

Web Appendix A Ignoring Brand Preference Confounding

In this Appendix, we present the results of naïve approaches to estimating state dependence effects. In particular, we compare the results reported in the main paper with two alternative approaches:

1. Regress brand repurchase (Y) on random coupon redemption (D)
2. Brand choice models that do not account for preference heterogeneity

These approaches fail to address the confounding effect of brand preferences, so we call the resulting effect naïve ATEs.

Based on the expectation that brand preferences have a positive impact on current and future brand purchases, as well as extant brand choice literature (see, e.g., [Keane 1997](#), [Chintagunta et al. 2005](#), [Goldfarb 2006](#)), ignoring brand preferences in the estimation of the state dependence effects should lead to an *upward bias*. To make results fully comparable, we use our Bayesian hierarchical model for evidence aggregation across brands for each specification.

Figure A1: Comparison of Shrunk Estimates

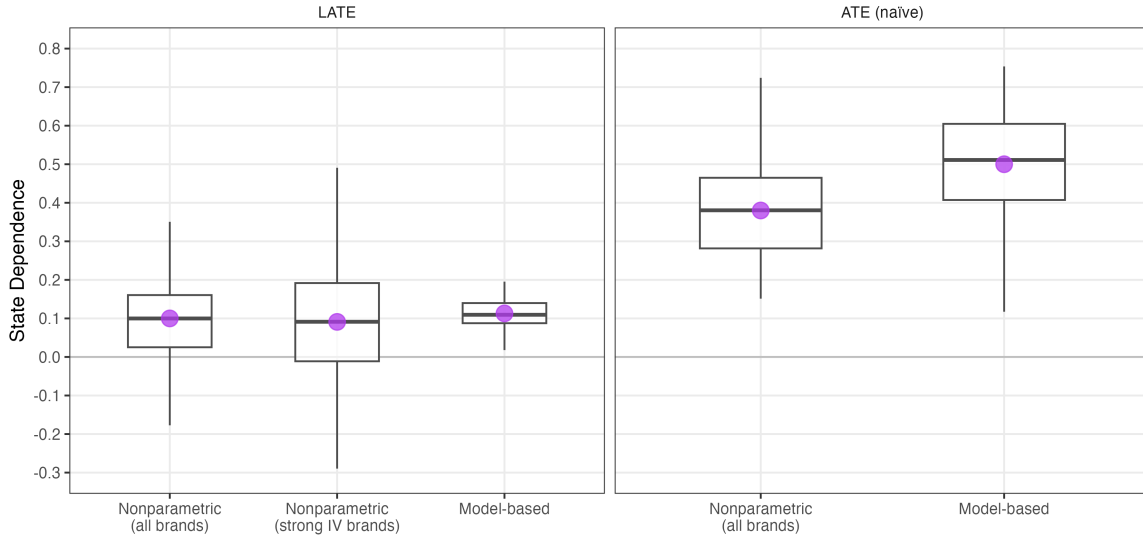


Figure A1 shows the distributions of the results as boxplots. The magenta points highlight the posterior means of the average LATE (or ATE) values across brands. The comparison confirms the concern that unobserved brand preferences are easily mistaken for true state dependence.

When we regress future brand purchase on current brand purchase (or estimate a discrete choice model without preference heterogeneity), the resulting naïve ATE is *roughly four to five times larger* than the causal estimate, irrespective of the approach to obtain the results

(i.e., experiment vs. choice model). Once we either (i) isolate the exogenous variation in purchases that is induced by random coupons (yielding the experimental LATE), or (ii) allow for preference heterogeneity in the structural choice model (yielding the choice model LATE), the average effect sizes are very similar across approaches (close to 0.1).

The convergence in results further increases our confidence in the validity of the results and lends credibility to the choice model's results, which rely on stronger assumptions.

Web Appendix B State Dependence and Consumer Learning

The pioneering work of [Erdem and Keane \(1996\)](#) spawned a large literature on consumer learning. They develop a choice model where consumers learn unknown product qualities from consumption, and contend that this learning process offers a micro-founded explanation for structural state dependence. In this section, we examine whether consumer learning is the main source of the structural state dependence that we documented in the previous sections. If state dependence generates inertia in brand choice beyond what can be explained by learning, then marketing investments are justified even for well-informed consumers, because a purchase increases switching costs and thereby raises the likelihood of repurchase.

Directly testing for learning as a causal mechanism is not possible because the quality beliefs of consumers are unobserved. Instead, we pursue a more indirect approach and implement a falsification test (cf. [Goldfarb et al. 2022](#)). Consider two consumer types: (i) imperfectly informed consumers, who may update beliefs after purchasing and consuming; and (ii) approximately perfectly informed consumers, who no longer update beliefs after purchase. If learning were the sole driver of long-run effects, we should observe no positive state-dependence effect among (ii). Evidence of positive effects in (ii) implies that learning cannot fully account for the state dependence.²⁶

To operationalize this idea, we proxy informed consumers in two ways. First, we classify customers with at least one purchase of the brand in the past 90 days as informed. Second, we use the retailer’s recommender system and calculate treatment effects for customers targeted by the retailer’s recommender engine. By design, the algorithm targets coupons to customers who are a good match and most likely to purchase the brand (see Section 3.2). Targeted customers have substantially higher baseline purchase rates for the focal brand than non-targeted customers ($\Delta r = 0.089$, $SE = 0.005$), suggesting they are, on average, better informed. We therefore analyze the intended targets ($\hat{T} = 1$) as a second informed group. Both classifications are proxies and thus rest on an explicit *assumption* that recent purchasers and intended targets are (approximately) informed.

Building on the identification approach in Section 4, we estimate the LATEs of the two focal populations, $LATE_{D \rightarrow Y, \sum_{90d} Y > 0}$ and $LATE_{D \rightarrow Y, \hat{T}=1}$.²⁷ Table A1 shows the results. For both groups, and regardless of the brand set (all brands vs. strong IV brands), we find evidence for structural state dependence, i.e., a posterior probability of a positive mean effect, $Pr\{\hat{\tau} > 0\}$, close to one. The effect sizes are similar to, and for intended targets possibly larger than, the average effect sizes reported in Section 6. Hence, assuming that learning among customers included in these analyses is minimal, the results imply that consumer learning is *not* the sole mechanism behind the observed state dependence.

²⁶[Dubé et al. \(2010\)](#) perform an analysis based on a similar idea, but the implementation differs from ours. [Ching et al. \(2013\)](#) critique their specific test; however, this criticism does not apply to our analysis.

²⁷See Web Appendix C for the details and a formal proof.

Table A1: Bayesian Meta-Analysis Results for Informed Customers

	All brands			Strong IV brands		
	Post. mean	95% CI	PD	Post. mean	95% CI	PD
Customers with brand purchase in past 90 days						
Mean (τ)	0.080	[−0.001, 0.161]	0.973	0.104	[−0.012, 0.214]	0.959
Std. dev. (σ)	0.142	[0.017, 0.264]	–	0.199	[0.047, 0.361]	–
Customers targeted by the recommender engine						
Mean (τ)	0.193	[0.025, 0.363]	0.987	0.192	[−0.050, 0.426]	0.938
Std. dev. (σ)	0.364	[0.198, 0.551]	–	0.442	[0.238, 0.688]	–

Notes: The sample sizes are 221 and 223 for “all brands” and 36 and 27 for “strong IV brands.” A brand is classified as strong IV if $ATE_{Z \rightarrow D}$ is statistically different from zero with an F -statistic greater than 10 (Stock and Yogo 2005). PD (probability of direction) is the proportion of the posterior distribution that is of the median’s sign.

Web Appendix C Identification and Inference

C.1 Data-Generating Process: Details

Before presenting the proof of the main identification proposition, we add some details to the overview of the data-generating process in Section (4.2).

Recall that $\hat{T} \in \{0, 1\}$ indicates whether the focal brand coupon is included among a customer's *intended* target coupons. Further, the *i.i.d.* Bernoulli draw $N \in \{0, 1\}$ captures whether a random coupon is drawn ($N = 1$ if there is **no** random coupon).

We use potential outcome notation for $R(n)$, $H(n)$, and $T(n)$, $n \in \{0, 1\}$, defined with respect to N . In the randomization condition ($N = 0$), $R(0)$ indicates the receipt of a random focal brand coupon. Further, $H(0)$ indicates if a focal brand target coupon is *hit* and replaced by a focal or non-focal brand random coupon. $T(0)$ denotes exposure to a focal brand target coupon.

Without randomization ($N = 1$), the realized target coupon is the intended target coupon, $T(1) = \hat{T}$, and $R(1) = H(1) = 0$. Exposure to a focal brand target coupon when $N = 0$ is given by

$$T(0) = (1 - H(0)) \cdot (1 - R(0)) \cdot T(1)$$

Note that $R(0)$ is an *i.i.d.* Bernoulli draw. However, $H(0)$ is an *i.i.d.* Bernoulli draw only if $T(1) = 1$, whereas $H(0) = 0$ if $T(1) = 0$ (if there is no intended target coupon it cannot be hit). Hence, $T(1)$ and $H(0)$ are dependent.

In the data, we observe N and

$$\begin{aligned} T &= NT(1) + (1 - N)T(0) \\ R &= NR(1) + (1 - N)R(0) \end{aligned}$$

However, $\hat{T}$ and $H(0)$ are unobserved.

C.2 Identification Proof

C.2.1 Identifying Conditions

In Section 4, we already noted the **exclusion** and **monotonicity** conditions: $Y(d, z) = Y(d)$ for all $d, z \in \{0, 1\}$, and $D(1) \geq D(0)$.

Further, exposure to a focal brand target coupon is monotonic in N , $T(1) \geq T(0)$, because intended target coupons can only be replaced but not randomly added when $N = 0$.

The focal brand target and random coupon assignment mechanism discussed in Section C.1 implies the following conditional independence conditions:

- (1) $Y(0), Y(1), D(0), D(1), H(0), T(1) \perp\!\!\!\perp N, R(0)$
- (2) $Y(0), Y(1), D(0), D(1) \perp\!\!\!\perp N, R(0), H(0), T(0) \mid T(1) = 1$
- (3) $Y(0), Y(1), D(0), D(1), H(0), T(0), T(1) \perp\!\!\!\perp N \mid R(0)$
- (4) $Y(0), Y(1), D(0), D(1), H(0), T(1) \perp\!\!\!\perp N, R(0), R$
- (5) $Y(0), Y(1), D(0), D(1) \perp\!\!\!\perp N, R(0), R, H(0) \mid T(1) = 1, T(0)$

Regarding condition (2), recall that $T(1)$ and $H(0)$ are dependent. Hence, in general, $H(0)$ and the potential outcomes $Y(d)$ and $D(z)$ are dependent, too. However, conditional on $T(1) = 1$, $H(0)$ is an *i.i.d.* Bernoulli draw. Further, the conditional independence of $T(0)$ from the potential outcomes follows because $T(0) = (1 - H(0)) \cdot (1 - R(0))$ conditional on $T(1) = 1$.

The third condition follows because $T(0) = (1 - H(0)) \cdot (1 - R(0)) \cdot T(1)$ is a function of $H(0)$ and $T(1)$ conditional on $R(0)$.

(4) follows from (1) because $R = NR(1) + (1 - N)R(0) = (1 - N)R(0)$. Similarly, (5) follows from (2), the weak union rule, and because R is a function of N and $R(0)$.

C.2.2 Identification of the Average Treatment Effect of the Treated: $ATT_{T=1, Z \rightarrow D}$

Given the conditional independence conditions, the average treatment effect of Z on D for the subgroup of customers who are exposed to a focal brand target coupon, $T = 1$, is identified.

Proposition.

$$\begin{aligned} ATT_{T=1, Z \rightarrow D} &\equiv \mathbb{E}[D(1) - D(0) \mid T = 1] \\ &= \mathbb{E}[D(1) - D(0) \mid T(1) > T(0), R(0) = 0] \end{aligned}$$

Proof. Note that customers are exposed to a target coupon for the focal brand, $T = 1$, if and only if either $T(0) = T(1) = 1$ or if $T(0) = 0, T(1) = 1$, and $N = 1$. Formally,

$$\{T = 1\} = \{T(0) = 1, T(1) = 1\} \cup \{T(0) = 0, T(1) = 1, N = 1\}$$

Then

$$\begin{aligned} ATT_{T=1, Z \rightarrow D} &\equiv \mathbb{E}[D(1) - D(0) \mid T = 1] \\ &= \mathbb{E}[D(1) - D(0) \mid T(0) = 1, T(1) = 1] \cdot \Pr\{T(0) = 1, T(1) = 1 \mid T = 1\} \\ &\quad + \mathbb{E}[D(1) - D(0) \mid T(0) = 0, T(1) = 1, N = 1] \cdot \Pr\{T(0) = 0, T(1) = 1, N = 1 \mid T = 1\} \\ &= \mathbb{E}[D(1) - D(0) \mid T(0) = 1, T(1) = 1] \cdot \Pr\{T(0) = 1, T(1) = 1 \mid T = 1\} \\ &\quad + \mathbb{E}[D(1) - D(0) \mid T(0) = 0, T(1) = 1] \cdot \Pr\{T(0) = 0, T(1) = 1, N = 1 \mid T = 1\} \\ &= \mathbb{E}[D(1) - D(0) \mid T(1) = 1] \cdot \Pr\{T(0) = 1, T(1) = 1 \mid T = 1\} \\ &\quad + \mathbb{E}[D(1) - D(0) \mid T(1) = 1] \cdot \Pr\{T(0) = 0, T(1) = 1, N = 1 \mid T = 1\} \\ &= \mathbb{E}[D(1) - D(0) \mid T(1) = 1] \\ &= \mathbb{E}[D(1) - D(0) \mid T(1) = 1, T(0) = 0] \\ &= \mathbb{E}[D(1) - D(0) \mid T(1) = 1, T(0) = 0, R(0) = 0] \\ &= \mathbb{E}[D(1) - D(0) \mid T(1) > T(0), R(0) = 0] \end{aligned}$$

□

The first equality uses the law of iterated expectations. Note that $T = 1$ is redundant given $T(0)$, $T(1)$, and N and hence omitted from the conditioning variables in the two conditional expectations. The second equality follows from the weak union rule and condition (2). The third and fifth equalities follow from condition (2). The fourth equality uses the fact noted above,

$$\{T = 1\} = \{T(0) = 1, T(1) = 1\} \cup \{T(0) = 0, T(1) = 1, N = 1\}$$

The sixth equality follows from condition (5). Note that in the last two lines of the proof $R(0)$ can be replaced with R due to condition (5).

