Marketing Science Institute Working Paper Series 2021

Report No. 21-129

How has COVID-19 Impacted Customer Relationship: Dynamics at Restaurant Food Delivery Businesses?

Elliot Shin Oblander and Daniel Minh McCarthy

How has COVID-19 Impacted Customer Relationship Dynamics at Restaurant Food Delivery Businesses?

Elliot Shin Oblander
Columbia University, EOblander23@gsb.columbia.edu

Daniel Minh McCarthy
Emory University, daniel.mccarthy@emory.edu

In this paper, we quantify the impact of COVID-19 on customer purchase behaviors – customer acquisition, retention, ordering, and spending – within the restaurant food delivery category in the United States and assess the mechanisms through which these effects have arisen using a unique collection of data sources. Our results suggest that pre-pandemic customer purchase trends were unfavorable, with falling acquisitions and weakening cross-cohort repeat purchase dynamics. COVID-19's impact has been significant, creating $19.3 billion in incremental sales for the category in 2020, or 69% of the overall year-on-year increase in sales. This increase was primarily due to higher purchase frequency from already-active pre-COVID customers and an increase in average order size, not due to changes in customer acquisition and retention. Turning to mechanisms, we find that this growth is primarily attributable to substitution away from restaurant dine-in; while increased stay-at-home behavior has increased customer adoption and order size, it has actually dampened overall sales growth. These results call into question the long-run sustainability of the pandemic-fueled growth in delivery sales, should on-premise dining meaningfully recover after it returns to being a safe activity.

Key words: customer acquisition; customer retention; customer relationship management; marketing-finance interface; COVID-19

---

1. Introduction

Since COVID-19 began to overtake the United States in early 2020, the economy has undergone sudden and extreme shifts. Most restaurants around the country experienced severe restrictions to on-premise dining, and consumers were instructed to stay at home whenever possible. These changes have shifted consumers' food consumption patterns, with meal delivery services enjoying an increased share of wallet as substitutes for restaurants that cannot offer on-premise dining, appealing to consumers reluctant to go outside.

These shifts have been severe. According to the National Restaurant Association, nearly 17% of restaurants in the US were closed as of December 2020 (National Restaurant Association 2020), with another 14% of restaurant operators saying they would close within the next 3 months in the absence of additional government support. These stresses have heightened tensions with restaurant food delivery companies, whose sales have grown dramatically during the crisis, leading to legislation limiting delivery commissions in many large cities (Kelso 2020). Knowing the magnitude of the economic surplus that the pandemic created for the food delivery category is of central importance when setting regulations and economic policy to support the restaurant industry, as regulations like commission caps serve to redistribute economic surplus from delivery companies to restaurants.

These shifts have also created significant uncertainty among company executives, investors, and other stakeholders regarding the future profitability of restaurant food delivery companies, whose prospects hinge upon post-pandemic consumer demand. For example, DoorDash, which went public in December 2020, closed on its first day of trading with a $72 billion valuation. While this valuation implies continued strong consumer demand, some analysts are less optimistic, arguing that revenue may fall sharply after the vaccine roll-out is underway (Roberts 2020). Understanding pre-pandemic trends, and the magnitude and durability of the impact that COVID-19 has had upon them, clarifies the fair valuation of companies in the category.

Our goal in this paper is to shed light upon these timely issues by quantifying the magnitude of the pandemic's impacts and evaluating the mechanisms through which these impacts have come about. We consider multiple aspects of customer behavior beyond overall sales. From the lens of customer relationship management literature (e.g. Gupta et al. 2006), multiple processes determine overall sales: when customers first adopt a food delivery service (acquisition), how often and for how long they continue using the service (repeat purchase and retention), and how much they spend on each order (spend). Decomposing these different processes provides insight into the unit economics of the category, and may lead to different inferences than what would be inferred through sales growth alone (Gupta et al. 2004). To the extent that these processes are more predictable than aggregate trends, this "bottom-up" decomposition also yields more accurate forward-looking assessments of overall sales (Schulze et al. 2012).

In addition to analyzing the absolute effects of COVID-19 on customer base dynamics, we analyze the extent to which four major mechanisms have mediated these effects:

1. Disposable income: many individuals have experienced negative income shocks, and so are likely to reduce their discretionary spending and opt for cheaper groceries rather than delivery.

2. Stay-at-home behavior: many individuals are spending more time at home due to working from home, the perceived danger in going outside, and/or self-quarantining. This drives consumers towards food options that do not require going outside, but also endows them with more time to cook at home.

3. Restaurant supply: many restaurants have closed due to economic shocks from the pandemic. This hurts the supply capacity of food for delivery services, potentially hurting sales.

4. Restaurant dining: many states have restricted restaurant dine-in, and consumers are hesitant to dine in due to potential health risk, which can drive consumers towards delivery as a substitute.

In sum, we infer pre-pandemic "baseline" trends in customer base dynamics and analyze how the pandemic has impacted those trends – both the magnitude of its impacts and the mechanisms driving them. This research contributes to a growing body of literature studying the impact of COVID-19 on consumer behavior. Within the marketing literature, Sim et al. (2021) study how the pandemic changed consumer preferences for music streaming services. Outside of marketing, the literature stream most relevant to our work studies the direct impact of the pandemic on overall consumer spending, and heterogeneity in its effects with respect to demographic variables and purchase characteristics (Chetty et al. 2020, Alexander and Karger 2020, Baker et al. 2020, Chen et al. 2020, Dunn et al. 2020). Our paper is complementary to this literature in that we (1) study one category in detail rather than many categories at a high level, (2) analyze customer relationship dynamics instead of aggregate-level spending, and (3) use "gold standard" population-level data so that our estimates of impact have external validity.[1]

Our results suggest that the category is already largely penetrated, that acquisitions had been falling pre-COVID, and are likely to resume falling post-COVID. As of the end of 2019, sales growth was on a trajectory to slow significantly, largely due to falling acquisitions

[1] Our emphasis on customer relationship dynamics is similar in spirit to Baker et al. (2020), who include sales per transaction as a dependent variable. We also note that Dunn et al. (2020) use a re-weighting scheme from Aladangady et al. (2019) to achieve external validity in their estimate of overall sales impact.

and unfavorable cross-cohort repeat buying trends. These negative growth prospects were forestalled by the pandemic, which had a large impact upon customer behavior. It caused an estimated \$19.3 billion in additional spending to occur in 2020, implying that COVID-19 caused the vast majority (69%) of category-level sales growth in 2020. The uplift was primarily due to existing customers purchasing more frequently and a general increase in average order value (AOV), rather than improvements in customer adoption or retention. We find that these increases are primarily attributable to substitution away from restaurant dining. Increased stay-at-home behavior led consumers to adopt delivery services at higher rates and place larger orders, but also led them to order less frequently, such that the overall effect of stay-at-home behavior on sales is negative.

Since much of this surplus comes from substitution away from restaurant dine-in business, sales growth is likely to fall should government dine-in restrictions be lifted and dine-in activity revert back to pre-pandemic levels, calling into question the long-term sustainability of the pandemic-driven growth in restaurant food delivery. These results also highlight the nuanced relationship between restaurants and the delivery category: substitution away from dine-in has been the primary driver of COVID-related gains for the delivery category, but these gains trade off against rising store closures in the long run. Restaurants are both the main competitor and main supplier of delivery companies, and as such, delivery companies have a vested interest in restaurants remaining open even as they capture market share.

2. Data

For our analysis, we synthesize multiple data sources about customer purchase patterns, geographic market coverage for the delivery category, economic impact, and stay-at-home behavior. We summarize these sources below, with more detailed information available in Table 1 and Web Appendix 1:

- *Customer purchase behavior*: Earnest Research, a leading data analytics company, provided us with daily credit/debit card transaction data at all major US restaurant delivery companies (27 companies in total) for 1.83 million panel members from January 1 2016 to December 31 2020. The monthly location associated with each member is also available at the core-based statistical area (CBSA) level.

- *Geographic market entry*: We collect historical market entry data for the delivery category using the Wayback Machine[2], data from YipitData (an alternative data provider which collects restaurant listings on delivery platforms), and the aforementioned credit/debit card panel data.
- *Employment*: We collect unemployment rate data at the county-month level, available through the Bureau of Labor Statistics.
- *Stay at home behavior*: We observe measures of stay-at-home behavior through mobile location data for 18 million devices at the census block group (CBG)-day level from SafeGraph, a leading geolocation data provider.
- *Restaurant supply and dine-in*: We obtain restaurant supply and dine-in level variables using daily visit duration data from SafeGraph on the approximately 945 thousand restaurants they track.

Table 1 Summary of Data Sources

Category	Data Source	Time Interval	Temporal Granularity	Data Comments
Customer purchasing	Earnest Research	1/1/2016 - 12/31/2020	Daily	Credit and debit card transaction activity, possibly across multiple cards, for 1.83 million panel members in the restaurant delivery category
Market entry	Wayback Machine	6/1/2004 - 12/31/2015	Daily	Web scrapes of every city served by DoorDash, Uber Eats, GrubHub, and Postmates (when available)
	YipitData	10/2018 - 12/2020	Monthly	Name and location of all restaurants listed on all major food delivery platforms
	Earnest Research	1/1/2016 - 12/31/2020	Monthly	Location associated with customers making purchases in category
Restaurant supply, dine-in activity, and stay-at-home behavior	SafeGraph	1/1/2020-12/31/2020	Daily	Mobile location data for 18 million devices. Daily visit duration data for approximately 945 thousand restaurants. Daily stay-at-home statistics by CBG.

Jointly, these data sources provide us with rich transactional data through which we observe customer behavior as well as proxies for each of our mechanisms of interest. Our

[2] `http://web.archive.org/`

spending data is one of the largest – both in terms of number of panel members and observation window – in extant COVID spending impact literature.

Next, we discuss the two methodological approaches that we employ to analyze the effect of COVID-19 on the food delivery category.

3. Methodological Approaches

The goals of our analyses are fourfold: (1) to predict what customer base activity in the restaurant delivery category would have been had COVID-19 not occurred; (2) to quantify the overall impact of COVID-19 on the category; (3) to decompose the overall effect into customer base dynamics; and (4) to assess the mechanisms driving these effects. To achieve these goals, we employ two methodological approaches: an event study and a panel regression. We briefly motivate and describe these methods below.

Consider an outcome of interest Y_t. For the purposes of this paper, Y_t will consist of different statistics about food choice in time t, such as the total sales in the delivery category.

For goals (1)-(3), we wish to estimate Y_t^0 and $Y_t^1 - Y_t^0$, where Y_t^1 denotes the true value that Y_t takes on, while Y_t^0 denotes the value that Y_t would have taken on in a counterfactual world where COVID-19 had not affected the US. The difficulty is that we cannot observe Y_t^0 because COVID-19 afflicted the entire US largely simultaneously, leaving no "control units" to which to compare.

We employ an event study approach to assess the overall impact of COVID-19 (Corrado 2011). Intuitively, Y_t^1 and Y_t^0 coincide prior to the occurrence of the "event" being studied (here, COVID-19), only diverging in the post-event period. Thus, we estimate a predictive model of Y_t^0 based on pre-COVID data, extrapolating from the model to form predictions $\hat{Y}_t^0$ that serve as proxies of the counterfactual no-COVID baseline in the post-event period, with the difference $Y_t^1 - \hat{Y}_t^0$ forming an estimate of the effect of COVID-19 on Y_t.

While this approach requires specifying a parametric model to impute Y_t^0, it does not require specifying a parametric form for the effect size $Y_t^1 - \hat{Y}_t^0$. As such, as long as the baseline Y_t^0 is well-characterized by a parametric model, we can flexibly estimate the time-varying effects of COVID-19. We combine parametric mixture models of customer acquisition, retention, repeat purchase, and spend to model and impute Y_t^0. These models are appropriate for our use case as they have repeatedly been shown to validate well out-of-sample, providing accurate predictions years into the future (McCarthy et al. 2017).

This approach allows us to obtain valid inferences about the overall impact of COVID-19, conditional on the validity of the predictive model used to impute the baseline. As we will demonstrate, our model yields good out-of-sample predictions, lending credibility to our estimates. We formally define the event study causal model and the assumptions required for valid inference in Web Appendix 2.

This approach provides us with an estimate of the overall impact of COVID-19, but does not provide us with insight into the mechanisms driving changes in customer behavior. For goal (4), we employ a panel regression which utilizes spatiotemporal variation to identify how our mechanisms of interest amplified or dampened the effect of COVID-19. We make use of region and time fixed effects, such that identification is driven by on-the-margin variation in how different regions of the US have been affected by COVID-19.

In relying only on marginal spatiotemporal variation for identification, the panel regression approach avoids the parametric extrapolation required for the event study approach. However, this also means that it does not measure the absolute magnitude of the national effect of COVID-19. See Section 6 for more discussion of this distinction.

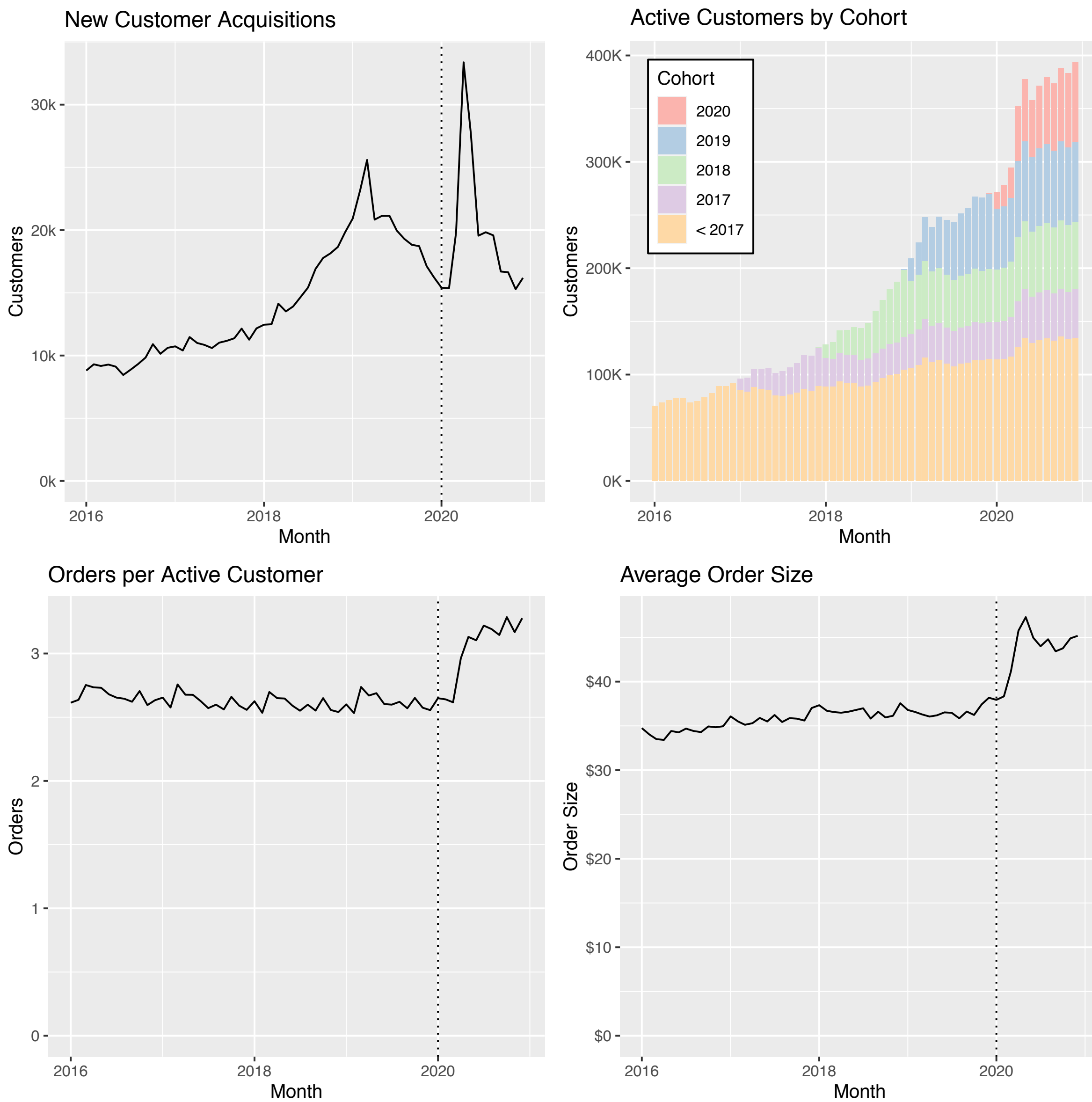
As such, these two approaches are complementary – the event study approach yields valid estimates of pre-pandemic trends and the overall impact of COVID-19, while the fixed effects regression approach provides us with insight into the mechanisms driving these effects on the margin.

4. Analysis of impact magnitude

We first bring the event study approach to life. We would like to understand the overall impact that COVID-19 had upon how customers adopt, churn, order, and spend. We first provide model-free evidence in Figure 1, which summarizes aggregated customer-level activity along the aforementioned four dimensions. We see the following:

- Customer acquisitions (top-left) were steadily rising, peaking in March 2019 before steadily falling. In 2020, we observe a sharp but transitory increase in customer acquisitions.
- Using the count of active customers by acquisition cohort (top-right) as a proxy for retention, we observe little change in retention for pre-COVID cohorts during the COVID period.
- Orders per active customer (bottom-left) and AOV (bottom-right), which had been fairly stable through the end of 2019, exhibited sharp, sustained increases in 2020.

Figure 1 Summary of customer acquisition, active customer, orders per active customer, and average order size by month



Category-level behavior was relatively well-behaved pre-COVID, lending itself well to parametric customer base modeling to infer the counterfactual baseline. The varied nature of COVID's effect – across processes and over time – supports our use of an event study approach, which captures these COVID-related dynamics without requiring explicit parameterization.

