A Probability Modelling Approach To Assess COVID 19: A Instacart Case Study

STAT 476 Project 2

I. Introduction

Executive Summary

The COVID-19 global pandemic caused a shock to the world. This impact extends from the frontlines of a battle for global health to daily activities that were formerly taken for granted. In particular, the subsequent policies and self-isolating measures taken by the vast population in the face of COVID-19 brought about an unprecedented change in consumer behavior and demand for online based food/grocery delivery softwares.

Problem Statement

In line with those events, this paper **aims to use adoption-based probability models to quantify and, in particular, breakdown such changes in order to answer critically important questions** that may be asked by the public, the companies and the marketers.

Data Source and Description

The main data utilized for this analysis is the daily downloads of the popular online grocery delivery app, Instacart. The data is tracked and supplied by Apptopia, a service that specializes in application tracking. The data at hand consists of Instacart app downloads over 41 days from February 15, 2020 to March 26, 2020. This data is also further supplemented by data gathered from Washington Post on their daily COVID 19 cases updates on the 50 states, the nuance of such addition will be discussed shortly.

Project Goals

This project overall has three main goals:

1. **Accurate Model Creation**
   This project aims to specify a model architecture that will most effectively capture the underlying drivers pushing Instacart downloads during the COVID-19 crisis. This section will include the underlying stories behind the models, the model assumptions, the parameter choices, and the model evaluation/performance.

2. **Quantitative Impact Assessment**
   Using the aforementioned model, this project aims to offer immediates and extended inferences as to the extent of changes that COVID-19 has brought upon Instacart downloads. This section will focus on a number of key metrics and answer a number of key questions that need to be addressed.

3. **Customer Base Exploration**
   Capitalizing on the properties of probability models, this paper will also seek to analyze the underlying customer base of instacart including the customer polarity, customer segment and overall customer base size.

Problem Statement

As indicated, most parts of this project will proceed under the assumption that February 15, 2020 is the first day of instacart's product launch instead of treating the adoption process as a left censored one. Furthermore, the panel size of this dataset is set to be 120 million, which is roughly equivalent to the number of households in the United States with internet access.

Furthermore, this is clearly a timing adoption dataset, whereby the time is in fact continuous. However, for the purpose of this analysis and since continuous timing model is really the generalization of discrete timing model, I will still refer to the underlying propensity to adopt using the coin analogy as it is most straightforward. Note that for this case, every coin toss doesn't occur in discrete time units like geometric, but instead it happens at infinitesimally small time slices (practically every mini second, thus virtually continuous in nature). In case of a simple exponential, the lambda estimated is expected to have a direct relationship with the underlying coin, since a higher lambda indicates a faster adoption time and thus a higher propensity.

II. Hypotheses

The Story Today

Prior to building the model, I want to first qualitatively lay out the kind of impact that we'd expect COVID to have on Instacart downloads, then later convert these qualitative hypotheses and conjectures into the model.

Examples of Instacart Customer Stories

1. **Hypothesis 1**

   "Amanda is a Mom who lives with her family in the suburbs; she is a frequent shopper of groceries, and while she knows about Instacart, she previously found it too wasteful to download and an excessive use of money on groceries (i.e., limited choice, delivery fee). However, when COVID started storming in, Amanda, being a good mom, realized that going to the grocery store could threaten her health and potentially even cause her kids' health to be in jeopardy. In order to protect herself and her family, she decided to finally download Instacart."

   COVID-19 and its media exposure have caused a greater need and emphasis for self-protection. That is to say, due to COVID-19, people are less inclined to go to supermarkets and risk getting the disease and will have greater reliance on delivery services. With more cases of COVID-19 being seen by people who look at broadcast and social media, the propensity to download apps such as Instacart will likely increase, thereby accelerating the adoption cycle.

2. **Hypothesis 2**

   "Rose is in her 20s and lives by herself in New York. She, much like other white-collar workers in their 20s, tends to eat out at restaurants often. Being a Gen Z, she is familiar with Instacart, so far as to say that it is a grocery delivery app, but never found a really compelling need to download it. When COVID-19 struck, the local government issued a warning and declared a State of Emergency, causing dine-in restaurants to start closing slowly one after another. Rose suddenly didn't have the option to eat out anymore, so she decided to look into some delivery options, and while checking out typical food delivery apps like Uber Eats and GrubHub, she also decided to download Instacart to potentially try out cooking or buying frozen, reheatable foods."