Intuitively, the group of customers for whom $T(1) > T(0)$ is a random subset of the group $T(1) = 1$, and the group $T(1) > T(0)$ and $R(0) = 0$ is a random subset of the $T(1) > T(0)$ group. Because the treated customers, $T = 1$, are also a random subset of $T(1) = 1$, the distribution of $D(z)$ and hence the expectation of $D(1) - D(0)$ is identical across these groups.

We now show that $\text{ATT}_{T=1, Z \rightarrow D}$ is identified.

Define

$$\begin{aligned} p_0 &\equiv \Pr\{T(1) = T(0) | R(0) = 0\} \\ p_1 &\equiv \Pr\{T(1) > T(0) | R(0) = 0\} \end{aligned}$$

Then

$$\begin{aligned} \mathbb{E}[T | N = 1, R = 0] - \mathbb{E}[T | N = 0, R = 0] &= \mathbb{E}[T(1) | N = 1, R = 0] - \mathbb{E}[T(0) | N = 0, R(0) = 0] \\ &= \mathbb{E}[T(1) | N = 1] - \mathbb{E}[T(0) | N = 0, R(0) = 0] \\ &= \mathbb{E}[T(1)] - \mathbb{E}[T(0) | R(0) = 0] \\ &= \mathbb{E}[T(1) | R(0) = 0] - \mathbb{E}[T(0) | R(0) = 0] \\ &= \mathbb{E}[T(1) - T(0) | R(0) = 0] \\ &= \mathbb{E}[T(1) - T(0) | T(1) = T(0), R(0) = 0] \cdot p_0 \\ &\quad + \mathbb{E}[T(1) - T(0) | T(1) > T(0), R(0) = 0] \cdot p_1 \\ &= 0 \cdot p_0 + \mathbb{E}[1 | T(1) > T(0), R(0) = 0] \cdot p_1 \\ &= \Pr\{T(1) > T(0) | R(0) = 0\} \end{aligned}$$

The first equality follows from consistency, and the second equality follows because $\{N = 1, R = 0\} = \{N = 1\}$. The third and fourth equalities follow from conditions (1) and (3), and the sixth equality uses the law of iterated expectations and the monotonicity of $T(\cdot)$ in N .

Similarly,

$$\begin{aligned}
& \mathbb{E}[D|N = 1, R = 0] - \mathbb{E}[D|N = 0, R = 0] \\
&= \mathbb{E}[D(T(1))|N = 1, R = 0] - \mathbb{E}[D(T(0))|N = 0, R(0) = 0] \\
&= \mathbb{E}[D(T(1))|N = 1] - \mathbb{E}[D(T(0))|N = 0, R(0) = 0] \\
&= \mathbb{E}[D(T(1))] - \mathbb{E}[D(T(0))|R(0) = 0] \\
&= \mathbb{E}[D(T(1))|R(0) = 0] - \mathbb{E}[D(T(0))|R(0) = 0] \\
&= \mathbb{E}[D(T(1)) - D(T(0))|R(0) = 0] \\
&= \mathbb{E}[D(T(1)) - D(T(0))|T(1) = T(0), R(0) = 0] \cdot p_0 \\
&\quad + \mathbb{E}[D(T(1)) - D(T(0))|T(1) > T(0), R(0) = 0] \cdot p_1 \\
&= 0 \cdot p_0 + \mathbb{E}[D(1) - D(0)|T(1) > T(0), R(0) = 0] \cdot p_1 \\
&= \mathbb{E}[D(1) - D(0)|T(1) > T(0), R(0) = 0] \cdot \Pr\{T(1) > T(0)|R(0) = 0\}
\end{aligned}$$

It follows that $ATT_{T=1, Z \rightarrow D}$ is identified from the observed data (D, T, N, R) :

$$\begin{aligned}
ATT_{T=1, Z \rightarrow D} &= \mathbb{E}[D(1) - D(0)|T(1) > T(0), R(0) = 0] \\
&= \frac{\mathbb{E}[D|N = 1, R = 0] - \mathbb{E}[D|N = 0, R = 0]}{\mathbb{E}[T|N = 1, R = 0] - \mathbb{E}[T|N = 0, R = 0]}
\end{aligned}$$

C.2.3 Identification of the Intent-To-Treat Effect of the Treated: $ITT_{T=1, Z \rightarrow Y}$

The analysis is analogous to the derivation of the $ATT_{T=1, Z \rightarrow D}$ above, and we can show that

$$\begin{aligned}
ITT_{T=1, D \rightarrow Y} &\equiv \mathbb{E}[Y(D(1)) - Y(D(0))|T = 1] \\
&= \mathbb{E}[Y(D(1)) - Y(D(0))|T(1) > T(0), R(0) = 0]
\end{aligned}$$

Moreover, it follows that $ITT_{T=1, Z \rightarrow Y}$ is identified from the observed data (Y, T, N, R) :

$$\begin{aligned}
ITT_{T=1, Z \rightarrow Y} &= \mathbb{E}[Y(D(1)) - Y(D(0))|T(1) > T(0), R(0) = 0] \\
&= \frac{\mathbb{E}[Y|N = 1, R = 0] - \mathbb{E}[Y|N = 0, R = 0]}{\mathbb{E}[T|N = 1, R = 0] - \mathbb{E}[T|N = 0, R = 0]}
\end{aligned}$$

C.2.4 Identification of the Average Treatment Effect: $ATE_{Z \rightarrow D}$

Proposition.

$$\begin{aligned}
ATE_{Z \rightarrow D} &\equiv \mathbb{E}[D(1) - D(0)] \\
&= \mathbb{E}[D|R = 1] - \mathbb{E}[D|R = 0] + \Pr\{T = 1|R = 0\} \cdot ATT_{T=1, Z \rightarrow D}
\end{aligned}$$

Proof.

$$\begin{aligned}
\mathbb{E}[D|R = 1] &= \mathbb{E}[D(1)|R = 1] \\
&= \mathbb{E}[D(1)] \\
\mathbb{E}[D|R = 0] &= \mathbb{E}[D|T = 1, R = 0] \cdot \Pr\{T = 1|R = 0\} \\
&\quad + \mathbb{E}[D|T = 0, R = 0] \cdot \Pr\{T = 0|R = 0\} \\
&= \mathbb{E}[D(1)|T = 1, R = 0] \cdot \Pr\{T = 1|R = 0\} \\
&\quad + \mathbb{E}[D(0)|T = 0, R = 0] \cdot \Pr\{T = 0|R = 0\} \\
&= \mathbb{E}[D(1) - D(0)|T = 1, R = 0] \cdot \Pr\{T = 1|R = 0\} \\
&\quad + \mathbb{E}[D(0)|T = 1, R = 0] \cdot \Pr\{T = 1|R = 0\} \\
&\quad + \mathbb{E}[D(0)|T = 0, R = 0] \cdot \Pr\{T = 0|R = 0\} \\
&= \mathbb{E}[D(0)] + \mathbb{E}[D(1) - D(0)|T = 1, R = 0] \cdot \Pr\{T = 1|R = 0\} \\
&= \mathbb{E}[D(0)] + \mathbb{E}[D(1) - D(0)|T = 1] \cdot \Pr\{T = 1|R = 0\} \\
&= \mathbb{E}[D(0)] + \Pr\{T = 1|R = 0\} \cdot \text{ATT}_{T=1, Z \rightarrow D}
\end{aligned}$$

The independence of $D(0)$ and $D(1)$ from R follows from condition (4). The second-to-last equality follows from condition (5).

It follows that

$$\begin{aligned}
\mathbb{E}[D(1)] &= \mathbb{E}[D|R = 1] \\
\mathbb{E}[D(0)] &= \mathbb{E}[D|R = 0] - \Pr\{T = 1|R = 0\} \cdot \text{ATT}_{T=1, Z \rightarrow D}
\end{aligned}$$

Subtracting the second from the first equation proves the proposition. $\square$

Identification follows because the proposition shows how the $\text{ATT}_{T=1, Z \rightarrow D}$ can be constructed from four identified quantities.

C.2.5 Identification of the Intent-To-Treat Effect: $\text{ITT}_{Z \rightarrow Y}$

As in the previous subsection, we can prove that:

$$\begin{aligned}
\text{ITT}_{Z \rightarrow Y} &\equiv \mathbb{E}[Y(D(1)) - Y(D(0))] \\
&= \mathbb{E}[Y|R = 1] - \mathbb{E}[Y|R = 0] + \Pr\{T = 1|R = 0\} \cdot \text{ITT}_{T=1, Z \rightarrow Y}
\end{aligned}$$

The proof follows the same logic as the identification proof for $\text{ATE}_{Z \rightarrow D}$. Again, identification follows because all the quantities to construct $\text{ITT}_{Z \rightarrow Y}$ are identified.

C.2.6 Identification of the Local Average Treatment Effect: $\text{LATE}_{D \rightarrow Y}$

The monotonicity condition, $D(1) \geq D(0)$, implies that there are two groups of customers. The first group, called *compliers*, purchases with a coupon but not without a coupon, $D(1) > D(0)$. The second group does not change their purchasing behavior in response to a coupon, $D(1) = D(0)$.

Hence, the intent-to-treat effect is

$$\begin{aligned}
\text{ITT}_{Z \rightarrow Y} &\equiv \mathbb{E}[Y(D(1)) - Y(D(0))] \\
&= \mathbb{E}[Y(D(1)) - Y(D(0)) | D(1) > D(0)] \cdot \Pr\{D(1) > D(0)\} \\
&\quad + \mathbb{E}[Y(D(1)) - Y(D(0)) | D(1) = D(0)] \cdot \Pr\{D(1) = D(0)\} \\
&= \mathbb{E}[Y(1) - Y(0) | D(1) > D(0)] \cdot \Pr\{D(1) > D(0)\}
\end{aligned}$$

Similarly,

$$\begin{aligned}
\text{ATE}_{Z \rightarrow D} &\equiv \mathbb{E}[D(1) - D(0)] \\
&= \mathbb{E}[D(1) - D(0) | D(1) > D(0)] \cdot \Pr\{D(1) > D(0)\} \\
&\quad + \mathbb{E}[D(1) - D(0) | D(1) = D(0)] \cdot \Pr\{D(1) = D(0)\} \\
&= \mathbb{E}[1 | D(1) > D(0)] \cdot \Pr\{D(1) > D(0)\} \\
&= \Pr\{D(1) > D(0)\}
\end{aligned}$$

Taking ratios of the $\text{ITT}_{Z \rightarrow Y}$ and $\text{ATE}_{Z \rightarrow D}$ we obtain a general expression for the LATE:

$$\begin{aligned}
\text{LATE}_{D \rightarrow Y} &\equiv \mathbb{E}[Y(1) - Y(0) | D(1) > D(0)] \\
&= \frac{\text{ITT}_{Z \rightarrow Y}}{\text{ATE}_{Z \rightarrow D}} \tag{14}
\end{aligned}$$

(14) applies generally, both in the [Imbens and Angrist \(1994\)](#) framework and in our context.

Specifically, using the expressions for $\text{ITT}_{Z \rightarrow Y}$ and $\text{ATE}_{Z \rightarrow D}$ derived above, we obtain the LATE estimator for our application that we presented in [Section 4.3](#):

$$\text{LATE}_{D \rightarrow Y} = \frac{\mathbb{E}[Y | R = 1] - \mathbb{E}[Y | R = 0] + \Pr\{T = 1 | R = 0\} \cdot \text{ITT}_{T=1, Z \rightarrow Y}}{\mathbb{E}[D | R = 1] - \mathbb{E}[D | R = 0] + \Pr\{T = 1 | R = 0\} \cdot \text{ATT}_{T=1, Z \rightarrow D}}$$

C.3 Additional Details

Note that we can further simplify the equation above using the following fact of our experimental design. Let p be the probability that a focal target coupon is replaced by any random coupon, i.e., $p \equiv \Pr\{H(0) = 1 | T(1) = 1\}$. From the identifying conditions above, we know that the following holds.

$$\begin{aligned}
\Pr\{T = 1 | N = 0, R = 0\} &= \Pr\{T = 1 | N = 0, R(0) = 0\} \\
&= \Pr\{T(0) = 1, T(1) = 1 | N = 0, R(0) = 0\} \\
&= \Pr\{H(0) = 0, T(1) = 1 | N = 0, R(0) = 0\} \\
&= \Pr\{H(0) = 0 | N = 0, R(0) = 0, T(1) = 1\} \\
&\quad \times \Pr\{T(1) = 1 | N = 0, R(0) = 0\} \\
&= (1 - p) \times \Pr\{T = 1 | N = 1\}
\end{aligned}$$

The third equality holds because $T(0) = (1 - H(0)) \cdot (1 - R(0))$ when $T(1) = 1$. The final equality follows from conditions (1) and (2).

Therefore,

$$\frac{Pr\{T = 1|N = 0, R = 0\}}{Pr\{T = 1|N = 1\}} = 1 - p$$

Now, let's simplify $LATE_{D \rightarrow Y}$. We further define the following quantities,

$$\begin{aligned} A &\equiv \mathbb{E}[D|R = 1] - \mathbb{E}[D|N = 1] \\ B &\equiv \mathbb{E}[D|N = 1] - \mathbb{E}[D|N = 0, R = 0] \end{aligned}$$

Then, the denominator of $LATE_{D \rightarrow Y}$ becomes,

$$\begin{aligned} \text{Denominator} &= \mathbb{E}[D|R = 1] - \mathbb{E}[D|R = 0] + Pr\{T = 1|R = 0\} \cdot ATT_{T=1, Z \rightarrow D} \\ &= \mathbb{E}[D|R = 1] - [Pr\{N = 0|R = 0\} \cdot \mathbb{E}[D|N = 0, R = 0] \\ &\quad + (1 - Pr\{N = 0|R = 0\}) \cdot \mathbb{E}[D|N = 1]] + Pr\{T = 1|R = 0\} \cdot ATT_{T=1, Z \rightarrow D} \\ &= A + Pr\{N = 0|R = 0\} [\mathbb{E}[D|N = 1] - \mathbb{E}[D|N = 0, R = 0]] \\ &\quad + Pr\{T = 1|R = 0\} \frac{E[D|N = 1] - E[D|R = 0, N = 0]}{E[T|N = 1] - E[T|R = 0, N = 0]} \\ &= A + B \left[Pr\{N = 0|R = 0\} \right. \\ &\quad \left. + \frac{Pr\{N = 0|R = 0\} \cdot E[T|N = 0, R = 0] + (1 - Pr\{N = 0|R = 0\}) \cdot E[T|N = 1]}{E[T|N = 1] - E[T|R = 0, N = 0]} \right] \\ &= A + B \left[\frac{\mathbb{E}[T|N = 1]}{\mathbb{E}[T|N = 1] - \mathbb{E}[T|N = 0, R = 0]} \right] \\ &= A + B \cdot \frac{1}{p} \\ &= \mathbb{E}[D|R = 1] - \mathbb{E}[D|N = 1] + \frac{1}{p} \cdot (\mathbb{E}[D|N = 1] - \mathbb{E}[D|N = 0, R = 0]) \end{aligned}$$

Note that the second, third, fourth, and fifth equalities follow from the law of iterated expectations, the identification result for $ATT_{T=1, Z \rightarrow D}$ and the fact that $N = 1 \implies R = 0$. And the second-to-last equality holds because

$$\frac{Pr\{T = 1|N = 0, R = 0\}}{Pr\{T = 1|N = 1\}} = 1 - p$$

Therefore,

$$LATE_{D \rightarrow Y} = \frac{\mathbb{E}[Y|R = 1] - \mathbb{E}[Y|N = 1] + \frac{1}{p} \cdot (\mathbb{E}[Y|N = 1] - \mathbb{E}[Y|N = 0, R = 0])}{\mathbb{E}[D|R = 1] - \mathbb{E}[D|N = 1] + \frac{1}{p} \cdot (\mathbb{E}[D|N = 1] - \mathbb{E}[D|N = 0, R = 0])}$$

Note that the denominator of $LATE_{D \rightarrow Y}$ follows the approach of [Gabel et al. \(2025, p. 15, eq. \(1\)\)](#).

