Online Appendix

Appendix A Construction of trip characteristics groups

In order to size the origin rectangles to have similar numbers of trips, we first divide the sample into two equally sized groups by origin latitude. Each group is then subdivided into two equally sized groups by origin longitude. Then we divide each group again by latitude, and repeat this process 6 times.[24] We follow an analogous process to divide the sample into 128 groups by destination coordinates.

The 15 hour-of-the-week intervals are: 7:00-9:00 am Mon-Fri, 9:00-11:00 am Mon-Fri, 11:00 am-1:00 pm Mon-Fri, 1:00-4:00 pm Mon-Fri, 4:00-6:00 pm Mon-Thu, 6:00-8:00 pm Mon-Thu, 8:00-10:00 pm Mon-Thu, 10:00 pm-1:00 am Mon-Thu, 4:00-8:00 pm Fri, 8:00 pm-midnight Fri-Sat, midnight-4:00 am Sat-Sun, 9:00 am-2:00 pm Sat-Sun, 2:00 pm-8:00 pm Sat, 2:00 pm-8:00 pm Sun, and all remaining times.

Figure 8 shows a histogram of the number of trips in the group each trip in our sample belongs to. We can see that the majority of trips are in groups with more than five trips.

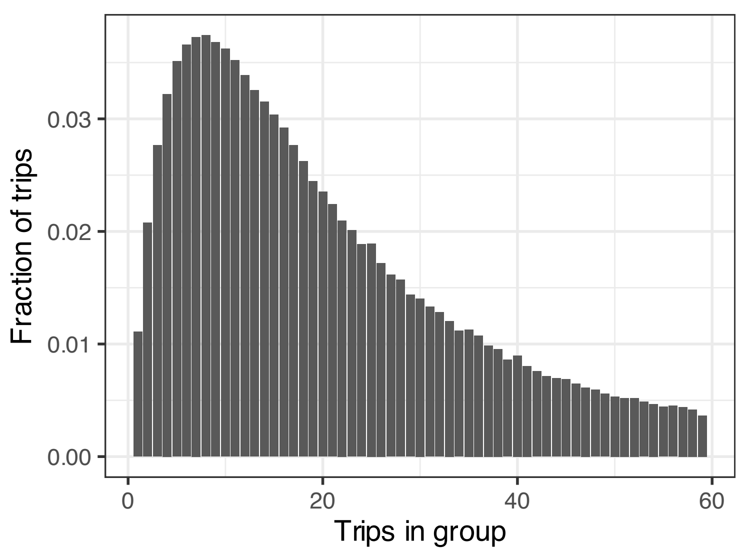

Figure 8: Distribution of trip group size

Note: We split the sample into groups by the origin and destination coordinates and by hour of the week. This figure shows the distribution of the number of trips in each group, weighted by number of trips.

[24]We are grateful to John Horton for suggesting this procedure.

Table 12: Selection of driving metrics

Metric	Penalty
Cont. speed 70	0.00001
Cont. speed mean	0.00003
Cont. speed 40	0.00003
Cont. speed 50	0.00004
Cont. speed 20	0.00004
Accelerations 3 m/s^2	0.0001
Brakes 3 m/s^2	0.0002
Cont. speed 100	0.0002
Cont. speed 30	0.0002
Accelerations 2.5 m/s^2	0.0003
Mounted	0.0008
Cont. speed 0	0.0012
Cont. speed 90	0.0012
Cont. speed 60	0.0013
Cont. speed 10	0.0014
Brakes 2.5 m/s^2	0.0023
Handling	0.0030
Brakes 2 m/s^2	0.0033
Accelerations 2 m/s^2	0.0033
Avg. speed when moving	0.0037
Cont. speed 80	0.0048
Dropoff gap	0.0093
Pickup gap	0.0093
Excess distance	0.0112
Excess duration	0.0178

Note: We fit a lasso regression of trip rating including all candidate metrics. The table shows the level of the penalty at which each variable is dropped.

Appendix B Variable selection and score construction

B.1 Selecting driving metrics

In order to select our main variables, we run a lasso regression of trip rating that includes all candidate metrics as well as their squares. We also include driver and trip characteristics fixed effects without penalization. The candidate metrics include metrics for brakes and accelerations using thresholds of 2, 2.5 and 3.06 m/s^2 (the industry standard is 3.06 m/s^2), and metrics for 12 different moments of the distribution of contextualized speeds within each trip (percentiles 0, 10, 20, ..., 100, as well as the mean).

Table 12 shows the order in which variables are dropped as we start increasing the penalty, and the value for the penalty at which they are dropped. Distance and duration are the most predictive variables. We choose the most predictive accelerations and brakes variables, those that use a threshold of 2 m/s^2. The most predictive speed variables are percentiles 80 and 10.

B.2 Score model

We tried a variety of ways of regularizing our model. In order to test them, we split our sample into three sets. The first is a train set with 47.5% of observations that we use to choose penalty parameters. We also have an estimation set with 47.5% of the data to estimate the model parameters. We set apart the remaining 5% of our observations as a test set. Our selection criterion is test-set mean square error (MSE).

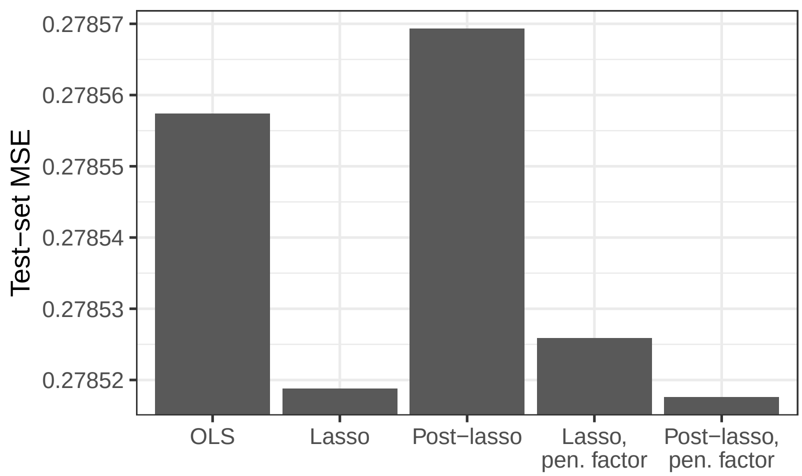

Figure 9: Performance of different models for score.

Note: In this figure we show the test-set mean-squared error for a variety of candidate models for trip rating as a function of a high-dimensional polynomial of driving metrics.

Our baseline model is an OLS regression with no penalization. We also run a lasso model with no penalties on fixed effects, as well as a post-lasso model that keeps all terms with nonzero coefficients in the lasso regression. We choose the penalty by 10-fold cross validation within the train set. We also run a lasso model with an increasing penalty factor. In other words, the penalty factor for an n-th order term is $\lambda\mu^n$, where λ is the base penalty for the model and μ is the penalty factor. We choose μ by 20-fold cross validation within the train set, and we choose λ by 10-fold cross validation within the remaining data in each fold.

