Appendix: Can Improvements to Mobile Internet Service Help Reduce Digital Inequality? An Empirical Analysis of Education and Overall Data Consumption

Karthik Babu Nattamai Kannan (Cox School of Business, Southern Methodist University),
Eric Overby (Scheller College of Business, Georgia Institute of Technology),
Sridhar Narasimhan (Scheller College of Business, Georgia Institute of Technology).

All figures and tables are described in the main text.

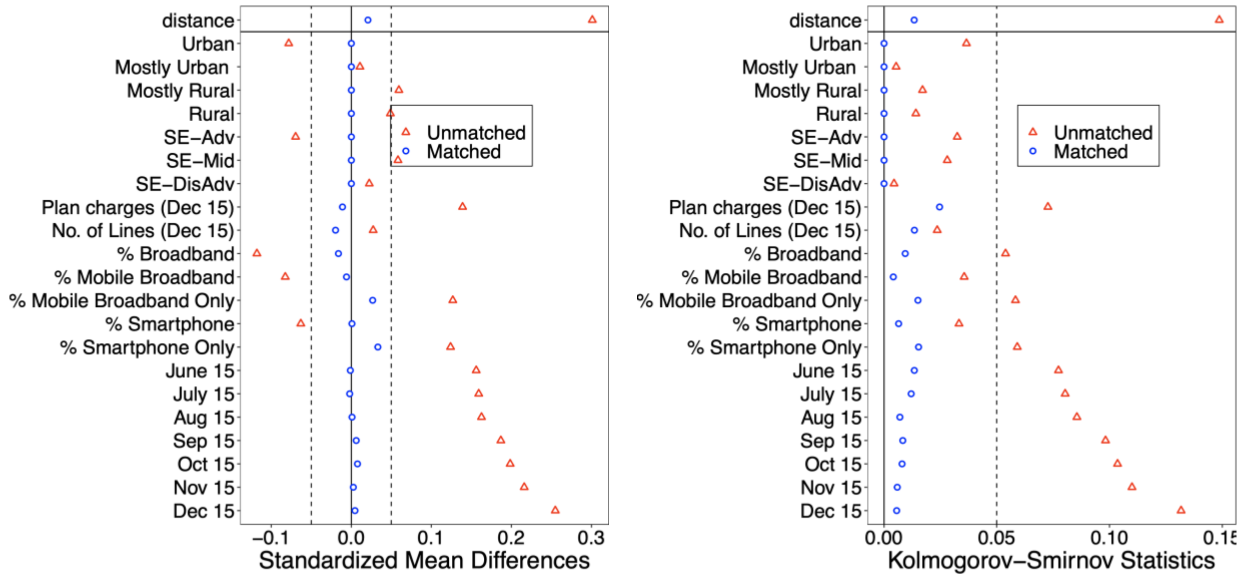
Table A1: Summary statistics and balance measures for the early-16 (treated) and late-16 (control) households in the matched sample for the January cohort

	Mean					
	Late-16 households	Early-16 households	Difference	Standardized Mean Difference	Variance Ratio	Kolmogorov - Smirnov Statistic
$Urban_i$	0.33	0.33	0.00	0.00	1.00	0.00
Mostly $Urban_i$	0.51	0.51	0.00	0.00	1.00	0.00
Mostly $Rural_i$	0.08	0.08	0.00	0.00	1.00	0.00
$Rural_i$	0.08	0.08	0.00	0.00	1.00	0.00
High SES_i	0.32	0.32	0.00	0.00	1.00	0.00
Mid SES_i	0.65	0.65	0.00	0.00	1.00	0.00
Low SES_i	0.03	0.03	0.00	0.00	1.00	0.00
Dec15 Plan $Charges_i$ <=$50/month	0.24	0.25	0.01	0.01	1.00	0.01
Dec15 Plan $Charges_i$ >$50; < $100/month	0.43	0.43	0.01	0.01	1.00	0.01
Dec15 Plan $Charges_i$ >= $100/month	0.33	0.32	-0.02	-0.02	1.00	0.02
Dec15 No. of $Lines_i$	3.68	3.65	-0.03	-0.02	0.98	0.01
% $Broadband_k$	79.09	78.90	-0.19	-0.02	1.03	0.01
% Mobile and Other $Broadband_k$	51.30	51.23	-0.07	-0.01	1.01	0.00
% Mobile Broadband $Only_k$	8.24	8.37	0.13	0.03	1.06	0.02
% Smartphone and Other $Device_k$	73.56	73.57	0.01	0.00	1.01	0.01
% Smartphone $Only_k$	4.09	4.19	0.10	0.03	1.06	0.02
$Data_{it}$ (June 2015)	12.80	12.78	-0.02	0.00	1.04	0.01
$Data_{it}$ (July 2015)	13.92	13.89	-0.03	0.00	1.04	0.01
$Data_{it}$ (Aug 2015)	13.49	13.50	0.01	0.00	1.02	0.01
$Data_{it}$ (Sep 2015)	13.59	13.67	0.08	0.01	1.04	0.01
$Data_{it}$ (Oct 2015)	14.90	15.01	0.10	0.01	1.05	0.01
$Data_{it}$ (Nov 2015)	14.47	14.50	0.03	0.00	1.01	0.01
$Data_{it}$ (Dec 2015)	16.41	16.48	0.07	0.00	1.02	0.01

