The Welfare Impact of Delegating Choice to Large Language Models*

Yann Delaprez† Ali Hortaçsu‡

March 26, 2026

[*Preliminary draft — incomplete. Please do not circulate.*]

Abstract

What are the equilibrium and welfare effects of delegating decision-making to large language models (LLMs)? We study this question by comparing human and LLM-generated purchase choices around the introduction of menu-labeling laws in a major restaurant market. We find that menu labeling had limited effects on calorie intake. By contrast, delegating purchase decisions to ChatGPT would reduce total calorie intake by up to 18%. Delegation also induces substantial exit of unsold product varieties and increases market concentration. We develop a structural model of delegation to quantify the underlying mechanisms and welfare effects. The net effect of delegation on total welfare is positive but has important distributional consequences: consumer surplus rises substantially for those who delegate to large language models, while the surplus of non-delegating consumers declines slightly due to equilibrium effects. Firm profits also increase.

Keywords: Large Language Models and Food Choices; Consumer Food Choices; Delegation

*We thank Li You and Sania Zeb for excellent research assistance. We thank John Rust for helpful comments. The conclusions drawn from the Numerator data are those of the researcher(s) and do not reflect the views of Numerator. Numerator is not responsible for and had no role in analyzing or preparing the results reported herein. Researcher(s)' own analyses calculated (or derived) based in part on data from Market Track, LLC dba Numerator and marketing databases provided through the Numerator Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. At no point during this research project were any Numerator or Syndigo data, or any data provided by the Kilts Center, shared with ChatGPT.

†University of Chicago: delaprez@uchicago.edu.

‡University of Chicago: hortacsu@uchicago.edu.

1 Motivation

Large Language Models (LLMs) contain extensive social and informational knowledge, often outperforming humans on specific tasks Horton (2023). While existing research has primarily focused on the role of AI assistance in workplace settings—for instance, examining its impact on worker productivity Brynjolfsson et al. (2025); Cui et al. (2025)—this article shifts the focus to consumers, investigating whether delegating consumption decisions to LLMs could lead to welfare improvements and healthier food choices.

Food choices have important implications for public policy through their effects on health. In the United States, obesity affects over 40% of adults and is associated with more than $170 billion in annual medical expenditures, accounting for a substantial share of total health care spending Centers for Disease Control and Prevention (2022); Finkelstein et al. (2009). Poor diet quality independently contributes to disease burden, with diets high in sodium, calories, and saturated fats linked to nearly $50 billion per year in cardiometabolic health care costs Jardim et al. (2019). These externalities motivate policy interventions aimed at improving dietary choices, including mandatory calorie disclosure in restaurant under menu labeling laws.

This paper begins by studying consumption patterns in pizza restaurants. We document a substantial misalignment between observed purchases and dietary recommendations. On average, purchased items contain 1,797 kcal and 4,381 mg of sodium, and a large share exceed standard daily reference thresholds: 65.1% of items exceed 2,000 kcal and 76.0% exceed 2,300 mg of sodium, as defined by the U.S. Department of Agriculture and U.S. Department of Health and Human Services (2020). Two broad explanations may account for this misalignment. First, consumers may have imperfect information about nutritional attributes (Allcott et al. (2019)). Second, consumers may face a self-control problem, whereby short-term preferences diverge from long-run health goals (Gul and Pesendorfer (2001)).

To shed light on these explanations, we exploit variation from the nationwide menu-labeling policy in pizza restaurants. Since May 2018, the law has required major restaurant chains with more than 20 locations to display calorie information on menus to promote more informed choices. We find little evidence that menu labeling significantly altered consumers' dietary choices, suggesting that information frictions are unlikely to be the primary driver and that self-control problems may be more relevant. This also raises the question of whether delegating food choices to an agent better aligned with nutritional guidelines could improve outcomes by acting as a commitment device. Our premise is that large language models (LLMs) may play such a role. Despite the substantial public health implications, this possibility has received limited attention in the economic literature Stanaway et al. (2018).