4.1. Model specification

Next, we specify the predictive models we use to estimate the pre-COVID baseline for customer acquisition, purchasing, and AOV. For concision, and because these models were largely conceived in prior literature, our descriptions are brief, with the full model specifications provided in Web Appendix 3.

Acquisition. We model customer acquisition using a minor modification of the "time of mass awareness" (TMA) model of McCarthy and Fader (2018). This model characterizes the duration of time that elapses from market entry to adoption of delivery services through two Weibull distributions. The second Weibull process begins at a time $t^{\star}$ to be estimated, allowing the trajectory of customer acquisitions to increase thereafter. We extend this model to account for geographic expansion in delivery services over time, because individuals cannot be prospects before the category serves their area.

Repeat Purchasing. We model repeat purchase behavior using the extended Pareto/NBD (EPNBD) model of Bachmann et al. (2021) This model is an extension of the Pareto/NBD model (Schmittlein et al. 1987), the most widely-recognized latent attrition model used to forecast repeat customer purchasing in non-subscription settings. The EPNBD model allows for time-varying covariates to enter the purchase and attrition propensities through proportional hazards. We empirically observe strong cross-cohort and tenure effects, so we include linear covariate terms for the customer's time of adoption (cohort) and time since adoption (tenure).[3] Latent attrition models have a long history of successful use for aggregate-level repeat purchase predictions, making this model a natural choice (Ascarza et al. 2017).

Spending. We model AOV using a simple time series model, assuming AOV for customer i at time t is a homogeneous log-linear function of customer tenure:

$$\log \mathrm{AOV}_{it} = \beta_{s,0} + \beta_{s,tenure}\mathrm{Tenure}_{it} + \varepsilon_{it}$$

This is a suitable functional form for customer spending because very little in the way of dynamics are evident in the pre-COVID data.

[3] We include the cohort variable for both the attrition and purchase process, but include the tenure variable only for the purchase process, as duration dependence is generally indistinguishable from latent heterogeneity in survival models (Heckman 1991).

4.2. Implementation

We apply the models described in the previous section to the credit/debit card panel data described in Section 2. We removed Uber Eats from our main analysis because of missing data from June to August 2019.[4] We include Uber Eats as a robustness check in Web Appendix 7. We conservatively set the event date to January 1 2020, estimating our model through the end of 2019 and forecasting outcomes for all of 2020.

We estimate our models by maximum likelihood, computing confidence intervals for our predictions via non-parametric bootstrap with bias correction (Efron and Tibshirani 1994). For further detail, see Web Appendix 4.

Figure 2 shows the resulting counterfactual baseline fits and forecasts for total acquisitions (top-left), total orders (top-right), AOV (bottom-left), and total sales (bottom-right). We provide parameters estimates and associated standard errors in Web Appendix 5.

The model does a good job of capturing baseline trends for each of these processes. Predictive validation through a holdout analysis, training upon all data through the end of 2018 and predicting 2019 outcomes, is included in Web Appendix 6. This validation further supports the proposed model's ability to accurately forecast future category-level customer behavior in the pre-COVID period.

Pre-pandemic trends were mixed at best. The category is saturated, with approximately 50% of all panel members having adopted into the category by the end of 2019 and total acquisitions having steadily fallen for 9 months. The baseline trend in customer acquisition, which had been a tailwind for growth, became a headwind. Moreover, customers acquired more recently are of significantly lower value, largely driven by higher churn propensities. While repeat purchasing tends to increase as a function of customer tenure, worsening cross-cohort dynamics more than offset this trend for recently acquired customers. To the extent that these patterns resume after the pandemic ends, growth trends will come under pressure. We include a more detailed discussion of these patterns in Web Appendix 5.

Figure 2 indicates that the pandemic caused a sharp deviation away from these baseline dynamics. There was a sharp but transitory spike in customer acquisition, while the corresponding increases in total orders and total sales have been more enduring. We discuss this in more detail in Section 6.

[4] Over this period, Uber experimented with allowing people to order Uber Eats through the main Uber app (O'Kane 2019), making it infeasible to disambiguate ridesharing transactions from Uber Eats transactions off of credit/debit card data.

Figure 2 **Actual versus expected total acquisitions, total orders, AOV, and total spend**

Note: Vertical dotted lines represent the end of calendar year 2019. Confidence intervals are obtained via nonparametric bootstrap.

4.3. Moving from panel to population

As a final step in our inference procedure, we translate our results from the credit/debit card panel to the broader US population, since credit/debit card holders' behavior may not be representative of all US consumers; accordingly, we must establish the external validity of our results (Aladangady et al. 2019).

Aggregate statistics disclosed by publicly traded companies in Securities and Exchange Commission (SEC) filings can serve as ground truth population-level data on which to assess and correct for selection bias in credit/debit card panels (McCarthy and Oblander 2021). Two of the three major players in the restaurant delivery aggregator market (GrubHub and DoorDash) disclose their US sales on a quarterly basis, providing ground truth population-level data. We compare gross food sales reported by GrubHub and DoorDash to gross food sales at both companies recorded in the panel data (all in per-capita units) in a simple log-log regression:

$$\log \text{Sales}_q^{Population} = \beta_0 + \beta_1 \log \text{Sales}_q^{Panel} + \varepsilon_q$$

This regression yields $R^2 = 99.8\%$, indicating near-perfect correspondence between panel and population sales; accordingly, we expect the inferences from the panel to be directionally consistent with the true population-level effects.

We estimate $\hat{\beta}_0 = -1.515$ (SE 0.053) and $\hat{\beta}_1 = 1.146$ (SE 0.015).[5] The negative intercept indicates that the panel tends to oversample high-spending customers, while the slope of slightly over 1 indicates that trends in the population are somewhat underrepresented in the panel (on proportional scale).

We use the estimated coefficients from this regression to perform a simple selection correction, plugging inferred panel sales into the regression equation to obtain population sales. All results reported in the remainder of this section are at the level of the entire US population, based on this selection correction.

We present the full regression results, additional discussion of panel representativeness, and details on the implementation of the selection correction in Web Appendix 8.

4.4. Results

Our inferred national actual and counterfactual sales are shown in Figure 4.4. We observe that sales first significantly diverged from the model predictions in the second quarter of 2020. The gap has been persistent, with the pandemic inferred to add over $6 billion in sales per quarter through the latter three quarters of 2020.

Over the full 2020 calendar year, we estimate that COVID-19 generated $19.3 billion (SE $2.8 billion) in sales, relative to an estimated $50.6 billion (SE $0.8 billion) in overall category sales. We estimate that category-level sales were $22.7 billion (SE $0.2 billion) in 2019, implying COVID-19 caused the vast majority (69%) of category-level sales growth in 2020. Category sales would have grown by 38% in the absence of the pandemic, significantly less than the 122% that was actually observed.

We decompose this impact into its underlying customer behaviors. We separately estimate the impact attributable to pre-COVID customers (i.e., to customers that were acquired before 2020) versus post-COVID customers; then, we decompose these impacts

[5] Standard errors are heteroskedasticity and autocorrelation consistent.

Figure 3 **Actual versus expected national sales with 95% confidence bands**

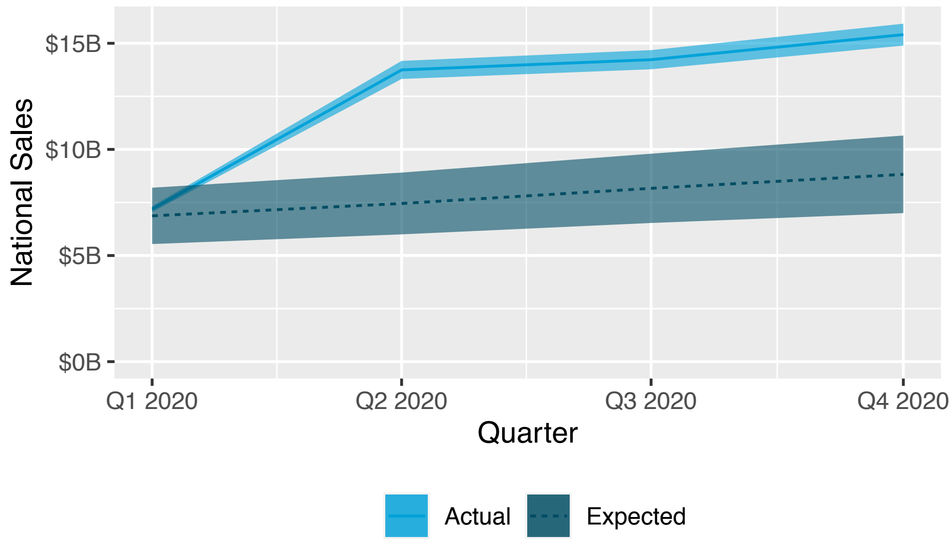
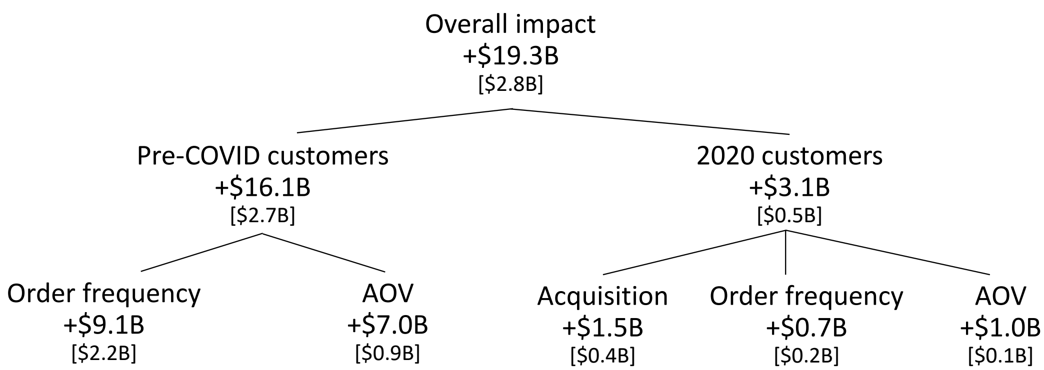


Figure 4 **Decomposition of the impact of COVID in the US in 2020 [standard error]**



Note: Standard errors are provided beneath point estimates. Decomposition of the impact attributable to 2020 customers should be interpreted as the (1) impact of the increase in the volume of customer acquisitions, assuming order frequency and AOV are fixed at baseline levels; (2) impact of the increase in order frequency, assuming AOV is fixed at its baseline level; and (3) impact of the increase in AOV.

into the impact due to an increase in customer acquisition volume (for post-COVID customers), order frequency, and AOV. The results, including details of the procedure, are reported in Figure 4.

The decomposition suggests that most (84%) of this impact was attributable to pre-COVID customers, roughly equally due to an increase in purchase frequency versus an increase in AOV. The limited impact attributable to post-COVID customers was primarily due to a higher volume of customers acquired (47%) and an increase in AOV (32%).

In sum, the impact of COVID is significant, largely due to an increase in AOV and order frequency from pre-COVID customers, rather than expansion in the size of the customer base. However, to understand whether these effects will persist into the future, it is instructive to understand the mechanisms through which these increases are coming about, and what is likely to happen to those mechanisms in the coming months.

5. Analysis of mechanisms

We focus on four mechanisms – income shocks, stay-at-home behavior, restaurant supply, and restaurant dine-in activity. We exploit spatiotemporal variation in the economic impacts of COVID-19 identify the extent to which each of these mechanisms has mediated the impact on the delivery category. We detail our model specification and identification strategy below.

5.1. Model specification

We model observations at the CBSA-day level with a two-way fixed effects (FE) log-log regression model:[6]

$$\log\left(Y_{ct}\right) = \alpha_c + \alpha_t + \vec{\beta}' \log\left(\vec{X}_{ct}\right) + \varepsilon_{ct}$$

where Y_{ct} represents the dependent variable of interest and $\vec{X}_{ct}$ represents the vector of regressors/mechanisms of interest for CBSA c on day t. We perform analysis at the CBSA-day level.

We consider four dependent variables, which are computed using the credit/debit card panel:

1. Acquisitions: the proportion of panel members in CBSA c acquired on day t.
2. Orders: the number of delivery orders placed by panel members (per capita) in CBSA c on day t.
3. Order size: the AOV of orders placed by panel members in CBSA c on day t.

[6] Since the dependent variables are sparse (i.e. there are many CBSA-day observations where zero purchases are observed), in practice we use $\log\left(a+Y_{ct}\right)$ as our dependent variable for a small positive value of a to avoid missing values from taking the logarithm of zero. We report results with $a = 10^{-4}$, but we obtain very similar results with other values such as 10^{-3} and 10^{-5}.

4. Sales: the dollar amount of all orders placed by panel members (per capita) in CBSA c on day t.

The last variable is the most relevant for determining the mechanisms of overall sales impact; the other three variables are diagnostic for decomposing the mechanisms for different customer base dynamics.

We consider four regressors, which are proxies for the four mechanisms laid out above. We define how we operationalize these regressors in Table 2.

Table 2 Summary of Regressors

Mechanism	Proxy variable	Definition
Income shock	Unemployment rate	Unemployment rate by CBSA is reported by the BLS. Data is only available at the monthly level
Stay-at-home behavior	Proportion of people at home	Proportion of people staying completely at home by CBSA-day, as reported by SafeGraph
Restaurant supply	Restaurant employee shifts	An employee shift is proxied for by restaurant visits of over four hours. It is measured on CBSA-day per capita basis, as reported by SafeGraph
Restaurant dine-in	Number of restaurant dine-in visits	A dine-in visit is proxied for by restaurant visits between 20 minutes and four hours. It is measured on a CBSA-day per per capita basis, as reported by SafeGraph

5.2. Identification

Having specified our FE model, we next discuss our identification strategy. Intuitively, while COVID-19 impacted all regions of the US simultaneously, the severity of initial impact and trajectory of recovery has differed, enabling identification.

The demographics and economies of CBSAs are likely to differ from each other substantially; such heterogeneity between CBSAs is likely to be correlated both with our regressors and delivery behavior. The CBSA fixed effects control for all static heterogeneity between CBSAs, such that only within-CBSA time series variation in regressors is used to identify the coefficients.

Additionally, while COVID-19 outbreaks occurred in different regions at different times, much of state government response and changes in consumer behavior occurred abruptly and near-simultaneously, especially early on in the pandemic. These “common shocks” may be the result of federal government guidance, national news coverage, and other such drivers of national consumer sentiment; in turn, these shocks may have affected other unobserved variables that are correlated with our regressors and delivery behavior. The day fixed effects in our model control for such unobserved common shocks.

Consequently, our model identifies the coefficients of interest using within-CBSA, within-day variation. For instance, while unemployment increased significantly for most CBSAs in April, our identification of the effect of unemployment on delivery spending comes from comparing CBSAs whose unemployment rates jumped, on the margin, more or less than the national average.

This discussion gives only a high-level overview, and some endogeneity concerns may remain even after controlling for fixed effects. In Web Appendix 9, we provide a more thorough discussion of the source of identification of each coefficient, and discuss and run additional robustness checks for a number of potential endogeneity concerns. These concerns include but are not limited to simultaneity of dine-in activity with delivery sales, of dine-in activity with stay-at-home behavior, and of restaurant employment with delivery sales, as well as a number of confounds, including socioeconomic status, political beliefs, population density, stay-at-home restrictions, strategic targeted marketing by delivery companies, and the number of restaurant listings on delivery platforms. The discussion and results further support the validity of the results we summarize in the next section.

5.3. Results

We estimate our model by weighted least squares (WLS).[7] We compute heteroskedasticity and cluster robust standard errors with two-way clustering by CBSA and day (Cameron et al. 2012). Table 3 reports the results.

All of our regressors have a statistically and economically significant impact on sales. The unemployment coefficients imply that the negative shock on disposable income from the pandemic has hurt overall delivery sales, and that this effect is attributable to decreased order frequency rather than changes in new customer adoption. There is a slight positive impact on order size, which could be an artifact of, for instance, unemployed individuals saving money on delivery fees by placing one large order instead of multiple small orders.

[7] Since CBSAs differ greatly in size, this introduces substantial heteroskedasticity in our dependent variables; accordingly, we weight observations by CBSA panel population.

Table 3 FE Regression Model Estimates

DV	Acquisitions		Orders		Order Size		Sales	
Regressor	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Unemployment	0.011	(0.017)	−0.175***	(0.044)	0.015**	(0.005)	−0.213***	(0.045)
Stay-at-home	0.261**	(0.054)	−0.352**	(0.125)	0.075***	(0.015)	−0.344***	(0.117)
Restaurant supply	0.128***	(0.030)	0.204***	(0.042)	0.001	(0.008)	0.250***	(0.050)
Restaurant dine-in	−0.124***	(0.028)	−0.200***	(0.056)	0.004	(0.009)	−0.216***	(0.053)
N	289,872		289,872		209,438		289,872	

Note: asterisks denote level of significance (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$). All specifications include CBSA and day fixed effects. Standard errors are two-way clustered, robust to within-CBSA and within-day dependence. N is smaller for the order size regression as average order size is ill-defined for CBSA-day pairs with zero observed delivery orders.

Increases in stay-at-home behavior have driven higher customer acquisitions and larger order sizes, but have substantially lowered order frequency, resulting in an overall negative effect on sales. This reflects the multiple ways in which stay-at-home behavior can affect delivery ordering. Staying at home may increase acquisitions, since people who otherwise would not have tried delivery services now have added incentive to do so due to safety and convenience. Additionally, more family members at home from work or school could lead to larger order sizes. Conversely, professionals who otherwise may have ordered delivery due to commute times are now working from home, freeing up time and resources to cook. These varying forces all appear to be at play, but the net result on sales is negative.

The restaurant supply coefficients imply that the supply of restaurant food is essential for delivery services; restaurant closures and decreases in staffing levels lead to fewer options and longer wait times for delivery, making consumers less likely to adopt and utilize delivery services.

Lastly, the coefficients on dine-in demonstrate the strong degree of substitution between dine-in and delivery. When on-premises restaurant dining is disallowed or unsafe, consumers turn to delivery as the next best alternative. This manifests both in increased adoption and order frequency.

