How Do Platform Participants Respond to an Unfair Rating? An Analysis of a Ride-Sharing Platform Using a Quasi-Experiment

Anuj Kapoor[1] Catherine Tucker[2]

[1]David Eccles School of Business
University of Utah

[2]MIT Sloan School of Management

Research Question

AUG 14, 2014 @ 12:31 PM 104,881 THE LITTLE BLACK BOOK OF BILLIONAIRE SECRETS

Uber's Ratings Terrorize Drivers And Trick Riders. Why Not Fix Them?

Jeff Bercovici, FORBES STAFF
I cover technology with an emphasis on social and digital media. FULL BIO

A few days ago, I had a novel Uber experience: I took a ride in a car as grimy and musty-smelling as a typical yellow cab. The driver was friendly and knew his way around, but he was clearly falling short of Uber's standard that the sedans in its UberX fleet be in "excellent condition." Since customer feedback is "important to insuring a high-quality experience," according to the company, when it came time to rate my trip, I only gave the driver three out of five stars.

Just kidding. I gave him five stars, of course. What do you think I am, a psychopath?

Figure:

Research Question

- What happens when sellers receive an unfairly negative rating?

What we Do

- Detailed panel on ratings of ride-share drivers over time.
- Platform informed each driver after each ride (before they could take a new passenger) what rating they received from the earlier ride via a smart phone application.
- To identify the causal effect of an unfair bad rating, we looked for an exogenous shifter in the form of cancellations by other drivers.
- Find that unfair negative rating in turn leads to a worse driver rating in the subsequent ride and may lead them to eventually leave the platform
- Behavioral evidence suggests it is an emotional response

How this research contributes to the academic literature

- Early literature shows that reviews matter
- Later literature shows potential for bias in reviews both due to fake reviews and selection in decision about whether to provide a review.
- Concerns about consumer welfare
- Take opposite tack and think about review's affect on the receiver of the review

How this research illuminates managerial practices

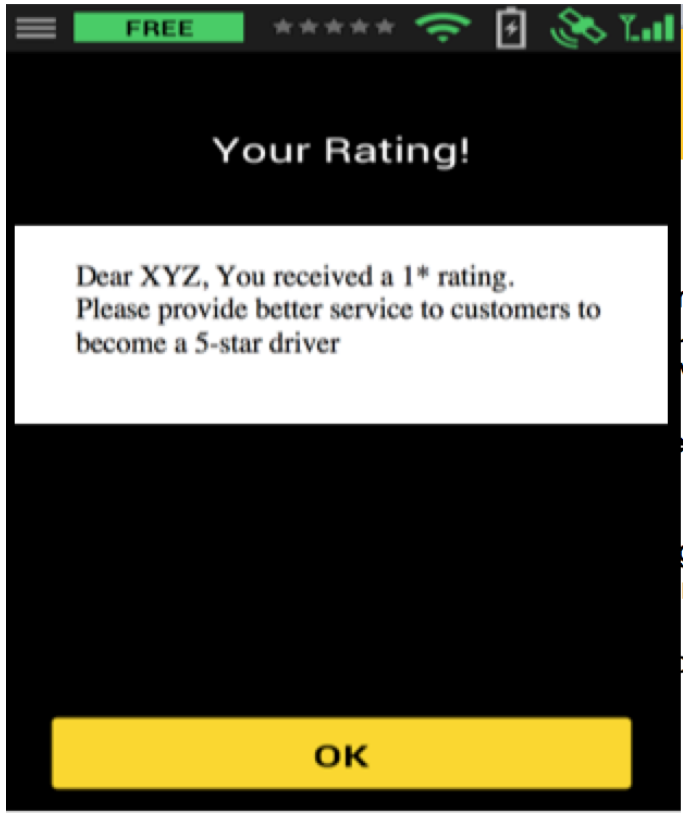
- Help managers inform reputation system design in online platforms.
- Most focus has been on maximizing ease and convenience of supply of reviews
- Unfair negative reviews can also affect participation decisions and platform performance.
- Platform operators need to be careful when designing their platforms so that there is no potential for mis attribution of errors due to factors outside the platform participants control.