   Due to state sanctioned action around COVID 19, people who have previously known but were on the fence about downloading Instacart (perhaps just not cooking much at home in general) are now more compelled to download due to quarantine and closure of businesses such as restaurants. While also an increase in propensity, this increase is independent from that of hypothesis 1, as it is not one driven by fear or emotional stimuli, but one driven by diminishing alternatives.

3. **Hypothesis 3**

   "Jack is in his 40s and has been working as a mechanic for most of his life. He is not the most familiar with trendy technologies and has actually never heard of Instacart before. COVID-19 certainly disrupted his life as he could no longer leave his house to get groceries or food. For a couple of days, he went around asking some friends and his daughter, who just came back from college, to see if they had any suggestions to get groceries. Someone recommended downloading the app Instacart, so he decided, why not?"

   The aforementioned closures and the saturated grocery market may have also converted individuals who previously were never in the market of Instacart. These may be people who lived very close to grocery stores and may have had other subscriptions such as Amazon Fresh, for instance. The crucial point here is that this does not concern the propensity to download, but instead the size of the capturable market there is.

These hypotheses may not be mutually exclusive, but instead are likely complementary and need to be taken into account when building up the model. In this project, I sought after an iterative approach whereby each of these key hypotheses were considered and built upon to the eventual final model.

III. First Approach

A 3-Story Approach

In this section, I will go through the 3 main stories that sequentially reflect and consider the hypothesis mentioned above. For each model, a brief discussion of parameters will be made; however, more discussion will follow after the final specification is being decided. For the models in this project, the estimation was done by minimizing over the in-sample Mean Absolute Percentage Error. I found this method to be more stable and almost always lead to the smallest in-sample log likelihood as well. Furthermore all model fitting done here is fitted without the last week of data. The last week's data is then used as a holdout to evaluate the model performance.

Purchase Acceleration

To bring in Hypothesis 1 and 2 does not require much of a complicated model. It is instead evident in the phrasing that both hypothesis 1 and 2 refer to an increase in propensity as a reaction to some externalities. In both these hypotheses, we are referring to individuals who have been spinning their coin, waiting to adopt, and their propensity increasing either due to fear/self protection or necessity in the face of quarantine.

This naturally corresponds to a covariate based story. However a couple of key components that concern heterogeneity and purchasing behavior must be settled prior to building the model. In particular the selection of those building blocks for me follow two main criteria. 1. Can such a specification ever make sense (can there be a justifiable story). 2. Does the specification have a strong enough effect (does the model pick up the specification). Thus instead of simply going straight to model building with every possible specification and throwing about pieces that models reject, I would first want to examine which specification actually makes sense to begin with.

- **Covariate Choices**
  Certainly, the most important building block of a covariate model will be the covariate itself. For this analysis, two covariates were chosen to support the first two hypotheses mentioned above.

  For the first hypothesis, a covariate is needed to reflect the rising need for self-protection that people may feel due to the increase in COVID-19.

  Given that people generally obtain self-quarantine information and coronavirus situation updates from mainstream media, I decided to scrape the daily COVID count update of all 50 states from the Washington Post in order to capture the pandemic's severity. However, as opposed to directly including this covariate as is, a couple more subtleties were considered.

  People's action on day T rarely depends on the news at day T, especially because COVID case counts really only become solidified by the end of day. Instead, I added a one day lag in the cases to accomodate for such delay in information.

- **Capturing Memory Span**
  People also don't just respond to yesterday's news, but instead, people's memory captures a longer span than that when making a decision. However, for COVID, unlike other regular exponential moving average based cases, the information in the day beforehand is likely going to be orders of magnitude more important than the preceding days and the look back window is rather small. Thus I added a short term smoothing factor that brings the covariates value at time T to be equal to COVID Cases at: **T - 1 * 0.7 + COVID Cases at T - 2 * 0.2 + COVID Cases at T - 3 * 0.1.** The specific factor used was an estimate based on personal experience and some minor tuning on the data.

