year changes in rate variation over survey years. The differing rate variation in data reflects quick and dynamic changes in the consumer market of fixed broadband services and inadequate sample size to capture all possible services that are offered in the market.

Our rate estimates are based on service plans with different combinations of download bandwidth, upload bandwidth, and monthly capacity allowance. The data collected present several difficulties to build such a model.

1. Not all bandwidth tiers have rate samples.
2. Some bandwidth tiers have very few samples.
3. Download bandwidth and upload bandwidth are correlated in practice. For example, it is common that a carrier provides a 10/1 service, but it is rare that a carrier would provide a 1000/1 service. Therefore, we do not have rate samples for services with a high download bandwidth and low upload bandwidth combination.

For these reasons, all models are subject to high risk of overfitting. Underlying data are essential for all model building and validation. In addition, we also have observed wider spread of service plan choices and geographic differentiations that have added and will continue to add more variation in our estimates over time.

The total number of unique service plan choices on the market with different combinations of download bandwidth, upload bandwidth, and monthly capacity allowance is about 360 in both the 2018 and 2019 surveys, though there are more high-bandwidth plans in the 2019 data. Among the 360 combinations, 30 were high bandwidth (download bandwidth greater than or equal to 500 Mbps) service plan choices in the 2018 survey and 40 were high bandwidth service plan choices in the 2019 survey. The rest of the service plan choices remain relatively steady. In other words, the total number of unique service plans available nationwide is fairly steady, though with an increase in the number of high-bandwidth plans. Despite few changes in unique service plan choices on the market, the 2019 URS received about 400 responses of service provider census tract pairs offering high bandwidth service plan choices, while the 2018 URS received only about 260 responses offering high bandwidth service plan choices. This shows that more service providers provided different service plan choices to more customers, particularly for high-bandwidth services. Although more high bandwidth service plan choices have been offered over time, other service plan choices are still available for consumers. We suspect this trend will continue.

We have also observed that the rate varies based on geographic locations because services are not all equally deployed. Therefore, it is reasonable to see a 4/1 service that charges more than a 10/1 service at a different geographic location. Within most strata, it is more likely to see the rate as a systematic function of download bandwidth, upload bandwidth, and monthly capacity allowance. However, this relationship does not hold at a national level because of the geographic differentiations. With the shifting in broadband service plans over time, the URS sample size is no longer sufficient to capture the geographic variation of urban rates.

The URS sample is designed to estimate rates for combinations of download bandwidth, upload bandwidth and capacity allowance, across urban geographies, providers and technologies. Given the variety of offers and rates, the current URS sample design and sample size are not sufficient to capture all variations. Adding additional strata without increasing sample size only exacerbates the issue of having a too-small sample size. The result of that too-small sample is the possibility of having a sample in the URS that does not reflect the overall population of urban rates, or which has an inconsistent rate variation (standard deviation) from year to year.

**APPENDIX A**

**Rates Modified or Excluded for the Analysis**

Respondents stated that 63 service plans were designed for business data users and thus were ineligible for the URS. Providers offering these plans are: Alamo Broadband Inc, American Wireless, Inc., Blue Spring Broadband, Brazos WiFi, Central Cellular LLC/dba Cotc Connections, EPB Fiber Optics, InfoWest Inc, LiteWire Internet Services, Inc., Martineer Wireless, Inc., McMinnville Access Company, Nuvisions, Ranch WiFI LLC, Razzo Link, Inc., Service Electric Cable TV, Inc., Softcom Internet Communications, Inc.

**APPENDIX B**

The 2019 URS modeled rates by download bandwidth and by upload bandwidth. Over this large range of bandwidths, the rates are not linear functions of download bandwidth and upload bandwidth. The size of the circles in the plots below represents the weights of the sample rates. Sampled rates represent common services provided to the customers and do not include all possible combinations of download bandwidth, upload bandwidth, and monthly capacity allowance.

 

**APPENDIX C**

A Generalized Boosted Model (GBM) is a machine learning algorithm that combines regression trees and gradient boosting techniques. The GBM framework does not assume a specific pattern between the independent variables and the dependent variable. It illustrates nonlinearity and interactions well without the need to define complex mathematical equations.