In this article, we address this question using a causal and quantitative framework. We conduct a purchase experiment in which ChatGPT, a state-of-the-art LLM, makes decisions over choice sets comparable to those faced by human consumers, and we benchmark these choices

against observed human responses to the menu-labeling law. Our experiment identifies three key margins through which delegating food decisions to ChatGPT affects welfare and nutrition. First, delegation triggers the exit of up to 10% of unsold product varieties. This is consistent with the view that ChatGPT acts as a commitment device by systematically steering consumers away from certain products. Second, it increases market concentration. Third, delegation has a mixed nutritional impact: on average, it lowers total calorie intake but also reduces fiber consumption, a nutrient that is already under-consumed. Taken together, these findings show that delegation not only alters how closely consumption aligns with nutritional guidelines, but also generates important indirect equilibrium effects that must be quantified.

We quantify the welfare impact of delegation by incorporating our causal estimates into a novel Nash–Bertrand pricing model with partial delegation, which enables counterfactual analysis. In the model, market demand is a weighted mixture of non-delegated and delegated choices, governed by the share of consumers who delegate. Our modeling of the U.S. pizza restaurant market highlights several general features that we believe are relevant for understanding the effects of LLM delegation in retail purchasing more broadly. First, delegation reduces the number of products each restaurant offers, limiting firms' ability to raise prices by exploiting substitution within their own portfolios. In isolation, this mechanism lowers average markups and prices. Second, delegation generates a reallocation effect that increases prices and markups. We find that this latter force dominates. Nevertheless, the net effect on total welfare is positive, driven by lower own-price elasticities of demand and improved diet quality under delegation. This pattern has important distributional implications: while the consumer surplus of consumers who delegate to large language models rises substantially, the surplus of non-delegating consumers declines slightly because equilibrium prices increase for all, even for those whose preferences remain unchanged.

Our results fill an important gap in the economics of artificial intelligence. To our knowledge, this paper is the first to empirically assess whether consumers benefit from delegating food purchase decisions to large language models in real-world settings. Prior work by Chen et al. (2023) studies whether ChatGPT behaves rationally using highly stylized choice experiments, such as binary decisions between tomatoes and meat. In contrast, we analyze ChatGPT's purchasing behavior when it faces the same choice sets as human consumers, enabling a direct comparison of substitution patterns between humans and an LLM. We find that ChatGPT is less responsive to prices and exhibits substantially narrower substitution patterns than humans. In the extreme case where firms set prices facing an LLM as the consumer, this behavior would imply significantly higher markups. Related work on the market implications of AI includes Demirer et al. (2025), who examine the evolution of LLM markets for business applications rather than consumer decision-making, and Metcalfe et al. (2025), who study whether AI adoption improves demand-side energy efficiency in electricity markets.

We focus on the U.S. pizza market and on large language models used for consumer decision-

making because this setting is both policy relevant and well suited to fill key gaps in the literatures on public health and the persistence of food habits. First, a substantial share of pizza orders are placed online, making delegation to an LLM practically feasible and the question immediately policy-relevant. Second, evidence on the nutritional effects of labeling policies in restaurants remains sparse relative to supermarkets. Studies that examine restaurants typically focus on a single chain, limiting the ability to assess how competition shapes policy outcomes Wisdom et al. (2010); Finkelstein et al. (2011); Bollinger et al. (2011). More broadly, food markets with important health implications are predominantly studied using supermarket purchase data, largely due to data availability Barahona et al. (2025); Dubois et al. (2018). We overcome this limitation by exploiting a novel dataset that covers purchases across the major U.S. pizza chains. Because pizza restaurants constitute a core segment of the fast-food sector, estimates that jointly inform competition and nutrition in this market are of first-order policy relevance. Finally, the muted behavioral responses we document contribute to a growing literature showing that dietary choices are highly persistent (see Hut and Oster (2022)). Our findings complement recent evidence from Barahona et al. (2025), who study a front-of-package labeling policy in Chilean supermarkets and find a 6% reduction in caloric intake. In contrast, our results suggest that dietary habits are substantially more rigid in restaurant settings.