In summary, we find that the windfall gains of the category due to COVID-19 are primarily attributable to substitution away from dine-in behavior, with the other mechanisms applying downward pressure on the category.

6. Discussion

Our event study and regression analyses both offer mixed inferences about the health of the restaurant delivery category. The former analysis suggests that pre-COVID sales growth trends were weakening significantly and that COVID accounted for the majority of overall growth in 2020, while the latter analysis implies that the decrease in restaurant dine-in was the primary driver of that growth. To the extent that there is a the return to restaurant dine-in as part of the recovery from the pandemic, this suggests negative future prospects for the delivery category.

Based on the regression results, we estimate that, if each CBSA's average dine-in levels in November-December 2020 were to revert to their respective January-February 2020 dine-in levels, this would have lowered national November-December delivery sales by 9.7% (SE 2.1%). This represents approximately one quarter of the pandemic-driven gains inferred in Section 4.4. While this figure is stylized, it demonstrates that if consumer demand for restaurant dine-in returns after government restrictions are lifted and consumers feel safe dining on-premise again, sales may fall sharply.

The extent to which stay-at-home behavior will return to pre-pandemic levels in the long run is an open question. Long-term persistence of the work-from-home lifestyle would imply further headwinds for the delivery category. The sharp increase in companies allowing employees to work from home for an extended period of time make it more likely that the trend towards staying at home will not be transitory. A Conference Board survey conducted in October 2020 reported that 40% or more of employers' workforces will be primarily remote in the long term (Cappelli and Bonet 2021).

We now address some limitations of our analyses and interpretations, and discuss related questions for future work.

First, our event study analysis cannot separate out effects of COVID-19 from other effects that may have occurred simultaneously. As such, while it is appropriate for inferences over the short to medium term (when it is reasonable to believe that the pandemic is the primary cause of disruption), its estimates should be treated with caution over longer time horizons, when other unrelated events are more likely to significantly shift customer behavior and contaminate our results.

Second, as noted in Section 3, the effect sizes in the regression analysis do not necessarily translate to overall national effect sizes. Accordingly, we cannot make statements about the overall national effect of, for example, unemployment increases on delivery sales without assuming that the unemployment effect is invariant to the source of variation (i.e., common national trends versus idiosyncratic regional trends). Accordingly, while we believe our regression inferences reflect the overall sign and approximate magnitude of the effects of interest, the exact coefficients and resulting counterfactual implications are stylized.

Third, our FE model only captures contemporaneous (within-day) effects: for instance, it is also possible that a consumer substituting from dine-in to delivery on a given day then becomes habituated to ordering delivery, resulting in a persistent boost to delivery sales. However, much of the variation in the data that identifies the dine-in coefficient comes from the reopening period, where we observed delivery orders being replaced by dine-in visits as restaurants resumed on-premises dining. If habituation had been a dominant factor over the reopening period, we would not have observed such strong substitution away from delivery back to dine-in on the margin.

Finally, we readily acknowledge that we have not incorporated competition between restaurant delivery platforms into our model. Our target of interest in the event study analysis is category-level impact, making competitive analysis out of scope. The competitive structure of the category, the network effects of its constituents (Gupta 2009), and whether/how these factors were transformed by the pandemic, are interesting questions that we leave to future work.

More broadly, this research contributes to the growing body of research studying the varied effects of COVID-19 on customer spending behavior, providing a useful framework for future researchers to build upon. While our modeling approach should be portable to other industries, two unique aspects of the restaurant delivery category merit mention. First, restaurants are both suppliers and competitors of delivery companies. As such, restaurant store closures decrease delivery demand. This is different from the market structure of other digitally native companies in verticals such as apparel and furniture, for whom brick and mortar stores are only competitors; such companies should benefit significantly from store closures. Second, delivery orders are generally paid for by card, making credit/debit card panels highly diagnostic. Some industries have a sizeable portion of cash transactions, making analysis with a credit/debit card panel difficult; this issue is especially salient during COVID, which may have changed payment method mix.

In sum, while there are limitations to our analyses, our results nevertheless suggest a decline in category sales growth after the pandemic subsides if demand for on-premise dining returns, which could lead to a substantial downward revaluation of companies in the category. Had pre-pandemic trends been more benign, the absolute impact of COVID been smaller or less sensitive to dine-in substitution, this decline would be less consequential, but our results suggest otherwise.

Acknowledgments

We thank Earnest Research for providing access to their credit/debit card panel data. We also thank YipitData and SafeGraph for providing us access to their data. We acknowledge financial support provided by a research grant from the Goizueta Business School of Emory University. We are very grateful to excellent research assistance from Kivan Polimis, as well as the data collection efforts of Travis Mays and Alex Mette. We are thankful to Don Lehmann, Oded Netzer, Asim Ansari, Peter Fader, Eric Bradlow, Davide Proserpio, and participants of the Columbia Business School and UC Davis Marketing seminar series for helpful suggestions. We also thank Markus Meierer, Patrick Bachmann, and Patrik Schilter for advice on implementation of the EPNBD model.

References

Aladangady A, Aron-Dine S, Dunn W, Feiveson L, Lengermann P, Sahm C (2019) From transactions data to economic statistics: Constructing real-time, high-frequency, geographic measures of consumer spending. *National Bureau of Economic Research Working Paper Number 26253* .

Alexander D, Karger E (2020) Do stay-at-home orders cause people to stay at home? Effects of stay-at-home orders on consumer behavior. *Federal Reserve Board of Chicago Working Paper Number 2020-12* .

Ascarza E, Fader PS, Hardie BG (2017) Marketing models for the customer-centric firm. *Handbook of marketing decision models*, 297–329 (Springer).

Bachmann P, Meierer M, Näf J (2021) The role of time-varying contextual factors in latent attrition models for customer base analysis. *Marketing Science (Articles in Advance)* .

Baker SR, Farrokhnia RA, Meyer S, Pagel M, Yannelis C (2020) Income, liquidity, and the consumption response to the 2020 economic stimulus payments. *National Bureau of Economic Research Working Paper Number 27097* .

Cameron AC, Gelbach JB, Miller DL (2012) Robust inference with multiway clustering. *Journal of Business & Economic Statistics* 29(2):238–249.

Cappelli P, Bonet R (2021) After COVID, should you keep working from home? Here's how to decide. *Wall Street Journal* .

Chen H, Qian W, Wen Q (2020) The impact of the covid-19 pandemic on consumption: Learning from high frequency transaction data. *Available at SSRN: https://www.ssrn.com/abstract=3568574* .

Chetty R, Friedman JN, Hendren N, Stepner M, Team TOI (2020) The economic impacts of COVID-19: Evidence from a new public database built using private sector data. *NBER Working Paper Number 27629* .

Corrado CJ (2011) Event studies: A methodology review. *Accounting & Finance* 51(1):207–234.

Dunn A, Hood K, Driessen A (2020) Measuring the effects of the covid-19 pandemic on consumer spending using card transaction data. *US Bureau of Economic Analysis Working Paper Number 2020-5* .

Efron B, Tibshirani RJ (1994) *An introduction to the bootstrap*, volume 57 (Boca Raton, FL: CRC press).

Gupta S (2009) Customer-based valuation. *Journal of Interactive Marketing* 23(2):169–178.

Gupta S, Hanssens D, Hardie B, Kahn W, Kumar V, Lin N, Ravishanker N, Sriram S (2006) Modeling customer lifetime value. *Journal of Service Research* 9(2):139–155.

Gupta S, Lehmann DR, Stuart JA (2004) Valuing customers. *Journal of marketing research* 41(1):7–18.

Heckman JJ (1991) Identifying the hand of past: Distinguishing state dependence from heterogeneity. *The American Economic Review* 81(2):75–79.

Kelso A (2020) New York, Los Angeles extend delivery commission fee caps. URL `https://www.restaurantdive.com/news/new-york-los-angeles-extend-delivery-commission-fee-caps/584385/`.

McCarthy DM, Fader PS (2018) Customer-based corporate valuation for publicly traded noncontractual firms. *Journal of Marketing Research* 55(5):617–635.

McCarthy DM, Fader PS, Hardie BG (2017) Valuing subscription-based businesses using publicly disclosed customer data. *Journal of Marketing* 81(1):17–35.

McCarthy DM, Oblander ES (2021) Scalable data fusion with selection correction: An application to customer base analysis. *Marketing Science (Articles in Advance)* .

National Restaurant Association (2020) Restaurant industry in free fall; 10,000 close in three months. URL `https://restaurant.org/news/pressroom/press-releases/restaurant-industry-in-free-fall-10000-close-in`.

O'Kane S (2019) Uber starts allowing Eats orders inside its main app. URL `https://www.theverge.com/2019/6/5/18654128/uber-eats-main-app-food-delivery-scooters-bikes`.

Roberts D (2020) DoorDash IPO is 'most ridiculous of 2020' and 'holds no value'. URL `https://finance.yahoo.com/news/door-dash-ipo-is-most-ridiculous-of-2020-and-holds-no-value-analyst-125054305.html`.

Schmittlein DC, Morrison DG, Colombo R (1987) Counting your customers: Who are they and what will they do next? *Management Science* 33(1):1–24.

Schulze C, Skiera B, Wiesel T (2012) Linking customer and financial metrics to shareholder value: The leverage effect in customer-based valuation. *Journal of Marketing* 76(2):17–32.

Sim J, Cho D, Hwang Y, Telang R (2021) Virus shook the streaming star: Estimating the COVID-19 impact on music consumption. *Forthcoming at Marketing Science* .

Web Appendix: How has COVID-19 Impacted Customer Relationship Dynamics at Restaurant Food Delivery Businesses?

Elliot S. Oblander
Columbia University, eoblander23@gsb.columbia.edu

Daniel M. McCarthy
Emory University, daniel.mccarthy@emory.edu

Appendix 1: Data descriptions and pre-processing

In this appendix, we provide additional details regarding the pre-processing steps that we performed to arrive at the data set used to perform our analyses. We organize this section by data source.

1.1. SafeGraph data

SafeGraph[1] provides aggregated mobile location data at the daily level based on a panel of over 18 million devices. We use two different data products for our analysis: Social Distancing Metrics and Weekly Patterns data.

The Social Distancing Metrics product provides stay-at-home data based on each device's location: the device's home location is inferred to be the modal nighttime (6 PM to 7 AM) location of the device over the preceding 6 weeks. Home locations are divided up to the geohash-7 level of precision (a grid of boxes of approximately 150 meters by 150 meters). A device is said to stay completely at home if it did not have a location ping outside this home location within a given day. SafeGraph aggregates stay-at-home data to the census block group (CBG) level, reporting the number of devices in a given CBG that stayed completely at home (along with the total number of devices with home location at the CBG level). CBGs are nested within CBSAs, and so to further aggregate these statistics to the day level, we simply sum across CBGs within a CBSA to obtain the total number of devices staying at home, and the total number of devices overall, within that CBSA. The ratio of these figures gives the stay-at-home rates that we use in our regressions.

The Weekly Patterns product provides store-level visit data for about 4.4 million retail locations in the United States across many industries, which we use to construct our proxies for restaurant supply/employment and restaurant dine-in levels. The data includes National American Industry Classification

[1] `https://www.safegraph.com/`

System (NAICS) codes for industry classifications; as such, we filter down to only locations with 4-digit NAICS code 7225 ("restaurants and other eating places"), leaving about 945,000 locations, hereafter referred to as "restaurants." The data provides the number of visits to each restaurant (defined as a device being at the restaurant for at least four minutes) at the daily level, and the number of visits segmented by visit length (less than 5 minutes, 5 to 20 minutes, 21 to 60 minutes, 61 to 240 minutes, and greater than 240 minutes) at the weekly level. As noted in Section 5.1 of the main text, we proxy dine-in visits by visits of 21 to 240 minutes, while we proxy employee shifts by visits of over 240 minutes. To approximate the daily dine-in visits and employee shifts for a given restaurant, we multiply total daily visits by the weekly proportion of visits that were 21 to 240 minutes (for dine-in) or over 240 minutes (for employment). We then sum these estimated daily visit counts across all restaurants within a given CBSA to obtain the total number of dine-in visits and employee shifts in that CBSA (within the SafeGraph panel). Lastly we normalize these figures to a per-capita level by dividing by the number of devices with home location in a given CBSA on the same day. These normalized figures give the dine-in visit levels and restaurant employee shift levels that we use in our regressions.

1.2. Bureau of Labor Statistics data

We obtained county-month unemployment statistics – in particular, the total number of people who are unemployed and the total size of the labor force – for 1,880 counties from the Bureau of Labor Statistics (BLS).[2] Recognizing that the unemployment rate is equal to the number of people who are unemployed divided by the size of the labor force, we obtain the unemployment rate in each CBSA-month by summing the total number of people who are unemployed across all counties within each CBSA, summing the total size of the labor force across all counties within each CBSA, then dividing the former by the latter. The resulting final data set consists of monthly unemployment rate data for 929 CBSAs.

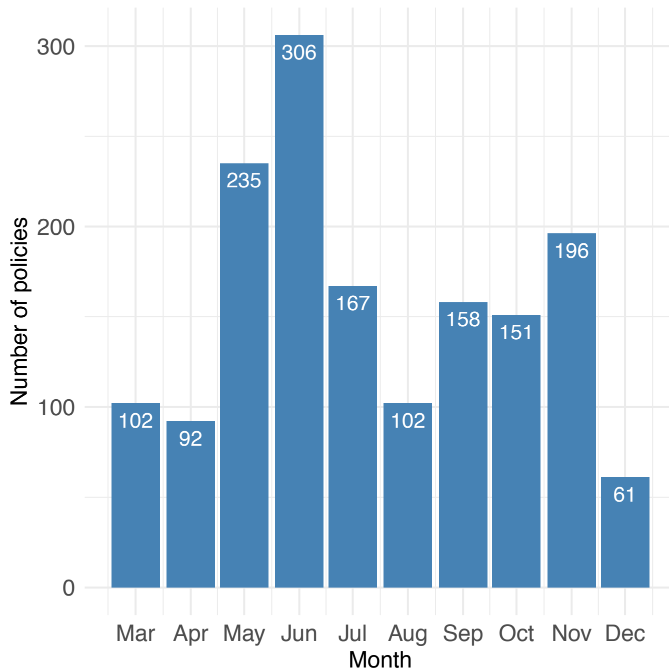
1.3. YipitData

YipitData,[3] a firm that specializes in collecting, processing, and analyzing alternative data, provided data regarding restaurants listed on six delivery platforms over a two year period. YipitData provided us with monthly data from October 2018 to December 2020. Each month, we observe every restaurant listed on the following platforms: Bite Squad, DoorDash, Grubhub, Postmates, UberEats, and Waitr. We observe the name, street address, and the platforms that each restaurant are listed on each month. We first map the zip code associated with each restaurant to the CBSA that zip code is in. In each given month, we infer whether a platform is operating in a given CBSA based on whether there is at least 1 restaurant listed in that CBSA in that month, providing geographic coverage data from October 2018 and December 2020 for these six platforms, which cover the vast majority of customer activity in our data. There are 873 CBSAs with restaurant listings as of December 2020, and 1,082,640 restaurants listed on at least one of the aforementioned six restaurant delivery services during our observation period.

[2] `https://www.bls.gov/web/metro/laucntycur14.txt`

[3] `http://www.yipitdata.com`

Figure 1 **Total number of state government policies issued in 2020**



1.4. Dine-in restrictions data

We created county-level COVID-19 restaurant restriction data by transforming state-level restaurant restriction data available from a repository maintained by researchers at the University of Washington (Fullman et al. 2021). The data repository is a managed collection of social distancing policies primarily sourced from individual state government websites and supplemented by governor social media, news articles, and other compilations of state-level policy actions (e.g., National Governors Association and the Kaiser Family Foundation). There are 1,570 unique policies over the ten-month period from March to December 2020. Figure 1 plots the total number of unique policies each month over this time period. 516 of these policies were state-wide orders, while the other 1,054 policies differed by county.

Our data pre-processing procedure is as follows. First, for each state-level restaurant policy order, we programmatically extracted three restaurant restriction measures and the dates the order applied to. We record three binary measures for each policy:

1. whether indoor dining is allowed
2. whether outdoor dining is allowed
3. whether indoor dining capacity is reduced, e.g. through spacing requirements or capacity caps.

Our policy date extraction creates state-day policy data by identifying what policy was in effect each day. For example, if a policy was issued on May 15th, applying to the period of time from May 22nd to June 15th, and another policy is issued on June 1st, applying to the period of time from June 7th to June 15th, we assume the former policy was in effect from May 22nd to June 6th, while the latter policy was in effect from June 7th to 15th.

For each state-day policy, we manually identified which counties were affected by the policy in the 1,054 policies that were not state-wide. Lastly, we map the county-day policy data to CBSA-day data by taking the population-weighted average of each of the three variables across all counties within a CBSA.

1.5. Earnest Research data

Earnest Research, one of the largest credit and debit card panel data companies provided credit/debit card transaction data. The data includes 926 CBSAs and spans 7 years, from January 1st 2014 to December 31st 2020. Our data includes 1,909,445 panel members with consistent shopping behavior over that time period who made at least one observed purchase.[4] Notably, this data covers multiple debit/credit cards associated with the same individual, such that delivery orders placed on multiple cards by one individual will correctly be attributed to the same individual. As we describe below, we use the data from 2014 and 2015 to address left-censoring in the dataset, such that in practice we use only the 5 years of data from 2016 to 2020 for analysis.

1.5.1. Inferring panel member locations and filtering panel members Each transaction has an associated CBSA which is either inferred directly from the transaction description ("known location" transactions) or imputed based on other transactions occurring in the same time period ("guessed location" transactions). We use this data to infer the modal CBSA where each panel member resided in each month. Specifically, the CBSA c_{im} where panel member i resided in month m is inferred to be the CBSA where they had the most known location transactions in that month m. In cases where there are no known location transactions by panel member i in month m, or there is a tie between multiple CBSAs, ties are broken by guessed location transactions. For the purposes of this procedure, all locations that are outside of any CBSA (i.e. rural areas) are classified into a single "other" CBSA.