- **Log Transformation**
  Lastly, a log transformation is used on the cases variable, as an increase of 100 cases would mean significantly more when there are only 20 cases versus when there are 20000 cases. Since the perception of such an increase is likely relative to the total number of cases, a log nicely captures such a proportional nature of evident increases.

IV. Approach Continued

Second Hypothesis

For the second hypothesis, a covariate that reflects some state based action would be extremely helpful in reflecting such a necessity. Personally, I believe while more direct policies such as restaurant closures and stay at home orders need to be taken into account, the propensity and necessity of people have likely begun to increase as soon as state of emergencies/school closures are declared within each state. While these declarations do not directly impose a quarantine upon individuals, more people have begun their self-motivated quarantines as a result of the declaration.

To accomodate all that, I manually scraped the dates at which each state issued either a (state of emergency, restaurant closure or stay at home order), and added a cumulative counter for all these policies combined (If state A issued a state of emergency today, and state B issued a dine in restaunrant closure at day T, the counter for day T will be 2, if state A issued a stay at home order at day T+1, the counter for T + 1 will be 3). The covariate is a cumulative sum since after any of such an action is imposed, the effect is likely going to and assumed to last until the policy ends. That is to say, while there is a stay at home order, people are likely going to continually have a higher propensity to download, up until such an order gets relinquished.

The assumption for a typical covariate in the timing model still holds, such that I expect the covariate effect to be the same throughout the day.

- **Mixture Models vs Homogeneous Models**
  The heterogeneity of the user base is entirely a reasonable specification to test for this model given that there is no reason to reject possible differences in the underlying customers.

- **HCNB**
  Given that there exist people unfamiliar with Instacart and competitors such as Amazon Fresh, it is entirely possible that not all of the 120 million households with internet access will be subscribers of Instacart. The existence of Hard Core Never Downloaders in this case needs to be accommodated and tested as well.

- **Duration Dependence**
  Duration dependence in this case demands an action to be more or less probable purely and solely due to the passage of time. This begs the question, will someone be more/less willing to download Instacart just because they didn't download yesterday? To me, this answer is largely no, especially because we are likely way past the novelty phase of the product (while we are assuming the product is just being rolled out this is a key characteristic, but I don't think we can assume that for this analysis). While, in reality, the duration dependence may capture certain effects (network effect or the partial effect of the covariate) and induce an artificially better fit, it is not likely something baked within the downloading calculus when someone is approaching Instacart and is therefore not something that can be generalized in the future. Since I don't think there is a justifiable story and generalizability behind this specification, I ultimately decided to not include it.

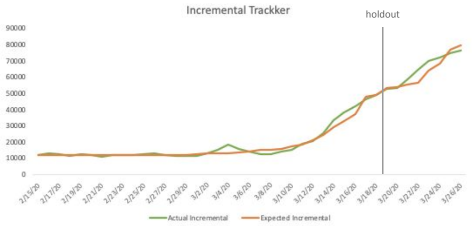
V. Model Results

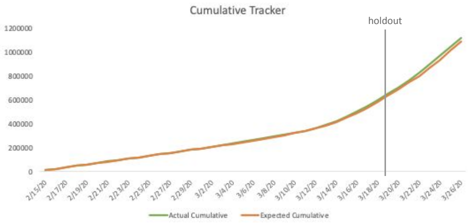
IS	In Sample
OS	Out of Sample

Story 1: Purchasing Acceleration

Model Type	# Of Parameters	IS Incremental MAPE	OS Incremental MAPE	IS Cumulative MAPE	OS Cumulative MAPE	IS Log Likelihood	BIC	Note
EG + HCNB + 2 Covariates	5	0.07	0.06	0.01	0.03	-6256158.7	12512410.5	Continous Heterogeniety with HCNB and Both Cov
EG + 2 Covariates	4	0.07	0.06	0.01	0.03	-6256158.7	12512391.9	Continous Heterogeniety with Both Cov
E + 1 Covariates (Case)	3	0.18	0.66	0.05	0.33	-6325389.1	12650815.4	Homogeniety with US Cases Cov
E + 1 Covariates (State)	2	0.07	0.06	0.02	0.04	-6256390.5	12512818.3	Homogeniety with State Action Cov
E + 2 Covariates	2	0.07	0.06	0.02	0.03	-6256164.8	12512385.3	Homogeniety with Both Cov

Story 1 Final Model	
lambda_1	0.00010102
B_case	0.00735046
B_state	0.01704886





Iterative Nested Model Results

An iterative nested model is being constructed from the full model above with all the following specifications with the exception of duration dependence, the key model results are listed above.

