Driver Gender: *Female drivers are tipped 10%-12% more than male drivers*

Table 5: Sample stats for tips by driver gender

Trip statistics across driver genders are reported in Table 5. There are more male drivers than female drivers, and male drivers take more trips on average. In the raw means, we find that women get tipped 5.7 cents (12%) more on average. This difference is driven by a higher percentage of trips tipped—women receive tips nearly 10% more often than men.

Table 6: Robustness analysis

…add RHS variables to address possible endogeneity (omitted variable impact/bias)

We test the robustness of these results via regression analysis, successively adding controls for the pick-up location, date of the trip, hour of the week, and drop-off location. Empirical estimates are presented in Table 6 and are clustered by driver. The results in Table 6 paint a picture similar to the raw data: female drivers receive considerably more tips than men, roughly 4.8 to 5.7 cents more on the average trip. …

Of course, differences in tipping may still be driven by variation in trip distance, fare, or characteristics that differ between men and women, such as age, experience, or rating. … Even after account[ing] for these factors, the gender tip gap remains statistically significant (see Appendix Table 5). In sum, we report a first supply side result:

Summary stats:

Table 5: Summary statistics by gender for drivers on Uber

	#Drivers	%Drivers	#Trips	%Trips	%Tipped	Mean Tip | Tip	Mean Tip
Male	513,410	78.0%	19,302,308	83.4%	0.153	3.083	0.470
Female	144,502	22.0%	3,843,859	16.6%	0.168	3.143	0.527

Regressions:

Table 6: Regression output for tip differences between male and female drivers.

	Dependent variable					
	Tip Amount					
	(1)	(2)	(3)	(4)	(5)	(6)
Female Driver	0.057 ***	0.057 ***	0.056 ***	0.047 ***	0.046 ***	0.048
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	0.47 ***					
	(0.0005)					
Controls/Fixed Effects						
Date		X	X	X	X	X
Hour of Week			X		X	X
Pick-up Geo				X	X	X
Drop-off Geo						X
Observations	23,146,167	23,146,167	23,146,167	23,146,167	23,146,167	23,146,167
R^2	0.0002	0.0003	0.001	0.020	0.020	0.028
Adjusted R^2	0.0002	0.0003	0.001	0.020	0.020	0.028
Residual Std. Error	1.388	1.388	1.388	1.375	1.375	1.369
df	23,146,165	23,146,152	23,145,991	23,140,106	23,139,945	23,133,162
Obs-df	2	15	176	6,061	6,222	13,005
Note						*p<.1; **p<.05; ***p<.01
#RHS Vars (w/ _cons)	2	15	176	6,061	6,222	13,005
@Margin		Date	Hr of Wk	PickUp Geo	PickUp Geo	DropOff Geo
# Vars @ Margin		13	161	6,046	6,046	6,944

OLS/SLR (Simple Linear Regression) Analysis - Looking @ Model (1)

SLR results:

Female Driver	0.057	estimated slope coefficient (**sign?** magnitude?)
	(0.001)	*p value* (**statistical significance?** *… std. error; t-stat*)
		*** indicates that $p < .01$ (see Note)
	0.47	estimated intercept (constant) coefficient
	(0.0005)	p value (statistical significance)
		*** indicates that $p < .01$ (see Note)

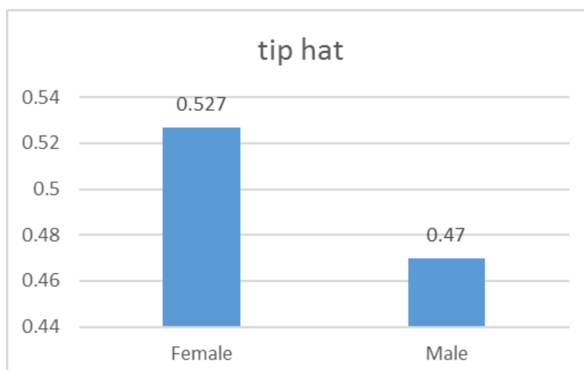
Observations	23,146,167	#trips in the model; why isn't this 46 million?
Goodness of Fit		
R^2	0.0002	*Coefficient of Determination*; % variation in the dependent variable *explained* by the model; always between 0 and 1; bigger is better
Adjusted R^2	0.0002	*Adjusted R-squared* (adjusted for # RHS variables); always less than R-squared; bigger is better
Residual Std. Error	1.388	aka *RMSE (Root Mean Squared Error)*… sort of an average residual (prediction error); smaller is better

Ordinary Least Squares: Estimate parameters by minimizing Sum Squared Residuals (SSRs)

- SLR: Data: $\{x_i, y_i\}$. Predict each y_i as a linear function of the single explanatory variable, x_i:

  ***actual**$_i$*: y_i, ***predicted**$_i$*: $b_0 + b_1 x_i$, given coefficients b_0 *and* b_1, and ***residual**$_i$*: $y_i - (b_0 + b_1 x_i)$
- OLS/SLR: Find the b_0 *and* b_1 that minimize $SSR = \sum (residual_i)^2 = \sum (y_i - (b_0 + b_1 x_i))^2$

Predicted SLR values: Sample Regression Function (SRF)



- $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$
- $\hat{y}$: predicted *tip amount*
- $\hat{\beta}_0$: estimated y-intercept, constant coefficient
- $\hat{\beta}_1$: estimated slope coefficient
- x: RHS variable… in this case *female driver* gender; a dummy variable (1: female driver; 0: male driver)
- SRF: $\widehat{\$tip} = \$0.47 + \$.057\ female\ driver$
  - $female\ driver = 1$: $\widehat{\$tip} = \$0.47 + \$.057 \cdot 1 = \$0.527$
  - $female\ driver = 0$: $\widehat{\$tip} = \$0.47 + \$.057 \cdot 0 = \$0.470$
  - Note that predicted values are also the sample means by driver gender.

Looking at estimated coefficients… focus on:

- signs (direction of estimated effects; Is $\hat{y}$ increasing (+) or decreasing (-) with increases in x?)
- statistical significance (explanatory power; *t stats*, *p values*, and *hypothesis testing*)
- point estimates v. interval estimates (*confidence intervals*)
- economic significance (*meaningfulness*; Are the critics impressed? Do they laugh at you when you brag about your results?)
- … and maybe magnitudes (But beware units of measurement impacts!)

And always worry/worry about #1 data integrity… and #2 omitted variable bias (endogeneity)!