Carter Sun
Devin Sze
Flora Zeng
Laurene Hui
Queeny Chan

Economic Drivers Final Report

Problem Statement

The last decade has seen a surge in the number of gig workers, indicating broad economic and demographic shifts. However, inflation might put pressure on the gig workers. Companies have to come up with strategies to deal with rising gas prices and unemployment rate. What are limits for a growth of a grocery delivery company and constraints in the current environment around its operations?

Background Research

The first constraint that we observed was universal with all grocery delivery operations: the perishability of their foods. The USDA reported that perishable foods like meat could only be kept at room temperature for two hours.[1] This places an upper bound on all deliveries by limiting the range of delivery to within a two-hour drivable distance. This perishability limitation is reflected in the delivery times for many of Shipt's competitors; for example, each trip an Instacart driver makes is a maximum of two hours.[2]

An additional constraint for grocery delivery is its customer base. Unlike countries like South Korea where there are a few large population centers that make distribution more centralized and easier to manage, the population distribution across the US makes it difficult to remain profitable when expanding into rural areas.[3] Another demographic concern is that many potential costumes are wary of e-commerce or lack the technical know-how. For example, only 9.5% of American adults have used a fresh grocery delivery service.[4] Combined with how software solutions are designed for able-bodied audiences,[5] software adjustments for ADA accessibility, instructional tutorials with step-by-step walkthroughs, and including niche stores or products in Shipt's selection are required to expand the customer base.

A third constraint is the issue of the rising number of competing grocery delivery services, such as InstaCart, Gopuff, Amazon Fresh, Uber Eats, etc. These competitors can largely affect Shipt's customer loyalty, especially when some of these companies already have existing well-established services (such as Uber and Amazon). Hence, customer retention as well as acquisition, become difficult without concrete incentives that motivate customers to stick to one delivery platform, especially since switching to alternate services is made incredibly easy. To mitigate this, customer loyalty programs could be implemented. Additionally, an overall

---

[1] USDA: https://ask.usda.gov/s/article/Is-food-safe-if-left-out-overnight
[2] Instacart: https://tech.instacart.com/how-instacart-delivers-on-time-using-quantile-regression-2383e2e03edb
[3] LinkedIn: https://www.linkedin.com/pulse/why-people-still-dont-buy-groceries-online-we-shop-almost-cabrera/
[4] Marist Poll: https://maristpoll.marist.edu/wp-content/misc/usapolls/us180423_NPR/NPR_Marist%20Poll_Tables%20of%20Questions_May%202018.pdf
[5] FoodInsight: https://foodinsight.org/survey-finds-that-few-older-americans-grocery-shop-online-but-lowering-hurdles-could-sway-many/

improvement in the platform's user experience through investing in UI/UX teams, could also result in higher customer satisfaction and thereby retention rate.

The last challenge we researched is grocery delivery companies struggle with staying on top of market trends. The market is always evolving as customer expectations are constantly changing. For example, fresh fruits and vegetables saw the highest growth in online grocery sales from 2020 to 2021 due to social distancing during the COVID-19 pandemic.[6] According to leading grocery retailer Kroger, grocery companies can expect to see a boost in the sale of food and beverage products that feature functional benefits that "support immune and cognitive health, digestion, energy levels, and stress management."[7] In order to stay competitive, it is important for Shipt to develop sustainable marketing strategies to adapt to market trends. Shipt can categorize their products as fast movers, which are products that are popular and have more demand, and slow movers, which are products that are more niche with less demand. Fast movers are already popular, so a little extra marketing on these products can retain a lot of new customers. Slow movers are important to market as well since these products can help distinguish Shipt's selection from competitors.[8] Shipt can consider marketing these products with reviews, photos, and short recipe videos uploaded by users that promote a specific product.

