Habits in consumer purchases: Evidence from store closures*

Amir Tohidi
Laboratory for Information and Decision Systems, Institute for Data, Systems, and Society, MIT, atohidi@mit.edu

Dean Eckles
Sloan School of Management, Institute for Data, Systems, and Society, MIT, eckles@mit.edu

Ali Jadbabaie
Laboratory for Information and Decision Systems, Institute for Data, Systems, and Society, MIT, jadbabai@mit.edu

In-store shopping involves recurring behaviors often happening in a familiar environment, potentially facilitating habit formation. Habit formation in retail purchases is broadly consistent with the finding that consumers' choices exhibit inertia. However, much prior work has left open the question of whether much of this inertia should be attributed to habits. Here, using panel data on households' purchases across many product categories, we use store closures as a shock that partially disrupts households' shopping to identify the role of formed habits in repeated brand purchases. Closures cause people to purchase products elsewhere, where we posit they are more likely to engage in deliberative decision-making processes — driving them to explore some options that are normally ignored in a familiar store because of learned habits. Following a closure at which they frequently purchased products in a given category, households are more likely to purchase something other than their modal brand in that product category. This effect persists even after accounting for reduced availability of their modal brand. Over time, households return to higher levels of repetitive choices of a single brand, consistent with the formation of new habits. Indeed, this temporary decrease in repeated choice of a modal brand results in a lasting impact on households' brand choices, leading to lower rates of purchase of their baseline modal brands in the long run. The work sheds light on how habit formation, while advantageous for consumers in other ways, can lead to persistence of sub-optimal choices, with implications for brands and retailers.

Key words: habits, brand loyalty, state dependence, retail, difference-in-differences
History: This version: 17 August 2022

---

1. Introduction

During the every-day experience of shopping in grocery-stores consumers are faced with a multitude of brand options across numerous product categories. Choosing what to buy is a complex decision-

* This paper reports the authors' own analyses calculated based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the authors and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

making process involving large, varying choice sets and various marketing parameters such as prices and promotions. As a result, consumers may rely on heuristics and habits as means of faster and less-effortful decision making. Importantly, this is a repetitive behavior happening in the same context. According to psychological theory, habits are likely to form when a rewarding behavior is repeated over and over in the same context (Wood and Rünger 2016, Verplanken and Aarts 1999), where in the long-term contextual cues trigger an automatic response (Orbell and Verplanken 2010); see Gardner (2015) for a review. As a result, we could expect that for product categories where consumers repeatedly make purchases in similar settings (e.g., the same store) these purchases may be habitual.

Habitual behavior in shopping could be one explanation of the empirical regularity that purchase decisions exhibit substantial temporal-dependency or inertia. (e.g., Carrasco *et al.* 2004, Dubé *et al.* 2010). Following work in psychology, we hypothesize that such repeated behaviors in stable contexts are often the result of slow-learning, fast-acting (i.e. System 1) processes (Wood and Rünger 2016, Mazar and Wood 2018). Of course, other varieties of inertia can also be present, such as learning and preference formation that are felicitously described through standard choice models. Here we aim to detect and estimate a role for habits specifically in explaining observed inertia in consumer purchase behavior. Hence, part of the contribution of this paper is to provide robust and rigorous empirical evidence to bridge this gap and bring insights from the psychological definition to a setting of broad interest in marketing, economics, public policy, and decision science. Strong consumption habits are at the core of marketing strategies designed to attract consumers from other brands, e.g., by temporary price reductions or free sampling, in order to benefit from their choice inertia in the long-run. Furthermore, understanding the habitual mechanisms behind repeated purchases could also be helpful for policy makers and individuals seeking behavioral change.

1.1. Related work

Habit formation, persistence, and disruption have been studied from a variety of perspectives across many disciplines. One aim of this paper is to adapt a theory of habit developed in the psychology and consumer behavior literatures to one of the most widely-studied settings in quantitative marketing, in this process clarifying the causes of stylized facts in quantitative marketing and economics. So we first consider definitions and theories of habit in psychology and then empirical work in quantitative marketing.

Since William James (1890) argued that "habit covers a very large part of life", psychological theory has posited a large role for habits, with Wood *et al.* (2002) concluding that more than a third of people's daily decisions could be considered to be habitual. While there is not universal agreement on the details, multiple competing accounts converge on defining habits with particular forms of automaticity. Wood *et al.* (2014) characterize habits as being activated by recurring contextual cues and being insensitive to short-term changes in goals. On this and related accounts, the repeated performance of a behavior in the presence of the same contextual cues creates an association between the context and the behavior, making performing it in that context automatic or proponent, thereby reducing deliberation prior to choice (Wood and Rünger 2016).[1] Second, this association is not immediately severed by a change in goals (or payoffs). That is, habits are relatively insensitive to changes that devalue taking the habitual action. This is reflected in experimental paradigms that test for habits (or the strength of habits) by examining persistence of the behavior even when the reward is absent or no longer desired (Dickinson 1985, Adams and Dickinson 1981, Beshears *et al.* 2021). This account thus predicts both "inertia" in habitual behaviors even as payoffs change, but also potentially rapid change when there is a change in the context or the habitual choice is unavailable. In an influential empirical study related to the present work, Wood *et al.* (2005) had participants report their frequency of performing exercise and media consumption behaviors before and after transferring to a university, which changed the context for some of those behaviors for some participants; following the transfer, behaviors in changed contexts changed to more closely match participants' self-reported goals, and they changed more than behaviors that continued to be performed in the same context before and after the transfer. This account of habits — in which deliberative processing of attributes of many available options is often absent — is consistent with studies that have probed visual attention of in-store shopping with mobile eye-tracking devices. For example, in a study of grocery shoppers in Uruguay, Machín *et al.* (2020) found that 67% of shoppers directly chose the product they were looking for and put it in their basket without any comparison to other products. Thus, psychological theory and empirical studies suggest that in-store shopping may be quite habitual, in the sense just articulated; this motivates our hypothesis that sudden disruptions to the context (e.g., the closure of a frequently-visited store) will cause consumers to engage in more deliberative decision-making, perhaps choosing products that may match their current, considered preferences.

In quantitative marketing and economics, researchers have long identified strong autocorrelation or "inertia" in consumption patterns across various product categories (Cunningham 1956, Guadagni and Little 1983, Keane 1997, Seetharaman *et al.* 1999, Dubé *et al.* 2010). Classically, the strength of this inertia has been captured by a *state dependence* (or loyalty) variable that captures

[1] To capture the key constructs involved in the definition of habitual behavior, various measurement schemes have been proposed. Inspired by the idea that habits are formed by repetition of certain actions in a stable context, Wood *et al.* (2005) suggested the frequency-in-context measure. To capture other features of habits such as automaticity of behavior, researchers have proposed other habit measures such as Self-Reported Habit Index (SRHI) (Verplanken and Orbell 2003), Self-Reported Behavioral Automaticity Index (SRBAI) Gardner *et al.* (2012), and the Slip of Action paradigm de Wit *et al.* (2012).

the dependence of one's current decision on her past purchases. This method was first introduced by Guadagni and Little (1983). They showed that the inclusion of past purchases as a measure of loyalty in a choice model can significantly improve the model fit. In much subsequent work, researchers added a Markovian assumption and used the immediate past purchase as a proxy for the "state" of the consumer (e.g., Dubé *et al.* 2010, Simonov *et al.* 2020, Levine and Seiler 2021). Hence, the phenomenon of positive inertia or loyalty has also been called state dependence in the literature.[2]

Much of the focus in the state dependence literature has been on validation of the estimation procedure and using flexible models to account for unobserved heterogeneity in consumer preferences (Dubé *et al.* 2010). Overall, the apparent consensus in the literature is that there is some structural state dependence even after using fairly complex models controlling for possible heterogeneity in preferences (cf. Levine and Seiler 2021), although the estimated size of state dependence shrinks after controlling for, e.g., consumer heterogeneity (Keane 1997, Dubé *et al.* 2010, Simonov *et al.* 2020). In particular, Dubé *et al.* (2010) observe that the estimated state dependence disappears after permuting the order of shopping trips. If past decisions did not temporally affect future ones, and estimated state dependence was only an artifact of unobserved heterogeneity, it should have remained significant for the random reordering of trips. Moreover, Seetharaman *et al.* (1999) investigates different product categories and finds that households' state dependence is correlated across categories, again indicating that this is a real phenomenon beyond model misspecification.

Inertia in purchases can have multiple underlying causes (Liu-Thompkins and Tam 2013). One typical explanation in the literature is psychological switching costs (Klemperer 1987, Farrell and Klemperer 2007, Dubé *et al.* 2010), although this does not articulate underlying mechanisms which generate these mental costs. However, considering this established evidence for state dependence with respect to a psychological theory of habit, we may want to distinguish the associative, slow-learning, fast-acting process underlying habits from a deliberative, fast-learning, slow-acting process. In the next section, we describe how our empirical strategy is chosen for this purpose. Here we first consider some other decompositions of state dependence proposed in the quantitative marketing and economics literatures; see Thomadsen and Seetharaman (2018) for a review. Note that sometimes these have used the term "habit", but in ways that diverge both from our use and each other. In a latent utility framework, Roy *et al.* (1996) incorporate both what they call "habit persistence" and "structural state dependence" in their model.[3] While structural state dependence

[2] Negative state dependence or variety seeking (McAlister 1982) has also been observed in some cases but to a lesser extent (Adamowicz and Swait 2013).

[3] In this model, *structural state dependence* is defined as a direct boost in utility at time t coming from purchasing the same brand as time $t-1$, while "habit persistence" is modeled as the serial correlation between consequent choices

only depends on realized past choices, "habit persistence" takes into account how prior propensities to choose a brand affect current choices (Heckman 1981). So if the household has high evaluation of brand j in trip t but purchases brand i, the high evaluation of brand j would persist in trip $t+1$ even though it was not purchased. Note that this might be interpreted as a reversal of terminology from the psychological account described above, which centrally features (multiple) prior choices (not just positive evaluations).

Moreover, Seetharaman (2004) builds on the work of Roy *et al.* (1996) and allows for more complicated forms of "habit persistence" in a utility-based framework. Seetharaman (2004) defines habit persistence type 1 as "serially correlated error terms in the random utility function". This form of habit accounts for persistence in choices for reasons unknown to the researcher such as long holidays or having guests which might require successive purchases of the same brands. Habit persistence type 2 is then defined as "serial correlations between utility-maximizing alternatives on successive purchase occasions of a household", which accounts for temporal dependencies in successive brand choices due to unobserved information signals such as billboards or television advertisements. However, the link between the notion of habits in these papers and the psychological view as cue–response associations in memory is not clear. In particular, these models capture persistence around specific brands and implicitly assume the shopping environment to be fixed. As a result, they are silent on any changes in choices if the decision is being made in a quite different context with the same choice set.

While developing considered preferences for specific brands can contribute to overall inertia, so can consequences of repeated choices in a stable context. In settings from voting (Cantoni and Pons 2022) to food consumption (Privitera and Zuraikat 2014), empirical researchers have argued that context effects have a substantial impact on individuals choices (Amir and Levav 2008). The importance of contextual cues in triggering habits, despite conflict with current goals, has been extensively studied by social psychologists (Neal *et al.* 2011). However, there is no clear link between psychological measures of habits and the state-dependence literature. For the most part, these two areas have evolved separately. Although Tam *et al.* (2014) discuss the conceptual distinction between brand loyalty and habits, they do not provide any empirical evidence to demonstrate this difference. We provide new empirical evidence consistent with a substantial role for habits in state dependence in consumer purchases.

in a Markov process that can be present even in the absence of structural state dependence. Similar to many other paper in this literature, Roy *et al.* (1996) impose a Markov assumption and consider the immediate past purchase as a proxy for past behavior. However, this modeling choice is questionable regarding modeling habits, which typically form and change slowly overtime.

1.2. Overview

This paper studies in-store purchase behavior and uses store closures as a shock to the context in which consumers make these purchases. Here we articulate how this approach contributes to our understanding of the role of habits in repeated purchases.

We need to distinguish between habits — defined as an association in memory between purchase responses and contextual cues — and both any complementarity in repeated consumption and other learned preferences for particular products and brands. That is, in addition to distinguishing inertia from consumer heterogeneity (Pakes *et al.* 2021), in order to attribute inertia to habits, we should also distinguish it from other processes. Repeated consumption of the same brand may reflect that consumers have developed expectations or preferences such that they get higher expected utility for that item fixing other features of the purchase occasion such as store context. On the other hand, customers can form *shopping habits*, by repeatedly purchasing the same brands in the same store context. Shopping habits could exist irrespective of the nature of the purchased product. They could be formed to avoid search costs (Dong *et al.* 2020) or as a form of decision heuristic to free up mental resources for more important tasks (Macrae *et al.* 1994).

In this paper, we aim to identify and measure how shopping habits can affect consumers' in-store decisions. To this end, we leverage store closures as a shock that disrupts part of the households' shopping behavior. The key idea is that each store closure can potentially force households to explore new store environments, where previous contextual cues are no longer present and consumers are engaged in a more thoughtful and deliberative decision-making process — driving them to explore other options that are normally ignored in a familiar store. This research design, in which a severe change of context is used to measure the strength of habits, is also related to the habit discontinuity hypothesis (Verplanken *et al.* 2008, Verplanken and Wood 2006). The idea is that when a habit is blocked or suspended due to a change of context, the person may need to search for information or advice, and be open to alternative options. Some examples of change of circumstances that can disrupt people's habits include: transitions from school to work (Busch-Geertsema and Lanzendorf 2017), residential relocation (Clark *et al.* 2016), changes in retail contexts (Poortinga *et al.* 2013, Figueroa *et al.* 2019), and lifestyle changes due to COVID-19 restrictions (Oblander and McCarthy 2022).

We argue that the use of the store closures increases the credibility of causal inference about habits. One might alternatively consider any two adjacent trips by the same household. Along these lines, Thomadsen (2016) finds some evidence that consumers exhibit higher levels of state dependence if the store they are shopping from is the same store they visited last time. However, the choice of the store could potentially confound the choice of brands, i.e., people might have chosen to go to a different store in the first place in order to buy a different brand. As a result, one cannot simply consider changes in purchase locations because the choice of the store could be correlated with brand choices. In our framework, the closure induces a relative increase in visits to new stores or at least newly shopping for a particular product category in a store. The set of exposed households is not all impacted equally by store closures. We posit each household's purchase behavior is primarily affected for the subset of categories which they used to buy from the closing store and the intensity of the effect increases with the frequency of visits.[4] From this perspective, the set of household–category pairs for which the household had never purchased that product category in the closing store can be considered as the control group for a difference-in-differences (DID) causal identification strategy. Hence, our identification strategy is based on a combination of different households being exposed to store closures at different times, as well as within-household variation in how much that household is exposed to a closure for a particular product category.

Modeling brand choices involves a complicated multi-choice problem with varying choice sets. One conventional modeling approach in the literature considers various forms of latent utility choice models. However, these models are readily applicable only to a single product category, and researchers typically limit their analysis to a few frequently purchased brands and ignore changes in the households' choice sets. Instead, we conduct a "reduced-form" analysis where the outcome is a binary variable indicating whether the household is choosing their most frequent brand option (i.e. modal brand), following the approach in Larcom, Rauch, and Willems (2017). This simplification will allow us to detect changes in purchase patterns after the closures by simultaneously modeling households' purchases across multiple product categories, use regression machinery to estimate causal effects, and avoid the difficulties of estimating a model with a high-dimensional categorical outcome. Following this framework, we first show that the subset of households with a higher proxy for antecedent habits (i.e. higher frequency of visiting the closing store) experience a temporary disruption in shopping habits right after the closure. They then apparently form new habits over time in the newly visited stores. This observation is robust to accounting for unavailability of their favorite brands. Furthermore, the induced temporary disruption in shopping habits results in lasting changes in households' *modal* brands (brands most often purchased) suggesting that formation of shopping habits.

These results augment our understanding of the state dependant consumers purchase behavior by demonstrating the importance of shopping habits, in addition to pure brand loyalty. Our findings

[4] Frequency and stable context does not measure all aspects of habitual behaviors. In particular, studies show that simpler tasks could become habitual faster than more complex tasks, despite similar frequency in a stable context (Verplanken 2006). We do not measure complexity of purchase decisions. But if they are somewhat similar in complexity, then frequency in a stable context can be a good proxy to properly capture the strength of shopping habits.

have immediate implications for firms who could benefit from understanding (or discovery) of these shopping habits by incentivizing stores to keep the placement of their brands consistent inside the store. However, depending on the brands for which habits are formed (for any specific store), competing firms could have conflicting interests regarding keeping the product placements constant. For a less popular brand, the firm has an incentive to pay the store to change their product placement in order to disrupt existing habits. This would be most effective if it complements other marketing strategies such as providing free samples or different forms of advertisement.

The rest of the paper is structured as follows: Section 2 describes the Nielsen scanner data and explains how closing stores and corresponding exposed households were identified. Section 3 discusses the problem formulation, in particular, how treatment exposure level and outcome variables used in the regression model are defined. Section 4 provides the results for various two-way fixed effects (TWFE) and event study models using different outcome measures. It also presents results for a Bacon decomposition analysis (Goodman-Bacon 2021) to explore any bias in TWFE estimation due to differential treatment timing. Section 5 concludes and discusses potential implications of our findings.

2. Data

We use Nielsen retail scanner and consumer panel data, containing detailed shopping information for more than 50,000 American households and 35,000 stores across the US between January 2006 and December 2018. We utilized the retail scanner data to identify closing stores, and the consumer panel data to detect changes in households' purchase decisions after one of their local stores closes.

2.1. Retail Scanner Data

The retail scanner data contains weekly pricing, volume, and store merchandising conditions generated by retail store point-of-sale systems. The data includes approximately 35,000 stores including grocery, drug, and mass merchandiser stores. The data is available from January 2004, but we only used from 2006 onward since we only needed the store closures relevant to panelists in the consumer panel data. All stores have unique anonymized identifiers, so we could track the sales of each store even if the retail chain changes, although more than 96% of the identified closing stores in the data operate under a single retailer.

2.2. Consumer Panel Data

The Consumer Panel Data represents a longitudinal panel of approximately 40,000–60,000 US households who use hand-held scanner devices to continually provide information to Nielsen about their purchases. Products include all Nielsen-tracked categories of food and non-food items, across all retail outlets in all US markets. Nielsen samples all states and major markets so panelists are geographically dispersed and demographically balanced. Importantly, the consumer panel data can be linked to the retail scanner data using unique store identifiers. Since we need to follow households' purchased brands, we do not use the "Magnet" data which includes non-barcoded products such as fresh fruit.

2.3. Store closures

In order to identify closing stores, we compute the aggregate store weekly sales using retail scanner data.[5] Then, we single out the stores whose sales drop to zero at a certain time and remain zero afterward. We found 7,847 such permanent store closures during a 13-year period starting in 2006. We also investigated potential temporary store closures. Considering stores whose sales drops to zero and remains zero for at least the duration of a year, we found only 83 such cases. Varying the required zero sales duration would change the number slightly, but overall, there were very few temporary closures. Furthermore, there was no instance of multiple closures for any store in the data set. As a result, given the small number of temporary closures, we decided to drop them altogether and only consider permanent closures to avoid further complications in the causal analysis.