Identification of the Local Average Treatment Effect of the Treated: $\text{LATE}_{\hat{T}=1, D \rightarrow Y}$

In the identification proof for $\text{ATT}_{T=1, Z \rightarrow D}$, we showed that

$$\begin{aligned}\text{ATT}_{T=1, Z \rightarrow D} &= \mathbb{E}[D(1) - D(0) | T(1) > T(0), R(0) = 0] \\ &= \mathbb{E}[D(1) - D(0) | T(1) = 1, T(0) = 0, R(0) = 0]\end{aligned}$$

Therefore, by the condition (2),

$$\text{ATT}_{T=1, Z \rightarrow D} \equiv \mathbb{E}[D(1) - D(0) | T = 1] = \mathbb{E}[D(1) - D(0) | T(1) = 1]$$

Because $T(1) = \hat{T}$, it follows that

$$\text{ATT}_{T=1, Z \rightarrow D} = \mathbb{E}[D(1) - D(0) | \hat{T} = 1]$$

Define

$$\begin{aligned}\pi_{0, \hat{T}=1} &\equiv \Pr\{D(1) = D(0) | \hat{T} = 1\} \\ \pi_{1, \hat{T}=1} &\equiv \Pr\{D(1) > D(0) | \hat{T} = 1\}\end{aligned}$$

Then:

$$\begin{aligned}\text{ATT}_{T=1, Z \rightarrow D} &= \mathbb{E}[D(1) - D(0) | \hat{T} = 1] \\ &= \mathbb{E}[D(1) - D(0) | D(1) = D(0), \hat{T} = 1] \cdot \pi_{0, \hat{T}=1} \\ &\quad + \mathbb{E}[D(1) - D(0) | D(1) > D(0), \hat{T} = 1] \cdot \pi_{1, \hat{T}=1} \\ &= 0 \cdot \pi_{0, \hat{T}=1} + \mathbb{E}[1 | D(1) > D(0), \hat{T} = 1] \cdot \pi_{1, \hat{T}=1} \\ &= \Pr\{D(1) > D(0) | \hat{T} = 1\}\end{aligned}$$

Further,

$$\begin{aligned}\text{ITT}_{T=1, Z \rightarrow Y} &= \mathbb{E}[Y(D(1)) - Y(D(0)) | \hat{T} = 1] \\ &= \mathbb{E}[Y(D(1)) - Y(D(0)) | D(1) = D(0), \hat{T} = 1] \cdot \pi_{0, \hat{T}=1} \\ &\quad + \mathbb{E}[Y(D(1)) - Y(D(0)) | D(1) > D(0), \hat{T} = 1] \cdot \pi_{1, \hat{T}=1} \\ &= 0 \cdot \pi_{0, \hat{T}=1} + \mathbb{E}[Y(1) - Y(0) | D(1) > D(0), \hat{T} = 1] \cdot \pi_{1, \hat{T}=1} \\ &= \mathbb{E}[Y(1) - Y(0) | D(1) > D(0), \hat{T} = 1] \cdot \Pr\{D(1) > D(0) | \hat{T} = 1\}\end{aligned}$$

Therefore,

$$\begin{aligned}\text{LATE}_{\hat{T}=1, D \rightarrow Y} &= \mathbb{E}[Y(1) - Y(0) | D(1) > D(0), \hat{T} = 1] \\ &= \frac{\text{ITT}_{T=1, Z \rightarrow Y}}{\text{ATT}_{T=1, Z \rightarrow D}} \\ &= \frac{\mathbb{E}[Y | N = 1, R = 0] - \mathbb{E}[Y | N = 0, R = 0]}{\mathbb{E}[D | N = 1, R = 0] - \mathbb{E}[D | N = 0, R = 0]}\end{aligned}$$

C.4 Inference

In this section, we define the population mean as $\mathbb{E}[\cdot]$ and the sample mean as $\mathbb{E}_n[\cdot]$. We first focus on deriving the asymptotic distribution of $A\hat{T}E_{Z \rightarrow D}$, $I\hat{T}T_{Z \rightarrow Y}$, and $L\hat{A}TE_{D \rightarrow Y}$. A similar derivation holds for the average treatment effects of the treated and untreated. The proof is inspired by Chernozhukov et al. (2024). For inference purposes, we assume that the probability of a focal target coupon being replaced by a random coupon, $p \equiv Pr\{H(0) = 1|T(1) = 1\}$, remains constant throughout our sample. This assumption holds because the number of promotional coupons was generally fixed at seven during the sample period, although it briefly increased to nine in the later period. Our main results remain consistent even if we exclude this later period.

LATE

We define a random variable G as follows:

$$\{G = 1\} = \{R = 1\}, \{G = 2\} = \{N = 1\}, \{G = 3\} = \{R = 0, N = 0\}.$$

For simplicity, let $w_1 \equiv 1$, $w_2 \equiv (1 - p)/p$, and $w_3 \equiv (-1/p)$. Also, consider the simplified identification of $ATE_{Z \rightarrow D}$ from the previous subsection and rewrite it in terms of the G 's. Then we have,

$$\begin{aligned} ATE_{Z \rightarrow D} &= \mathbb{E}[D|G = 1] + \frac{1-p}{p}\mathbb{E}[D|G = 2] + \frac{-1}{p}\mathbb{E}[D|G = 3] \\ &= \frac{\mathbb{E}[D\mathbf{1}\{G = 1\}]}{\mathbb{E}[\mathbf{1}\{G = 1\}]} + \frac{1-p}{p} \frac{\mathbb{E}[D\mathbf{1}\{G = 2\}]}{\mathbb{E}[\mathbf{1}\{G = 2\}]} + \frac{-1}{p} \frac{\mathbb{E}[D\mathbf{1}\{G = 3\}]}{\mathbb{E}[\mathbf{1}\{G = 3\}]} \\ &= \sum_{g=1}^3 w_g \frac{\mathbb{E}[D\mathbf{1}\{G = g\}]}{\mathbb{E}[\mathbf{1}\{G = g\}]} \end{aligned}$$

Let,

$$\theta_g \equiv w_g \mathbb{E}[D(g)] = w_g \frac{\mathbb{E}[D\mathbf{1}\{G = g\}]}{\mathbb{E}[\mathbf{1}\{G = g\}]}$$

$$\begin{aligned} \implies \sqrt{n}(\hat{\theta}_g - \theta_g) &= \sqrt{n}\hat{w}_g \frac{\mathbb{E}_n[(D_i(g) - \mathbb{E}[D(g)])\mathbf{1}\{G_i = g\}]}{E_n[\mathbf{1}\{G_i = g\}]} \\ &= \sqrt{n}w_g \frac{\mathbb{E}_n[(D_i(g) - \mathbb{E}[D(g)])\mathbf{1}\{G_i = g\}]}{Pr\{G = g\}} + o_p(1) \end{aligned}$$

Note that the last line holds because of the Weak Law of Large Numbers and Slutsky's theorem. The term

$$\frac{(D_i(g) - \mathbb{E}[D(g)])\mathbf{1}\{G_i = g\}}{Pr\{G = g\}}$$

has zero mean and variance

$$\frac{\mathbb{E}\left[\left(D_i(g) - \mathbb{E}[D(g)]\right)^2 \mathbf{1}\{G_i = g\}^2\right]}{Pr\{G = g\}^2} = \frac{Var(D|G = g)}{Pr\{G = g\}}$$

Moreover, note that for $G = g_j \neq g_k$, the covariance term,

$$\mathbb{E}\left[\frac{(D_i(g_j) - \mathbb{E}[D(g_j)])\mathbf{1}\{G_i = g_j\} \times (D_i(g_k) - \mathbb{E}[D(g_k)])\mathbf{1}\{G_i = g_k\}}{Pr\{G = g_j\} \times Pr\{G = g_k\}}\right] = 0$$

Therefore, by the CLT and the delta method,

$$\sqrt{n}(\hat{ATE} - ATE) \xrightarrow{d} N\left(0, \frac{Var(D|R=1)}{Pr\{R=1\}} + \left(\frac{1-p}{p}\right)^2 \frac{Var(D|N=1)}{Pr\{N=1\}} + \left(\frac{1}{p}\right)^2 \frac{Var(D|R=0, N=0)}{Pr\{R=0, N=0\}}\right)$$

By a similar argument, we have

$$\sqrt{n}(\hat{ITT} - ITT) \xrightarrow{d} N\left(0, \frac{Var(Y|R=1)}{Pr\{R=1\}} + \left(\frac{1-p}{p}\right)^2 \frac{Var(Y|N=1)}{Pr\{N=1\}} + \left(\frac{1}{p}\right)^2 \frac{Var(Y|R=0, N=0)}{Pr\{R=0, N=0\}}\right)$$

Moreover, by the CLT and the delta method, we can show that (a similar derivation holds for $LATE_{\hat{T}=1, D \rightarrow Y}$ and $LATE_{\hat{T}=0, D \rightarrow Y}$)

$$\sqrt{n}(\hat{LATE} - LATE) \xrightarrow{d} N(0, \tilde{\Sigma})$$

where,

$$\tilde{\Sigma} = \frac{1}{ATE^2} \Sigma_{11} - 2 \frac{ITT}{ATE^3} \Sigma_{12} + \frac{ITT^2}{ATE^4} \Sigma_{22}$$

$$\Sigma_{11} = \sum_{g=1}^3 w_g^2 \frac{Var(Y|G=g)}{Pr\{G=g\}}, \quad \Sigma_{22} = \sum_{g=1}^3 w_g^2 \frac{Var(D|G=g)}{Pr\{G=g\}}$$

$$\begin{aligned} \Sigma_{12} = \Sigma_{21} &= E\left[\left(\sum_{g=1}^3 w_g \frac{(Y_i(g) - \mathbb{E}[Y(g)])\mathbf{1}\{G_i = g\}}{Pr\{G = g\}}\right)\left(\sum_{g=1}^3 w_g \frac{(D_i(g) - \mathbb{E}[D(g)])\mathbf{1}\{G_i = g\}}{Pr\{G = g\}}\right)\right] \\ &= \sum_{g=1}^3 w_g^2 \frac{Cov(Y, D|G=g)}{Pr\{G=g\}} \end{aligned}$$

Web Appendix D Simulation Study

To demonstrate that our estimator is consistent and identifies the true LATE, we conduct a Monte Carlo study with synthetic data.

We simulate data for $I = 1,000,000$ customers. Mirroring the implementation at the retailer that provided the data for our study, we implement a couponing system that promotes $J = 50$ brands from different categories. Every time customers interact with the couponing system, they receive three coupons. The simulation is calibrated such that the characteristics of the output data are similar to those of the empirical data. To assess the uncertainty in the results, we use $B = 1,000$ replications.

The data-generating process is based on a probit choice model. We assume that customer i chooses brand j if

$$u_{ijt} = \alpha_{ij} + \beta_i c_{ijt} + \gamma_i \cdot \mathbb{I}\{s_{it} = j\} + \epsilon_{ijt} > 0,$$

where c_{ijt} is a coupon indicator and s_{it} is the customer’s loyalty state. The error term ϵ_{ijt} is an i.i.d. draw from a normal distribution. Heterogeneity of the parameters α_{ij} , β_i , and γ_i is continuous (normal distribution). For the intercepts α_{ij} , the average values $\bar{\alpha}_j$ range from -5.5 to -2.5 , and the focal brand in the simulation has an average intercept of -3.25 , indicating a medium-high preference. All variances are set to 1, and we allow for small-to-medium correlations between brands using an $LKJ(10)$ prior. This setup, combined with individual-level heterogeneity, yields reasonable variation in brand relevance across customers.

The coupon sensitivity parameters β_i and the state dependence parameters γ_i are independently drawn from $N(1, 0.1)$. For *target coupons*, we construct a noisy score by adding standard normal noise to α_{ij} , then apply a threshold rule to assign an intended target coupon ($\hat{T} = 1$). The probability of receiving a *random coupon* is 0.15, and conditional on getting a random coupon, the corresponding brand is drawn from all $J = 50$ brands with equal probability.

Based on this setup, we created five scenarios to study the performance of our proposed approach:

Scenario 1: No target coupons.

Scenario 2: Use the setup and parameter values described above: $\beta_i, \gamma_i \sim N(1, 0.1)$.

Scenario 3: Same as Scenario 2, but higher mean coupon sensitivity: $\beta_i \sim N(1.5, 0.1)$.

Scenario 4: Same as Scenario 2, but higher mean state dependence: $\gamma_i \sim N(1.5, 0.1)$.

Scenario 5: Same as Scenario 2, but no state dependence for all customers: $\gamma_i = 0$.

This simulation setup enables us to compute the true values of the treatment effect, including the true LATE of state dependence. For each replication, we estimate $\widehat{\text{LATE}}$, $\widehat{\text{ITT}}$, and $\widehat{\text{ATE}}$ with and without adjustment for the presence of target coupons (see Section 4 and Web

Appendix C) in the control group and calculate the bias with respect to the true treatment effects (e.g., $\widehat{\text{LATE}} - \text{true LATE}$). Web Appendix M explains in detail how to derive the treatment effects from the choice model.

