Persistence of Consumer Lifestyle Choices: Evidence from Restaurant Delivery During COVID-19

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

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

We study the impact of the COVID-19 pandemic on customer purchasing behavior in the restaurant delivery category in the United States and, particularly, the extent to which pandemic-driven shocks to purchasing behavior have persisted even as consumers' lifestyles are returning to a "new normal." We apply age-period-cohort (APC) models, commonly used in sociology but with limited adoption in marketing, to nonparametrically decompose customer behavior into acquisition cohort, tenure, and calendar time effects. This approach, in conjunction with an event study approach, allows us to flexibly estimate the time-varying effects of ubiquitous events such as COVID, where no contemporaneous "control group" is available, by comparing the behavior of cohorts acquired at different times. We estimate that pandemic-related disruptions initially more than doubled customers' aggregate spending. As of Summer 2021, the lift in sales was still approximately one-third of its early pandemic peak, despite evidence that the pandemic-driven shocks driving the initial lift had abated, due to a sizeable segment of customers continuing to purchase at elevated rates. This persistent behavior is consistent with some consumers having formed habits around delivery. However, this persistence appears to gradually decay over time, suggesting that habits acquired during COVID may not last indefinitely.

Key words: customer relationship management; COVID-19; age-period-cohort model; persistence; habit formation

---

1. Introduction

When COVID-19 began to overtake the United States in early 2020, consumers went through sudden and extreme shifts in their lifestyles: with consumers reluctant or unable to go to restaurants, entertainment venues, and brick-and-mortar retail, industries enabling at-home lifestyles such as delivery services and home entertainment saw unprecedented growth in consumer demand (Valinsky 2020).

Though the unprecedented effects of the pandemic are still being felt nearly two years after its onset, many aspects of consumer lifestyle have returned to a "new normal." An important and highly timely question in the minds of firms, policymakers, investors, and consumer behavior researchers is the extent to which consumer lifestyle changes driven by COVID are persistent: are the effects of the pandemic transient, or do they represent a secular change in consumers' way of life?

We study this question in the context of the restaurant delivery category: with consumers reluctant to leave their homes and most state and local governments implementing restrictions on on-premise dining, many turned to delivery as a convenient and safe alternative. With restaurants reopened and many consumers reverting to pre-pandemic lifestyles, it is natural to wonder the extent to which consumers have made persistent habits out of ordering food delivery.

Our goal in this paper is to shed light upon these questions by quantifying the magnitude of the pandemic's impacts on the restaurant delivery category, and assessing how much pandemic-induced changes in lifestyle have persisted even after their causes have largely dissipated. We do so through a novel application of age-period-cohort (APC) models, which have seen limited adoption within marketing but are commonly used in other social sciences, in combination with an event study approach. In doing so, we find that the pandemic generated massive growth for the restaurant delivery category, which was primarily driven by pre-existing customers. Early on in the pandemic, pre-existing customers more than doubled their spending relative to the counterfactual baseline where there was no pandemic, due to a large lift in the number of active customers and smaller lifts in order frequency and average order size. Analysis of the mechanisms of these changes suggests that they are in part attributable to substitution away from restaurant dine-in.

Since February 2021, as the economy has reopened, new customer adoption has fallen to below pre-pandemic levels, newly acquired customers are spending as if the pandemic had never occurred, and restaurant dine-in activity have reverted to nearly pre-pandemic levels; despite this, some customers continue to purchase at elevated rates. In particular, the lift in pre-existing customer sales as of July 2021 was about one-third of what it was at the peak of the pandemic.

We take these results to suggest that some consumers have formed habits around delivery, leading to a persistent lift in sales despite the initial pandemic-related shocks that pushed consumers towards delivery having worn off. Nonetheless, this persistence appears to gradually decay over time, suggesting that the habits are not permanent, eventually being supplanted by other life events that lead consumers to change their lifestyles. These findings suggest that the long-term impact of the pandemic on the delivery category is likely to be limited; more broadly, changes in consumption patterns spurred on by the pandemic may be similarly persistent but not indefinite.

2. Relevant Literature

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 (Baker et al. 2021, Chen et al. 2021, Chetty et al. 2020, Dunn et al. 2021). 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 (e.g., decomposing the sales effect into customer adoption, purchase frequency, and average order value) beyond aggregate-level spending, and (3) study the longer-term persistent impacts of the pandemic.

Beyond COVID-19, our research contributes to the broader marketing and economics literature on consumer habit formation and persistence of behavior. In experimental settings, behavioral and health economics researchers have found that short-term interventions can sometimes promote persistent lifestyle changes through habit formation (Loewenstein et al. 2016, Charness and Gneezy 2009, Yang and Long Lim 2018). Research on state dependence in brand choice similarly finds that consumers exhibit persistence in brand choice beyond what is explained by unobserved heterogeneity (Dubé et al. 2010). We add to these bases of knowledge by demonstrating the extent to which the pandemic induced consumers to form persistent habits around food delivery.

We use age-period-cohort (APC) models to decompose changes in consumer preferences over time into age effects (changes associated with customer tenure, such as customer attrition and duration dependence), period effects (changes associated with calendar time, such as seasonality and COVID), and cohort effects (time-invariant differences between acquisition cohorts). These models have a long history of use within the social sciences, developed by demographers and primarily used by sociologists (Fosse and Winship 2019).

While APC models have been used in prior marketing literature, they have mostly been used in a demographic sense, with cohort and age representing the literal birth year and biological age of the consumers (Fukuda 2010, Rentz and Reynolds 1991). Separately, many papers within the customer relationship management (CRM) literature address similar questions involving cross-cohort dynamics with calendar time and customer tenure effects, in which "cohort" and "age" are relative to time of adoption (Gopalakrishnan et al. 2017, Schweidel et al. 2008). However, to the best of our knowledge, no extant CRM literature has used an explicit APC modeling approach. We extend upon prior literature by merging these two areas, applying an APC model to customer acquisition cohorts, then combining it with an event study approach (Corrado 2011) to understand the causal impact of a major regime shift, the COVID-19 pandemic, on customer behavior.

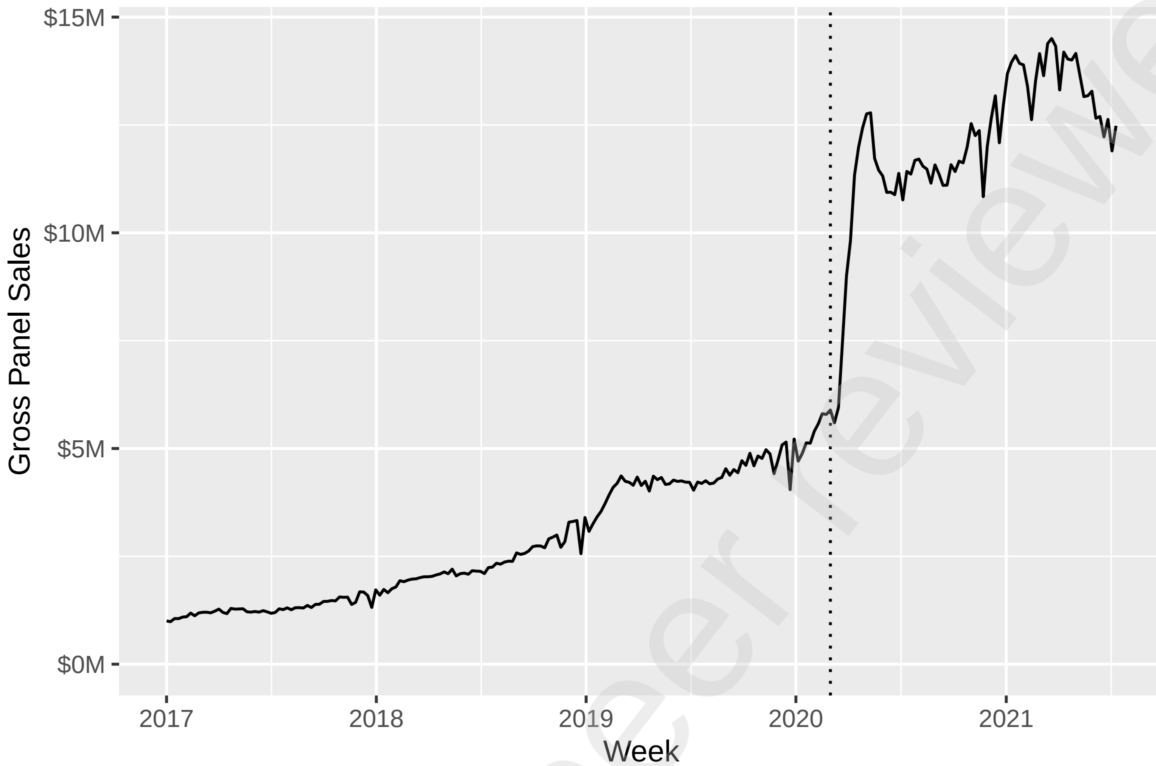
3. Descriptive Trends

Our primary data source for this research is credit/debit card transaction data from Earnest Research, a leading data analytics company. Through this data, we observe all restaurant delivery transactions, possibly across multiple credit/debit cards, for 1.59 million panel members in the United States, at the daily level from January 1st, 2017, to July 17th, 2021. Additionally, we observe the monthly modal location of each panel member at the core-based statistical area (CBSA) level. Further detail on this dataset, other supplementary datasets used in our analyses, and data processing steps are given in Web Appendix A. Our spending data is one of the largest – both in terms of number of panel members and observation window – in the extant COVID spending impact literature.

Figure 1 shows weekly restaurant delivery spending observed in our panel over time. We see a dramatic shift in consumer behavior early on in the pandemic, with weekly delivery spending more than doubling in a matter of weeks. While delivery spending fell slightly in the early Summer of 2020, it did not fall to anywhere near pre-pandemic levels and even moved above its previous peak in early 2021. Evidently, the pandemic generated a significant lift in the restaurant delivery category. This lift appears to have been sustained, to some extent, for over a year.

While the time series of aggregate spending shows descriptive evidence of the magnitude of COVID's impact on the restaurant delivery category, richer insights on the nature of the COVID effect and its change over time can be gained by breaking down spending into

Figure 1 **National Trends in Restaurant Delivery Spending**



Note: Figure depicts total weekly sales within the restaurant delivery category across all members of the credit/debit card panel from January 2017 to July 2021. Vertical dotted line represents March 1st, 2020.

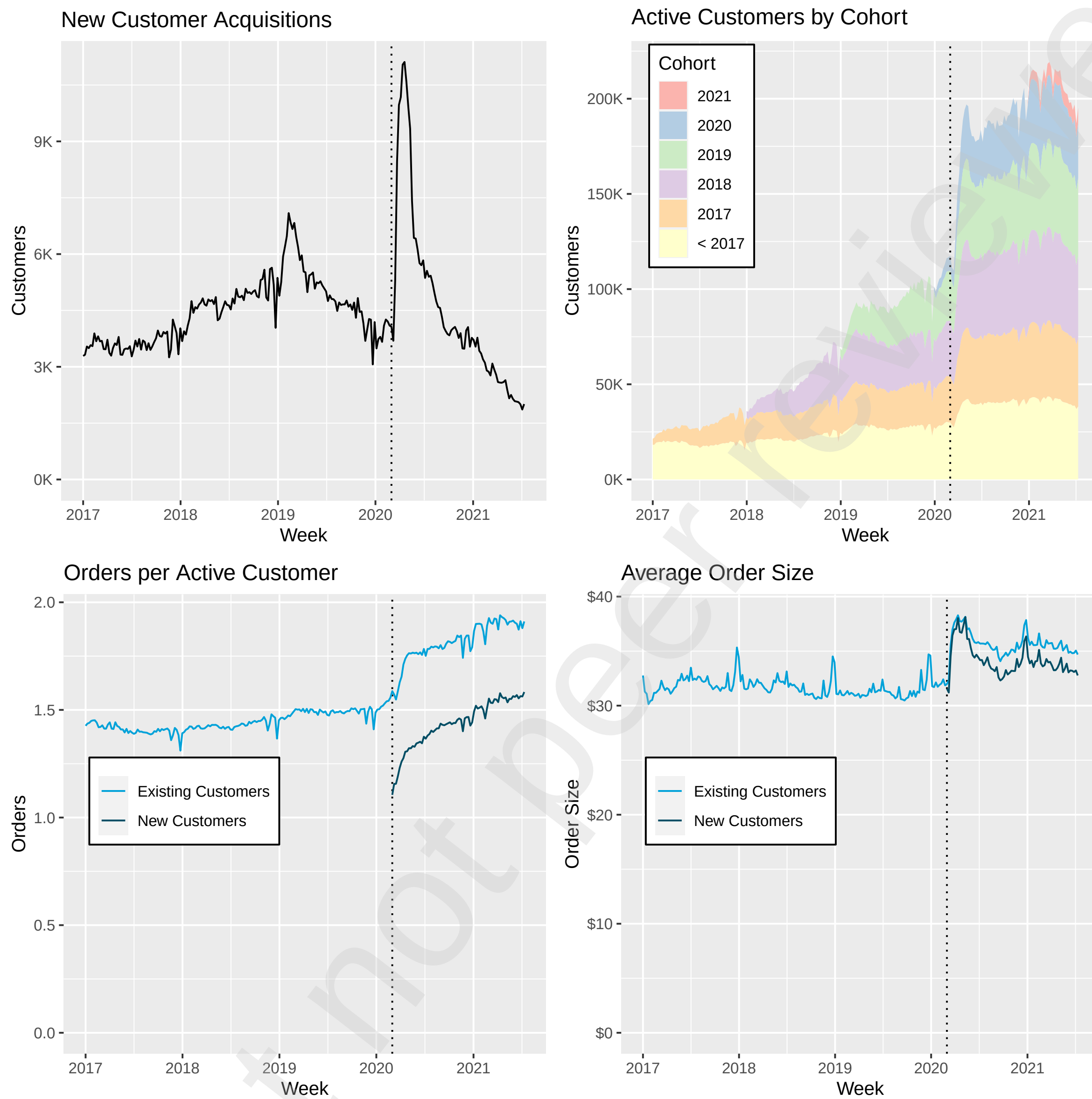
component customer behaviors. Figure 2 summarizes restaurant delivery category activity in terms of customer adoption, active customers, order frequency, and order size.

The weekly count of new customers ordering delivery for the first time (top-left) shows that, while many people who had never tried restaurant delivery before were spurred to do so early on in the pandemic, new customer adoption has subsequently plummeted to well below pre-pandemic levels.

Conversely, other shifts in delivery ordering behavior appear less transient. A far greater number of customers are active each week compared to pre-pandemic levels (top-right). Additionally, customers continue to place a greater number of orders each week they are active (bottom-left) and spend more when they buy (bottom-right).

Trends across cohorts reveal that most of the COVID-related increase in revenue came from customers acquired before the onset of the pandemic. Customers adopting delivery during the pandemic represented a small proportion of the lift in active customers (top-right), with lower order frequencies (bottom-left) and sizes (bottom-right) than existing

Figure 2 Summary of Customer Base Dynamics by Week



Note: Vertical dotted lines represent March 1st, 2020. In the top-right panel, “cohorts” are defined in terms of acquisition year (year of initial adoption of delivery services). In the bottom panels, “Existing Customers” are customers acquired prior to March 1st, 2021; “New Customers” are customers acquired thereafter.

customers, limiting the contribution of new customers to the overall spending growth observed in Figure 1.

Overall, the aggregate trends show a few clear patterns. Spending on restaurant delivery more than doubled in a matter of weeks early on in the pandemic, without a subsequent commensurate drop. Though the pandemic initially drew in significantly more new customers, longer-term gains in sales have been sustained by existing customers ordering more frequently and in larger amounts.