We expected that some of these stores simply stopped participating in the panel, while remaining open. To exclude such false closure identifications, we used the consumer panel data and ruled out any store for which there was a reported purchase trip after the closure date. After this correction, 3,243 closing stores remained. We could also identify a false closure if a certain store and all the related customers in the panel opt-out of reporting to Nielsen simultaneously. Even though there is a low chance of this incident happening, it will not affect our results because we only consider households who are active both before and after their corresponding closure to measure the change in their behavior.[6]

Figure 1 shows log-weekly sales for the closing stores corresponding to the set of exposed households. Weekly sales by retailer, the distribution of closing stores over years, and their geographical dispersion are shown in Appendix 6.1.

2.4. Exposed households

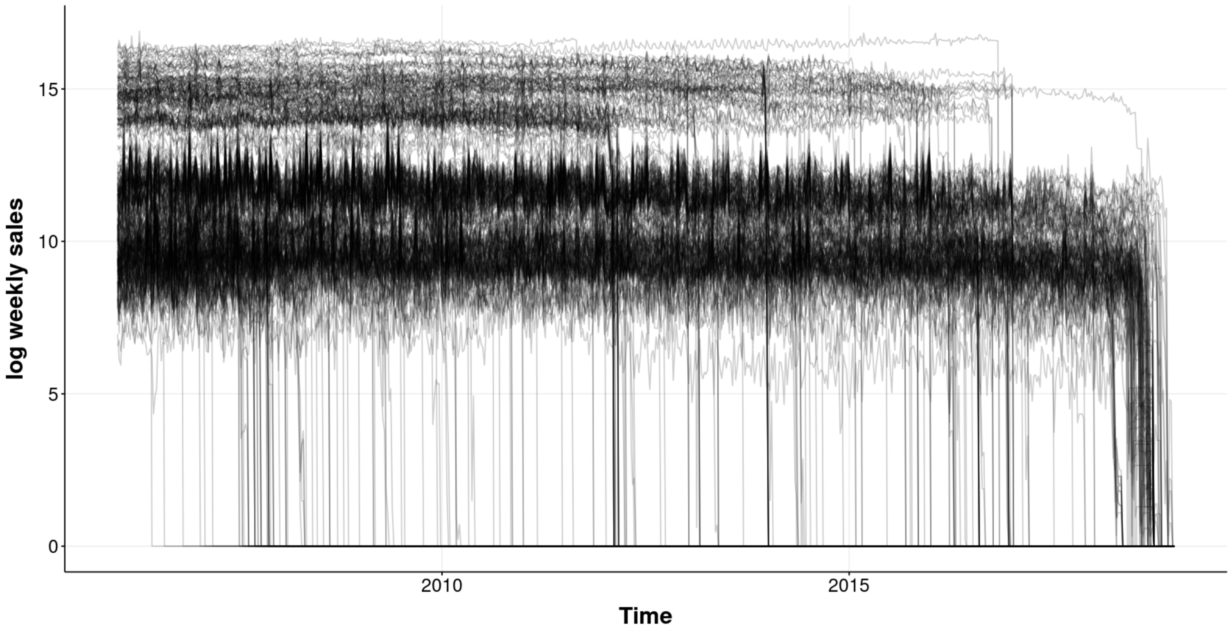
After matching the set of closing stores with the consumer panel data, we find 14,406 households who at some point in time visited one of the closing stores.[7] However, not all of these households

[5] For computational simplicity we only compute the overall weekly sales of each store for top-5 purchased product categories: Refrigerated milk, refrigerated yogurt, fresh bakery bread, cereal, and canned soup.

[6] Finally, some households might stop going to a store around the time it drops out of the retail data, and therefore cause a false closure identification. Although this could potentially happen, it is less likely to happen for households who are frequently shopping from the closing store. And as we see in the following results, the main effect is driven by these more frequent visitors.

[7] In the final analysis, we only consider top-30 product categories and remove infrequent ones. This leaves us with 14,360 households.

Figure 1 Log weekly sales of the closing stores



Note. The figure shows log-weekly sales in the top-5 product categories by national purchase volume for the closing stores corresponding to the set of exposed households from January 2006 until December 2018.

were *exposed* to the exogenous closure shocks. Some of them might have visited the closing store months or years before it closed. We consider only households who were still shopping from the closing store near its closing time to be *exposed* by the closure. Therefore, we marked households who had at least a shopping trip to the closing store within a 4-month interval prior to the corresponding closure date as the *exposed* set, which included 684 households with a total of 407,630 distinct shopping trips.

3. Framework and Definitions

In this section, first, we specify the treatment each household is receiving due to the store closures, and then define two distinct outcome variables which are used in our models.

3.1. Treatment exposure

The set of exposed households are not all equally affected by a store closure. First, we posit each household's purchase behavior is primarily affected within the subset of categories that they used to buy from the closing store. For example, if someone frequently bought yogurt but not cereal from the closing store , we expect the closure affects their yogurt purchase behavior, with effects on cereal, if any, being much smaller. Second, we expect the effect to vary based on the prevalence of the shopping trips to the closing store. In order to capture both of these dimensions, we define the *treatment exposure level* ($e_{i,c}$) for household i and category[8] c as the relative fraction of household i shopping trips to the closing store in which a product in category c was purchased:

$$e_{i,c}(T_e) = \frac{\text{trips to the closing store by household } i \text{, purchasing category } c \text{, } T_e \text{ years before closure}}{\text{trips by household } i \text{, purchasing category } c \text{, } T_e \text{ years before closure}}. \tag{1}$$

The exposure level is a function of the pre-closure time period on which it is defined, T_e. The shorter we define this period, the better the fraction would capture the true impact of the closure because someone could have a many shopping trips to the closing store many months before the closure but only a few such trips right before the closure. This household would be less likely affected by the closure. However, at the limit of $T_e \to 0$, we have zero observations to define the fraction, and with small values of T_e, measured exposure would be sensitive to a small number of recent trips. Since it is not obvious ex-ante what would be the optimal time period, we do the analysis for a range of values and show that the main results are robust to the choice of the T_e parameter. Results presented in the main text are for $T_e = 1$ year, and result for $T_e = 2$ and $T_e = \frac{1}{2}$ year are presented in Appendix 6.3.

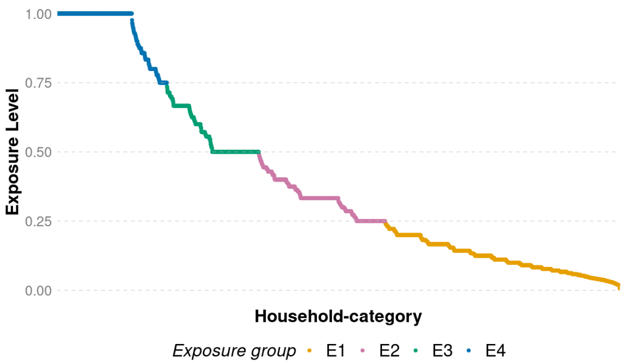
Figure 2(a) displays the treatment exposure levels for all household–category pairs. There are a total of 9,338 units (not shown) with zero exposure level; these household–category pairs can be used as the set of control units in a difference-in-differences framework with differential treatment timing.

Furthermore, to flexibly allow for potentially heterogeneous effects among household–category pairs, we partition treated pairs into four groups based on their exposure levels: $E_1 = \{(i,c)|e_{i,c} \leq 0.25\}$, $E_2 = \{(i,c)|0.25 < e_{i,c} \leq 0.5\}$, $E_3 = \{(i,c)|0.5 < e_{i,c} \leq 0.75\}$, $E_4 = \{(i,c)|0.75 < e_{i,c}\}$, and the control group is defined as $C = \{(i,c)|e_{i,c} = 0\}$. These groups are separated by dotted grey lines in Figure 2(a) where each point stands for a household–category pair. There are 1,636, 881, 641, 768 household–category pairs in E_1–E_4 correspondingly.

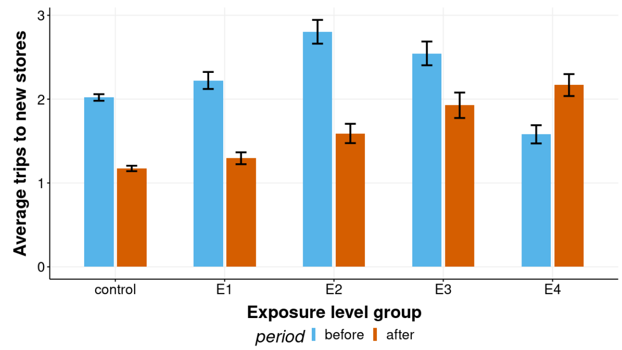
We expect household–category pairs with higher treatment exposure level to have more significant disruption in their purchasing behavior for two main reasons. First, according to psychological theory, more frequent trips make stronger shopping habits, and hence the disruption in brand choices could be more substantial (Wood *et al.* 2005). Second, the main channel through which the closure is affecting households purchasing behavior is the resulting forced exploration in visits to

[8] Nielsen has a 3 level hierarchy for categorizing different products. There are 10 Departments, 125 product groups, and about 1100 product modules. For example, within the *Frozen foods* department, there are multiple groups including *frozen vegetables* or *frozen breakfast foods*. And within each group there could be multiple modules such as *frozen beans*, *frozen toaster items*, etc. Throughout the paper by product category we mean the grouping at the product module level.

Figure 2 (a) Treatment exposure levels, (b) and the average number of trips to new store–category pairs



(a)



(b)

Note. (a) Treatment exposure levels ($e_{i,c}$) for household–category pairs sorted in decreasing order, where each point shows a unique pair. Control group ($e_{i,c} = 0$) pairs are not shown in the figure. Dashed grey lines show how household–category pairs are categorized into four treatment exposure level groups. There are a total of 9,338 pairs in the control group, and 1,636, 881, 641, 768 pairs in E_1-E_4 correspondingly.

(b) The average number of new store–category pairs visited by each exposure group during an L trips before and after the corresponding closures. A purchase occasion at store s in category c is counted as *new store-category* visit for household–category (i, c) if household i purchases category c in store s during her first L trips after her corresponding closure date, while she was not purchasing any items in category c from store s during L trips prior to closure. Error bars show the 95% confidence intervals. The highest exposed group E_4 has relatively more new visits after the closure, which makes it more likely to observe a significant disruption effect in their purchasing behavior caused by the closure.

new store–category pairs[9] where old habitual cues are no longer present and people are susceptible to exploration and formation of new habits. As you can see in Figure 2(b), the average number of visits to new store–category pairs during the post-closure period increases with exposure level. A purchase occasion at store s is counted as *new store* visit for household–category (i, c) if household i purchases category c in store s during her first L trips after her corresponding closure date, while she had not purchased any items in category c from store s during L trips prior to closure. We compare these average new visits with the same quantity defined based on the period of L trips prior to each closure, while here a visit is counted as new if the customer had not purchased any item between $2L$ and L trips prior to closure. The comparison shows that relatively the new store visits increases only for E_4. This is another reason to expect effects of closures to be concentrated in the fourth exposure group.

Finally, the timing of closures and whether a household is exposed to a closure at a given time is determined by households' own choices and hence is potentially endogenous. This motivates using

[9] A purchase occasion at store s in category c is counted as *new store-category* visit for household–category (i, c) if household i purchases category c in store s during her first L trips after her corresponding closure date, while she was not purchasing any items in category c from store s during L trips prior to closure.

differences-in-differences, whereby before–after closure changes for household–category pairs are compared with those changes in the control group of household–category pairs, which consists of households that did not purchase that category at the closing store. We will return to this issue later in Section 4.1.3 and show evidence consistent with parallel pre-treatment trends comparing each of the exposure groups with the control group.

3.2. Outcome Variables

Our goal is to detect changes in households' brand choice patterns that are indicative of habits and closure-induced search. This is a complicated multi-choice problem with varying choice sets. In order to simplify the problem and provide interpretable estimates, we follow the approach in Larcom, Rauch, and Willems (2017), and conduct a relatively "reduced-form" analysis where the outcome is a binary variable indicating whether the household is choosing their modal brand option. This simplification has multiple advantages. First, it allows us to use regression machinery to estimate causal effects and avoid the difficulties of estimating a model with a high-dimensional categorical outcome. Second, it allows us to readily pool information across different product categories, hence giving a more comprehensive view of shopping behavior. We separately define recent and baseline modal brands to capture different aspects of the changes in households' behavior.

3.2.1. Recent modal brand We expect the effect of the habit disruption on purchase patterns to be, in some sense, temporary because customers will soon form new shopping habits in the new store environments they visit. Therefore, to measure the temporary effect of habit discontinuity on purchase decisions, we define the recent modal brand using a moving window, based on households' L-most-recent shopping trips for each category.[10] More precisely, let $b_{i,t,c}$ be the brand purchase by household i at trip t in product category c. The *L-recent modal brand* for the triple (i,t,c) is defined as $\tilde{b}^r_{i,t,c} = \mod (b_{i,\tau,c})\big|^{t}_{\tau_c(L)}$, where $\tau_c(L)$ specifies L previous trips in which category c was purchased. Since different categories are purchased with different frequencies, the time duration in which modal brand is computed would be different for each category. Although time duration is not entirely irrelevant, habits are understood as persistent over time (Wood and Neal 2016). Therefore, we decided to define the modal brand based on the number of visits to each store because what matters most is the repetition and frequency of purchase behavior.[11]

[10] Results presented in the main text use $L = 20$, but they are robust to variations in L. More details can be found in Appendix 6.3

[11] This choice was primarily based on the frequency-in-context measure of habits (Labrecque and Wood 2015). Since people could have very different rates for visiting stores, considering same time frames could result in very different number of trips. In particular, habits are resilient to the passage of time and could be triggered even with the loss of memory (Bayley *et al.* 2005, Knowlton *et al.* 1996). As a result, we concluded the number of repetitions could matter more than the frequency over time. Although one could imagine the time passed between shopping trips could also play a role, this data set did not provide us with enough variation to study both of these phenomena simultaneously. This is an interesting research question that can be studied using carefully designed experiments.

As our first outcome variable, we define the recent modal brand indicator $y^r_{i,t,c}$ as a binary variable indicating buying the recent modal brand $\tilde{b}^r_{i,t,c}$, where i specifies the household, t the trip number, and c the corresponding category of the purchased product:

$$y^r_{i,t,c} = \begin{cases} 1, & b_{i,t,c} = \tilde{b}^r_{i,t,c} \\ 0, & b_{i,t,c} \neq \tilde{b}^r_{i,t,c} \end{cases}. \tag{2}$$

We hypothesize that when shopping habits are disrupted, households purchase decisions deviate more often from their modal options (so all the effects are expected to be negative) because the old contextual cues that used to trigger the behavior are no longer present. However, we expect the effect to be only temporary because after a while the recent modal brand is defined based on post-closure trips. Some deviations could also be caused by the unavailability of a household's modal brand in stores they visit after the closure; we return to this issue by conducting analyses that condition on brand availability.

3.2.2. Baseline modal brand Another interesting question to explore here is whether this disruption causes a lasting change in households' brand choices, or after doing some exploration they would return to their prior modal brands. To answer this question, we use all trips prior to closures to specify households' baseline modal brand and then measure deviations from that after the closures.[12] For each household–category pair, the *baseline modal brand* $\tilde{b}^b_{i,c}$ is defined as the most frequently purchased brand during all trips before the household i's corresponding store closure date τ_i.[13] Note that the long-term modal brand is fixed for each household–category and independent of trip number, unlike the recent modal brand which is defined on a rolling basis. Similar to Equation 2, we define the long-term modal brand indicator $y^b_{i,t,c}$ as a binary variable indicating whether household i is buying her baseline modal brand in category c during trip t:

$$\tilde{b}^b_{i,c} = \quad \bmod \left(b_{i,t,c}\right)\Big|_{t=-\infty}^{\tau_i}, \quad y^b_{i,t,c} = \begin{cases} 1, & b_{i,t,c} = \tilde{b}^b_{i,c} \\ 0, & b_{i,t,c} \neq \tilde{b}^b_{i,c} \end{cases}. \tag{3}$$

3.3. Variation in treatment timing

Our identification strategy is based on a combination of different households being exposed to store closures at different times, as well as within-household variation in how much that household is exposed to a closure for a particular product category. Until recently, two-way fixed effects (TWFE)

[12] Since the panel is not balanced, the number of pre-closure trips could be highly variable for different household–category pairs. The full panel, as well as the histogram of pre-closure trips can be found in Appendix 6.1. So in order to make analysis comparable across different units, we check the robustness of results using a fixed length of 40 trips to define the baseline modal brand; note that the average number of pre-closure trips is 42. All of the results are qualitatively the same as you can see in Appendix 6.4.

[13] Households exposed to the same store closure might stop visiting the store on slightly different dates. So in practice, we set the corresponding closure date to be the last visit by that household.

estimators would be the standard method for estimating treatment effects in such difference-in-differences (DID) settings. However, with variation in treatment timing (i.e. differential timing, staggered adoption), the estimated coefficient is more difficult to interpret and generally does not equal the average treatment effect (ATE), or the ATE on the treated; and need not be any weighted ATE either. Recent work has addressed this issue (Sun and Abraham 2021, Imai and Kim 2019, Goodman-Bacon 2021, Callaway and Sant'Anna 2020), including highlighting that, in some cases, this can make resulting estimates quite biased if interpreted as treatment effects (Baker *et al.* 2022).

Here we use TWFE as our primary estimator, but show that this choice is not so consequential. In particular, we use a decomposition (Goodman-Bacon 2021) of the estimates into a weighted average of individual 2×2 DID estimators with the weights proportional to group sizes and variance of treatment duration. The decomposition shows that the TWFE estimator consists of three comparisons and gives the corresponding weights for each: treated vs. untreated, lately-treated vs. early-treated, and early-treated vs. lately-treated.