Analysis[9]

<u>Datasets</u>:

We searched for publicly-available datasets online to make appropriate analyses with. We determined that the biggest economic problem facing gig companies like Shipt is rising gas prices, which heavily affect costs incurred by Shipt's driver-partners. With that in mind, we found the following datasets and used them in our analyses:

1. uber-raw-data-janjune-15.csv[10]
    a. Each row is an Uber order in New York City between January and June 2015
    b. Contains data on how many orders were placed each day. Used as the outcome variable in our causal inference analyses and as a predictor in our ride pricing model
2. "uber fares".csv[11]
    a. Each row is an Uber order between 2009 and 2015
    b. Contains data on Uber ride prices. Used as a confounding variable in our causal inference analyses and as the outcome variable in our ride pricing model
3. "sf gas data".csv[12]

---

6 https://www.supermarketnews.com/online-retail/online-grocery-shopping-grows-amid-pandemic-induced-channel-stickiness

7 https://www.eatthis.com/top-grocery-trends-for-2021-according-to-leading-grocer/

8 Concept drawn from SCM II

9 The Jupyter notebook called "project.ipynb" containing the data cleaning and analyses are in the Dropbox

10 Kaggle: https://www.kaggle.com/datasets/fivethirtyeight/uber-pickups-in-new-york-city

11 Kaggle: https://www.kaggle.com/datasets/yasserh/uber-fares-dataset

12 EIA: https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=EMM_EPMR_PTE_Y05SF_DPG&f=M

   a. Each row is the weekly gas price in San Francisco from 2000 to present
   b. Used as the treatment variable in our causal inference analyses and as a predictor in our ride pricing model
4. inflation_expectation.csv[13]
   a. Each row is the monthly expected inflation rate in the US from 1978 to present
   b. Used as an instrument variable in our causal inference analyses and as a predictor in our ride pricing model
5. unemployment_rate.csv[14]
   a. Each row is the monthly unemployment rate in the US from 1948 to present
   b. Used as a confounding variable in our causal inference analyses and as a predictor in our ride pricing model

We cleaned the data for each of the CSVs (eg. extracting date features, grouping by day/month, etc.) and merged the data frames into a single table containing information from each of the datasets. We used this final table to conduct the following analyses.

Causal Inference Analysis:
Since rising gas prices can have a significant impact on the operations surrounding transportation-centered gig companies like Uber and Shipt, we wanted to investigate the causal effect of gas prices on the number of orders placed on their platforms. The following variables were used to set up the two stage least squares regression (2SLS), which is a method of determining the causal effect of a treatment variable on an outcome variable subject to potential confounding covariates:
1. Outcome variable: Total number of Uber orders per day
2. Treatment variable: Weekly SF gas prices
3. Confounding variables: Uber prices and the unemployment rate
4. Instrument variable: Weekly inflationary expectations

Key assumptions: Businesses only use the inflation rate displayed on the news in their decision-making process, not necessarily inflationary expectations. This assumption allows our instrument variable to not have a causal relationship with our confounding variables. Additionally, we assume that the relationship between the variables are generally linear, which enables us to use linear regression as our model.

Following the 2SLS algorithm, I first fit the weekly gas prices variable using our confounding and instrument variables to remove the biased variation explained by our confounders. In other words, the output is a newly-predicted vector of values for weekly gas prices, controlling for confounders.

[13] FRED: https://fred.stlouisfed.org/series/MICH
[14] FRED: https://fred.stlouisfed.org/series/UNRATE

Carter Sun
Devin Sze
Flora Zeng
Laurene Hui
Queeny Chan

Using the unbiased gas prices, I finally attempted to predict the number of Uber orders, controlling for confounders. This was done using linear regression with the unbiased gas prices, Uber prices, and the unemployment rate as predictors. The causal interpretation of gas prices on Uber orders from this method yielded a coefficient of 6633.3149. This suggests that a dollar increase in gas prices causes approximately 6633 more Uber orders to be placed. This could be due to a number of factors; for example, higher gas prices may dissuade people from driving themselves and instead take Uber rides so that they can pass on the gas costs to Uber drivers instead.