4. Quantifying the Effect of COVID-19 Over Time

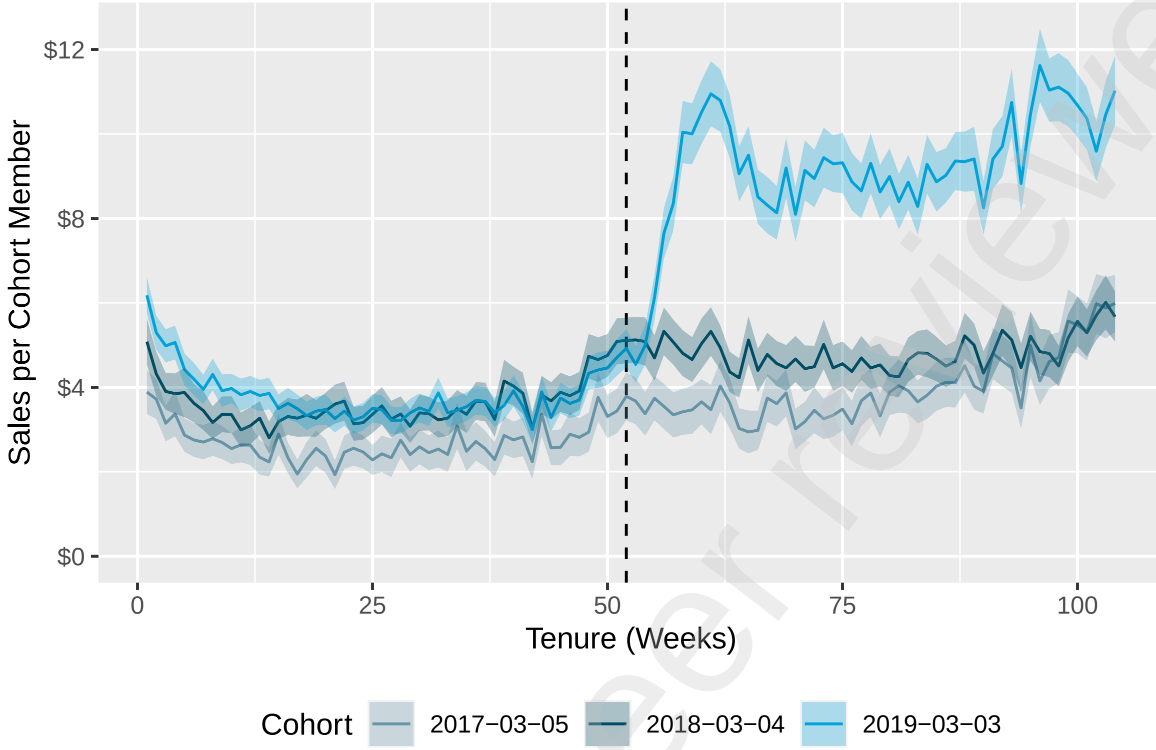
4.1. Methodological approach

While it is visually evident from Figure 1 that the pandemic had a large, immediate impact on delivery sales in the first few months of the pandemic, it is more difficult to extrapolate what the counterfactual no-pandemic baseline would have been over a longer horizon. Ideally, to quantify the longer-run effects of the pandemic, we would compare a group of "treated" consumers affected by the pandemic to a contemporaneous group of "control" consumers who were unaffected by the pandemic. However, due to the ubiquitous nature of the pandemic, no suitable contemporaneous control group exists. While COVID has affected regions differently (which we exploit in Section 5.1 to assess potential mechanisms driving the impact of COVID), even consumers in regions with low COVID incidence and no government-imposed restrictions still empirically exhibited behavior similar to regions with higher COVID incidence and stricter restrictions (e.g., increased stay-at-home rates and restaurant closures), leaving no truly untreated group for comparison.

In lieu of a contemporaneous control group, we consider comparisons across acquisition cohorts from different time periods to quantify COVID-driven lift for existing customers. Empirically, cohorts tend to exhibit highly regular behavior as a function of tenure (time since first order) after accounting for cross-cohort level shifts and seasonality (Schweidel et al. 2008). This regularity allows us to identify the effect of COVID by comparing cohorts affected versus unaffected by COVID at the same period of tenure. The intuition for this approach is illustrated in Figure 3, which shows the weekly sales per cohort member for three cohorts (defined on a weekly basis) acquired in early March of 2017, 2018, and 2019.

The impact of COVID is visually clear from this figure: the three cohorts follow very similar trajectories for their first year, but after a year, when COVID takes effect for the 2019 cohort but not the 2017 and 2018 cohorts, the trajectories diverge greatly, with sales more than doubling for the 2017 cohort relative to the other two. Intuitively, if we assume that the 2019 cohort would have followed a trend similar to the 2017 and 2018 cohorts in a counterfactual world where COVID had not occurred, then we can use these older cohorts as a "control group" against which to compare the 2019 cohort. This is analogous to assuming parallel trends in a typical difference-in-differences analysis, except that parallel

Figure 3 **Cohort Comparisons, Unadjusted for Calendar Time Trends**



Note: Cohorts are defined at the weekly level, with the cohort date representing the first day of the week of acquisition (e.g., the 2019-03-03 cohort consists of panel members who first ordered delivery between March 3rd and March 9th, 2019). Vertical dotted line represents 52 weeks after acquisition, which is March 1st, 2020 for the 2019 cohort. The intervals represent 95% pointwise confidence bands.

trends are assumed as a function of tenure instead of calendar time. This example also makes clear the lack of validity of lift estimates based upon naive extrapolation in situations such as this one, especially over longer time horizons, due to the highly nonlinear curvature of baseline spending.

In practice, such a parallel trends assumption does not hold exactly due to other calendar time effects besides COVID which lead to imperfect correspondence between cohorts. As seen in Figure 3, the 2017 and 2018 cohorts track closely, but their trends are not exactly parallel. Evidently, there are other non-COVID calendar time trends that make it inappropriate to compare cohorts directly. Accordingly, we extend this simple cohort comparison approach to flexibly account for baseline calendar time shifts in consumption propensities.

In particular, we employ an age-period-cohort (APC) model (Fosse and Winship 2019) to nonparametrically decompose cohort-time observations into cohort, tenure, and time effects. Letting Y_{ct} denote the per-capita sales from cohort c on week t, we model observations as a three-way fixed effects model:

$$\log(Y_{ct}) = \alpha_c + \gamma_{t-c} + \delta_t + \varepsilon_{ct} \tag{1}$$

where α_c captures static differences in behavior across cohorts, γ_{t-c} captures within-cohort trends as a function of tenure, and δ_t captures overall calendar-time trends. We discuss the appropriateness of the additivity assumption in our empirical setting in the next section. While we use sales as our running example of Y_{ct}, we also apply our model to three other behavioral variables of interest – proportion of cohort active (i.e., placing at least one order in that week), orders per active customer, and average order size – to decompose the behaviors driving the lift in overall sales.

To further isolate the impact of COVID, we residualize δ_t of cyclical calendar time effects such as seasonality and holidays. We refer to this residualized time series as ξ_t.

We can think of COVID as a regime change that disrupted the evolution of ξ_t:

$$\xi_t = \begin{cases} \xi_t^0 & \text{for } t < t^* \\ \xi_t^0 + \tau_t & \text{for } t \geq t^* \end{cases}$$

where the time t^* represents the onset of COVID. That is, (residualized) calendar time effects evolve according to some baseline stochastic process ξ_t^0; following the onset of COVID, ξ_t^0 represents the counterfactual baseline trend in ξ_t had the pandemic not occurred, while τ_t represents the deviation from this baseline attributable to COVID.

Naturally, τ_t cannot be directly observed; instead, we observe ξ_t, the sum of the baseline trend ξ_t^0 and the effect of COVID τ_t. Accordingly, we adopt an event study approach (Corrado 2011), wherein we fit a time series model to the *pre-pandemic* observations of ξ_t (i.e., $[\xi_1, \ldots, \xi_{t^*-1}]$, over which period $\xi_t = \xi_t^0$), then use this model to forecast the distribution of probable values of ξ_t^0 during the pandemic, $P(\xi_t^0 | \xi_1, \ldots, \xi_{t^*-1})$ for $t \geq t^*$, of what would have happened in the no-pandemic counterfactual baseline. The deviation of ξ_t from the conditional expectation of its baseline forms an estimate of the COVID effect:

$$\hat{\tau}_t = \xi_t - \mathbb{E}\left(\xi_t^0 | \xi_1, \ldots, \xi_{t^*-1}\right) \quad \text{for } t \geq t^*,$$

and the corresponding predictive interval around the forecast of ξ_t^0 allows us to make statistical statements about the effect size. This is in contrast to direct comparison of cohorts as in Figure 3, wherein a naive parallel trends assumption would impose that $\xi_t^0 = 0$.

In the APC modeling literature, each of the functions α_c, γ_{t-c}, δ_t (and thus ξ_t) are well-known to only be identified up to an additive linear term due to cohort, tenure, and time being linearly dependent (Fosse and Winship 2019). This form of degeneracy is not a problem in our context since our quantity of interest is the *nonlinear* departure of ξ_t from its pre-COVID trend (which is estimable): adding a linear term to the entire time series does not change the absolute gap of post-pandemic values of ξ_t from the pre-pandemic trend, which is our estimand of interest. Without loss of generality, we constrain ξ_t to have no first-order trend in the pre-pandemic period to achieve identification. See Web Appendix B for further detail on the identifiability of this model. We describe how we apply this model in our setting detail below.

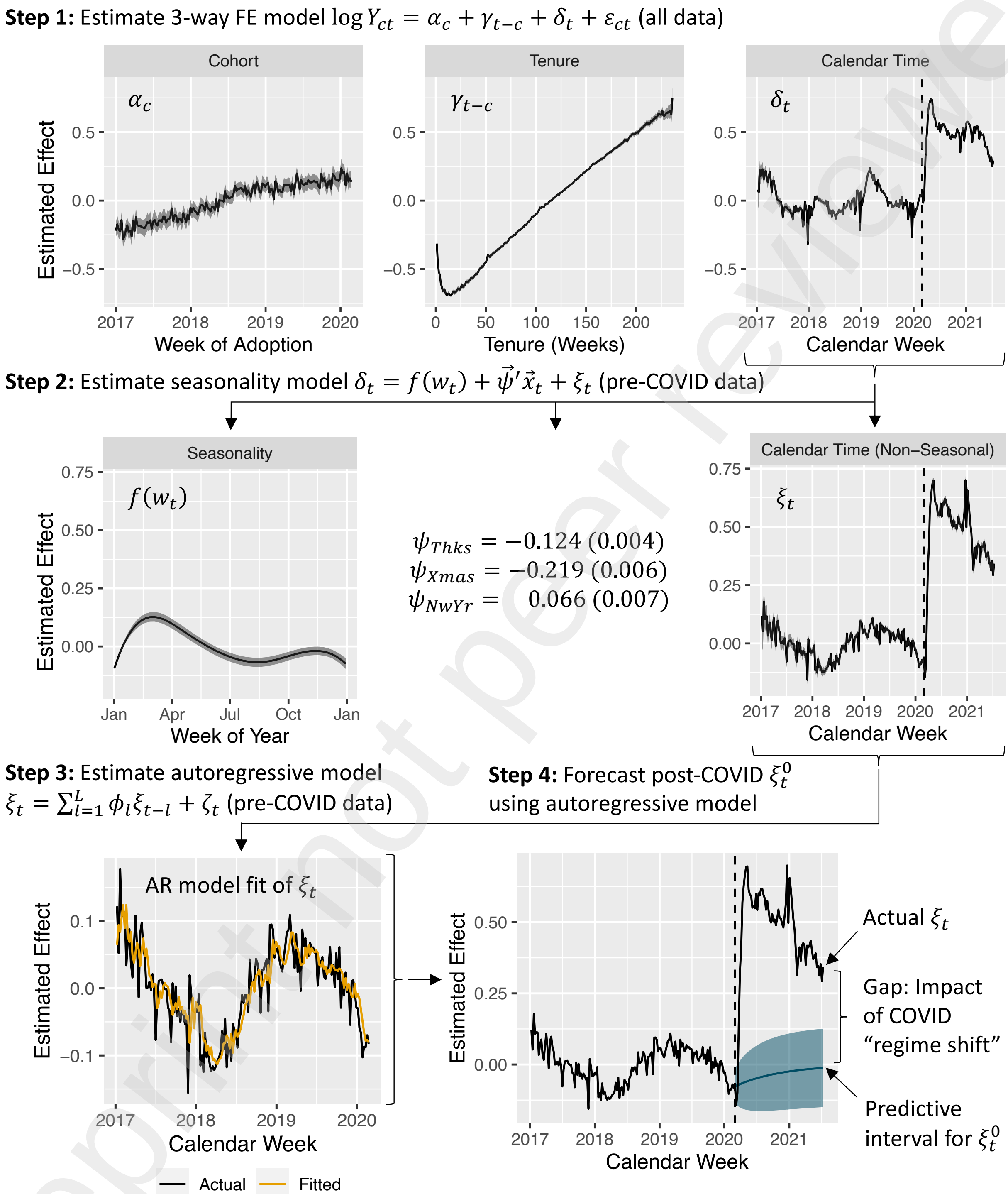
4.2. Model implementation and estimates

We illustrate how we bring this procedure to life in our empirical application in Figure 4 with per-capita sales as the dependent variable (analogous estimates for our other dependent variables are presented in Web Appendix B).

First, we estimate the three-way fixed effects model in Equation 1. We use data on all cohorts acquired prior to March 1st, 2020, with data up to the week of July 11th, 2021. This yields nonparametric estimates of cohort, tenure, and calendar time trends (Step 1 of Figure 4). Since δ_t shows clear seasonality and holiday effects, we residualize δ_t of seasonality and holidays to isolate non-seasonal effects, denoting the residualized effects by ξ_t (rightmost figure, Step 2).

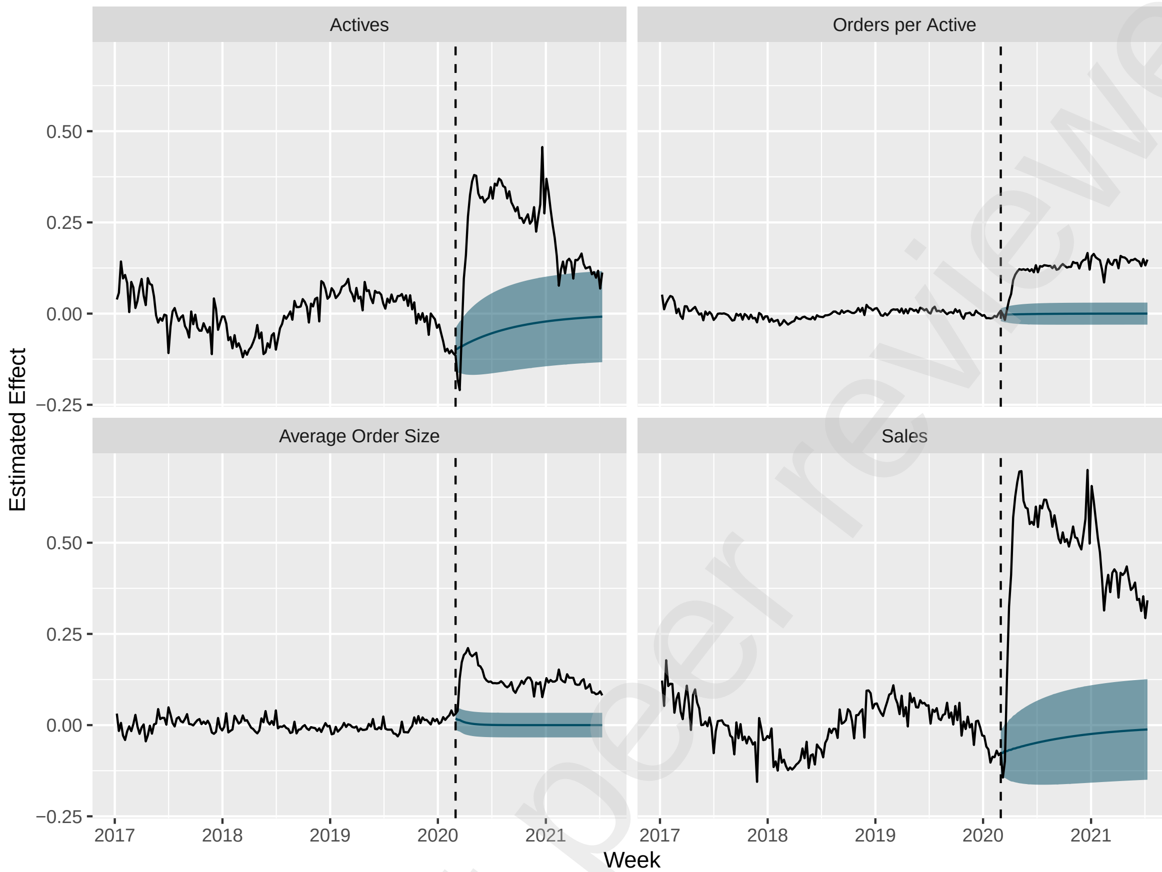
ξ_t was clearly non-constant in the pre-COVID period, highlighting the need for a model capturing the evolution of the baseline trend rather than simply assuming parallel trends. Accordingly, we fit a time series model to ξ_t over the pre-COVID period; substantial autocorrelation is visually evident in ξ_t, so we apply an autoregressive model to it, selecting an AR(4) model by AIC. The in-sample conditional mean of this AR model over the pre-COVID period is shown in Step 3. Lastly, we use this AR model to obtain the conditional expectation and predictive distribution of the counterfactual baseline over the post-COVID period, $P\left(\xi_t^0 | \xi_1, \ldots, \xi_{t^*-1}\right)$ (Step 4).