In the spirit of balancing parsimony with the in-sample and out-of-sample fit, the final model used is the 3 parameter Homogeneous Exponential model with both covariates. The in-sample and out-of-sample, incremental and cumulative tracking plot looks as displayed above.

To quickly discuss the model parameters, it seems like the customer base is much more homogenous than otherwise expected. The coefficients attached with both covariates are positive, much as expected. HCNB in this case is strongly rejected by the end.

VI. Second Approach

HCNB Turned Adopter

While the fit for the first model is rather good, it certainly feels like it's missing a piece of the story. However, given that HCNB is rejected, does that mean hypothesis 3 is untrue? Personally, I don't believe so. Instead, while I don't think there may be HCNB when spanning through the horizon, I have indeed witnessed quite a bit of individuals, including my family members that had never heard of Instacart before, adopt Instacart after the start of the COVID-19 pandemic. This interesting previously HCNB turned buyer story is something that I believe will be quite fascinating to experiment with.

Such a story like the one mentioned above can easily be translated into a delayed Latent Class model, whereby one segment (let us call them the original buyers) starts being part of Instacart's capturable market since the very first day, while the other segment (let us call them the new adopters) has a 0 propensity of purchase in the beginning but enters the market at a later time.

Much like the earlier model, some technicalities must also be resolved before proceeding to finally construct the model.

- **Covariate Choices**
  The same Covariate will be used for this model as the purchase acceleration model above. Given that now there are two segments, it is entirely possible that they may respond to the covariates in a different manner (for instance, the original adapter may experience a slower acceleration of their adoption process compared to the new adopter since they were already in consideration of downloading Instacart). In order to accomodate for such a possibility, 4 coefficients will be used at the onset to estimate the Covariate effect.

- **Latent Class Exponential vs Latent Class EG**
  Similar to the previous model as well, it is hard to say a priori that there will be no heterogeneity within either of the Latent Classes. Thus, again to accomodate for such possibility, a Latent Class EG will be used in the initial construction of the model.

- **HCNB & Duration Dependence**
  Similar reasoning as the first model - given the competition and alternatives, it is possible for never converted HCNB to exist. Due to lack of coherent story and generalizability, Duration Dependence is excluded.

- **New Adopter Latent Class Entry Date**
  Given the above analysis, unlike individuals who are already in the market segment, the new adopters are assumed to be unaware and truly beyond the reach of Instacart in the very beginning. Thus, this segment is likely to have entered not immediately after, for instance, the first state of emergency has been declared. For the purpose of this analysis, I set the market entry date to be March 10, 2020, which is roughly one week after the first state of emergency was declared and roughly around the time when universities started suspending operations. This gives ample time for households to receive new information as family members who don't typically live at home start to come home, but also for people to possibly engage in online searches to gain insights about the product.

VII. Model Results

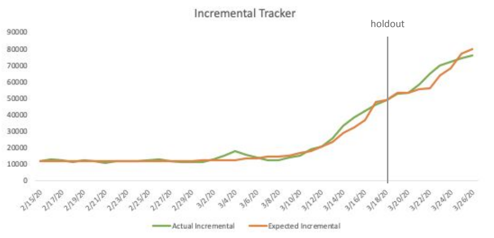
IS	In Sample
OS	Out of Sample

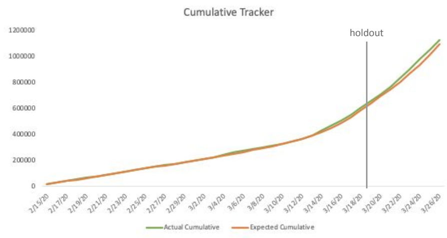
Story 2: HCNB turned Adopters

Model Type	# Of Parameters	IS Incremental MAPE	OS Incremental MAPE	IS Cumulative MAPE	OS Cumulative MAPE	IS Log Likelihood	BIC	Note
LC + EG + HCNB + 2 Covariates	10	0.07	0.06	0.02	0.03	-6256303.1	12512792.2	Heterogenous LC with HCNB and Both Cov
LC + EG + HCNB + 1 Covariates (State)	8	0.07	0.06	0.02	0.03	-6256346.4	12512841.6	Heterogenous LC with HCNB and State Action Cov
LC + EG + 1 Covariates (State)	7	0.07	0.06	0.02	0.03	-6256346.4	12512823	Heterogenous LC with State Action Cov
LC + E + EG + 1 Covariates (State)	6	0.07	0.06	0.02	0.03	-6256348.8	12512809.2	One Homogenous and One Heterogenous LC with State Action Cov