4. Results

In this section, we present the results for many difference-in-differences and event study models using the outcomes defined in the previous section. We do the following analyses for each of the outcome variables discussed in the previous section:

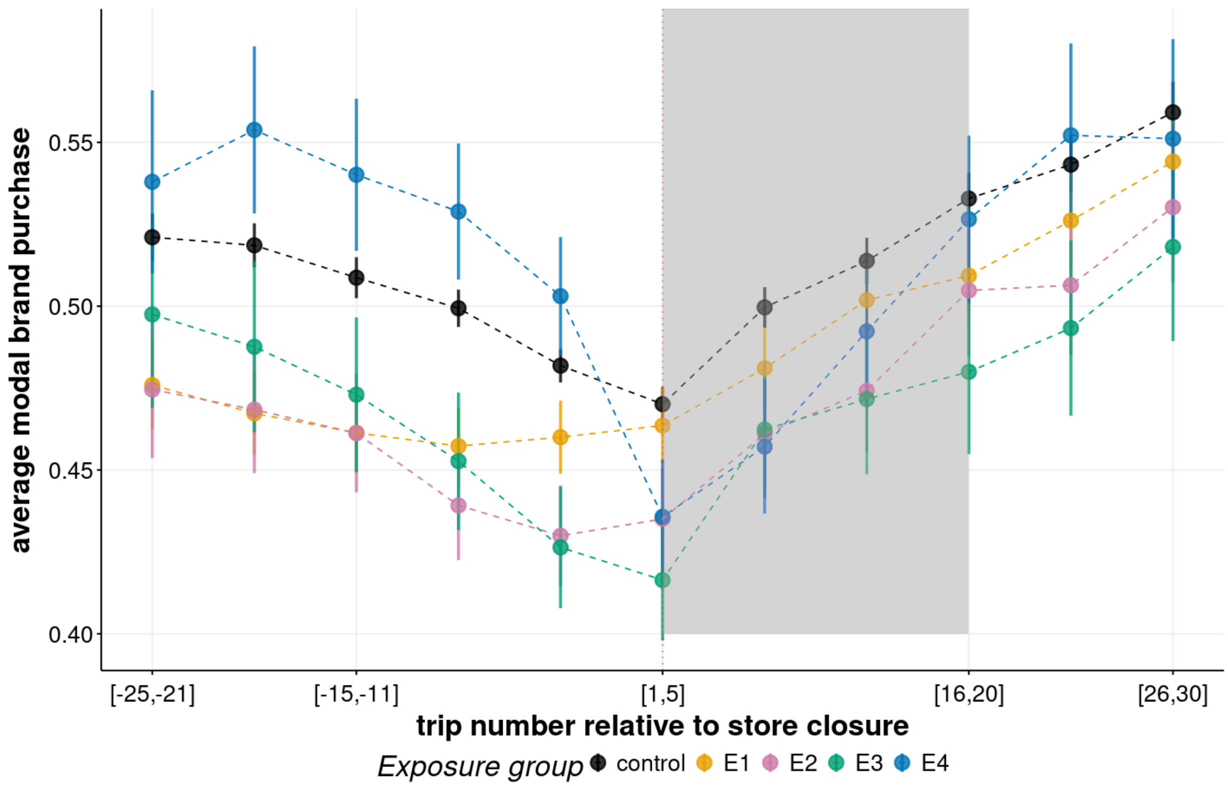
1. Estimate a TWFE model to find the aggregated treatment effect.
2. Estimate a heterogeneous fixed effects model to explore possible heterogeneity in effects across household–category pairs in different exposure-level groups.
3. Estimate an event study model to examine the testable implication of the parallel trends assumption, and also examine how treatment effects change over time.
4. Estimate a conditional TWFE model to measure to what extent the post-closure unavailability of brands is driving the treatment effects.
5. Use the Goodman-Bacon (2021) decomposition to find the extent of the bias in TWFE estimators.

4.1. Temporary disruption in shopping habits

Here we use the recent modal brand indicator to measure the short-term effect of store closures on households' purchase behavior.

4.1.1. Descriptive analysis First, we examine the dynamics of modal brand choices by plotting the average recent modal brand purchase rates across all exposure groups. For each household, we consider the trip number relative to the corresponding store closure date and compute the average recent modal brand choice for blocks of 5 trips (Figure 3). There is a substantial drop in modal

Figure 3 Dynamics of the average recent modal brand purchases



Note. Dynamics of recent modal brand purchase rate for every 5 trips relative to closure date. The grey area shows the duration in which we are considering households to be treated. The subplot (a) compares all treated household–category pairs with the control group, and subplot (b) shows the averages separated by exposure level groups.

brand purchase rate for E_4 relative to the control group shortly after the closures happen, which shrinks over subsequent trips. This observation is consistent with our hypothesis that purchase behavior of higher exposed household–category pairs is more strongly affected. Moreover, the figure shows approximately parallel pre-trends between different exposure groups and the control group (maybe except for E_1). This analyses does not yet account for household–category or seasonal patterns in the panel data.

4.1.2. Difference-in-differences Here we use TWFE to estimate the effects of closures on modal brand purchase rate. Further, as explained previously, we separately estimate the effect for short-term and long-term treatment variables. The short-term treatment variable is active only for the first L trips after the closure for each category so that it can capture the temporary effect shortly after the closure. Moreover, the treatment intensity is equal to the corresponding household–category exposure level. As a result, the treatment vectors for household i in trip t and category c are defined as:

Table 1 DID results for recent modal brand.

	Dependent variable:			
	recent modal brand indicator (×100)			
	(1)	(2)	(3)	(4)
Overall, short-term	-4.616^{**}		-2.732^{***}	
	(1.514)		(0.714)	
Overall, long-term	0.128		-1.943^{*}	
	(2.210)		(0.971)	
E_1, short-term		0.849		-0.111
		(0.632)		(0.507)
E_2, short-term		1.072		-0.495
		(0.878)		(0.704)
E_3, short-term		-2.274		-2.117^{**}
		(1.397)		(0.772)
E_4, short-term		-6.350^{***}		-2.687^{***}
		(1.493)		(0.738)
E_1, long-term		1.634^{*}		-0.201
		(0.769)		(0.504)
E_2, long-term		3.629^{**}		0.153
		(1.361)		(0.960)
E_3, long-term		0.983		-0.879
		(2.176)		(1.059)
E_4, long-term		-3.686^{*}		-2.389^{**}
		(1.809)		(0.854)
Continuous treatment	✓		✓	
Conditioned on modal brand availability			✓	✓
Observations	887,544	887,544	535,313	535,313
R^2	0.279	0.279	0.149	0.149
Adjusted R^2	0.268	0.268	0.129	0.129
Note:				*p<0.05; **p<0.01; ***p<0.001

Estimation results for TWFE models with the recent modal brand indicator as the outcome variable. Columns 1 & 2 show the corresponding β parameter(s) in Equations 5, 6. These coefficients measure short-term and long-term rates of recent modal brand purchases, compared with control household–category pairs. Columns 3 & 4 contain the same parameters conditional on trips in which the recent modal brand was available. All standard errors are clustered at the closing store level, and all numbers are multiplied by 100.

$$T^{r_1}_{i,t,c} = \begin{cases} e_{i,c}, & \tau_i \le t \le \tau_i + L \\ 0, & t < \tau_i,\ t > \tau_i + L \end{cases},\ T^{r_2}_{i,t,c} = \begin{cases} e_{i,c}, & t > \tau_i + L \\ 0, & t \le \tau_i + L \end{cases}. \tag{4}$$

Using the recent modal brand outcome (Equation 2) and these treatment vectors, the fixed effects regression model can be formulated as:

$$y^r_{i,c,t} = \alpha_{i,c} + \gamma_t + X^T_{i,t}\theta + \beta_1 T^{r_1}_{i,c,t} + \beta_2 T^{r_2}_{i,c,t} + \epsilon_{i,c,t}, \tag{5}$$

where $\alpha_{i,c}$ are household–category fixed effects, and γ_t are the temporal (monthly) fixed effects. $X_{i,t}$ is a set of household covariates that are varying over time and could potentially affect shopping behavior; these include dummy variables for household income level, size, and composition.[14]

Furthermore, we estimate a heterogeneous TWFE model to capture the treatment effect for each exposure level group. Note that, unlike the model in Equation 5 in which the treatment was defined proportional to the exposure level, in the following model the treatment is a binary indicator for the household–category pair belonging to each exposure group,

$$y^r_{i,c,t} = \alpha_{i,c} + \gamma_t + X^T_{i,t}\theta + \sum_{j=1}^{4} \beta_{1,j}\, \mathbb{1}_{i\in E_j}\, \mathbb{1}_{T^{r1}_{i,c,t}>0} + \sum_{j=1}^{4} \beta_{2,j}\, \mathbb{1}_{i\in E_j}\, \mathbb{1}_{T^{r2}_{i,c,t}>0} + \epsilon_{i,c,t}. \tag{6}$$

Table 1 summarizes the effects ($\hat{\beta}$s) estimated in Equations 5 and 6. All of the standard errors are clustered at the closing store level, thereby allowing for dependence among all household–category pairs associated with the same store closure. The estimated overall short-term effect (column 1) is negative and statistically significant, consistent with our hypothesis about the disruption caused by store closures. Furthermore, the estimates for different exposure groups (column 2) show that the effect magnitude is increasing for higher exposed household–category pairs and primarily driven by E_4, consistent with our expectation that household–category pairs in which the household frequently purchased from the closing store are most affected. There is no statistically significant long-term effect detected. Notice that there is an opposite significant short-term effect for E_1 and E_2. This is probably because the closing store did not play a major role in their modal brand purchases, but may have been causing deviations from their modal brands. So the closure makes them visit other stores that they already visited more often, which results in an increase in their modal brand purchase rates.

4.1.3. Event study analysis A common robustness check for the TWFE model is an *event study analysis* (Granger 1969, Sun and Abraham 2021, Roth 2019). To this end, we estimate a model similar to Equation 5, but with lags and leads of the treatment variable:

$$y^r_{i,c,t} = \alpha_{i,c} + \gamma_t + X^T_{i,t}\theta + \sum_{\tau=1}^{4} \beta^{lead}_{\tau}\, T^r_{i,c,t-\tau} + \sum_{\tau=0}^{5} \beta^{lag}_{\tau}\, T^r_{i,c,t+\tau} + \epsilon_{i,c,t}. \tag{7}$$

If the lead estimates were statistically significant, it would be a violation of the parallel trends assumption because it would imply the cause is preceding the effect. Moreover, the estimated lagged effects are informative about the dynamics of post-treatment treatment effects. Similar to the analysis in Section 4.1.1, we estimate the leads and lags for every 5 trips grouped together. We

[14] Note that these covariates only include yearly changes and the Nielsen data does not provide more accurate temporal information on panelists.

added five leads and six lags, where the first and last include all trips whose relative trip number is less than -20 and more than 25. The fifth lead variable is the omitted baseline category in estimating the model.

The estimates for the event study model (Equation 7) with corresponding confidence intervals can be seen in Figure 4 (top). The lead parameters are not rejected at a 95% confidence level, which indicates the control and treatment units are indistinguishable prior to the treatment and is consistent with the parallel trends assumption. Furthermore, the point estimates are significantly negative for the next 15 trips after the closure with a diminishing magnitude. This is exactly what we expected since the recent modal brands are defined on a rolling basis (e.g., the point at $x = 4$, is entirely based on post-closure trips). Also, the last lagged variable which measures the long-term treatment effect is almost zero. Both of these observations support our hypothesis that store closures do not have a lasting effect on the degree to which households' purchases eventually concentrate into a model brand. The rate at which the effect goes to zero also gives us a sense of the required number of visits to new stores for that to happen.[15]

We do a similar event study analysis for heterogeneous effects across exposure groups to validate the parallel trends assumption for Equation 6 and also observe the dynamics of the post-treatment effects,

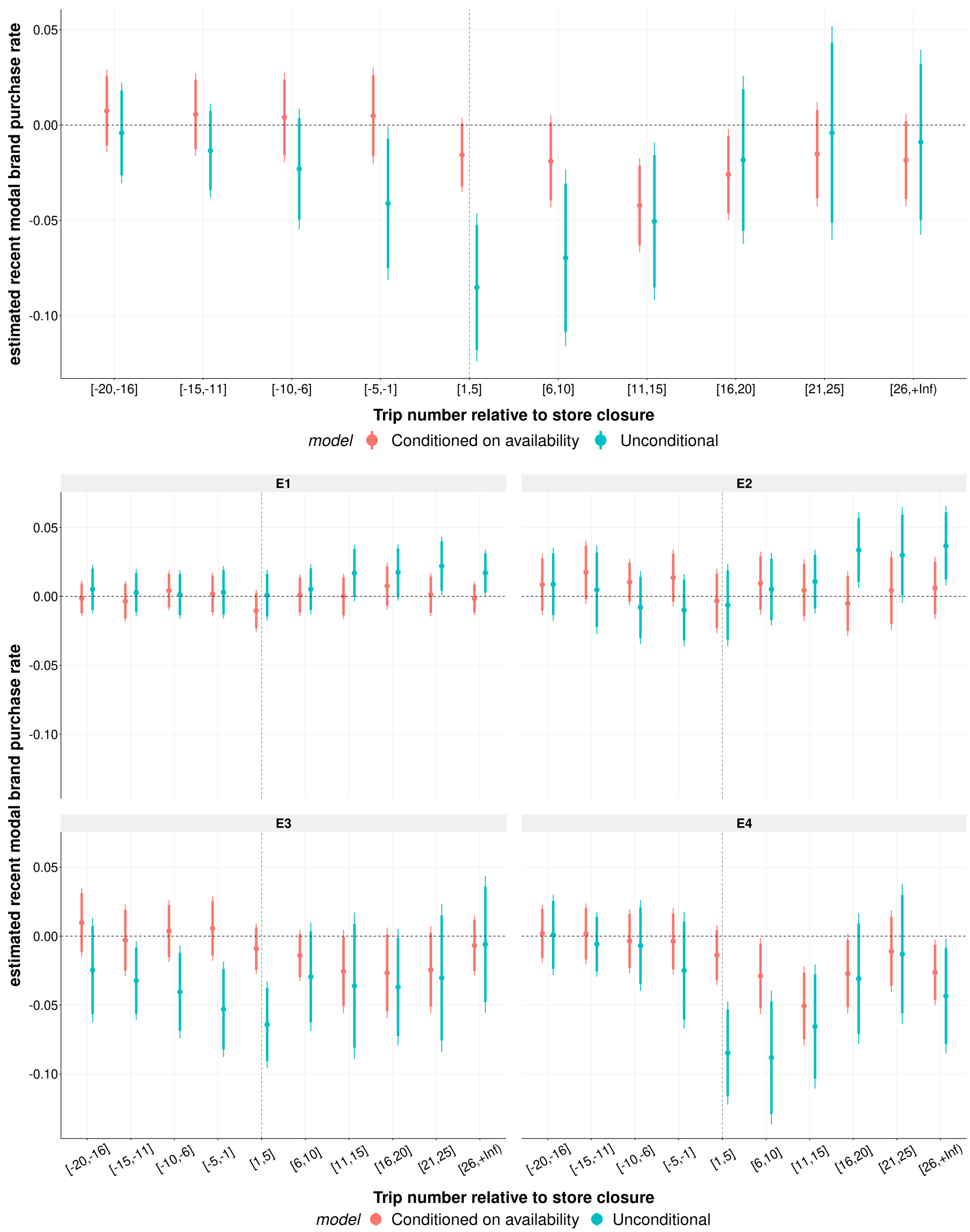
$$y^r_{i,c,t} = \alpha_{i,c} + \gamma_t + X^T_{i,t}\theta + \sum_{j=1}^{4}\sum_{\tau=1}^{4} \beta^{lead}_{j,\tau}\, \mathbb{1}_{i\in E_j}\, \mathbb{1}_{T^r_{i,c,t-\tau}>0} + \sum_{j=1}^{4}\sum_{\tau=0}^{5} \beta^{lag}_{j,\tau}\, \mathbb{1}_{i\in E_j}\, \mathbb{1}_{T^r_{i,c,t+\tau}>0} + \epsilon_{i,c,t}. \quad (8)$$

The results are shown in Figure 4 (bottom). Except for leads coefficients of E_3, the rest of the leads are statistically non-significant, consistent with the parallel trends assumption for the heterogeneous fixed-effects model. The post-treatment trends also have a diminishing magnitude similar to the previous model, while the highest exposed group E_4 has the most significant and lasting effects.

4.1.4. Availability of modal brands One potential source of the change in brand choices could be the lack of availability of prior modal brands in newly explored stores after the closure. In this section, we show that only part of the observed effect can be explained by unavailability of a modal brand on a given trip. Nielsen data does not directly provide information on all of the available brands in each store over time, so we need to infer that from retail scanner and consumer panel data. The list of stores in the retail scanner data does not have a full overlap with stores in the consumer panel data. We therefore used purchases from other households in the consumer panel data to identify available brands. For each week, we mark a modal brand as available in

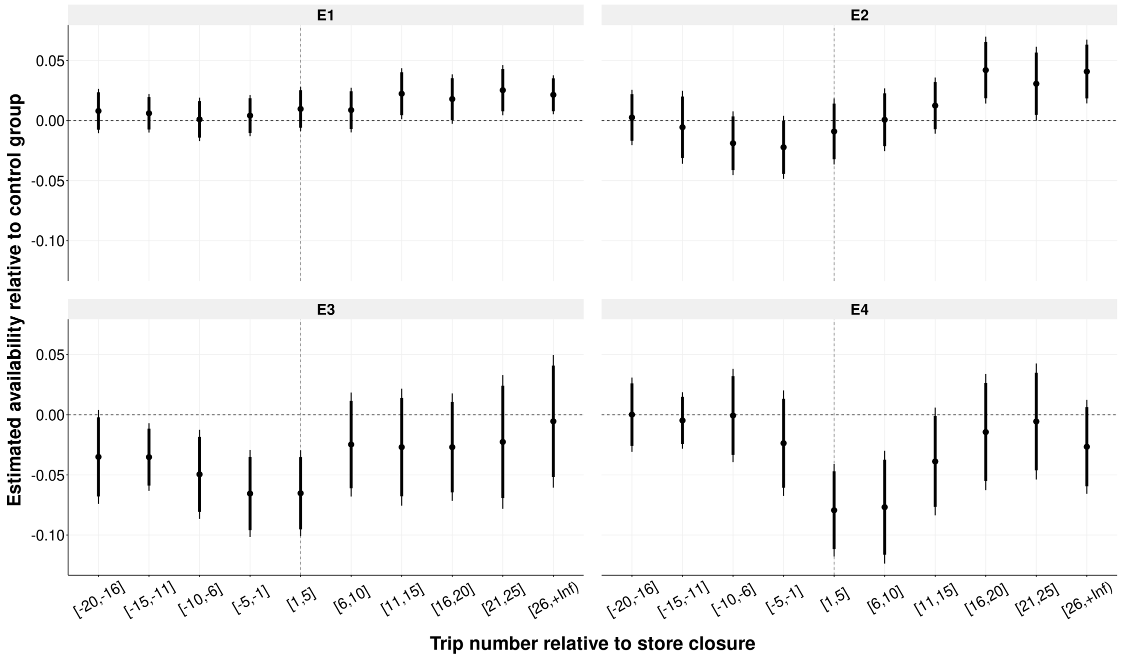
[15] This is more like an upper limit for the required number of trips to form new habits because not of the trips in our data set are at new stores.

Figure 4 **Event study analyses results for aggregate (top), and different exposure groups (bottom) with the recent modal brand indicator as the outcome variable**



Note. The plot shows the estimated lead and lag coefficients in Equations 7 & 8. Point estimates are computed for every 5 trips grouped together, the thin and thick error bars show the corresponding 95% and 90% confidence intervals. The blue color indicates the unconditional model, and red shows estimated coefficients conditional on recent modal brand availability. 5 lead and 6 lag variables are used, where the last lead (lag) includes all trips whose relative trip number is less than (more than) -20 (25). The fifth lead variable is used as the baseline in estimating the model and hence not shown in the figure.