In some months, there may be 0 observed transactions by a given customer, and thus the location c_{im} is missing; in these cases, we impute it as the previous month's location $c_{im} \leftarrow c_{i(m-1)}$; if this location is also missing, we in turn recursively impute it as $c_{i(m-1)} \leftarrow c_{i(m-2)}$. When location is missing at the beginning of the dataset, we impute in the opposite direction, i.e. imputing $c_{i1} \leftarrow c_{i2}$, imputing $c_{i2} \leftarrow c_{i3}$, and so on. We impute up to 12 consecutive months of missing location data; if an individual is missing 13 or more consecutive months of location data, we exclude that individual from the dataset. This results in removing 41,022 panel members (2.15%) from the dataset, leaving 1,868,423 panel members with complete monthly location data.

Lastly, there are many CBSAs in which no major delivery platform operates, or where a delivery platform entered but subsequently exited. Accordingly, we restrict our attention to panel members who primarily reside in a CBSA where at least one delivery platform (among those covered by Yipit Data) operated as of December 2020 (843 CBSAs in total). To do so, we construct a static measure of location for each panel member as the mode of their monthly locations across all months; if an individual's primary location over the length of the dataset was not in one of the 843 retained CBSAs, that individual was omitted from analysis.

[4] Consistent shopping behavior is defined using proprietary logic from the data provider to filter down to panel members who are inferred to use the credit and debit cards on record for most of their purchases.

This results in the removal of 42,152 panel members (2.26%), leaving 1,826,271 panel members who were eligible to adopt a delivery platform. These panel members are used for all of our analyses.[5]

1.5.2. Pre-processing transactions We then filtered down the transactions to be used in our analysis. First, we identified all food delivery platforms in the spend transactions data and filtered down to these transactions. The full list of delivery platforms considered, and their respect market shares, is given in Table 1. This resulted in 52,382,711 observed transactions among the 1,826,271 retained panel members.

We omitted the platform EzCater, since it is a B2B business that caters to corporate clients. We further filtered out transactions of negative amounts (i.e. refunds) and transactions of $1,000 or more (since these orders are likely to be for B2B catering rather than individual consumers). This results in the removal of 1,474,257 transactions (2.81%), leaving 50,908,454 transactions.

Lastly, we aggregated together orders placed on the same day by the same individual, such that our inferred order sizes reflect daily spend amounts (i.e. total spending on delivery across all orders). This smooths over changes over time in within-day order frequency (e.g. changes in whether people tend to order twice in one day versus ordering for two meals at once) and simplifies modeling by binarizing order counts at the daily level. Modeling purchase incidence at the sub-day level is infeasible since our credit/debit card data only identifies transaction dates, not times. This results in 41,732,099 unique panel member × day pairs with non-zero transaction amounts.

1.5.3. Analysis-specific filtering For the event study model, we omit Uber Eats from our analysis due to a data quality issue. In particular, starting in late April of 2019, Uber began allowing certain customers in some markets to place Uber Eats orders from within the main Uber ridesharing app (rather than the separate Uber Eats app).[6] At first, the way in which transactions placed on the main Uber app were described on credit card bills made these transactions indistinguishable from rideshares, meaning that some percentage of Uber Eats transactions were missing from our data. This issue was resolved in August of 2019, such that orders placed through the main Uber app were correctly accounted for thereafter.

Figure 2 shows the weekly sales attributed to Uber Eats in the panel in 2019. Visually interpolating, it appears that roughly a quarter of sales were missing between May and August. Given the substantial magnitude of this missingness, we elect to exclude Uber Eats from our main event study model. Nonetheless, in Web Appendix 7, we re-estimate the event study model including Uber Eats; we find that the resulting implications of the effect of COVID-19 are qualitatively very consistent with our main analysis.

Additionally, the left-censored nature of the data poses problem for event study modeling: in particular, we do not know whether the first observed transaction for a given panel member is truly that member's first transaction in the delivery category. This complicates analysis, since our acquisition model models the day at which a panel members adopts a delivery service for the first time, and our repeat purchase model depends on the time of acquisition through the cohort and tenure covariates.

[5] For analyses where we exclude Uber Eats (namely the focal event study model), we further exclude 5 CBSAs where Uber Eats was the only platform operating. This results in the exclusion of another 315 panel members.

[6] `https://techcrunch.com/2019/06/04/uber-eats-uber-eats/`

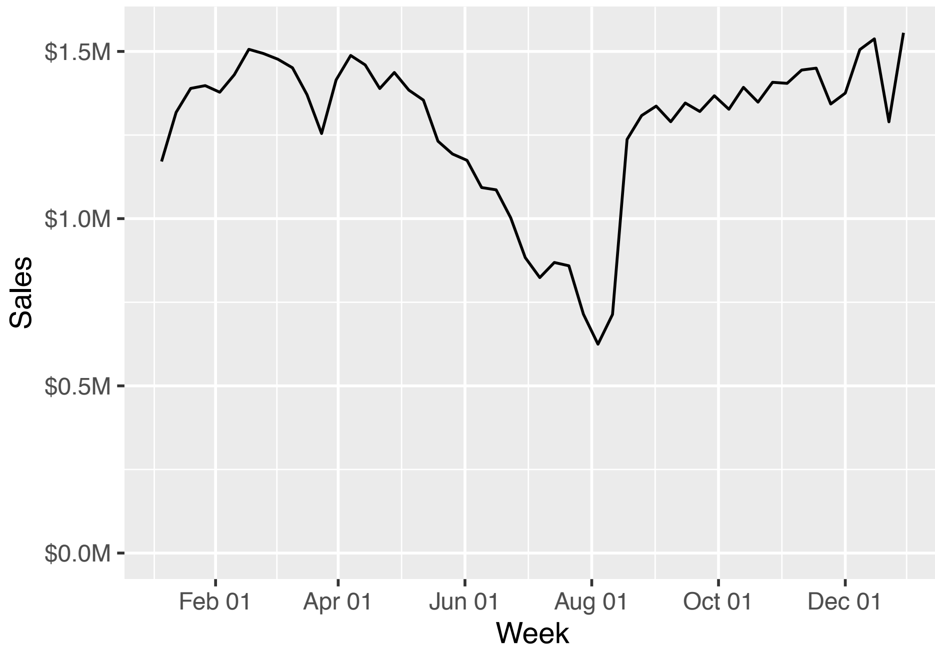
Table 1 Market Shares by Platform

Platform	Market Share	Notes
GrubHub	34.313%	
DoorDash	30.115%	
Uber Eats	17.954%	
Postmates	6.275%	Acquired by Uber Eats in 2020
Eat24	1.886%	Acquired by GrubHub in 2017
Slice	1.269%	
Waitr	1.261%	
Caviar	1.208%	Acquired by DoorDash in 2019
Delivery.com	0.734%	
Bite Squad	0.723%	Acquired by Waitr in 2018
Square	0.629%	
EatStreet	0.546%	
OrderUp	0.505%	Acquired by GrubHub in 2017-2018
Territory Foods	0.313%	
Goldbelly	0.286%	
Amazon Restaurants	0.266%	Discontinued in 2019
Foodler	0.206%	Acquired by GrubHub in 2017
Fooda	0.174%	
Delivery Dudes	0.174%	Acquired by Waitr in 2021
EzCater	0.169%	Omitted from analysis (B2B)
Munchery	0.166%	Discontinued in 2019
Ritual	0.157%	
MealPal	0.152%	
Foodsby	0.144%	
Tapingo	0.122%	Acquired by GrubHub in 2018
Thistle	0.097%	
Just Eat	0.088%	
Food Dudes Delivery	0.046%	
Chowbus	0.015%	
Hungry Panda	0.006%	

Note: Market shares are in terms of sales over the entire 7-year period of the panel data.

Accordingly, to address left-censoring, we follow McCarthy and Oblander (2021) in using the beginning of our dataset to filter out panel members with left-censored acquisition dates. In particular, we omit panel members whose first observed delivery transaction was between January 1 2014 and December 31 2015 from the acquisition model and from the main cross-cohort model, assuming that first transactions from January 1 2016 onwards are genuine acquisitions. Since very long interpurchase times are rare (in our data, 99.6% of interpurchase times are 2 years or less) and the customer base was relatively small prior to 2014 (32,926 panel members placed at least one delivery order in January 2014, as opposed to 471,944 in December 2020),

Figure 2 **Uber Eats weekly panel sales (2019)**



the probability that an initial order in 2016 or later is actually a repeat order (with the true initial order taking place prior to 2014), is very small.

Thus, in our event study model, we only include data from January 1 2016 to December 31 2020 (and only estimate the model using data up to December 31 2019). As detailed in Web Appendix 3, we modify our model specification to account for the customers with left-censored acquisition dates.

Additionally, we found that consumers with an excessively large number of transactions introduced numerical instabilities in the repeat purchase model, so to improve stability during estimation we excluded panel members who placed over 750 orders over the estimation period (i.e. between January 1 2016 and December 31 2020). This results in the exclusion of of 105 panel members who collectively account for only 0.45% of transactions, and so do not materially influence our inferences.

For our panel regression, left-censoring is not a concern, since we do not specify a generative model of purchasing behavior, and we only use data from 2020. Similarly, the missingness for Uber Eats in 2019 is not a concern, since we only use data from 2020, so we retain Uber Eats transactions for this analysis. We use the retained transactions and inferred locations to perform CBSA-level aggregations for the panel regression. Since we include data for the entirety of 2020, we only include CBSAs where, for each month in 2020, at least one delivery platform (among those covered by the Yipit Data) was operating. This results in daily observations for 792 CBSAs, slightly smaller than the 843 CBSAs used in the event study model, since some CBSAs were not serviced by a delivery platform until partway through 2020. The CBSAs that this filters out are generally very small - the 792 CBSAs collectively represent 99.7% of total spending across all 843 CBSAs - so this filter is unlikely to affect our inferences as well.

Appendix 2: Event study causal model and assumptions

In this section, we formally define the event study causal model and the assumptions required for valid inference. For now, we focus on inference within the credit/debit card panel population, then later discuss generalization to the US population in Web Appendix 8.

Denote by Y_{it}^0 the behavior of individual i at time t in a counterfactual world where the event of interest does not occur, i.e. where COVID-19 did not impact the United States. The behavior Y_{it}^0 considered depends on the process being modeled: e.g. it could be a binary indicator of whether i first adopted a delivery service at time t, the count of delivery orders placed by i at time t, or the total dollar amount spent on delivery orders by i at time t. Analogously, denote by Y_{it}^1 the behavior of individual i at time t in the (factual) world where the event of interest does occur.

Denote the time of event onset as t^* (in our case, January 1 2020). Under our event study setup, $Y_{it}^0 = Y_{it}^1$ for $t < t^*$: the worlds where the event do and do not occur only diverge upon onset of the event. The causal estimand we seek to estimate is the average treatment effect (ATE) of the event at time t, i.e. the incremental change in behavior attributable to COVID-19:

$$\tau_t = E\left[Y_{it}^1 - Y_{it}^0\right], \, t \geq t^*$$

The first term is estimable by sample analog: the sample average of observed behavior in the post-event period, $\bar{Y}_{it}^1$, is a consistent estimate of $E\left[Y_{it}^1\right]$.

The difficulty in estimating τ_t stems from estimation of the second term $E\left[Y_{it}^0\right]$, which has no sample analog: we have no observations that directly indicate what would have happened in 2020 had the pandemic not occurred. The idea behind the event study approach is that Y_{it}^0 is observable for $t < t^*$, such that we can estimate a predictive model of Y_{it}^0 on the pre-event period and then extrapolate the predictions to the post-event period $t \geq t^*$. Intuitively, if the model of Y_{it}^0 is well-specified, we can accurately forecast the counterfactual baseline under no event, recovering $E\left[Y_{it}^0\right]$ for the post-event period and allowing for consistent estimation of τ_t.

Formally, assume that the time series vector of behavior $(Y_{i1}^0, Y_{i2}^0, \dots, Y_{iT}^0)'$ is independently and identically distributed according to a distribution P^0 for all i (all expectations are with respect to the distribution P^0 unless otherwise indicated by a subscript); we propose a parametric model class $\{P_\theta^0 | \theta \in \Theta\}$ indexed by parameter θ. We denote this model as "the event study model," and the full specification is given below in Web Appendix 3. Our aim is to estimate parameter $\hat{\theta}$ to accurately capture the baseline trends in Y_{it}^0, such that we can use the predictions $E_{P_{\hat{\theta}}^0}\left[Y_{it}^0\right]$ to proxy for the true unobservable counterfactual baseline $E\left[Y_{it}^0\right]$ in the post-event period.

Since Y_{it}^0 is unobserved for $t \geq t^*$, we marginalize out $\left(Y_{it^*}^0, Y_{i(t^*+1)}^0, \dots, Y_{iT}^0\right)'$ and estimate $\hat{\theta}$ based on the pre-event observations $\vec{Y}_i^0 := \left(Y_{i1}^0, Y_{i2}^0, \dots, Y_{i(t^*-1)}^0\right)'$. In particular, we construct our estimate of $\hat{\theta}$ as an M-estimator under a loss function $\ell : \mathrm{Supp}\left(\vec{Y}_i^0\right) \times \Theta \to \mathbb{R}$:

$$\hat{\theta}\left(\vec{Y}_1^0, \vec{Y}_2^0, \dots, \vec{Y}_N^0\right) = \arg\min_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} \ell\left(\vec{Y}_i^0; \theta\right)$$

In our context, we perform (weighted) maximum likelihood estimation, so the loss function is the (weighted) negative log-likelihood. Using our estimated parameter, we then compute the predicted values $E_{P^0_{\hat{\theta}}}[Y^0_{it}]$ to approximate the true counterfactual baseline $E[Y^0_{it}]$, yielding an estimator for the desired ATE:

$$\hat{\tau}_t = \bar{Y}^1_{it} - E_{P^0_{\hat{\theta}}}\left[Y^0_{it}\right]$$

The first term is simply the empirical mean of Y^1_{it} in the post-event period, while the second term is the predicted value of Y^0_{it} in the post-event period based on model parameter $\hat{\theta}$ estimated on the pre-event data.

The key question is under what conditions $\hat{\tau}_t$ is a good approximation of the true ATE τ_t. We turn to this question next.

Denote by $\theta^\dagger$ the population optimal model parameter:

$$\theta^\dagger = \arg\min_{\theta \in \Theta} E\left[\ell\left(\vec{Y}^0_i; \theta\right)\right]$$

and denote by $\tilde{\tau}_t$ the estimand of the ATE under the assumed model:

$$\tilde{\tau}_t = E\left[Y^1_{it}\right] - E_{P^0_{\theta^\dagger}}\left[Y^0_{it}\right]$$

Under mild regularity conditions on the model class and loss function, $\hat{\theta} \to_p \theta^\dagger$, i.e. $\hat{\theta}$ is a consistent estimator of $\theta^\dagger$ as $N \to \infty$ (Newey and McFadden 1994). This implies that $E_{P^0_{\hat{\theta}}}[Y^0_{it}] \to_p E_{P^0_{\theta^\dagger}}[Y^0_{it}]$, so long as $E_{P_\theta}[Y^0_{it}]$ is smooth in θ, which in turn implies that $\hat{\tau} \to_p \tilde{\tau}$. That is, $\hat{\tau}$ is a consistent estimator of the quantity $\tilde{\tau}$.

The question that remains is whether the model-based estimand $\tilde{\tau}_t$ is a good approximation of the true causal estimand of interest τ_t. If the model is correctly specified, i.e. $\exists\ \theta^* \in \Theta$ such that $P^0_{\theta^*} = P^0$, then $\theta^\dagger = \theta^*$ and accordingly $E_{P^0_{\theta^\dagger}}[Y^0_{it}] = E[Y^0_{it}]$ and in turn $\tilde{\tau}_t = \tau_t$; thus, asymptotically, our estimator correctly recovers the true causal estimand of interest.

Of course, with any parametric model, some degree of misspecification is inevitable in practice. For instance, we expect that there will be some degree of "common shocks" that perturb the distribution Y^0_{it} in each time period t such as seasonality; our simple model smooths over such perturbations. When the model is misspecified, $P^0_{\theta^\dagger} \neq P^0$. Instead, under maximum likelihood estimation, $P_{\theta^\dagger}$ corresponds to the closest distribution to P^0 in an information theoretic sense: specifically, $\theta^\dagger$ minimizes the Kullback-Leibler divergence of the true distribution P^0 from the model distribution P^0_θ *in terms of the distribution of pre-event behavior* (White 1982, Buja et al. 2019).

Thus, intuitively, we are finding the distribution within our model class that most closely approximates the true distribution in the pre-event period. As long as this approximation is good, and the approximation continues to be good in the post-event period, then we have that $\tilde{\tau}_t \approx \tau_t$, even if the quantities are not exactly equal. For instance, if our model captures the overall shape of the curve of $E[Y^0_{it}]$ over time correctly, but fails to account for transient perturbations around the mean (e.g. due to seasonality or weather events), then we would accordingly expect $\tilde{\tau}_t$ to correctly capture τ_t on average (averaged across time periods), although the two quantities may differ period-by-period due to these transient perturbations.

As seen in Section 4.2, our model captures the overall empirical trends well, although it misses some apparent common shocks. Additionally, in Web Appendix 6 we show that in-sample closeness of the approximation does appear to generalize out-of-sample, since estimating the model up to end-of-year 2018 still results in good predictions of 2019 behavior, though again missing some apparent common shocks. Accordingly, we believe our model approximates the overall shape of the counterfactual baseline $E\left[Y_{it}^0\right]$ well, though not perfectly; in turn, we expect our causal effect estimates $\hat{\tau}$ to be approximately correct, although they may be biased in individual time periods; that is, day-by-day they may not precisely estimate treatment effects, but over longer time horizons (especially the full year of 2020), biases due to misspecification are smoothed over such that our estimates of the overall treatment effects are valid.