The remainder of the article is organized as follows. Section 2 describes the institutional setting and the data. Section 3 examines the alignment between consumption choices and dietary habits. Section 4 analyzes the impact of the menu-labeling law on calorie intake. Section 5 provides causal estimates of the effects of delegating food choices to large language models (LLMs). Section 6 develops and estimates a model of consumer demand and firm supply. Section 7 combines the estimated model and causal evidence to conduct counterfactual analyses, including the welfare and consumption effects of delegating food choices to ChatGPT. Section 8 explores potential underlying mechanisms. Section 9 concludes.

2 Institutional Setting and Data

2.1 The Food Menu Labeling Policy

The Affordable Care Act (ACA) included a provision (Section 4205) requiring chain restaurants and similar retail food establishments with 20 or more locations to post calorie information on menus and menu boards. After May 2018, it is compulsory to report daily calorie intake, indicating that "2,000 calories a day is used for general nutrition advice, but calorie needs vary". The FDA began enforcing the rule on May 7, 2018. However, Pizza Hut had already implemented menu calorie labeling before the policy took effect. Also before the nation wide policy some localities already implemented the policy, in our descriptive evidences these localites are excluded

from the analysis.[1] Figure 1 provides an illustrative example of a menu display complying with the menu labeling law, where calorie information is posted alongside menu items at the point of purchase.



Figure 1. Example of Menu Display under the Menu Labeling Law

Notes: The figure illustrates a representative menu display showing calorie information as required under the U.S. menu labeling regulation. Calorie counts are posted alongside menu items at the point of purchase.

2.2 Relevant Data

We rely on two primary data sources. The datasets used to analyze ChatGPT's choices are compiled from publicly available online sources, while comparisons with actual consumer purchases rely on transaction data from Numerator.

ChatGPT Experiments We conduct our experiment via the OpenAI API, prompting ChatGPT to make food purchase decisions across different scenarios. We collected nutritional information and product prices from publicly available sources. Price data were obtained from the official websites of each chain using postal code 60611 corresponding to downtown Chicago. Data collection occurred between October and December 2025. Next, prices were deflated using the monthly Consumer Price Index for food (CPIUFDSL), normalized to 2018, using monthly data

[1] See. Table 1 in Courtemanche et al. (2025) for a detailed summary.

from FRED to generate real price variation.[2] In Appendix A, we present descriptive statistics for these data.

Actual Purchase We use purchase data from U.S. consumers covering 2018, drawn from the Numerator Dataset. Our goal is to compare choice behavior and substitution patterns between ChatGPT and human consumers. Hence, in the Numerator dataset, we select the products for which we were able to collect information in our experiments. The dataset captures transactions across both e-commerce and traditional retail channels, including detailed information on retailer identity, product quantity, unit price, and product description. For the structural estimation and experimental analysis, given our focus on the national menu labeling law, we restrict attention to purchases from the three largest U.S. pizza restaurant chains operating at the national level—Domino's, Papa John's, and Pizza Hut. These are widely recognized as the main national pizza chains; for example, in its 2018 Annual Report, Domino's explicitly identifies Pizza Hut® and Papa John's® as its primary national competitors Domino's Pizza (2019). We exclude sales from local pizza restaurants, as they fall outside the scope of the federal policy. Our analysis considers product categories that are plausibly substitutable, namely *Pizza*, *Wings*, *Pasta*, *Salad*, and *Sandwiches*. Descriptive statistics are shown in Appendix B. For the descriptive analysis, we broaden the sample to include purchases from the two largest retailers in our sample, Walmart and Kroger, which serve as the control group in the event study presented in Section 4. Nutritional information for products sold by retailers is obtained from the Syndigo database, while nutritional information for restaurant items is collected from the official websites of the pizza chains.

3 Alignment of Consumption with Dietary Guidelines

We first study the extent to which individuals' consumption aligns with nutritional guidelines. To do so we assess the nutritional content of items purchased at pizza restaurants under two complementary measurement conventions: a *per-item* perspective and a *per-user* (*household-adjusted*) perspective. The analysis covers the full set of food items in the main menu categories typically offered at pizza restaurants.