Figure 5 Event study analyses results with availability indicator as the outcome variable



Note. The plot shows the estimated lead and lag coefficients in Equation 7 with availability indicator as the outcome variable. Point estimates are computed for every 5 trips grouped together, the thin and thick error bars show the corresponding 95% and 90% confidence intervals. 5 lead and 6 lag variables are used, where the last lead (lag) includes all trips whose relative trip number is less than (more than) -20 (25). The fifth lead variable is used as the baseline in estimating the model and hence not shown in the figure.

a store, if there is at least one purchase occasion by any household in the entire consumer panel data. In order to compare the availability among different exposure groups relative to control, we estimate event study models similar to Equations 7 and 8 using the availability indicator as the outcome variable.

As it might be expected, the overall percentage of available modal brands drops after the closure, both for the treated and control units. Figure 15 (Appendix 6.3) shows the percentage of trips with available modal brands across exposure level groups and time periods. Nevertheless, these are averages over extended periods of time, and do not account for seasonal variations. To follow how modal brand availability varies in each exposure group over time, relative to the control, we estimate a similar event study model considering availability indicator as the outcome variable. The resulting coefficients can be seen in Figure 5. There are significant negative pre-trends for E_3, which can explain the negative lead coefficients in Figure 4. More importantly, the availability rates significantly drops (up to 8 percentage points) for E_4 right after the closure, which is expected since these household–category pairs are more likely to be purchasing in a new store (Figure 2b).

The decreased availability for E_4 could account for some or all of the significant short-term effect, so in order to be able to attribute the observed effect to disrupted habits we need to adequately account for that. To this end, we estimate similar DID models (Equations 5 and 6) conditional on the subset of trips in which the recent modal brand was available. The idea is that in trips where the households' modal brand in a certain category is available, increased average deviation from that brand would reveal the impact of the habits formed around the store environment.[16]

Results for the conditional model can be seen in Table 1, columns 3 and 4. We also estimate the same event study models (Equations 7 and 8) conditional on modal brand availability. As shown in Figure 4 (red points), the scarcity of modal brands can explain only part of the observed effect, and a substantial effect remains. For example, the short-term effect for E_4 is about a third of the unconditional effect. This remaining part could be attributed to what we called *shopping habits*, which happens due to the absence of previous contextual cues. Notice that in the long-term after the closure period, there is a significant increase in availability for E_1 and E_2. This can explain the positive long-term effect observed in the unconditional model for E_1 and E_2 in Table 1. This effect disappears as we account for brand availability, so there is no behavioral factor causing the effect. For lower exposed household–category pairs, by definition, the closing store plays a smaller role in their shopping. Hence the increased availability is plausibly because of the fact that they are increasingly often visiting the set of other store–category pairs they used to shop from — places where they bought their modal brands before the closure. Moreover, there is a marginal negative long-term effect for E_4, which could suggest there might be a lasting increase in variety-seeing. However, this effect is not robust to different choices of hyper-parameters (L and T_e), so we would not draw any conclusions based on that (see Appendix 6.3 for more details).

4.1.5. Effects without temporary unavailability The previous analysis does not entirely rule out the impact of modal brand unavailability. Temporary unavailability of a certain brand could force the household to explore new ones which can cause increased information about alternative options leading to changed brand preferences. Such a process is consistent with our broad account of how disruptions to choice environments can have lasting consequences by changing habits; however, we wish to also characterize whether some of this is attributable to changes in the store environment along, even if the modal brand remains continually available.

To this end, we now analyze subsets of household–category pairs in which we restrict any unavailability in the sequence of households' trips after the closure. So, for each household–category pair, we estimate the short-term effect for the first L_a trips in which the modal brand was always available, where $0 < L_a \leq L$. We include the remaining trips after L_a among the long-term effect. The

[16] Note that, to keep estimates comparable, we define the short-term/long-term periods the same way using all trips. So the short-term coefficients would contain less than or equal to L trips. We could have alternatively used the first L trips in which the modal brand was available to estimate the short-term effect (which would indeed result in more substantial effect size), but then it would cover a longer period of time and one might worry about other factors affecting households' behavior.

long-term effect would not have a similar clear interpretation since it includes trips with unavailable brands, however, we still include them in the estimation because they help with estimation of fixed effects and hence improve the precision of the desired short-term parameter.

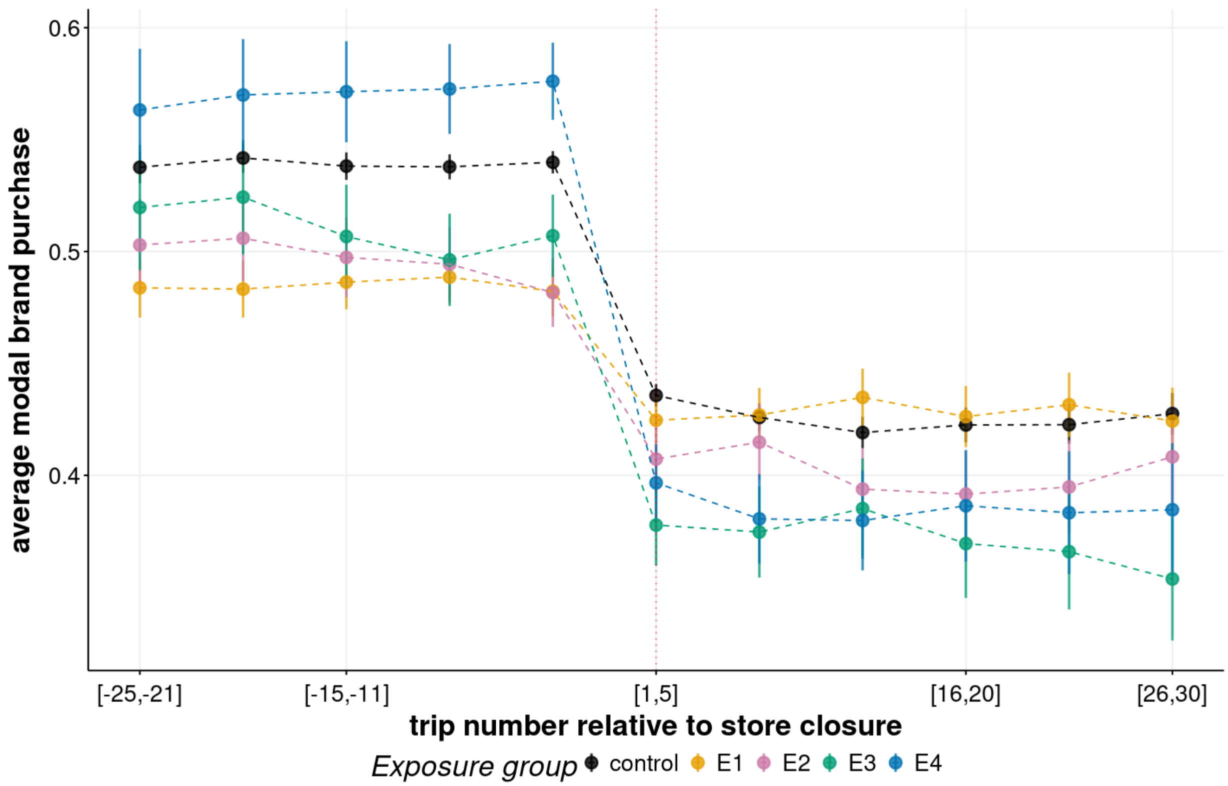
If we take the subset of the data for which $L_a > 0$, we are left with 6,084 household–category pairs and about 10,000 short-term trips; the full distribution of L_a can be seen in Appendix 6.3, Figure 16. Estimating Equation 6 gives the following short-term effect for the forth exposure group: $\hat{\beta}_{1,4} = -0.870$, $SE = 0.748$, $p = 0.245$. The estimate is not statistically significant, mostly because the condition requiring the full sequence of trips to have available modal brands is very restrictive and leaves us with very few observations.[17] For example, there could be many units for which there is only one trip where the modal brand is unavailable, and the information effect of exploration is minimal, but the previous condition would drop the sequence altogether. However, if we slightly loosen the conditions by allowing the sequence of short-term post-closure trips to have at most one trip with unavailable modal brand there are 7,825 household–category pairs with about 15,000 short-term trips remaining, and the estimated coefficient would be: $\hat{\beta}_{1,4} = -1.743$, $SE = 0.784$, $p = 0.027$. Here, we have included post-closure trips up to the point where the second trips with unavailable modal brand appears. This would minimize the effect of unavailable brands on households' information about alternative brands, while leaving enough data to be able to precisely estimate the parameter of interest. These results are broadly consistent with both an unavailability-driven mechanism and other effects of changing contexts for choices, even as modal brands remain largely available.

4.2. The effect of store closures on baseline modal brands

In this section, we utilize the baseline modal brand indicator as the outcome variable (Equation 3), and use TWFE to estimate effects of closures on consumers' rate of choosing their baseline modal brand. In particular, the difference of the current analysis is that it tells us whether people return to their prior modal brand options, or the disruption will lead to lasting changes in brand choices. As we discussed in the previous section, the disruption can be caused by various mechanisms including unavailability of brands and discontinuity in shopping habits. To capture the treatment effect on baseline choices, we define all post-closure trips to be treated because we want to see the overall effect and there is no reason to expect the effect to be temporary. The follow-up event study analysis will further justify this assumption. Note that this is different from how we defined the treatment vector for the recent modal brand (Equation 4). The baseline treatment variable is defined as:

$$T^b_{i,t,c} = \begin{cases} e_{i,c}, & \tau_i \geq t \\ 0, & t < \tau_i \end{cases}. \tag{9}$$

Figure 6 Dynamics of long-term modal brand purchases



Note. Dynamics of average long-term modal brand purchase rate for every 5 trips relative to closure date, across four exposure groups and the control group. The figure supports the parallel trends assumption between different groups and the control. It also suggests a constant lasting effect after the closure, unlike the temporary effect on recent modal brand choices.

4.2.1. Descriptive analysis Again, we first plot the average modal brand purchase rates for different treated and control groups to compare the trends around store closures. Figure 6 shows the comparison for different exposure groups vs. control. Similar to the previous outcome, a differetial change for E_4 is detectable even from comparing raw means without controlling for fixed effects. The average rate of purchasing modal brand for E_4 is always more than the control group prior to the treatment, while it drops below the control curve right after the closure time. Furthermore, pre-treatment trends are parallel for exposure groups and the control group, although we will later illustrate this point more rigorously using event study analysis.

4.2.2. Difference-in-differences Similar to the recent modal brand analyses, we estimate two fixed-effects models to estimate the average and heterogeneous treatment effects. The outcome

[17] Indeed, the same estimated parameter is statistically significant if we choose $T_e = 2$ years: $\hat{\beta}_{1,4} = -1.804$, $SE = 0.833$, $P = 0.031$, and $T_e = \frac{1}{2}$ year: $\hat{\beta}_{1,4} = -1.875$, $SE = 0.826$, $p = 0.024$.

Table 2 DID results for long-term modal brand.

	Dependent variable:		
	baseline modal brand indicator (×100)		
	(1)	(2)	(3)
Overall	-14.415^{***}	-5.870^{***}	
	(2.021)	(0.894)	
E_1			-4.269^{***}
			(1.063)
E_2			-4.630^{***}
			(1.510)
E_3			-9.110^{***}
			(2.149)
E_4			-12.432^{***}
			(1.938)
Continuous treatment	✓		
Observations	895,035	895,035	895,035
R^2	0.343	0.342	0.342
Adjusted R^2	0.333	0.332	0.333
Note:			$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Estimation results for TWFE models with the baseline modal brand indicator as the outcome variable. Column 1 shows the estimated β parameter in Equation 10, and column 2 shows the same parameter where instead of a continuous treatment, a binary treatment indicator has been used. Column 3 displays corresponding β parameters in Equation 11. These coefficients measure the extent of deviation from baseline modal brands during the entire post-closures period, compared with entire pre-treatment period, relative to control household–category pairs. The fact that all coefficients are negative shows that disruption caused by the closure leads households to new brand options that are on average different from what they used to buy, and the effect becomes stronger for units with higher exposure. All standard errors are clustered at the closing store level, and all estimates are multiplied by 100.

variable and treatment are defined differently as shown in Equations 3 and 9. The overall and heterogeneous TWFE models are:

$$y^b_{i,c,t} = \alpha_{i,c} + \gamma_t + X^T_{i,t}\theta + \beta\, T^b_{i,c,t} + \epsilon_{i,c,t} \tag{10}$$

$$y^b_{i,c,t} = \alpha_{i,c} + \gamma_t + X^T_{i,t}\theta + \sum_{j=1}^{4} \beta_j\, \mathbb{1}_{i\in E_j}\, \mathbb{1}_{T^b_{i,c,t}>0} + \epsilon_{i,c,t} \tag{11}$$

where $\alpha_{i,c}$ are the household–category, and γ_t the temporal (monthly) fixed effects. $X_{i,t}$ is a set of household covariates that are varying over time and could potentially affect shopping behavior; these include dummy variables for household income level, size, and composition.

The estimation results for these equations are shown in Table 2. As it can be seen, the overall effect is negative and statistically significant. This result shows that the disruption due to store closures will cause a permanent change in modal brands for exposed household–category pairs, compared with the control units. So, although the disruption in habits is, in some sense, temporary in that people form new habits in their new environments, they converge to a set of brands that is on average different from what they used to buy. The second column shows the same parameter where instead of a continuous treatment, a binary treatment indicator has been used.[18] Finally, the third column contains heterogeneous treatment effects across exposure groups. The effect sizes are stronger for higher exposed groups, inline with our theoretical predictions and results from previous analysis. Note that although the identified temporary effect on recent modal brand was not statistically significant for lower exposed groups, the lasting effect on baseline brands is still statistically significant and substantial even for these lower exposed units.

4.2.3. Event study analysis Next, we estimate two event study models for the baseline outcome both to test implications of the parallel trends assumptions used in difference-in-differences models and explore the dynamics of the treatment effect after closures. The estimated models are defined similarly to Section 4.1.3, but with two changes. First, the outcome and treatment variables are the baseline version defined in Equations 3 and 9. Second, we used the first lead coefficient as the reference point to estimate the model because we would like to know the difference in treated and control units' behavior compared with how it was right before the closure happens. The event study models for the overall and heterogeneous effects are as follows:

$$y^b_{i,c,t} = \alpha_{i,c} + \gamma_t + X^T_{i,t}\theta + \sum_{\tau=2}^{5} \beta^{lead}_{\tau} T^b_{i,c,t-\tau} + \sum_{\tau=0}^{5} \beta^{lag}_{\tau} T^b_{i,c,t+\tau} + \epsilon_{i,c,t}. \tag{12}$$

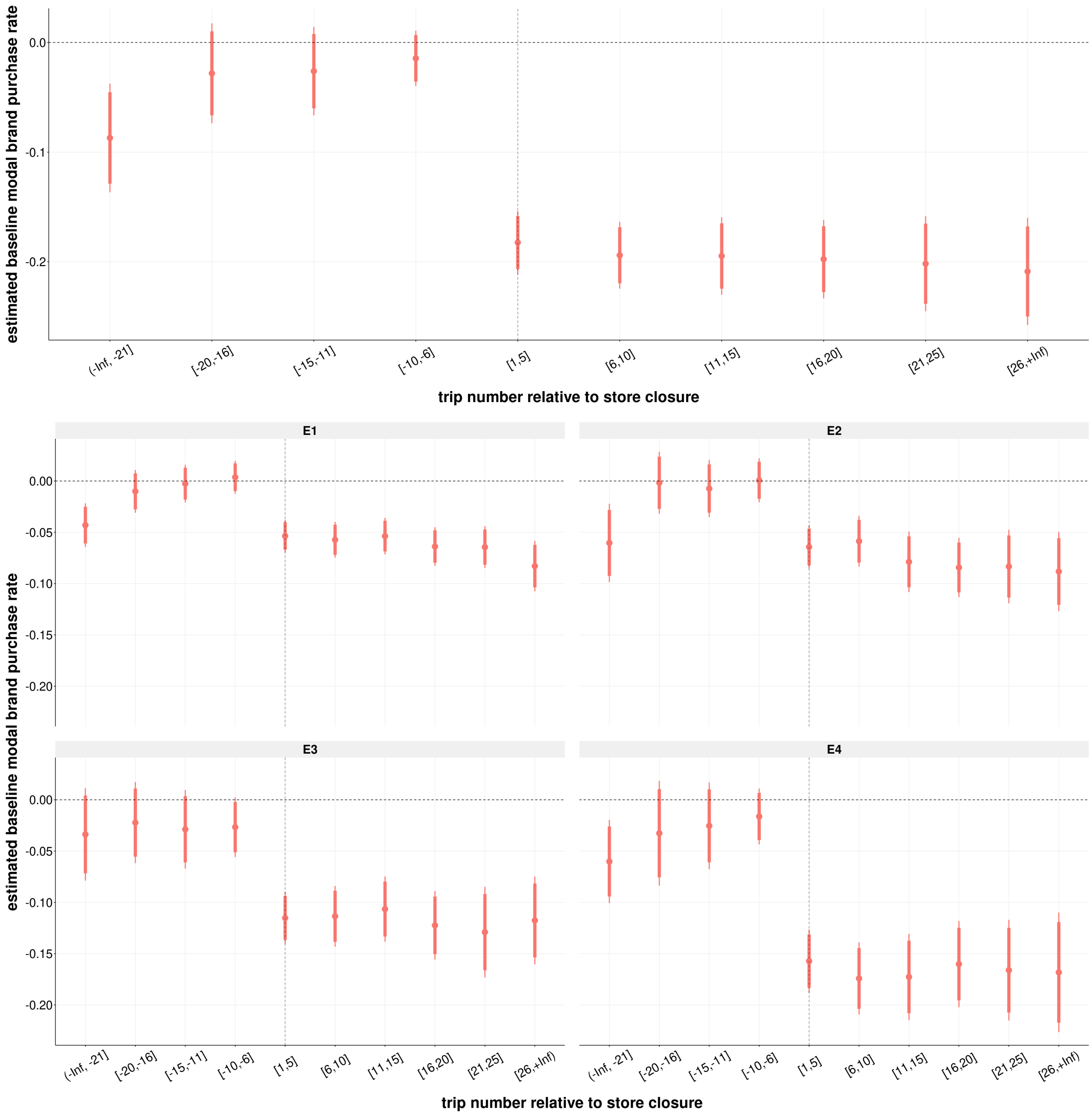
$$y^b_{i,c,t} = \alpha_{i,c} + \gamma_t + X^T_{i,t}\theta + \sum_{j=1}^{4}\sum_{\tau=2}^{5} \beta^{lead}_{j,\tau} \mathbb{1}_{i\in E_j} \mathbb{1}_{T^b_{i,c,t-\tau}>0} + \sum_{j=1}^{4}\sum_{\tau=0}^{5} \beta^{lag}_{j,\tau} \mathbb{1}_{i\in E_j} \mathbb{1}_{T^b_{i,c,t+\tau}>0} + \epsilon_{i,c,t}. \tag{13}$$

The results for these models can be seen in Figure 7. In the aggregate model, the first six lead parameters (up to trip 35 trips before closure) are not statistically significant, which means the treatment and control groups are behaving in parallel for a long time before closure in terms of purchasing baseline modal brands. However, coefficients for the last lead variable is non-zero showing that treated and control groups behave differently as we go further from the closure event. This is not unexpected since there is heterogeneity in the length of pre-closure panel across household-categories. The same observations is true for different exposure groups with some of them having parallel trends for a slightly longer period. Furthermore, the lag coefficients display a non-diminishing effect on baseline brands after the closures, unlike the recent modal brand model. The significance of the last lag coefficient ($t > 4$) shows there is a lasting effect on deviation from baseline modal brands, despite the disruption on shopping habits being temporary.