Figure 4 Estimation Procedure Illustrated for Cohort Sales



Note: Bands represents 95% pointwise intervals. For the calendar time component, the vertical line represents March 1st, 2020. w_t represents the week of year at week t. $\vec{x}_t$ includes dummy variables representing Thanksgiving week, Christmas week, and New Year's Eve/Day week(s). $f(w_t)$ is parameterized by cubic b-splines with 2 interior knots. The autoregressive model presented has AR order $L = 4$. We repeat this procedure with active customers, orders per active customer, and average order size as the Y_{ct} variable.

Figure 5 Predictive Intervals for Estimated Calendar Time Effects from AR(4) Model By Process



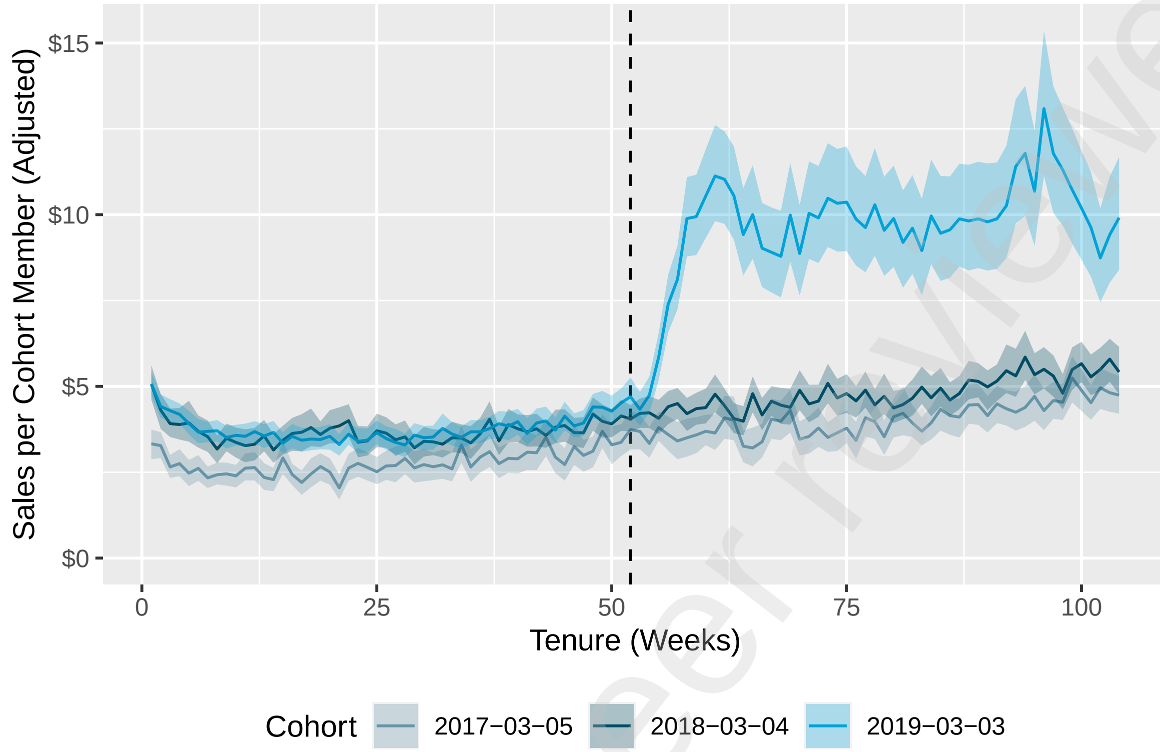
Note: blue bands represent 95% pointwise predictive intervals for each process. Vertical lines represent March 1st, 2020.

This four-step procedure gives us a predictive interval for the counterfactual baseline value of ξ_t^0 over the post-COVID period. Comparing this interval to ξ_t allows us to make statistical inferences about how much spending has deviated from baseline during the pandemic. We repeat this process for each of our four dependent variables.

ξ_t, along with point estimates and predictive intervals for ξ_t^0 over the post-COVID period, are shown in Figure 5 for each of the four dependent variables. The predictive intervals appear sensible given the pre-COVID evolution of ξ_t. Naturally, the intervals become wider further into the pandemic due to the increasing uncertainty in the evolution of the baseline trend in ξ_t^0 for larger values of t. We provide full estimates for the fixed effects, seasonality, and time series models, along with further details about model computation and time series model selection, in Web Appendix B.

The in-sample R^2 of the three-way fixed effects model of cohort sales is 98.8%, with the other three dependent variables all having an R^2 in excess of 88%, indicating that

Figure 6 **Cohort Comparisons, Adjusted for Calendar Time Trends**



Note: Vertical dotted line represents 52 weeks after acquisition, which is March 1st, 2020 for the 2019 cohort. The intervals represent 95% pointwise confidence bands. Logged sales figures are adjusted by subtracting non-COVID calendar time effects from each observation (seasonality effects $f(w_t)$, holidays $\vec{\psi}'\vec{x}_t$, and non-seasonal baseline trends ξ_t^0). Confidence bands are larger for the 2019 cohort after 52 weeks due to the uncertainty in the evolution of ξ_t^0 in the post-COVID period (where ξ_t^0 is forecasted from the AR model).

this specification captures cross-cohort and cross-time variation in behavior exceptionally well. Additionally, the AR(4) model effectively captures the serial correlation structure in ξ_t. While ξ_t violates standard tests for time series non-stationarity, the AR model's residuals have no statistically significant residual autocorrelation structure for each of the four dependent variables, as measured by the Augmented Dickey-Fuller and Ljung-Box tests (for details, see Web Appendix B). Autocorrelation explains 72%, 74%, 60%, and 25% of the variation in ξ_t for sales, active customers, orders per active customer, and average order value, respectively.

The validity of using the additive APC model to compare cohorts is demonstrated by Figure 6, which shows the same three cohorts as in Figure 3 after adjusting for baseline calendar time trends. After adjustment, the 2019 cohort is nearly perfectly parallel with the older cohorts in the pre-COVID period, making the impact of COVID over time visually clear. The two older cohorts are similarly parallel throughout the entire series. These parallel trends provide reassurance that the cohort comparison strategy yields valid estimates of the impact of COVID, with additive baseline calendar time trends being sufficient to correct for parallel trends violations across cohorts.

4.3. Results

We can calculate lift figures as a simple transformation of the gap between estimated and predicted counterfactual baseline curves seen in Figure 5. Since our model is in logarithmic terms, the percentage lift in sales is estimated as:

$$\widehat{\text{Lift}}(t) = \exp(\hat{\tau}_t) - 1$$

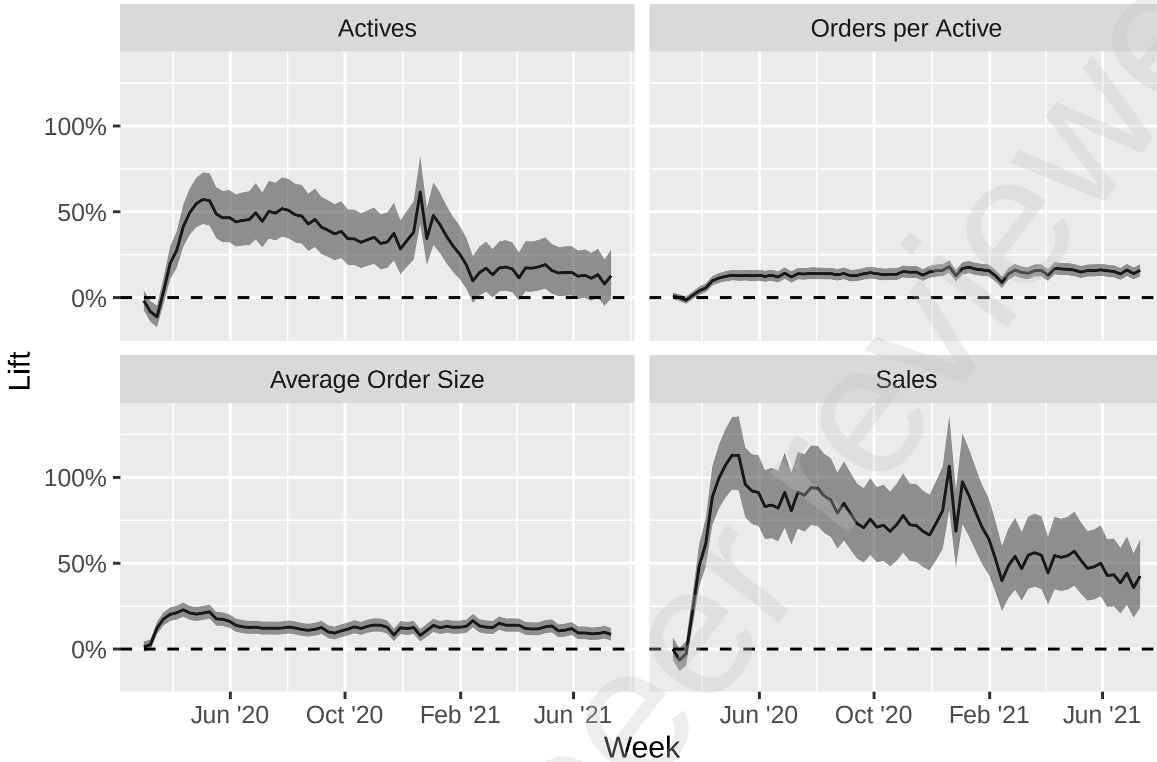
The estimated lifts and corresponding confidence bands are given in Figure 7. Overall, we see that customer spending more than doubled relative to baseline in the early weeks of the pandemic, peaking at a 113% lift (SE: 11%) at the beginning of May (bottom-right of Figure 7). This early peak was followed by a small dip in the Summer and Fall of 2020, then a second wave in Winter 2020 corresponding to a major resurgence in COVID cases. Since Spring 2021, the lift in sales has dropped, reaching a smaller but still substantial 42% (SE: 10%) in July. Turning to the other variables, we find that the Spring 2021 drop in spending lift appears attributable mainly to a large drop in excess active customers (top-left); however, the remaining active customers are still ordering at elevated rates (top-right). Meanwhile, order sizes increased by 22.8% (SE: 2.1%) over baseline initially but by July 2021 had dropped to only 8.5% (SE: 1.9%) over baseline (bottom-left).

Taken together, these figures suggest that the pandemic spurred a broad segment of the existing customer base to begin more actively ordering, at greater frequencies, and in larger sizes; after February 2021, a significant fraction of this segment dropped off, and the remaining customers also appear to be slowly dropping off; still, those who remain continue to order at elevated frequencies and (to a lesser extent) sizes.

5. Persistence of Behavioral Changes

As alluded to in the previous section, a year and four months into the pandemic, a sizeable segment of pre-existing customers continue to order delivery who would not have in absence of a pandemic. This raises an important question: is this sustained boost in customer

Figure 7 Estimated Sales Lift from AR(4) Model



Note: bands represents 95% pointwise confidence intervals.

activity due to continuing pandemic-related disruptions, or do they represent a persistent change in customer behavior, e.g., due to habit formation?

We hypothesize the following: the COVID-19 pandemic disrupted consumers' lifestyles (e.g., removing the option of dining at restaurants), which shifted their consumption patterns towards restaurant delivery services; over the course of the pandemic, many consumers formed habits around ordering restaurant delivery, continuing to order delivery at elevated rates even once the initial shocks that shifted their consumption patterns subsided.

To substantiate this hypothesis, it is necessary to demonstrate that the initial shocks that pushed consumers towards restaurant delivery early on in the pandemic are no longer driving the sustained lift in spending: if the shocks themselves are still ongoing, then Figure 7 could simply reflect the continuing effects of pandemic-related disruptions, rather than a meaningful shift in consumer habits/preferences. We support this hypothesis by verifying three empirical implications of it below.

5.1. Evaluating continued existence of shocks to restaurant delivery

We evaluate whether the shocks driving customer lifestyle shifts for existing customers are still present in three ways – by analyzing the behavior of prospective and new customers, and by analyzing the underlying mechanisms driving the shocks.

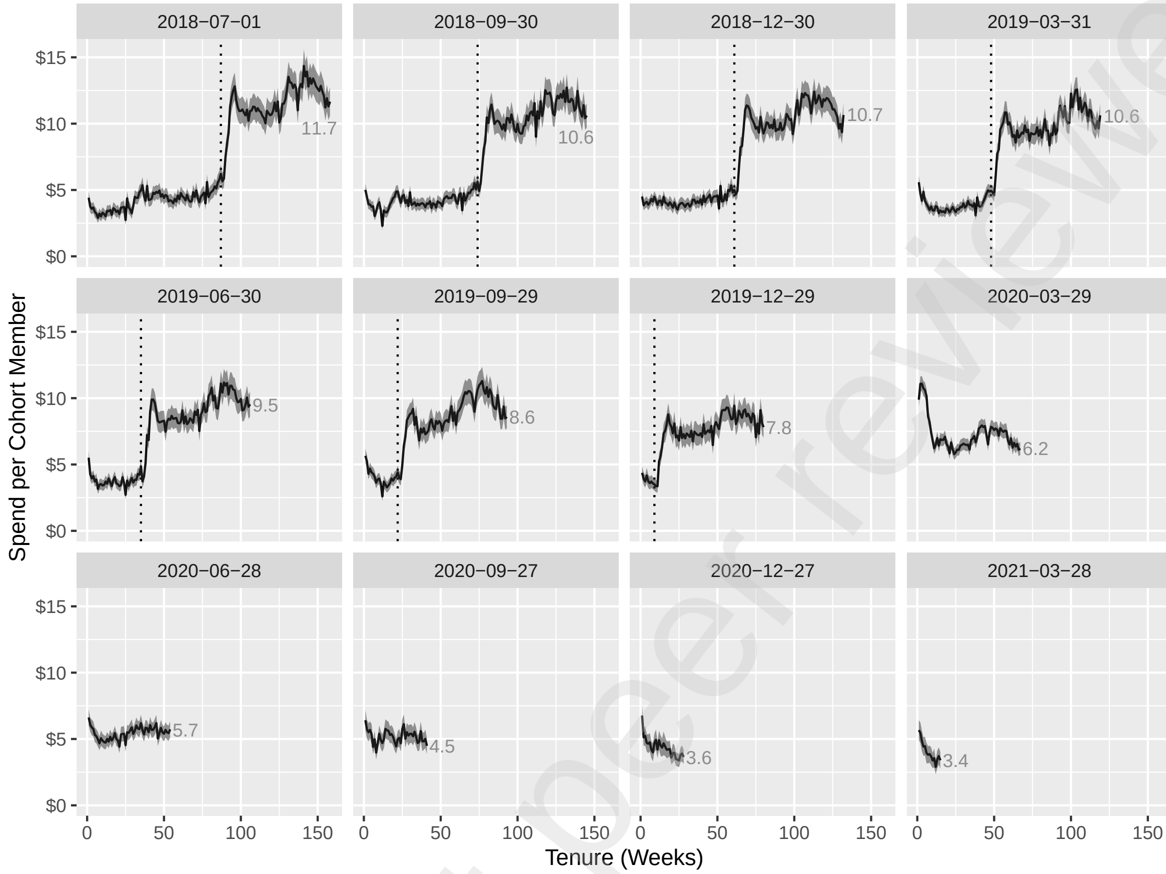
Behavior of prospective customers. First, if the shocks driving pre-existing customers to order delivery are still present, we would expect them to affect prospective customers as well. The top-left panel in Figure 2 shows that new customer adoption is down to well below where it was prior to the pandemic; thus, it appears that prospective customers are no longer being pushed to enter the market.