Figure 9 compares the performance of all these models. The one that performs best is the post-lasso model with a penalty factor. The lasso model without a penalty factor performs almost as well. We prefer the post-lasso model with a penalty factor since the final model has no penalty, which means our coefficients have no asymptotic bias.

The final score we create uses the same procedure as the post-lasso model with a penalty factor, but we use all the data to estimate it. In other words, we split the sample into two equally sized training and estimation sets (without leaving out any data in a test set).

While we use this procedure to choose our main methodology, if we use a different methodology the coefficients we measure throughout our paper change very little and the interpretation of all our results stays the same.

B.3 Imputing duration and distance metrics for UberTaxi trips

Since UberTaxi trips do not have estimated trip distance and duration, we cannot compute our duration and distance metrics. For that reason, we impute their estimated trip distance and duration based on random forests that we train on our sample of UberX trips. The covariates we use are the timestamp, hour of the day, day of the week, origin coordinates, and the destination coordinates. We also include several transformations of these variables to improve the fit of the random forest.[25]

Both random forests result in good predictions. The out-of-sample R^2 for the distance random forest is 0.958, and for the duration random forest it is 0.936. We use these random forests to impute the distance and duration of UberTaxi trips. We then use these imputed values to compute our driving metrics and scores for UberTaxi trips. In all results that use UberTaxi trips, we use these imputed metrics for UberX trips as well to ensure we use the same metrics.

B.4 Alternative samples for score construction

In order to measure how our results are affected by estimating preferences based on UberX trips only, we estimate regressions similar to those in Table 1, based on the sample of all trips (UberX and UberTaxi). Columns (1)-(2) in Table 13 present those results. We also reestimate our main results from Table 10 using scores that are constructed based on the full sample of trips (see Columns (1)-(2) in Table 14). All results look very similar to our main results in the paper.

[25]We include the sine and cosine of the day of the week times $2/7\pi$ (so that they lie on a unit circle) as well as a 45° rotation of those two variables. Finally, we include rotations of the origin and destination by multiples of 22.5°. We also include the difference between the origin and destination coordinates and rotations thereof, the straight-line distance between them, and the angle of the direction of the trip.

Table 13: Rating response to driving metrics for the full sample of trips and for UberTaxi trips

	Dependent variable: Rating					
	All trips		UberTaxi			
	(1)	(2)	(3)	(4)	(5)	(6)
Mounted	0.0099^{***}	0.0019	-0.0038	-0.0111	0.0044	0.0148^{**}
	(0.0009)	(0.0012)	(0.0093)	(0.0111)	(0.0040)	(0.0067)
Handling	-0.0018^{**}	-0.0056^{***}	-0.0165	-0.0020	-0.0119^{**}	-0.0059
	(0.0008)	(0.0008)	(0.0146)	(0.0105)	(0.0047)	(0.0057)
Brakes	-0.0084^{***}	-0.0030^{***}	0.0074	0.0091	-0.0037	-0.0025
	(0.0007)	(0.0005)	(0.0100)	(0.0058)	(0.0039)	(0.0038)
Accelerations	0.0010	-0.0035^{***}	0.0001	0.0045	0.00002	-0.0014
	(0.0007)	(0.0006)	(0.0090)	(0.0054)	(0.0038)	(0.0037)
Speed low	0.0041^{***}	0.0006	-0.0037	-0.0073	-0.0014	-0.0033
	(0.0006)	(0.0005)	(0.0113)	(0.0061)	(0.0040)	(0.0038)
Speed high	-0.0029^{***}	-0.0071^{***}	-0.0073	-0.0123	0.0052	0.0092^{*}
	(0.0007)	(0.0006)	(0.0140)	(0.0083)	(0.0056)	(0.0052)
Distance	-0.0148^{***}	-0.0148^{***}	-0.0172	-0.0125	-0.0014	0.0013
	(0.0007)	(0.0006)	(0.0175)	(0.0094)	(0.0048)	(0.0044)
Duration	-0.0284^{***}	-0.0252^{***}	-0.0293^{**}	-0.0223^{***}	-0.0220^{***}	-0.0190^{***}
	(0.0007)	(0.0006)	(0.0124)	(0.0068)	(0.0046)	(0.0041)
Pickup	-0.0127^{***}	-0.0110^{***}	-0.0073	-0.0033	-0.0135^{***}	-0.0122^{***}
	(0.0006)	(0.0005)	(0.0114)	(0.0062)	(0.0039)	(0.0036)
Dropoff	-0.0139^{***}	-0.0120^{***}	-0.0030	-0.0066	-0.0084^{**}	-0.0087^{**}
	(0.0007)	(0.0006)	(0.0121)	(0.0067)	(0.0039)	(0.0035)
Trip characteristics FE fine	✓	✓	✓	✓		
Trip characteristics FE coarse					✓	✓
Driver FE		✓		✓		✓
Observations	2,126,993	2,126,993	36,115	36,115	36,115	36,115

Note: This table shows results of regressions of five-star rating on driving metrics. Columns 1 and 2 show results for the full sample that combines UberX and UberTaxi trips. Columns 3 and 4 show results for UberTaxi trips using granular fixed effects. Columns 5 and 6 show results for UberTaxi trips using coarser fixed effects. Standard errors are clustered by driver. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table 14: Comparison of UberX and UberTaxi trips for alternative scores

	Dependent variable:							
	Pooled F	Pooled C	Pooled R	Pooled S	Taxi F	Taxi C	Taxi R	Taxi S
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Matching estimator							
UberX	0.0753^{***}	0.2958^{***}	-0.2811^{***}	0.2790^{***}	-2.0838^{***}	-0.4644^{***}	-0.1468^{***}	-0.2803^{***}
	(0.0232)	(0.0232)	(0.0188)	(0.0091)	(0.0573)	(0.0357)	(0.0178)	(0.0148)
Observations	139,716	139,716	139,716	139,716	139,716	139,716	139,716	139,716
	Panel B: Trip characteristics fixed effects							
UberX	0.0990^{***}	0.2790^{***}	-0.2503^{***}	0.2690^{***}	-2.0055^{***}	-0.4213^{***}	-0.1317^{***}	-0.2651^{***}
	(0.0242)	(0.0241)	(0.0200)	(0.0094)	(0.0601)	(0.0373)	(0.0193)	(0.0148)
Observations	7,147,729	7,147,729	7,147,729	7,147,729	7,147,729	7,147,729	7,147,729	7,147,729
	Panel C: Trip characteristics and rider fixed effects							
UberX	0.2005^{***}	0.3144^{***}	-0.1520^{***}	0.2782^{***}	-1.8475^{***}	-0.4859^{***}	-0.0131	-0.2869^{***}
	(0.0329)	(0.0265)	(0.0293)	(0.0121)	(0.0738)	(0.0490)	(0.0338)	(0.0257)
Observations	7,147,729	7,147,729	7,147,729	7,147,729	7,147,729	7,147,729	7,147,729	7,147,729