Table A2 summarizes the mean value of each true treatment effect across the $B = 1,000$ replications, demonstrating how the specific parameter values of the choice model-based DGP translate into interpretable treatment effect values. The simulation design covers both a null-effect setting and economically meaningful state-dependence effects of magnitudes comparable to those in our empirical application; the true LATE ranges from no state-dependence effect in Scenario 5 to medium-sized effects of 0.164 in Scenarios 1–2 and 0.138 and in Scenario 3, and a larger effect of 0.294 in Scenario 4. The true ITT values are smaller, reflecting that random coupon assignment affects only a relatively small share of relevant purchase occasions ($\leq 10\%$). The ITT depends on both the coupon sensitivity parameter β_i and the state-dependence parameter γ_i , whereas the ATE is only affected by the coupon sensitivity β_i . Since the LATE is identified as a Wald ratio of the ITT and the (first-stage) ATE, it is a nonlinear function of both choice-model parameters. When $\gamma_i = 0$, however, state dependence does not enter utility; the corresponding ITT is therefore zero, and so is the LATE.

Table A2: True Treatment Effects by Scenario

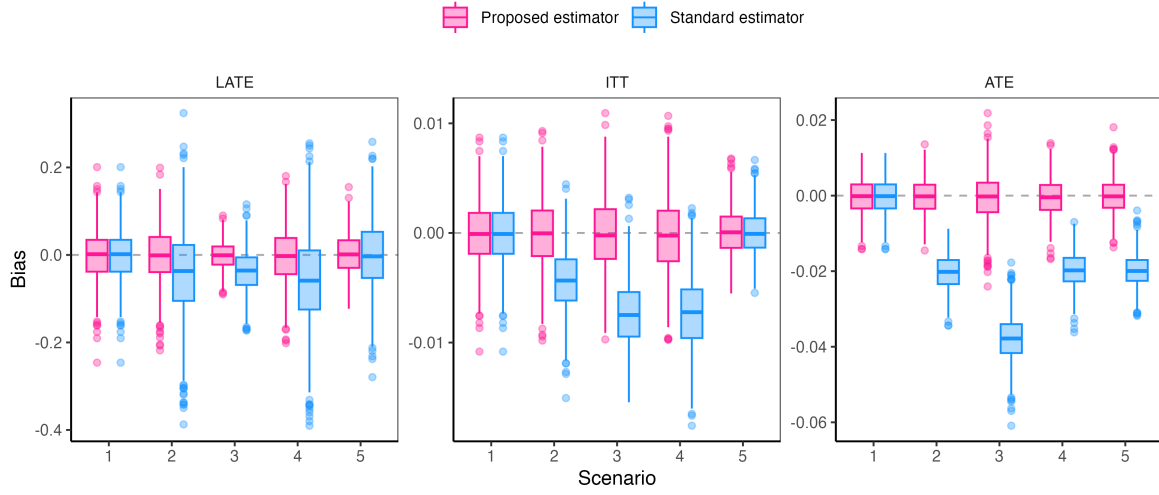
Scenario	LATE	ITT	ATE
1	0.164	0.008	0.047
2	0.164	0.008	0.047
3	0.138	0.014	0.100
4	0.294	0.014	0.046
5	0.000	0.000	0.045

Figure A2 shows the bias across the $B = 1,000$ replications for each approach and treatment effect (i.e., the difference between estimated and true values). In Scenario 1, there are no target coupons. In the absence of a target coupon, the adjusted Wald estimator reduces to the standard Wald estimator; the two results are identical with very low bias (median < 0.002). The LATE only uses the variation in random coupons for the focal brands to identify the LATE (as well as ITT and ATE).

In Scenario 2, the same holds for our proposed estimator in the presence of target coupons (absolute median bias < 0.002). Interestingly, the variation in the results is only slightly higher than in the case without target coupons (+15%). If target coupons are part of the data-generating process but ignored during estimation, the LATE is biased (−26.9%). This bias results from substantial biases in the ITT (−56.8%) and ATE (−43.2%). Not accounting for target coupons also increases uncertainty in the results, with variance increasing by 341%.

The higher coupon sensitivity in Scenario 3 increases the biases for ITT and ATE (in absolute terms), but produces a similar bias for the LATE compared to Scenario 2 (in relative terms: −26.6%). The increase in complier rates results in tighter LATE bias distributions. With adjustment, the bias is very small (abs. median bias < 0.002). Also here, the increase in

Figure A2: Simulation Results



complier rates results in tighter LATE bias distributions.

The larger state dependence parameter in Scenario 4 does not affect the ATE bias (compared to Scenario 2) but increases the ITT bias (similar to Scenario 3), and the LATE bias is also similar to Scenario 2. Again, the adjusted estimator has a very small bias.

The results for Scenario 5 show that when there is no utility from state dependence, the biases for ITT and LATE are close to zero, even though the ATE remains biased (as in Scenarios 2 and 4). This leads to a wider LATE bias distribution without adjustment than with the adjusted estimator. Hence, the adjustment is useful for getting a “tight” null effect.

We conclude that our proposed estimator identifies the true treatment effects and performs well across a variety of reasonable scenarios.

Web Appendix E Complier Profiling

This appendix describes how we profile compliers in our setting. We first derive the average brand purchase rate in the absence of treatment for compliers, $Y(0)$. We then show how the same logic extends to the characterization of compliers using pre-encouragement variables X .

E.1 Average Brand Purchase Rate for Compliers

To estimate the brand purchase rate conditional on category purchase in the absence of treatment for compliers, $Y(0)$, we use the identification argument for $Y(d)$ in the standard encouragement design.

Based on Lemma 2.1 from [Abadie \(2002\)](#), we can derive these quantities by modifying the ITT expressions in equation (5) in the main paper.²⁸

For the expected value of $Y(0)$ among compliers, we first consider $ITT_{Z \rightarrow (1-D)Y}$:

$$\begin{aligned} ITT_{Z \rightarrow (1-D)Y} &\equiv \mathbb{E}[(1 - D(1))Y(D(1)) - (1 - D(0))Y(D(0))] \\ &= \mathbb{E}[(1 - D(1))Y(D(1)) - (1 - D(0))Y(D(0)) \mid D(1) > D(0)] \cdot \Pr(D(1) > D(0)) \\ &\quad + \mathbb{E}[(1 - D(1))Y(D(1)) - (1 - D(0))Y(D(0)) \mid D(1) = D(0)] \cdot \Pr(D(1) = D(0)) \\ &= -\mathbb{E}[Y(0) \mid D(1) > D(0)] \cdot \Pr(D(1) > D(0)). \end{aligned}$$

It follows that

$$\mathbb{E}[Y(0) \mid D(1) > D(0)] = -\frac{ITT_{Z \rightarrow (1-D)Y}}{ATE_{Z \rightarrow D}}.$$

For estimation, we replace expectations with sample analogues:

$$\mathbb{E}[Y(0) \mid D(1) > D(0)] = -\frac{\mathbb{E}[(1 - D)Y \mid Z = 1] - \mathbb{E}[(1 - D)Y \mid Z = 0]}{\mathbb{E}[D \mid Z = 1] - \mathbb{E}[D \mid Z = 0]}.$$

Note that this is again a Wald-style estimator. Relative to the LATE formula, we only replace the outcome Y by $(1 - D)Y$ and flip the sign of the result.

As before, target coupons complicate the analysis in our setting. However, we can use exactly the same identification arguments as for the LATE in the presence of target coupons. Hence, we can use our proposed estimator based on random variation in non-focal brands' coupons. Consequently, we only need to change the numerator in eq. (4). We obtain the ITT of Z on $(1 - D)Y$:

$$\begin{aligned} ITT_{Z \rightarrow (1-D)Y} &= \mathbb{E}[(1 - D(1))Y(D(1)) - (1 - D(0))Y(D(0))] \\ &= \mathbb{E}[(1 - D)Y \mid R = 1] - \mathbb{E}[(1 - D)Y \mid R = 0] \\ &\quad + \Pr(T = 1 \mid R = 0) \cdot ITT_{T=1, Z \rightarrow (1-D)Y}, \end{aligned}$$

²⁸Lemma 2.1 in [Abadie \(2002\)](#) is more general, as it allows the potential outcome $Y(d)$ to be transformed by a measurable function on the real line $h(\cdot)$ such that $\mathbb{E}[h(Y_i)] < \infty$.

with

$$\text{ITT}_{T=1, Z \rightarrow (1-D)Y} = \frac{\mathbb{E}[(1-D)Y \mid N=1, R=0] - \mathbb{E}[(1-D)Y \mid N=0, R=0]}{\mathbb{E}[T \mid N=1, R=0] - \mathbb{E}[T \mid N=0, R=0]}.$$

Applying the adjusted estimator for $Y(0)$ to our data with 223 brands and using the same meta-analytic evidence aggregation with the Bayesian hierarchical model yields the results reported in Table A3.

Table A3: Average Brand Purchase Rate for Compliers

	Posterior mean	95% CI
Mean (τ)	0.135	[0.080, 0.186]
Standard Deviation (σ)	0.234	[0.176, 0.297]

Now, τ represents the average value of $Y(0)$ in the population of compliers, and σ measures the corresponding heterogeneity across all brands. The posterior mean of τ is 0.135, i.e., across brands we find an average brand purchase rate for compliers of about 13.5%. This is considerably higher than the descriptive value of 7.9% in the loyalty card data for all customers (i.e., not just compliers).

We compute the state dependence lift as $\frac{LATE}{Y(0)} = 74.1\%$. Instead of computing the ratio of average values, we could also first compute brand-specific lifts using the shrunk estimates for brand-specific LATEs and $Y(0)$ values and then aggregate the results. Because the shrunk estimates for $Y(0)$ are still noisy, some are less than or equal to zero, and the corresponding lifts are therefore not meaningful. However, if we restrict the analysis to brands with $Y(0)$ greater than zero (i.e., 196 out of 223 brands, 88%), the median value across the brand-specific lifts is 68.9%. Both approaches lead to similar results, especially once we account for the corresponding uncertainty. The posterior distribution of $\frac{LATE}{Y(0)}$ has a 95% credible interval of [31.0%, 146.0%].

E.2 Complier Characterization with Pre-Encouragement Variables

The same argument used above for $Y(0)$ also yields complier moments for pre-encouragement variables. Let X denote any variable measured before coupon exposure, such as a lagged purchase indicator, prior purchase frequency, or another baseline customer characteristic. Because X is measured before encouragement, it is unaffected by coupon exposure.

In the standard encouragement design, replacing the transformed outcome $(1-D)Y$ by DX implies

$$\begin{aligned} \text{ITT}_{Z \rightarrow DX} &\equiv \mathbb{E}[D(1)X - D(0)X] \\ &= \mathbb{E}[X \mid D(1) > D(0)] \cdot \Pr(D(1) > D(0)). \end{aligned}$$

Hence,

$$\mathbb{E}[X \mid D(1) > D(0)] = \frac{\text{ITT}_{Z \rightarrow DX}}{\text{ATE}_{Z \rightarrow D}} = \frac{\mathbb{E}[DX \mid Z = 1] - \mathbb{E}[DX \mid Z = 0]}{\mathbb{E}[D \mid Z = 1] - \mathbb{E}[D \mid Z = 0]}.$$

Thus, relative to the Wald-style estimator for the LATE, we only replace the numerator outcome Y by DX , while the denominator remains the first stage $\text{ATE}_{Z \rightarrow D}$.

As before, target coupons complicate the analysis in our setting. However, the same identification argument as in the previous subsection applies. Specifically, we replace the numerator in eq. (4) by the corresponding adjusted ITT for DX :

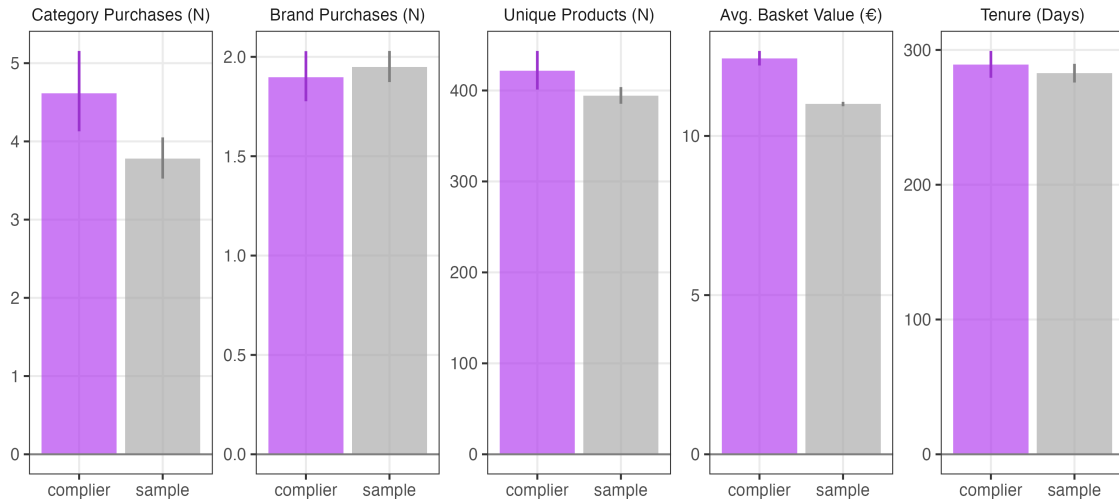
$$\begin{aligned} \text{ITT}_{Z \rightarrow DX} &= \mathbb{E}[DX \mid R = 1] - \mathbb{E}[DX \mid R = 0] \\ &\quad + \Pr(T = 1 \mid R = 0) \cdot \text{ITT}_{T=1, Z \rightarrow DX}, \end{aligned}$$

with

$$\text{ITT}_{T=1, Z \rightarrow DX} = \frac{\mathbb{E}[DX \mid N = 1, R = 0] - \mathbb{E}[DX \mid N = 0, R = 0]}{\mathbb{E}[T \mid N = 1, R = 0] - \mathbb{E}[T \mid N = 0, R = 0]}.$$

We apply the adjusted estimator for $\mathbb{E}[X \mid D(1) > D(0)]$ to five pre-encouragement variables in our data: the *number of prior category purchases*, the *number of prior brand purchases*, the *number of unique products purchased*, the *average basket value* (in Euro), and the loyalty card membership *tenure* (in days). As before, we summarize the results across the 223 brands using the Bayesian meta-analytic evidence aggregation approach. To deal with right-skewed data distributions, we estimate the model in log-space for all variables except *tenure*. Figure A3 reports the posterior means and corresponding 95% credible intervals for the average values of each variable and group (i.e., compliers versus the full sample). For the variables estimated in log-space, we report back-transformed median values.

Figure A3: Complier Characterization for Pre-Encouragement Variables



The characterization results indicate that compliers are a small group, with a median complier

share across brands of about 2% (max 8.9%), but otherwise fairly similar to the overall sample. This suggests that the customers whose purchases are affected by the coupon of the focal brand are not an unusual subgroup. Relative to the sample, compliers have a modestly higher average basket value (€12.44 vs. €11.00), somewhat more prior category purchases (4.62 vs. 3.78), and a slightly larger number of unique products overall purchased (422 vs. 394 unique products). At the same time, prior brand purchases are very similar across the two groups and statistically indistinguishable and, if anything, slightly lower for compliers (1.90 vs. 1.95), while tenure is only marginally higher (289 vs. 283 days).