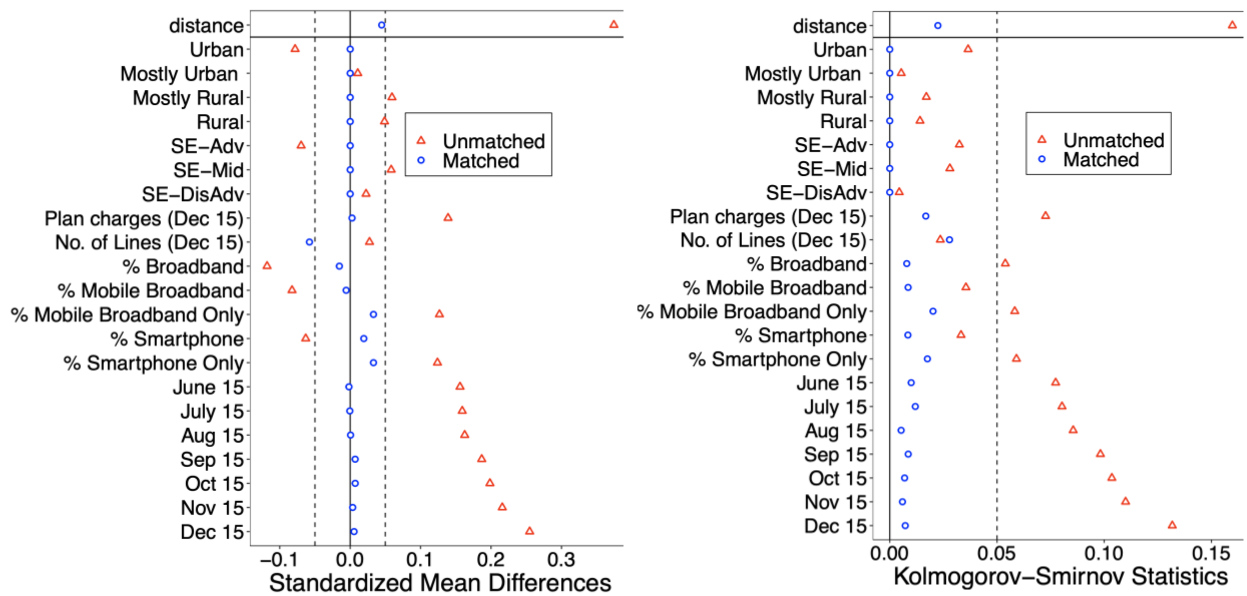
Notes: Variable means and the raw difference in means for the late-16 and early-16 households appear in columns 2-4. The standardized mean difference (SMD, column 5) for each variable is close to 0, indicating good balance for the late-16 and early-16 households. Austin (2009) notes that an SMD of greater than 0.1 (or 10%) reflects significant imbalance. The variance ratio and Kolmogorov-Smirnov statistic for each variable are close to 1 and 0 respectively (columns 6 and 7), indicating similar distributions and good balance. Because our NDA precludes us from reporting actual prices, we divide *Dec15 Plan* $Charges_i$ into three categories: 1) less than or equal to $50/month; 2) greater than $50/month and less than $100/month; and 3) greater than $100/month. $Data_{it}$ is measured in gigabytes.

Figure A1: Covariate balance for households that switched to unlimited mobile data in January, before and after matching. Propensity score matching was done via different link functions. The figures show standardized mean differences (left column) and Kolmogorov−Smirnov Statistics (right column)

Balance after matching via logit link function



Balance after matching via random forest link function



Ho et al. (2011) define covariate balance as “the degree to which the distribution of covariates is similar across levels of the treatment”. In a randomized experiment, covariate balance is achieved automatically in expectation, ensuring there is little systematic difference between the treated and control units. However, in an observational setting like ours, we achieve a similar level of covariate balance through a matching procedure. The objective is for the matched sample to have similar levels of covariate balance as in a randomized control trial. Hence, estimates from a well-balanced matched sample are likely to be less sensitive to model misspecification.

In order to optimally match the samples and assess the quality of resulting matches we use the following measures to check covariate balance as described by Ho et al. (2011). These results are generated using the open-source package MatchIt and summarized in Table A1 and Figure A1.

Standardized mean difference (SMD) is defined as "the difference in the means of each covariate between treatment groups standardized by a standardization factor so that it is on the same scale for all covariates"[1]. Austin (2009) explains that SMD has become a popular measure in assessing balance in matching procedure due to its unique properties (Rosenbaum and Rubin 1985). Because the SMD compares the differences in means in units of the pooled standard deviation, it is not influenced by sample size like t-tests. Also, it allows for the assessment of relative balance between variables measured in different units. Austin suggests that based on the trends in the matching literature, an SMD of greater than 0.1 (or 10%) is considered a significant imbalance among the covariates. In Figure A1 (top) we can observe that all the SMDs of the matched sample are close to zero, indicating a good balance. More importantly, we observe that there is a significant reduction in the imbalance from the unmatched (red triangles) to the matched samples (blue circles). Table A1 shows the means of matched control group, treated group, and the SMDs. The SMDs across all the matched variables are close to zero, indicating a good match.

The **Kolmogorov-Smirnov statistic** is defined as the "maximum difference in the empirical cumulative density functions (eCDFs) of each covariate between the treatment groups"[1]. While the SMD checks if the means between the treatment groups are similar, this statistic measures if the distributions of each covariate across the treatment groups are similar. In Figure A1 (bottom) all the matched covariates have this metric close to zero, indicating an effective match. The Kolmogorov-Smirnov statistics shown in Table A1 are effectively close to 0, indicating that there is very little difference in the distributions for the matched variables between the matched controls and treated group.

Variance Ratio is defined as "the ratio of the variance of a covariate in one group to that in the others" [1]. Austin (2009) argues that comparing variances between the treated and matched control groups can help us in understanding the similarity of the distribution of a continuous covariate between two groups. Variance ratios in Table A1 are effectively close to 1 as the variance for all the matched variables between the matched controls and treated group are nearly the same, indicating a good balance between the distributions of the matched continuous variables.

[1] https://kosukeimai.github.io/MatchIt/articles/assessing-balance.html

Table A2: Regression results for specification (1), using the matched sample obtained via the random forest link function

	DV: $Data_{it}$			DV: $Data\_Edu_{it}$		
	(1)	(2)	(3)	(4)	(5)	(6)
$Unlimited_{it}$	9.88***	8.66***	8.48***	6.05***	5.02***	4.95***
	(0.04)	(0.06)	(0.06)	(0.17)	(0.25)	(0.24)
$Unlimited_{it}$ X Mostly $Urban_i$		1.49***			1.11***	
		(0.08)			(0.33)	
$Unlimited_{it}$ X Mostly $Rural_i$		2.92***			3.17***	
		(0.15)			(0.67)	
$Unlimited_{it}$ X $Rural_i$		3.47***			3.40***	
		(0.15)			(0.66)	
$Unlimited_{it}$ X Mid SES_i			2.05***			1.55***
			(0.07)			(0.31)
$Unlimited_{it}$ X Low SES_i			3.07***			3.82***
			(0.23)			(1.10)
Constant	14.1***	14.2***	14.2***	16.9***	17.1***	17.2***
	(0.09)	(0.09)	(0.09)	(0.48)	(0.48)	(0.49)
n (household-months)	4,254,522	4,254,522	4,254,522	1,751,862	1,751,862	1,751,862
n (households)	250,266	250,266	250,266	250,266	250,266	250,266
Number of months	17	17	17	7	7	7
R2 (w/household fixed effects)	0.72	0.72	0.72	0.53	0.53	0.53

Notes: Sample and regression details are as discussed in the Table 3 notes. Each of the interaction coefficients are statistically different from the others, except for the Mostly Rural and Rural interaction coefficients in column 5 (p=0.79). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A3: Correlation matrix for variables in the analysis sample