Note: This table compares the driving behavior of UberX and UberTaxi trips according to scores constructed using alternative samples. Columns 1 through 4 present results for scores constructed on the full sample of UberX and UberTaxi trips. Columns 5-8 present results for scores constructed using only UberTaxi trips (to construct these scores, we fix the set of variables that remain after the variable selection for the main scores from section 3.2). As in Table 10, Panel A uses a matching estimator, while Panel B presents results from a linear regression that includes trip characteristics fixed effects. Panel C presents results from a linear regression that includes trip characteristics as well as rider fixed effects. All driving metrics are normalized to mean zero and variance one. Standard errors are adjusted as in Abadie and Imbens (2005) and clustered by driver. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

We also estimate similar results based on the sample of UberTaxi trips, although the very small number of UberTaxi trips limits us in ways that make us less confident about our results. In particular, we face a tradeoff between bias (failing to control granularly for trip characteristics) and variance (noisy estimates resulting from a small sample available to estimate effects when granular trip characteristics are included). Columns 3 and 4 in Table 13 show that if we run the exact same specification as in Table 1 on that sample, the coefficients we estimate on driving metrics have wide confidence intervals. This is not surprising since 87% of the trip characteristics groups constructed in Appendix A have no UberTaxi trips, and conditional on having at least one trip on average there are only 1.8 UberTaxi trips per group. In Columns 5 and 6 we present results based on coarser fixed effects.[26] The precision of the coefficients is higher, but the only coefficients that are significantly different from zero are those for handling, speed high, duration, pickup, and dropoff, suggesting that riders may prefer faster trips and better routing. Overall, these results suggest, with the caveats about omitted variable bias we have mentioned, that riders who take UberTaxi have preferences for higher speed.

Columns 4-6 in Table 14 present results based on scores that are estimated on the sample of UberTaxi trips. The variable selection is unstable if we use the same procedure as when creating our baseline scores, so we fix the set of covariates that remain after the variable selection for the baseline scores. Our findings suggest that UberTaxi riders may in fact prefer UberTaxi driving behavior. Note, however, that UberTaxi trips account for only 2% of our sample. Thus, even though this specific sample of riders may prefer the driving behavior of UberTaxi, the overwhelming majority of Uber riders prefer the behavior of UberX drivers or at least view them comparably.

[26] We divide the sample into 8 longitude groups by 8 latitude groups, in comparison to 128 x 128 groups for our baseline groups in Appendix A. We continue to use 15 hour of the week intervals as in Appendix A. For these coarser groups, 87% of groups have at least one UberTaxi trip, and conditional on having at least one there are on average 16 taxi trips. These frequencies are comparable to those for UberX using our baseline trip characteristics.

Appendix C Additional results

C.1 Response to riders preferences on scores

Table 15 presents results similar to those in Table 8, but where the outcome variables are scores. The results are almost identical if we use scores that are constructed using the sample of morning rush hour trips.

Table 15: Response to rider preferences

	Dependent variable:			
	Score F	Score C	Score R	Score S
	(1)	(2)	(3)	(4)
	Panel A: Main effect			
Mean by rider	0.151***	0.170***	0.066***	0.279***
	(0.006)	(0.003)	(0.006)	(0.003)
Observations	3,231,440	3,231,440	3,231,440	3,231,440
	Panel B: Heterogeneous effects by time of the week			
Mean by rider	0.129***	0.139***	0.053***	0.229***
× Off-peak	(0.008)	(0.003)	(0.008)	(0.003)
Mean by rider	0.275***	0.308***	0.141***	0.488***
× AM rush	(0.015)	(0.007)	(0.014)	(0.007)
Mean by rider	0.118***	0.150***	0.043***	0.243***
× PM rush	(0.014)	(0.006)	(0.013)	(0.006)
Observations	3,231,440	3,231,440	3,231,440	3,231,440

Note: This table shows regressions of quality metrics on leave-out means by rider. Only riders with more than 20 trips are included. All regressions control for driver and trip characteristics fixed effects. Standard errors are clustered by driver. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

C.2 Heterogeneity in UberX effect on scores

Table 16 shows the heterogeneity in effect of UberX by morning and evening rush hours.

C.3 Distribution of distance metric for UberX and UberTaxi trips

Figure 10 shows the distribution of distance metrics for UberX and UberTaxi trips.

Table 16: Heterogeneity in effect of UberX using a matching estimator

	Dependent variable:			
	Score F	Score C	Score R	Score S
	(1)	(2)	(3)	(4)
UberX (Off-peak)	0.022	0.167^{***}	-0.314^{***}	0.305^{***}
	(0.027)	(0.027)	(0.021)	(0.010)
UberX (AM rush)	-0.162^{***}	0.014	-0.263^{***}	0.281^{***}
	(0.034)	(0.033)	(0.028)	(0.013)
UberX (PM rush)	0.041	0.127^{***}	-0.177^{***}	0.298^{***}
	(0.038)	(0.034)	(0.033)	(0.014)
Observations	139,716	139,716	139,716	139,716

Note: This table how the effect of UberX on scores varies during rush hours. In order to do so, we split the sample into three subsamples, off-peak trips, morning rush hour trips, and afternoon rush hour trips. Each column presents results for a different response variable. We use a matching estimator, where every UberTaxi trip is compared to the ten nearest UberX trips within the same subsample, based on origin and destination coordinates and time of the week. All driving metrics are normalized to mean zero and variance one. Standard errors are clustered by driver. $^{*}p$ <0.1; $^{**}p$ <0.05; $^{***}p$ <0.01.

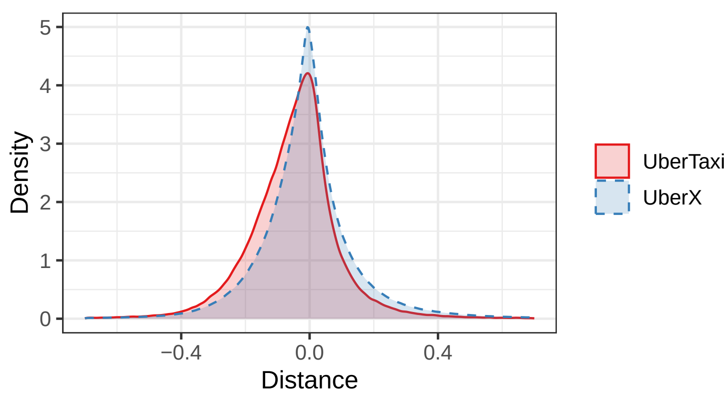

Figure 10: Distribution of distance metric for UberX and UberTaxi trips

Note: This figure presents kernel densities of the distribution of the distance metric for UberX and UberTaxi trips.