Taken together, these results suggest that compliers are broadly similar to the overall customer base, with modest evidence that they are somewhat more active shoppers at the category and store level. This pattern is intuitive: compliers are, by definition, customers whose brand choice can be shifted by a coupon, which requires meaningful engagement with the category but no strong pre-existing commitment to the focal brand. This makes them a particularly relevant subgroup from a managerial perspective, since they are precisely the customers whose brand choice a coupon can plausibly influence.

Web Appendix F Balance Checks

Our encouragement design approach for measuring the local average treatment effect of state dependence relies on conditional independence conditions implied by the experimentation layer (see Section 4). To assess the validity of this assumption, we report summary statistics for six (customer-level) covariates by encouragement status in this appendix.

The six covariates are basket size, brand purchase amount per basket (in Euros), brand purchase frequency, category purchase frequency, kiosk usage frequency, and shopping trip frequency. We chose the variables to describe different aspects of customers’ shopping behavior and interactions with the loyalty program and retailer. We only use loyalty card data *before* the experiment (i.e., pre-encouragement data) to calculate the variables.

We follow [Imbens and Wooldridge \(2009\)](#) and calculate the normalized difference in averages by encouragement status for the six covariates. This statistic is a scale-free measure of the difference in distributions. We report the percentages of brands exceeding an absolute test statistic of 0.25 in Table A4 for the two instruments used in the main text: coupon exposure (see Section 6) and discount depth (see Section 7). In addition, we also calculate the percentages of brands for which the test statistic exceeds 0.10. For both thresholds, the test statistics exceed the threshold for less than 5% of the brands.

Table A4: Balance Check Results

Variable	Z = Coupon exposure		Z = Discount depth	
	% > 0.25	% > 0.1	% > 0.25	% > 0.1
Basket Size	0.45	4.04	0.00	2.24
Basket Purchase Amount	0.45	2.24	0.00	1.35
Brand Purchase Frequency	0.45	1.79	0.00	1.79
Category Purchase Frequency	2.24	4.48	0.00	1.35
Kiosk Usage Frequency	0.00	0.90	0.00	2.69
Shopping Trip Frequency	0.00	3.59	0.00	4.04

Note: The table reports the percentages of brands with a scaled difference in pre-encouragement variables between the encouragement status levels above a threshold of 0.25 and 0.10.

Similarly, we also checked whether random coupons affect the probability that customers are targeted with coupons for the same brand (e.g., a purchase caused by a coupon can influence the targeting algorithm and future target coupon assignment). To evaluate this concern, we compare *target* coupon rates at the *next* kiosk visit for customers with different encouragement statuses.

As before, we calculate the scaled mean difference separately for each brand and calculate the percentages of brands with a mean scaled difference larger than 0.25. We find values exceeding the threshold for only 1.86% of all brands. We replicate the same analysis for high vs. low discounts; here, all scaled mean differences are below the threshold of 0.25. In our data, only

around 1% of customers and 10% of compliers²⁹ receive a focal-brand coupon on their next category trip, bounding any contamination at a small fraction of the estimated LATE. We acknowledge that no observational test can fully rule out complier-specific violations of the exclusion restriction.

We conclude that the covariate distributions are sufficiently similar across the two encouragement status values. Neither instrument (random coupons vs. no coupons, high vs. low discounts) affects future (target) coupon assignments.

²⁹Estimated using the complier characterization approach presented in Web Appendix [E.2](#).

Web Appendix G LATE Results Based on a Time Window

In the main analysis (Section 6), we define Y relative to the next category-purchase occasion. If such an occasion is observed, $Y = 1$ when the focal brand is purchased and $Y = 0$ otherwise. Customers without a subsequent focal-category purchase are retained with $Y = 0$, so the estimation sample is not restricted based on a post-encouragement category purchase. This definition focuses on brand choice at the next category-purchase occasion and thereby reduces mechanical concerns about purchase acceleration or stockpiling. As a robustness check, we also use fixed-window outcomes that do not refer to future category-purchase occasions.

To assess the robustness of our findings, we construct an alternative set of outcome variables Y based on fixed post-encouragement time windows. Specifically, we define Y as a binary indicator that equals 1 if the customer purchases the focal brand at least once within a time window following the encouragement (7, 14, 28, or 84 days), and 0 otherwise. As product categories differ in inter-purchase times (IPTs), we additionally estimate a specification that uses the median IPT as the time-window length. This window-based specification does not condition on observed category purchases; instead, it captures the treatment effect on the brand purchase incidence within economically relevant horizons.

We emphasize that the LATE estimates across outcome specifications are not directly comparable in magnitude, because each definition of Y measures a different object. The main specification measures repurchase using the next-category-purchase outcome described above. The fixed-window specifications instead measure whether any focal-brand purchase occurs within a given time window, irrespective of whether and when the next category-purchase occasion occurs. Shorter windows mechanically provide fewer repurchase opportunities, especially in slow-moving categories, whereas longer windows provide more such opportunities. Our goal is therefore not to benchmark magnitudes against the main results, but to assess whether the qualitative conclusion regarding positive structural state dependence is robust to outcome definitions that do not refer to subsequent category-purchase occasions.

Table A5 shows the results of the LATE estimates from the hierarchical Bayesian meta-analysis for all brands and the strong IV brands. The first part of the table reports results for brand-specific time windows set to the median IPT within each category. The estimated mean LATEs have a similar magnitude, and there is considerable heterogeneity in LATEs across brands. Most importantly, the average LATE values for both brand sets are positive and significantly different from zero (all brands: 0.043; strong IV brands 0.046).

Hence, we find clear empirical evidence for state dependence under an outcome definition that does not refer to the next category-purchase occasion. The magnitude of the effect appears to be lower than the result reported in Section 6, which is expected because the window-based outcomes measure brand purchase incidence over finite calendar windows rather than brand choice at the next category-purchase occasion.

Table A5 also shows the results for the different time windows (7, 14, 28, and 84 days after encouragement), and they mirror the previous results. We find clear empirical evidence for state dependence (even for the 7-day window, $Pr\{\hat{\tau} > 0\} = 89\%$). As expected, longer time

Table A5: Bayesian Meta-Analysis Results for Different Outcome Variables

	All brands ($N = 223$)		Strong IV brands ($N = 126$)	
	Posterior mean	95% CI	Posterior mean	95% CI
Category specific IPT				
Mean (τ)	0.043	[0.013, 0.073]	0.046	[0.013, 0.080]
Standard Deviation (σ)	0.121	[0.088, 0.157]	0.126	[0.091, 0.164]
7 days after encouragement				
Mean (τ)	0.013	[-0.007, 0.034]	0.020	[-0.001, 0.042]
Standard Deviation (σ)	0.074	[0.048, 0.101]	0.072	[0.045, 0.103]
14 days after encouragement				
Mean (τ)	0.047	[0.016, 0.079]	0.057	[0.024, 0.092]
Standard Deviation (σ)	0.132	[0.097, 0.171]	0.131	[0.095, 0.171]
28 days after encouragement				
Mean (τ)	0.071	[0.029, 0.115]	0.073	[0.026, 0.118]
Standard Deviation (σ)	0.191	[0.144, 0.244]	0.194	[0.143, 0.251]
84 days after encouragement				
Mean (τ)	0.110	[0.039, 0.184]	0.097	[0.020, 0.175]
Standard Deviation (σ)	0.340	[0.266, 0.424]	0.355	[0.273, 0.446]

Notes: A brand is classified as strong IV if $ATE_{Z \rightarrow D}$ is statistically different from zero with an F -statistic greater than 10 (Stock and Yogo 2005).

windows lead to larger (mean) LATE values (τ).

In summary, we conclude that the additional results are consistent with our main analysis (Section 6). The average effect remains positive across short and long time horizons, which is not consistent with short-run stockpiling or purchase acceleration being the primary explanation for the main results. Overall, the window-based results reinforce the main conclusion that the coupon induces genuine state dependence rather than primarily intertemporal substitution.

Web Appendix H Prior Specification in Bayesian Meta-Analysis

For the evidence aggregation approach used in the main text (see Section 6), we follow current literature on statistical aggregation and rely on a Hierarchical Bayes approach that uses weakly informative priors. This appendix evaluates the robustness of our findings to different prior choices.

The model provides a principled approach to achieve shrinkage of each individual $\widehat{\text{LATE}}_j$ estimate towards τ based on its statistical uncertainty, $\hat{s}_j^2$, and *partially pools* the results across brands. Specifically, the distribution of τ_j has a mean of $\lambda_j \cdot \widehat{\text{LATE}}_j + (1 - \lambda_j) \cdot \tau$ and a variance of $\lambda_j \cdot \hat{s}_j^2$, where $1 - \lambda_j = \hat{s}_j^2 / (\hat{s}_j^2 + \sigma^2)$ is the pooling factor that determines how much the posterior mean for brand j is “shrunk” toward the overall mean. Because λ_j approaches 1 as the brand-specific sampling variance $\hat{s}_j^2$ decreases, brands that are estimated precisely (small $\hat{s}_j^2$) retain most of their own results, while imprecisely estimated brands (large $\hat{s}_j^2$) borrow more strength from the other brands.

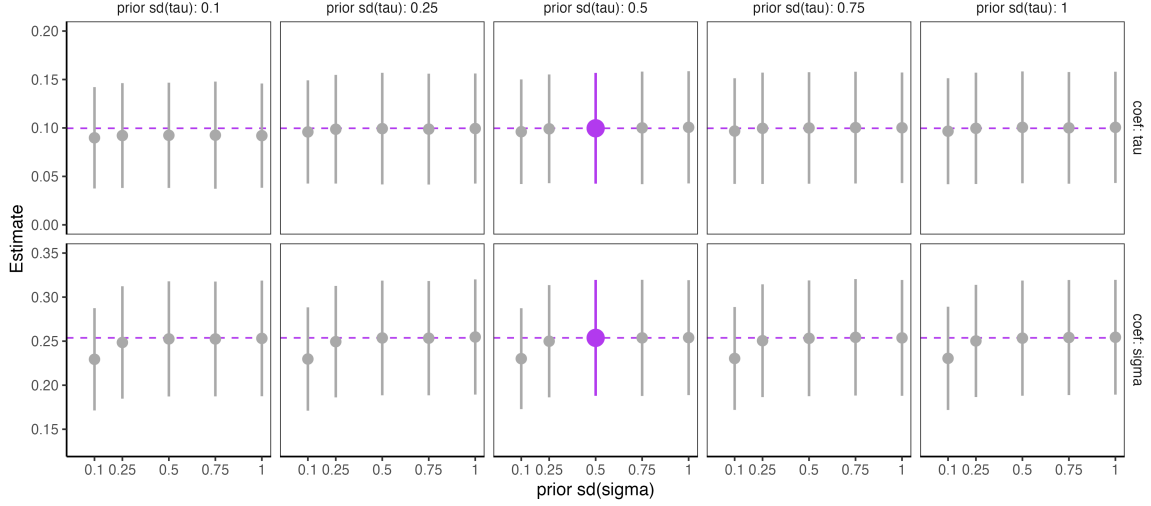
The weight on the overall mean exceeds 0.5 only when the within-brand noise dominates the between-brand heterogeneity ($\hat{s}_j^2 > \sigma^2$); otherwise, the brand’s own estimate gets more than half of the weight. Except for the edge case $\hat{s}_j^2 = 0$, every shrunk estimate τ_j therefore blends information from all brands, resulting in an empirically optimal trade-off between overfitting the idiosyncrasies of a single experiment and ignoring brand-level evidence altogether. $\lambda_j < 1$ implies that the shrunk LATE estimate of a brand is more precise than $\widehat{\text{LATE}}_j$. Hence, given the results across all brands, the shrunk LATE estimates τ_j are the best we can learn at the brand level.

For the prior specification used in the main paper (all brands), we obtain an average value across brands for the pooling factor $(1 - \lambda_j)$ of 0.653 ($SD = 0.291$) and corresponding percentiles of $q_{10} = 0.180$, $q_{25} = 0.451$, $q_{50} = 0.698$, $q_{75} = 0.921$, and $q_{90} = 0.985$. Thus, on average, the posterior means for brands are shrunk about 65% towards $\tau = 0.100$ and the corresponding sampling variances are reduced to about 35% of their original values. However, the percentiles highlight that there is a considerable amount of heterogeneity in the brand-specific shrinkage (because of differences in the brand-specific sampling variances). 10% of the brands with the most precise LATE estimates are shrunk by only 18% (or less); 25% of the brands with the largest sampling variances are shrunk by at least 95%.

In the main paper, we use prior variances of 0.5^2 for τ and σ (i.e., mean and standard deviation of the normal distribution of the heterogeneity model). For τ , this implies a 95% prior interval of $[-0.98, 0.98]$. For σ , a half-normal prior with the same variance has expectation $0.5\sqrt{2/\pi} \approx 0.40$, which corresponds to a 95% central interval for the implied heterogeneity distribution of $[-0.78, 0.78]$. The chosen prior is intentionally non-restrictive: it covers almost the entire range of plausible LATE values in $[-1, 1]$, while allowing substantial between-brand variation even when the mean effect is near zero. To assess robustness, we re-estimate the hierarchical model for all 25 combinations of prior variances 0.1^2 , 0.25^2 , 0.5^2 , 0.75^2 , and 1 applied independently to τ and σ .

Figure A4 shows the estimates for all meta analyses (i.e., posterior means for τ and σ to-

Figure A4: Robustness of $\hat{\tau}$ and $\hat{\sigma}$ for Different Prior Values



Note: Purple color indicates the prior choice in the main paper.

gether with 95% credible intervals). The reported values in Table 2 of the main paper are highlighted in purple color, with specific values of $\hat{\tau} = 0.100$ and $\hat{\sigma} = 0.252$. The results for the alternative prior specifications are very stable across prior specifications. Across all specifications we obtain $\hat{\tau} \in [0.090, 0.101]$ and $\hat{\sigma} \in [0.229, 0.255]$ (i.e., change $\leq 10\%$). Only the most informative prior (variance 0.1^2), which restricts τ to $[-0.196, 0.196]$ with 95% prior probability, induces a minimal visible change of the mean. Furthermore, a tight prior on τ (σ) affects primarily the posterior mean of the corresponding parameter. The very similar results across prior specifications are also reflected in the almost identical amounts of shrinkage. The average values for the pooling factor $(1 - \bar{\lambda}_j)$ do not vary for different priors (range $[0.651, 0.681]$), and the overall median of 0.653 aligns exactly with the average value of our preferred prior specification.