Story 2 Final Model

lambda_1	0.00010796
alpha_2	464121.733
r_2	0.0001
B_state1	0.01766441
B_state2	0.04668801
pi	0.94170577

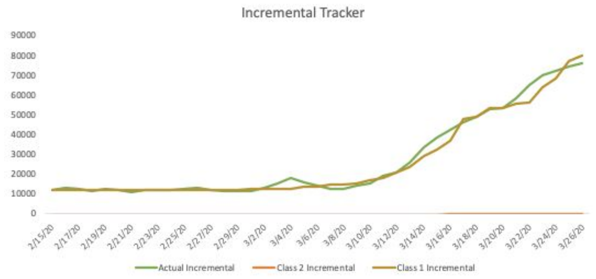


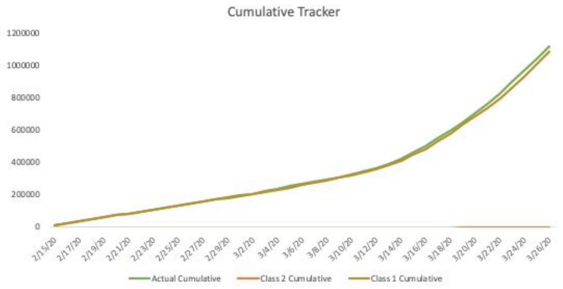


Iterative Nested Model Results

The model was once again built iteratively, with the results seen above.

After once again simplifying the model, the final mode chosen was a 6 parameter model with the above incremental and cumulative tracking plot.It would be extremely interesting to break down this model into both segments to understand how each of them have contributed to the overall curve:





Interestingly upon breaking them down, it seems as if the new adopter segment practically didn't contribute to the overall curve at all. The finite mixture even seems to be rejected; does not mean no one is really converted? Or perhaps we are making a wrong assumption? Investigating the initial graphics above regarding the capturable market, it is evident that underlying the delayed Latent Class is that it will only likely fit nicely if there exists a natural and immediate spike in adoption, which doesn't exist in this dataset. But it also begs the questions, should we expect such a spike for Latent Class?

VI. Third Approach

Gradual New Wave

In the implementation of the previous model, we are assuming that on March 10, everyone who used to be a HCNB suddenly all started spinning the coin at the same time. There is no graduality nor variance, but instead the whole Latent Class gets converted right away. This next model intends to build upon the previous model, except it integrates a smoothing element that allows the Latent Class to be integrated gradually as opposed to immediately into the capturable market.

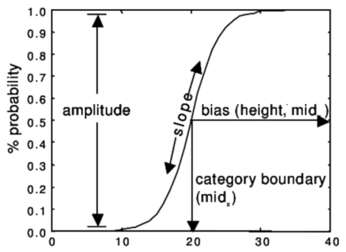
This idea is conceptually sound since echoing our example earlier, there may exist more than one Jack (from story 3) in the world, with varying circumstances. For instance, maybe one individual in the New Adopter Latent Class lives in Washington (which was hit earlier by coronavirus) and is therefore likely to have entered the market earlier, while another may have been in New York and thus may have entered during the peak season as the situation differed in both places. The graduality in some way approximates the heterogeneity in the situation of individuals in the Latent Class and avoids the immediate spike problem that existed before. Most of the technicalities of this model will follow the HCNB turned adopter model with the nuanced addition of a Latent Class Smoothing.