The algorithm first selects a portion of data to “train” a regression tree model (regression tree phase). The regression tree model used in GBM is usually a stump-only model or with only very few branches. Then, it uses the unselected data to “validate” the model and output a user defined performance statistic or loss function (validation phase). The algorithm repeats the same procedure on the residuals from the previous modeling phases until the performance gain stabilizes or loss function optimizes (gradient boosting phase). The outputs of a GBM are model fits from a series of regression tree models. Therefore, conventional coefficients are not applicable. Independent variable collinearity and data outliners have very little impact on the model fit because only the most influential variables are selected during each regression tree phase (only one most influential variable is selected if fitting a stump-only model). The interactions are naturally embedded in the structure of a series of regression tree models. Overfitting is safeguarded by inserting a cross-validation technique. Therefore, the GBM algorithm is considered to have high predictive accuracy. However, its predictive performance is weakened when the relationship between an independent variable and the dependent variable is very linear. More information about GBM can be found in the following references:

Y. Freund and R.E. Schapire. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences. 55(1):119-139.

G. Ridgeway. 1999. The state of boosting. Computing Science and Statistics. 31:172-181.

J.H. Friedman, T. Hastie, and R. Tibshirani. 2000. Additive Logistic Regression: a Statistical View of Boosting. Annals of Statistics. 28(2):337-374.

J.H. Friedman. 2001. Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics. 29(5):1189-1232.

J.H. Friedman. 2002. Stochastic Gradient Boosting. Computational Statistics and Data Analysis. 38(4):367-378.

1. *Connect America Fund*, WC Docket No. 10-90, Order, 28 FCC Rcd 4242 (WCB/WTB 2013). [↑](#footnote-ref-2)
2. The 2018 reasonable comparability benchmarks for fixed broadband services are higher than the 2017 reasonable comparability benchmarks because the 2018 data contain higher rate variation from wider range of fixed broadband services. [↑](#footnote-ref-3)
3. A frame is an inventory that lists all sampling units from which we select our samples. [↑](#footnote-ref-4)
4. The number of potential subscribers for the terrestrial fixed wireless providers is calculated as 2 x number of residential subscribers, assuming that such providers could no more than double their number of existing residential customers within a few months. [↑](#footnote-ref-5)
5. There were 10 high bandwidth strata in last year’s URS: AT&T, CenturyLink, Comcast, Cox, Verizon, WideOpenWest, Windstream, Terrestrial Fixed Wireless, Major, and Minor. [↑](#footnote-ref-6)
6. Bright House Networks, LLC, Time Warner Cable Inc., and Charter Communications, Inc. have merged and operate as Charter Communications, Inc. [↑](#footnote-ref-7)
7. The Major and Minor strata are divided based on the number of potential subscribers, the number of occupied housing units to which the provider offers service, and the Provider Presence Ratio. The algorithm used to divide the sampling units into the Major and Minor strata is “Partitioning Around Medoids.” Partitioning Around Medoids is a type of cluster analysis to identify data clusters based on dissimilarities between clusters. Medoids are the medians for multi-dimensional data. *See* Kaufman, L. and Rousseeuw, P.J. 1990. Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, New York; Park, H.S. and C.H. Jun. 2009. A simple and fast algorithm for K-medoids clustering. Expert Systems with Applications. 36(2):3336–3341. [↑](#footnote-ref-8)
8. The selection probability of a Probability Proportional to Size (PPS) sampling is a function of measure of “size.” Measure of size is the number of potential subscribers from a provider in a given tract. The selection probability is higher for sampling units with higher number of potential subscribers. [↑](#footnote-ref-9)
9. FTTH stands for fiber to the home. [↑](#footnote-ref-10)
10. Such a situation could arise when a provider uses different technologies to provide similar services to customers in different parts of a census tract. [↑](#footnote-ref-11)
11. The 2018 broadband average rate model was built with Generalized Additive Model (GAM), which is also a machine learning method that allows nonlinearity in estimation. However, GAM dramatically overfits the 2019 URS data, which results in uncomfortable negative average rate estimates for download bandwidth, upload bandwidth, and capacity allowance combinations that do not have samples. This is one of the indications that the current URS has inadequate sample size for our purposes. [↑](#footnote-ref-12)
12. Ideally, we would calculate directly the weighted means and the weighted standard deviations of rates for all services. However, our samples do not cover all possible combinations of services provided to consumers nationwide. Therefore, we use a statistical model to estimate rates for all possible services. [↑](#footnote-ref-13)
13. The 2018 broadband average rate model was the first year to include data with download bandwidths between 2 and 1000 Mbps. The 2017 broadband linear regression only models average rate between 2 and 50 Mbps. [↑](#footnote-ref-14)
14. The average rate model based on a weighted GBM for the 2019 URS allows nonlinearity in rate per download bandwidth and rate per upload bandwidth by stratum groups. For further information, see Appendix C. [↑](#footnote-ref-15)
15. We used the R package “gbm: Generalized Boosted Regression Models” to perform model fitting. We used random 50% of data as training set and 50% of data as validation set for each regression tree phase. Multiple GBM models were constructed and compared. Our final model was selected based on best performance of root mean square errors. The optimal number of trees of our final model is 19,576 based on 10-fold cross-validation results. [↑](#footnote-ref-16)
16. RWMSR is the square root of the weighted average of the square of residuals (observed rate minus average rate as defined by the Average Monthly Rate equation) plus the square of the spreads divided by 12. [↑](#footnote-ref-17)