For comparison, a simple estimate using the original biased gas prices gives a causal estimate of about 14000. Since our 2SLS estimate is lower than the naive estimate, then there generally are upward biases that inflate the effects of gas prices on Uber orders.

One limitation of this study is that there are many potential confounding variables that exist between gas prices and Uber orders. Controlling for all of these variables would be extremely difficult and there are likely more biases that exist in my estimate even after applying the two stage algorithm.

<u>Uber Ride Pricing Model</u>:
Using the macroeconomic factors data identified as important to gig companies like Uber and Shipt, we next wanted to see to what extent these variables can be used to predict the daily average Uber prices. The following variables were used as predictor variables in our linear regression model:

1. Average passenger count per day
2. Number of rides per day
3. Monthly expected inflation
4. Monthly unemployment rate
5. Weekly SF gas prices

We first split the data frame into training and testing splits, with our test set containing 30% of the original data. We then fit a Scikit-Learn linear regression model using our training data. To prevent overfitting, we used the testing data to validate the root mean squared error (RMSE) of our model. The RMSE of our model is $1.70.

One limitation of our price prediction model is that our analyses are limited to 2015. Therefore, our findings are most accurately reflective of the gig economy environment in 2015, which may be different from those today; for example, the pandemic and the high inflation economy today may affect Uber prices differently. If we had data closer to the present, then we may be more confident that our model can translate to the current world.

Carter Sun
Devin Sze
Flora Zeng
Laurene Hui
Queeny Chan

Recommendations

We believe improvements should be invested into our analyses on order pricing and the causal effect of gas prices on orders since they can be used to inform decisions on delivery operations, which affects much of Shipt's business model.

One such improvement would be to collect more granular data which pertain specifically to Shipt's business operations. We had to source our Uber data using only what was publicly available, which wasn't enough to make an adequate pricing model with. More data on confounding variables can also be identified and scraped to account for some of the biases in our causal effect study. Which other confounders to control for can be identified using **cause-effect diagrams**[15] so that we can find which variables are adequately correlated with one another. Additionally, other models, such as neural networks, may fit nonlinear trends in the data better and should be experimented with.

Otherwise, Shipt can also design their own randomized control experiment and **A/B test**[16] varying levels of gas subsidies to determine its causal effect on the number of orders on their platform. This has the advantage of controlling for all biases besides the treatment variable, assuming the sample size is large.

For the Uber Fare Amounts line plot, a strategy that can be used to flatten the peaks would be to increase capacity for new users also known as **network effects**[17]. Therefore, when we have more Uber users/passengers, rather than increasing the fare prices, we increase the number of drivers. This way we are able to retain our old users as well as increase the amount of new users, which benefits the company with a growing business.

[15] Concept drawn from Quality II
[16] Concept drawn from Strategy III
[17] Concept drawn from Strategy II

Carter Sun
Devin Sze
Flora Zeng
Laurene Hui
Queeny Chan

Exhibit 1: Datasets

	Dispatching_base_num	Pickup_date	Affiliated_base_num	locationID
2	B02617	5/17/2015 9:47	B02617	141
3	B02617	5/17/2015 9:47	B02617	65
4	B02617	5/17/2015 9:47	B02617	100
5	B02617	5/17/2015 9:47	B02774	80
6	B02617	5/17/2015 9:47	B02617	90
7	B02617	5/17/2015 9:47	B02617	228
8	B02617	5/17/2015 9:47	B02617	7
9	B02617	5/17/2015 9:47	B02764	74
10	B02617	5/17/2015 9:47	B02617	249
11	B02617	5/17/2015 9:47	B02764	22
12	B02617	5/17/2015 9:48	B02617	263
13	B02617	5/17/2015 9:48	B02617	61
14	B02617	5/17/2015 9:49	B02617	229
15	B02617	5/17/2015 9:49	B02617	164
16	B02617	5/17/2015 9:49	B02617	237
17	B02617	5/17/2015 9:49	B02617	142
18	B02617	5/17/2015 9:49	B02617	188
19	B02617	5/17/2015 9:49	B02617	237
20	B02617	5/17/2015 9:49	B02617	224
21	B02617	5/17/2015 9:49	B02617	238