Outline

There are three important institutional features of our data

1. Only drivers were rated and they saw their rating immediately. Driver training emphasized that reviews were a function of driver effort
2. If a driver canceled, that passenger was reassigned to another driver. The second driver was not told that there had already been a cancellation.
3. All data from town of Chandigarh which has unusual geographical makeup - this gives us exogenous variation in driving conditions

Figure: Screen shot of the App from the Drivers Perspective



Data Description

- Six months of data from June 2015 to December 2015 on 2,197 drivers from the platform.
- Average rating (if given) was 4.35 stars but many trips were not rated (Only 23% of rides earned a 5 star rating).
- Reviews on the platform are predominantly positive if they are actually given. More than 70% of reviewers leave a 5 star rating.
- The average fare was 73 rupees which is around $1.15.
- The average length of trip was around 25 minutes. Customers were reassigned to a new driver for 3% of trips.
- Over 500,000 rides, and more than 15,000 rides were reassigned.

Figure: Distribution of Ratings that the Drivers Received during our Observation Period

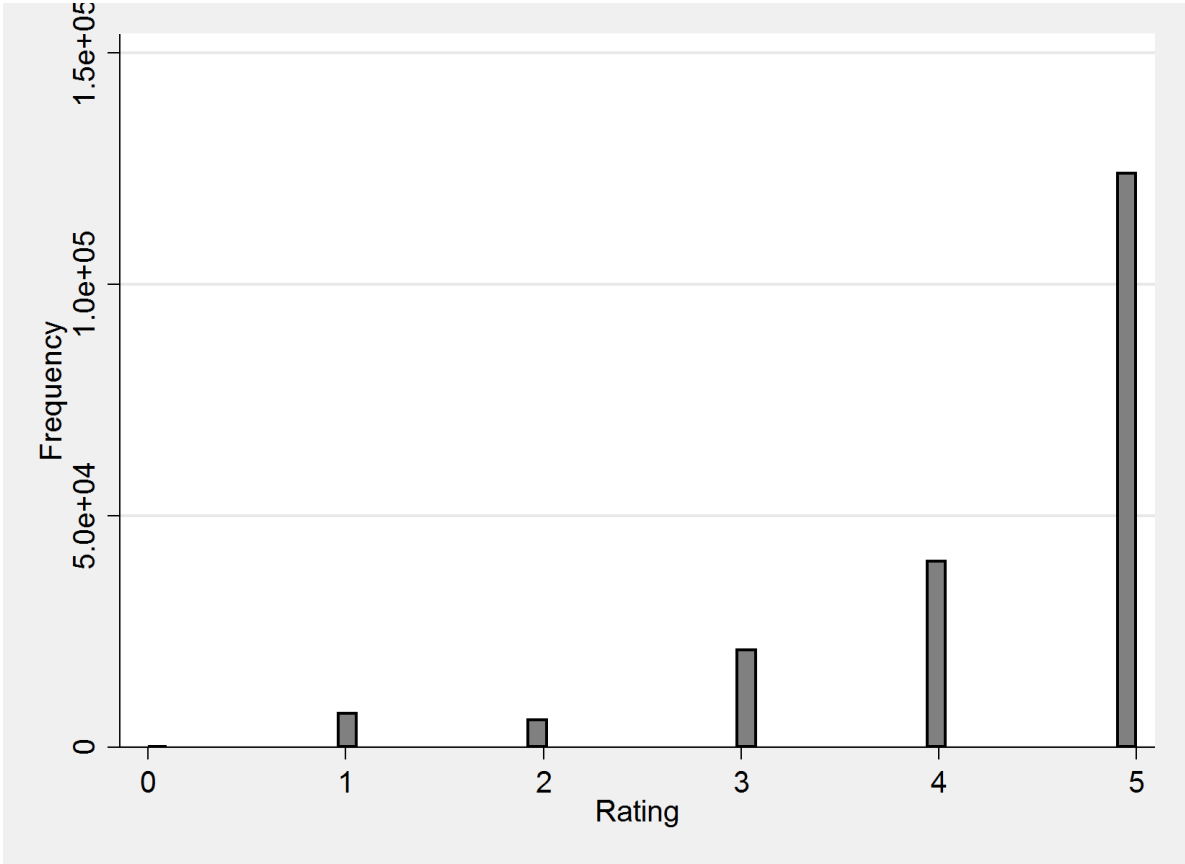


Table: Summary Statistics

	Mean	Std Dev	Min	Max	Observations
Rating	4.35	1.03	0	5	198578
5* Rating	0.23	0.42	0	1	529469
Customer Reassigned	0.028	0.17	0	1	529469
Trip Length	24.2	127.2	-713.8	59216.6	529469
Rural	0.56	0.50	0	1	529469
Fare	73.0	34.1	40	3738	529469
Observations	529469				