The per-item measure reflects the nutritional load associated with selecting a single menu item, consistent with information disclosed under U.S. menu labeling regulations. However, because many pizza-restaurant items are shared within households, per-item measures likely *overstate* individual nutritional intake. To address this, we also construct a per-user measure by dividing menu-reported nutrients by household size reported in the transaction data. This household-adjusted measure provides a conservative approximation to nutritional exposure

[2] See https://fred.stlouisfed.org/series/CPIUFDSL.

per person and likely *understates* individual intake when consumption within the household is uneven. Together, the two measures bracket plausible individual nutritional exposure.

To assess alignment with dietary guidance, we compute saturated fat intensity, defined as the share of total calories attributable to saturated fat, and indicators for whether nutrient levels exceed commonly used daily reference thresholds. Saturated fat intensity is calculated as

$$\%kcal_satfat_i = 100 \times \frac{9 \cdot SatFat_i}{Cal_i},$$

where $SatFat_i$ denotes grams of saturated fat and Cal_i denotes total calories. The factor of 9 corresponds to the Atwater conversion factor of 9 kilocalories per gram of fat. We flag whether saturated fat exceeds 10% of total calories and whether sodium and calories exceed 2,300 mg and 2,000 kcal, respectively, consistent with the *Dietary Guidelines for Americans, 2025–2030*, which recommend limiting saturated fat intake to less than 10% of total energy and sodium intake to less than 2,300 mg per day U.S. Department of Health and Human Services and U.S. Department of Agriculture (2026). Because both saturated fat and calories are scaled by household size, saturated fat intensity is invariant to the per-user transformation.

Table 1 reports summary statistics under both measurement conventions. On a per-item basis, the average purchased item contains 1,797 kcal and 4,381 mg of sodium, with 65.1% of items exceeding 2,000 kcal and 76.0% exceeding 2,300 mg of sodium. On a per-user basis, average intake declines substantially to 717 kcal and 1,749 mg of sodium, and the shares exceeding daily reference thresholds fall to 7.7% for calories and 23.9% for sodium. In contrast, saturated fat intensity remains high under both measures: the mean intensity is 18.4% of calories, and over 96% of purchases exceed the 10% guideline. Even under the conservative per-user measure, a substantial share of pizza-restaurant purchases remain misaligned with federal dietary guidelines. To assess whether this reflects an information problem or a self-control problem, we next exploit the menu-labeling policy as an exogenous information shock and test whether consumers respond to the provision of information.

Table 1. Nutritional characteristics of consumption at pizza restaurants

	Per item	Per user (household-adjusted)
Mean (SD)		
Calories (kcal)	1,797 (925)	717 (619)
Sodium (mg)	4,381 (2,068)	1,749 (1,449)
Saturated fat (g)	37.42 (20.48)	14.91 (13.28)
Saturated fat (% of calories)	18.41 (3.51)	18.40 (3.51)
Cholesterol (mg)	231.58 (102.83)	92.51 (74.63)
Share exceeding threshold		
Saturated fat $>$ 10% of calories	0.9635	0.9632
Sodium $\geq 2{,}300$ mg	0.7596	0.2388
Calories $\geq 2{,}000$ kcal	0.6511	0.0767
Trans fat > 0	0.5057	0.5026
Observations	1,946,615	1,742,306

Notes: Per-item nutrients are based on menu-reported values for each purchased item. Per-user nutrients divide per-item values by transaction household size; observations with missing household size are excluded. Saturated fat as a share of calories is computed using the Atwater conversion factor of 9 kcal per gram of fat and is invariant to the household adjustment. Thresholds correspond to the *Dietary Guidelines for Americans, 2025–2030* U.S. Department of Health and Human Services and U.S. Department of Agriculture (2026).