We conclude that our baseline priors reflect a sensible balance between informativeness and flexibility, and that the substantive results are very robust to a wide range of alternative prior choices.

Web Appendix I Heterogeneity Distribution in Bayesian Meta-Analysis

The evidence aggregation used in Section 6 follows current literature on statistical aggregation by using a normal distribution to model the first stage (i.e., the heterogeneity distribution). The empirical distributions of the LATE estimates in Figure 5b show no signs of boundedness, skewness, or multimodality, so this choice seems appropriate. Nonetheless, the normal distribution has thin tails, which could lead to *overshrinkage* (Willwerscheid et al. 2021). In this web appendix, we evaluate the robustness of our findings to the choice of the heterogeneity distribution. We compare the main model specification to an alternative Hierarchical Bayes (HB) and four Empirical Bayes (EB) meta-analysis approaches:

Normal HB: Base model with a normal distribution for first stages used in Section 6.

Student-*t* HB: A specification based on a Student-*t* distribution to address the concerns about overshrinkage. We implement the normal-gamma-mixture representation in Stan. Specifically, we use $\tau_j \sim N(\tau, \sigma^2/\kappa_j)$, with $\kappa_j \sim \Gamma(\nu/2, \nu/2)$ and $\nu \sim \Gamma(2, 0.2)$, where ν is the degrees of freedom parameter of the Student-*t* distribution.

Normal EB: This specification uses a normal distribution for the first stage to isolate the effect of changing the estimation methods (EB vs. HB) for models with the same distributional assumptions.

Normal-Scale-Mixture EB: A mixture-of-normals approach with the constraint that only the scales can vary across components. Willwerscheid et al. (2021) consider this a constrained nonparametric approach.

Nonparametric EB: A finite mixture specification as a (unconstrained) nonparametric approach that only uses point estimates and weights.

Deconvolver EB: The smoothed nonparametric spline approach of Efron (2016). Deconvolver EB uses *z*-scores as input (i.e., estimates divided by their standard errors). To achieve comparability with the other results, we rescale the results for the shrunk estimates such that the overall mean of Deconvolver EB and Normal EB align.

We use the R package `ebnm` of Willwerscheid et al. (2021) to implement the EB specifications. We refer to Efron (2016) and Willwerscheid et al. (2021) for details about these methods and to Wernerfelt et al. (2025) for an application of the Deconvolver EB approach. Table A6 summarizes the resulting shrunk LATE estimates for each approach.

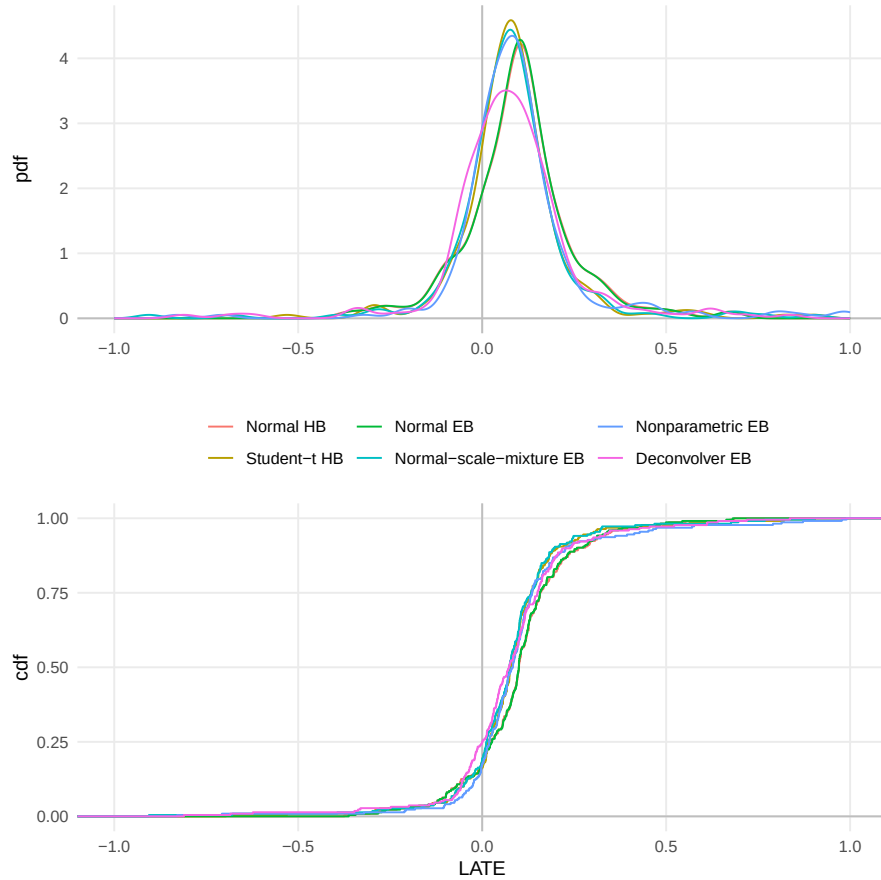
The results are very similar. Especially symmetric distributions, particularly the normal distribution, show high correlations close to 1, making hierarchical priors less critical (see also Figure A5). For models with a normal distribution, HB and EB yield nearly identical results. This highlights that the priors in our HB approach are not very informative, so our particular choice does not drive the results. The Student-*t* distribution model leads to results similar to nonparametric models, suggesting this parametric model helps address fatter tails and leptokurtic distributions. More flexible models have similar average values

Table A6: Descriptives for Shrunk LATEs From Different Approaches

Approach		Mean	SD	Percentiles							Frac > 0
				5%	10%	25%	50%	75%	90%	95%	
HB	Normal	0.101	0.148	-0.121	-0.070	0.024	0.100	0.159	0.261	0.342	0.812
	Student- <i>t</i>	0.092	0.191	-0.094	-0.050	0.018	0.078	0.135	0.218	0.309	0.839
EB	Normal	0.100	0.146	-0.121	-0.069	0.027	0.100	0.157	0.256	0.335	0.825
	Normal-scale-mixt.	0.092	0.215	-0.098	-0.050	0.013	0.074	0.136	0.211	0.322	0.812
	Nonparametric	0.112	0.236	-0.073	-0.026	0.020	0.084	0.137	0.235	0.449	0.830
	Deconvolver	0.103	0.241	-0.103	-0.057	0.004	0.076	0.154	0.257	0.452	0.758

(and medians) compared to the normal distribution. This also holds for the fraction of brands with a shrunk LATE value > 0.

Figure A5: Densities of Shrunk LATEs From Different Approaches



In conclusion, more flexible models do not provide substantially different insights than the normal distribution model. Robust positive average structural state dependence effects are observed across brands, indicating considerable heterogeneity and significant measurement uncertainty at the individual brand level without collapsing the aggregated results to a single point estimate.

Web Appendix J Treatment Effect Heterogeneity Across Brands

In Section 6, we documented substantial heterogeneity in structural state dependence across 223 brands in 90 categories, similar to prior analyses in the literature (e.g., [Seetharaman et al. 1999](#)). In this web appendix, we examine whether observable factors can explain some of the heterogeneity by generalizing our Bayesian hierarchical meta-analysis framework. Specifically, we assume that the brand-level LATEs are linearly related to observed covariates, $\mathbb{E}[\tau_j | x_j] = \tau + x_j' \delta$; x_j includes (standardized) brand and category-level covariates, and δ is a parameter vector.

We include eleven covariates in x_j that are plausibly related to differences in the LATEs across brands. The five category-level features include the impulse-buying and stockpiling scores from [Narasimhan et al. \(1996\)](#), category concentration, category purchase frequency, and customer penetration; the six brand-level features include basket penetration, market share, relative price position, average weekly brand revenue, average campaign attractiveness, and campaign size. We operationalized category concentration as the Herfindahl index, average campaign attractiveness as the recommender system score, and campaign size as target coupon penetration.

We use the retailer’s sales and loyalty card data as the source of information for constructing all variables, except for the impulse buying and ability-to-stockpile scales. The data for the impulse buying and ability-to-stockpile scores were collected in July 2017 on Amazon Mechanical Turk (MTurk); 614 respondents rated eight (randomly chosen) categories each. We used a respondent’s rating only if they purchased in the category at least once in the last six months. Therefore, the number of observations differs across categories (min = 7, median = 41, max = 58). The final values for both variables are simple sum scores. We validated the MTurk results using ratings on the same scales for a subset of 20 categories generated by ten German retail experts (correlations close to 0.8, $p < 0.01$).

Table A7 presents descriptive statistics for the eleven variables. By tendency, the categories of 223 brands in our data score above average on the impulse score and below average on the stockpile score. The average purchase frequency implies an inter-purchase time of 3.9 weeks. We find reasonable variation in each variable across brands, and the correlations between variables are low to medium. Only the correlation between the basket and the coupon penetration of a brand is slightly above 0.5. To facilitate interpretation, we standardize each variable before running the Bayesian (hierarchical) meta-regression in Section J.

Table A7: Descriptive Statistics and Correlations of the Variables

Variable	Mean	SD	Correlation									
			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
<i>Category:</i>												
Impulse (1)	0.475	1.047										
Stockpile (2)	−0.132	0.790	−0.339									
Category concentration (3)	0.201	0.156	−0.131	−0.025								
Purchase frequency (4)	0.037	0.014	−0.211	0.000	−0.147							
Customer penetration (5)	0.267	0.138	0.384	−0.245	−0.261	−0.027						
<i>Brand:</i>												
Basket penetration (6)	0.005	0.006	0.032	−0.208	0.208	0.045	0.315					
Market share (7)	0.149	0.161	−0.115	−0.032	0.506	−0.125	−0.372	0.504				
Rel. price position (8)	0.996	0.340	−0.086	0.000	−0.005	0.113	0.005	−0.079	0.027			
Weekly revenue (9)	0.518	0.323	−0.188	0.109	0.152	0.188	−0.183	0.186	0.244	0.276		
Campaign attractiveness (10)	0.130	0.049	−0.050	−0.055	−0.118	0.299	0.235	0.118	−0.166	0.040	−0.148	
Campaign size (11)	0.051	0.046	0.213	−0.249	0.096	−0.056	0.307	0.537	0.262	−0.110	0.039	−0.032

Table A8 presents the regression results, including the posterior means of all parameters (τ, δ, σ). The posterior mean of the grand mean effect is $\tau = 0.111$, reaffirming positive structural state dependence on average. For context, the overall mean of the LATE distribution reported in Section 6 is $\tau = 0.100$ (see Table 2).

Table A8: Evidence Aggregation – Bayesian Meta-Regression

	Posterior Mean	95% CI	PD
Grand mean effect (τ)	0.111	[0.060, 0.162]	0.999
<i>Category:</i>			
Impulse (δ_1)	-0.073	[-0.136, -0.011]	0.989
Stockpile (δ_2)	-0.007	[-0.065, 0.051]	0.601
Category concentration (δ_3)	0.053	[0.005, 0.110]	0.976
Purchase frequency (δ_4)	0.155	[0.102, 0.210]	0.999
Customer penetration (δ_5)	-0.058	[-0.124, 0.011]	0.953
<i>Brand:</i>			
Basket penetration (δ_6)	-0.045	[-0.135, 0.043]	0.840
Market share (δ_7)	-0.013	[-0.110, 0.086]	0.607
Rel. price position (δ_8)	0.024	[-0.025, 0.073]	0.832
Weekly revenue (δ_9)	-0.088	[-0.173, -0.002]	0.978
Campaign attractiveness (δ_{10})	-0.062	[-0.124, -0.001]	0.978
Campaign size (δ_{11})	0.055	[0.006, 0.105]	0.985
<i>Heterogeneity:</i>			
Standard deviation (σ)	0.148	[0.088, 0.214]	–

Notes: PD (probability of direction) is the proportion of the posterior distribution that is of the median’s sign and can be mapped to frequentist p -values: $p_{\text{two-sided}} = 2 \times (1 - PD)$.

For category-level covariates, impulse buying score is negatively associated with state dependence, consistent with our prior that impulse buying is unlikely to foster loyalty. There is no evidence that stockpiling is associated with state dependence, lending credibility to our approach, which measures a future repurchase conditional on a same-category purchase, to control for stockpiling. Further, category concentration has a positive effect, suggesting that brand switching is easier in less-concentrated categories, which lends face validity. Purchase frequency is positively associated with state dependence, whereas the association is negative for customer penetration. The former association is plausibly due to higher customer involvement in frequently considered categories. The latter may reflect systematic differences between customer types purchasing across different categories. Specifically, customers in low-penetration categories may exhibit strong category preferences and brand loyalty. High-penetration categories, on the other hand, often include “necessities,” such as toilet paper or laundry detergent, and accordingly attract a more diverse customer base. This more diverse pool of customers may—on average—exhibit less brand loyalty than the customers in low penetration categories. We note that the latter two explanations are post-hoc, with correspondingly low scientific validity.

For the brand-level covariates, we find that weekly revenue and campaign attractiveness are negatively related to state dependence. This suggests that larger or more popular brands may be less associated with increasing habit formation. On the other hand, coupon penetra-

tion is associated with greater state dependence, suggesting that brands with higher coupon penetration may have more marginal consumers (i.e., compliers) who begin buying when exposed to coupons. The remaining variables are weakly correlated with the estimated state dependence effects, with a probability of direction < 0.85 .

The posterior mean of the standard deviation (σ) of the LATEs is 0.148, indicating unobserved heterogeneity across brands after controlling for the 11 explanatory variables in the regression. However, we observe a substantial reduction in the degree of unobserved heterogeneity compared to the base specification (see Table 2). Specifically, 65.5% of the variance in the brand-level LATEs is explained by the 11 covariates in the analysis.

Our results indicate that differences across brand-level LATEs are predictable based on category- and brand-level covariates. This analysis is useful for firms because it identifies observable brand and category characteristics that systematically relate to state dependence, allowing them to anticipate whether their brand is likely to generate repeat purchases. These insights can guide decisions on promotion design, coupon targeting, and brand positioning by highlighting which brand environments are more conducive to habit formation.

Web Appendix K Choice Model Parameters and Switching Costs

In this appendix, we summarize the mean and standard deviation of the price and the state dependence parameter for each estimated brand choice model. The relevant categories were apple juice, black tea (bags), butter, cat food (dry), cat food (wet), chocolate (tablets), coffee beans, condensed milk, cream, dog food (wet), energy drinks, fabric softener, frozen pizza, laundry detergent, margarine, mustard, orange juice, paper tissues, paper towels, toast, and toilet paper. The categories contain between 2 and 7 brands, and at least one brand per category is included in the coupon data. Table A9 reports relative switching costs by category and lists two popular brands per category. Figure A6 shows the estimates (μ and σ) for the state dependence (γ_i) and price parameters (β_i), including the 95% CI.