Table: Driver Level Summary Statistics

	Mean	Std Dev	Min	Max	Observations
Cumulative Rating	4.34	0.40	1	5	2142
Trips During Dataset	241.0	276.0	1	1559	2197
Average Daily Trips	5.23	2.69	1	16.0	2197
Observations	2197				

Econometric Analysis is Straightforward

$$
\begin{aligned}
5 * \text{Rating}_{i,j} = \alpha &+ \beta_1 5 * \text{Rating}_{i,j\ \ 1} \\
&+ \beta_2 \text{Fare}_j \\
&+ \beta_3 \text{Rural}_j \\
&+ \beta_4 \text{TripLength}_j \\
&+ \beta_5 \text{CustomerReassigned}_j \\
&+ \beta_6 \text{Day}_d \\
&+ \beta_7 \text{Time}_t \\
&+ \beta_8 \text{Driver}_i + \epsilon_k
\end{aligned}
$$

Table: Ordinary Least Squares Estimates: Correlation in Ratings Over Time

	(1) 5* Rating	(2) 5* Rating	(3) 5* Rating	(4) 5* Rating
Previous Rating 5*	0.0325	0.0315	0.0310	0.0305
	(0.0014)	(0.0014)	(0.0014)	(0.0014)
Customer Reassigned		-0.0633	-0.0637	-0.0637
		(0.0035)	(0.0035)	(0.0035)
Trip Length		-0.0000	-0.0000	-0.0000
		(0.0000)	(0.0000)	(0.0000)
Rural		-0.0026	-0.0028	-0.0027
		(0.0012)	(0.0012)	(0.0012)
Fare		-0.0006	-0.0006	-0.0006
		(0.0000)	(0.0000)	(0.0000)
Day of Week FE	No	No	No	Yes
Time Fixed Effects	No	No	Yes	Yes
Driver Fixed Effects	Yes	Yes	Yes	Yes
Observations	527272	527272	527272	527272
Log-Likelihood	-291056	-290148	-289875	-289750

Standard errors in parentheses

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

Our Instrument

- To identify the causal relationship between ratings, we need a plausibly exogenous source that may influence ratings that is not correlated with the drivers and consumers unobserved attributes.
- We use as our instrument whether or not the customer was reassigned from another driver who had canceled their ride.