Behavior of new customers. Second, there is evidence when we compare new customers to pre-existing customers: since habits are formed over time through repeated consumption experiences, if COVID-related disruptions have dissipated, recently acquired customers should not exhibit elevated ordering activity, since they would not have had the time to develop the habits driven by these disruptions. Figure 8 shows the trajectory of weekly spending for a selection of weekly cohorts. We see that pre-COVID cohort members typically spent \$3-5 per week prior to COVID, followed by a sharp spike to \$8-12 per week. The early COVID cohorts have settled into spending \$5-8 per week, while the more recent cohorts have been spending less and less, with the later cohorts' spending comparable to older cohorts prior to COVID, contrary to a pre-pandemic upward trend across cohorts. See Web Appendix C for further discussion of these cross-cohort trends.

In other words, pre-existing customers, who had the full duration of the pandemic to get in the habit of ordering delivery, still show a large and persistent lift in spending; customers who started ordering delivery in the first months of the pandemic exhibit smaller but still meaningful lifts, since they experienced COVID-related disruptions to a lesser and briefer extent. Lastly, very recent cohorts behave as if the pandemic never happened, as they did not experience enough disruption to form habits around them.

Mechanisms driving behavioral changes. Third, to the extent that we can observe COVID-related disruptions directly, we can explicitly measure the extent to which they have dissipated. To this end, we estimate a separate model to infer the potential mechanisms driving the behavioral changes in the delivery category and assess whether these mechanisms have returned to pre-pandemic levels.

Figure 8 Weekly Cohort Spending by Tenure



Note: Vertical dotted lines represent March 1st, 2020. Facet labels indicate the start date of the cohort's acquisition week (e.g., the 2018-12-30 label indicates that the plot corresponds to the cohort acquired during the week of December 30th, 2018). Gray text indicates the value of the time series as of the final week of the data (July 11th, 2021). These cohorts are chosen to be spaced quarterly; results for other cohorts are substantively the same as for the cohorts shown here.

In this supplementary analysis, we explain total CBSA-level sales per capita as a function of four such mechanisms: income shocks, stay-at-home behavior, restaurant supply, and restaurant dine-in activity (as measured through the local unemployment rate, proportion of the population staying completely at home, total restaurant employee shifts, and total restaurant dine-in visits, respectively), estimating a two-way fixed effects (FE) log-log regression model using observations at the CBSA-day level.

This two-way FE model identifies the coefficients of interest using within-CBSA, within-day variation. For instance, while unemployment increased significantly for most CBSAs in April 2020, 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 intuition is presented visually in Figure 9, which shows the national average trends in our regressors with CBSA-specific trends for a subset of CBSAs overlaid (the specific CBSAs are chosen to highlight spatiotemporal variation in our regressors). While most CBSAs correlate strongly with the national trends, they also exhibit idiosyncratic and sometimes even countercyclical deviations from the overall trend; it is these deviations which drive our identification. For brevity, details of this analysis, including a detailed discussion of endogeneity issues and other robustness checks, are included in Web Appendix D.

The main result from this regression is that all four regressors have statistically and economically significant impacts on sales, but three of them – higher unemployment, increased stay-at-home rates, and decreased restaurant supply – all had negative effects upon restaurant delivery sales during the pandemic. The only variable that had a positive effect is restaurant dine-in – as individuals physically dined in less, they ordered delivery more.

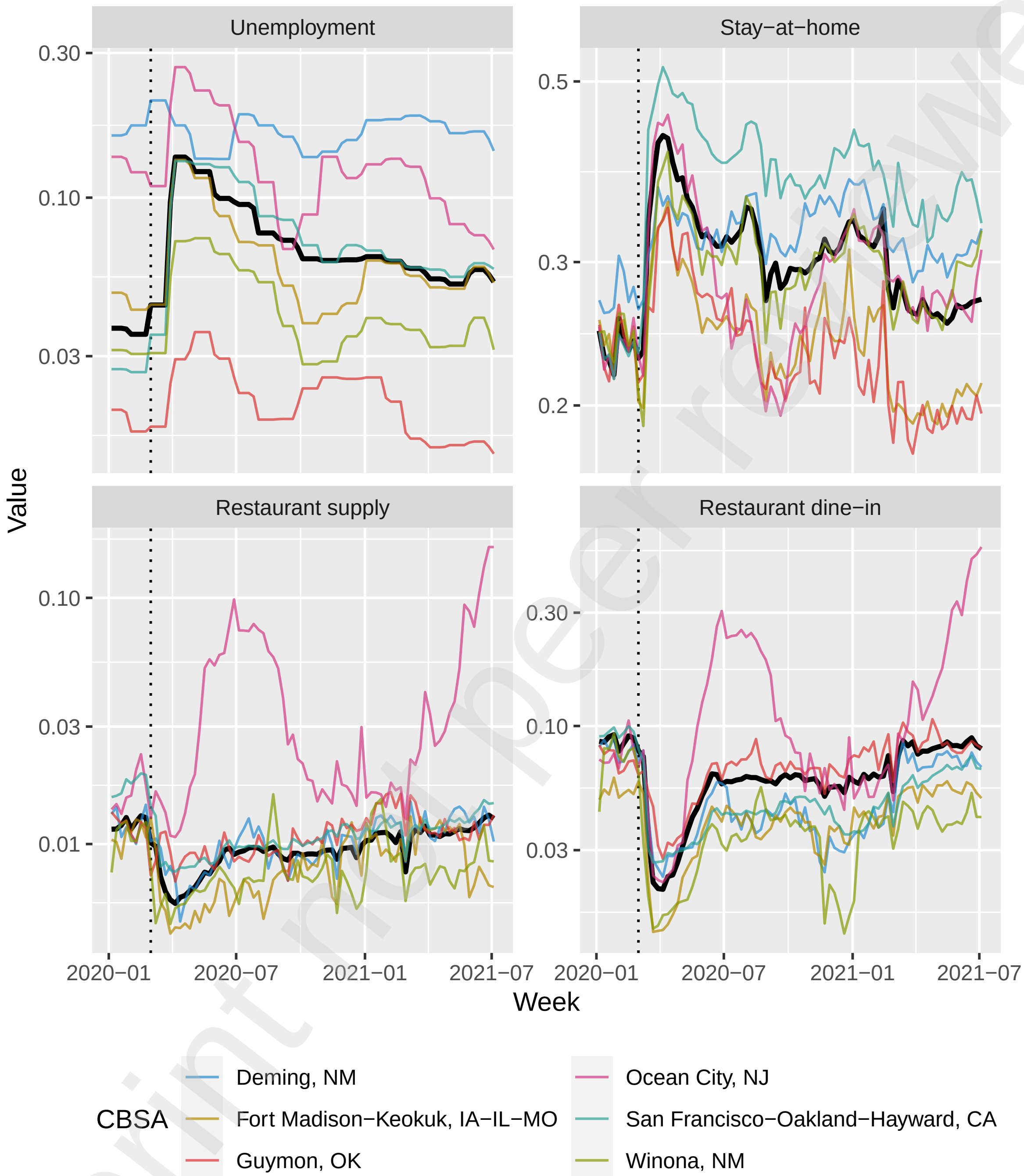
Notably, Figure 9 shows that, as of early Spring 2021, national dine-in volume had recovered to pre-pandemic levels (bottom-right panel, thick black line), coinciding with the drop in active customer and spending lifts shown in Figure 7. Thus, the shift away from dine-in that drove customers to delivery appears to be largely over.

5.2. Longer-term persistence of habits

In sum, we find that since February 2021, new customer adoption has fallen to below pre-pandemic levels, newly acquired customers are spending as if the pandemic had never occurred, and restaurant dine-in levels are largely back to normal. These findings all point towards the initial COVID-related disruptions that drove customers towards ordering delivery having largely dissipated.

Despite this, Figure 7 shows a persistent sales lift for pre-existing customers that, as of July 2021, is about one-third of its peak effect. We take these results to mean that a segment of consumers has formed habits around food delivery, persistently ordering in greater frequencies and sizes despite no longer having external forces driving them to do so. The drop in active customers in Spring 2021 may indicate that, as the US entered a "new normal" with vaccines and reopening, many consumers reverted to their pre-pandemic lifestyles, leaving behind a smaller but highly habituated segment.

Figure 9 **Time Series of Regressors (National Average and Select CBSAs)**



Note: Vertical dotted lines represent March 1st, 2020. The thick black line represents the national average, while the thin colored lines represent specific CBSAs.

Nonetheless, habits are not permanent: though the lifestyle shocks of the pandemic may have led many consumers to make habits of ordering delivery, subsequent life events are likely to spur further lifestyle changes, eventually supplanting pandemic-driven habits. This is evidenced by the gradual decline in active customer and sales lifts discussed in Section 4.3; the fact that the drop in sales lift is driven by a drop in active customer lift seems to suggest that, rather than a uniform drop in sales lift across individuals, customers who formed habits during the pandemic are gradually "dropping out" as their habits are shifted away by other events. Though it is difficult to extrapolate how fast habits fade based on this limited time series, it seems unlikely that the sales lift will meaningfully persist over a horizon much beyond the end of our data, barring further lifestyle shocks due to resurgences of new COVID variants.

6. Discussion

The COVID-19 pandemic has generated significant growth for the restaurant delivery category, with sales more than doubling in a matter of weeks early on in the pandemic. This growth was primarily driven by a large increase in the number of active pre-existing customers, who also increased the frequency and size of their orders.

Despite a steep drop in the lift in active customers since February 2021, sales remained roughly 40% above baseline in July 2021 (about one-third of peak lift early in the pandemic), even though the shocks that drove consumers towards delivery had subsided as of early 2021, suggesting that these behavioral changes are persistent. We take these findings as evidence that a segment of consumers has formed habits around food delivery, continuing to order at elevated rates despite no external force driving them to do so.

Still, we infer that this segment is shrinking over time, to the point that the lift in active customers is only marginally significant by July 2021. Consequently, barring further lifestyle shocks from new COVID variants, it seems unlikely that this segment will remain sizeable for many months beyond the end of our data.

Our findings are consistent with behavioral economics literature on the persistence of behavioral interventions: much as prior experiments have found that several weeks of incentive-based interventions can lead some consumers to form habits that persist weeks or months after intervention, we find that several months of external disruption to one's lifestyle—such as having limited opportunity to dine in at restaurants—has led a segment of consumers to form habits around ordering food delivery, continuing to order delivery at elevated rates for several months despite the external disruption abating. We further find evidence of gradual attrition as the lifestyle shocks of COVID are supplanted by other life events, suggesting that the habit-forming effects of COVID on consumer lifestyle habits decay over time and are unlikely to endure over a long time horizon. We now address some limitations of our analyses and interpretations, and discuss related questions for future work.

First, our lift analysis cannot separate out effects of COVID-19 from other major regime shifts that disrupt the baseline evolution of calendar effects. As such, while it can be used to infer longer-term effects of regime shifts, its estimates should be treated with caution over long time horizons if there is reason to believe another major regime shift has occurred that could significantly shift customer behavior and enter the results.

We also readily acknowledge that we have not analyzed how the pandemic has impacted competition between restaurant delivery platforms. Our target of interest in this research is the first-order effect of the pandemic on overall delivery spending, making competitive analysis and related issues out of scope. Relatedly, our finding that restaurant dine-in is back to normal while restaurant delivery demand remains above normal motivates the question of what food-related category, if any, is below normal (e.g., grocery). However, such an inter-category competition analysis is similarly out of scope. We leave these interesting questions for future work.

Lastly, we note that our study is limited to a single product category, and so our analysis does not yield empirical generalizations. That said, consistent with extensive prior research documenting persistence in consumer behavior across many domains, we expect to see qualitatively similar patterns across categories; however, the magnitude and durability of persistence is likely to differ by how conducive the category is to habit formation. For instance, whereas consumers must eat and often exercise at regular intervals, making it easier to form routines and habits around food delivery and in-home fitness products, discretionary online shopping is typically done sporadically and thus less conducive to habit formation, suggesting online retail may see less durable effects of the pandemic. Conversely, subscription-based categories such as streaming and in-home fitness services may enjoy more durable gains since, unlike food delivery where consumers must actively place orders, consumers may passively fail to cancel subscriptions that they adopted during the pandemic. Further analysis of the heterogeneity in persistence across categories is an interesting direction for future work.

Towards this end, the methodological approach we have employed is transferable to these and other marketing and economics problems. While we focus on the impact of COVID-19 upon the food delivery category, our approach could easily be used to measure the extent to which other consumption categories boosted by the pandemic—e.g., online retail or home entertainment—have also seen persistent changes in consumer behavior. Beyond COVID, many papers within the empirical CRM literature streams (e.g., Gopalakrishnan et al. 2017, McCarthy and Oblander 2021, Schweidel et al. 2008) model cross-cohort effects jointly with calendar time and tenure effects. The APC approach could provide useful insight in such situations, with far fewer parametric assumptions than are typically made in these papers. More generally, our combined APC model/event study approach could be used to study the impact of other events for which researchers do not have a control group but do have customer base data on multiple cohorts.

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 research grants from the Goizueta Business School of Emory University and the Marketing Science Institute. 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 for the guidance of Eric Bradlow, as well as helpful comments from Don Lehmann, Oded Netzer, Asim Ansari, Peter Fader, Eric Bradlow, Davide Proserpio, Lan Luo (Columbia) and participants of the Columbia Business School, UC Davis, and Dartmouth Marketing seminar series.

References

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

Baker SR, Farrokhnia RA, Meyer S, Pagel M, Yannelis C (2021) Income, liquidity, and the consumption response to the 2020 economic stimulus payments. *Forthcoming at the Review of Finance* .

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

Charness G, Gneezy U (2009) Incentives to exercise. *Econometrica* 77(3):909–931.

Chen H, Qian W, Wen Q (2021) The impact of the COVID-19 pandemic on consumption: Learning from high-frequency transaction data. *AEA Papers and Proceedings*, volume 111, 307–11.

Chetty R, Friedman JN, Hendren N, Stepner M, the Opportunity Insights Team (2020) The economic impacts of COVID-19: Evidence from a new public database built using private sector data. *NBER Working Paper Number 27431* .

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

Dubé JP, Hitsch GJ, Rossi PE (2010) State dependence and alternative explanations for consumer inertia. *The RAND Journal of Economics* 41(3):417–445.

Dunn A, Hood K, Batch A, Driessen A (2021) Measuring consumer spending using card transaction data: Lessons from the COVID-19 pandemic. *AEA Papers and Proceedings*, volume 111, 321–25.

Fosse E, Winship C (2019) Analyzing age-period-cohort data: A review and critique. *Annual Review of Sociology* 45:467–492.

Fukuda K (2010) A cohort analysis of household vehicle expenditure in the US and Japan: A possibility of generational marketing. *Marketing Letters* 21(1):53–64.

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/`.

Gopalakrishnan A, Bradlow ET, Fader PS (2017) A cross-cohort changepoint model for customer-base analysis. *Marketing Science* 36(2):195–213.

Hastie TJ (1992) *Statistical models in S*, chapter 7, 249–307 (Wadsworth and Brooks).