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Unlimited$_{it}$	1.00													
2	Income$_{i}$	-0.01***	1.00												
3	% Rural$_{i}$	0.01***	-0.25***	1.00											
4	Dec15 Plan Charges$_{i}$	-0.00	-0.04***	0.05***	1.00										
5	Dec15 No. of Lines$_{i}$	-0.00***	0.06***	0.03***	0.32***	1.00									
6	Dec15 Above_Cap$_{i}$	0.03***	0.03***	-0.01***	-0.36***	-0.07***	1.00								
7	% Broadband$_{k}$	-0.02***	0.42***	-0.48***	-0.07***	0.00***	0.02***	1.00							
8	% Mobile Broadband Only$_{k}$	0.03***	-0.27***	0.43***	0.07***	0.00***	-0.01***	-0.46***	1.00						
9	% Mobile and Other Broadband$_{k}$	-0.02***	0.42***	-0.49***	-0.07***	-0.00***	0.02***	0.89***	-0.54***	1.00					
10	% Smartphone Only$_{k}$	0.03***	-0.35***	0.28***	0.07***	-0.01***	-0.01***	-0.72***	0.60***	-0.69***	1.00				
11	% Smartphone and Other Device$_{k}$	-0.01***	0.42***	-0.48***	-0.07***	-0.00***	0.02***	0.91***	-0.44***	0.87***	-0.68***	1.00			
12	% Bachelors Degree or higher$_{k}$	-0.02***	0.43***	-0.41***	-0.08***	-0.02***	0.02***	0.72***	-0.52***	0.77***	-0.66***	0.74***	1.00		
13	Estimated No. of Children$_{i}$	0.00***	0.08***	0.03***	0.07***	0.15***	-0.01***	0.02***	-0.01***	0.01***	-0.03***	0.01***	0.00	1.00	
14	Data$_{it}$ (GB)	0.41***	-0.04***	0.07***	0.36***	0.28***	-0.00***	-0.09***	0.09***	-0.09***	0.09***	-0.07***	-0.10***	0.08***	1.00
15	Data_Edu$_{it}$ (MB)	0.06***	-0.04***	0.01***	0.02***	0.01***	-0.00***	-0.03***	0.02***	-0.03***	0.03***	-0.03***	-0.04***	0.02***	0.09***

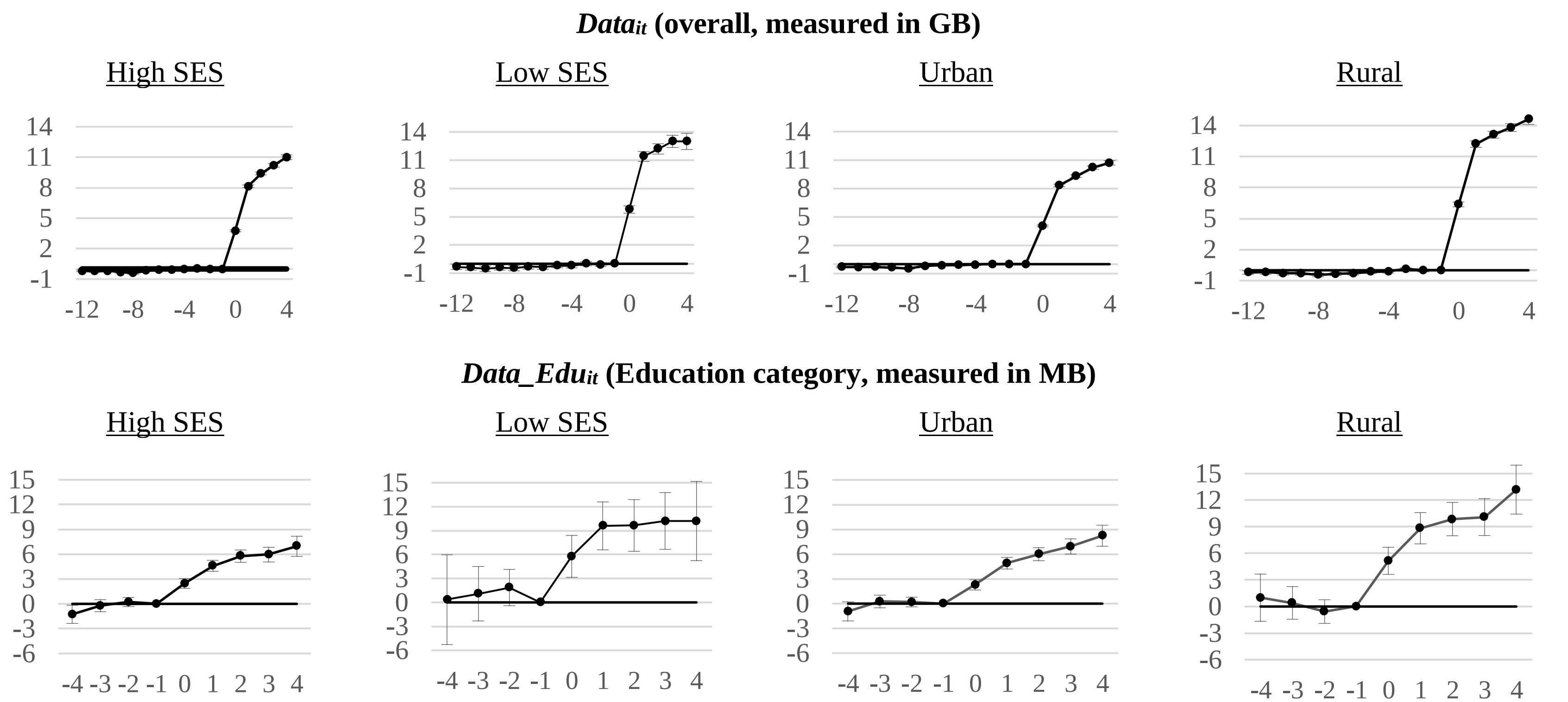
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table A4: Regression results for specification (1), with alternate fixed effects specifications

	DV: $Data_{it}$			DV: $Data\_Edu_{it}$			DV: $Data_{it}$			DV: $Data\_Edu_{it}$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Unlimited$_{it}$	9.28***	7.89***	7.79***	5.41***	4.53***	4.23***	9.82***	8.55***	8.39***	6.14***	4.99***	4.95***
	(0.04)	(0.06)	(0.06)	(0.15)	(0.23)	(0.22)	(0.04)	(0.06)	(0.06)	(0.17)	(0.27)	(0.26)
Unlimited$_{it}$ X Mostly Urban$_i$		1.70***			1.06***			1.57***			1.28***	
		(0.08)			(0.30)			(0.08)			(0.37)	
Unlimited$_{it}$ X Mostly Rural$_i$		3.10***			2.40***			2.96***			3.34***	
		(0.15)			(0.59)			(0.16)			(0.73)	
Unlimited$_{it}$ X Rural$_i$		3.91***			2.31***			3.52***			3.62***	
		(0.15)			(0.61)			(0.15)			(0.73)	
Unlimited$_{it}$ X Mid SES$_i$			2.18***			1.70***			2.09***			1.71***
			(0.07)			(0.28)			(0.08)			(0.34)
Unlimited$_{it}$ X Low SES$_i$			3.27***			3.15***			3.09***			3.13***
			(0.24)			(1.00)			(0.25)			(1.20)
Constant	9.09***	9.09***	9.09***	12.9***	12.9***	12.9***	17.5***	17.5***	17.5***	17.7***	17.6***	17.6***
	(0.01)	(0.01)	(0.01)	(0.08)	(0.08)	(0.08)	(0.02)	(0.02)	(0.02)	(0.11)	(0.11)	(0.11)
n (household-months)	4,265,368	4,265,368	4,265,368	1,756,328	1,756,328	1,756,328	4,265,368	4,265,368	4,265,368	1,756,328	1,756,328	1,756,328
n (households)	250,904	250,904	250,904	250,904	250,904	250,904	250,904	250,904	250,904	250,904	250,904	250,904
Number of months	17	17	17	7	7	7	17	17	17	7	7	7
Month FE	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N
Cohort-Geosocial-Month FE	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y
R2 (w/household fixed effects)	0.72	0.72	0.72	0.53	0.53	0.53	0.72	0.72	0.72	0.53	0.53	0.53