The fact that the event study model reveals non-zero and negative pre-trends long before the closures happens, could bias the results from the DID models towards zero, since those models compare the post-closure period with all pre-closure. That's is why the effect sizes in Table 2 are generally smaller than the coefficients in Figure 7.

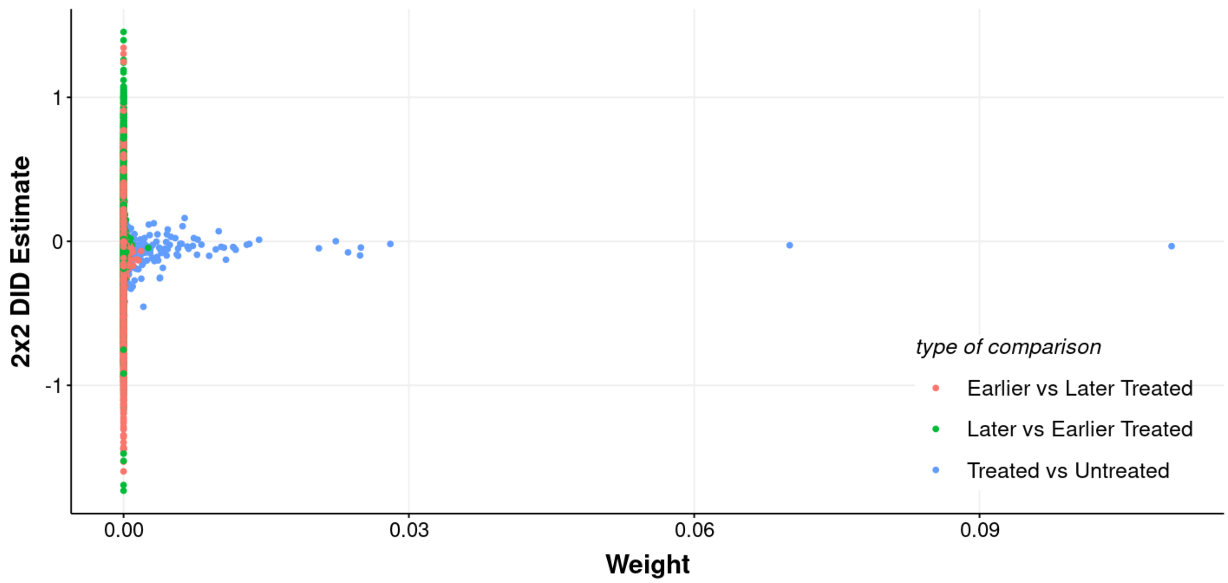
[18] We also included this result because the follow-up Bacon decomposition analysis (Goodman-Bacon 2021) does not work for continuous or heterogeneous treatments.

Figure 7 **Event study analyses for aggregate (top), and different exposure groups (bottom) with baseline modal brand indicator as the outcome variable**



Note. The plot shows the estimated lead and lag coefficients in Equations 12 & 13. Point estimates are computed for every 5 trips grouped together, the thin and thick error bars show the corresponding 95% and 90% confidence intervals. Ten lead and six lag variables are used, where the last lead (lag) includes all trips whose relative trip number is less than (more than) -21 (26). The first lead coefficient ($t = [-5, -1]$) is used as the reference level in estimating the parameters and hence not shown in the figure.

Figure 8 Weights for individual 2×2 DID estimates



Note. Individual 2×2 DID estimates and corresponding weights as defined in Equation 16. As you can see, most of the weights associated with earlier vs. later, and later vs. earlier treated units are close to zero.

4.3. Robustness: Bacon Decomposition

In this section, we use the decomposition proposed by Goodman-Bacon (2021) to access the extent of bias caused by variation in treatment timing (i.e. store closures). One can decompose the TWFE estimate into a weighted sum of individual 2×2 DID estimates (Goodman-Bacon 2021). Consider the group k of household–category pairs that face a store closure, and let u be the set of untreated pairs. An individual 2×2 DID estimate is the difference in means for between pre and post-treatment periods:

$$\hat{\beta}^{k}_{ku} = (\bar{y}^{post(k)}_{k} - \bar{y}^{pre(k)}_{k}) - (\bar{y}^{post(k)}_{u} - \bar{y}^{pre(k)}_{u}), \tag{14}$$

where the subscript shows comparison groups, and the superscript stands for the time of the treatment. For example, for two treated groups of early treated k and later treated l, $\hat{\beta}^{k}_{kl}$ is the 2×2 DID estimate between k and l when k get treated while l is still not affected, and $\hat{\beta}^{k}_{kl}$ captures the case where l get treated while k has already been treated before.

All household–category pairs facing a single store closure are gathered together in a group. If we have C store closures, there are C groups and one never treated group which consists of all household–category pairs that never face any closure. Therefore, there would be C 2×2 DID between treated and untreated units, and $C(C-1)$ 2×2 DID estimates between early treated and lately treated, and vice-versa. Following the notation in Goodman-Bacon (2021), the TWFE estimator can be decomposed as follows:

$$\hat{\beta}_{\text{TWFE}} = \sum_{k \neq u} s_{ku} \hat{\beta}^k_{ku} + \sum_{k \neq u} \sum_{l > k} s_{kl} \left[\mu_{kl} \hat{\beta}^k_{kl} + (1 - \mu_{kl}) \hat{\beta}^l_{kl} \right], \tag{15}$$

with the weights are shown in the following equations:

$$\begin{aligned} s_{ku} &= \frac{n_k n_u \bar{D}_k (1 - \bar{D}_k)}{\widehat{\text{Var}}(\tilde{D}_{i,t})} \\ s_{kl} &= \frac{n_k n_l (\bar{D}_k - \bar{D}_l)(1 - (\bar{D}_k - \bar{D}_l))}{\widehat{\text{Var}}(\tilde{D}_{i,t})} \\ \mu_{kl} &= \frac{1 - \bar{D}_k}{1 - (\bar{D}_k - \bar{D}_l)}, \end{aligned} \tag{16}$$

where n_k is the number of units in group k (i.e. the number of household–category pairs experiencing store closure k), n_u is the number of units in the control group, $\bar{D}_k$ is the fraction of the panel length group k is treated, and $\tilde{D}_{i,t}$ is the residual of the treatment vector after partialling out individual and time fixed effects.

Figure 8 show the $C^2 = 121,104$ separate 2×2 DID estimates and their weights. As it can be seen in the figure, most of the weight is contrasting earlier vs. later treated, while weight for the later vs. earlier treated contrast is close to zero. So we expect the bias to be small. Further, Table 3 shows the aggregated weights and corresponding average estimates for each comparison. This decomposition shows that there is a small, $\frac{(-5.78)-(-4.87)}{(-4.87)} \times 100 = 5.8\%$ bias in the TWFE estimator..

There are a number of implicit assumptions used in the decomposition results presented in the previous section. Note that the TWFE estimate here is different from the aggregate model in Table 2. The reason is that so far we have been using a continuous treatment vector where the treatment intensity was proportional to the exposure level as defined in Equation 1. However, Bacon decomposition only works for binary treatment variables, so in this section, we defined the treatment as an indicator that equals one in the post-closure period if the corresponding exposure level is not zero. The estimated parameter for a binary treatment is reported in Table 2 column 2.

Also, Bacon decomposition assumes a non-decreasing treatment assignment, which matches our baseline model in (Equation 10) but not the recent modal brand model (Equation 5). However, according to Equation 16, the weights for early and late cross-comparisons s_{kl} depend on the differential amount of time units get treated $\bar{D}_k - \bar{D}_l$. While, in the recent modal brand model by definition (Equation 4), units are treated only temporarily, so the difference in duration of treatment timings are much less, although they are not exactly zero if the next closure happens before the L trips duration for the consumers of the previous closing store. As a result, we expect the bias inferred from Table 3 be an upper bound for any bias for the recent modal brand model.

Table 3 Aggregate weights and estimates from Bacon decomposition

Type of comparison	Weight	Avg. estimate × 100
Earlier vs. Later Treated	0.131	-11.91
Later vs. Earlier Treated	0.109	-4.65
Treated vs. Untreated	0.759	-4.35

Note: The aggregated weights and associated average estimates for each comparison in the Bacon decomposition (Equation 15). The majority of the weight is attributed to the treated vs. untreated comparison. The middle comparison (Later vs. Earlier) is the one that can cause bias since it is comparing later treated units with already treated ones. The weighted average of the other two comparisons is -4.87 which is very close to the TWFE estimate -5.78 (Table 2 column 2) indicating small (5.8%) bias due to unwanted comparisons.

5. Discussion

In summary, we showed that households on average choose different brands in product categories after facing a the closure of a store where they used to purchase in that category, thereby exposing them to new shopping contexts. This pattern is robust to accounting for the lack of availability of their prior modal brands. Furthermore, we show that the temporary disruption in shopping habits results in lasting changes in households' modal brands suggesting that formation of shopping habits could lead to sub-optimal behavior. These findings provide positive empirical evidence for the effect of shopping habits on consumers' in-store decision making, and that the measured inertia in the literature is not only due to brand loyalty, but also attributable to learned habits in the stable shopping contexts. We attribute the observed behavior to habits for two main reasons. First, the heterogeneous effect models show that the effect sizes increase with the frequency of prior purchases in that category at the closing store. This is in line with the prior literature on habits where it has been shown that frequent actions in a stable context lead to habit formation (Wood *et al.* 2005). Second, the observed effect is robust to accounting for the unavailability of brand options. Since store closures are not dependant on an individual consumer's tastes, there is no reason to believe consumers suddenly would have changed their preferred brands in the absence of the closure — and only in the categories purchased at that closing store. So different choices are a result of exposure to a new context where old habits are no longer present.

These results have consequences both for firms and customers. These kinds of behavioral factors in consumers' decision-making can have managerial implications for optimal pricing, advertising strategies, and allocation and location of goods inside stores. Strong consumption habits are at the core of marketing campaigns designed to attract consumers from other brands, e.g., by temporary price reductions or free sampling, in order to benefit from their choice inertia in the long-run. The average state dependence can affect firms' decisions about promotions or temporary price discounts. If people are very state-dependent, it would be an extra incentive for firms to attract customers sooner than later. So they have incentives to lower prices. This phenomenon has been studied in form of a dynamic game among firms to lower prices to exploit consumers' state dependence. The equilibrium of this game would depend on consumers' state dependence, and it can be shown that in some cases consumers can benefit from this competition (Klemperer 1987, Seetharaman and Che 2009). Therefore, correctly estimating the degree of consumer brand loyalty can be crucial for firms' decisions and biased estimates could adversely affect their total profit.

Having a better understanding of the psychological determinants of consumers shopping behavior can benefit both individuals and firms. On the consumer side, it can help us design more effective interventions to improve people's health, e.g., by nudging them to choose healthier options (Leonard 2008). Most habits are typically formed to help achieve particular goals (Aarts and Dijksterhuis 2000, Wood and Neal 2007). But these habits could continue to persist even after changes in outcome structure or reward devaluation (Wood 2017, Neal *et al.* 2011, Adams and Dickinson 1981). As a result, although in the short run habits could be advantageous by automating repetitive tasks and freeing up mental resources, in the long run they could lead to non-optimal behavior. For example, Larcom *et al.* (2017) show that a significant fraction of commuters on the London Underground used to take non-optimal routes, and an exogenous disruption such as a strike brought lasting behavior change in commuters' routing behavior. The same phenomenon could also happen for in-store purchase behavior where a shock causing extended periods of brand availability could bring lasting changes in brand choices (Figueroa *et al.* 2019). This inertia of habits could have adverse healthcare consequences when habits of buying unhealthy products are formed. As a result, it is important to understand the extent to which habit formation influences consumers' in-store shopping behavior. Being aware of the role of context can help those individuals seeking a change in their consumption habits. When they change the typical place they visit for shopping, they are less affected by exiting contextual cues and have an opportunity to start buying healthier products. Also, this could help policymakers design and implement more effective policies.

Finally, our findings shows a substantial role for shopping habits in people's purchasing decisions. This poses an immediate question for retailers regarding the effect of commonly practiced in-store re-arrangement of items on the store's aggregate sales and profits. One could imagine competing mechanisms in play here which could make these actions beneficial or damaging for the store. On the one hand, re-arrangement of items would nudge people to explore the store further and find things which could have been ignored previously due to existing shopping habits. On the other hand, these explorations are not free and require extra time and cognitive effort. These search costs could ultimately make people give up on buying an item. Perhaps future work will address this question through analysis of field experiments with changes to stores. Beyond implications for managers, shopping habits may importantly moderate effects of policies Khan *et al.*, Hinnosaar, Seiler *et al.*, such as taxes on particular product, on short- and long-run purchase patterns.

References

Aarts H, Dijksterhuis A (2000) Habits as knowledge structures: automaticity in goal-directed behavior. *Journal of Personality and Social Psychology* 78(1):53.

Adamowicz WL, Swait JD (2013) Are food choices really habitual? Integrating habits, variety-seeking, and compensatory choice in a utility-maximizing framework. *American Journal of Agricultural Economics* 95(1):17–41.

Adams CD, Dickinson A (1981) Instrumental responding following reinforcer devaluation. *The Quarterly Journal of Experimental Psychology Section B* 33(2b):109–121.

Amir O, Levav J (2008) Choice construction versus preference construction: The instability of preferences learned in context. *Journal of Marketing Research* 45(2):145–158.

Baker AC, Larcker DF, Wang CC (2022) How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics* 144(2):370–395.

Bayley PJ, Frascino JC, Squire LR (2005) Robust habit learning in the absence of awareness and independent of the medial temporal lobe. *Nature* 436(7050):550–553.

Beshears J, Lee HN, Milkman KL, Mislavsky R, Wisdom J (2021) Creating exercise habits using incentives: The trade-off between flexibility and routinization. *Management science* 67(7):4139–4171.

Busch-Geertsema A, Lanzendorf M (2017) From university to work life – Jumping behind the wheel? Explaining mode change of students making the transition to professional life. *Transportation Research Part A: Policy and Practice* 106:181–196, ISSN 09658564, URL `http://dx.doi.org/10.1016/j.tra.2017.09.016`.

Callaway B, Sant'Anna PH (2020) Difference-in-differences with multiple time periods. *Journal of Econometrics* .

Cantoni E, Pons V (2022) Does context outweigh individual characteristics in driving voting behavior? Evidence from relocations within the united states. *American Economic Review* 112(4):1226–72.

Carrasco R, Labeaga JM, David López-Salido J (2004) Consumption and Habits: Evidence from Panel Data. *The Economic Journal* 115(500):144–165, ISSN 0013-0133, URL `http://dx.doi.org/10.1111/j.1468-0297.2004.00963.x`.

Clark B, Chatterjee K, Melia S (2016) Changes to commute mode: The role of life events, spatial context and environmental attitude. *Transportation Research Part A: Policy and Practice* 89:89–105.

Cunningham RM (1956) Brand loyalty — What, where, how much. *Harvard Business Review* 34(1):116–128.

de Wit S, Watson P, Harsay HA, Cohen MX, van de Vijver I, Ridderinkhof KR (2012) Corticostriatal connectivity underlies individual differences in the balance between habitual and goal-directed action control. *Journal of Neuroscience* 32(35):12066–12075.

Dickinson A (1985) Actions and habits: The development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 308(1135):67–78.

Dong X, Morozov I, Seiler S, Hou L (2020) Estimation of preference heterogeneity in markets with costly search. Technical report, Working paper.

Dubé JP, Hitsch GJ, Rossi PE (2010) State dependence and alternative explanations for consumer inertia. *The RAND Journal of Economics* 41(3):417–445, ISSN 07416261, URL `http://dx.doi.org/10.1111/j.1756-2171.2010.00106.x`.

Farrell J, Klemperer P (2007) Coordination and lock-in: Competition with switching costs and network effects. *Handbook of Industrial Organization* 3:1967–2072.

Figueroa C, Musalem A, Noton C (2019) Earthquakes and brand loyalty: Beyond the short-term effects of stockouts. *Available at SSRN 3325923* .

Gardner B (2015) A review and analysis of the use of 'habit'in understanding, predicting and influencing health-related behaviour. *Health psychology review* 9(3):277–295.

Gardner B, Abraham C, Lally P, de Bruijn GJ (2012) Towards parsimony in habit measurement: Testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. *International Journal of Behavioral Nutrition and Physical Activity* 9(1):102, ISSN 1479-5868, URL `http://dx.doi.org/10.1186/1479-5868-9-102`.

Goodman-Bacon A (2021) Difference-in-differences with variation in treatment timing. *Journal of Econometrics* .

Granger CW (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 424–438.

Guadagni PM, Little JDC (1983) A Logit Model of Brand Choice Calibrated on Scanner Data. *Marketing Science* 27(1):29–48, ISSN 0732-2399, 1526-548X, URL `http://dx.doi.org/10.1287/mksc.1070.0331`.

Heckman JJ (1981) Statistical models for discrete panel data. *Structural Analysis of Discrete Data with Econometric Applications* 114:178.

Hinnosaar M (2021) The persistence of healthy behaviors in food purchasing. *Available at SSRN 3143853* .

Imai K, Kim IS (2019) When should we use unit fixed effects regression models for causal inference with longitudinal data? *American Journal of Political Science* 63(2):467–490.

James W (1890) *The Principles of Psychology* (Macmillan London).

Keane MP (1997) Modeling Heterogeneity and State Dependence in Consumer Choice Behavior. *Journal of Business & Economic Statistics* 15(3):310–327, ISSN 0735-0015, 1537-2707, URL `http://dx.doi.org/10.1080/07350015.1997.10524709`.

Khan R, Misra K, Singh V (2016) Will a fat tax work? *Marketing Science* 35(1):10–26.

Klemperer P (1987) Markets with consumer switching costs. *The Quarterly Journal of Economics* 102(2):375–394.

Knowlton BJ, Mangels JA, Squire LR (1996) A neostriatal habit learning system in humans. *Science* 273(5280):1399–1402.

Labrecque JS, Wood W (2015) What measures of habit strength to use? Comment on Gardner (2015). *Health Psychology Review* 9(3):303–310.

Larcom S, Rauch F, Willems T (2017) The benefits of forced experimentation: Striking evidence from the London Underground network. *The Quarterly Journal of Economics* 132(4):2019–2055, ISSN 0033-5533, 1531-4650, URL `http://dx.doi.org/10.1093/qje/qjx020`.