Loewenstein G, Price J, Volpp K (2016) Habit formation in children: Evidence from incentives for healthy eating. *Journal of Health Economics* 45:47–54.

McCarthy DM, Oblander ES (2021) Scalable data fusion with selection correction: An application to customer base analysis. *Marketing Science* 40(3):459–480.

Rentz JO, Reynolds FD (1991) Forecasting the effects of an aging population on product consumption: An age-period-cohort framework. *Journal of Marketing Research* 28(3):355–360.

Schweidel DA, Fader PS, Bradlow ET (2008) Understanding service retention within and across cohorts using limited information. *Journal of Marketing* 72(1):82–94.

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* .

Stammann A (2017) Fast and feasible estimation of generalized linear models with high-dimensional k-way fixed effects. *arXiv preprint arXiv:1707.01815* .

Valinsky J (2020) Peloton sales surge 172% as pandemic bolsters home fitness industry. URL `https://www.cnn.com/2020/09/11/business/peloton-stock-earnings/index.html`.

Yang N, Long Lim Y (2018) Temporary incentives change daily routines: Evidence from a field experiment on Singapore's subways. *Management Science* 64(7):3365–3379.

Appendix A: 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 sets used to perform our analyses. We organize this section by data source. The focal analysis in the main text (APC and event study analysis) is based on the Earnest data, described in Web Appendix A.1, while the other data sources are only used in our mechanism analysis mentioned briefly in Section 5.1 of the main text and covered in full detail in Web Appendix D.

A.1. Earnest Research data

Earnest Research, one of the largest credit and debit card panel data companies, provided us with credit/debit card transaction data. The data includes 926 CBSAs and spans 5.5 years, from January 1st, 2016 to July 17th, 2021. Our data includes 1,629,818 panel members with consistent shopping behavior over that time period who made at least one observed purchase.[1] 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 2016 to address left-censoring in the dataset, such that in practice we use only the 4.5 years of data from 2017 to 2021 for analysis.

A.1.1. 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 43,329 panel members (2.66%) from the dataset, leaving 1,586,489 panel members with complete monthly location data.

[1] 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.

A.1.2. 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. This resulted in 55,047,588 observed transactions among the 1,586,489 retained panel members.

We omitted the platform EzCater, since it is a B2B business that caters to corporate clients (EzCater accounts for only 14,151 transactions, or 0.026% of total transactions). 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 1,820,071 transactions (3.31%), leaving 53,227,517 transactions.

The full list of delivery platforms considered, and their respective market shares after the above filtering, is given in Table 1.

A.1.3. Analysis-specific filtering For the APC 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).[2] 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 10 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 APC model.

Additionally, the left-censored nature of the data poses problem for APC 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 APC model is based on segmentation of cohorts by the week of first transaction.

Accordingly, to address left-censoring, we use 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, 2016 and December 31, 2016 from our APC model, assuming that first transactions from January 1, 2017 onwards are genuine acquisitions. Since very long interpurchase times are rare (in our data, 99.1% of interpurchase times are 1 year or less) and the customer base was relatively small prior to 2016 (33,125 panel members placed at least one delivery order in January 2016, as opposed to 492,441 in June 2021), the probability that an initial order in 2017 or later is actually a repeat order (with the true initial order taking place prior to 2016), is very small. Thus, in our APC model, we only include data for cohorts acquired since January 1, 2017.

For our analysis of mechanisms, left-censoring is not a concern, since we only use data from 2020 and 2021. 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

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

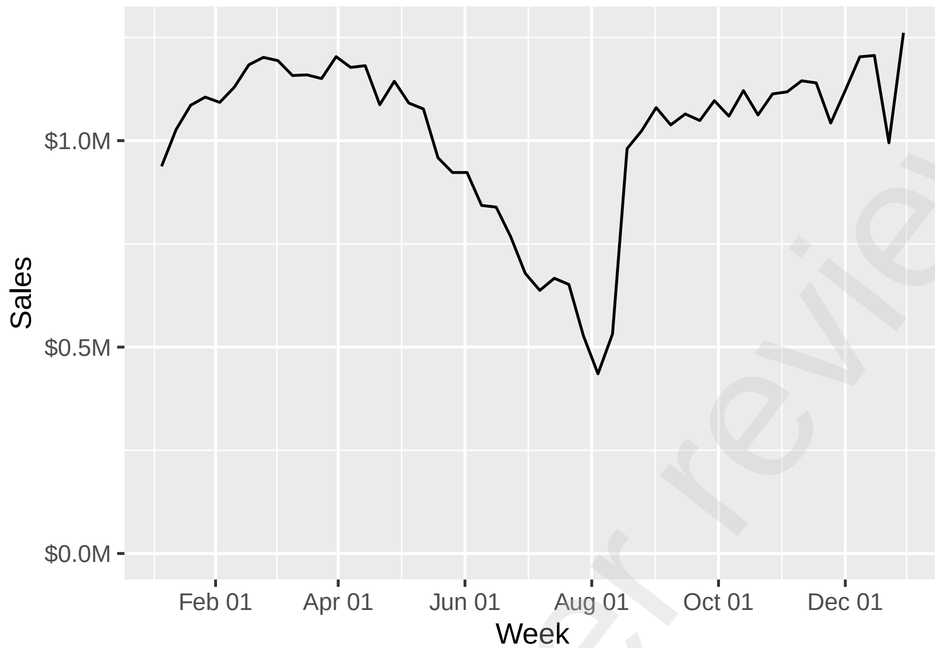
Table 1 Market Shares by Platform

Platform	Market Share	Notes
DoorDash	43.126%	
Uber Eats	20.112%	
GrubHub	16.016%	
Olo	7.736%	
Postmates	5.207%	Acquired by Uber Eats in 2020
Waitr	1.919%	
Slice	1.065%	
Eat24	0.795%	Acquired by GrubHub in 2017
Bite Squad	0.698%	Acquired by Waitr in 2018
Square	0.683%	
EatStreet	0.491%	
OrderUp	0.244%	Acquired by GrubHub in 2017-2018
Goldbelly	0.236%	
Amazon Restaurants	0.231%	Discontinued in 2019
Caviar	0.210%	Acquired by DoorDash in 2019
Delivery.com	0.195%	
Foodsby	0.166%	Acquired by GrubHub in 2017
Fooda	0.140%	
Tapingo	0.129%	Acquired by GrubHub in 2018
Ritual	0.116%	
Delivery Dudes	0.082%	Acquired by Waitr in 2021
Foodler	0.080%	Acquired by GrubHub in 2017
Just Eat	0.075%	
Thistle	0.072%	
Food Dudes Delivery	0.062%	
Munchery	0.061%	Discontinued in 2019
MealPal	0.046%	
Chowbus	0.005%	
Hungry Panda	0.003%	
Fantuan	0.001%	

Note: Market shares are in terms of sales over the entire 5.5-year period of the panel data.

perform CBSA-level aggregations for the panel regression. Since we include data from January 2020 to July 2021, we only include CBSAs where, for each month in this range, at least one delivery platform (among those covered by the platform restaurant listing data described in Web Appendix A.4) was operating. This results in daily observations for 780 CBSAs.

Figure 10 **Uber Eats weekly panel sales (2019)**



A.2. SafeGraph data

SafeGraph[3] provides aggregated mobile location data at the daily level based on a panel of over 18 million devices, which we use for the mechanism analysis in Web Appendix D. 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 7.5 million locations of interest 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 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 963,000 locations, hereafter referred to as "restaurants." The data provides the number of visits to each restaurant (defined as a device being at

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

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 Table 4, 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.

SafeGraph discontinued updating the Social Distancing Metrics dataset on April 17, 2021 as social distancing guidelines were temporarily starting to be relaxed, so the aforementioned stay-at-home rate data is observed for each CBSA from January 1, 2019 through April 16, 2021. However, the Weekly Patterns data continued to get updated through July 4, 2021, and as noted by SafeGraph, can provide similar insights to the Social Distancing Metrics dataset.[4] We use the Weekly Patterns data to impute CBSA-specific stay-at-home patterns from April 17, 2021 through July 4, 2021. In particular, the mechanism analysis we perform in Web Appendix D uses the log-transformed stay-at-home rate as a regressor and includes CBSA and day fixed effects, so for consistency, we impute the log-transformed stay-at-home rate residual of these fixed effects as well.

Our imputation procedure consists of the following four steps:

1. First, we obtain log-transformed CBSA-specific visit rates (i.e., the natural logarithm of the total number of visits, normalized by the total number of devices within that CBSA), for all 83 3-digit NAICS codes in the Weekly Patterns data from January 1, 2019 to July 4, 2021.[5]

2. Second, we partial out CBSA and day fixed effects by exploiting the Frisch-Waugh-Lovell theorem, residualizing both observed (log-transformed) stay-at-home rates and the (log-transformed) visit rates for each NAICS code before regressing the former against the latter.

3. Third, we run separate CBSA-specific ridge regressions for each of the 929 CBSAs in our dataset over the period of time over which both datasets overlap (i.e., from January 1, 2019 to April 16, 2021), modeling (log-transformed and residualized) stay-at-home rates within each CBSA as a function of contemporaneous (log-transformed and residualized) visit rates at the 83 3-digit NAICS codes.[6]

[4] https://colab.research.google.com/drive/1ETZo3KBhcwUikLRos5YmMZkf4aPweG44#scrollTo=W_rYMrvJBLCq

[5] There were a few likely data errors within particular CBSA-NAICS codes. For example, in CBSA 40140, NAICS code 481 registered zero visits on all but one day (in which there was one visit) over the period from January 1, 2019 to April 27, 2021, after which it abruptly and consistently registered more than 100 visits per day. To avoid contaminating our results with these apparent outliers, we clipped outlier CBSA-NAICS code visit counts over the period from April 17, 2021 to July 4, 2021 to be within the historically observed range (i.e., its range over the period from January 1, 2019 to April 16, 2021).

[6] We perform ridge regressions instead of simple regressions because the ratio of sample size to parameters is approximately 10-to-1, motivating regularization. We had alternatively run one homogeneous model, explaining variation in stay-at-home rates as a function of visit rates for each NAICS code across all CBSAs, but the partial R^2 of this model on a random 5% holdout sample of CBSA-days was 45.0% (i.e., the visit rate data explains 45% of the variation in log-transformed stay-at-home rates that was not already explained by CBSA and day fixed effects). The CBSA-specific ridge regressions perform substantially better over the same random holdout sample, as described below.

4. Fourth, we predict (log-transformed and residualized) stay-at-home rates using the fitted model from step 3 on the (log-transformed and residualized) visit rate data for each NAICS code over the period from April 17, 2021 to July 4, 2021, for which we do not observe stay-at-home data.

To assess the validity of this four-step procedure, we perform a holdout validation analysis. We hold out a random 5% of CBSA-days over the period from January 1, 2019 to April 16, 2021. In these held out CBSA-days, we assume that visit rate data by NAICS code is observed, but stay-at-home data is not. We then perform the above four-step procedure upon the remaining 95% of CBSA-days, which we use in conjunction with the visit rate data from the 5% holdout sample to impute (log-transformed and residualized) stay-at-home rates over the holdout sample. The goodness of fit of our predictions is high – for example, the R^2 of a no-intercept regression of the held out data against our predictions is 72.8% – giving us confidence that the imputed values are a reasonable proxy for what the stay-at-home measure was likely to have been over this period of time.

The final data object which we include in the fixed effects regression in Web Appendix D is the concatenation of the observed (log-transformed and residualized) stay-at-home rate data from January 1, 2019 to April 16, 2021 with the imputed (log-transformed and residualized) stay-at-home rate data from April 17, 2021 to July 4, 2021.

As a robustness check, we re-run the mechanisms analysis in Web Appendix D excluding data after April 16, 2021 (i.e., without any imputed stay-at-home data) and the results were nearly identical, implying that the results are not an artifact of the imputation procedure.

A.3. 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).[7] We use this data in our mechanism analysis in Web Appendix D. 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.

A.4. YipitData

YipitData,[8] 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 July 2021, which we use in our mechanism analysis in Web Appendix D to determine which CBSAs had delivery services available during out data period, and for the platform breadth robustness check in Web Appendix D.4.3.

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

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

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 to July 2021 for these six platforms. There are 858 CBSAs with restaurant listings as of July 2021, and 1,265,614 restaurants listed on at least one of the aforementioned six restaurant delivery services during our observation period. For our analysis of mechanisms, we subset our data to the CBSAs that had at least one restaurant listing for the entire period of the regression data (January 2020 to July 2021).

A.5. 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), which we use for our restaurant dining simultaneity robustness check in Web Appendix D.4.1. 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 1,082 unique policies from March 15th, 2020 to July 4th, 2021. Figure 11 plots the total number of unique policies each month over this time period. 722 of these policies were state-wide orders, while the other 360 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.

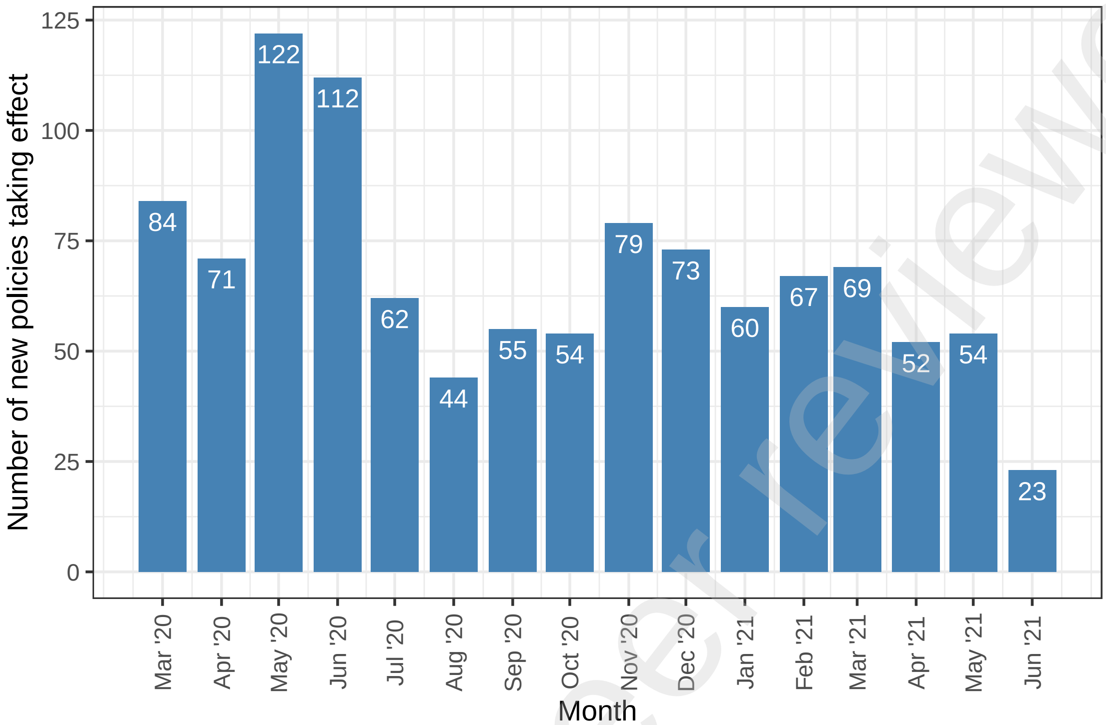
Our policy date extraction creates state-day policy data by identifying what policy was in effect each day. For example, if a policy was issued on May 15th, applying to the period of time from May 22nd to June 15th, and another policy is issued on June 1st, applying to the period of time from June 7th 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 360 policies that were not state-wide. Lastly, we mapped 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.

Appendix B: Details of lift analysis model implementation and estimates

In this section, we detail the specification and computational implementation for the APC, seasonality, and time series models, along with discussion of other subtleties such as the identification of the APC model and the autocorrelation order selection for the time series model. We also present the full estimates from these models for all four dependent variables of interest.