uber-raw-data-janjune-15

uber-raw-data-janjune-15.csv

		key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_latitude	passenger_count
2	28165919	40:39.0	27	2015-06-30 23:40:39 UTC	-73.98479462	40.75141144	-73.92776489	40.70628738	1
3	15173476	33:33.0	8.5	2015-06-30 23:33:33 UTC	-73.98098755	40.76202011	-73.96008301	40.7705307	1
4	22269752	31:06.0	20	2015-06-30 23:31:06 UTC	-73.99996185	40.73313522	-73.96244812	40.77304077	4
5	44430308	16:42.0	25.5	2015-06-30 23:16:42 UTC	-74.00109863	40.73096085	-73.9571228	40.80690765	2
6	37789542	57:53.0	18.5	2015-06-30 22:57:53 UTC	-73.97170258	40.78220749	-73.94367981	40.82799149	2
7	48165790	43:48.0	6	2015-06-30 22:43:48 UTC	-73.97847748	40.78305054	-73.97008514	40.79961395	5
8	2336540	39:38.0	6	2015-06-30 22:39:38 UTC	-73.98280334	40.7509079	-73.99261475	40.73732376	1
9	13023720	27:28.0	4.5	2015-06-30 22:27:28 UTC	-73.98415375	40.74900818	-73.97805023	40.75159836	1
10	7476774	09:32.0	9	2015-06-30 22:09:32 UTC	-74.00530243	40.74579239	-73.98091125	40.73075104	1
11	26339509	55:11.0	11	2015-06-30 21:55:11 UTC	-73.96551514	40.71360016	-73.9602356	40.68661499	1
12	22788649	46:23.0	8	2015-06-30 21:46:23 UTC	-73.99742126	40.72578049	-74.01216125	40.71258926	1
13	23535251	39:29.0	9.5	2015-06-30 21:39:29 UTC	-73.98246765	40.77226639	-73.97678375	40.79032516	2
14	6789139	02:16.0	13.5	2015-06-30 21:02:16 UTC	-74.01002502	40.71207047	-73.98612213	40.7558136	4
15	32020814	46:20.0	6	2015-06-30 20:46:20 UTC	-73.93757629	40.80451965	-73.93860626	40.81261826	4
16	20222977	46:03.0	7	2015-06-30 20:46:03 UTC	0	0	0	0	4
17	12764580	39:11.0	6	2015-06-30 20:39:11 UTC	-73.97572327	40.78689194	-73.96601868	40.80399323	1
18	41015040	05:47.0	8.5	2015-06-30 20:05:47 UTC	-73.95531464	40.80456161	-73.94232178	40.82852936	1
19	41448516	47:07.0	4.5	2015-06-30 19:47:07 UTC	-74.00789642	40.71228409	-74.01365662	40.70231628	1
20	49946638	38:21.0	20.5	2015-06-30 19:38:21 UTC	-73.96788788	40.79241562	-74.00370789	40.73246384	1
21	44063474	34:38.0	8.5	2015-06-30 19:34:38 UTC	-73.99794769	40.72594833	-73.97704315	40.7338295	1