Leonard TC (2008) Richard h. thaler, cass r. sunstein, nudge: Improving decisions about health, wealth, and happiness.

Levine J, Seiler S (2021) Identifying state dependence in brand choice: Evidence from hurricanes, CEPR Discussion Paper No. DP16632.

Liu-Thompkins Y, Tam L (2013) Not all repeat customers are the same: Designing effective cross-selling promotion on the basis of attitudinal loyalty and habit. *Journal of Marketing* 77(5):21–36.

Machín L, Curutchet MR, Gugliucci V, Vitola A, Otterbring T, de Alcantara M, Ares G (2020) The habitual nature of food purchases at the supermarket: Implications for policy making. *Appetite* 104844, ISSN 01956663, URL `http://dx.doi.org/10.1016/j.appet.2020.104844`.

Macrae CN, Milne AB, Bodenhausen GV (1994) Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of personality and Social Psychology* 66(1):37.

Mazar A, Wood W (2018) Defining habit in psychology. *The Psychology of Habit*, 13–29 (Springer).

McAlister L (1982) A dynamic attribute satiation model of variety-seeking behavior. *Journal of consumer research* 9(2):141–150.

Neal DT, Wood W, Wu M, Kurlander D (2011) The pull of the past: When do habits persist despite conflict with motives? *Personality and Social Psychology Bulletin* 37(11):1428–1437.

Oblander ES, McCarthy DM (2022) Persistence of consumer lifestyle choices: Evidence from restaurant delivery during COVID-19, available at https://dx.doi.org/10.2139/ssrn.3836262.

Orbell S, Verplanken B (2010) The automatic component of habit in health behavior: Habit as cue-contingent automaticity. *Health Psychology* 29(4):374–383, ISSN 1930-7810, 0278-6133, URL `http://dx.doi.org/10.1037/a0019596`.

Pakes A, Porter JR, Shepard M, Calder-Wang S (2021) Unobserved heterogeneity, state dependence, and health plan choices. Technical report, National Bureau of Economic Research.

Poortinga W, Whitmarsh L, Suffolk C (2013) The introduction of a single-use carrier bag charge in Wales: Attitude change and behavioural spillover effects. *Journal of Environmental Psychology* 36:240–247, ISSN 02724944, URL `http://dx.doi.org/10.1016/j.jenvp.2013.09.001`.

Privitera GJ, Zuraikat FM (2014) Proximity of foods in a competitive food environment influences consumption of a low calorie and a high calorie food. *Appetite* 76:175–179, ISSN 01956663, URL `http://dx.doi.org/10.1016/j.appet.2014.02.004`.

Roth J (2019) Pre-test with caution: Event-study estimates after testing for parallel trends. *Department of Economics, Harvard University, Unpublished manuscript* .

Roy R, Chintagunta PK, Haldar S (1996) A framework for investigating habits,"the hand of the past," and heterogeneity in dynamic brand choice. *Marketing Science* 15(3):280–299.

Seetharaman P (2004) Modeling multiple sources of state dependence in random utility models: A distributed lag approach. *Marketing Science* 23(2):263–271.

Seetharaman P, Ainslie A, Chintagunta PK (1999) Investigating household state dependence effects across categories. *Journal of Marketing Research* 36(4):488–500.

Seetharaman P, Che H (2009) Price competition in markets with consumer variety seeking. *Marketing Science* 28(3):516–525.

Seiler S, Tuchman A, Yao S (2021) The impact of soda taxes: Pass-through, tax avoidance, and nutritional effects. *Journal of Marketing Research* 58(1):22–49.

Simonov A, Dubé JP, Hitsch G, Rossi P (2020) State-dependent demand estimation with initial conditions correction. *Journal of Marketing Research* 57(5):789–809.

Sun L, Abraham S (2021) Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics* 225(2):175–199.

Tam L, Wood W, Ji MF (2014) Brand loyalty is not habitual. *Handbook of Brand Relationships* 65–84.

Thomadsen R (2016) The Impact of Switching Stores on State Dependence in Brand Choice. *SSRN Electronic Journal* ISSN 1556-5068, URL `http://dx.doi.org/10.2139/ssrn.2759868`.

Thomadsen R, Seetharaman PB (2018) The strategic effects of state-dependent consumer preferences: The roles of habits and variety seeking. *The Psychology of Habit*, 131–150 (Springer).

Verplanken B (2006) Beyond frequency: Habit as mental construct. *British Journal of Social Psychology* 45(3):639–656, ISSN 01446665, URL `http://dx.doi.org/10.1348/014466605X49122`.

Verplanken B, Aarts H (1999) Habit, Attitude, and Planned Behaviour: Is Habit an Empty Construct or an Interesting Case of Goal-directed Automaticity? *European Review of Social Psychology* 10(1):101–134, ISSN 1046-3283, 1479-277X, URL `http://dx.doi.org/10.1080/14792779943000035`.

Verplanken B, Orbell S (2003) Reflections on Past Behavior: A Self-Report Index of Habit Strength [1]. *Journal of Applied Social Psychology* 33(6):1313–1330, ISSN 00219029, URL `http://dx.doi.org/10.1111/j.1559-1816.2003.tb01951.x`.

Verplanken B, Walker I, Davis A, Jurasek M (2008) Context change and travel mode choice: Combining the habit discontinuity and self-activation hypotheses. *Journal of Environmental Psychology* 28(2):121–127, ISSN 02724944, URL `http://dx.doi.org/10.1016/j.jenvp.2007.10.005`.

Verplanken B, Wood W (2006) Interventions to Break and Create Consumer Habits. *Journal of Public Policy & Marketing* 25(1):90–103, ISSN 0748-6766, 1547-7207, URL http://dx.doi.org/10.1509/jppm.25.1.90.

Wood W (2017) Habit in Personality and Social Psychology. *Personality and Social Psychology Review* 21(4):389–403, ISSN 1088-8683, 1532-7957, URL http://dx.doi.org/10.1177/1088868317720362.

Wood W, Labrecque JS, Lin PY, Rünger D (2014) Habits in dual process models. Sherman JW, Gawronski B, Trope Y, eds., *Dual Process Theories of the Social Mind*, 371–385 (Guilford Press New York, NY).

Wood W, Neal DT (2007) A new look at habits and the habit-goal interface. *Psychological Review* 114(4):843–863, ISSN 1939-1471, 0033-295X, URL http://dx.doi.org/10.1037/0033-295X.114.4.843.

Wood W, Neal DT (2016) Healthy through habit: Interventions for initiating & maintaining health behavior change. *Behavioral Science & Policy* 2(1):71–83.

Wood W, Quinn JM, Kashy DA (2002) Habits in everyday life: Thought, emotion, and action. *Journal of Personality and Social Psychology* 83(6):1281–1297, ISSN 1939-1315, 0022-3514, URL http://dx.doi.org/10.1037/0022-3514.83.6.1281.

Wood W, Rünger D (2016) Psychology of Habit. *Annual Review of Psychology* 67(1):289–314, ISSN 0066-4308, 1545-2085, URL http://dx.doi.org/10.1146/annurev-psych-122414-033417.

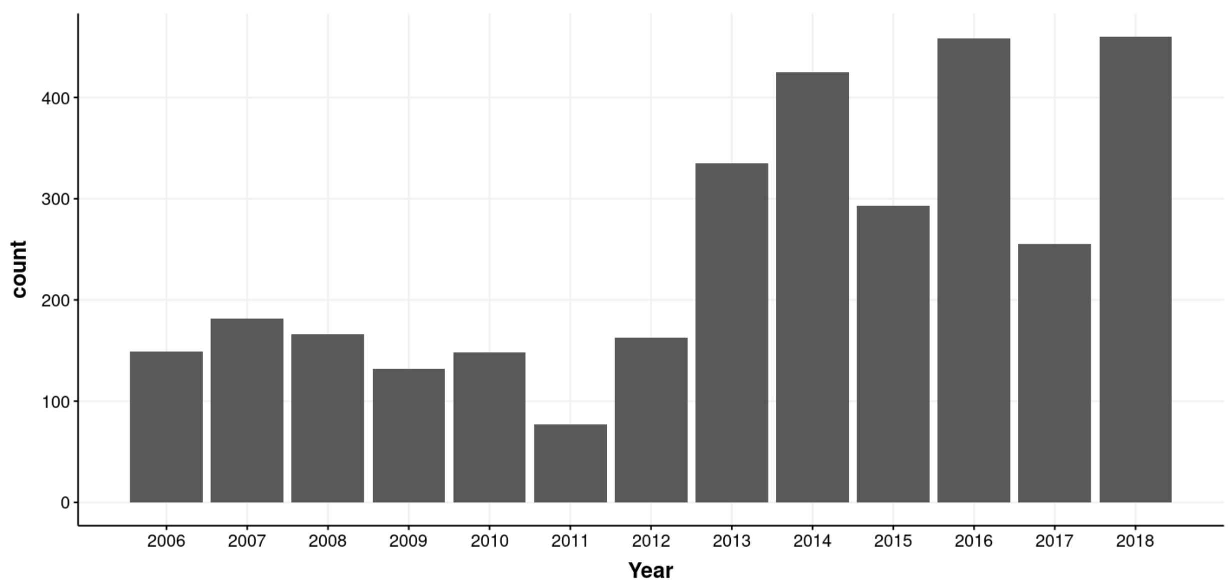
Wood W, Tam L, Witt MG (2005) Changing circumstances, disrupting habits. *Journal of Personality and Social Psychology* 88(6):918–933, ISSN 1939-1315, 0022-3514, URL http://dx.doi.org/10.1037/0022-3514.88.6.918.

6. APPENDIX

6.1. Additional Statistics

In this appendix, we include some additional figures to illustrate the data. Figure 9 shows the number of identified closing stores over years, and Figure 10 shows the geographic dispersion of closing stores over the US at the county level. Figure 11 displays log-weekly sales in the top-5 product categories by national purchase volume for the closing stores corresponding to the set of exposed households from January 2006 until December 2018, separated based on retailers.

Figure 9 **Distribution of closing stores over years**



Note. Number of closing stores in each year. The fact that there are more identified closings later in the timeline is not necessarily indicator of more closures, but could rather be due to change in the Nielsen data sample over time.

6.2. Treatment distribution

Figure 10 **The geographic dispersion of closing stores in the US at the county level.**

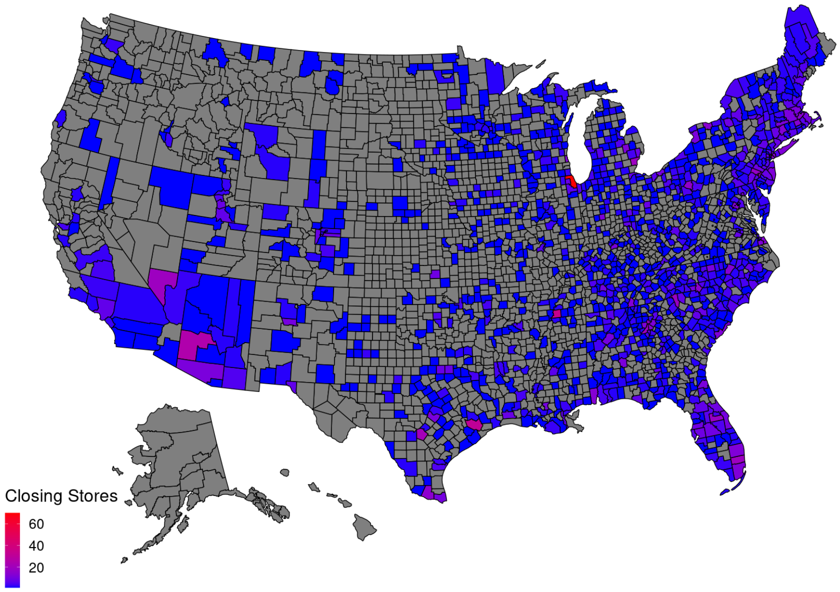
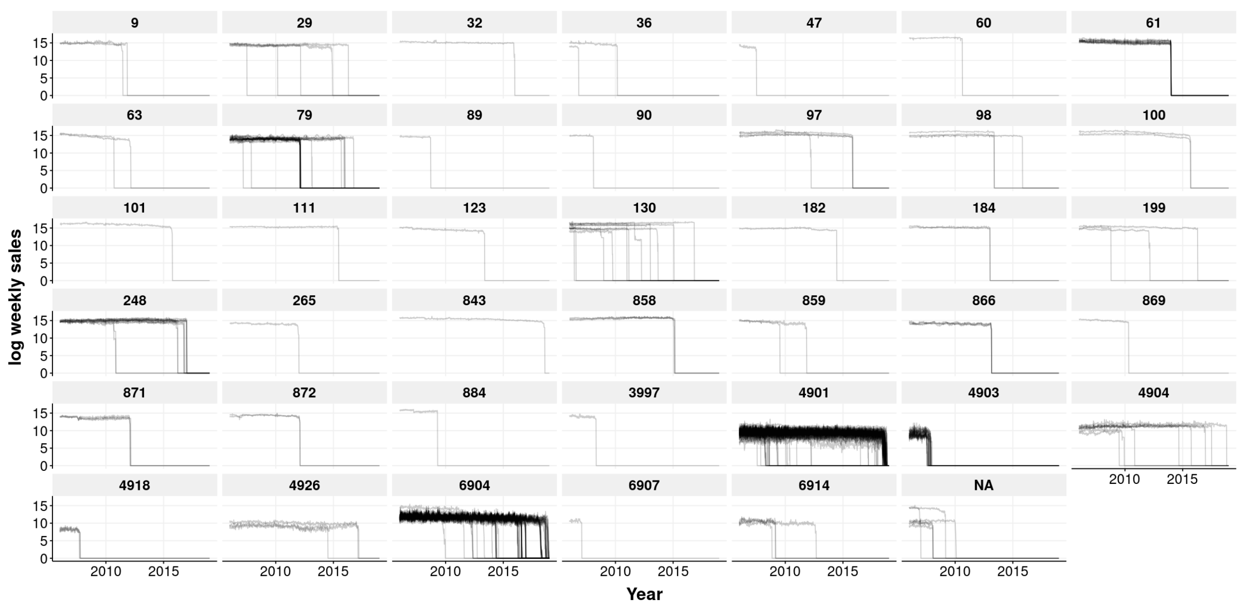


Figure 11 **Log weekly sales of the closing stores corresponding to exposed households, separated by retailers.**

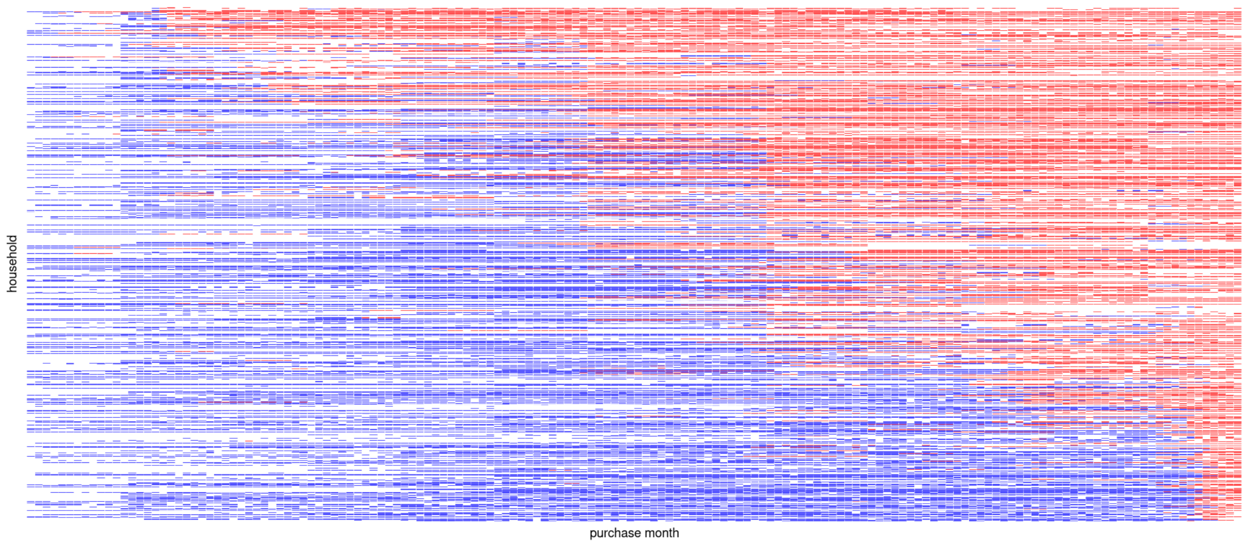


Note. The figure shows log-weekly sales in the top-5 product categories by national purchase volume for the closing stores corresponding to the set of exposed households from January 2006 until December 2018, separated based on retailers. Numbers on top of each facet show the corresponding retailer code in the Nielsen data.

6.3. Recent modal brand replication results

In this appendix we replicate the results for all DID and event study models for different values of L and T_e. L determines the length of the window for used to define the recent modal brand

Figure 12 Timing of household exposure to store closures



Note. Each row stands for a household, and columns corresponds to different purchase months. Blue rectangles show pre-closure and red rectangles show post-closure purchase occasions. Note that each household could be both in treatment and control groups, based on the household–category exposure level. Closures are staggered over time, and due to Bacon decomposition, units for which the closure happens in the middle of the panel have a higher weight in the TWFE fixed effects estimate.

(Equation 2), and T_e specifies the pre-closure duration used to define the exposure levels (Equation 1).