Figure 11 **State government policies enacted by month**



B.1. Age-period-cohort (APC) model

As discussed in Section 4.1 of the main text, we use an age-period-cohort (APC) model to decompose cohort-week observations into cohort, tenure, and calendar time effects. The model specification takes the form of a three-way fixed effects model:

$$\log(Y_{ct}) = \alpha_c + \gamma_{t-c} + \delta_t + \varepsilon_{ct} \tag{2}$$

We estimate this model using weighted least squares as implemented in the `lfe` package in `R`, which efficiently estimates multi-way fixed effects models using the method of alternating projections, iterating between centering by each group of fixed effects until converging to a fixed point (Stammann 2017). We weight observations by cohort size (number of members of the cohort). Throughout our lift analysis, we exclude the first week of each cohort (the week in which the cohort members placed their first order) due to these weeks being on a different order of magnitude as they contain the first orders for every customer in the cohort.

As noted in Section 4.1 of the main text, the APC model as laid out above is not point identified without further constraints due to cohort, tenure, and calendar time being linearly dependent. We can observe this by defining the following perturbed parameters:

$$\alpha_c^\dagger = \alpha_c + \varphi \cdot c$$

$$\gamma_{t-c}^\dagger = \gamma_{t-c} + \varphi \cdot (t-c)$$

$$\delta_t^\dagger = \delta_t - \varphi \cdot t$$

for any scalar $\varphi \neq 0$. We can see that this perturbed set of functions results in an observationally identical model to the original set of parameters α_c, γ_{t-c}, δ_t. Specifically, due to tenure being a linear function of cohort and tenure, the linear trends in α_c, γ_{t-c}, δ_t are not identified—however, nonlinear deviations from the linear trends are fully identified (Fosse and Winship 2019). As such, putting a single constraint on the linear trend of either α_c, γ_{t-c}, or δ_t is sufficient to achieve point identification.

Conceptually, as discussed in the main text, there is no issue with this identification problem, because our estimand of interest is the deviation of δ_t from its pre-COVID trend, which is a nonlinear quantity unaffected by adding a linear term to the entire series. Without loss of generality, we achieve point identification by imposing the constraint that δ_t has no linear trend in the pre-COVID period, which makes the effect of COVID visually easier to discern.

Computationally, because the estimation procedure we use is based on iterative centering rather than explicitly solving the closed-form least squares formula, there is no matrix inversion and so the procedure still converges to a fixed point without imposing explicit constraints despite the model being singular. The fixed point is not unique, since as seen above, perturbing the solution by a linear term also results in an equivalent solution; but, after subtracting a linear term to impose a post-hoc linear constraint, the solution is unique. Thus, in practice, we simply estimate the model with no constraints, then impose the constraint post-hoc by estimating the linear trend in δ_t based on pre-COVID observations and subtracting this linear term from the estimated δ_ts (and adding the offsetting linear term to α_c and γ_{t-c}).

Though our observations for estimation are aggregated to the cohort level, the aggregations are based on observations at the level of individual panel members; as such, we calculate confidence intervals around the estimated APC model using bootstrapping, where we bootstrap by individual panel member.

Figures 12, 13, and 14 present the estimated cohort effects (α_c), tenure effects (γ_{t-c}), and calendar time effects (δ_t) respectively for each of our four dependent variables of interest. The bands represent 95% pointwise confidence intervals calculated based on 250 bootstrap samples. The dashed vertical line in Figure 14 represents March 1st, 2020, which we use as the delineating date between the pre-COVID and post-COVID period; that is, the week of March 1st, 2020 is considered the first week of the post-COVID period. The adjusted R^2 values for the actives, orders per active, order size, and sales APC models are 98.2%, 93.7%, 88.2%, and 98.8%, respectively.

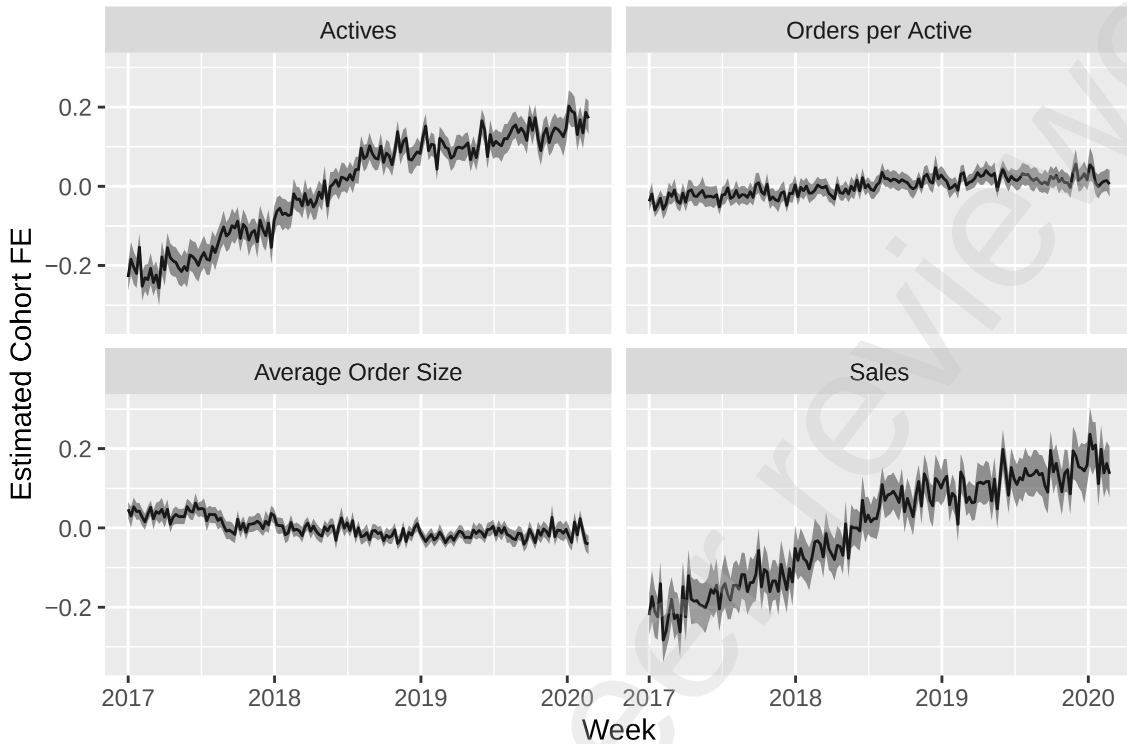
B.2. Seasonality model

The estimated calendar time effects in Figure 14 show clear evidence of seasonality and holiday effects, and so we residualize these effects out to further isolate the effect of COVID. Taking the estimated of δ_t from Figure 14, we further decompose this calendar time effect into a seasonal and non-seasonal component:

$$\delta_t = f(w_t) + \vec{\psi}'\vec{x}_t + \xi_t$$

where w_t represents the week of year (from 1 to 53) and $\vec{x}_t$ is a vector of dummy variables for holidays. We parameterize $f(w_t)$ using cubic b-splines with two interior knots in April and August (Hastie 1992), resulting

Figure 12 APC Model Estimated Cohort Effects



Note: bands represent 95% pointwise confidence intervals.

in five degrees of freedom and allowing a smooth seasonal effect over the course of the year.[9] For $\vec{x}_t$, we include dummy variables for Thanksgiving week, Christmas week, and New Year's Eve and/or Day week.[10]

We estimate this seasonality model using ordinary least squares based on the pre-COVID estimates of δ_t. We compute the residuals, subtracting the predicted values for all periods (including post-COVID periods), to obtain estimates of the non-seasonal calendar time effect ξ_t. Since the estimates of this model depend on the APC model decomposition estimated previously, we propagate uncertainty from the APC model by bootstrapping at the panel member level, re-estimating the APC model and then re-estimating the seasonality model based on the new APC model at each bootstrap.

Figure 15, Table 2, and Figure 16 present the estimated seasonality effects ($f(w_t)$), holiday effects ($\vec{\psi}$), and non-seasonal calendar time effects (ξ_t) respectively for each of our four dependent variables of interest. The bands represent 95% pointwise confidence intervals calculated based on 250 bootstrap samples. The dashed vertical line in Figure 16 represents March 1st, 2020.

B.3. Time series model of calendar time effect evolution

As noted in Section 4.1 in the main body, we measure the impact of COVID on within-cohort customer behavior by forecasting the counterfactual no-COVID baseline for the non-seasonal calendar time effects (ξ_t),

[9] The estimated seasonality effects are substantively the same using a larger number of knots.

[10] We collapse these into one variable since the two days usually fall in the same week, making them difficult to separate out.

Figure 13 APC Model Estimated Tenure Effects

Note: bands represent 95% pointwise confidence intervals.

Table 2 Holiday Coefficients from Seasonality Model

Variable	Active Customers		Orders per Active Cust.		Avg. Order Size		Total Spending	
Regressor	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Thanksgiving	−0.145***	(0.004)	−0.039***	(0.002)	0.061***	(0.003)	−0.124***	(0.004)
Christmas	−0.230***	(0.005)	−0.053***	(0.003)	0.064***	(0.003)	−0.219***	(0.006)
New Year's Eve/Day	0.022***	(0.005)	0.008*	(0.004)	0.036***	(0.004)	0.066***	(0.007)

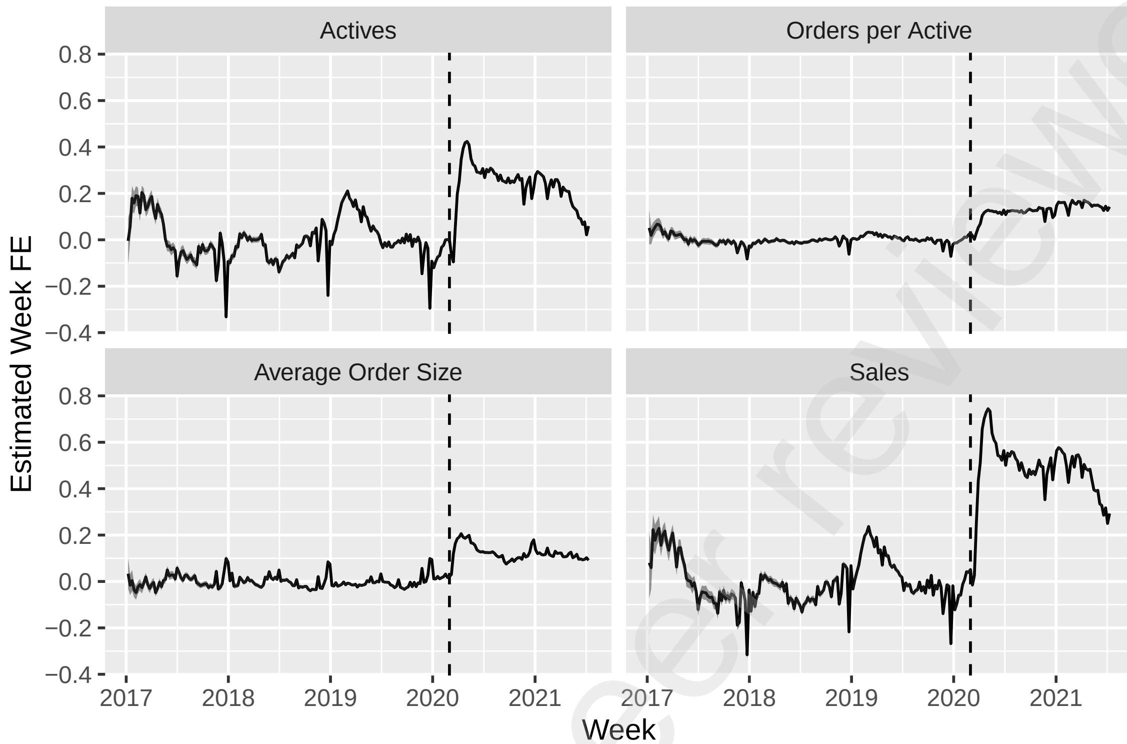
Note: asterisks denote level of significance (†: $p < 0.1$, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$). Standard errors are clustered at the cohort level. The sample size on all regressions is 13,530 with 164 unique cohorts and 164 unique tenure levels.

conditional upon ξ_t over the pre-COVID period, for each of our four dependent variables. To obtain valid inference for our lift estimates, we must specify appropriate time series models from which to obtain these forecasts.

It is visually evident from Figure 16 that ξ_t over the pre-COVID period – especially for the active customer and sales dependent variables – is significantly autocorrelated. We address this autocorrelation by selecting an autoregressive model of the form

$$\xi_t = \sum_{l=1}^{L} \phi_l \xi_{t-l} + \zeta_t$$

Figure 14 **APC Model Estimated Calendar Time Effects**



Note: bands represent 95% pointwise confidence intervals.

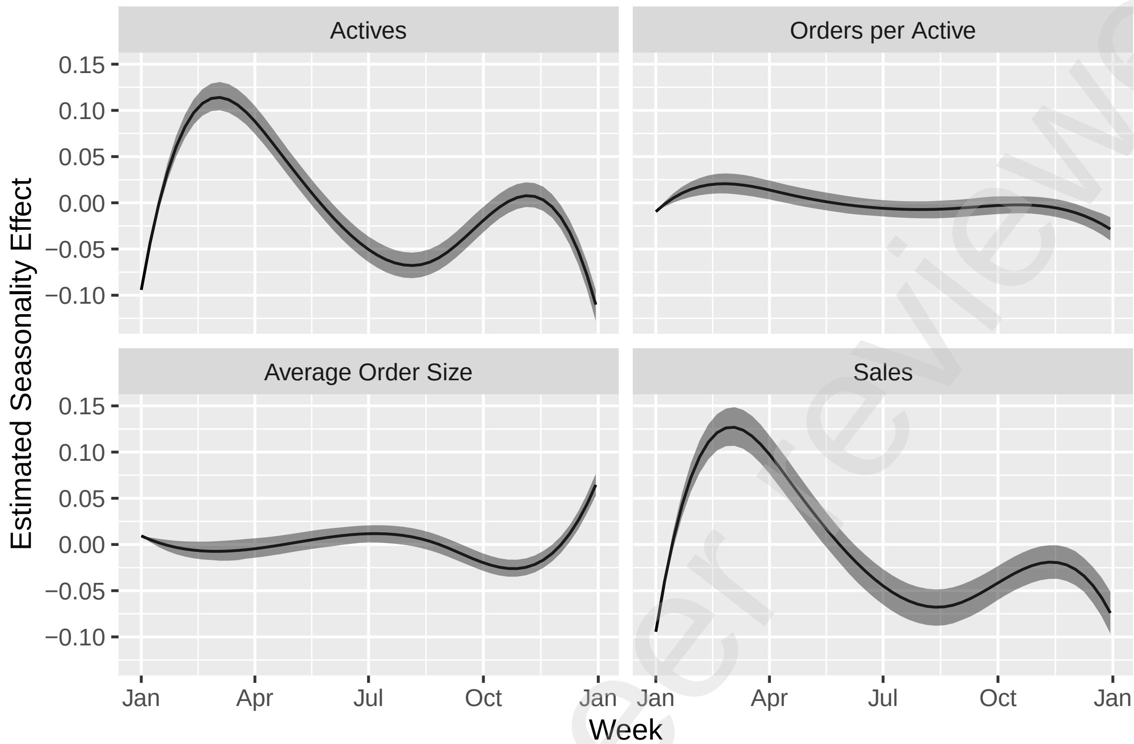
We select L, the order of the AR model, by obtaining the AIC associated with each possible value of L from 0 to 10 for each of the four dependent variables, then selecting the value of L associated with the lowest AIC value, on average, across those dependent variables, using the `arima` function in `R`. This procedure results in our selecting an L of 4 (average AIC: -821.9).[11] Coefficient estimates (standard errors) for each dependent variable are shown in Table 3.

The residuals of the time series models pass common tests for time series non-stationarity and autocorrelation. For example, the Augmented Dickey-Fuller Test, applied to the residuals of the AR(4) model, rejects the null hypothesis of nonstationarity across all types 1 to 3, and all lags 1 to 10, for all four dependent variables of interest at the 1% significance level. Similarly, the Ljung-Box test does not reject the null hypothesis that the data are independently distributed for all four processes, with p-values in excess of 17% for all lags between 1 and 10.