There is considerable variation in the results across categories for both parameters and in terms of the means and standard deviations of the heterogeneity distributions. Most mean values of the state dependence parameter are between 0.5 and 1, with corresponding standard deviations that imply that the overwhelming majority of the customers have a positive state dependence (i.e., are not variety seekers). This is in line with the results reported in Simonov et al. (2020) and Levine and Seiler (2023). Except for the category of condensed milk, all results imply price-sensitive customers, but the specific values depend, of course, on the scaling of the price variable.

To better understand the magnitude of the state dependence across (brands and) categories and to gauge the economic implications, we follow Dubé et al. (2010) and compute the “monetary value of loyalty” in Euros. We calculate the ratio between the state dependence and (negative) price parameter ($\gamma / -\beta$) to express the additional utility from buying a brand on the last category purchase occasion in monetary terms. This ratio can be interpreted as a switching cost (Dubé et al. 2010), even though there is no real monetary cost associated with switching brands in these categories (Kong et al. 2025, refer to “psychological switching costs”).

As different product categories have different price levels, we compute the ratio of the switching costs to the average category price and report the quantiles of the corresponding distribution. Specifically, for each of the 1,000 posterior draws of the population distribution, we simulate 10,000 customers and compute their switching costs, divided by the average category price. Across the simulated customers, we compute the quantiles of the distribution, and for each category, we report the posterior mean of each quantile.

The results in Table A9 show striking similarities with those reported by Dubé et al. (2010). While the lowest quantile shows slightly negative values, the median of the distributions often lies between 0.1 and 0.2 (average value 0.13), and for the highest quantile, we see values of about 0.5 to 0.7 for many categories (average value 0.66).

We also see meaningful variation across categories that makes intuitive sense and has face validity. For example, chocolate (tablets), as an impulse-driven category, has switching costs with a monetary value of loyalty of 3.3% of the average category price, whereas for laundry detergent, the (median) ratio is much higher with 27.0%. The latter value is remarkably

Table A9: Relative Switching Costs

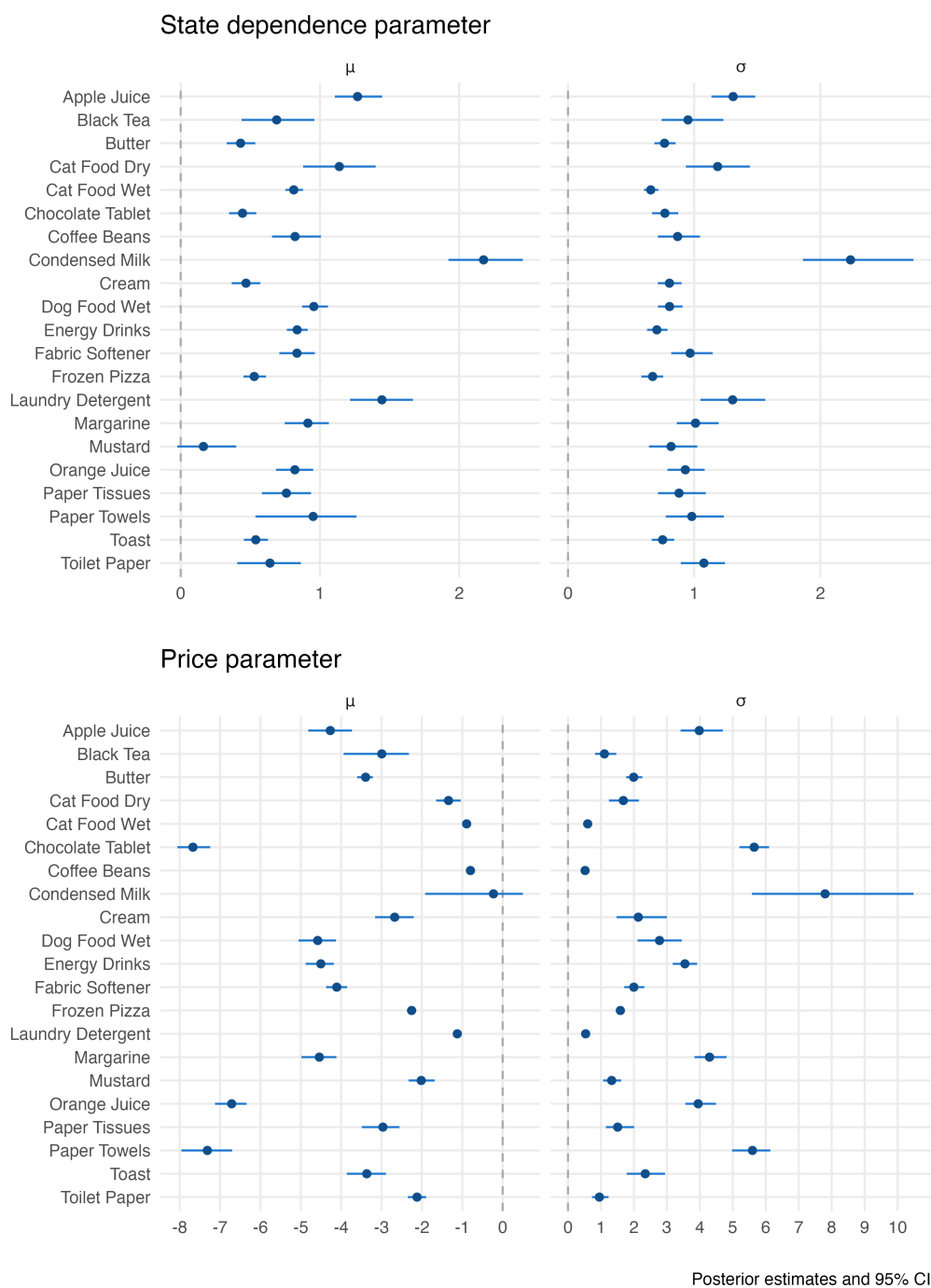
Category (<i>Popular Brands</i>)	Quantile (%)				
	10	25	50	75	90
Apple Juice (<i>Hohes C, Albi, ...</i>)	-0.417	-0.020	0.154	0.391	0.876
Black Tea (<i>Meßmer, Teekanne, ...</i>)	-0.117	0.007	0.132	0.272	0.445
Butter (<i>Kerrygold, Greenfields, ...</i>)	-0.172	-0.035	0.089	0.249	0.517
Cat Food Dry (<i>Whiskas, Purina, ...</i>)	-0.369	-0.032	0.114	0.281	0.627
Cat Food Wet (<i>Gourmet, Sheba, ...</i>)	-0.063	0.055	0.154	0.294	0.557
Chocolate Tablets (<i>Ritter Sport, Milka, ...</i>)	-0.106	-0.023	0.033	0.111	0.264
Coffee Beans (<i>Tchibo, Lavazza, ...</i>)	-0.065	0.011	0.074	0.160	0.320
Condensed Milk* (<i>Bärenmarke, Jeden Tag</i>)	-0.232	-0.040	0.237	0.813	2.100
Cream (<i>Weihenstephan, Frischgold, ...</i>)	-0.551	-0.109	0.171	0.516	1.131
Dog Food Wet (<i>Pedigree, Cesar, ...</i>)	-0.091	0.088	0.261	0.509	0.955
Energy Drinks (<i>Red Bull, Effect, ...</i>)	-0.166	0.029	0.143	0.314	0.667
Fabric Softener (<i>Lenor, Vernel, ...</i>)	-0.053	0.014	0.089	0.193	0.361
Frozen Pizza (<i>Dr. Oetker, Wagner, ...</i>)	-0.107	-0.010	0.066	0.180	0.405
Laundry Detergent (<i>Persil, Ariel, ...</i>)	-0.061	0.112	0.270	0.446	0.685
Margarine (<i>Arla, Becel, ...</i>)	-0.279	-0.022	0.087	0.233	0.538
Mustard (<i>Bautzner, Thomy, ...</i>)	-0.678	-0.232	0.081	0.411	0.911
Orange Juice (<i>Hohes C, Valensina, ...</i>)	-0.059	0.009	0.075	0.166	0.328
Paper Tissues (<i>Tempo, Regina, ...</i>)	-0.124	0.033	0.177	0.350	0.607
Paper Towels (<i>Zewa, Regina, ...</i>)	-0.045	0.027	0.071	0.121	0.209
Toast (<i>Golden Toast, Harry, ...</i>)	-0.266	-0.025	0.151	0.379	0.785
Toilet Paper (<i>Zewa, Hakle, ...</i>)	-0.128	-0.020	0.106	0.275	0.535

Notes: * With sign-constraint for the price parameter. Two popular brands are listed in parentheses (*italics*) for each category.

close to the reported 30% for the same category based on US data from the 1980s in [Osborne \(2011\)](#).

In summary, we find meaningful switching costs with relevant economic implications for the different categories in our analysis. We find heterogeneity in switching costs across customers but also across categories, and the magnitudes of the results align with the literature.

Figure A6: Selected Estimation Results of the Brand Choice Model for Each Category



Web Appendix L Sensitivity of Brand Choice Model LATEs to Discount Depth

To assess the sensitivity of the brand-choice LATEs to the assumed discount depth in the simulation (i.e., 25%), we conduct a robustness check for one illustrative category: apple juice.

This category contains multiple brands, ranging from private labels to large national brands (e.g., “Hohes C”). Based on the estimated parameters, most customers are price-sensitive and exhibit state dependence, and the simulated LATEs for the baseline 25% discount lie between 0.1 and 0.2 across brands. We then vary the discount depth from 5% to 50% in 5-percentage-point increments and recompute LATEs for each brand.

Figure A7: Choice Model LATEs for Different Discount Values in the Apple Juice Category

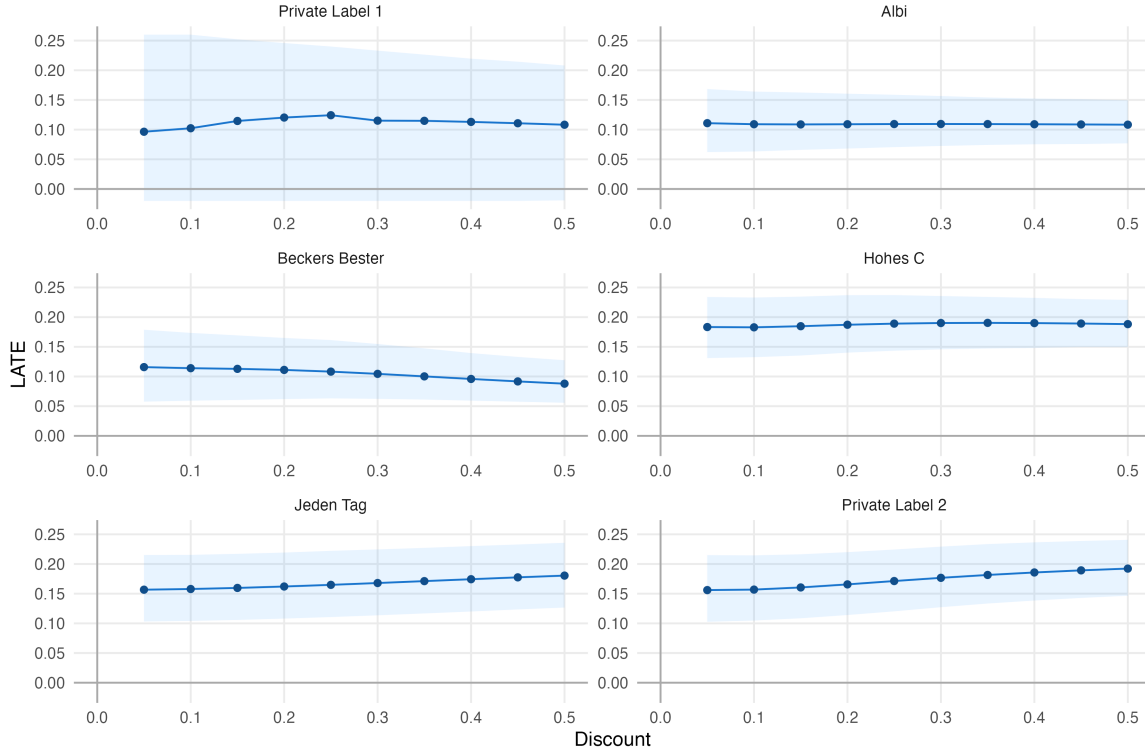


Figure A7 reports the resulting point estimates and 95% credible intervals. The LATEs are essentially flat in the discount depth: for all six brands, we do not observe a statistically meaningful deviation from a horizontal line, implying no systematic relationship between LATEs and discount size.

Web Appendix M LATE From Brand Choice Model Estimates

To compare the results of the discrete choice model with our nonparametric estimates for state dependence, we derive model-based LATEs for state dependence through simulations that are based on the estimated parametric brand choice models. This web appendix summarizes our approach.

We start the simulation of local average treatment effects by drawing from the population distribution for μ and Σ . Because we need the full distribution of all parameters across all brands, we first adjust the heterogeneity parameters of the intercepts (of brands with target coupons). We standardize our control function to simplify interpretation. Assuming average price reductions (i.e., the control functions) r_i and brand preferences α_{ij} for each brand j to follow a joint normal distribution with correlation ρ_j , we can write the utility function of our extended model as

$$u_{ijt} = a_j + \beta_i \cdot p_{ijt} + \gamma_i \cdot \mathbb{I}\{s_{it} = j\} + \sigma_{\alpha_j} \cdot \rho_j \cdot \sigma_{r_j}^{-1} \cdot r_{ij} + \sigma_{\alpha_j} \cdot \sqrt{1 - \rho_j^2} \cdot \iota_{ij} + \epsilon_{ijt},$$

where α_{ij} is decomposed into the brand-specific mean value a_j and the customer-specific standard-normal deviation from the mean ι_{ij} . Conditional on the r_{ij} , the net price p_{ijt} and the brand preference α_{ij} are uncorrelated. The product terms $\sigma_{\alpha_j} \cdot \rho_j$ and $\sigma_{\alpha_j} \cdot \sqrt{1 - \rho_j^2}$ are additional parameters. We use posterior draws for both product terms to recover the posterior draws of σ_{α_j} (and ρ_j).

Next, we assume all brand prices to be at the average observed value (without discounts). For each posterior draw of μ and Σ , we simulate 10,000 customers and compute the steady-state distribution of s_{it} for each customer. Then, we compute the effect of a 25% price reduction for the same period and for the next one. Based on these conditional brand choice probabilities and the steady-state distribution of the loyalty state, we can compute the local average treatment effect. Repeating this procedure for all 1,000 posterior draws gives us the posterior distribution of the brand choice model-based local average treatment effects. We repeat this computation for each brand in a category, irrespective of whether there were experiments with random coupons.