Lastly, we note that this event study approach works only when the distribution of Y_{it}^0 is sufficiently smooth and regular over time, and the event study model is not overly flexible. This is for two reasons. First, if the distribution is highly erratic over time, a parametric extrapolation is likely to fail: even if the model approximates the distribution closely in the pre-event period, if the distribution is prone to erratic shifts, this well-approximation may fail in the post-event period. Second, if the model is overly flexible (particularly if it allows for distributions that may shift in a non-smooth or irregular manner over time), identification may fail: estimation is performed based pre-event data; if the model allows, for instance, for purchase propensities to shift arbitrarily each year, then the parameters governing the 2020 shift would be unidentified. Thus, this approach works in our empirical setting because the behaviors that we model tend to be highly regular over time, as seen in Figure 1 of the main paper, and because we are able to capture these patterns using a highly parsimonious model; these factors result in our model being able to extrapolate out-of-sample with acceptable accuracy. Additionally, we note that while this approach relies on a parametric model, our estimates of τ_t are semiparametric in the sense that, while the baseline $E\left[Y_{it}^0\right]$ is modeled in a parametric fashion, we do not assume a parametric form for the "gap" τ_t over time.

Appendix 3: Full Model Specification

In this section, we fully specify the proposed models for customer acquisition, repeat purchasing, and spend summarized in Section 4.1 of the main paper.

3.1. Acquisition

We first specify a model that governs the duration of time that elapses from when individuals are "born" as prospects to when they are acquired into the delivery category. Unlike previous work, which has assumed all individuals in a geographic region (e.g., the United States) become prospects when commercial operations begins or when those individuals are born (Gupta et al. 2004, Schulze et al. 2012, McCarthy et al. 2017, McCarthy and Fader 2018), we explicitly account for the fact that individuals cannot be prospects until the category serves the area the prospect lives in. As we will discuss in detail in Web Appendix 4, we do so by utilizing a collection of alternative data sources to directly observe city-specific geographic expansion over time for all major delivery companies.

We denote all individuals who first become prospects in a particular month to be a "prospect pool." At the inception of the category, there is an initial prospect pool M(0) which is equal to the population size in the markets served when commercial operations begins, POP(0). Individuals from this prospect pool may adopt in future days $t = 1, 2$, and so on. The size of the prospect pool in future periods t are equal to the population sizes of the markets that the category entered into in those periods:[7]

$$\mathrm{M}(t) = \mathrm{POP}(t) - \mathrm{POP}(t-1), \quad t = 1, 2, \ldots \tag{1}$$

The marginal probability of customer acquisition at time t is then

$$\mathrm{A}(t) = \sum_{i=0}^{t-1} \mathrm{M}(i) \times [F_A(t-i|i) - F_A(t-i-1|i)]/\mathrm{POP}(T), \tag{2}$$

where $F_A(t-i|i)$ is the probability that an individual from prospect pool i is acquired by day t.

The duration of time from when each prospect pool is "born" to when individuals first adopt into the category is governed by a small extension of the "time of mass awareness" model of McCarthy and Fader (2018). The acquisition process differs based upon whether the individual became a prospect before or after the "time of mass awareness," denoted by $t^\star$. Individuals who become prospects before and after $t^\star$ are referred to as "early prospects" and "late prospects," respectively, with adoption propensities governed by the following set of assumptions:

- A proportion π_1 of all (i.e., early and late) prospects are "first wave intenders," with times until acquisition characterized by a Weibull(λ_1, c_1) distribution.
- Early prospects who are not first wave intenders have zero probability of acquisition before $t^\star$.
- At time $t^\star$, a proportion π_2 of early prospects who were not first wave intenders become "second wave intenders," with times until acquisition characterized by a Weibull(λ_2, c_2) distribution.
- A proportion $(1 - \pi_1) \times \pi_2$ of late prospects are also second wave intenders, with Weibull(λ_2, c_2)-distributed adoption times.
- All prospects who are not intenders after time $t^\star$ will never be acquired.

Given a prospect's homogeneous baseline propensities to be acquired (λ_1 and λ_2), their corresponding homogeneous acquisition shape parameters (c_1 and c_2), time-varying acquisition covariates ($\mathbf{X}_A(t+1,t') = [\mathbf{x}_A(t+1), \mathbf{x}_A(t+2), \ldots, \mathbf{x}_A(t')]$), and the coefficients associated with those acquisition covariates ($\boldsymbol{\beta}_A$), the probability that an individual from prospect pool t is acquired by the end of week t' is equal to

$$\begin{aligned} &\mathrm{F}_A[t'-t|t, \mathbf{X}_A(t+1,t'); t^\star, \pi_1, \pi_2, \lambda_1, \lambda_2, c_1, c_2, \boldsymbol{\beta}_A] \\ &= \begin{cases} \pi_1\left(1 - e^{-\lambda_1 B_1(t,t')}\right), & t < t^\star \text{ and } t' \le t^\star, \\ \pi_1\left(1 - e^{-\lambda_1 B_1(t,t')}\right) + (1-\pi_1)\pi_2\left(1 - e^{-\lambda_2 B_2(t^\star,t')}\right), & t < t^\star \text{ and } t' > t^\star, \\ \pi_1\left(1 - e^{-\lambda_1 B_1(t,t')}\right) + (1-\pi_1)\pi_2\left(1 - e^{-\lambda_2 B_2(t,t')}\right) & \text{otherwise,} \end{cases} \end{aligned} \tag{3}$$

where

$$B_n(t,t') = \sum_{i=t+1}^{t'} \left[(i-t)^{c_n} - (i-t-1)^{c_n}\right] e^{\boldsymbol{\beta}_A^T \mathbf{x}_A(i)}, \quad n \in \{1, 2\}. \tag{4}$$

We insert this expression into Equation 2 to perform estimation via maximum marginal likelihood.

[7] We assume that prospect pools form monthly, since our data is not at a high enough frequency to reliably infer market entry at a more granular frequency.

There are two differences between this specification and that of McCarthy and Fader (2018). First, late prospects may be a part of both "first wave" and "second wave" segments. Second, $t^\star$ is a continuous-valued parameter. In the original TMA model, late prospects could only be a part of the "second wave" segment, and $t^\star$ was a discrete valued parameter. These modifications make the resulting likelihood function differentiable, allowing us to use gradient-based optimization methods to estimate model parameters. In contrast, McCarthy and Fader (2018) use a profile likelihood approach, estimating all parameters excluding $t^\star$ repeatedly over a grid of many possible values of $t^\star$, which is much more computationally intensive.

3.2. Repeat purchasing

As noted in Section 4.1 of the main paper, we use the Extended Pareto/NBD (EPNBD) model of Bachmann et al. (2021b). The EPNBD model extends the Pareto/NBD model (PNBD, Schmittlein et al. 1987, Fader et al. 2005), the most widely-recognized latent attrition model used to forecast repeat customer purchasing in non-subscription settings, by allowing the baseline purchase and retention propensities of customers to vary over time through proportional hazards. The appeal of this model is that it allows for time-varying covariates while retaining the computational efficiency associated with having a closed-form model likelihood.

The Pareto/NBD model assumes that customers have two states - they are "alive" for some period of time, after which they churn (i.e., become permanently inactive), where churn is assumed to be an absorbing state. The time until customer i churns is governed by an Exponential(μ_i) distribution. While alive, customer ordering follows a Poisson Process with intensity λ_i, with both propensities λ_i and μ_i varying across customers according to independent gamma distributions.

The EPNBD model extends the PNBD model by allowing λ_i and μ_i to vary over time. Customer i's time until churn is governed by a Cox-Exponential(μ_i, γ_c) distribution, and the number of purchases customer i makes over that period is governed by an inhomogeneous Poisson Process with parameters (λ_i, γ_p), where γ_c and γ_p are the parameters corresponding to the time varying covariates for the two processes. All other assumptions of the EPNBD are the same as those of the PNBD. As such, denoting customer i's time invariant baseline purchase and churn propensities by λ_{i0} and μ_{i0}, respectively,

$$\lambda_i(t) = \lambda_{i0} \exp\left(\mathbf{x}_i(t)_p^T \gamma_p\right) \qquad \text{and} \qquad \mu_i(t) = \mu_{i0} \exp\left(\mathbf{x}_i(t)_c^T \gamma_c\right)$$

where

$$\lambda_{i0} \sim \text{Gamma}(r, \alpha) \qquad \text{and} \qquad \mu_{i0} \sim \text{Gamma}(s, \beta)$$

As we discuss in more detail in the next section, we use the EPNBD to allow for two covariates – one time invariant covariate for the customer's time of adoption (i.e., their "cohort number"), and one time varying covariate for the number of days that have elapsed since the customer was acquired (i.e., their "tenure").

We implement this model using a custom modification of the `CLVTools` package (Bachmann et al. 2021a). The `CLVTools` package made the EPNBD model publicly available in the `R` programming language. Our extensions decreased the compute time of the likelihood function by a factor of over 1,000 times in our empirical application.

3.3. Spending given purchase

We model customer spend given a purchase (i.e., AOV) took place as a homogeneous log-linear function of customer cohort and tenure. That is,

$$\log \mathrm{AOV}_{it} = \beta_{s,0} + \beta_{s,cohort}\mathrm{Cohort}_{it} + \beta_{s,tenure}\mathrm{Tenure}_{it} + \epsilon_{it}$$

where, denoting customer i's acquisition time by Cohort_i,

$$\mathrm{Tenure}_{it} = t - \mathrm{Cohort}_i.$$

While it would be straightforward to specify a more complex model for customer spending, we do not do so because our empirical application does not require it. As is visually evident from Figure 1 in the main paper, category-level AOV before the onset of COVID-19 was very well behaved. Indeed, the proposed model specification is accurate to within 1.1% and 2.3% in-sample and out-of-sample, respectively, over a one-year holdout period. Moreover, allowing for unobserved heterogeneity in customers' baseline spending propensities would not change our unconditional spending forecasts, nor would they change our estimates of the impact of COVID-19.

Appendix 4: Implementation details: event study approach

In this section, we describe in detail the implementation steps we performed to carry out the event study approach, process by process. For each process, we obtain bias corrected parameter estimates and standard errors via non-parametric bootstrap. We sample with replacement 1.83 million times from the 1.83 million customers in our panel data set. For each bootstrapped data set, we estimate each of the aforementioned three processes. We obtain 1,000 bootstrapped parameter estimates in this way. The bias-corrected parameter estimates and standard errors are equal to the empirical mean and standard deviation of the bootstrapped parameter estimates, respectively.

Acquisitions. As mentioned in Section 3.1, we estimate the parameters of the acquisition model by maximizing the marginal likelihood of the daily category-level customer acquisitions data from January 1 2016 through December 31 2019. The category effectively began commercial operations in June 2004, when a subsidiary of GrubHub began commercial operations in Boston. As such, there are 4,202 days over which the data is left censored, and 1,461 days of observable daily acquisitions data.

As alluded to in Section 2 of the main paper, we use three data sources to determine geographic market entry, which we discuss in turn below.

1. We systematically obtain historical versions of the websites of DoorDash, Uber Eats, GrubHub and Postmates using the Wayback Machine.[8] From June 2004 to December 2015, with some variation by company, these websites provided complete listings of every city they operated in. The evolution of these listings over time determines which markets were covered at which times.

[8] `https://archive.org/web/`

2. We use data from YipitData, a leading alternative data firm, to observe market coverage more recently. Yipit provided us with the name and location of every restaurant listed on a number of food delivery platforms – DoorDash, Uber Eats, GrubHub, Postmates, Waitr, and Bite Squad – on a monthly basis from October 2018 through December 2020. We obtained the CBSA associated with each zip code, assuming a CBSA is entered/covered when there is at least one restaurant delivery company serving restaurants in that CBSA. In this way, we obtain all CBSAs served each month for each major platform in more recent periods.

3. A number of markets were right censored and left censored using the Wayback Machine and Yipit data sets, respectively. For example, no delivery platform had listed Somerset Pennsylvania (CBSA 43740) as a market served as per the Wayback Machine, implying we had not observed market entry by December 2015. However, Somerset was being served as of the beginning of the Yipit data in October 2018 – through Yipit, we observe that both GrubHub and UberEats served restaurants in this CBSA that month. In Somerset, then, delivery must have occurred after December 2015 but may have occurred before October 2018. For all CBSAs that are right and left censored in this way, we impute their market entry date using data from Earnest Research, as this data allows us to observe the date and location associated with purchases at all restaurant delivery companies over the intermediate period from January 2016 through September 2018. We assumed that a particular market was entered when two conditions were met: (1) there were at least 10 orders that had been placed in that CBSA in the panel and (2) at least one percent of the panel residing in the area had been acquired. A minimum level of purchase activity in the delivery category must occur within a particular CBSA to ensure that restaurant delivery companies actually entered it, and that the purchase activity was not an artifact of measurement error (e.g., a panel member living in an uncovered market who orders delivery while on a short trip to a covered market). In the case of Somerset, for example, these criteria were first met in April 2017, more than a year after the end of the Wayback Machine data, and more than a year before the beginning of the Yipit data.

In summary, the date a particular CBSA was entered is equal to:

1. the earliest date that CBSA was covered in the Wayback Machine data, for CBSAs represented in the Wayback Machine data

2. the earliest date that CBSA was covered in the Yipit data, for CBSAs not represented in the Wayback Machine data and not present in the Yipit data in October 2018 (i.e., were not left censored in the Yipit data)

3. the earliest date that CBSA had a sufficient amount of purchase activity in the Earnest Research data (as described above), for CBSAs not represented in the Wayback Machine data and present in the Yipit data October 2018 onwards (i.e., were left censored in the Yipit data)

For each individual in our panel, we obtain their modal CBSA, as measured through the Earnest Research data. They are first assumed to be prospects on the market entry date associated with their modal CBSA.

The likelihood associated with customers who were not acquired during the observed data window accounts for the possibilities that they were either acquired prior to 2016 (left-censoring) or were not acquired by the end of 2019 (right-censoring).

Repeat purchasing. To account for left-censoring in our data, we first partition customers into two mutually exclusive, collectively exhaustive sets of customers – customers who were first acquired before January 1

2016 ("legacy" customers) and customers who were first acquired January 1 2016 onwards ("non-legacy" customers). We have complete transactional data for all non-legacy customers. Legacy customers, in contrast, are left censored, making their aggregate-level behavioral patterns different from those of non-legacy customers. There are 218,206 and 914,369 legacy and non-legacy customers, respectively. We estimate separate repeat purchasing models for legacy and non-legacy customers.

We include two covariates in the repeat purchase model – the customer's tenure and cohort number. The former represents the number of years that had elapsed since the customer was born, and as such, is a time-varying, individual-specific covariate. The latter represents the number of years that elapsed from December 31st 2015 until when the customer was born, and thus is individual-specific but time invariant. Both covariates are discretized to the weekly level for computational efficiency.

For non-legacy customers, we include cohort number as a covariate in both the purchase and attrition processes of the EPNBD, but only include customer tenure as a covariate for the purchase process. As alluded to in Heckman (1991) (and empirically observed in Braun et al. 2015), duration dependence and unobserved heterogeneity are not separately empirically identifiable in hazard models. Our inclusion of these covariates further necessitated our partitioning of the customer base.

For non-legacy customers, we also observe that younger cohorts contribute very little the marginal likelihood, because the length of the observation period is short for young customers. As a result, parameter estimates obtained via MLE tend to overweight older cohorts relative to younger ones, resulting in very good fits for the former cohorts but relatively poor fits for the latter. This is problematic in our empirical setting, because the validity of the impact decomposition we perform in Section 4.4 in the main paper is predicated upon sensible fits for young cohorts. We re-weight the likelihood, inverse weighting each customer by the length of that customer's observation period, to put approximately the same weight upon the likelihoods of all customers. That is, denoting the length of calibration period and the observable repeat purchase data for customer i by T_i and $\mathbf{D}_{i,purch}$, respectively, and the parameters of the repeat purchase model by $\boldsymbol{\theta}_{purch}$, instead of estimating the parameters that maximize the marginal loglikelihood of the data,

$$\hat{\boldsymbol{\theta}}_{purch} = \text{argmax}_{\boldsymbol{\theta}_{purch}} \sum_{i=1}^{I} \ell(\boldsymbol{\theta}_{purch}|\mathbf{D}_{purch}),$$

we instead maximize the weighted loglikehood:

$$\hat{\boldsymbol{\theta}}_{purch} = \text{argmax}_{\boldsymbol{\theta}_{purch}} \sum_{i=1}^{I} \frac{1}{T_i} \ell(\boldsymbol{\theta}_{purch}|\mathbf{D}_{i,purch}).$$

Because legacy customers are all within the same cohort, we only include customer tenure (i.e., we do not include cohort number) as a covariate in our model specification for their repeat purchasing.

Spending given purchase. As with the repeat purchasing model, all customers are categorized as legacy and non-legacy customers. For each transaction, we obtain the customer's tenure and cohort number, defined as in the repeat purchase model. For non-legacy customers, we model log(AOV) as a linear function of cohort number and tenure, as in the repeat purchase model, and only include tenure as a covariate when modeling AOV for legacy customers because legacy customers share the same cohort. We estimate the model via ordinary least squares, which is equivalent to maximum likelihood estimation assuming a homoskedastic normal error distribution.

Appendix 5: Parameter estimates and in-sample fits

In this section, we provide the parameter estimates and associated standard errors for the acquisition, repeat purchase, and spend models underlying the main event study analysis, performing estimation using all data through the end of calendar year 2019 (Table 2). We also provide additional figures summarizing the in-sample fit of our model.