Notes: Columns 1-6 include month fixed effects without interactions with cohort or geosocial group. Columns 7-12 include three-way interactions between the month fixed effects, cohort indicators, and geosocial group indicators. Sample and regression details are as discussed in the Table 3 notes.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure A2: Coefficients and 95% confidence intervals for the lead and lag terms from regressions for specific socioeconomic status (SES) and geographic groups



Notes: We withhold the −1 lead dummy to avoid the dummy variable trap, thereby using −1 as the “baseline”. The confidence intervals for the $Data_{it}$ analysis are narrow, so much so that the error bars that represent them are barely visible in the figure. Given the January 2015 to May 2016 analysis period for the $Data_{it}$ analysis, the earliest possible lead dummy is −14 and the latest lag dummy is +4. This is because we can go back 14 months for households that switched to unlimited data in March 2016 and forward 4 months for households that switched in January 2016. We aggregate all leads 12 or more months before switching into the −12 lead. For the $Data\_Edu_{it}$ analysis, the analysis period is November 2015 to May 2016, yielding leads/lags from −4 to +4.

Table A5: Regression results for data per line for subset of households with 1 or more children

	DV: $DataPerLine_{it}$ (GB)		DV: $DataPerLine\_Edu_{it}$ (MB)	
	(1)	(2)	(3)	(4)
$Unlimited_{it}$	3.12 (0.02)***	2.93 (0.03)***	2.00 (0.09)***	1.78 (0.14)***
$Unlimited_{it}$ X Estimated No. of $Children_i$		0.07 (0.01)***		0.08 (0.04)***
Constant	4.13 (0.07)***	4.12 (0.07)***	4.93 (0.30)***	4.93 (0.30)***
n (household-months)	1,987,164	1,987,164	818,244	818,244
R^2 (w/household fixed effects)	0.69	0.69	0.54	0.54

Notes: Sample and regression details are as discussed in the Table 3 notes. *** p < 0.01, ** p < 0.05, * p < 0.10.

Data cap utilization prior to switching to the unlimited plan: To explore selection bias concerns, we leveraged the *Dec15 Above* Cap_i variable, which indicates whether household *i* was below or above its data cap in December 2015, just prior to the implementation of the unlimited plan. Consider two groups of households that switched to the unlimited mobile data plan: “below cap” switchers and “above cap” switchers. “Below cap” switchers may switch because they anticipate that they will use more data in the future than their existing cap allows. If that is the case, then our estimate of the increase in data consumption after switching might be biased upward for these switchers, because we would expect some of the increase to have happened anyway. By contrast, because “above cap” switchers are already exceeding their data caps, our estimate of the increase for them is less likely to be biased in this way.

Columns 1, 2, 4, and 5 of Table A6 show the results after splitting the sample into “above cap” households and “below cap” households – each with their matched controls – based on the *Dec15 Above* Cap_i variable. In Columns 3 and 6, we used the full sample and included an $Above\_Cap_i$ interaction term, which is negative and significant but small in magnitude. The similarity in the results across groups suggests that any upward bias due to this selection concern is likely to be small.

Table A6: Regression results based on data cap utilization in December 2015

	DV: $Data_{it}$ (GB)			DV: $Data\_Edu_{it}$ (MB)		
	Below cap	Above cap	Full sample	Below cap	Above cap	Full sample
$Unlimited_{it}$	9.83***	9.76***	9.84***	6.22***	5.24***	6.22***
	(0.04)	(0.13)	(0.04)	(0.18)	(0.57)	(0.18)
$Unlimited_{it}$ X $Above\_Cap_i$			-0.21*			-0.94**
			(0.12)			(0.45)
Constant	14.1***	13.8***	14.1***	16.9***	16.7***	16.9***
	(0.09)	(0.34)	(0.09)	(0.52)	(1.08)	(0.49)
n (household-months)	3,942,096	322,949	4,265,045	1,623,216	132,979	1,756,195
Adjusted R^2	0.74	0.71	0.73	0.57	0.59	0.57

Notes: “Below cap” and “Above cap” represent sub-samples defined by whether the switching household exceeded its data cap in December 2015. We omitted 19 households from this analysis because our data do not report their December 2015 data cap. Other sample and regression details are as discussed in the Table 3 notes. *** p < 0.01, ** p < 0.05, * p < 0.10.

Growth in data consumption prior to switching to the unlimited plan: Another related possibility is that households that were already trending upward in their data consumption were the first to switch to the unlimited plan. Essentially, it is possible that these households switch to the unlimited mobile data plan because of growing data consumption, rather than having their data consumption grow because they switched to the unlimited plan. Although our matching process accounts for household-level data use patterns prior to switching to the unlimited plan, we conducted the following analysis to consider this further.

For each household *i*, we calculate the growth in data consumption from September and October 2015 to November and December 2015. In other words, we compute $Data\ Change_i = (Nov\ Data_i + Dec\ Data_i) - (Sept\ Data_i + Oct\ Data_i)$. We divide households into four quartiles based on this variable. Quartile 1 contains households with the smallest (including negative) growth in data consumption; quartile 4 contains households with the largest growth. We posit that the quartile 4 households are those most likely to switch to the unlimited plan based on growing data consumption. If each quartile experiences similar increases in data consumption after switching to unlimited data, then this would lessen our concerns about selection bias.

Table A7 shows the regression results after splitting the sample into the four quartiles. The increases in data use are large and significant for each quartile, indicating that our results hold regardless of the growth trajectory of a household's data use prior to switching. The coefficients with the largest absolute values are those for quartile 4. This suggests that our main results may be biased upward by a selection effect attributable to these households. However, the *percentage* increases in data use – based on comparing the $Unlimited_{it}$ coefficient to the constant – for quartile 4 are smaller than those for quartiles 2 and 3, which suggests that any selection bias is not particularly severe.