Figure: Model Free Evidence for Instrumental Variables Approach

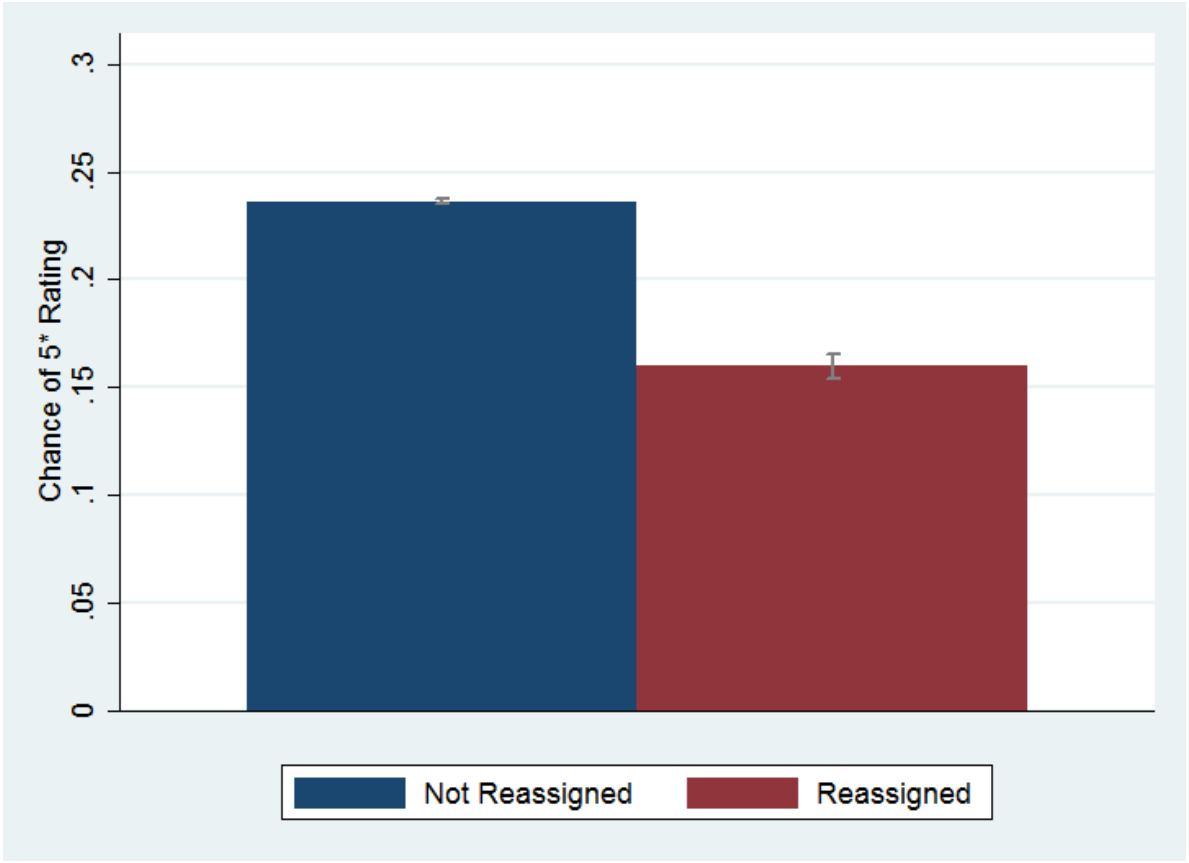


Table: Instrumental Variables: The Correlation in Ratings is More Pronounced if Shifted by Something Outside the Driver's Control

	(1) 5* Rating	(2) 5* Rating	(3) 5* Rating
Previous Rating 5*	0.1469	0.1232	0.1282
	(0.0493)	(0.0495)	(0.0493)
Customer Reassigned		-0.0626	-0.0629
		(0.0035)	(0.0035)
Trip Length		-0.0000	-0.0000
		(0.0000)	(0.0000)
Rural		-0.0025	-0.0026
		(0.0012)	(0.0012)
Fare		-0.0006	-0.0006
		(0.0000)	(0.0000)
Day of Week FE	No	No	Yes
Time Fixed Effects	No	No	Yes
Driver Fixed Effects	Yes	Yes	Yes
Observations	527244	527244	527244
Log-Likelihood	-294485	-292363	-292261
Anderson Rubin F-Stat	8.99	6.24	6.82
Anderson Rubin p-value	0.0027	0.013	0.0090
Anderson canonical correlations LR	416.5	409.8	413.7
Anderson canonical correlations LR p-value	1.4e-92	4.0e-91	5.7e-92

Standard errors in parentheses

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

First Stage Results for Table-4

- Positive relationship between receiving a five star rating and receiving a five star rating on a subsequent trip.
- Rating received was influenced by external forces - and consequently potentially unfair - drivers are far more likely to receive a negative rating than otherwise for the current trip.
- Result holds when controlling for the characteristics of that trip, such as how long and expensive it was and whether it was rural in origin and controlling additionally for variation in ratings attributable to time and day of the week.

Table: First-Stage Results for Table

	(1) L.5* Rating	(2) L.5* Rating	(3) L.5* Rating
Prior Trip Reassigned	-0.0460	-0.0456	-0.0458
	(0.0023)	(0.0023)	(0.0022)
Customer Reassigned		-0.0079	-0.0080
		(0.0035)	(0.0035)
Trip Length		0.0000	0.0000
		(0.0000)	(0.0000)
Rural		-0.0012	-0.0012
		(0.0012)	(0.0012)
Fare		-0.0002	-0.0002
		(0.0000)	(0.0000)
Day of Week FE	No	No	Yes
Time Fixed Effects	No	No	Yes
Driver Fixed Effects	Yes	Yes	Yes
Observations	527244	527244	527244

Standard errors in parentheses

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

Investigating whether our results are robust to various challenges to the exclusion

- Results hold if we exclude new (and inexperienced) drivers.
- When we exclude late night drivers, our results still hold.
- Reassigned customers are systematically different, and more likely to be negative in their rating behavior - for example if there was negative word of mouth about a particular customer which led more experienced drivers to reject them.
- Exacting customers, that is customers who on average give a rating below three stars, are more likely to be reassigned. We show that when we exclude such customers, our results hold.
- Only new customers, who potentially be less experienced about the need to give five star ratings, are more likely to be reassigned. We show that when we exclude these new customers, our results hold.