Appendix D Robustness checks

D.1 Rider fixed effects

In Table 17 we run the regressions from Table 1, but we also include rider fixed effects. The numbers change somewhat, but the interpretation of the coefficients does not change.

D.2 Traffic conditions

To control for traffic conditions, we exploit our detailed trip data to construct a metric for traffic conditions. For every trip, we residualize the average trip speed on granular origin

Table 17: Rating response to driving metrics

	Dependent variable:		
	Rating	Rating is 5	Rated
	(1)	(2)	(3)
Mounted	0.0005	0.0012***	−0.0009**
	(0.0010)	(0.0005)	(0.0004)
Handling	−0.0039***	−0.0015***	0.0002
	(0.0006)	(0.0003)	(0.0002)
Brakes	−0.0022***	−0.0009***	0.0016***
	(0.0004)	(0.0002)	(0.0002)
Accelerations	−0.0030***	−0.0016***	0.0019***
	(0.0005)	(0.0002)	(0.0002)
Speed low	0.0016***	0.0008***	−0.0022***
	(0.0004)	(0.0002)	(0.0002)
Speed high	−0.0048***	−0.0024***	−0.0004**
	(0.0005)	(0.0002)	(0.0002)
Distance	−0.0124***	−0.0041***	0.0007***
	(0.0005)	(0.0002)	(0.0002)
Duration	−0.0197***	−0.0077***	0.0050***
	(0.0005)	(0.0002)	(0.0002)
Pickup	−0.0109***	−0.0052***	0.0018***
	(0.0004)	(0.0002)	(0.0002)
Dropoff	−0.0123***	−0.0051***	0.0002
	(0.0005)	(0.0002)	(0.0002)
Observations	1,991,742	1,991,742	6,901,200

Note: This table shows results of regressions of rating variables—five-star rating, a dummy for the rating being five, and a dummy for the trips being rated—on driving metrics. All regressions include three-way fixed effects by rider, driver, and trip characteristics. All metrics are normalized to mean zero and variance one. Standard errors are clustered by driver. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

by destination fixed effects, which we create by following the same procedure described in Appendix A to split the our sample into 64 origin and 64 destination rectangles. We call this residual of trip speed the *speed deviation*, which tells us whether the trip speed was atypically high or low given the origin and destination. We then use this variable to construct a traffic variable for every trip: the leave-out mean of the speed deviation of other trips with similar origins and destinations that took place *during the exact same hour*. To define similar trips, we split Chicago into 8 origin groups and 8 destination groups, also following the procedure described in Section 3.1; similar trips are those that start and end in the same groups.[27]

[27] To prevent issues arising from groups of similar trips with too few trips, we also define coarser origin and destination groups, with 4 groups each. We use this second definition whenever the finer group has

Table 18 shows results similar to those in Table 1, but we also include this traffic variable as a control. In columns 1, 3, and 5, we simply control for traffic. In columns 2, 4, and 6, we control for a cubic function of traffic. In both cases we observe that the coefficients on our driving metrics are virtually unchanged after controlling for traffic.

Table 18: Rating response to driving metrics

	Dependent variable:					
	Rating		Rating is 5		Rated	
	(1)	(2)	(3)	(4)	(5)	(6)
Mounted	0.0015	0.0015	0.0010	0.0010	-0.0011^{**}	-0.0011^{**}
	(0.0013)	(0.0013)	(0.0006)	(0.0006)	(0.0005)	(0.0005)
Handling	-0.0050^{***}	-0.0051^{***}	-0.0016^{***}	-0.0017^{***}	-0.0003	-0.0003
	(0.0009)	(0.0009)	(0.0004)	(0.0004)	(0.0003)	(0.0003)
Brakes	-0.0035^{***}	-0.0035^{***}	-0.0016^{***}	-0.0016^{***}	0.0012^{***}	0.0012^{***}
	(0.0006)	(0.0006)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Accelerations	-0.0038^{***}	-0.0038^{***}	-0.0017^{***}	-0.0017^{***}	0.0012^{***}	0.0012^{***}
	(0.0006)	(0.0006)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Speed low	0.0030^{***}	0.0030^{***}	0.0014^{***}	0.0014^{***}	-0.0029^{***}	-0.0029^{***}
	(0.0006)	(0.0006)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Speed high	-0.0067^{***}	-0.0067^{***}	-0.0037^{***}	-0.0037^{***}	-0.0027^{***}	-0.0027^{***}
	(0.0006)	(0.0006)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Distance	-0.0159^{***}	-0.0159^{***}	-0.0049^{***}	-0.0049^{***}	0.0004^{*}	0.0003
	(0.0007)	(0.0007)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Duration	-0.0224^{***}	-0.0222^{***}	-0.0085^{***}	-0.0084^{***}	0.0063^{***}	0.0064^{***}
	(0.0007)	(0.0007)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Pickup	-0.0115^{***}	-0.0115^{***}	-0.0058^{***}	-0.0058^{***}	0.0080^{***}	0.0080^{***}
	(0.0006)	(0.0006)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Dropoff	-0.0138^{***}	-0.0138^{***}	-0.0056^{***}	-0.0056^{***}	-0.0008^{***}	-0.0008^{***}
	(0.0006)	(0.0006)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Traffic	-0.0019^{***}	-0.0023^{***}	-0.0010^{***}	-0.0012^{***}	-0.0008^{***}	-0.0011^{***}
	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Traffic^2		-0.00004^{***}		-0.00002^{**}		-0.00003^{***}
		(0.00001)		(0.00001)		(0.00001)
Traffic^3		0.000003^{***}		0.000001^{***}		0.000002^{***}
		(0.000001)		(0.000000)		(0.000000)
Trip characteristics FE	✓	✓	✓	✓	✓	✓
Driver FE	✓	✓	✓	✓	✓	✓
Observations	1,991,657	1,991,657	1,991,657	1,991,657	6,900,944	6,900,944

Note: This table shows results of regressions of rating variables—five-star rating, a dummy for the rating being five, and a dummy for the trips being rated—on driving metrics and traffic conditions. All metrics are normalized to mean zero and variance one. Standard errors are clustered by driver. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

fewer than 100 trips.

D.3 Alternative trip characteristics fixed effects

In Table 19 we run the regressions from Table 1, but we use different fixed effects for trip characteristics. In columns 1, 3, and 5, we split space into 64 origin groups and 64 destination groups, and we split time into $24 \times 7 = 168$ hours of the week. In columns 2, 4, and 6, we split space into 32 origin groups and 832 destination groups, and we split time into hours. Columns (1)-(5) look almost identical to those in Table 1. We observe some changes in column (6), although the interpretation of the coefficients does not change.