Table A7: Regression results, by quartiles based on growth in data consumption prior to switching

	DV: $Data_{it}$ (GB)				DV: $Data\_Edu_{it}$ (MB)			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
$Unlimited_{it}$	9.66***	8.50***	9.67***	11.4***	3.57***	6.34***	6.97***	7.66***
	(0.07)	(0.07)	(0.07)	(0.08)	(0.39)	(0.27)	(0.28)	(0.39)
Constant	14.5***	11.1***	13.3***	17.5***	25.5***	11.0***	10.4***	20.7***
	(0.20)	(0.17)	(0.18)	(0.18)	(1.18)	(0.62)	(0.93)	(1.22)
n (household-months)	1,066,376	1,066,376	1,066,308	1,066,308	439,096	439,096	439,068	439,068
Adjusted R^2	0.71	0.73	0.73	0.71	0.58	0.53	0.53	0.58

Note: "Q1", "Q2", etc. represent sub-samples defined by the quartile of a switching household in terms of its growth in data consumption from September to December 2015. Other sample and regression details are as discussed in the Table 3 notes. *** p < 0.01, ** p < 0.05, * p < 0.10.

Random implementation test: The standard errors in difference-in-differences estimation can be inconsistent due to serial correlation in the dependent variable. Following Bertrand et al. (2004) and Burtch et al. (2018), we implement a random implementation test in which we shuffle the treatment indicators – thereby randomly reassigning treatment to households – and then re-estimate the model. We store the coefficient from the randomly-assigned variable and repeat this 1,000 times.

Results are shown in Table A8. The mean of the coefficients (i.e., μ of the β's) from the random shuffle tests is close to zero. This indicates that our main result is not driven by serial correlation in the dependent variable. We also approximated the distribution of the coefficients from the random shuffle tests using their mean and standard deviation (σ). We then assessed the likelihood that the coefficient from the main model could come from that distribution by computing its Z-score. As shown in Table A8, there is very little chance (p<0.00001) that our main results could have arisen by chance in this way.

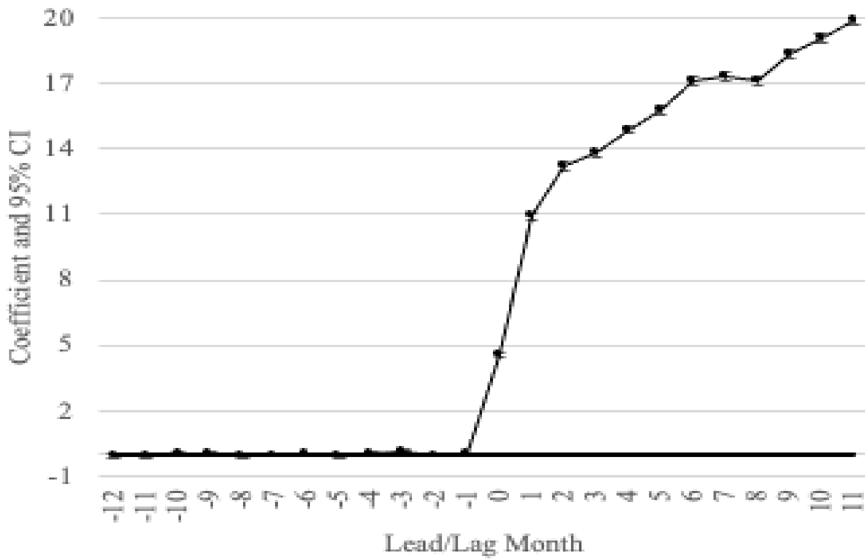
Table A8: Random implementation test

Estimation	DV: $Data_{it}$ (GB)	DV: $Data\_Edu_{it}$ (MB)
	(1)	(2)
μ of β's	-0.00	-0.00
σ of β's	0.01	0.07
Replications	1,000	1,000
Estimated β from model 1	9.82	6.14
Z-score	1054.76	90.77
p-value	p < 0.00001	p < 0.00001

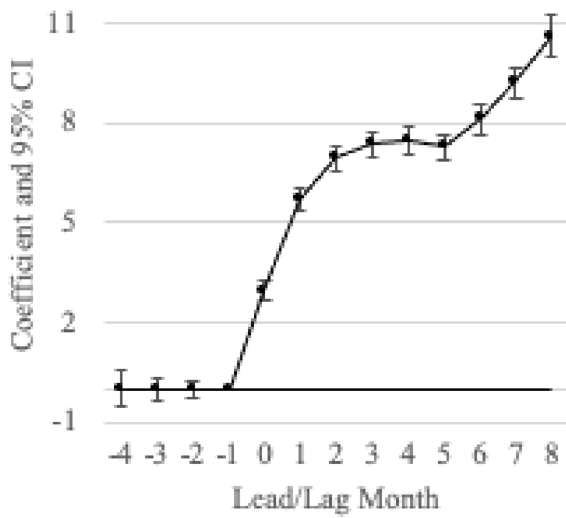
Notes: The β's in this test are estimated after randomly shuffling the treatment indicators across households. The Z-score is for the β coefficient from specification 1 (see Table 3), if it came from the distribution of the β's resulting from the random implementation test. This is highly unlikely, as reflected by the p-value.

Figure A3: Switchers matched with non-switchers: Coefficients and 95% confidence intervals for the lead and lag terms