4 Menu Labeling: Impact on Calories Consumed

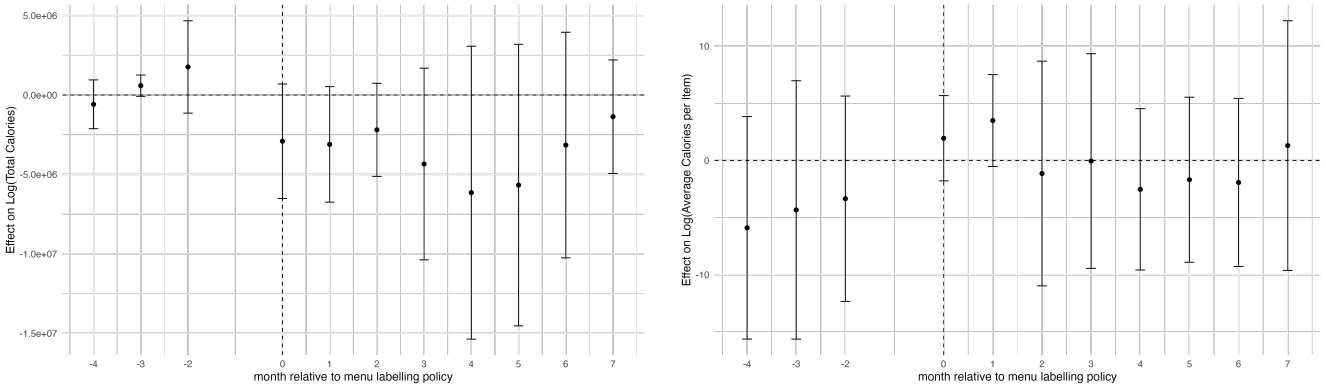
We begin by estimating the effect of the menu labeling policy on calorie intake. We estimate the following specification at the *banner* × *product-group* × *month* level:

$$y_{bgt} = \sum_{k \neq -1} \beta_k \, \mathbb{1}\{\text{rel_time}_t = k\} \times \text{Treat}_b \; + \; \lambda_t \; + \; \mu_{bg} \; + \; \varepsilon_{bgt}, \tag{1}$$

where y_{bgt} is either $\log(\text{total calories})$ or $\log(\text{average calories per item})$ for banner b (either restaurants or retailers), product group g, and month t. Treat_b is an indicator equal to one for treated restaurant chains (Domino's and Papa John's) and zero otherwise. Pizza Hut began displaying calorie information prior to the federal menu labeling deadline of May 7, 2018. Because it had already implemented calorie labeling before the policy went into effect, we exclude Pizza Hut from our treated sample to preserve the pre-policy control period. The control group consists of purchases of *similar product categories* (pizza, sandwiches, and wings) sold through supermarkets—specifically Kroger and Walmart. These products are likely similar to those offered by restaurants; however, unlike restaurant menu items, calorie information is already displayed on packaged supermarket products both before and after the policy. Another advantage of comparing products sold in restaurants and supermarkets is that spillover between treated and untreated groups is likely to be limited. The specification includes calendar-month fixed effects, λ_t, which absorb common shocks and seasonality in food purchasing, and banner-by-group fixed effects, μ_{bg}, which control for time-invariant differences in product mix and purchasing intensity across retailers and product categories. Event time rel_time_t is measured in months relative to the policy month (May 2018), and the coefficient for $k = -1$ (the month immediately preceding the policy) is omitted and serves as the reference period. Standard errors are clustered at the banner-product group level.

Figure 2 presents the event-study estimates. Panel A reports results using $\log(\text{total calories})$ as the outcome, where total calories are defined as the sum of (quantity × calories) within a banner–group–month. Panel B reports results using $\log(\text{average calories per item})$, defined as total calories divided by total items purchased within a month. Across both panels, pre-policy coefficients are small and statistically indistinguishable from zero, providing no evidence of differential pre-trends between treated restaurants and supermarket controls. Following the policy, Panel A shows no clear evidence of a reduction in total calories purchased. Similarly, Panel B shows no evidence of changes in average calories per item. This contrasts with the findings of Barahona et al. (2025), who study the impact of a front-of-package labeling policy implemented in Chile in supermarkets and find that consumers reduced their caloric intake by 6%. In Appendix C, we also document a lack of evidence of changes in caloric intake per dollar spent. One possible explanation for the greater persistence of dietary habits in restaurant settings is that restaurant purchases are more spontaneous and often made under time constraints, leaving consumers less responsive to nutritional information than in supermarket settings Wisdom

et al. (2010). Because menu labeling is intended to increase nutritional transparency at the point of purchase—and ultimately to shift choices toward lower-calorie items—the weak evidence of a consumer response motivates an alternative benchmark. We therefore examine whether delegating food purchase decisions to large language models yields different nutritional outcomes, which we study in the next sections.