D.4 Heterogeneity by time of the week

We run regressions similar to Equation (1), but we interact our metrics with dummies for whether the trip took place during the morning rush hour, the afternoon rush hour, or during off-peak hours. Table 20 shows the results when we use rating as our dependent variable. Every row corresponds to one metric, and every column represents one dummy for time of the week. The main difference across columns is that riders have stronger preferences for faster trips during the morning rush hour, consistent with people having to arrive to work on time. In some alternative specifications we also included dummies for trips that start and end at airports, but we did not find any noticeable difference with non-airport trips.

D.5 Response to car prices

One potential concern is the effect that car quality might have on riders' preferences. In order to address that issue, we construct a car price variable based on car make, model, year, and mileage. We do not observe mileage, so we assume that cars are driven twice the average mileage of 13,476 mi per year since the car was produced, given Uber cars are used more intensely than average cars. We use Kelley Blue Book data for prices that were collected manually by Uber. This is a time consuming task, so we only have prices for the most common car models, which account for roughly 60% of our trips.

Table 21 shows regressions of rating variables on driving metrics, as well as on prices. Columns (2), (4), and (6) also include the interaction of prices and driving metrics. Neither the car price nor its interactions seem to have any noticeable effect on ratings. Fur-

Table 19: Rating response to driving metrics

	Dependent variable:					
	Rating		Rating is 5		Rated	
	(1)	(2)	(3)	(4)	(5)	(6)
Mounted	0.0023*	0.0013	0.0012*	0.0011	−0.0006	−0.0005
	(0.0013)	(0.0014)	(0.0006)	(0.0007)	(0.0005)	(0.0005)
Handling	−0.0057***	−0.0065***	−0.0022***	−0.0021***	−0.0003	−0.0003
	(0.0008)	(0.0010)	(0.0004)	(0.0005)	(0.0003)	(0.0003)
Brakes	−0.0036***	−0.0045***	−0.0015***	−0.0017***	0.0016***	0.0003
	(0.0006)	(0.0007)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Accelerations	−0.0042***	−0.0044***	−0.0017***	−0.0016***	0.0013***	−0.0007***
	(0.0006)	(0.0007)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Speed low	0.0038***	0.0038***	0.0016***	0.0014***	−0.0027***	−0.0005**
	(0.0006)	(0.0007)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Speed high	−0.0056***	−0.0059***	−0.0029***	−0.0028***	−0.0033***	−0.0051***
	(0.0006)	(0.0007)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Distance	−0.0158***	−0.0131***	−0.0048***	−0.0038***	0.0008***	0.0016***
	(0.0007)	(0.0007)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Duration	−0.0233***	−0.0257***	−0.0089***	−0.0096***	0.0058***	0.0046***
	(0.0007)	(0.0008)	(0.0003)	(0.0004)	(0.0002)	(0.0003)
Pickup	−0.0115***	−0.0124***	−0.0058***	−0.0063***	0.0088***	0.0096***
	(0.0005)	(0.0006)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Dropoff	−0.0138***	−0.0141***	−0.0055***	−0.0057***	−0.0006***	−0.0004*
	(0.0006)	(0.0007)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Trip char. FE 1	✓		✓		✓	
Trip char. FE 2		✓		✓		✓
Driver FE	✓	✓	✓	✓	✓	✓
Observations	1,991,742	1,991,742	1,991,742	1,991,742	6,901,200	6,901,200

Note: This table shows results of regressions of rating variables—five-star rating, a dummy for the rating being five, and a dummy for the trips being rated—on driving metrics. All metrics are normalized to mean zero and variance one. In odd columns, trip characteristics fixed effects are based on 64 origin groups and 64 destination groups as well as on hours of the week. In odd columns, trip characteristics fixed effects are based on 32 origin groups and 32 destination groups as well as on the exact hour. Standard errors are clustered by driver. $^{*}p$ <0.1; $^{**}p$ <0.05; $^{***}p$ <0.01.

thermore, we do not observe any major changes from the main coefficients in Table 1.

D.6 Response to last rating

We conduct an exercise in which, instead of exploring how drivers respond to their app rating, we look at how they respond to the last rating they received. Let r_i represent the rating given to the driver after trip i. Let $l(i)$ represent the index of the last trip by driver $d(i)$ that received a rating before trip i takes place. Then $r_{l(i)}$ represents the last rating

Table 20: Heterogeneity in the response of rating to driving metrics by time of the week

	Dependent variable: Rating		
	Interaction of covariate with:		
	Off-peak	AM rush	PM rush
	(1)	(2)	(3)
Mounted	0.0008	0.0014	0.0036**
	(0.0013)	(0.0019)	(0.0018)
Handling	-0.0049^{***}	-0.0073^{***}	-0.0038^{***}
	(0.0009)	(0.0018)	(0.0015)
Brakes	-0.0032^{***}	-0.0031^{**}	-0.0041^{***}
	(0.0007)	(0.0015)	(0.0013)
Accelerations	-0.0036^{***}	-0.0039^{***}	-0.0033^{**}
	(0.0007)	(0.0015)	(0.0013)
Speed low	0.0030***	0.0031**	0.0036***
	(0.0007)	(0.0015)	(0.0013)
Speed high	-0.0065^{***}	-0.0044^{***}	-0.0090^{***}
	(0.0007)	(0.0016)	(0.0015)
Distance	-0.0147^{***}	-0.0217^{***}	-0.0164^{***}
	(0.0008)	(0.0019)	(0.0015)
Duration	-0.0210^{***}	-0.0351^{***}	-0.0200^{***}
	(0.0009)	(0.0017)	(0.0014)
Pickup	-0.0103^{***}	-0.0156^{***}	-0.0134^{***}
	(0.0007)	(0.0016)	(0.0012)
Dropoff	-0.0126^{***}	-0.0141^{***}	-0.0191^{***}
	(0.0008)	(0.0015)	(0.0015)

Obervations: 1,991,742

Note: This table shows the result of one regression of rating variables on quality metrics interacted with dummies for morning and afternoon rush hour trips. Rows represent quality metrics, and columns represent rush hour dummies. All ratings are normalized to mean zero and variance one. Standard errors are clustered by driver. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

that the driver received before the trip started. We run regressions of the form

$$y_i = \mu_{d(i)} + \nu_{c(i)} + \alpha r_{l(i)} + P_2(n_i;\beta) + \epsilon_{ijkt} \tag{11}$$

where $P_2(n_i;\beta)$ is quadratic function of the number of trips completed by the driver before trip i.