DV: $Data_{it}$



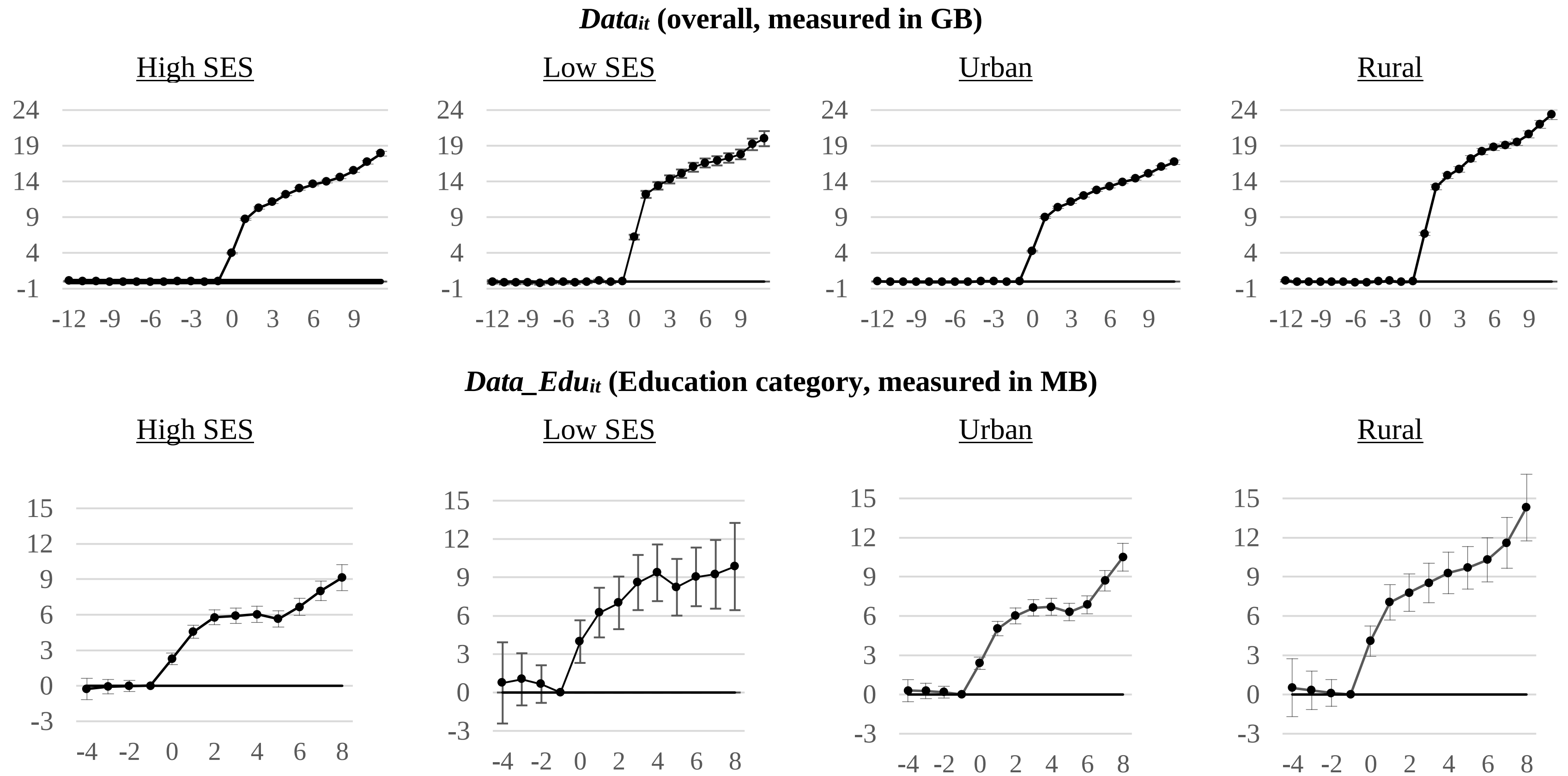
DV: $Data\_Edu_{it}$



Y Axis: Coefficient and 95% CI. X Axis: Lead/Lag Month

Notes: We withhold the −1 lead dummy to avoid the dummy variable trap, thereby using −1 as the "baseline". The confidence intervals for the $Data_{it}$ analysis are narrow, so much so that the error bars that represent them are barely visible in the figure. Given the January 2015 to December 2016 analysis period for the $Data_{it}$ analysis (note that we are using the entire sample period in this analysis), the earliest possible lead dummy is −14 and the latest lag dummy is +11. This is because we can go back 14 months for households that switched to unlimited data in March 2016 and forward 11 months for households that switched in January 2016. We aggregate all leads 12 or more months before switching into the −12 lead. For the $Data\_Edu_{it}$ analysis, the analysis period is November 2015 to September 2016, yielding leads/lags from −4 to +8.

Figure A4: Switchers matched with non-switchers: Coefficients and 95% confidence intervals for the lead and lag terms from regressions for specific socioeconomic status (SES) and geographic groups



Y Axis: Coefficient and 95% CI. X Axis: Lead/Lag Month

Notes: We withhold the −1 lead dummy to avoid the dummy variable trap, thereby using −1 as the "baseline". The confidence intervals for the $Data_{it}$ analysis are narrow, so much so that the error bars that represent them are barely visible in the figure. Given the January 2015 to December 2016 analysis period for the $Data_{it}$ analysis (note that we are using the entire sample period in this analysis), the earliest possible lead dummy is −14 and the latest lag dummy is +11. This is because we can go back 14 months for households that switched to unlimited data in March 2016 and forward 11 months for households that switched in January 2016. We aggregate all leads 12 or more months before switching into the −12 lead. For the $Data\_Edu_{it}$ analysis, the analysis period is November 2015 to September 2016, yielding leads/lags from −4 to +8.

Table A9: Switchers matched with non-switchers: Regression results

	DV: $Data_{it}$			DV: $Data\_Edu_{it}$		
	(1)	(2)	(3)	(4)	(5)	(6)
$Unlimited_{it}$	15.1***	13.2***	8.48***	7.12***	6.18***	6.17***
	(0.05)	(0.08)	(0.06)	(0.13)	(0.20)	(0.20)
$Unlimited_{it}$ X Mostly $Urban_{i}$		2.33***			1.15***	
		(0.10)			(0.26)	
$Unlimited_{it}$ X Mostly $Rural_{i}$		4.08***			2.41***	
		(0.20)			(0.52)	
$Unlimited_{it}$ X $Rural_{i}$		4.70***			2.50***	
		(0.20)			(0.53)	
$Unlimited_{it}$ X Mid SES_{i}			2.05***			1.34***
			(0.07)			(0.25)
$Unlimited_{it}$ X Low SES_{i}			3.07***			1.57**
			(0.23)			(0.70)
Constant	14.5***	14.7***	14.2***	16.1***	16.2***	16.1***
	(0.11)	(0.11)	(0.09)	(0.33)	(0.33)	(0.33)
n (household-months)	7,400,736	7,400,736	7,400,736	3,392,004	3,392,004	3,392,004
n (households)	308,364	308,364	308,364	308,364	308,364	308,364
Number of months	24	24	24	11	11	11
R^2 (w/household fixed effects)	0.67	0.67	0.67	0.50	0.50	0.50

Notes: Sample and regression details are as discussed in the Table 3 notes. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A10: Regression results using more granular income and geography variables

	DV: $Data_{it}$ (GB)		DV: $Data\_Edu_{it}$ (MB)	
Model	(1)	(2)	(3)	(4)
$Unlimited_{it}$	9.08***	8.42***	5.55***	5.11***
	(0.04)	(0.06)	(0.19)	(0.24)
$Unlimited_{it}$ * % $Rural_{i}$	0.04***		0.03***	
	(0.00)		(0.01)	
$Unlimited_{it}$ * Income $75,00-$124,999		1.46***		1.02**
		(0.09)		(0.36)
$Unlimited_{it}$ * Income $40,000-$74,999		2.33***		1.59***
		(0.09)		(0.38)
$Unlimited_{it}$ * Income $20,000-$39,999		2.97***		2.40***
		(0.13)		(0.54)
$Unlimited_{it}$ * Income Under $19,999		3.10***		2.89**
		(0.23)		(1.09)
Constant	14.4***	14.2***	17.1***	17.0***
	(0.09)	(0.09)	(0.49)	(0.50)
n (household-months)	4,265,368	4,265,368	1,756,328	1,756,328
n (households)	250,904	250,904	250,904	250,904
Number of months	17	17	7	7
R^2 (w/household fixed effects)	0.72	0.72	0.53	0.53