Table 2 Parameter Estimates

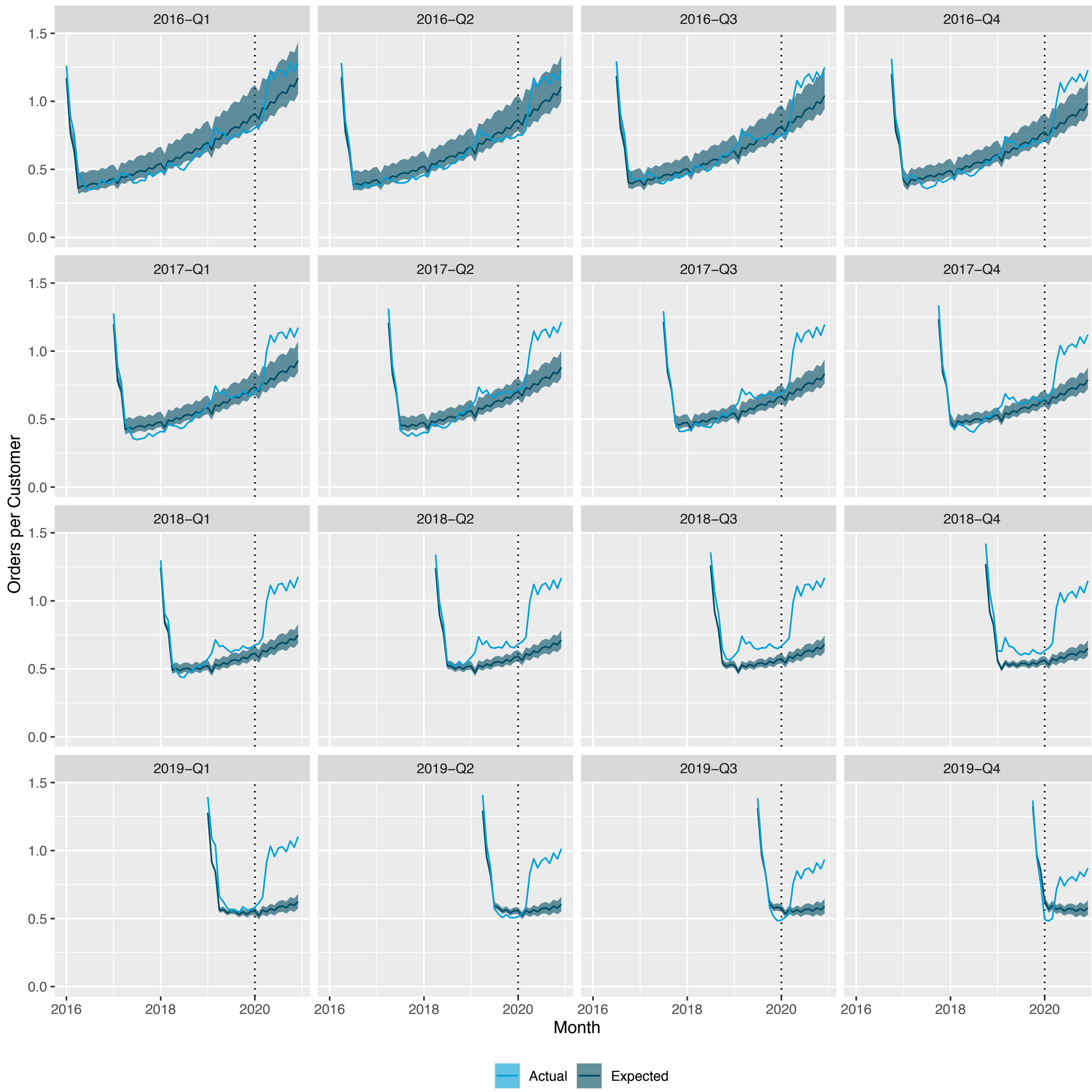
	Acquisition			Repeat Order			Spend	
	Est	SE		Est	SE		Est	SE
			Non-legacy customers			Non-legacy customers		
λ_1	.1923	2.1883	r	.4919	.0084	$\beta_{s,0}$	3.4961	.0025
c_1	5.2254	2.4571	α	.1198	.0119	$\beta_{s,cohort}$	.0001	.0009
π_1	.8337	.1259	$\gamma_{p,cohort}$	.1725	.0324	$\beta_{s,tenure}$	.0404	.0007
λ_2	1.2760	1.6275	$\gamma_{p,tenure}$	.3319	.0121			
c_1	3.3325	1.0697	s	.3626	.0540			
π_2	.9937	.0551	β	2.3425	.5874			
$t^\star$	9.6264	4.6264	$\gamma_{c,cohort}$	.5061	.0893			
			Legacy customers			Legacy customers		
			r	.6150	.0355	$\beta_{s,0}$	3.5327	.0019
			α	.0635	.0063	$\beta_{s,tenure}$	.0402	.0005
			$\gamma_{p,tenure}$	.1506	.0077			
			s	.1281	.0368			
			β	1.8233	.6257			

Note: for interpretability, we report the results assuming an annual unit of time. While model fits and forecasts are identical to those when we use a daily unit of time, some parameter estimates (and their associated standard errors) change by a multiplicative factor. We multiply λ_1 and λ_2 by 365.25^{c_1} and 365.25^{c_2}, respectively. We divide $t^\star$, α, and β by 365.25. Finally, we multiply all covariates – $\gamma_{p,cohort}$, $\gamma_{p,tenure}$, $\gamma_{c,cohort}$, $\beta_{s,cohort}$, and $\beta_{s,tenure}$ – by 365.25.

We infer that virtually all prospects will eventually be acquired, with 84% of all prospects being a part of the "first wave" and the remaining 16% being a part of the "second wave." Customers' baseline spending patterns improve significantly as a function of customer tenure – non-legacy customers that are still alive after one year have repeat purchase propensities that are 39% ($\gamma_{p,tenure}$) higher and AOVs that are 4% higher ($\beta_{s,tenure}$) than that of newly-acquired customers, all else equal. However, customers' repeat purchase patterns are significantly worsening across cohorts – more recently-acquired customers have a significantly higher baseline propensity to churn ($\gamma_{c,cohort}$), which is only partially offset by a marginally higher baseline propensity to purchase while alive ($\gamma_{p,cohort}$).

To visualize these trends and to validate the in-sample cohort-specific goodness of fit of the repeat purchase model, we plot in Figure 3 monthly tracking plots associated with all quarterly acquisition cohorts. There is

Figure 3 **Monthly actual and expected orders per customer by quarterly acquisition cohort**



a sharp decline in orders per customer after a cohort is first “born” because of customer churn. Thereafter, the number of orders per customer gradually increases for older cohorts, as the aforementioned favorable tenure dynamics more than outweigh continued customer churn. Younger cohorts (e.g., those born in 2019), in contrast, do not exhibit a noticeable increase in orders over time, as the favorable tenure dynamics are increasingly offset by unfavorable cross-cohort dynamics. This figure also underscores the robustness of COVID’s impact, as there was a significant increase in ordering across every cohort.

It is readily apparent that the repeat purchase model does not capture all cohort-specific variation in order dynamics. In particular, while the model adequately captures the underlying trends in ordering for the 2016, 2017, and 2019 acquisition cohorts, it slightly underpredicts order patterns for the 2018 cohort. While this may be the case, we nevertheless capture the aggregate trends very well, as evidenced through the top-right panel of Figure 2 in the main paper. Our main results hinge less upon capturing in-sample order patterns for every cohort, and more upon capturing the general cross-cohort trends on average so that our model provides valid predictions for what customers acquired in 2020 would have done had COVID not come about. Given our emphasis upon prediction, model parsimony is an important consideration. In this regard, the proposed model performs very well – it largely captures customer purchase dynamics across 16 quarterly cohorts over multiple years with one 7-parameter model, and does an adequate job of modeling cross-cohort trends in customer purchasing as we move from very young to very old cohorts. This supports the validity of the repeat purchase model, and in turn, the decomposition we perform in Section 4.4 of the main paper.

Appendix 6: Predictive validation analysis: event study approach

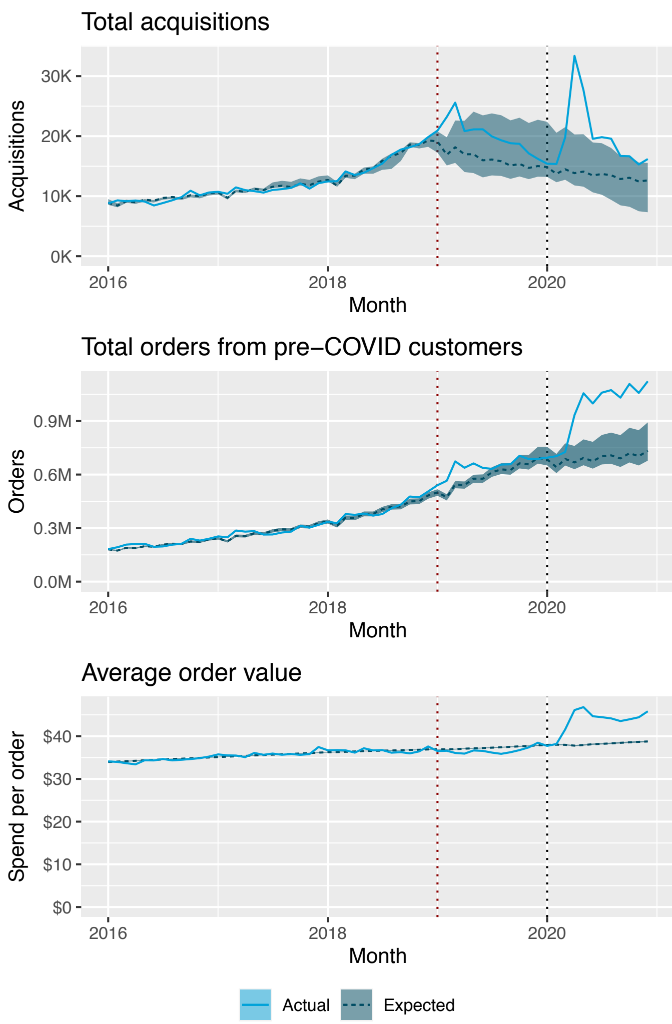
A key assumption underlying our event study analysis is that forecasts from our predictive model reliably capture the counterfactual baseline level of activity that we would have expected had COVID-19 not occurred. One way in which we can evaluate this is through a predictive validation exercise. That is, we train our predictive model upon some pre-COVID data (the 'calibration period'), forecast what will happen over the remaining pre-COVID data (the 'holdout period') conditional upon the calibration period data, then compare our forecasts to what we actually observed over the holdout period. In the analysis that follows, we train our model upon all data from January 1st 2016 through December 31st 2018, leaving all of calendar year 2019 as a holdout period.

We summarize the results of this predictive validation analysis in Figure 4. The upper, middle, and lower graphs in this figure plot actual and expected total customer acquisitions, total orders from pre-COVID customers, and average order value, respectively. We plot total orders from pre-COVID customers conditional upon actual acquisitions to evaluate the goodness of fit of the repeat purchase model on its own (i.e., without contamination from the acquisition model). In each plot, we overlay the 95% confidence interval, as in Figure 2 of the main paper, while the dotted red and black vertical lines denote the end of the calibration and holdout periods, respectively.

Figure 4 suggests that our forecasts, after taking into account their uncertainty, do a reasonable job of covering the observed data in the holdout period and capturing its underlying baseline trends. That said, the model slightly underpredicts acquisitions and orders in early 2019. The sharp increase in acquisitions (and thus in orders) that we observe during this period did not stem from any publicly disclosed corporate action. Its impact was transitory, and the observed data falls back to baseline levels by the end of the holdout period. While the spend model slightly overpredicts AOV in 2019, its out-of-sample mean absolute percentage error (MAPE) was nevertheless low, at 2.3%.

The customer acquisition forecast is highly uncertain (i.e., the confidence interval for the acquisition forecast is relatively wide). This is consistent with prior literature, which has shown that predicting future customer adoption prior to its observed peak is subject to a high degree of uncertainty and even bias (Heeler and Hustad 1980, Van den Bulte and Lilien 1997). Our model is not immune to these issues. However, as we calibrate upon more data after customer acquisition peaks, uncertainty generally decreases and the accuracy of our point estimate increases. This is evident when we consider other calibration periods between

Figure 4 **Monthly actual and expected customer acquisitions: December 31st 2018 end of calibration period**



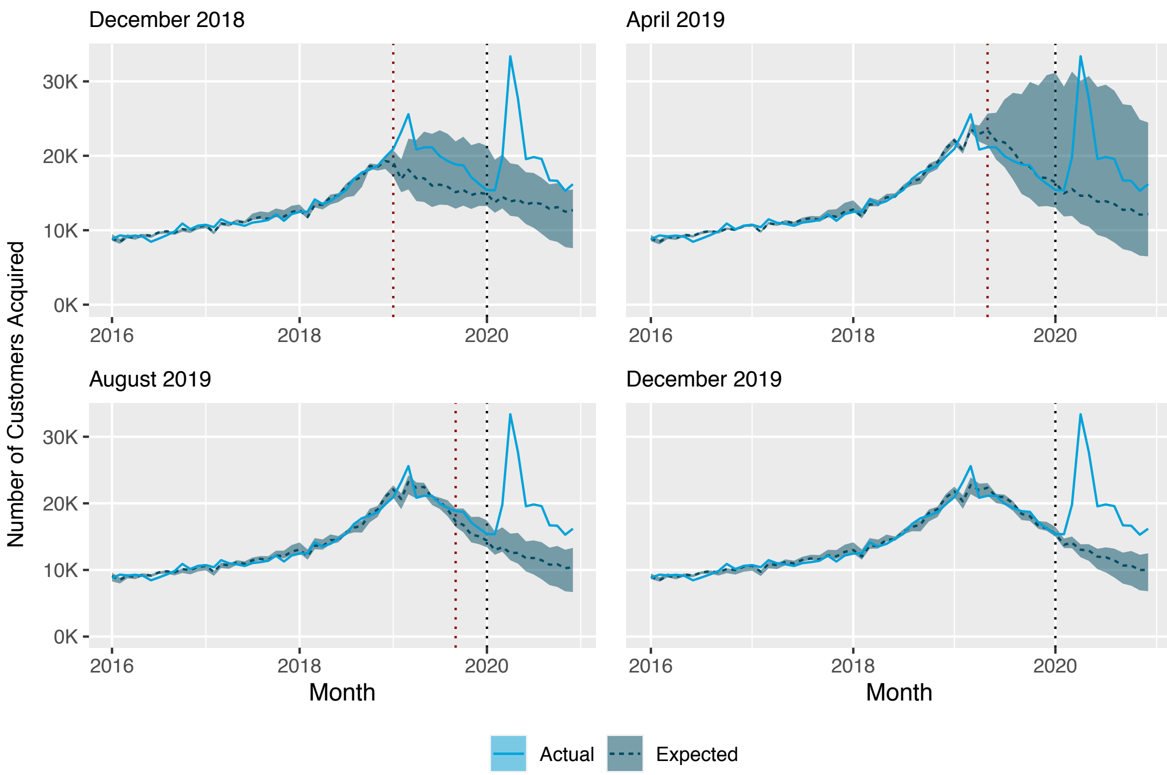
Note: 95% confidence bands are provided alongside point predictions. Dotted red and black vertical lines represent the end of the calibration and holdout periods, respectively.

December 2018 and December 2019, as shown in Figure 5. While the April 2019 forecast (top-right panel of Figure 5) is uncertain, as it is unclear whether March 2019 is a global or local maximum, its point estimate is largely unbiased. The August 2019 forecast (bottom-left panel of Figure 5) is also largely unbiased but with a significantly narrower confidence interval. Training on all data through the end of 2019 (bottom-right panel of Figure 5), when we actually predict the impact of COVID, acquisitions had fallen very steadily for nine months, making its underlying baseline trend self-evident.

Appendix 7: Robustness check: results including Uber Eats

While we did not include Uber Eats in our main analysis in Section 4.2 due to our inability to disambiguate Uber Eats transactions from Uber ridesharing transactions in part of 2019 because of a change that Uber made to its app at the time, we include them here as a robustness check. Analogous to Figure 2 in the main paper, Figure 6 plots actual total acquisitions (top-left), total orders (top-right), average spend per order (bottom-left), and total spend (bottom-right), as well as baseline fits and forecasts, with bootstrapped 95%

Figure 5 Actual versus expected customer acquisitions: rolling calibration periods



confidence intervals overlaid. We see that the general trends, and our goodness of fit, are largely the same whether we include Uber Eats transactions or not. As such, while including Uber Eats transactions could bias inferences of the absolute impact of COVID-19 (i.e., Section 4.4 in the main paper), the results when we include Uber Eats are nonetheless qualitatively consistent with our main results.