Panel A: $\log$(total calories) **Panel B:** $\log$(average calories per item)

Figure 2. Event-study estimates of menu labeling effects on calorie outcomes

Notes: This figure presents regression coefficient estimates from event-study specifications with calendar-month fixed effects and banner-by-group fixed effects. The unit of observation is banner-by-group-by-month. Panel A uses total calories as the outcome, where total calories are defined as the sum of (quantity $\times$ calories) within a month. Panel B uses average calories per item as the outcome, where average calories per item is defined as total calories divided by total items within a month. Treated banners are the restaurant chains (dominos, papajohns), and the control group consists of retail purchases from (kroger, walmart). For each panel, the estimation sample includes only group$\times$month cells in which both treated and control observations are present. The sample spans January 2018 to December 2018 (event time -4 to 7 months relative to the policy month). Each panel has 140 banner$\times$group$\times$month observations and 12 clusters. Observation counts by product group within each panel are: pizza ($n = 48$); sandwich ($n = 44$); wings ($n = 48$). Standard errors are clustered at the banner-by-group level.

5 Delegation as a Commitment Device: Purchase Experiment and Stylized Facts

We aim to examine the extent to which ChatGPT can function as a commitment device that helps consumers achieve their long-term health goals. Several studies show that people often rely on AI systems, following their recommendations instead of combining them with their own knowledge Buçinca et al. (2021). Building on this evidence, we study a setting in which consumers fully delegate their decisions to ChatGPT. We describe the experiment below.

5.1 Constructing Delegated Choice Sets Comparable to Actual Choice Sets

To evaluate how the restaurant calorie-labeling mandate would have affected consumer nutrition and economic welfare if individuals had delegated their purchase decisions to an AI assistant, we conduct a counterfactual experiment in which ChatGPT replaces the human decision maker at treated restaurants during the treatment period.

We begin by reorganizing product information collected online into discrete choice events. Choice sets are restricted to products that also appear in the transaction-level data, ensuring comparability between observed and simulated purchases. For each alternative, we retain the product's brand, name, category, restaurant, deflated price, and detailed nutritional information, including the policy-mandated attributes: calories, total fat, saturated fat, trans fat, cholesterol, sodium, carbohydrates, fiber, sugars, and protein. Each event represents a single consumer making one purchase among available products within a given calendar month. For each reconstructed event, ChatGPT receives the full choice set and is instructed to select exactly one option.

Finally, we use GPT-5.1 with temperature set to zero and implement four main prompt variants. Our favorite prompt is shown in Figure 3. In Appendix F we report robustness checks using alternative model versions and higher randomness.

Round 1: Initial Decision
You are a human decision maker making a single consumer choice among alternatives.
You receive one event with multiple options. Each option includes:

- identifiers: row_id, user, event_number, restaurant
- product info: brand, product, product_category
- price: price
- nutrition: calories, fat, saturated fat, trans fat, cholesterol, sodium, carbs, fiber, sugars, protein

Additional variables:

- i_want: previously purchased product category
- i_want_to_go: preferred restaurant

These are hints only; you are not forced to match them. Use ONLY provided fields.
Choose EXACTLY ONE option_id. Do not invent tie-breaking rules. Reply ONLY with the chosen option_id as an integer.

Round 2: Self-Critique and Revision
You previously chose option_id={initial_choice}. Self-critique using ONLY provided fields: mention weaknesses/tradeoffs and, if another option is better, explain why. Then choose again.
Output format (exactly):
SELF_CRITIQUE: critique text
FINAL_OPTION_ID: integer between 1 and {num_options}
No extra lines.

Round 3: Final Critique and Decision
You previously chose option_id={initial_choice} in Round 2. Critique again using ONLY provided fields, then make the FINAL choice.
Output format (exactly):
SELF_CRITIQUE: critique text
FINAL_OPTION_ID: integer between 1 and {num_options}
No extra lines.