A potential concern with giving this finding a causal interpretation is that there may be factors that lead to serial correlation in driver behavior. In order to isolate the effect of the rating, we use two different instrumental variables strategies. We wish to focus on variation in the previous rating that is not explained by the driver's own behavior or

Table 21: Response to ratings and notifications, including car prices

	Dependent variable:					
	Rating		Rating is 5		Rated	
	(1)	(2)	(3)	(4)	(5)	(6)
Mounted	0.0040^{**}	0.0030	0.0018^{**}	0.0015	-0.0005	-0.0006
	(0.0016)	(0.0032)	(0.0008)	(0.0015)	(0.0006)	(0.0012)
Handling	-0.0043^{***}	-0.0018	-0.0017^{***}	-0.0002	-0.0002	-0.0008
	(0.0010)	(0.0021)	(0.0005)	(0.0010)	(0.0004)	(0.0008)
Brakes	-0.0037^{***}	-0.0045^{***}	-0.0014^{***}	-0.0015^{**}	0.0014^{***}	0.0013^{**}
	(0.0007)	(0.0014)	(0.0003)	(0.0007)	(0.0003)	(0.0005)
Accelerations	-0.0027^{***}	-0.0030^{**}	-0.0012^{***}	-0.0007	0.0014^{***}	0.0019^{***}
	(0.0008)	(0.0015)	(0.0004)	(0.0007)	(0.0003)	(0.0006)
Speed low	0.0036^{***}	0.0039^{***}	0.0016^{***}	0.0019^{***}	-0.0028^{***}	-0.0026^{***}
	(0.0007)	(0.0014)	(0.0003)	(0.0007)	(0.0003)	(0.0005)
Speed high	-0.0059^{***}	-0.0081^{***}	-0.0032^{***}	-0.0041^{***}	-0.0026^{***}	-0.0023^{***}
	(0.0008)	(0.0014)	(0.0004)	(0.0007)	(0.0003)	(0.0005)
Distance	-0.0152^{***}	-0.0141^{***}	-0.0043^{***}	-0.0036^{***}	0.0003	0.0003
	(0.0008)	(0.0016)	(0.0004)	(0.0007)	(0.0003)	(0.0005)
Duration	-0.0237^{***}	-0.0260^{***}	-0.0091^{***}	-0.0108^{***}	0.0057^{***}	0.0056^{***}
	(0.0008)	(0.0016)	(0.0004)	(0.0008)	(0.0003)	(0.0006)
Pickup	-0.0121^{***}	-0.0112^{***}	-0.0061^{***}	-0.0056^{***}	0.0087^{***}	0.0080^{***}
	(0.0007)	(0.0013)	(0.0003)	(0.0006)	(0.0002)	(0.0005)
Dropoff	-0.0149^{***}	-0.0144^{***}	-0.0061^{***}	-0.0062^{***}	-0.0006^{**}	-0.0003
	(0.0008)	(0.0014)	(0.0004)	(0.0007)	(0.0003)	(0.0005)
Price	0.0015^{*}	0.0015^{*}	0.0006	0.0006	0.0004	0.0004
	(0.0008)	(0.0008)	(0.0004)	(0.0004)	(0.0003)	(0.0003)
Price × Mounted		0.0001		0.00003		0.00002
		(0.0004)		(0.0002)		(0.0001)
Price × Handling		-0.0003		-0.0002^{*}		0.0001
		(0.0002)		(0.0001)		(0.0001)
Price × Brakes		0.0001		0.00001		0.00001
		(0.0002)		(0.0001)		(0.0001)
Price × Accels.		0.00005		-0.0001		-0.0001
		(0.0002)		(0.0001)		(0.0001)
Price × Speed low		-0.00004		-0.00004		-0.00003
		(0.0002)		(0.0001)		(0.0001)
Price × Speed high		0.0003^{*}		0.0001		-0.00004
		(0.0002)		(0.0001)		(0.0001)
Price × Dist.		-0.0002		-0.0001		-0.00001
		(0.0002)		(0.0001)		(0.0001)
Price × Dur.		0.0003^{*}		0.0002^{***}		0.000004
		(0.0002)		(0.0001)		(0.0001)
Price × Pickup		-0.0001		-0.0001		0.0001^{*}
		(0.0002)		(0.0001)		(0.0001)
Price × Dropoff		-0.0001		0.00001		-0.00004
		(0.0002)		(0.0001)		(0.0001)
Trip characteristics FE	✓	✓	✓	✓	✓	✓
Driver FE	✓	✓	✓	✓	✓	✓
Observations	1,226,114	1,226,114	1,226,114	1,226,114	4,287,352	4,287,352

Note: This table shows regressions of the trip rating on driving metrics, car price, and their interaction. All metrics are normalized to mean zero and variance one. Standard errors are clustered by driver. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

changing characteristics (e.g. car condition). We instrument for the last rating using the average residual of other similar trips, where by similar we mean trips taken on the same calendar day and hour in the same location. Thus, if exogenous factors lead all drivers to deliver an experience that riders perceive as low quality (e.g. traffic accidents, weather), this shock to the driver's rating is unrelated to driver-specific changes over time.

To implement this, we first take the residual from the model in Equation (3). Then we group trips by 16 origin and destination areas and by calendar day and hour. The instrument for the previous rating is the average residual of all other trips that took place in the group corresponding to the previous trip. The second instrument is the leave-out average of all ratings given by the previous rider.[28]

Table 22 shows the result of these regressions. Panel A shows estimates from an OLS regression, and Panels B and C show results of 2SLS regressions. Throughout, our results are consistent with Tables 4 and 5: ratings have a negative effect on new ratings, but they do not have a large effect on our telemetry metrics or scores.

Table 22: Response to last rating

	Dependent variable:				
	Rating	Score F	Score C	Score R	Score S
	(1)	(2)	(3)	(4)	(5)
	Panel A: OLS				
Last rating	-0.0040^{***}	0.0026	0.0014	0.0008	-0.0008
	(0.0008)	(0.0025)	(0.0011)	(0.0024)	(0.0008)
Observations	1,979,500	6,861,226	6,861,226	6,861,226	6,861,226
	Panel B: IV, average rating by rider				
Last rating	-0.0047^{**}	-0.0001	0.0012	-0.0017	-0.0022
	(0.0021)	(0.0077)	(0.0035)	(0.0073)	(0.0025)
Observations	1,811,637	6,284,068	6,284,068	6,284,068	6,284,068
	Panel C: IV, both instruments				
Last rating	-0.0142^{***}	-0.0037	-0.0016	0.0003	-0.0052^{*}
	(0.0026)	(0.0086)	(0.0040)	(0.0083)	(0.0028)
Observations	771,482	2,689,979	2,689,979	2,689,979	2,689,979

All safety metrics are normalized to mean zero and variance one.