- **Latent Class Smoothing**
  The Latent Class Smoothing is done with the help of a truncated logistic function/sigmoid where the x-axis follows a passage of time while the y-axis refers to the proportion of Latent Class introduced. Unlike a full logistic function where the domain goes to negative infinity, the function in this case is truncated and will only take on positive domain value (time can only be positive). There are two things controlling the logistic function: first, the smoothing factor (slope) that controls the curvature of the logistic function, and second, the location factor that controls the median of the logistic function. The amplitude factor seen below is set to 1 and the height is set to 0.5 as it is equibiased. The parameters of the logistic function are also to be estimated along with the rest of the parameters to obtain the optimal.

Mathematical Function

$$\frac{\dfrac{1}{1+e^{-smooth(T-location)}}}{1-\dfrac{1}{1+e^{-smooth(0-location)}}}$$

Logistic Curve



The logistic function is used since it allows for an S-shaped curve, which is what I hypothesized for the Latent Class inclusion in reality. This is mainly because it is very much likely that the Latent Class will first enter in rather slowly (in the first states being hit) and later vastly accelerate as more states follow, then die down again as fewer states are left unaffected. With a logistic function in place, the capturable market graph will look more analogous to this as opposed to the one above.

The assumption here in the smoothed latent class is that every single day, a portion of the Latent Class is smoothed over and starts their purchases, much like the effect of the covariates. Moreover, while the Latent Class is being introduced gradually, I imposed that they are still the same segment characterized by the same underlying parameter.

However, even with such a simplifying assumption to fully implement the logistic function along with the Latent Class is trickier than expected due to the inclusion of the covariates. Effectively, in practice, by allowing for smoothing, I am essentially introducing “one Latent Class” at every time period since March 10th, 2020.

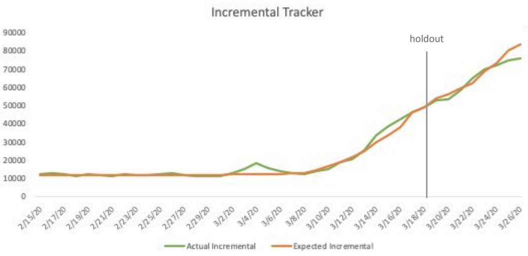
VIII. Model Results

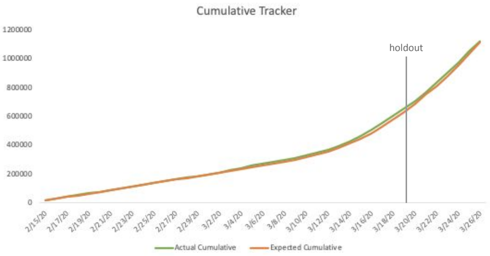
IS	In Sample
OS	Out of Sample

Story 3: Gradual Latent Class

Model Type	# Of Parameters	IS Incremental MAPE	OS Incremental MAPE	IS Cumulative MAPE	OS Cumulative MAPE	IS Log Likelihood	BIC	Note
LC + EG + HCNB + 2 Covariates + Smooth	12	0.06	0.05	0.03	0.02	-6255990.4	12512204	Smoothed Heterogenous LC with HCNB and Both Cov
LC + EG + HCNB + 1 Covariate (State) + Smooth	10	0.06	0.05	0.03	0.02	-6255997.4	12512180.7	Smoothed Heterogenous LC with HCNB and State Action Cov
LC + EG + 1 Covariate (State) + Smooth	9	0.06	0.05	0.03	0.02	-6255997.4	12512162.3	Heterogenous LC with State Action Cov
LC + E + 1 Covariate (State) + Smooth	7	0.06	0.05	0.03	0.02	-6255999.1	12512128.4	Smoothed Homogenous LC with State Action Cov

Story 3 Final Model	
lambda_1	0.00011676
lambda_2	0.01666666
B_state1	0.00485899
B_state2	0.012102
pi	0.86637465
smoothed	0.0112772
location	10.0009374





Thus, to calculate the probability of the entire New Latent Class purchasing before time T will require examining the probability of each of the smaller portions of the Latent Class performing said action. For example, to calculate the probability of the entire Latent Class purchasing before day 2, I will have to look at class A, which is, say, 10% of the Latent Class, that enters the market segment in day 1, and class B, say 15% of the latent class, that enters the market segment in day 2. From there, I'll have to extract the corresponding probabilities of both of these subclasses with their corresponding covariates at day 2, and sum these probabilities together weighted by their corresponding percentage (i.e., 10% for class A, 15% for Class B). The overall probability of the New Adopter Latent Class purchasing before time T will then have to summed across the probabilities of each of these segments purchasing before time T. The exact implementation of this case can be seen through the graphics below.