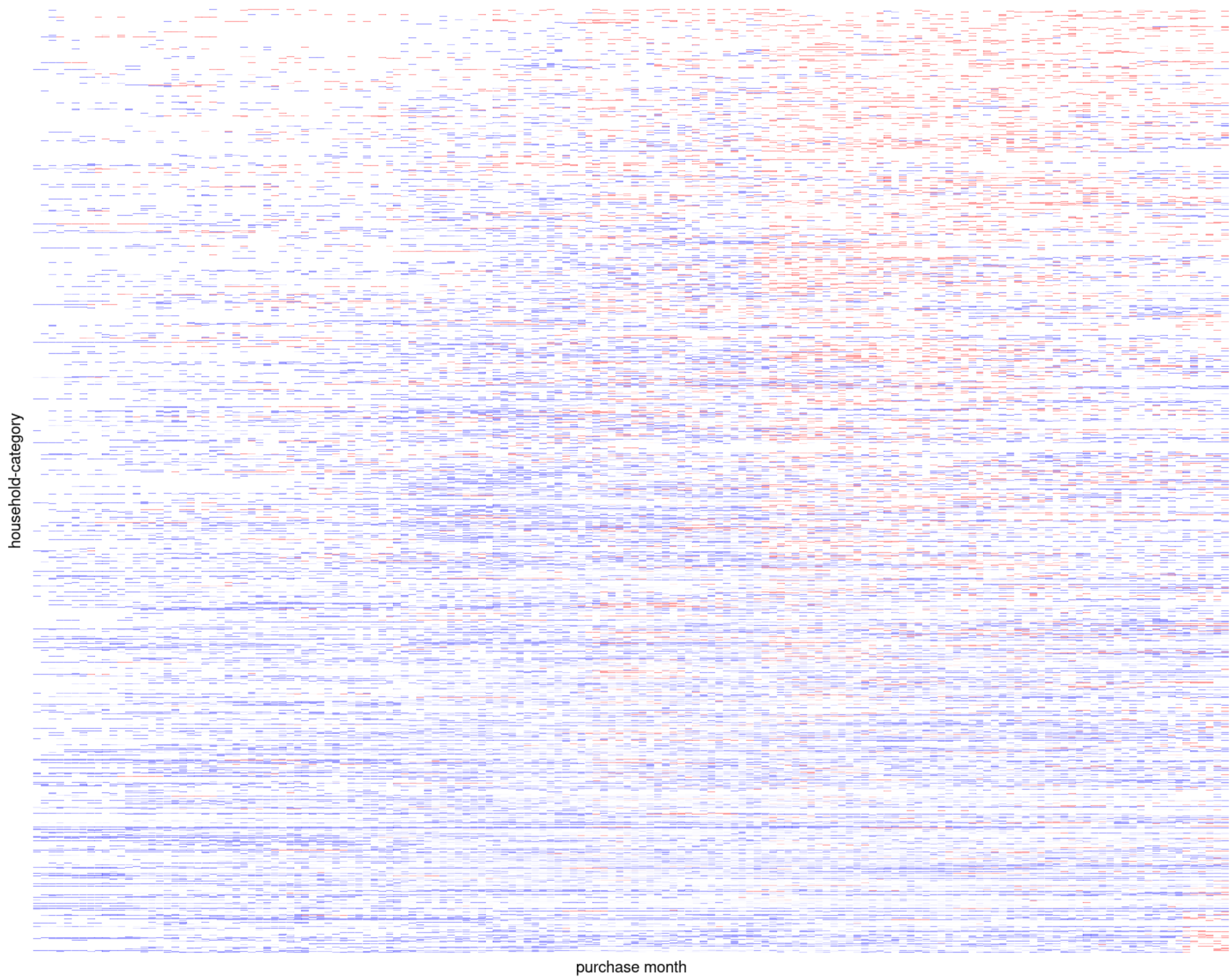
6.3.1. Changing T_e Here, we change the value of the period on which treatment exposure is defined,T_e, to show that results are qualitatively robust to choice of this parameter in a range of values. In the main text, we used $T_e = 1$ year. Here, we present results for $T_e = 1/2$ year, and $T_e = 2$ years.

6.3.2. Changing L Note that L enters the analysis in two ways, first through the definition of the L-recent modal brand, and second by the short-term treatment definition (Equation 4). The former specifies the width of the moving window based on which the modal brand is defined, and the latter determines the short-term period after closure for which we expect habits to be disrupted. One could imagine using two different L values, but we considered a symmetric case for simplicity of presenting results. Also, note that the dynamics of short-term treatment effect is estimated more systematically with the event study models, so what really matters is the width of the moving window.

6.4. Baseline modal brand replication results

In this appendix we replicate the results for the the baseline modal brand analysis where instead of all pre-closure period shopping trips, the baseline modal brand is defined based on the prior

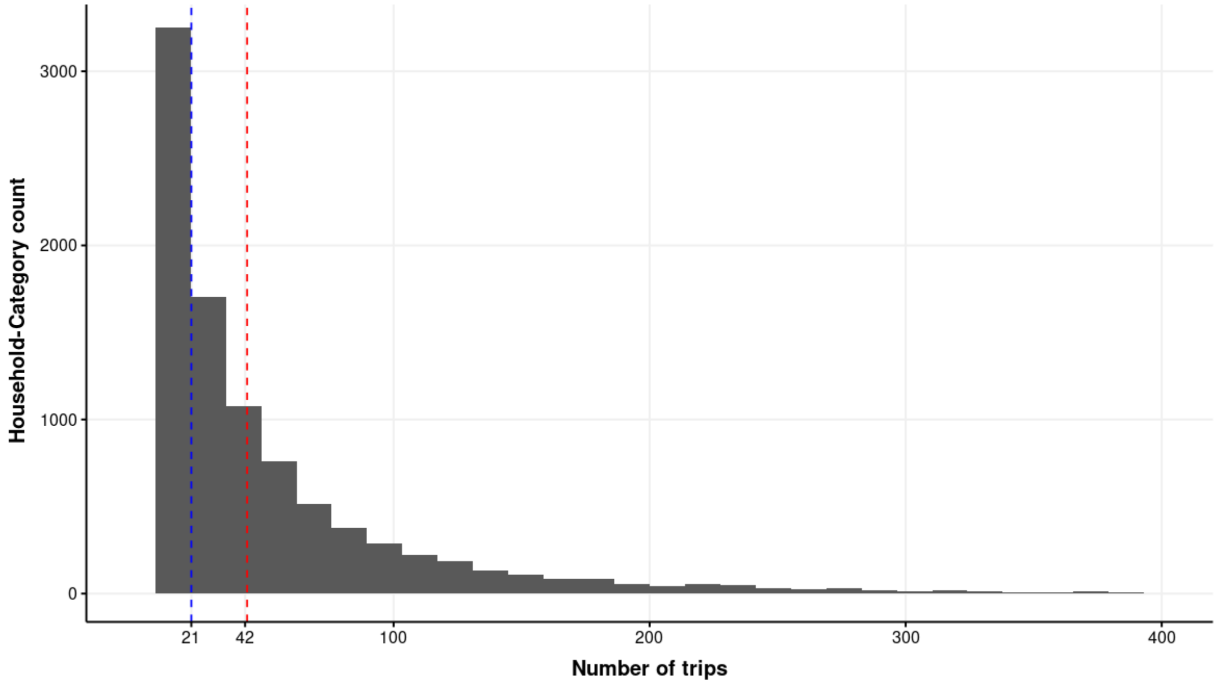
Figure 13 Distribution of short-term treatment across household–category pairs over time.



Note. In this figure each row corresponds to a household–category pair, and that is why is looks sparse. Also, in each category, only the first L trips after the closure are marked as treated (red) to reflect the short-term β_1 in Equation 4. Control units (household–category pairs with zero exposure level) are not shown in the figure.

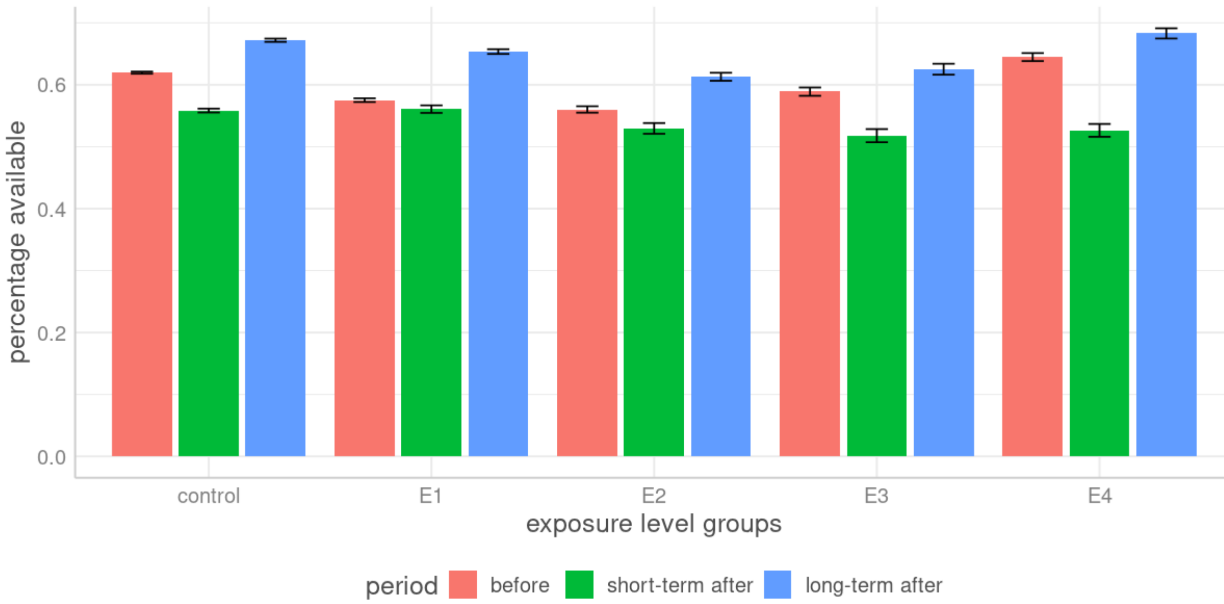
40 trips in each category. Since the panel is not balanced, the number of pre-closure trips could be highly variable for different household–category pairs. So in order to make analysis comparable across different units, we use a fix length of 40 trips to define the baseline modal brand. Note that the average number of pre-closure trips is 42.

Figure 14 Histogram of the number of pre-closure shopping trips across household–category pairs.



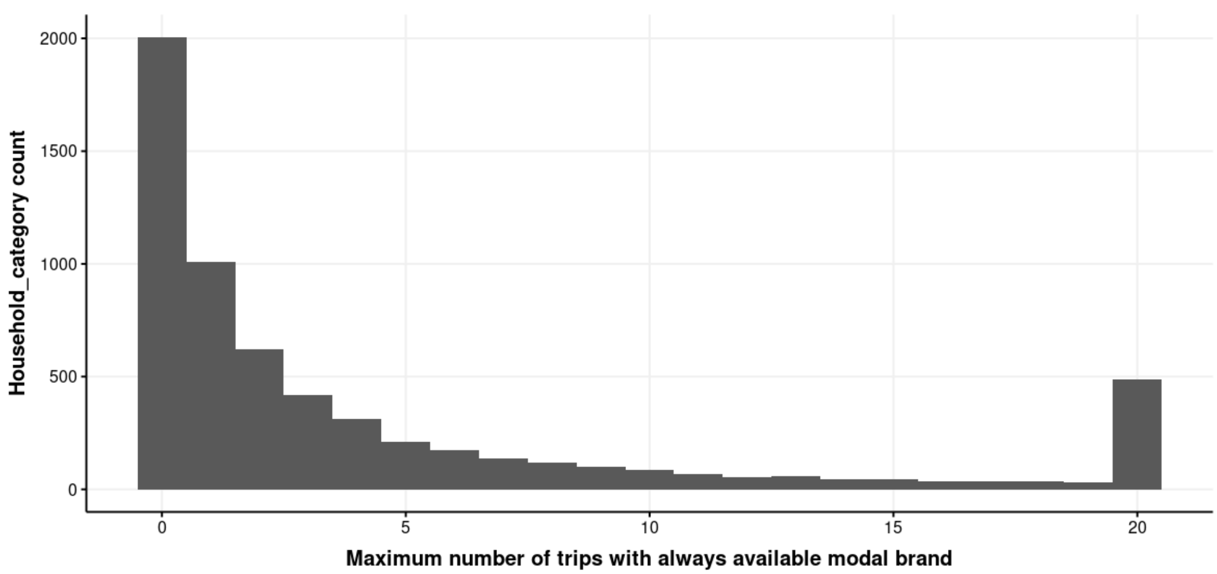
Note. The figure shows the histogram of number of shopping trips for all household–category pairs, before the corresponding store closure. Mean (red) of the distribution is 42 and the median (blue) equals 21.

Figure 15 **Recent modal brand availability for different treatment groups, before and after closures.**



Note. Different periods are defined based on the treatment definitions in Equation 4. The *before* period contains the whole pre-closure trips, *short-term after* period includes the first L trips for each category (where $T^{r1}_{i,t,c} \neq 0$), and *long-term after* marks the rest of the trips. Note that since periods are defined at the household–category level, for a certain household a trip could be in the *short-term after* period for category a, while in the *long-term after* period for category b. For each week, we mark a modal brand as available in a store if there is at least one purchase occasion by any household in the entire consumer panel data in the same store.

Figure 16 **Distribution of maximum number of trips with always available modal brand (L_a).**



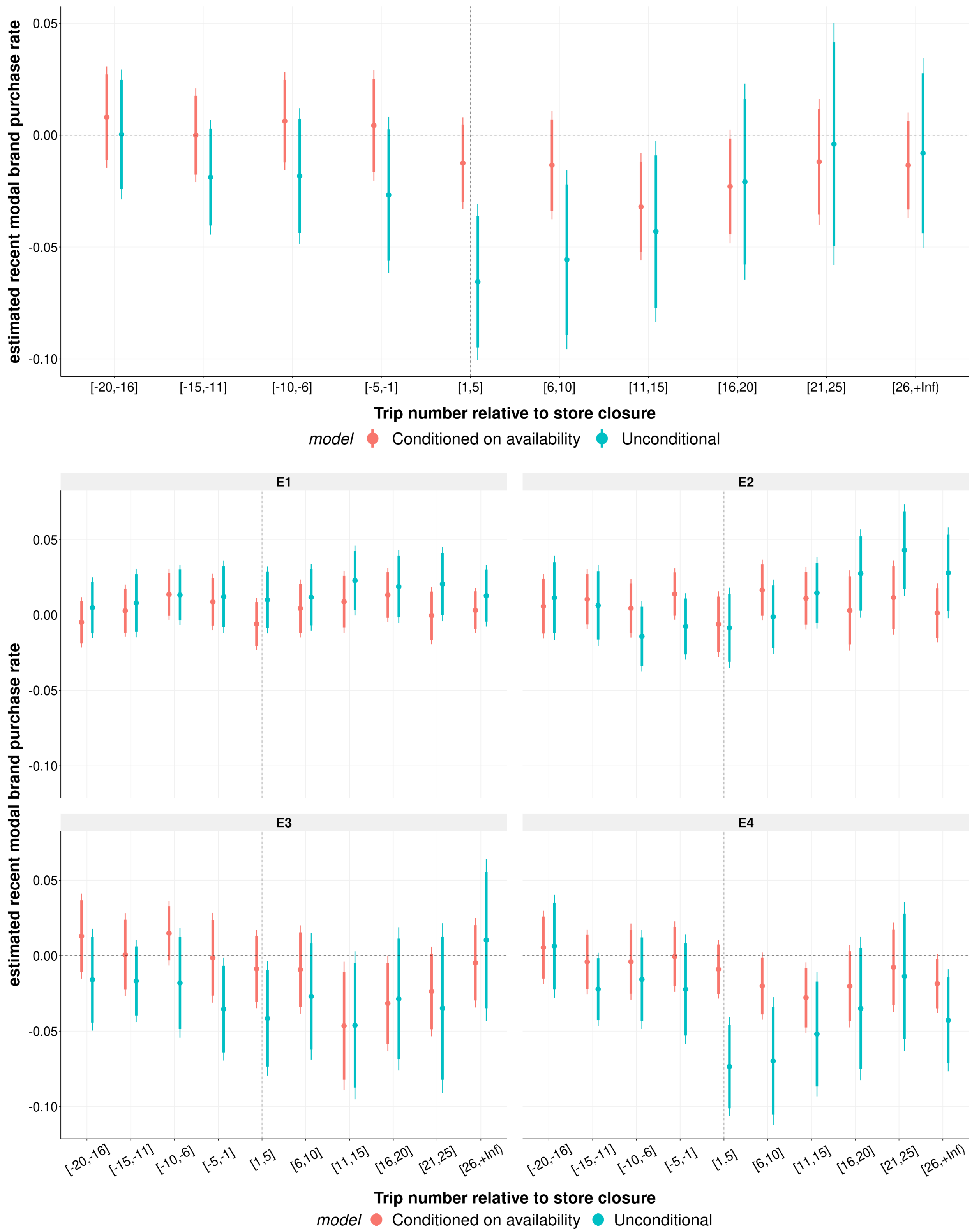
Note. Distribution of the maximum number of trips where modal brand is always available in a sequence, over all household–category pairs.

Table 4 DID results for recent modal brand.

	Dependent variable:			
	recent modal brand indicator (×100)			
	(1)	(2)	(3)	(4)
Overall, short-term	-3.867^{**}		-2.207^{**}	
	(1.330)		(0.676)	
Overall, long-term	-0.006		-1.402	
	(1.922)		(0.911)	
E_1, short-term			1.308	
0.292				
		(0.772)		(0.619)
E_2, short-term		0.757		0.055
		(0.874)		(0.699)
E_3, short-term		-2.553		-2.650^{**}
		(1.611)		(0.998)
E_4, short-term		-5.055^{***}		-1.742^{*}
		(1.276)		(0.715)
E_1, long-term		0.982		0.052
		(0.911)		(0.640)
E_2, long-term		3.022^{*}		-0.065
		(1.325)		(0.884)
E_3, long-term		1.613		-0.726
		(2.487)		(1.236)
E_4, long-term		-3.368^{*}		-1.648^{*}
		(1.414)		(0.712)
Continuous treatment	✓		✓	
Conditioned on modal brand availability			✓	✓
Observations	846,032	846,032	520,216	520,216
R^2	0.279	0.279	0.144	0.144
Adjusted R^2	0.269	0.269	0.126	0.126
Note:		*p<0.05; **p<0.01; ***p<0.001		

Replication of the results in table 1 for $T_e = \frac{1}{2}$ year and $L = 20$.