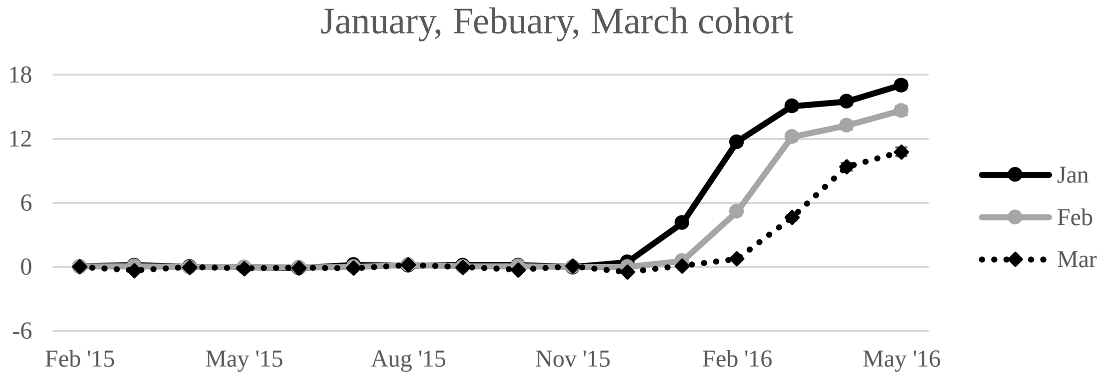
Notes: Sample and regression details are as discussed in the Table 3 notes. For models 2 and 4, the baseline reference is the group with income above $125,000. *** p < 0.01, ** p < 0.05, * p < 0.10

Table A11: Regression results for only the January cohort and their matched controls

	DV: $Data_{it}$			DV: $Data\_Edu_{it}$		
	(1)	(2)	(3)	(4)	(5)	(6)
$Unlimited_{it}$	10.5***	9.14***	9.08***	6.82***	5.56***	5.15***
	(0.06)	(0.10)	(0.10)	(0.28)	(0.46)	(0.43)
$Unlimited_{it}$ X Mostly $Urban_i$		1.63***			1.28**	
		(0.13)			(0.61)	
$Unlimited_{it}$ X Mostly $Rural_i$		3.26***			4.12***	
		(0.24)			(1.16)	
$Unlimited_{it}$ X $Rural_i$		3.65***			3.35***	
		(0.24)			(1.13)	
$Unlimited_{it}$ X Mid SES_i			2.12***			2.31***
			(0.12)			(0.56)
$Unlimited_{it}$ X Low SES_i			3.19***			5.51***
			(0.39)			(1.94)
Constant	19.8***	19.6***	19.6***	17.9***	17.7***	17.6***
	(0.08)	(0.07)	(0.07)	(0.43)	(0.44)	(0.46)
n (household-months)	1,646,484	1,646,484	1,646,484	677,964	677,964	677,964
n (households)	96,852	96,852	96,852	96,852	96,852	96,852
Number of months	17	17	17	7	7	7
R^2 (w/household fixed effects)	0.73	0.73	0.73	0.53	0.53	0.53

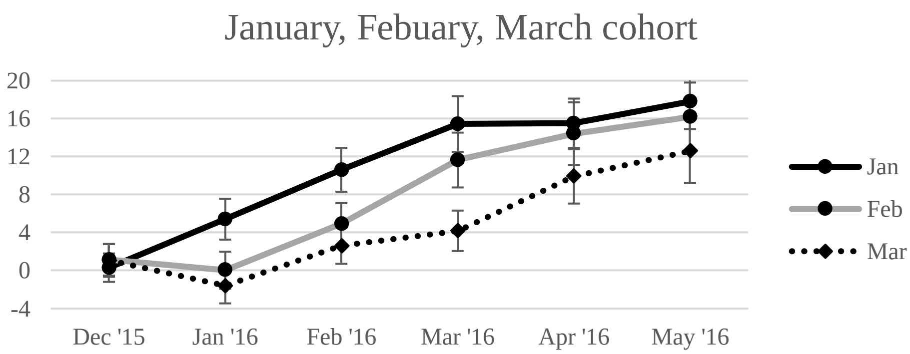
Notes: Sample and regression details are as discussed in the Table 3 notes. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure A5: Lead/lag coefficients for the $Data_{it}$ regression for the January, February, and March cohorts, as estimated via the Callaway and Sant'Anna approach



Notes: 95% confidence intervals are shown in the figure. However, they are narrow, so much so that the error bars that represent them are barely visible in the figure.

Figure A6: Lead/lag coefficients for the $Data\_Edu_{it}$ regression for the January, February, and March cohorts, as estimated via the Callaway and Sant'Anna approach



Notes: 95% confidence intervals are shown in the figure.

Table A12: Regression results for the subset of households that remained in the unlimited plan after switching

	DV: $Data_{it}$			DV: $Data\_Edu_{it}$		
	(1)	(2)	(3)	(4)	(5)	(6)
$Unlimited_{it}$	10.0***	8.78***	8.59***	6.54***	5.54***	5.28***
	(0.04)	(0.06)	(0.06)	(0.18)	(0.28)	(0.26)
$Unlimited_{it}$ X Mostly $Urban_i$		1.52***			1.01***	
		(0.08)			(0.35)	
$Unlimited_{it}$ X Mostly $Rural_i$		2.97***			3.12***	
		(0.15)			(0.70)	
$Unlimited_{it}$ X $Rural_i$		3.57***			3.56***	
		(0.15)			(0.70)	
$Unlimited_{it}$ X Mid SES_i			2.15***			1.82***
			(0.08)			(0.33)
$Unlimited_{it}$ X Low SES_i			3.29***			3.70***
			(0.25)			(1.17)
Constant	14.2***	14.4***	14.4***	16.8***	17.1***	17.1***
	(0.10)	(0.09)	(0.09)	(0.51)	(0.52)	(0.52)
n (household-months)	3,915,304	3,915,304	3,915,304	1,612,184	1,612,184	1,612,184
n (households)	230,312	230,312	230,312	230,312	230,312	230,312
Number of months	17	17	17	7	7	7
R^2 (w/household fixed effects)	0.72	0.72	0.72	0.53	0.53	0.53

Notes: Sample and regression details are as discussed in the Table 3 notes. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A13: Regression results with standard errors clustered by 3 digit ZIP code