Appendix 8: Panel representativeness and selection correction

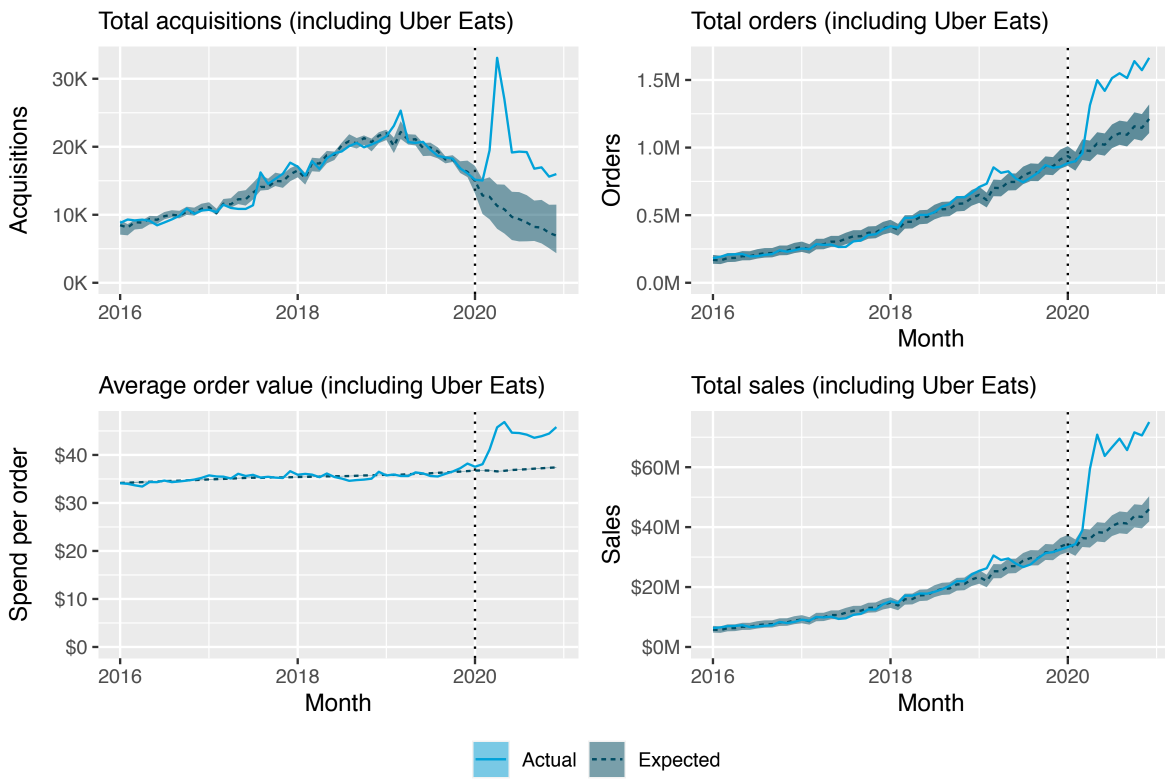
In this section, we present further results on the representativeness of the credit/debit card panel and describe how we implement the selection correction to translate our results to the population level.

8.1. Comparison of panel and population

We present a more detailed comparison of the panel data and quarterly aggregate data obtained from SEC filings (S-1 and 10-Q filings). GrubHub has disclosed their total sales on a quarterly basis dating back to Q1 2013, while DoorDash has disclosed their total sales on a quarterly basis dating back to Q1 2018. By calculating the equivalent aggregate statistics from our panel data, we can assess the nature and severity of selection bias in the panel data.[9]

[9] While the other major player in the market, Uber Eats, is also a publicly traded company that reports some quartertly statistics, its SEC data is not amenable to comparison with the panel. Uber Eats has substantial overseas business and does not separately report US and international sales, such that we cannot calculate comparable statistics from the panel, which only covers US consumers. Conversely, sales at GrubHub and DoorDash are almost entirely driven by domestic business, such that the panel and SEC data are comparable. Additionally, as mentioned in Web Appendix 1.5, we have data quality issues with Uber Eats due to some orders being indistinguishable from rideshares.

Figure 6 **Actual versus expected total acquisitions, total orders, AOV, and total spend, including Uber Eats transactions**



Note: Vertical dotted lines represent the end of calendar year 2019. Confidence intervals are obtained via nonparametric bootstrap.

To make the aggregate statistics comparable in scale between the two data sources, we normalize them by their respective relevant populations so that the statistics are on a per capita basis: we normalize the SEC data by the 2019 US Census population estimate (the most recent estimate available as of the writing of the paper) across all CBSAs covered by a major delivery platform, and the panel data by the total number of panel members inferred to primarily live in one of these CBSAs.

The resulting aggregate time series are shown in Figure 7 on logarithmic scale. Visually, it is clear that while the panel tends to oversample customer activity, there is a very strong correspondence between panel and population trends: the panel seems to be nearly proportional to the population. The strong correspondence between the panel and population provides reassurance that our inferences based on the panel are directionally consistent with the population at large.

As discussed in Section 4.3 of the main paper, we further benchmark the representativeness of the panel through a simple log-log regression for each of these time series, regressing the population sales for each quarter on the corresponding panel sales:

$$\log \text{Sales}_q^{Population} = \beta_0 + \beta_1 \log \text{Sales}_q^{Panel} + \varepsilon_q$$

Figure 7 **Panel and population sales per capita (GrubHub, DoorDash, and both combined)**

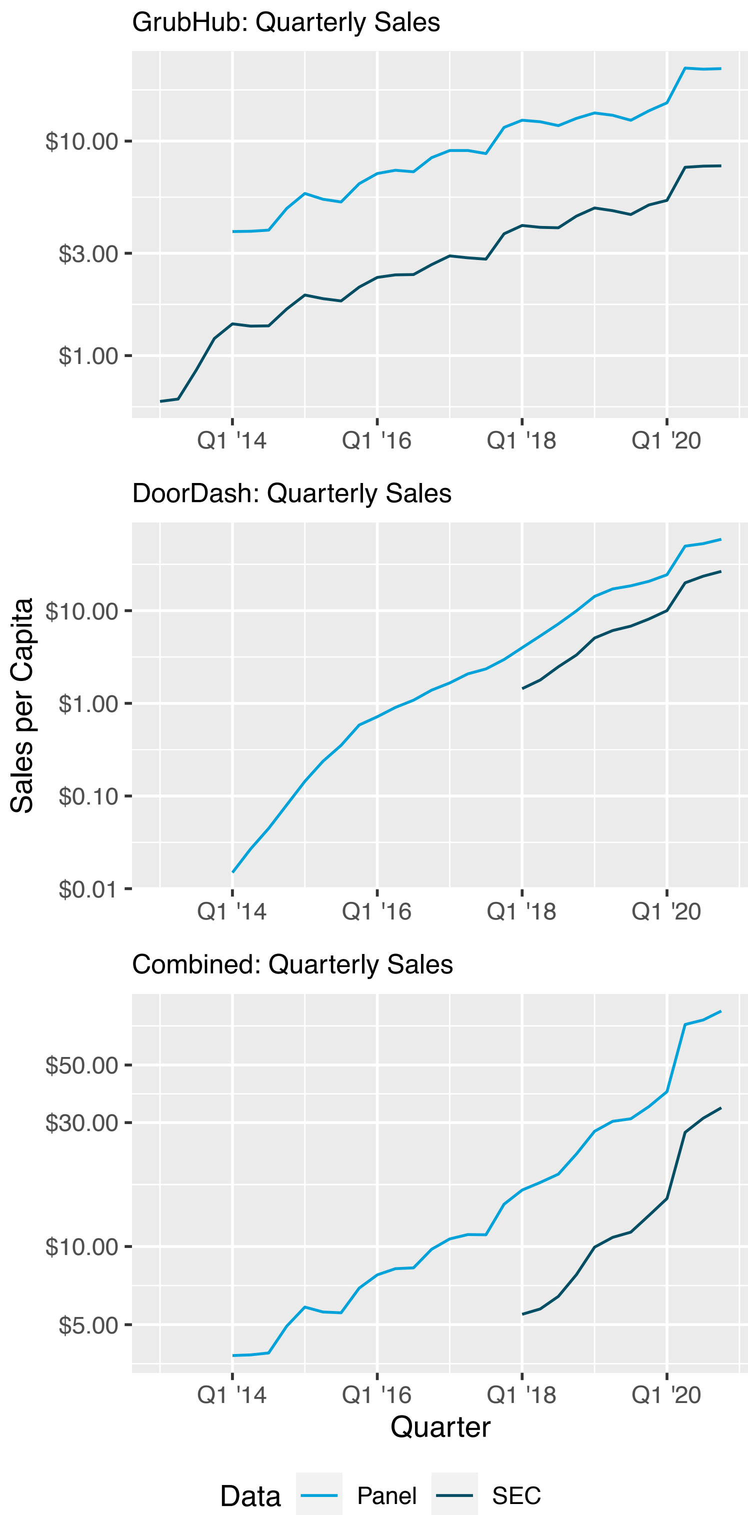


Table 3 Population-panel log-log regressions

Company	$\hat{\beta}_0$ (SE)	$\hat{\beta}_1$ (SE)	T	R^2
GrubHub	−1.079 (0.068)	1.001 (0.032)	28	99.1%
DoorDash	−1.257 (0.071)	1.100 (0.023)	12	99.7%
Combined	−1.516 (0.053)	1.146 (0.015)	12	99.8%

Note: standard errors are heteroskedasticity and autocorrelation robust (HAR) with autocorrelation bandwidth selected automatically as in Andrews (1991) using the implementation of Zeileis (2006).

The fit of the regression is informative as to how well the panel tracks the population trends, while the coefficient indicates whether the two time series are proportional: $\beta_1 = 1$ indicates that the panel is perfectly proportional to the population (up to the error term), $\beta_1 > 1$ indicates that trends in the population are underrepresented/dampened in the panel, and $\beta_1 < 1$ indicates that trends in the population are overrepresented/amplified in the panel.

Table 3 presents the results of these regressions. Notably, all three regressions yield good fits, with R^2 in excess of 99%; this reinforces that the panel data has strong directional correspondence with population trends.

In all cases, $\hat{\beta}_0 \ll 0$, indicating that the panel tends to overrepresent sales. Additionally, $\hat{\beta}_1 \gtrapprox 1$ slightly, indicating that the panel sales are nearly proportional to the population sales, with a slight tendency for changes at the population level to be understated in the panel.

Overall, these results suggest that while the panel evidently exhibits some selection bias, it has strong directional consistency with the population. As such, we are confident that our inferences are diagnostic of the US restaurant delivery industry as a whole. Next, we discuss how we use these results to generalize our event study inferences to the population as a whole.

8.2. Selection correction and effect decomposition

In addition to being diagnostic of selection bias, the log-log regressions presented above provide us with a means of correcting for selection bias and generalizing results to the population. In particular, the regressions provide a mapping from panel sales to population sales; as such, we can estimate the effect of COVID-19 at the population level by plugging our inferred sales under different counterfactual scenarios into the regression equation.

In particular, we start by computing the counterfactual sales per capita in each quarter of 2020 under six scenarios:

1. Baseline under no effect of COVID-19 (i.e. direct prediction from the event study model)
2. Conditional on observed 2020 purchase frequencies by customers acquired prior to 2020, holding average order value and new customer behavior fixed at event study model predictions
3. Conditional on observed 2020 behavior by customers acquired prior to 2020, holding new customer behavior fixed at event study model predictions
4. Conditional on 2020 behavior by customers acquired prior to 2020 and acquisitions of new customers in 2020, holding new customer purchase frequencies and average order values fixed at event study model predictions

5. Conditional on 2020 behavior by customers acquired prior to 2020, acquisitions of new customers in 2020, and purchase frequencies of new customers; holding new customer average order values fixed at event study model predictions
6. Conditional on all 2020 behavior (i.e. the empirically observed outcomes)

We then use the log-log regression estimates from Table 3 to estimate the corresponding population-level sales under each counterfactual scenario. We use the estimates for the combined regression (summing together the quarterly sales of GrubHub and DoorDash) since it is the closest to category-level that we can observe. In particular, assuming the linear log-log relationship is well-specified and the error variance is small, we calculate:

$$\widehat{\text{Sales}}_q^{Population} = e^{\hat{\beta}_0}\left(\widehat{\text{Sales}}_q^{Panel}\right)^{\hat{\beta}_1}$$

This yields population-level estimates of sales for each quarter of 2020 under different counterfactual scenarios, on a per capita basis and exclusive of Uber Eats. To translate to an absolute dollar value and account for the missing Uber Eats orders, we multiply these figures by the relevant population and divide by one minus the estimated market share of Uber Eats in 2020 (based on the credit/debit card panel data). These rescaled figures (summed across quarters where reported at the annual level) are the estimated population sales figures reported in-text in Section 4.4 and in Figure 4.4 of the main text, while the effect sizes reported in Figure 4 of the main text consist of the differences between these numbers (e.g. the overall effect of COVID-19 is the difference between the estimated sales figures for scenarios 6 and 1; the effect attributable to pre-COVID customers is the difference between scenarios 3 and 1; and so on).

These figures make two key assumptions to achieve valid generalization. First, we assume that the log-log regression relating the population and panel quarterly combined sales at GrubHub and DoorDash also holds for the entire category (i.e. all delivery companies combined). We cannot directly test this assumption, since we do not observe ground truth data for the category as a whole; however, given the fairly consistent results across the two companies, and given that these two companies account for the majority of restaurant delivery aggregator transactions (64.4% of total sales in our data), we expect that this mapping holds at least approximately for the category as a whole.

Second, we assume that the effect of COVID-19 on Uber Eats is comparable to the effect of COVID-19 on the rest of the category (such that simply rescaling by one minus Uber's market share corrects for its missingness). In our panel data, Uber Eats had a market share of 21.3% in 2020, compared to 18.6% in 2019; summer 2019 is also when Uber Eats was missing a proportion of transactions in the credit/debit card data, such that the actual market share was somewhat higher. As such, Uber Eats' market share appears to have been similar in 2019 and 2020, suggesting that their COVID-19 effect was comparable to other companies (if Uber Eats' COVID-19 effect had differed substantially from the rest of the category, their market share also would have changed substantially). Thus, we expect that this simple correction factor should be approximately correct.

Lastly, we need to compute standard errors to quantify the uncertainty in these estimates. There are two main sources of uncertainty which we need to propagate: uncertainty in the panel sales estimates (due to estimation uncertainty in the event study model parameters) and uncertainty in the log-log regression parameter estimates. We do so using the delta method.

The difficulty with the delta method is that it depends on the covariance between all pairs of input quantities; we cannot estimate the covariance between $\widehat{\text{Sales}}_q^{Panel}$ and $\hat{\beta}_0$ or $\hat{\beta}_1$, since these quantities and their variances are estimated using different procedures on different datasets: $\widehat{\text{Sales}}_q^{Panel}$ is estimated by maximum likelihood on the individual-level credit/debit card panel data with bootstrapped standard errors, while $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimated by ordinary least squares on aggregate time series data (SEC filings and aggregated credit/debit card panel data) with autocorrelation-robust sandwich standard errors.

Accordingly, to obtain valid standard errors, we assume the worst case covariance structure where the unknown covariances indicate perfect positive correlation (when the partial derivatives of the output quantity with respect to the two inputs have the same sign) or perfect negative correlation (when the partial derivatives have the opposite sign). This yields conservative standard errors that asymptotically upper bound standard errors under any covariance structure. All standard errors with respect to population-level quantities reported in the paper are calculated in this way, using the delta method while assuming the worst case perfect positive/negative correlations between quantities with unknown covariances.

Appendix 9: Identification of FE regression

In this section, we elaborate upon the identifying assumptions of our panel regressions. As stated in the main text, spatiotemporal variation in the trajectory of COVID-19's economic and behavioral impacts enable us to identify the differential effects of different mechanisms on delivery ordering behavior. We discuss the exogenous sources of variation in our mechanisms of interest, as well as possible endogenous sources, in turn.

First, as described in the main text, our identification of the effect of unemployment on delivery spending comes from comparing CBSAs whose unemployment rates jumped, on the margin, more or less than the national average (and thereafter have recovered slower or faster than the national average). Exogenous variation comes from the composition of occupations differing across CBSAs: CBSAs where many jobs are in industries hard-hit by the pandemic will have a higher jump in unemployment compared to CBSAs where most people work in industries that were less affected by the pandemic.

For this variation to be exogenous, we require that these differences in job composition are uncorrelated with other time-varying factors that affect delivery ordering behavior, conditional on our controls. The main effect of baseline variation in socioeconomic status and behavioral traits of consumers who self-select into different industries is controlled for by CBSA fixed effects; only the change in employment status during the pandemic factors into identification. Additionally, while unemployment also includes unemployment from the restaurant industry, which determines the supply side of restaurant food delivery, since we also include restaurant employment in our regression, the variation in unemployment used to identify unemployment effects will be residual of that explained by the restaurant industry. Apart from this, there could be second-order supply side effects in that, in CBSAs with higher unemployment, workers who are out of a job could turn to the gig economy as a short-term alternative, increasing the supply of delivery drivers and shortening waiting times (thus positively impacting delivery). While we expect such second-orders effects to be small in magnitude, to the extent that such effects are present, they would bias us towards more conservative estimates of the unemployment effect (since the effect is negative overall, while this supply effect would be positive).

Second, our identification of the effect of stay-at-home behavior comes from comparing CBSAs where stay-at-home behavior has jumped more or less, relative both to the national trend and to CBSA-specific baseline stay-at-home rates. This residual variation comes from differences in the perceived danger of going outside and the extent to which consumers work in industries where working from home is feasible; in turn, perceived danger may be driven by factors such as population density, local government restrictions, and political beliefs.

The main effects of static variables such as population density and political beliefs are controlled for by CBSA fixed effects, so these sources of variation are exogenous so long as they do not correlate with other time-varying factors that affect delivery ordering. Local government restrictions are time-varying and may affect not only stay-at-home behavior but also the restaurant industry through restrictions on dine-in. Similarly, stay-at-home is simultaneously determined with dine-in since, by definition, if a consumer goes to dine in at a restaurant, they did not stay at home. However, given that we also include dine-in and restaurant employment in our regression, the variation used to identify the effect of stay-at-home is residual of dine-in levels and other changes in the restaurant industry due to government regulations.

Third, our identification of the effect of restaurant employment comes from variation across CBSAs and over time in how restaurant owners and managers responded to COVID-19 related government restrictions and economic shocks; many restaurants decided to close temporarily or permanently, and those that were open adjusted their staffing levels to accommodate financial constraints and negative shocks to demand.

While restaurant employment captures both dine-in capacity and capacity to prepare food for delivery (the latter being our mechanism of interest), since we include dine-in levels in our regression, the variation identifying the employment effect is residual of dine-in capacity. Additionally, we note a potential simultaneity issue: managers may adjust staffing levels based on anticipated demand for delivery, such that increased employment is caused by delivery demand rather than the other way around. However, given that our unit of analysis is at the daily level, such concerns are somewhat alleviated: managers are unlikely to be able to anticipate ahead of time daily-level demand shocks in excess of common shocks such as holidays (which are captured by our day fixed effects), and presumably have limited capacity to adjust staffing levels on the same day, since employee schedules are typically determined ahead of time.