Note: This table shows regressions of quality metrics and scores on the rating for the last trip completed by the driver. Panel A presents an OLS estimator. Panel B presents a 2SLS estimator, where we use the rating of trips that took place nearby and during the same time as an instrument. Panel B also presents a 2SLS estimator, where we use the leave-out average rating by rider as an instrument. All driving metrics are normalized to mean zero and variance one. Standard errors are clustered by driver. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

[28]Our results are very similar if we exclude trips with riders with fewer than 10 trips

Table 23: Response to last rating

	Dependent variable:									
	Mounted	Handling	Brakes	Accels.	Speed low	Speed high	Distance	Duration	Pickup	Dropoff
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Panel A: OLS									
Last rating	0.0012***	−0.0011*	−0.0010*	−0.0004	−0.0010*	0.0006	0.0005	−0.0004	0.0004	−0.0005
	(0.0004)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0005)	(0.0006)	(0.0005)	(0.0005)
Observations	6,861,226	6,861,226	6,861,226	6,861,226	6,861,226	6,861,226	6,861,226	6,861,226	6,861,226	6,861,226
	Panel B: IV, average rating by rider									
Last rating	0.0016	−0.0006	−0.0034*	−0.0039**	0.0015	0.0021	0.0008	−0.0024	−0.0013	0.0004
	(0.0013)	(0.0016)	(0.0018)	(0.0018)	(0.0019)	(0.0018)	(0.0017)	(0.0017)	(0.0017)	(0.0015)
Observations	6,284,068	6,284,068	6,284,068	6,284,068	6,284,068	6,284,068	6,284,068	6,284,068	6,284,068	6,284,068
	Panel C: IV, both instruments									
Last rating	0.0031**	−0.0023	−0.0008	−0.0021	−0.0005	0.0023	−0.0008	−0.0019	0.0002	0.0019
	(0.0014)	(0.0018)	(0.0021)	(0.0020)	(0.0021)	(0.0019)	(0.0019)	(0.0020)	(0.0019)	(0.0017)
Observations	2,689,979	2,689,979	2,689,979	2,689,979	2,689,979	2,689,979	2,689,979	2,689,979	2,689,979	2,689,979

Note: This table shows regressions of quality metrics and scores on the rating for the last trip completed by the driver. Panel A presents an OLS estimator. Panel B presents a 2SLS estimator, where we use the rating of trips that took place nearby and during the same time as an instrument. Panel B also presents a 2SLS estimator, where we use the leave-out average rating by rider as an instrument. All driving metrics are normalized to mean zero and variance one. Standard errors are clustered by driver. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

D.7 Excluding trips before warnings

Tables 24 and 25 present similar results to those in Panel A on Tables 4 and 5, but in which the last few trips before receiving a warning are excluded from the sample. We observe very similar results.

Appendix E Further details about the deactivation process

Table 26 gives a sense of the number of drivers in each one of the ranges for the app rating (the average rating from the last 500 trips). This table also shows how many drivers already completed 500 rated trips, so that every additional trip only contributes one five hundredth to the rating after the trip.

To get a sense of how strong these incentives are, Figure 11 shows the percentage of drivers that are at risk of falling below each threshold. It shows how many 3-star trips the driver would have to complete in order to fall below the threshold. Only a very small number of drivers are likely to eventually reach the 4.4 threshold for deactivation. A somewhat more important fraction of drivers are close and even below the 4.6 threshold

Table 24: Response to ratings and notifications

	Dependent variable:				
	Rating	Score F	Score C	Score R	Score S
	(1)	(2)	(3)	(4)	(5)
	Panel A: Excluding 10 trips before warning				
Has received notif.	0.086***	0.096***	0.035*	0.083***	0.016**
	(0.007)	(0.022)	(0.019)	(0.019)	(0.008)
App rating	−0.296***	0.075**	0.043*	0.071**	0.028**
	(0.014)	(0.037)	(0.026)	(0.033)	(0.012)
Observations	1,973,515	6,843,534	6,843,534	6,843,534	6,843,534
	Panel B: Excluding 20 trips before warning				
Has received notif.	0.075***	0.087***	0.018	0.083***	0.013
	(0.008)	(0.024)	(0.021)	(0.021)	(0.008)
App rating	−0.310***	0.076**	0.037	0.075**	0.030**
	(0.014)	(0.038)	(0.026)	(0.034)	(0.013)
Observations	1,966,316	6,819,548	6,819,548	6,819,548	6,819,548

Note: This table shows regressions of ratings and scores on app ratings and dummies for having received notifications, similar to those in Panel A of table 4. Panel A excludes the last 10 trips before receiving a warning, and Panel B excludes the last 20 trips before receiving a warning. All regressions include driver and trip characteristics fixed effects, and control for a quadratic function of the number of Uber trips the driver has completed. Standard errors are clustered by driver. All driving metrics are normalized to mean zero and variance one. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table 25: Response to ratings and notifications

	Dependent variable:									
	Mounted	Handling	Brakes	Accels.	Speed low	Speed high	Distance	Duration	Pickup	Dropoff
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Panel A: Excluding 10 trips before warning									
Has received notif.	0.050***	−0.038***	0.004	−0.012	0.013**	−0.0001	−0.011***	−0.017***	−0.014***	−0.004
	(0.013)	(0.013)	(0.008)	(0.008)	(0.006)	(0.006)	(0.004)	(0.005)	(0.005)	(0.004)
App rating	0.040**	−0.023	0.009	0.002	0.044***	0.017*	0.009	−0.025***	−0.007	−0.013**
	(0.018)	(0.019)	(0.011)	(0.012)	(0.009)	(0.009)	(0.007)	(0.007)	(0.007)	(0.007)
Observations	6,843,534	6,843,534	6,843,534	6,843,534	6,843,534	6,843,534	6,843,534	6,843,534	6,843,534	6,843,534
	Panel B: Excluding 20 trips before warning									
Has received notif.	0.044***	−0.031**	0.007	−0.009	0.008	−0.001	−0.009**	−0.018***	−0.011**	−0.006
	(0.015)	(0.015)	(0.009)	(0.009)	(0.007)	(0.007)	(0.004)	(0.005)	(0.006)	(0.005)
App rating	0.038**	−0.015	0.012	0.001	0.045***	0.016	0.009	−0.025***	−0.003	−0.015**
	(0.018)	(0.019)	(0.012)	(0.012)	(0.009)	(0.010)	(0.007)	(0.007)	(0.007)	(0.007)
Observations	6,819,548	6,819,548	6,819,548	6,819,548	6,819,548	6,819,548	6,819,548	6,819,548	6,819,548	6,819,548

Note: This table shows regressions of metrics on app ratings and dummies for having received notifications, similar to those in Panel A of table 4. Panel A excludes the last 10 trips before receiving a warning, and Panel B excludes the last 20 trips before receiving a warning. All regressions include driver and trip characteristics fixed effects, and control for a quadratic function of the number of Uber trips the driver has completed. Standard errors are clustered by driver. All driving metrics are normalized to mean zero and variance one. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

for the first notification. This suggests that if this deactivation process has an effect on driving behavior it is most likely through behavioral nudges instead of through actual incentives.