Figure 3. Decision and self-critique prompt given to ChatGPT (example).

5.2 Delegation Protocol and Identification Strategy

We leverage the fact that we have pre and post policy data. In the pre-policy period, we retain the actual human purchase outcomes for both treated and control restaurants. In the post-policy period, however, we replace all purchases at the treated restaurants with counterfactual choices generated by ChatGPT. The control restaurant continues to follow actual human choices throughout, providing a contemporaneous benchmark unaffected by delegation. We treat each restaurant separately and run the same analysis.

After choices are generated for all events, we collapse the dataset to the product-category $\times$ restaurant $\times$ month level. The resulting panel records, for each month and restaurant-category pair, the number of unique products chosen and the aggregate nutrient content implied by the delegated choices. The goal is to estimate the causal effect of calorie labeling under a regime in which consumers delegate their decisions to an AI assistant. Because the intervention (delegation) is implemented only for the treated restaurants in the post-policy period, identification comes from the parallel trend assumption. Formally, for each restaurant $r \in \{\text{Domino's}, \text{Pizza Hut}, \text{Papa Johns}\}$, we estimate:

$$\log(Y_{rct}) = \sum_{k \neq -1} \beta_{rk} \, \mathbf{1}\{\text{time_event}_{rct} = k\} \; + \; p_{rct} \; + \; \alpha_{rc} \; + \; \lambda_t \; + \; \varepsilon_{rct},$$

where:

- Y_{rct} is the outcome for product category c at restaurant r in month t. Outcomes include (i) the number of unique products selected and (ii) total nutrient content (calories, fat, sodium, etc.).
- $\mathbf{1}\{\text{time_event}_{rct} = k\}$ are event-time indicators measuring k months relative to the implementation of the labeling mandate, with $k = -1$ omitted.
- We also control for prices to account for the fact that prices collected online are not exactly the same as actual prices.
- α_{rc} denotes product-category-by-restaurant fixed effects that absorb all time-invariant heterogeneity in menu design across restaurants and product categories.
- λ_t are calendar-month fixed effects capturing seasonal patterns and macroeconomic conditions common to all restaurants.

Each treated restaurant is compared to other rivals over time. For instance, when we treat users at Papajohns, the control groups is made of purchases at Dominos and PizzaHut. Because the delegated choices only replace treated restaurants' decisions in the post-policy period, the identifying variation arises from the differential break in treated vs. untreated decision processes induced by delegation under labeling. Next, we select some stylized facts relevant for our analysis.

5.3 Stylized Facts

In this section, we document three stylized facts with direct implications for economic welfare that support the view of ChatGPT as a commitment device. First, delegation leads to the exit of certain unsold product varieties. Second, it affects nutritional outcomes. These patterns are consistent with a commitment device that removes some alternatives and shifts short-term preferences toward long-term preferences. Third, delegation increases market concentration.

5.3.1 Delegation Triggers the Exit of Unsold Product Varieties

We begin by examining how delegating purchase decisions to ChatGPT affects realized product variety. Figure 4 reports event-study estimates of the log number of unique products purchased, relative to month $t = -1$, across prompt scenarios. We present results separately for each restaurant.

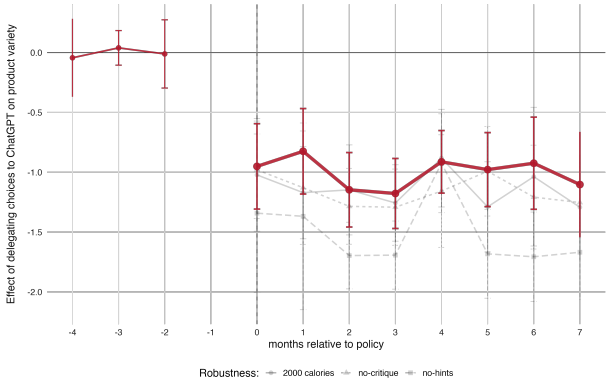
The red estimates correspond to our preferred experiment, as they provide the most conservative estimates. The prompt used in this experiment incorporates self-critique, exploiting evidence that iterative self-evaluation improves LLM reasoning (Li et al. (2025)). ChatGPT critiques and revises its choice twice before making a final decision. To make this concrete, we include an example of self-critique in the appendix E. Pre-policy coefficients ($t < 0$) are small and statistically insignificant across all chains, indicating no differential pre-trends. After implementation ($t \geq 0$), estimates become sharply negative. Delegating purchase decisions to ChatGPT results in a large and persistent reduction in product variety per product category, typically amounting to an 8–10 percent decline relative to the total variety observed in our sample. This pattern is highly consistent across chains and over time. Effect magnitudes differ by restaurant: Papa John's and Pizza Hut show the largest and most persistent declines, while Domino's exhibits smaller but still negative effects. Overall, delegation systematically narrows the set of products purchased rather than merely reallocating demand across items.