Suggestive evidence that the Mechanism is Driven by Emotional Response

- Rationally, a driver who receives a bad rating should invest more to compensate for that bad rating in subsequent trips. However, that is not we observe. Therefore, a potential explanation of our results is that what we are measuring is an emotional response to the perception of an unfair rating.
- To tease this apart, we look at instances where there is more likely to be emotion involved.

Table: Investigating whether our Results are Robust to Various Challenges to the Exclusion Restriction

	Not First Month	Not Late Night	Not Exacting Customer	Not New Customer
	(1)	(2)	(3)	(4)
	5* Rating	5* Rating	5* Rating	5* Rating
Previous Rating 5*	0.1401	0.1473	0.1290	0.1059
	(0.0508)	(0.0527)	(0.0511)	(0.0576)
Customer Reassigned	-0.0631	-0.0611	-0.0635	-0.0642
	(0.0037)	(0.0038)	(0.0037)	(0.0042)
Trip Length	-0.0000	-0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Rural	-0.0029	-0.0014	-0.0026	-0.0024
	(0.0012)	(0.0012)	(0.0012)	(0.0013)
Fare	-0.0006	-0.0006	-0.0006	-0.0005
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Day of Week FE	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes
Driver Fixed Effects	Yes	Yes	Yes	Yes
Observations	507315	473826	507513	431444
Log-Likelihood	-281708	-262610	-287174	-252067
Anderson Rubin F-Stat	7.69	7.92	6.43	3.40
Anderson Rubin p-value	0.0055	0.0049	0.011	0.065
Anderson canonical correlations LR	390.7	363.5	393.5	316.9
Anderson canonical correlations LR p-value	5.9e-87	4.8e-81	1.5e-87	6.9e-71

Standard errors in parentheses

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

Suggestive Evidence that the Mechanism is Driven by Emotional Response

- Temporal proximity enhances the effect we measure, which provides suggestive evidence that the negative shift in subsequent rating is to do with the drivers current emotional state.
- Instances where the previous unfair rating represents a larger deviation from the drivers average rating where we tend to see the largest negative effect.

Table: Suggestive Evidence that the Mechanism is Driven by Emotional Response (Part 1)

	Long Interval	Short Interval	Small Deviation	High Deviation
	(1)	(2)	(3)	(4)
	5* Rating	5* Rating	5* Rating	5* Rating
Previous Rating 5*	0.0705	0.1790	0.1063	0.2827
	(0.0633)	(0.0827)	(0.0486)	(0.1620)
Trip Length (Hr)	-0.0009	-0.0048	-0.0012	-0.0281
	(0.0003)	(0.0009)	(0.0004)	(0.0057)
Fare	-0.0006	-0.0006	-0.0006	-0.0006
	(0.0000)	(0.0000)	(0.0000)	(0.0001)
Customer Reassigned	-0.0639	-0.0600	-0.0580	-0.0750
	(0.0048)	(0.0053)	(0.0033)	(0.0114)
Rural	-0.0038	-0.0013	-0.0031	-0.0055
	(0.0016)	(0.0017)	(0.0012)	(0.0035)
Day of Week FE	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes
Driver Fixed Effects	Yes	Yes	Yes	Yes
Observations	261304	265877	445646	78791
Log-Likelihood	-137626	-153545	-225980	-55592

Suggestive Evidence that the Mechanism is Driven by Emotional Response

- Our proposed effect is indeed larger for younger drivers, again suggesting it is the potential for a negative response to to negative feedback which drivers our result.
- Another potential interpretation of this effect that it reflects experience - it is only the older (and more experienced) drivers who have the reflexive skill to keep on driving well even if they are feeling upset at an unfair negative rating.
- Instances where the rides were difficult, i.e. where they were more likely to encounter roundabouts, where we see the largest negative effect.
- This is suggestive that the mechanism is that the driver is less likely to exert themselves to avoid potential traffic hazards if they receive a potentially unfair review.