Figure 17 Event study analyses results for aggregate (top), and different exposure groups (bottom) with the recent modal brand indicator as the outcome variable



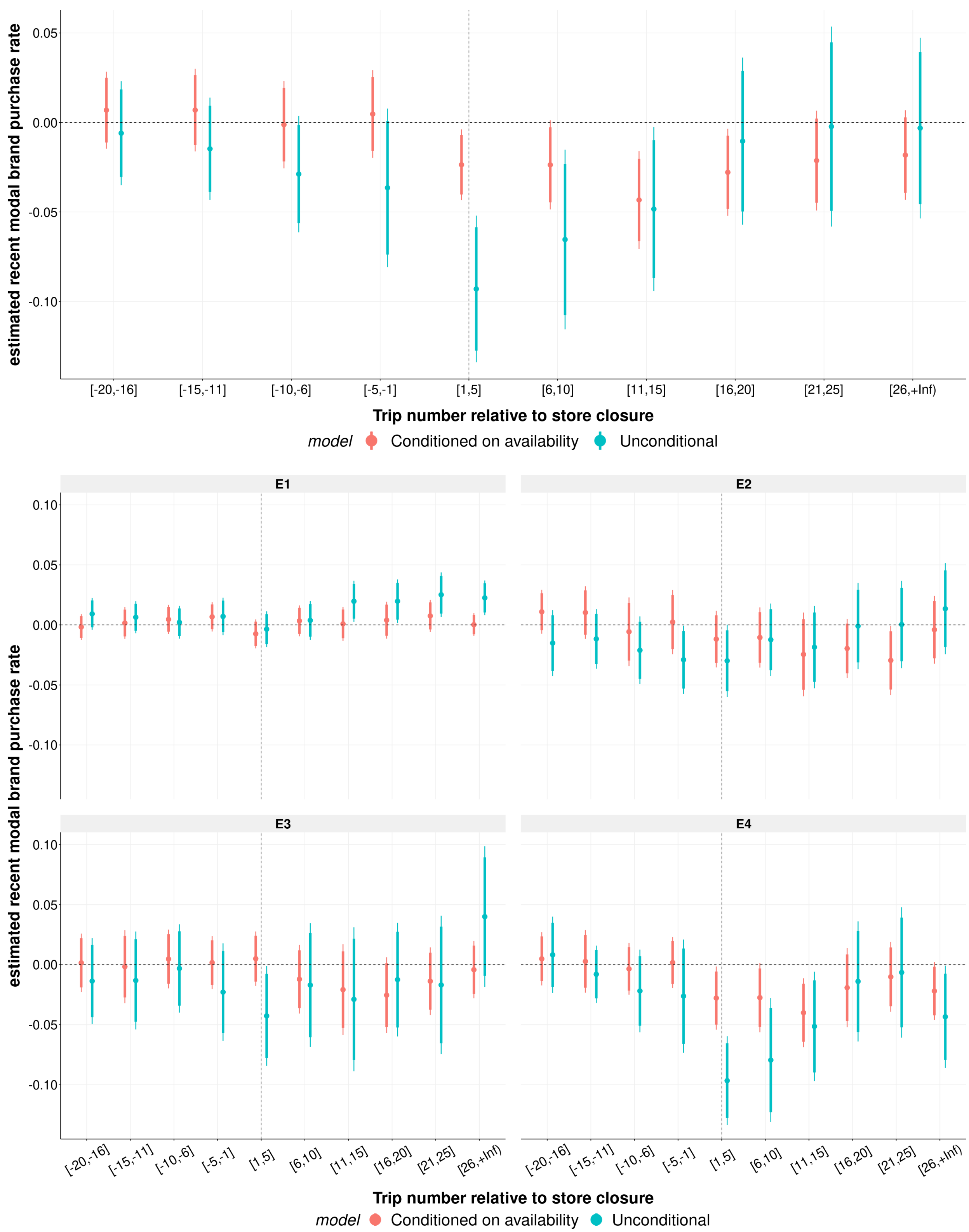
Note. Replication of the results in Figure 4 for $T_e = \frac{1}{2}$ year, and $L = 20$.

Table 5 DID results for recent modal brand.

	Dependent variable:			
	recent modal brand indicator (×100)			
	(1)	(2)	(3)	(4)
Overall, short-term	-4.568^{**}		-3.102^{***}	
	(1.715)		(0.764)	
Overall, long-term	0.582		-1.953	
	(2.300)		(1.037)	
E_1, short-term		0.621		-0.136
		(0.533)		(0.406)
E_2, short-term		-0.572		-1.767^{*}
		(1.039)		(0.826)
E_3, short-term		-2.084		-1.467
		(1.811)		(0.984)
E_4, short-term		-5.705^{***}		-2.855^{**}
		(1.621)		(0.869)
E_1, long-term		1.992^{**}		-0.076
		(0.644)		(0.400)
E_2, long-term		2.045		-0.775
		(1.604)		(1.183)
E_3, long-term		3.943		-0.340
		(2.677)		(1.059)
E_4, long-term		-3.579		-2.197^{*}
		(1.905)		(0.961)
Continuous treatment	✓		✓	
Conditioned on modal brand availability			✓	✓
Observations	899,487	899,487	547,614	547,614
R^2	0.279	0.279	0.145	0.145
Adjusted R^2	0.268	0.268	0.125	0.125
Note:				*p<0.05; **p<0.01; ***p<0.001

Replication of the results in table 1 for $T_e = 2$ years and $L = 20$.

Figure 18 Event study analyses results for aggregate (top), and different exposure groups (bottom) with the recent modal brand indicator as the outcome variable



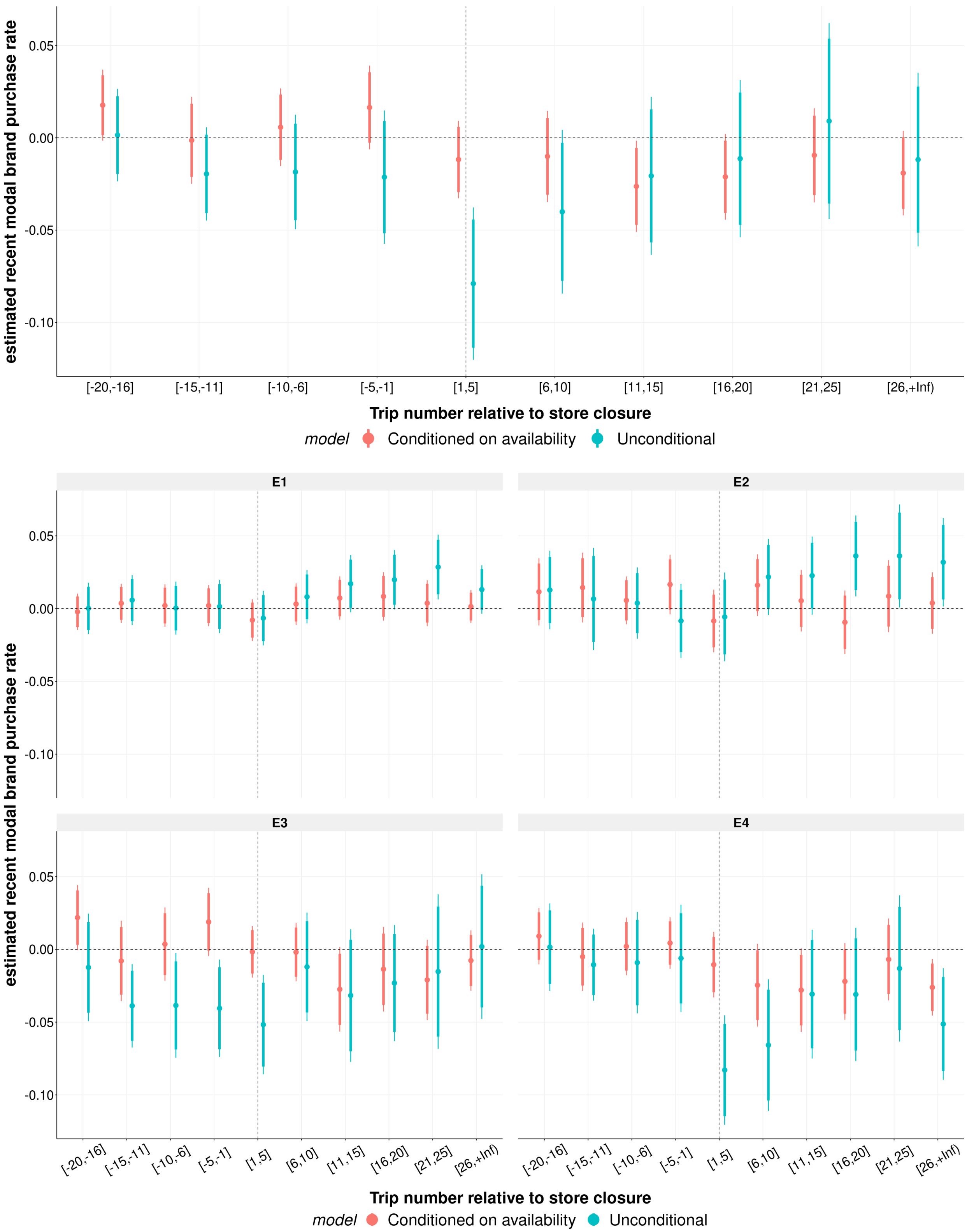
Note. Replication of the results in Figure 4 for $T_e = 2$ years, and $L = 20$.

Table 6 DID results for recent modal brand.

	Dependent variable:			
	recent modal brand indicator (×100)			
	output100			
	(1)	(2)	(3)	(4)
Overall, short-term	-4.900^{**}		-1.624^{*}	
	(1.626)		(0.789)	
Overall, long-term	-0.341		-2.457^{**}	
	(1.892)		(0.844)	
E_1, short-term		0.071		-0.235
		(0.677)		(0.509)
E_2, short-term		0.846		-0.153
		(0.953)		(0.679)
E_3, short-term		-1.589		-0.797
		(1.254)		(0.848)
E_4, short-term		-6.813^{***}		-1.844^{*}
		(1.551)		(0.893)
E_1, long-term		1.499^{*}		0.168
		(0.708)		(0.449)
E_2, long-term		3.047^{*}		-0.246
		(1.210)		(0.753)
E_3, long-term		0.967		-1.620
		(1.991)		(0.965)
E_4, long-term		-3.989^{*}		-2.588^{***}
		(1.553)		(0.711)
Continuous treatment	✓		✓	
Conditioned on modal brand availability			✓	✓
Observations	883,237	883,237	544,180	544,180
R^2	0.279	0.279	0.141	0.141
Adjusted R^2	0.268	0.268	0.121	0.121
Note:	*p<0.05; **p<0.01; ***p<0.001			

Replication of the results in Table 1 for $T_e = 1$ year, and $L = 10$.

Figure 19 **Event study analyses results for aggregate (top), and different exposure groups (bottom) with the recent modal brand indicator as the outcome variable**



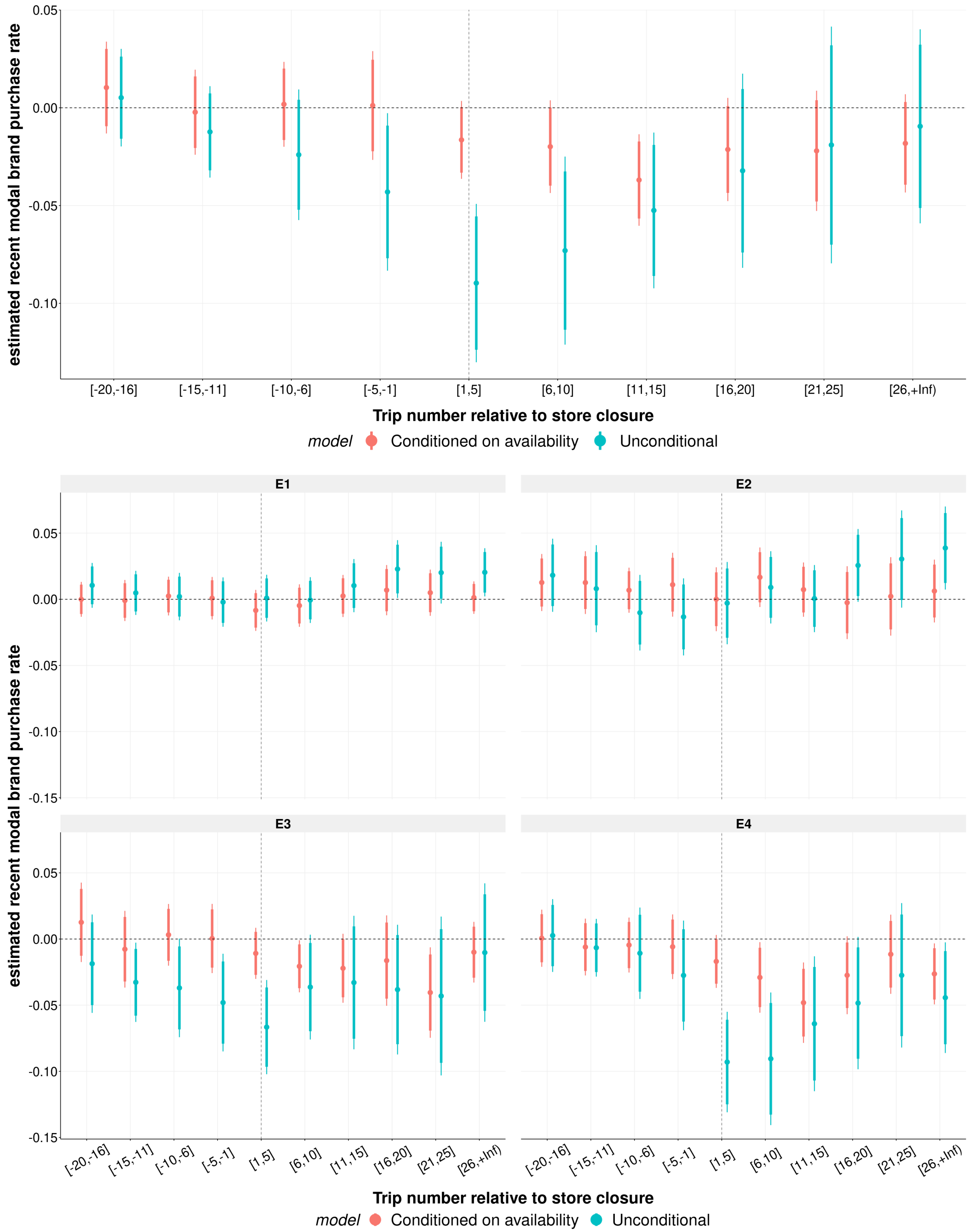
Note. Replication of the results in Figure 4 for $T_e = 1$ year, and $L = 10$.

Table 7 DID results for recent modal brand.

	Dependent variable:			
	recent modal brand indicator (×100)			
	(1)	(2)	(3)	(4)
Overall, short-term	-4.188^{*}		-2.364^{**}	
	(1.731)		(0.750)	
Overall, long-term	-0.400		-1.830	
	(2.206)		(1.066)	
E_1, short-term		1.123		0.118
		(0.684)		(0.518)
E_2, short-term		1.392		-0.191
		(1.025)		(0.818)
E_3, short-term		-2.393		-1.999^{*}
		(1.620)		(0.863)
E_4, short-term		-6.065^{***}		-2.495^{**}
		(1.615)		(0.783)
E_1, long-term		1.643		-0.119
		(0.866)		(0.572)
E_2, long-term		3.836^{**}		0.459
		(1.473)		(1.069)
E_3, long-term		0.662		-1.049
		(2.275)		(1.147)
E_4, long-term		-4.531^{*}		-2.393^{*}
		(2.013)		(0.981)
Continuous treatment	✓		✓	
Conditioned on modal brand availability			✓	✓
Observations	883,237	883,237	533,618	533,618
R^2	0.281	0.281	0.149	0.149
Adjusted R^2	0.270	0.270	0.129	0.129
Note:		$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$		

Replication of the results in Table 1 for $T_e = 1$ year and $L = 30$.

Figure 20 **Event study analyses results for aggregate (top), and different exposure groups (bottom) with the recent modal brand indicator as the outcome variable**



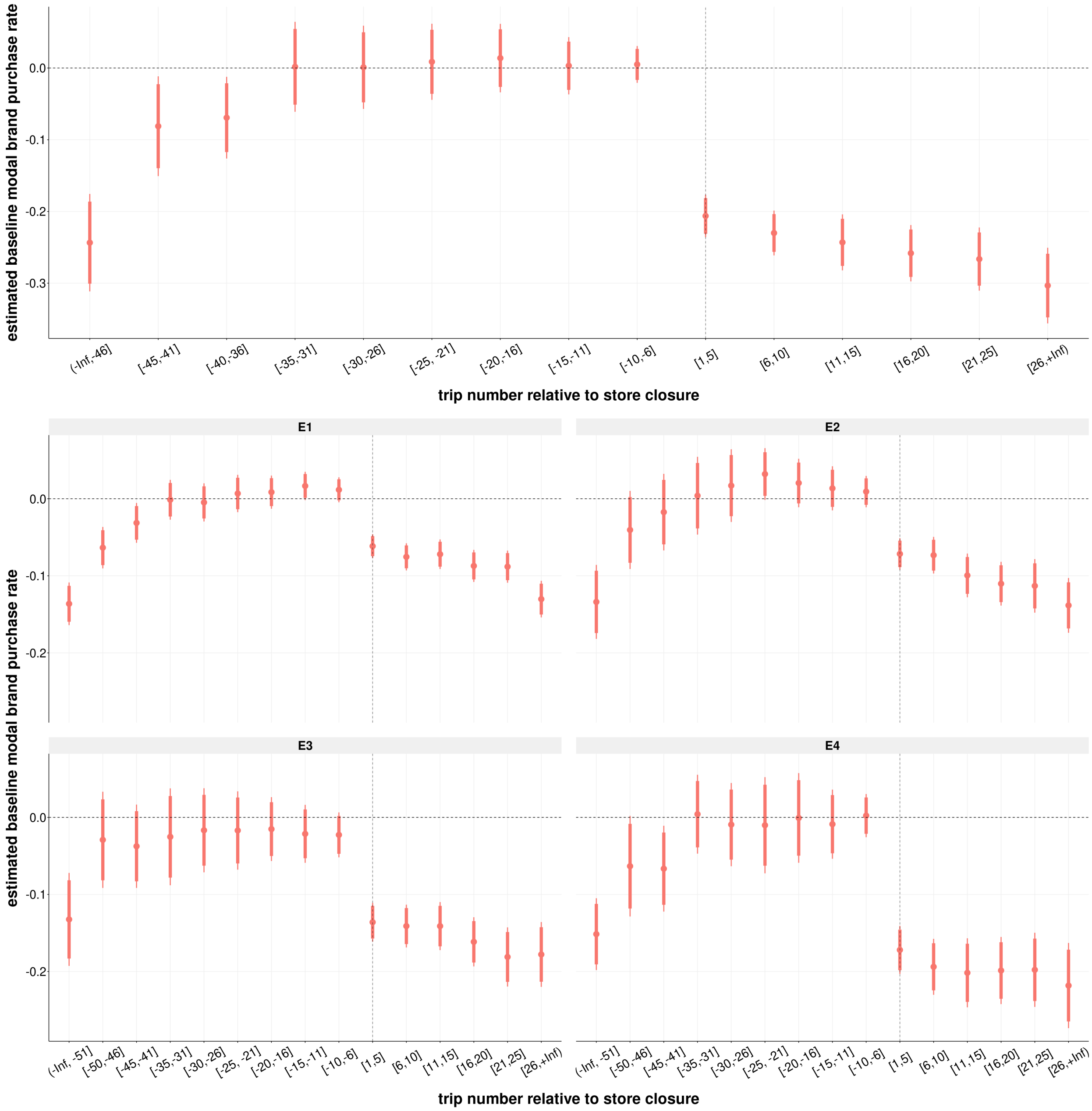
Note. Replication of the results in Figure 4 for $T_e = 1$ year, and $L = 30$.

Table 8 DID results for long-term modal brand.

	Dependent variable:		
	baseline modal brand indicator (×100)		
	(1)	(2)	(3)
Overall	-25.07^{***}	-8.09^{***}	
	(1.74)	(0.98)	
E_1			-5.47^{***}
			(1.07)
E_2			-8.10^{***}
			(1.59)
E_3			-11.62^{***}
			(2.20)
E_4			-15.99^{***}
			(1.78)
Continuous treatment	✓		
Observations	895,035	895,035	895,035
R^2	0.343	0.342	0.342
Adjusted R^2	0.333	0.332	0.333
Note:			$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Replication of the results in Table 2, with the baseline modal brand defined based on only 40 trips prior to each closure.

Figure 21 Event study analyses results for aggregate (top), and different exposure groups (bottom) with long-term modal brand indicator as the outcome variable



Note. Replication of the results in Figure 7, with the baseline modal brand defined based on only 40 trips prior to each closure.