Table 27 shows how frequently drivers cross one of these thresholds, both from above and from below. We see that there is a large number of events, even for the threshold at

Table 26: Number of trips during which driver ratings satisfy each condition

	Number	Fraction
Total	6, 901, 197	
Lifetime trips >500	4, 412, 839	0.639
Rating <4.6	507, 310	0.074
Rating <4.5	198, 536	0.029
Rating <4.4	89, 375	0.013

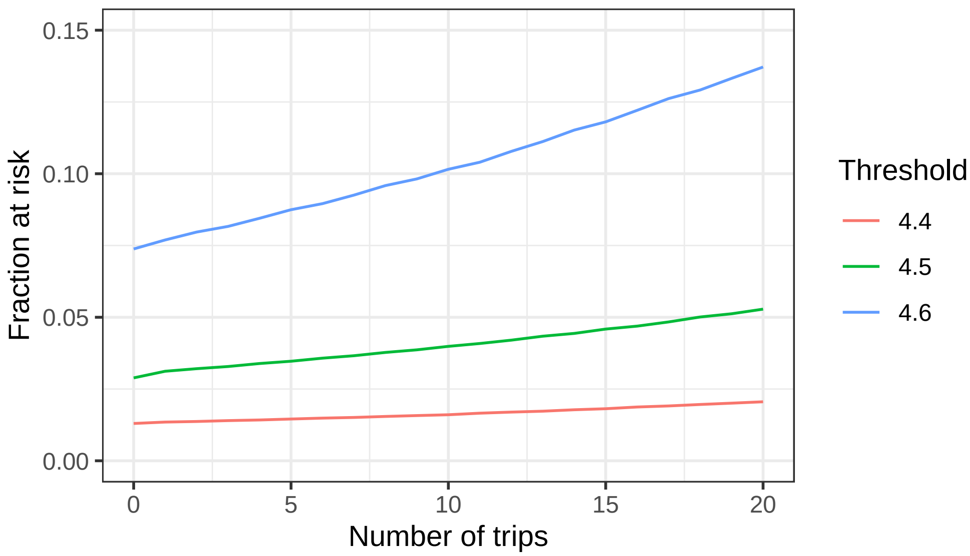

Figure 11: Fraction of drivers at risk of falling below rating thresholds

Note: This plots shows the fraction of drivers whose app rating would fall below a certain threshold if the next N consecutive trips received a 3-star rating, where N is displayed on the x-axis. The values corresponding to $N = 0$ represent the fraction whose app rating currently falls below the threshold.

4.4, close to which there are not that many drivers.

Table 27: Number of threshold crossings

	From above		From below	
Threshold	Crossings	Unique drivers	Crossings	Unique drivers
4.6	7, 201	4, 650	6, 896	4, 372
4.5	4, 758	3, 240	4, 181	2, 778
4.4	3, 476	2, 612	2, 670	1, 955

Note: Number of events in our dataset in which a driver's rating crosses one of the rating thresholds, either from above or from below.

Appendix F Additional experimental results

F.1 Balance of experimental sample

Table 28 shows results of a balance test for mean pre-experiment period outcomes for each driver. Estimates are across trips in the experimental period and are clustered by driver. While all of the non-speed metrics and scores have insignificant results, there seems to be some difference in the speed outcomes.

Table 28: Balance Test for Experiment

	Dependent variable:						
	Score F	Score C	Score R	Score S	Mounted	Handling	Brakes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	-0.009	-0.017	-0.007	-0.006	0.701^{***}	0.102^{***}	0.219^{***}
	(0.014)	(0.014)	(0.010)	(0.005)	(0.005)	(0.002)	(0.001)
Treatment	0.016	0.029	0.013	0.011^{*}	0.0003	-0.002	0.00001
	(0.018)	(0.018)	(0.013)	(0.006)	(0.007)	(0.003)	(0.001)
Observations	3,785,965	3,785,965	3,785,965	3,785,965	3,785,965	3,785,965	3,785,965
	Dependent variable:						
	Accels.	Speed low	Speed high	Distance	Duration	Pickup	Dropoff
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.186^{***}	23.735^{***}	86.038^{***}	0.014^{***}	0.056^{***}	-0.00001	-0.0001
	(0.001)	(0.060)	(0.053)	(0.0004)	(0.001)	(0.00004)	(0.0001)
Treatment	-0.001	0.234^{***}	0.159^{**}	-0.002^{***}	-0.0003	0.00002	0.0002^{*}
	(0.002)	(0.079)	(0.069)	(0.001)	(0.001)	(0.0001)	(0.0001)
Observations	3,785,965	3,785,965	3,785,965	3,785,965	3,785,965	3,785,965	3,785,965

Note: The table shows regressions of driving metrics and scores on a treatment dummy. We focus on trips that took place before the beginning of the experiment. All metrics are normalized to mean zero and variance one. Standard errors are clustered by driver. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

F.2 Effect of Dashboard Experiment on Metrics

Table 29 shows results for regressions of the form in Equations 6 and 7 where the dependent variables are the quality metrics.

Table 29: Results of experiment, metrics

	Dependent variable:									
	Mounted	Handling	Brakes	Accels.	Speed low	Speed high	Distance	Duration	Pickup	Dropoff
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Panel A: Intent to treat estimator									
Treatment	-0.004	-0.015^{**}	-0.0004	0.002	0.003	-0.002	-0.002	-0.006^{***}	-0.002	-0.0005
	(0.006)	(0.006)	(0.004)	(0.005)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Pre-Period Mean	2.081^{***}	3.697^{***}	4.679^{***}	4.836^{***}	0.051^{***}	0.074^{***}	1.691^{***}	1.871^{***}	13.136^{***}	5.392^{***}
	(0.009)	(0.036)	(0.031)	(0.030)	(0.0004)	(0.001)	(0.028)	(0.019)	(0.340)	(0.407)
Observations	3,785,965	3,785,965	3,785,965	3,785,965	3,785,965	3,785,965	3,682,194	3,692,289	3,785,965	3,785,965
	Panel B: 2SLS estimator									
Interaction	-0.005	-0.032^{***}	-0.010	-0.005	0.006	-0.004	-0.001	-0.007^{*}	-0.001	0.001
	(0.012)	(0.011)	(0.008)	(0.009)	(0.006)	(0.006)	(0.003)	(0.004)	(0.003)	(0.003)
Observations	3,785,965	3,785,965	3,785,965	3,785,965	3,785,965	3,785,965	3,682,194	3,692,289	3,785,965	3,785,965

Note: Panel A shows regressions of driving metrics on a dummy for being in the treatment group of the dashboard experiment. Panel B shows 2SLS regressions of driving metrics on a dummy for observing the dashboard, using the treatment dummy as an instrument. All metrics are normalized to mean zero and variance one. Standard errors are clustered by driver. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.