The resulting fits over the pre-COVID period, shown in Figure 17, appear sensible. Autocorrelation explains 74%, 60%, 25%, and 72% of the variation in ξ_t for active customers, orders per active customer, average order value, and sales respectively.

[11] We could alternatively allow for process-specific orders L. Doing so, selected by AIC, would result in L values of 4, 4, 3, and 7 for spend, average order value, orders per active customer, and active customers, respectively. We do not do so for expositional simplicity, and because the resulting fits and forecasts are nearly identical to the fits obtained assuming L equal to 4 for all process.

Figure 15 Seasonality Model Estimated Seasonality Effects



Note: bands represent 95% pointwise confidence intervals.

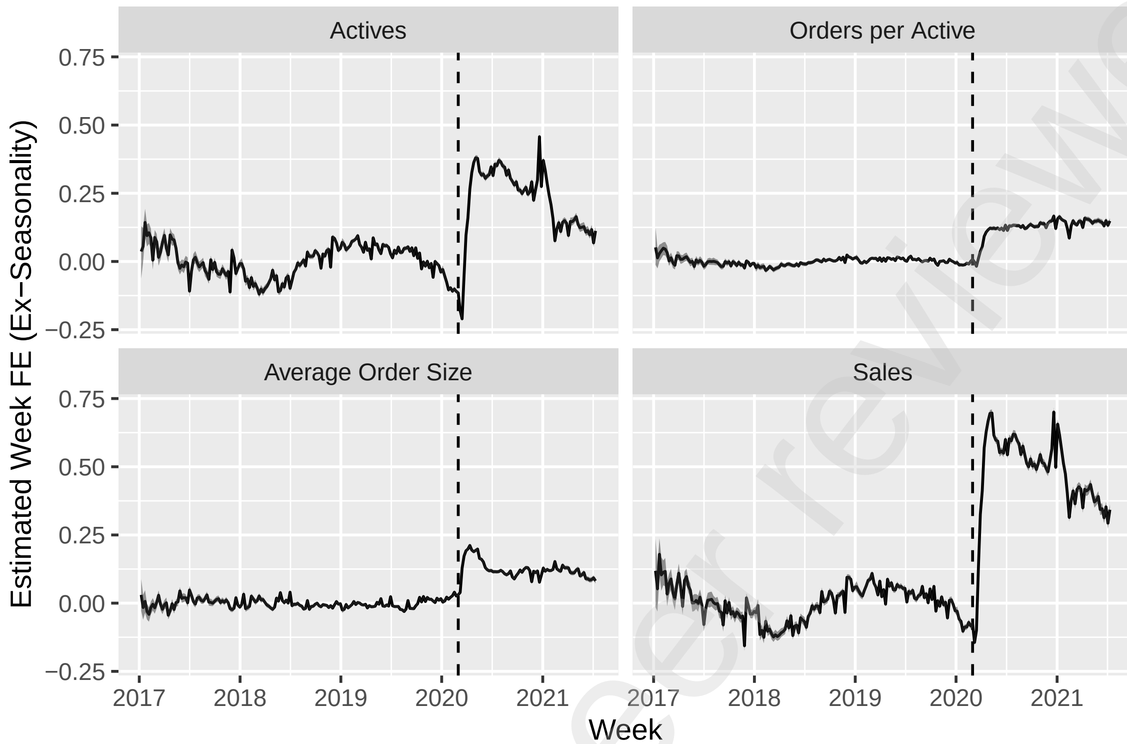
Table 3 Coefficient Estimates from AR(4) Model

Variable	Active Customers		Orders per Active Cust.		Avg. Order Size		Total Spending	
Regressor	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
ϕ_1	0.4917***	(0.0377)	0.4855***	(0.0960)	0.3584***	(0.0611)	0.3575***	(0.0527)
ϕ_2	0.1881***	(0.0449)	0.2050*	(0.0946)	$0.1146^{\dagger}$	(0.0599)	0.2852***	(0.0552)
ϕ_3	0.0839	(0.0515)	0.0889	(0.1016)	0.0427	(0.0601)	0.0235	(0.0510)
ϕ_4	0.1704***	(0.0460)	0.1090	(0.1065)	0.1335*	(0.0558)	0.2766***	(0.0453)
Loglikelihood	340.67		541.58		457.69		323.94	
AIC	-671.34		-1073.15		-905.39		-637.88	

Note: asterisks denote level of significance (†: $p < 0.1$, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$). The sample size on all regressions is 164, equal to the number of weeks in the pre-COVID period.

Given that the predictive uncertainty from the time series models dwarfs the uncertainty in estimation of ξ_t seen in Figure 16, for simplicity we present the predictive intervals directly from the estimated time series model without taking into account the uncertainty in the first-stage model estimates of ξ_t. Below, we verify for the sales dependent variable that propagating uncertainty from the estimation of ξ_t through bootstrapping results in nearly identical predictive intervals.

Figure 16 **Seasonality Model Estimated Non-Seasonal Calendar Time Effects**



Note: bands represent 95% pointwise confidence intervals.

Figure 17 **In-Sample Fit of AR(4) Model by Process**

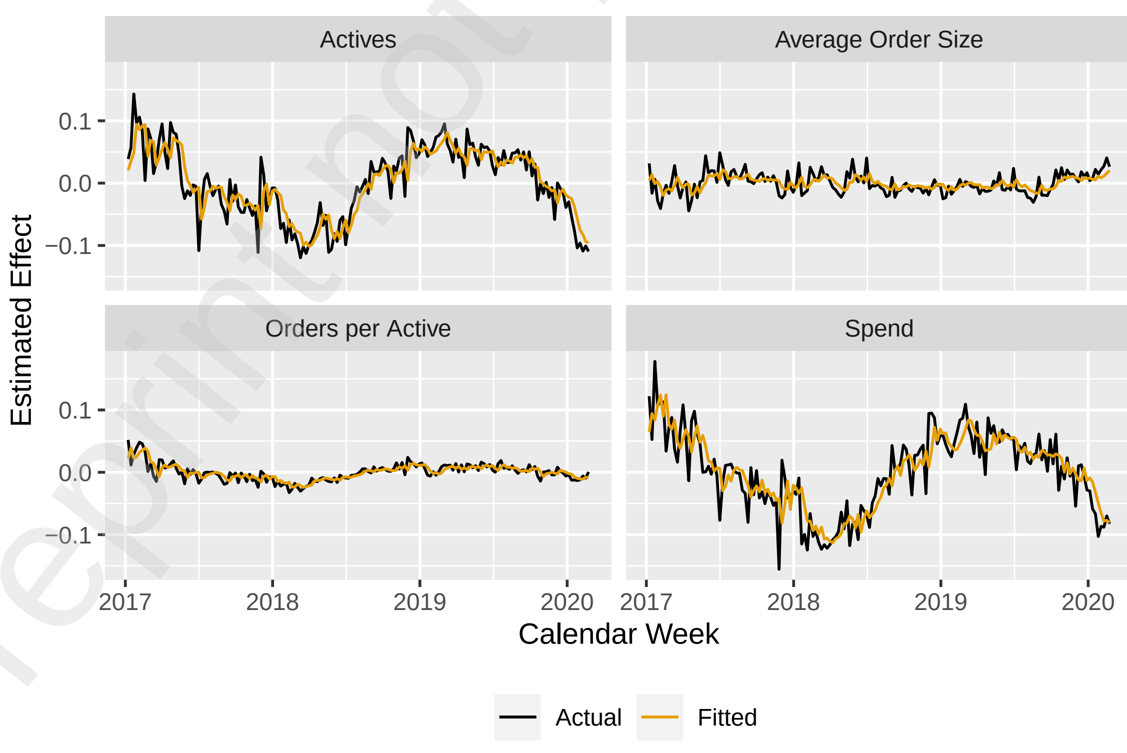
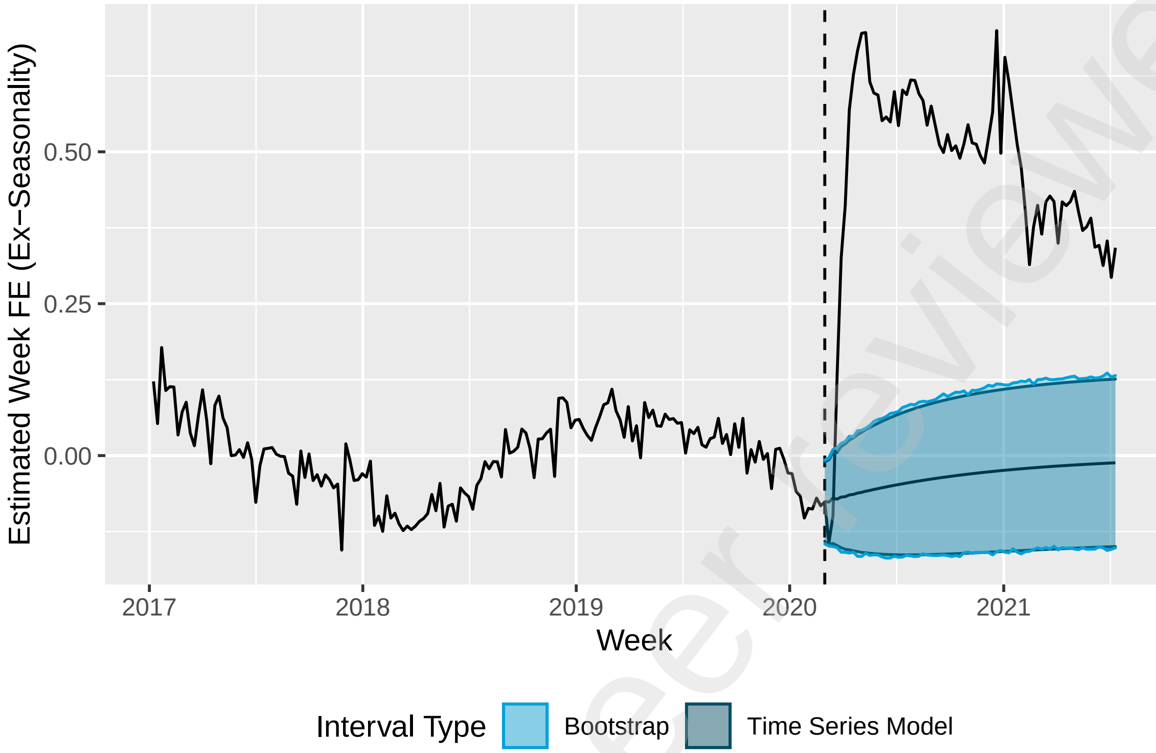


Figure 18 **Comparison of Predictive Intervals For Estimated Calendar Effect, Spend**



Note: dotted bands represent 95% pointwise predictive intervals obtained directly from the estimated time series model (dark blue) and non-parametric bootstrap (light blue).

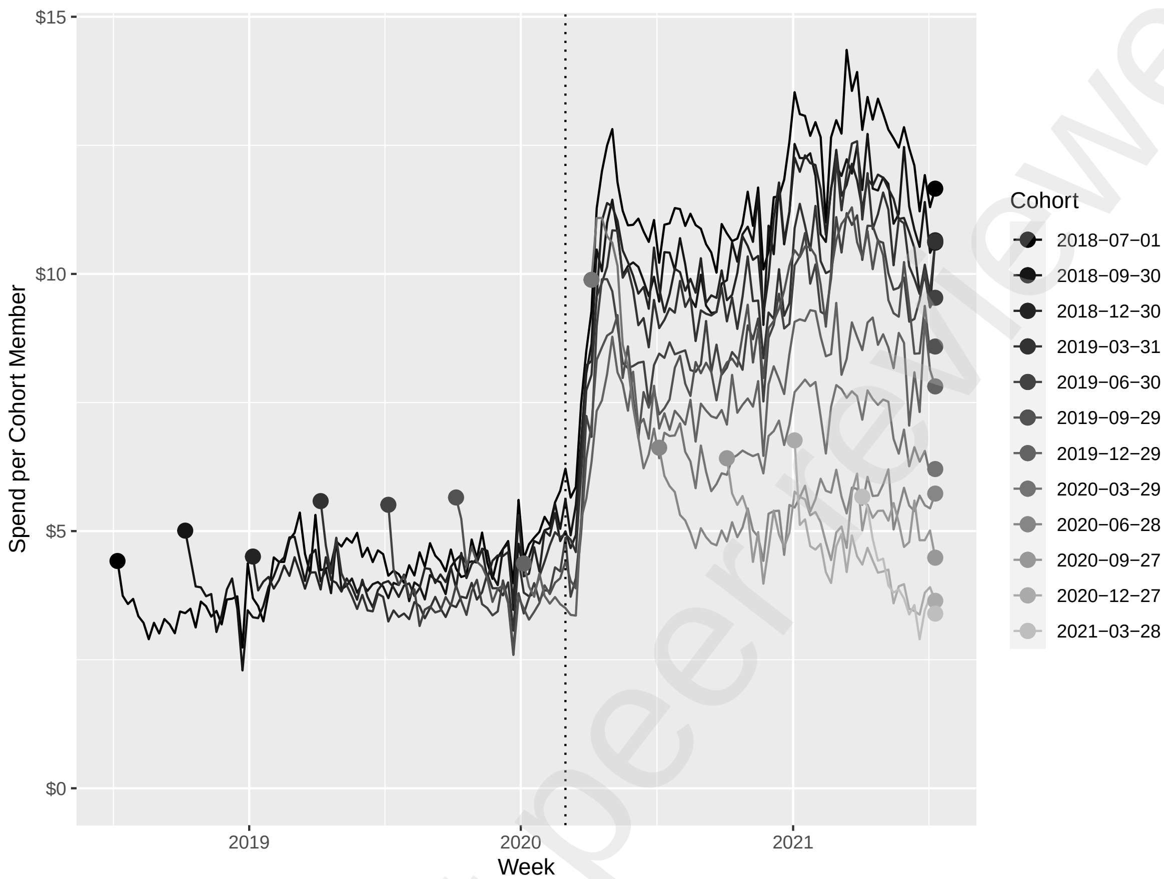
B.3.1. Alternative calculation of predictive interval As mentioned above, the interval shown in Figure 5 of the main body is a 95% pointwise predictive interval, obtained directly from the estimated time series model, and as such does not account for uncertainty in the first-stage model estimates of ξ_t. As a robustness check, we instead compute the predictive distribution via non-parametric bootstrap, to assess whether our results are substantively the same when we do take into account uncertainty in the first-stage model estimates of ξ_t.

We bootstrap each panel member within the population with replacement. Each time, we re-fit the AR(4) model, then forward-simulate, period-by-period, data over our forecasting horizon by adding to each future period's point forecast a random draw from a normal distribution with mean zero and variance equal to the in-sample variance of the residuals. The resulting comparison of theoretical and bootstrapped predictive intervals for total spending, shown in Figure 18, suggests that the predictive interval obtained using the two methods are effectively equivalent to one another. The predictive intervals are nearly identical because estimation uncertainty for ξ_t is very small, as seen in Figure 16.

Appendix C: Cohorted spending patterns in calendar time

In this section, we provide further insight into the spending patterns of customers acquired during COVID discussed in Section 5.1 of the main text. Figure 19 shows cohorted spending per week across the same cohorts shown in Figure 8 of the main text. However, instead of plotting each cohort in a separate panel as a

Figure 19 **Weekly Cohort Spending in Calendar Time**



Note: Vertical dotted lines represent March 1st, 2020. Legend dates indicate the start date of the cohort's acquisition week (e.g., the 2018-12-30 label indicates that the plot corresponds to the cohort acquired during the week of December 30th, 2018). The first and last spend per week by cohort are overplotted.

function of tenure, Figure 19 plots spending per week across all cohorts on one panel in calendar time. The first and last spend per week associated with each cohort are further emphasized through points overlaid on the line graph.