uber fares

“uber fares”.csv

3	Date	Weekly San Francisco, CA Regular Reformulated Retail Gasoline Prices (Dollars per Gallon)
4	Jun 05, 2000	1.759
5	Jun 12, 2000	1.766
6	Jun 19, 2000	1.769
7	Jun 26, 2000	1.785
8	Jul 03, 2000	1.836
9	Jul 10, 2000	1.861
10	Jul 17, 2000	1.865
11	Jul 24, 2000	1.866
12	Jul 31, 2000	1.852
13	Aug 07, 2000	1.843
14	Aug 14, 2000	1.836
15	Aug 21, 2000	1.856
16	Aug 28, 2000	1.88
17	Sep 04, 2000	1.928
18	Sep 11, 2000	1.972
19	Sep 18, 2000	1.966
20	Sep 25, 2000	1.963
21	Oct 02, 2000	1.952
22	Oct 09, 2000	1.956

Contents Data 1

“sf gas data”.csv

Carter Sun
Devin Sze
Flora Zeng
Laurene Hui
Queeny Chan

DATE	MICH
1/1/1978	5.2
2/1/1978	6.4
3/1/1978	6.3
4/1/1978	6.7
5/1/1978	6.9
6/1/1978	6.5
7/1/1978	6.6
8/1/1978	8.7
9/1/1978	6.9
10/1/1978	7.4
11/1/1978	7.5
12/1/1978	7.3
1/1/1979	7.8
2/1/1979	9.3
3/1/1979	8.8
4/1/1979	9.7
5/1/1979	9.8
6/1/1979	9.9
7/1/1979	9.9
8/1/1979	9.9

inflation_expectation

inflation_expectation.csv

DATE	UNRATE
1/1/1948	3.4
2/1/1948	3.8
3/1/1948	4
4/1/1948	3.9
5/1/1948	3.5
6/1/1948	3.6
7/1/1948	3.6
8/1/1948	3.9
9/1/1948	3.8
10/1/1948	3.7
11/1/1948	3.8
12/1/1948	4
1/1/1949	4.3
2/1/1949	4.7
3/1/1949	5
4/1/1949	5.3
5/1/1949	6.1
6/1/1949	6.2
7/1/1949	6.7
8/1/1949	6.8

unemployment_rate

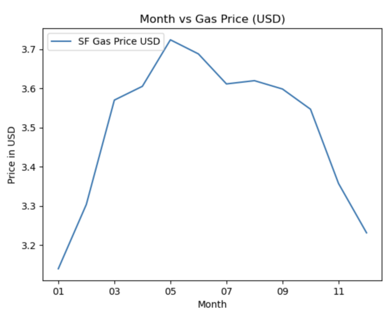
unemployment_rate.csv

Carter Sun
Devin Sze
Flora Zeng
Laurene Hui
Queeny Chan

Exhibit 2: 2SLS Statsmodels Output

	coef	std err	z	P>|z|	[0.025	0.975]
const	2.238e+05	1.1e+05	2.033	0.042	8062.725	4.4e+05
unemployment rate	-3.123e+04	1.58e+04	-1.978	0.048	-6.22e+04	-287.167
fare_amount	391.6351	1063.185	0.368	0.713	-1692.168	2475.438
gas price hat	6633.3149	1.1e+04	0.603	0.547	-1.49e+04	2.82e+04

Omnibus:	2.178	**Durbin-Watson:**	0.753
Prob(Omnibus):	0.337	**Jarque-Bera (JB):**	1.760
Skew:	0.192	**Prob(JB):**	0.415
Kurtosis:	3.294	**Cond. No.**	1.46e+03



We created two line plots to display how the gas prices and uber fare prices change over the months. From the Month vs Gas Price line plot, we observe a peak during the month of May which describes **peak loading pricing**[18]. This is likely due to an increase in travelers both by plane and car, causing a rise in gas prices. Similarly, from the the Month vs. Uber Price line plot, we observe two peaks in the Uber fare amounts in May and September which is likely due to an increase in passengers and not enough drivers. Overall, from looking at the two plots, we conclude that during the Summer months, we have more travelers and people using public transportation which likely caused the peak in gas and uber prices.

[18] Concept drawn from Queue I