Nonetheless, if demand shocks are strongly autocorrelated over time, and managers anticipate this autocorrelation, then simultaneity may still be present (i.e. managers observe higher than expected demand one day, so they increase staffing levels to plan for increased demand the next day), leading to an upward bias in the estimated effect of employment. To assess this possibility, we conduct a robustness check where we include lagged delivery sales as a regressor, such that the variation used to identify the employment effect is residual of the previous day's demand shock. The results of this robustness check are presented in Table 6 and discussed further in Web Appendix 10.2. While there is significant evidence for first-order autocorrelation in demand shocks, the extent of autocorrelation is fairly small in magnitude (we estimate that 10% higher demand than expected based on CBSA and day fixed effects translates to 1.1% higher demand the next day), and importantly the coefficient of restaurant employment on sales stays approximately the same. This suggests that the estimates are not simply driven by reverse causality from managers adjusting staffing levels based on past demand.

Fourth, identification of the effect of dine-in levels comes from variation across CBSAs and over time in government restrictions: while most state governments issued states of emergency and ordered the shut down of on-premises dining around the same time in March 2020, they differed substantially in the timings of when they allowed outdoor dining and indoor dining to reopen (and in if/when they reinstated restrictions amid resurgences of COVID-19 cases in the fall), enabling identification.

Changes in government restrictions on dine-in may co-occur with changes in stay-at-home restrictions; but, since we include stay-at-home rates in our regression, the dine-in effect is identified by residual variation in excess of what is explained by correlation with stay-at-home rates. Additionally, it is possible that delivery companies strategically targeted marketing efforts based on dine-in restrictions, e.g. by increasing advertising immediately after dine-in restrictions are put in place to target consumers who are missing the dine-in experience. However, GrubHub and DoorDash's quarterly marketing spends (as reported in their SEC filings) do not suggest this to be the case: the year-on-year change in marketing spending in the second calendar quarter of 2020 was not higher than in previous quarters for either GrubHub or DoorDash. Accordingly, strategic marketing does not appear to explain our results.

Beyond these possible confounders, there is an obvious simultaneity concern with dine-in: because dine-in is a substitute for delivery, a positive shock to delivery demand (unrelated to dine-in) could cause a spurious decrease in dine-in due to substitution. This concern is mitigated in our empirical setting because the dine-in market is temporarily in a shortage state: many restaurants have permanently closed, many are only allowing outdoor dining (whose capacity may be further constrained by weather), and most have some form of spacing requirements that drastically reduce the number of tables that they can serve. As a result, the primary determinants of dine-in levels during this time are shifters of supply-side constraints (e.g., government restrictions and weather preventing outdoor dining), alleviating demand-side endogeneity concerns.

To provide further credibility to the argument that dine-in is primarily driven by supply-side factors, we estimate a a fixed effects regression models with instrumental variables (FE-IV), instrumenting dine-in levels by policy variables summarizing government restrictions on restaurant dining (e.g., disallowing indoor dining) to isolate this supply-side source of variation from possible endogenous variation. The results are reported in Web Appendix 10.1. The estimated coefficients are nearly identical in terms of size and magnitude to the non-instrumented fixed effects regressions, providing empirical support for the notion that using an instrumental variable for dine-in is not necessary given the dine-in supply constraints that were evident in 2020.

Another potential mechanism is the number of restaurants listed on restaurant delivery platforms. If the onset of COVID caused many restaurants to list themselves on delivery platforms, the resulting "expansion of inventory" available to consumers could have driven their increased purchasing. To evaluate this, we use aforementioned restaurant listings data provided by Yipit to assess whether we observed a sharp increase in restaurant listings in the wake of the pandemic.

Figure 8 **Total number of restaurants listed in category, October 2018 - December 2020**

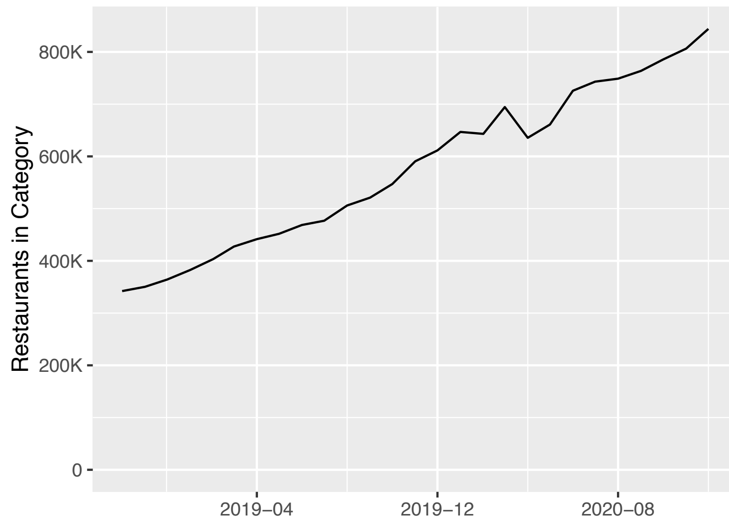


Figure 8 shows the evolution of total restaurant listings over time for the restaurant food delivery category as a whole. Before the pandemic, restaurant listings had steadily been increasing over time, and there is no evidence of change in the rate of growth after the pandemic began. This would suggest that an increase in the supply of restaurant listings was not a driver of the increase in sales during COVID. In fact, the opposite was temporarily true – there was a small *decline* in total listings in the immediate aftermath of the onset of COVID-19.

To provide insight into what drove the decline in listings at the beginning of the onset of the pandemic, Figure 9 shows the corresponding total number of restaurant listings over time at the four largest platforms. We can see from this figure that DoorDash and GrubHub had sizable declines in their respective restaurant listings at the beginning of the pandemic, while Uber Eats and Postmates did not. The number of listings declined at DoorDash and GrubHub, as confirmed with the data vendor, because these firms listed many restaurants on their platforms without explicit consent from the restaurants.

Thus, while there are many possible reasons why the mechanisms of interest may be confounded, we believe that the fixed effects and other control variables, as well as the accompanying robustness checks, demonstrate that we have plausibly exogenous variation for identification.

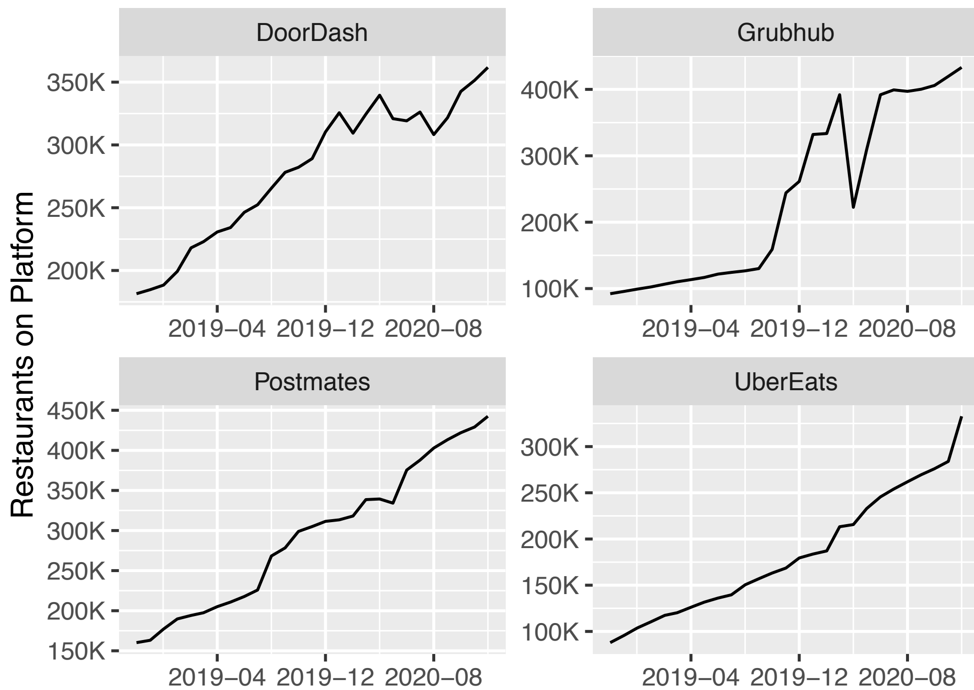
Appendix 10: Supplemental FE regression results

10.1. FE-IV model with policy instrumental variable

As discussed in Web Appendix 9, there is a potential simultaneity issue for the dine-in coefficient, which we argue is mitigated by the restaurant dine-in industry being in a temporary shortage state during the pandemic. To empirically support this argument, we apply instrumental variables for restaurant dine-in. Our excluded instruments capture government restrictions on dine-in, and are specified as follows:

1. Indoor dining allowed: a dummy variable for whether indoor dining was allowed in CBSA c on day t.

Figure 9 **Total number of restaurants listed by platform, October 2018 - December 2020**



2. Indoor dining allowed: a dummy variable for whether there were restrictions on indoor dining (e.g. spacing, capacity, and/or party size limits) in effect in CBSA c on day t.

3. Outdoor dining allowed: a dummy variable for whether outdoor dining was allowed in CBSA c on day t.

As described in Web Appendix 1.4, we use data on COVID-related state government mandates, collected by Fullman et al. (2021), to derive these instruments. These variables are supply-side shifters of restaurant capacity. Since we argue that capacity constraints obviate simultaneity issues, with government restrictions being a major driver of variation in capacity, analysis with these instruments allows us to isolate this supply-side variation and assess whether this argument is valid.

The model specification is identical to that shown in Section 5.1 of the main text, except that the dine-in coefficient is instrumented for with these three dummy variables as excluded instruments. We estimate all four FE-IV models by weighted two-stage least squares (W2SLS). Results are reported in Table 4, with the first stage model results reported in Table 5.

The first stage F-statistic on excluded instruments is 61.7 ($p \approx 0$), indicating that our instruments are relevant. The strongest instrument is whether indoor dining is allowed, though whether outdoor dining is allowed still significantly drives dine-in levels. Whether other indoor dining restrictions are in place (besides indoor dining being allowed/disallowed) does not appear to correlate with dine-in levels, but we leave it in the regression since we have no a priori reason to omit it. We obtain very similar results when we use the indicator for whether indoor dining is allowed as the sole instrument.

Table 4 FE-IV Regression Model Estimates

DV	Acquisitions		Orders		Order Size		Sales	
Regressor	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Unemployment	0.005	(0.017)	−0.174***	(0.043)	0.014**	(0.005)	−0.214***	(0.044)
Stay-at-home	0.183**	(0.062)	−0.347**	(0.111)	0.063***	(0.015)	−0.356***	(0.106)
Restaurant supply	0.247***	(0.051)	0.196***	(0.042)	0.020	(0.014)	0.268***	(0.051)
Restaurant dine-in	−0.306***	(0.071)	−0.189***	(0.046)	−0.026	(0.017)	−0.244***	(0.049)
N	289,872		289,872		209,438		289,872	

Note: asterisks denote level of significance (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$). All specifications include CBSA and day fixed effects. Standard errors are two-way clustered, robust to within-CBSA and within-day dependence. N is smaller for the order size regression as average order size is ill-defined for CBSA-day pairs with zero observed delivery orders.

The estimates in Table 4 are qualitatively very similar to our results from Section 5.3 of the main text, with the one exception of acquisition, where the IV estimate of the dine-in effect is substantially larger in magnitude than the WLS estimate; this could be due to IV procedures estimating the local average treatment effects (LATE), which would mean that the “compliers” (in this context, consumers who dine in at restaurants when restrictions are lifted) are more likely to adopt delivery services in response to being unable to dine in restaurants compared to the general population (Angrist and Imbens 1994).

The strong consistency between the WLS and IV results suggests that our argument about shortages mitigating simultaneity issues has some empirical support, in turn validating the use of using simple WLS estimates without the need for instruments.

10.2. FE model with lagged sales

As noted in Web Appendix 9, there is also a simultaneity concern with the restaurant supply effect, since restaurant managers may strategically set staffing levels to match expected demand for delivery food. Since managers presumably have limited ability to change staffing levels within-day, they may instead anticipate autocorrelation in demand shocks, setting staffing levels based on observed demand from prior days.

To test whether this anticipatory behavior explains our restaurant supply coefficient, we perform another robustness check where we include first-order lagged sales (more precisely, the log of sales per capita, the same variable we use as our main dependent variable of interest) in our regressions, to see if this substantially shifts the restaurant supply coefficient. The specification is identical to our main regression model except for this added regressor.

The results are given in Table 6. Lagged sales are a significant regressor in all cases, indicating that delivery behavior has significant serial correlation (although the coefficient is not large in magnitude: the largest coefficient is 0.115). Our other coefficients are largely unchanged. Although a first-order lag term does not capture all possible autocorrelation structures, the robustness of our estimates to this lagged term seems to indicate that, while sales exhibits modest autocorrelation, simultaneity due to managers strategically setting staffing levels based on this autocorrelation does not seem to be the primary driver of our restaurant supply coefficient.

Table 5 FE-IV First Stage Estimates

Regression	Acquisitions/Orders/Sales		Order Size	
Regressor	Coef.	SE	Coef.	SE
Unemployment	−0.016	(0.019)	−0.016	(0.020)
Stay-at-home	−0.280**	(0.101)	−0.286**	(0.103)
Restaurant supply	0.638***	(0.033)	0.639***	(0.034)
Indoor dining allowed	0.173***	(0.027)	0.172***	(0.027)
Outdoor dining allowed	0.063*	(0.029)	0.063*	(0.030)
Indoor dining reduced	−0.007	(0.012)	−0.007	(0.013)
N	289,872		209,438	
F(Excluded instruments)	61.73		59.84	

Note: asterisks denote level of significance (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$). All specifications include CBSA and day fixed effects, with dine-in levels as the dependent variable. Standard errors are two-way clustered, robust to within-CBSA and within-day dependence. The first stage estimates for acquisitions, orders, and sales are identical since the endogenous regressor and instruments are shared. The order size first stage is slightly different since some observations are excluded due to missingness in the second stage.

Table 6 Lagged Sales Model Estimates

Variable	Acquisitions		Orders		Order Size		Sales	
Regressor	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Unemployment	0.013	(0.017)	−0.159***	(0.041)	0.015**	(0.005)	−0.187***	(0.039)
Stay-at-home	0.263***	(0.053)	−0.329**	(0.116)	0.076***	(0.016)	−0.307**	(0.103)
Restaurant supply	0.128***	(0.030)	0.187***	(0.041)	0.001	(0.008)	0.223***	(0.044)
Indoor dining allowed	−0.123***	(0.028)	−0.185***	(0.051)	0.004	(0.009)	−0.192***	(0.046)
Lagged sales	0.008***	(0.002)	0.071***	(0.005)	0.003***	(0.001)	0.115***	(0.007)
N	289,080		289,080		208,942		289,080	

Note: asterisks denote level of significance (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$). All specifications include CBSA and day fixed effects. Standard errors are two-way clustered, robust to within-CBSA and within-day dependence. N is smaller for the order size regression as average order size is ill-defined for CBSA-day pairs with zero observed delivery orders.

References

Andrews DW (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society* 817–858.

Angrist J, Imbens G (1994) Identification and estimation of local average treatment effects. *Econometrica* 41(2):467–475.

Bachmann P, Kuebler N, Meierer M, Naef J, Oblander E, Schilter P (2021a) CLVTools: tools for customer lifetime value estimation. URL `https://cran.r-project.org/package=CLVTools`.

Bachmann P, Meierer M, Näf J (2021b) The role of time-varying contextual factors in latent attrition models for customer base analysis. *Marketing Science (Articles in Advance)* .

Braun M, Schweidel DA, Stein E (2015) Transaction attributes and customer valuation. *Journal of Marketing Research* 52(6):848–864.

Buja A, Brown L, Kuchibhotla AK, Berk R, George E, Zhao L, et al. (2019) Models as approximations ii: A model-free theory of parametric regression. *Statistical Science* 34(4):545–565.

Fader PS, Hardie BG, Lee KL (2005) RFM and CLV: Using iso-value curves for customer base analysis. *Journal of Marketing Research* 42(4):415–430.

Fullman N, Bang-Jensen B, Reinke G, Magistro B, Castellano R, Erickson M (2021) State-level social distancing policies in response to COVID-19 in the US, version 1.118. URL `http://www.covid19statepolicy.org/`.

Gupta S, Lehmann DR, Stuart JA (2004) Valuing customers. *Journal of Marketing Research* 41(1):7–18.

Heckman JJ (1991) Identifying the hand of past: Distinguishing state dependence from heterogeneity. *The American Economic Review* 81(2):75–79.

Heeler RM, Hustad TP (1980) Problems in predicting new product growth for consumer durables. *Management Science* 26(10):1007–1020.

McCarthy D, Fader P (2018) Customer-based corporate valuation for publicly traded noncontractual firms. *Journal of Marketing Research* 55(5):617–635.

McCarthy D, Fader P, Hardie B (2017) Valuing subscription-based businesses using publicly disclosed customer data. *Journal of Marketing* 81(1):17–35.

McCarthy DM, Oblander ES (2021) Scalable data fusion with selection correction: An application to customer base analysis. *Marketing Science (Articles in Advance)* .

Newey WK, McFadden D (1994) Large sample estimation and hypothesis testing. *Handbook of Econometrics* 4:2111–2245.

Schmittlein DC, Morrison DG, Colombo R (1987) Counting your customers: Who are they and what will they do next? *Management Science* 33(1):1–24.

Schulze C, Skiera B, Wiesel T (2012) Linking customer and financial metrics to shareholder value: The leverage effect in customer-based valuation. *Journal of Marketing* 76(2):17–32.

Van den Bulte C, Lilien GL (1997) Bias and systematic change in the parameter estimates of macro-level diffusion models. *Marketing Science* 16(4):338–353.

White H (1982) Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the econometric society* 1–25.

Zeileis A (2006) Object-oriented computation of sandwich estimators. *Journal of Statistical Software* 16(9).