We observe the calendar time evolution of initial spend per week across cohorts through the heights of the leftmost points, moving from left to right. Before COVID, cohorts' initial spends were gradually increasing over time, modulo a seasonal decline in average spending at the end of each year. This is also evident from the gradually increasing pattern in cohort fixed effects from the top-left panel of Figure 4 of the main text. Initial spending was significantly elevated for cohorts born shortly after the onset of COVID. However, initial spending then fell sharply back to the pre-COVID trend by the cohorts acquired in the second half of 2020, while cohorts acquired thus far in 2021 have been spending less than what would have been implied by the pre-pandemic trend.

We observe the cohort-based evolution of average spending the week of July 11th 2021, the final week of the dataset, through the heights of the rightmost points, moving from black (older cohorts) to grey (younger cohorts). Average spending at the end of the data period monotonically decreases as for more recent cohorts. Pre-COVID cohorts are all spending well in excess of their pre-pandemic trends, and have generally maintained spending close to the maximum absolute level generated during the peak of COVID. Secondly, cohorts born immediately after the onset of COVID had elevated levels of initial spending which subsequently declined to sharply below their initial spending levels, but to levels in excess of what would have been implied by pre-pandemic trends. More recently acquired cohorts have both below-trend initial spending, and below-trend spending in the final week of the data set.

Taken together, these patterns further support the hypothesis that behaviorally, the forces driving people towards restaurant delivery have largely dissipated. Indeed, they even suggest that we are now seeing overcompensation, with recent cohort spending falling to levels below what pre-pandemic data would have suggested.

Appendix D: Analyzing Mechanisms

Our main analysis provides us with an estimate of the overall impact of COVID-19 on the behavior of pre-existing customers, but does not provide us with insight into the mechanisms driving changes in customer behavior. Towards this end, we employ a panel regression which utilizes spatiotemporal variation to identify how our mechanisms of interest amplified or dampened the effect of COVID-19. While the intuition and results were discussed in Section 5.1 of the main text, this section discusses the full details of the model specification, identification, implementation, results, and robustness checks.

We consider four mechanisms of how COVID has affected customer behavior: income shocks (operationalized as unemployment rates), stay-at-home behavior, restaurant supply (operationalized as restaurant employee shifts), and restaurant dine-in behavior. Our reasoning for considering each of these variables, along with our data sources and definitions of our operationalizations of them, are given in Table 4.

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.

D.1. Model specification

We model observations at the CBSA-day level with a two-way fixed effects (FE) log-log regression model:[12]

$$\log\left(Y_{lt}\right) = \alpha_l + \alpha_t + \vec{\beta}' \log\left(\vec{X}_{lt}\right) + \varepsilon_{lt} \tag{3}$$

where Y_{lt} represents the dependent variable of interest and $\vec{X}_{lt}$ represents the vector of regressors/mechanisms of interest for CBSA location l on day t. We perform analysis at the CBSA-day level. As discussed below, the use of two-way fixed effects ensures that identification is driven by on-the-margin variation in how different regions of the US have been affected by COVID-19.

We primarily focus on sales per capita in CBSA l on day t as our dependent variable. We also consider three other dependent variables – proportion of panel members newly acquired, orders per capita, and

[12] Since the dependent variables are sparse (i.e., there are many CBSA-day observations where zero purchases are observed), in practice we use $\log(a+Y_{lt})$ 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}.

Table 4 Summary of Regressors

Mechanism	Reasoning	Definition
Income shocks	Many people experienced negative income shocks due to the pandemic, so are likely to reduce discretionary spending such as food delivery.	Unemployment rate by CBSA-month as reported by the Bureau of Labor Statistics.
Stay-at-home behavior	Many people were/are spending more time at home due to the pandemic. This drives consumers towards food options that do not require going outside, but also endows them with more time to cook at home.	Proportion of people staying completely at home by CBSA-day, as reported by SafeGraph, a mobile location panel covering 18.8 million devices.
Restaurant supply	Many restaurants, which supply food to delivery services, closed due to the pandemic, hurting the supply capacity of delivery services.	Number of employee shifts per capita worked at restaurants by CBSA-day, as proxied by restaurant visits of over four hours, as reported by SafeGraph based on 963 thousand restaurants.
Restaurant dine-in	Many states restricted restaurant dine-in, and consumers were/are hesitant to dine in due to potential health risks, driving consumers towards delivery as a substitute.	Number of restaurant dine-in visits per capita by CBSA-day, as proxied by restaurant visits of between 20 minutes and four hours, as reported by SafeGraph.

average order size – to decompose the effect of mechanisms on different components of overall sales.[13] For our regressors, we include the four variables laid out in Table 4.

D.2. Identification of Effects

D.2.1. Overview Intuitively, while COVID-19 impacted all regions of the US simultaneously, the severity of initial impact and trajectory of recovery has differed, enabling identification. However, there are a number of potential cross-sectional differences between CBSAs and aggregate shocks across CBSAs over time that may be sources of endogeneity, which motivates our use of fixed effects as detailed below.

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 2020, 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 intuition is presented visually in Figure 9 of the main text.

We note that while using two-way fixed effects allows for relatively clean identification of the marginal impacts of our mechanisms of interest, it also means that we cannot measure the absolute magnitude of the national effect of these mechanisms on customer behavior without assuming that the effect sizes of on-the-margin variations generalize to cross-sectional and aggregate time series variation in our mechanisms of interest. As such, while we believe our estimates capture the directional effects of our mechanisms of interest on delivery behavior, the specific magnitude of the estimates should be considered stylized.

This discussion gives only a high-level overview; below, we provide a more thorough discussion of the source of identification of each coefficient and present additional robustness checks for a number of potential endogeneity concerns.

D.2.2. Full details of identification In this section, we elaborate upon the identifying assumptions of our panel regressions. As stated above, 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.

[13] Here, we do not separate out the proportion of active customers from orders per active customer, since at the daily observation level almost all active customers place only one order.

First, 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 8 and discussed further in Web Appendix D.4.2. 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 on a given day translates to 1.74% 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 have shifted capacity towards outdoor dining (which may be further constrained by weather), and most have had some form of spacing requirements that drastically reduced the number of tables that they can serve. Even as capacity constraints have been lifted since vaccine rollout, a substantial labor shortage is still constraining the capacity of restaurants. 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 D.4.1. 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 and 2021.

Lastly, 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 20 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. Apart from an initial dip in listings in April and May of 2020, there is no visually discernible effect of COVID on restaurant listings. Listings plateaued in late 2020 and began declining in June 2021, but it is unclear whether this is pandemic-related.

Thus, while platform breadth (variety of restaurants available on delivery platform) is likely a driver of customer behavior, we do not consider it to be a mechanism of the COVID effect per se, due to it being unclear whether it was actually affected by COVID in the first place. Nevertheless, in Web Appendix D.4.3, we confirm that including platform breadth in our regressions does not change our results.

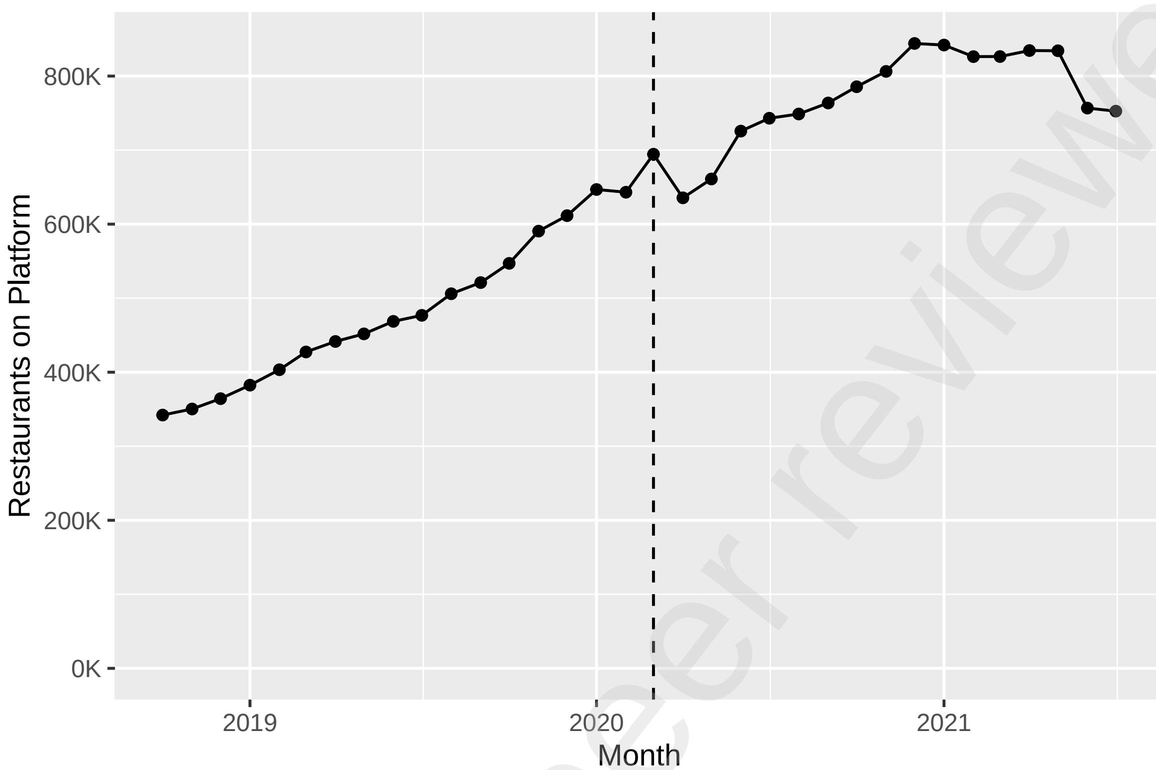
Thus, while there are many possible reasons why the mechanisms of interest may be confounded, we believe that the fixed effects and other control variables, as well as the accompanying robustness checks, demonstrate that we have plausibly exogenous variation for identification.

D.3. Results

We estimate our model by weighted least squares (WLS) using observations from January 1, 2020 to July 4, 2021 (the last date for which we have Safegraph data).[14] We compute heteroskedasticity and cluster robust standard errors with two-way clustering by CBSA and day (Cameron et al. 2012). Table 5 reports the results.

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

Figure 20 Total Restaurants Listed on at Least One Platform Nationally, October 2018 - July 2021



Note: Vertical line represents March 2020.

Table 5 FE Regression Model Estimates

DV	Acquisitions		Orders		Order Size		Sales	
Regressor	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Unemployment	$-0.041^{\dagger}$	(0.0228)	-0.105***	(0.024)	0.005	(0.004)	-0.143***	(0.027)
Stay-at-home	$0.088^{\dagger}$	(0.046)	-0.230***	(0.058)	0.045***	(0.009)	-0.230***	(0.064)
Restaurant supply	0.104***	(0.030)	0.161***	(0.029)	-0.005	(0.007)	0.191***	(0.033)
Restaurant dine-in	-0.117***	(0.028)	-0.086***	(0.023)	0.002	(0.005)	-0.100***	(0.026)
N	429,780		429,780		347,331		429,780	

Note: asterisks denote level of significance (†: $p < 0.1$, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$). 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.

All of our regressors have a statistically and economically significant impact on orders 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 mainly attributable to decreased order frequency rather than changes in new customer adoption or order size.

Increases in stay-at-home behavior have driven larger order sizes (and marginally, more customer adoption), 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 significant 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.

D.4. Supplemental FE regression results

D.4.1. FE-IV model with policy instrumental variable As discussed in Web Appendix D.2, 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.
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 A.5, 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 Web Appendix D.1, 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 6, with the first stage model results reported in Table 7.

Table 6 FE-IV Regression Model Estimates

DV	Acquisitions		Orders		Order Size		Sales	
Regressor	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Unemployment	−0.048*	(0.023)	−0.109***	(0.024)	0.005	(0.004)	−0.147***	(0.026)
Stay-at-home	−0.054	(0.062)	−0.313***	(0.069)	0.029**	(0.011)	−0.328***	(0.075)
Restaurant supply	0.279***	(0.053)	0.263***	(0.044)	0.015	(0.010)	0.311***	(0.048)
Restaurant dine-in	−0.380***	(0.064)	−0.241***	(0.040)	−0.027*	(0.012)	−0.283***	(0.045)
N	429,780		429,780		347,331		429,780	

Note: asterisks denote level of significance (†: $p < 0.1$, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$). 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 first stage F-statistic on excluded instruments is 47.55 ($p \approx 0$), 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.

The estimates in Table 6 are directionally very similar to our results from Web Appendix D.3, though the IV estimates of the dine-in effect are 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 change their behavior 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.

D.4.2. FE model with lagged sales As noted in Web Appendix D.2, 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.

Table 7 FE-IV First Stage Estimates

Regression	Acquisitions/Orders/Sales		Order Size	
Regressor	Coef.	SE	Coef.	SE
Unemployment	−0.018	(0.016)	−0.019	(0.016)
Stay-at-home	−0.497***	(0.040)	−0.502***	(0.041)
Restaurant supply	0.654***	(0.031)	0.656***	(0.032)
Indoor dining allowed	0.224***	(0.021)	0.223***	(0.021)
Outdoor dining allowed	−0.095***	(0.026)	−0.095***	(0.026)
Indoor dining reduced	$-0.016^{\dagger}$	(0.008)	$-0.015^{\dagger}$	(0.008)
N	429,780		347,331	
F(Excluded instruments)	47.55		46.75	

Note: asterisks denote level of significance (†: $p < 0.1$, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$). 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.

The results are given in Table 8. 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 0.174). 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.

D.4.3. FE model with platform breadth As noted in Web Appendix D.2, platform breadth (the availability of a variety of restaurants on a delivery platform) is likely to affect customer behavior. Though we rule this out as a mechanism of the COVID effect, due to COVID not having a clear effect on platform breadth, we further run a regression that includes platform breadth as a regressor to see whether platform breadth is confounding our results.

In particular, we include as an additional regressor the log count of total unique restaurants listed in a given CBSA-month across all six platforms covered by the Yipit data. The specification is identical to our main regression model except for this added regressor.

The results are given in Table 9. Platform breadth positively affects acquisitions, orders, and sales, with a marginally significant effect on order size: having a broader variety of restaurants available for delivery makes consumers more likely to adopt delivery services and to order more, but does not substantially effect the size of their orders. Importantly, the other coefficients are almost identical to our main results, showing that platform breadth is not an alternative explanation for our main findings.

Table 8 Lagged Sales Model Estimates

DV	Acquisitions		Orders		Order Size		Sales	
Regressor	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Unemployment	$-0.037^{\dagger}$	(0.023)	−0.088***	(0.021)	0.006	(0.004)	−0.116***	(0.022)
Stay-at-home	0.092*	(0.046)	−0.206***	(0.051)	0.046***	(0.009)	−0.192***	(0.054)
Restaurant supply	0.102***	(0.029)	0.142***	(0.026)	−0.006	(0.007)	0.160***	(0.028)
Restaurant dine-in	−0.115***	(0.028)	−0.073***	(0.021)	0.002	(0.005)	−0.079***	(0.022)
Lagged sales	0.025***	(0.003)	0.112***	(0.005)	0.004***	(0.001)	0.174***	(0.007)
N	429,000		429,000		346,796		429,000	

Note: asterisks denote level of significance (†: $p < 0.1$, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$). 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.

Table 9 Platform Breadth Model Estimates

DV	Acquisitions		Orders		Order Size		Sales	
Regressor	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Unemployment	−0.037	(0.023)	−0.096***	(0.024)	0.006	(0.004)	−0.127***	(0.026)
Stay-at-home	0.095*	(0.046)	−0.213***	(0.058)	0.046***	(0.009)	−0.202**	(0.063)
Restaurant supply	0.109***	(0.030)	0.171***	(0.028)	−0.004	(0.007)	0.206***	(0.032)
Restaurant dine-in	−0.119***	(0.028)	−0.092***	(0.021)	0.002	(0.005)	−0.109***	(0.026)
Platform breadth	0.123***	(0.025)	0.310***	(0.046)	$0.014^{\dagger}$	(0.008)	0.507***	(0.077)
N	429,780		429,780		347,331		429,780	

Note: asterisks denote level of significance (†: $p < 0.1$, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$). 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.