We use the following equations to simulate the relevant probabilities that we use to construct the LATE. The conditional choice probability of customer i for the brand j with coupon availability (or a price shock) for the focal brand $c = 0, 1$ and loyalty state $s = 1, \dots, J$ is:

$$Pr\{y_{it} = j \mid c_{it}, s_{it}, \theta_i\} = \sigma_{ij}(c, s).$$

To simplify interpretation, we define WLOG brand $j = 1$ as the focal brand. Hence, $\sigma_{i1}(0, 1)$ would be the conditional choice probability for the focal brand without a coupon and loyalty for this brand. On the other hand, $\sigma_{i2}(1, 3)$ would be the conditional choice probability for brand 2 when there is a coupon for the focal brand and loyalty for brand 3.

Let $j = 0$ be the outside option, and we assume – as in the main paper – that (i) there is no loyalty w.r.t. the outside option and (ii) the loyalty state does not change when the outside

option is chosen ($s_{t+1} = s_t$).

We further define $\pi_i(s)$ as the stationary distribution of the loyalty states. Given the assumptions about the market condition in the simulation, we solve for $\pi_i(s)$ for each posterior draw.

Our goal is to derive all quantities for

$$\begin{aligned} \text{LATE}_{D \rightarrow Y} &= \frac{\text{ITT}_{Z \rightarrow Y}}{\text{ATE}_{Z \rightarrow D}} \\ &= \frac{\mathbb{E}[Y(D(1)) - Y(D(0))]}{\mathbb{E}[D(1) - D(0)]} \end{aligned}$$

from our HB-MNL model.

We start with the case of $J = 2$:

$$\begin{aligned} \mathbb{E}[Y(D(1))] &= \frac{1}{I} \sum_{i=1}^I \sum_{s=1}^2 \pi_i(s) (\sigma_{i1}(1, s) \cdot \sigma_{i1}(0, 1) + \sigma_{i2}(1, s) \cdot \sigma_{i1}(0, 2) + \sigma_{i0}(1, s) \cdot \sigma_{i1}(0, s)), \\ \mathbb{E}[Y(D(0))] &= \frac{1}{I} \sum_{i=1}^I \sum_{s=1}^2 \pi_i(s) (\sigma_{i1}(0, s) \cdot \sigma_{i1}(0, 1) + \sigma_{i2}(0, s) \cdot \sigma_{i1}(0, 2) + \sigma_{i0}(0, s) \cdot \sigma_{i1}(0, s)) \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}[D(1)] &= \frac{1}{I} \sum_{i=1}^I \sum_{s=1}^2 \pi_i(s) \sigma_{i1}(1, s), \\ \mathbb{E}[D(0)] &= \frac{1}{I} \sum_{i=1}^I \sum_{s=1}^2 \pi_i(s) \sigma_{i1}(0, s). \end{aligned}$$

Next, extending to the general case of J brands, collecting terms, and combining everything leads to the following expression for the LATE of the focal brand $j = 1$:

$$\text{LATE} = \frac{\frac{1}{I} \sum_{i=1}^I \sum_{s=1}^J \pi_i(s) \cdot \left(\sum_{j=1}^J (\sigma_{ij}(1, s) - \sigma_{ij}(0, s)) \cdot \sigma_{i1}(0, j) + (\sigma_{i0}(1, s) - \sigma_{i0}(0, s)) \cdot \sigma_{i1}(0, s) \right)}{\frac{1}{I} \sum_{i=1}^I \sum_{s=1}^J \pi_i(s) \cdot (\sigma_{i1}(1, s) - \sigma_{i1}(0, s))}.$$

Web Appendix N ATE From Brand Choice Model Estimates

The brand choice model also allows us to compute ATE values directly. This is the difference in brand choice probability for the focal brand $j = 1$ with state dependence compared to a scenario without state dependence. However, compared to a binary case, where the focal brand either has or does not have the state dependence status, the multinomial case is more complex. Here, each of the other $J - 1$ serves as a benchmark.

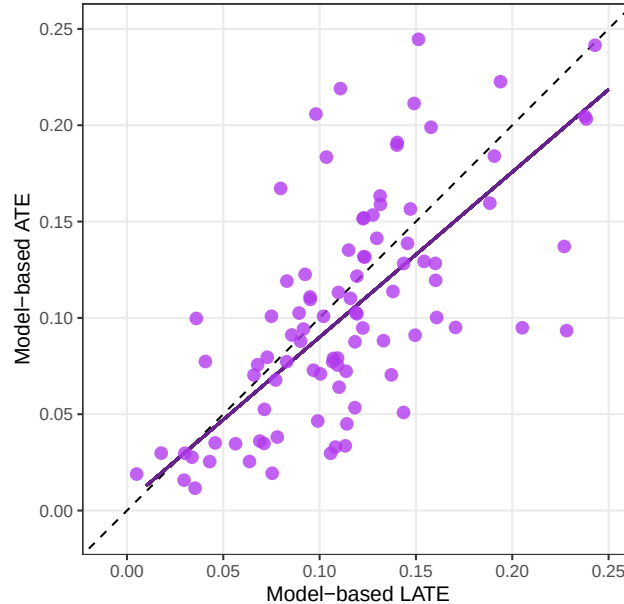
Using the same notation as in Web Appendix M, we define the ATE of state dependence as a weighted average over all pairwise comparisons of the loyalty state for the focal brand with all other loyalty states:

$$ATE = \mathbb{E}[Y(1) - Y(0)] = \frac{1}{I} \sum_{i=1}^I \sum_{j=2}^J \omega_i(j) \cdot (\sigma_{i1}(0, 1) - \sigma_{i1}(0, j)).$$

As weight for $\omega_i(j)$, we use $\frac{\pi_i(j)}{1 - \pi_i(1)}$, i.e., the normalized *other brand* steady state probabilities. Note that $\omega_i(j) = 1$ for the binary case with $J = 2$.

Figure A8 shows the relationship between the posterior means of the model-predicted LATEs (see main text) and ATEs computed as described above. The association is positive, with a correlation of $\rho = 0.748$ (95% CI [0.622, 0.845]) using the approach by [Matzke et al. \(2017\)](#) to account for measurement error. The mean model-based ATE is 0.101, which is about 10% smaller than the mean model-based LATE, 0.113. Hence, the model-based comparison reveals that the LATEs and ATEs have similar magnitudes.

Figure A8: Distribution of Model-Based LATE and ATE Estimates ($N = 93$)



Web Appendix O Bayesian Meta-Analysis: Extensions

We use a (hierarchical) Bayesian meta-analysis model to statistically aggregate the brand-level results. The main model specification is described in Section 6. Here, we briefly present three extensions used in Sections 8 and Web Appendix J.

We implemented all models in Stan and estimated the posterior distribution of unknown parameters using Hamiltonian Monte Carlo and the No-U-turn sampler. We ran five chains in parallel for 10,000 iterations, discarded the first 5,000 for warm-up, and saved every fifth draw to have 5,000 draws in total for posterior inference (ESS > 1,000). Convergence was assessed via visual inspection of trace plots as well as by checking the Gelman-Rubin $\hat{R}$ -statistic (all values < 1.01).

O.1 Three-Level Meta-Analysis

The first extension of the 2-level model is a specification with three levels. In this specification, the J brand-level experiments belong to a smaller number of K non-overlapping groups, in our case categories. In that case, we can add a third level to the hierarchical model as follows:

$$\begin{aligned}\widehat{\text{LATE}}_j &\sim N(\tau_j, \hat{s}_j^2), \text{ with} \\ \tau_j &\sim N(\tau_{k[j]}, \sigma_{\tau_j}^2), \\ \tau_k &\sim N(\tau, \sigma_{\tau_k}^2), \\ \tau &\sim N(0, 0.5^2) \text{ and } \sigma_{\tau_j}, \sigma_{\tau_k} \sim N_+(0, 0.5^2).\end{aligned}$$

σ_{τ_j} represents the amount of heterogeneity across brands within a category k , and σ_{τ_k} is the amount of heterogeneity across category means.

O.2 Meta-Analysis With Correlation

A second extension is necessary for comparing the results between different model specifications or approaches for obtaining the LATE values (e.g., nonparametric LATEs vs. choice model LATEs). Correlating the LATE point estimates is problematic because the measurement errors in both dimensions lead to attenuation bias. It might be tempting to run two meta-analyses separately and compare the correlation of the shrunken values $\hat{\tau}_{j1}$ and $\hat{\tau}_{j2}$ descriptively. However, $\hat{\tau}$ is still measured with error. Instead, we can combine both meta-analyses and use a bivariate normal distribution for the heterogeneity (see also [Matzke et al. 2017](#)):

$$\begin{bmatrix} \tau_{j1} \\ \tau_{j2} \end{bmatrix} \sim N \left(\begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix} \right).$$

ρ directly measures the correlation between the “shrunken” LATE estimates. We use a flat

hyper-prior for this correlation ($\rho \sim U(-1, 1)$).

O.3 Meta-Regression

A third extension is a Bayesian (hierarchical) meta-regression model, where the additional vector x contains category and brand-specific variables for each j , and the parameters in δ map the variables on the latent values τ_j :

$$\begin{aligned}\tau_j &\sim N(\tau + x'_j \delta, \sigma^2) \\ \tau, \delta &\sim N(0, 0.5^2) \text{ and } \sigma \sim N_+(0, 0.5^2) .\end{aligned}$$

Web Appendix P Combination of Structural and Experimental Estimates: Bias Estimation

Given our setup with only 39 brands, we use a simple and transparent approach to estimate b^2 from cross-brand variation. Observing J brands allows us to separate systematic bias from sampling noise, following the logic of the [DerSimonian and Laird \(1986\)](#) random-effects estimator.

For each brand j , we define the discrepancy

$$\delta_j = \hat{\tau}_{b,j} - \hat{\tau}_{u,j}$$

and model this as

$$\delta_j = b_j + \varepsilon_j,$$

with $\mathbb{E}[\varepsilon_j] = 0$, so that

$$\text{Var}(\delta_j) = \text{Var}(b_j) + \text{Var}(\varepsilon_j).$$

Here b_j denotes the (potentially heterogeneous) bias for brand j ; the scalar b^2 used in the shrinkage weight corresponds to the expected squared bias $\mathbb{E}[b_j^2]$ across brands.

The cross-sectional variance of the discrepancies is

$$S_\delta^2 = \frac{1}{J-1} \sum_{j=1}^J (\delta_j - \bar{\delta})^2.$$

We approximate $\text{Var}(\varepsilon_j)$ by

$$\bar{v} = \frac{1}{J} \sum_{j=1}^J (\hat{\sigma}_{b,j}^2 + \hat{\sigma}_{u,j}^2),$$

where $\hat{\sigma}_{b,j}^2$ and $\hat{\sigma}_{u,j}^2$ are the posterior sampling variances from the structural and experimental estimators, respectively.³⁰ The variance of the bias across brands is then estimated as

$$\widehat{\text{Var}}(b) = \max \{ S_\delta^2 - \bar{v}, 0 \}.$$

Because the expected squared bias for a randomly drawn brand satisfies

$$\mathbb{E}[b_j^2] = \text{Var}(b_j) + (\mathbb{E}[b_j])^2,$$

and as $\mathbb{E}[b_j]$ is typically small in our setting, we use

$$\widehat{\mathbb{E}[b^2]} \approx \widehat{\text{Var}}(b)$$

in the shrinkage weight formula.

This variance-decomposition estimator parallels the classic random-effects meta-analytic approach and provides a simple, data-driven measure of the magnitude of structural misspecification needed for our MSE-optimal combination of estimators.

³⁰This expression assumes zero sampling covariance between the structural and experimental estimators. With nonzero covariance, $\text{Var}(\varepsilon_j) = \hat{\sigma}_{b,j}^2 + \hat{\sigma}_{u,j}^2 - 2 \text{Cov}(\hat{\tau}_{b,j}, \hat{\tau}_{u,j})$.

Web Appendix References

- Abadie A (2002) Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the American Statistical Association* 97(457):284–292.
- Chernozhukov V, Hansen C, Kallus N, Spindler M, Syrgkanis V (2024) Applied Causal Inference Powered by ML and AI.
- Chintagunta P, Dubé JP, Goh KY (2005) Beyond the Endogeneity Bias: The Effect of Unmeasured Brand Characteristics on Household-Level Brand Choice Models. *Management Science* 51(5):832–849.
- DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. *Controlled Clinical Trials* 7(3):177–188.
- Dubé JP, Hitsch GJ, Rossi PE (2010) State dependence and alternative explanations for consumer inertia. *The RAND Journal of Economics* 41(3):417–445.
- Efron B (2016) Empirical Bayes deconvolution estimates. *Biometrika* 103(1):1–20.
- Gabel S, Simester D, Timoshenko A (2025) In-Store Coupon Effectiveness: A Large-Scale Field Experiment. Available at SSRN 4335525.
- Goldfarb A (2006) State Dependence at Internet Portals. *Journal of Economics & Management Strategy* 15(2):317–352.
- Imbens GW, Angrist JD (1994) Identification and estimation of Local Average Treatment Effects. *Econometrica* 62(2):467–475.
- Imbens GW, Wooldridge JM (2009) Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* 47(1):5–86.
- Keane MP (1997) Modeling Heterogeneity and State Dependence in Consumer Choice Behavior. *Journal of Business & Economic Statistics* 15(3):310–327.
- Kong X, Dubé JP, Daljord Ø (2025) Nonparametric Estimation of Demand with Switching Costs: The Case of Habitual Brand Loyalty. Available at SSRN 4070747.
- Levine J, Seiler S (2023) Identifying State Dependence in Brand Choice: Evidence from Hurricanes. *Marketing Science* 42(5):934–957.
- Matzke D, Ly A, Selker R, Weeda W, Scheibehenne B, Lee M, Wagenmakers EJ (2017) Bayesian Inference for Correlations in the Presence of Measurement Error and Estimation Uncertainty. *Collabra: Psychology* 3(1):25.
- Osborne M (2011) Consumer learning, switching costs, and heterogeneity: A structural examination. *Quantitative Marketing and Economics* 9:25–70.
- 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.
- Stock JH, Yogo M (2005) Testing for Weak Instruments in Linear IV Regression. Andrews DWK, Stock JH, eds., *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, 80–108 (Cambridge: Cambridge University Press).
- Wernerfelt N, Tuchman A, Shapiro B, Moakler R (2025) Estimating the Value of Offsite Tracking Data to Advertisers: Evidence from Meta. *Marketing Science* 44(2):268–286.
- Willwerscheid J, Carbonetto P, Stephens M (2021) ebnm: An R package for solving the empirical bayes normal means problem using a variety of prior families. <https://arxiv.org/abs/2110.00152>.