Figure: Difficult Rides: Presence of cows on Indian roads

Table: Suggestive Evidence that the Mechanism is Driven by Emotional Response (Part 2)

	Old	Young	Easy Driving	Difficult Driving
	(1)	(2)	(3)	(4)
	5* Rating	5* Rating	5* Rating	5* Rating
Previous Rating 5*	0.0196	0.1345	0.0826	0.1716
	(0.0878)	(0.0811)	(0.0689)	(0.0725)
Trip Length (Hr)	-0.0011	-0.0012	-0.0011	-0.0017
	(0.0004)	(0.0005)	(0.0003)	(0.0005)
Fare	-0.0007	-0.0006	-0.0007	-0.0005
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Customer Reassigned	-0.0564	-0.0700	-0.0604	-0.0663
	(0.0059)	(0.0061)	(0.0048)	(0.0053)
Rural	-0.0010	-0.0033	0.0050	-0.0045
	(0.0020)	(0.0020)	(0.0017)	(0.0018)
Day of Week FE	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes
Driver Fixed Effects	Yes	Yes	Yes	Yes
Observations	184935	184687	262235	264940
Log-Likelihood	-100418	-103942	-146122	-145416

Is situations where Ratings are unfair which lead Drivers to Leave the platform

- Drivers with a low rating are no more likely to leave the platform than the drivers with higher ratings just in terms of the raw correlation.
- The more rides a driver has been allocated that were reassigned, the more likely they are to leave the platform.
- Low ratings that are a function of unfair consequences such as accepting reassigned customers are more likely to make a driver leave the platform.

Table: It is Only Situations where Ratings are Unfair which Lead Drivers to Leave the Platform

	Cox Model		IV	IV:Exc.Final
	(1)	(2)	(3)	(4)
	Last Trip	Last Trip	Last Trip	Last Trip
Cumulative Average Rating	-0.1862		-0.0635	-0.0493
	(0.0903)		(0.0251)	(0.0214)
Cumulative Average Reassigned		1.9428		
		(0.3078)		
_cons			0.2821	0.2188
			(0.1125)	(0.0960)
Date Fixed Effects	No	No	Yes	Yes
Observations	89158	90324	89158	80187
Log-Likelihood	-18648	-19217	55387	63405
Anderson Rubin F-Stat			6.58	5.41
Anderson Rubin p-value			0.010	0.020
Anderson canonical correlations LR			245.0	225.1
Anderson canonical correlations LR p-value			3.2e-55	7.0e-51

Standard errors in parentheses

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

Limitations

- Results are limited to one city in India.
- Our vendor partner provides rating feedback to the driver straight after the end of the ride, and other vendors may not do so. Therefore, ratings may be more salient in our platform than in other platforms.
- The consumer behavior we rely on for exogenous variation, which is the transfer of blame for a delay from the canceling driver to the replacement driver, may not be universal.
- As consumers become more sophisticated, there is the chance that they will not transfer blame across platform providers in this way.

Punchline

- We use a natural experiment to identify the effect of unfair ratings on seller performance
- Exploit exogenous variation in ratings due to *other* drivers cancelling rides
- Unfair ratings can lead to worse performance on platforms and people to leave platforms '
- Suggests that the balance of always giving favoring the consumer in how ratings are managed may be misguided. Instead, steps need to be taken to avoid instances where there is a potential for unfair ratings.

Thank you

- I would love questions.
- cetucker@mit.edu

Table: Robustness to Binary Functional Form

	Probit	Bi Probit
	(1)	(2)
	5* Rating	5* Rating
5* Rating		
Previous Rating 5*	0.1368	1.3019
	(0.0044)	(0.0808)
Trip Length (Hr)	-0.0480	-0.0456
	(0.0044)	(0.0041)
Fare	-0.0024	-0.0021
	(0.0001)	(0.0001)
Customer Reassigned	-0.2336	-0.2082
	(0.0126)	(0.0118)
Rural	-0.0095	-0.0082
	(0.0038)	(0.0034)
Previous Rating 5*		
Prior Trip Reassigned		-0.1824
		(0.0086)
Day of Week FE	Yes	Yes
Time Fixed Effects	Yes	Yes
Observations	527272	527272
Log-Likelihood	-284652	-571354