Smoothing Model Results

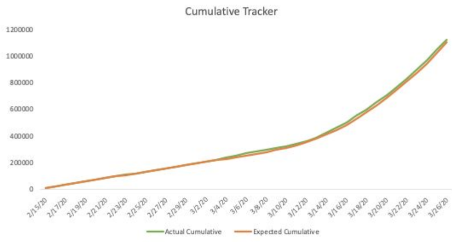
Using the smoothing introduced along with the covariates and other factors described in story 2, the model was once again built iteratively, with the results seen above.

After once again simplifying the model, the final model chosen was a 6 parameter model with the above incremental and cumulative tracking plot.

IX. Model Final Selection

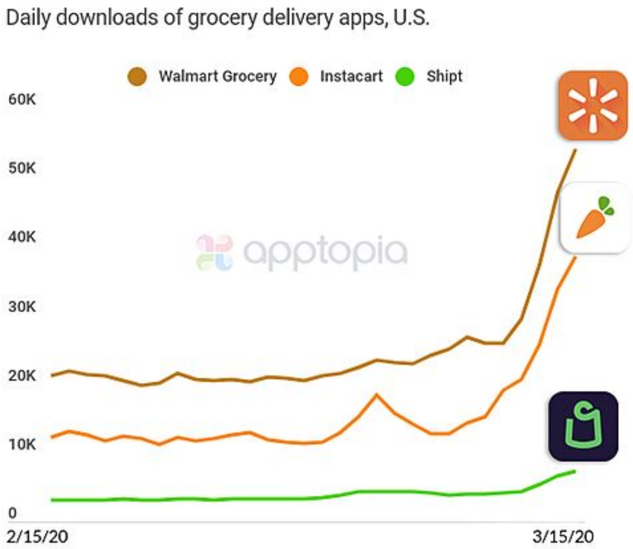
Model Type	# Of Parameters	IS Incremental MAPE	OS Incremental MAPE	Compare IS Cumulative MAPE	OS Cumulative MAPE	IS Log Likelihood	BIC	Note
Story 1: E + 2 Covariates	2	0.07	0.06	0.02	0.03	-6256164.8	12512385.3	Homogeniety with Both Cov
Story 2: LC + E + EG + 1 Covariates (State)	6	0.07	0.06	0.02	0.03	-6256348.8	12512809.2	One Homogenous and One Heterogenous LC with State Action Cov
Story 3: LC + E + 1 Covariate (State) + Smooth	7	0.06	0.05	0.03	0.02	-6255999.1	12512128.4	Smoothed Homogenous LC with State Action Cov

Overall Final Model	
lambda_1	0.00011276
lambda_2	0.02165518
B_state1	0.00461277
B_state2	0.01046687
pi	0.89658751
smoothed	0.01327621
location	10.0011653



Model Type	# Of Parameters	Incremental MAPE	Incremental R^2	Cumulative MAPE	Cumulative R^2	IS Log Likelihood	BIC	Note
Story 3: LC + E + 1 Covariate (State) + Smooth	7	0.06	0.99	0.02	1.00	-10236590	20473310.5	Smoothed Homogenous LC with State Action Cov

Comparing and contrasting the final model, the Latent Class Smoothing model seems to provide the best in-sample LL, BIC, and also OOS performance. For the subsequent analysis I will be using the Story 3 model train on the whole dataset. The final parameter estimates and performance metrics is as shown above.



Deal with Small Spike

With the multiple iterations of the model, the recurring problem seems to be the small spike around March 3rd to 4th. After reviewing more articles online, I notice that this trend is unique to Instacart as opposed to seen across all grocery delivery apps, like Amazon Fresh; thus, I believe this is not something that can be explained by macro-level covariates that in theory is supposed to impact all retailers at once. However, even after searching through the social media accounts of Instacart, I did not notice any aggressive marketing being done around this time frame; thus, I do think this is likely an artifact of chance as opposed to a capturable trend.