	DV: *$Data_{it}$* (GB)			DV: *$Data\_Edu_{it}$* (MB)		
	(1)	(2)	(3)	(4)	(5)	(6)
$Unlimited_{it}$	9.82^{***}	8.59^{***}	8.42^{***}	6.14^{***}	5.13^{***}	5.11^{***}
	(0.10)	(0.12)	(0.12)	(0.20)	(0.26)	(0.27)
$Unlimited_{it}$ X Mostly $Urban_i$		1.50^{***}			1.08^{***}	
		(0.14)			(0.32)	
$Unlimited_{it}$ X Mostly $Rural_i$		2.90^{***}			3.15^{***}	
		(0.19)			(0.62)	
$Unlimited_{it}$ X $Rural_i$		3.49^{***}			3.36^{***}	
		(0.17)			(0.67)	
$Unlimited_{it}$ X Mid SES_i			2.04^{***}			1.47^{***}
			(0.10)			(0.30)
$Unlimited_{it}$ X Low SES_i			3.10^{***}			2.89^{**}
			(0.26)			(0.99)
Constant	14.1^{***}	14.3^{***}	14.2^{***}	16.9^{***}	17.1^{***}	17.0^{***}
	(0.09)	(0.08)	(0.08)	(0.51)	(0.51)	(0.52)
n (household-months)	4,265,368	4,265,368	4,265,368	1,756,328	1,756,328	1,756,328
R^2 (w/household fixed effects)	0.72	0.72	0.72	0.53	0.53	0.53

Notes: Sample and regression details are as discussed in the Table 3 notes. Clustered standard errors by 3-digit ZIP codes shown in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.10.

Table A14: Regression results for inverse hyperbolic sine (IHS) transformed dependent variables

	DV: $Data_{it}$			DV: $Data\_Edu_{it}$		
	(1)	(2)	(3)	(4)	(5)	(6)
$Unlimited_{it}$	0.42***	0.38***	0.37***	0.29***	0.26***	0.26***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
$Unlimited_{it}$ X Mostly $Urban_i$		0.05***			0.03***	
		(0.00)			(0.01)	
$Unlimited_{it}$ X Mostly $Rural_i$		0.11***			0.08***	
		(0.01)			(0.02)	
$Unlimited_{it}$ X $Rural_i$		0.12***			0.10***	
		(0.01)			(0.02)	
$Unlimited_{it}$ X Mid SES_i			0.082***			0.04***
			(0.00)			(0.01)
$Unlimited_{it}$ X Low SES_i			0.16***			0.06**
			(0.01)			(0.03)
Constant	3.05***	3.06***	3.06***	1.52***	1.53***	1.52***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
n (household-months)	4,265,368	4,265,368	4,265,368	1,756,328	1,756,328	1,756,328
n (households)	250,904	250,904	250,904	250,904	250,904	250,904
Number of months	17	17	17	7	7	7
R^2 (w/household fixed effects)	0.80	0. 80	0. 80	0.46	0.46	0.46

Notes: Sample and regression details are as discussed in the Table 3 notes. *** p < 0.01, ** p < 0.05, * p < 0.10.

Table A15: Regression results for overall data consumption based on whether the household had below or above median data consumption in December 2015, before the unlimited plan

	DV: $Data_{it}$			DV: $Data_{it}$		
	Below median pre-treatment data consumption			Above median pre-treatment data consumption		
	(1)	(2)	(3)	(4)	(5)	(6)
$Unlimited_{it}$	7.26***	6.28***	6.07***	12.2***	10.9***	10.6***
	(0.05)	(0.07)	(0.07)	(0.05)	(0.09)	(0.08)
$Unlimited_{it}$ X Mostly $Urban_i$		1.22***			1.46***	
		(0.09)			(0.11)	
$Unlimited_{it}$ X Mostly $Rural_i$		2.40***			2.90***	
		(0.19)			(0.21)	
$Unlimited_{it}$ X $Rural_i$		3.02***			3.16***	
		(0.20)			(0.20)	
$Unlimited_{it}$ X Mid SES_i			1.71***			2.24***
			(0.09)			(0.10)
$Unlimited_{it}$ X Low SES_i			3.32***			3.14***
			(0.31)			(0.34)
Constant	9.32***	9.43***	9.43***	19.1***	19.3***	19.3***
	(0.09)	(0.09)	(0.08)	(0.16)	(0.16)	(0.16)
n (household-months)	2,132,684	2,132,684	2,132,684	2,132,684	2,132,684	2,132,684
n (households)	125,452	125,452	125,452	125,452	125,452	125,452
Number of months	17	17	17	17	17	17
R^2 (w/household fixed effects)	0.55	0.55	0.55	0.68	0.68	0.68

Notes: Sample and regression details are as discussed in the Table 3 notes. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A16: Regression results for education data consumption based on whether the household had below or above median data consumption in December 2015, before the unlimited plan

	DV: *Data_Edu*$_{it}$			DV: *Data_Edu*$_{it}$		
	Below median pre-treatment data consumption			Above median pre-treatment data consumption		
	(1)	(2)	(3)	(4)	(5)	(6)
Unlimited$_{it}$	4.59***	3.97***	3.71***	7.72***	6.37***	6.32***
	(0.18)	(0.26)	(0.26)	(0.29)	(0.45)	(0.39)
Unlimited$_{it}$ X Mostly Urban$_i$		0.59*			1.47***	
		(0.34)			(0.56)	
Unlimited$_{it}$ X Mostly Rural$_i$		1.64**			4.55***	
		(0.70)			(1.12)	
Unlimited$_{it}$ X Rural$_i$		2.88***			3.82***	
		(0.76)			(1.08)	
Unlimited$_{it}$ X Mid SES$_i$			1.21***			2.09***
			(0.33)			(0.51)
Unlimited$_{it}$ X Low SES$_i$			2.07**			4.64**
			(1.04)			(2.14)
Constant	8.62***	8.83***	8.71***	25.5***	25.7***	25.8***
	(0.49)	(0.49)	(0.48)	(0.96)	(0.96)	(0.99)
n (household-months)	878,164	878,164	878,164	878,164	878,164	878,164
n (households)	125,452	125,452	125,452	125,452	125,452	125,452
Number of months	7	17	17	7	7	7
R^2 (w/household fixed effects)	0.28	0.28	0.28	0.58	0.58	0.58

Notes: Sample and regression details are as discussed in the Table 3 notes. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

References

Austin PC (2009) Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med* 28(25):3083–3107.

Bertrand M, Duflo E, Mullainathan S (2004) How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics* 119(1):249–275.

Burtch G, Carnahan S, Greenwood BN (2018) Can you gig It? An empirical examination of the gig economy and entrepreneurial activity. *Management Science* 64(12):5497–5520.

Ho DE, Imai K, King G, Stuart EA (2011) MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software* 42(8):1–28.

Rosenbaum PR, Rubin DB (1985) Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *The American Statistician* 39(1):33–38.