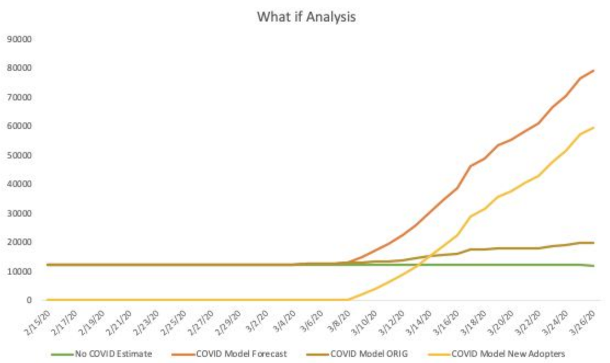
According to Final Model, what is the customer himself/herself like?

With the final smoothed homogenous Latent Class model, close to 90% of the customers fall within the original customer basket, while 10.3% fall within the new adopters. While they are both rather homogenous according to our model, they have vastly different propensities in terms of adoption. In particular, the new adopters have a lambda that is 192 times the size of the original segment, with a mean adoption time around 40 days as opposed to 8800 days. Both segments tend to be affected by state sanctioned actions, with the new adopters slightly more affected. Both phenomena could be due to a number of reasons, including the new adopters being more driven by necessity and thus more responsive when the need comes (higher initial rate and greater increase with more state action), or because the new adopters have only recently been exposed to the product and are thereby converting at a much faster pace.

X. Impact Analysis

Immediate Impact of COVID

Here I am moving on to the immediate impact assessment of COVID. In particular for this section, I seek to answer a couple of questions:



How many more adopters did we get due to COVID with the time span of Feb 15th to March 26th?

Comparing the model projections without the additional Latent Class and covariates and the model projection with those factors, COVID-19 seems to have pushed a grand total of 600k more downloads than if it weren't to have existed. The total downloads projected by the model with COVID is around 110129.54, while the expected downloads is about 496250. In relative terms, the total downloads is approximately 2.2 times the expected number otherwise. Instacart would typically take 50 days to achieve the same number of downloads brought by COVID in this period. Economically, these 50 days could mean that all these recent downloads could convert to actual activity much quicker and at a much bigger scale than otherwise.

For the original customers, how much faster did they adopt?

According to the model, the original customers accounted for 579162 downloads. Contrasting this to the earlier case, we can say due to COVID, 82911.88 more downloads came from the Original Customers than Instacart would have otherwise gotten. If it weren't for COVID, this volume of downloads would have only been reached after an additional week by the original customers.

Who couldn't we have gotten and what's up with them?

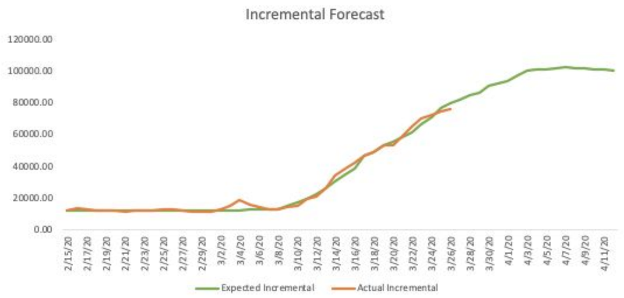
The new segment in just 18 days racked up about 520000 downloads, approximately half the total downloads observed. Despite being only 10% of the market, the strong purchasing propensity has led this segment to contribute greatly to the overall impact.

How did the new market segment enter?

The new market segment actually entered the market in a near linear function with the midpoint and the highest incremental gain around March 20th. This indicates a greater spread around the market entry, which is rather reasonable given that huge variance exists around the situation of every individual and unique household that may have caused people to either enter earlier or much later for instance.

X. Impact Analysis Continued

Extended Impact of COVID
Apart from the immediate impact assessment that can be made, more long-term projections can also be said based on the model that I have.



For instance, based on the forecast from March 26th to April 12th, according to this model, the peak download is expected to occur around April 7th. However, the general plateau will start on April 3rd and continue on till the 16th. Such a high propensity trend will likely stay afloat for an extended amount of time based on the observed trend.

Conclusion and Recommendation:
Overall, this paper offers a model specification that captures the adoption variation of Instacart during COVID quite well. However, it is hard to extend beyond April 12th to make firm inferences given that the covariates may be changing as time goes by. However, as an extension to this project, a sensitivity analysis can be constructed to understand different what if scenarios given future conditions.