Light-grey estimates present alternative prompts as robustness checks. The *no-hints* scenario produces the largest decline; here ChatGPT observes only product characteristics, including calories, but not the intended restaurant or product category. The *no-critique* scenario adds the intended restaurant and category. The *critique* scenario adds one round of self-critique. Finally, *2000 calories* includes the policy statement that "2,000 calories a day is used for general nutrition advice, but calorie needs vary," along with intentions, but omits self-critique.

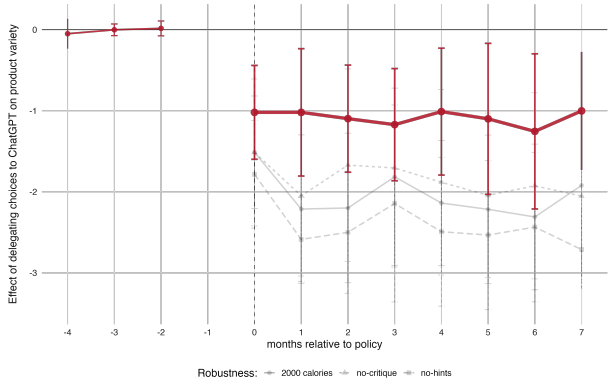
The observation that delegating choices to ChatGPT effectively removes certain products is consistent with the view that ChatGPT functions as a commitment device, as it compels consumers to disregard some alternatives. In theory, the welfare implications of this contraction in variety are ambiguous. On the one hand, reduced variety is typically associated with lower consumer surplus when consumers value differentiation. On the other hand, a narrower effective product set may weaken firms' ability to exploit within-portfolio substitution patterns.

To quantify these opposing forces, we next estimate a structural demand model that allows us to evaluate the welfare consequences of delegating food purchase decisions to ChatGPT (Section 6.1).

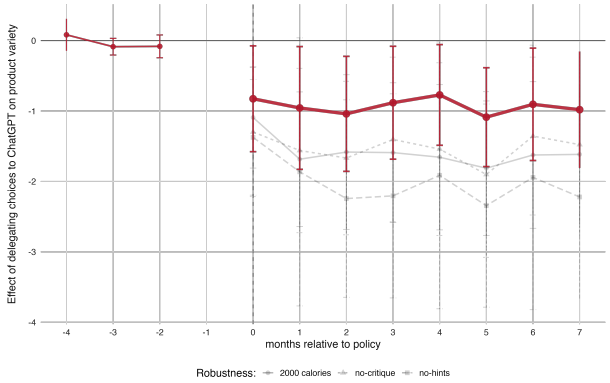
Figure 4. Event-study estimates of product variety (log unique products) after delegation



(a) Pizza Hut



(b) Papa John's



(c) Domino's

Notes: Each panel reports coefficient estimates from an event-study specification with product-category-by-restaurant fixed effects and event-month fixed effects. The unit of observation is restaurant-by-product-category-by-month, and the outcome is the logarithm of the number of unique products purchased. In each panel, the treated group corresponds to the chain named in the title when purchase decisions are delegated to ChatGPT, while the control group consists of the other major pizza chains. Event time is measured in months relative to the policy month ($t = 0$), with $t = -1$ omitted as the reference period. Points show estimated coefficients and vertical bars denote 95% confidence intervals. Pre-policy coefficients ($t < 0$) are shown only for the baseline scenario ("nohints"), while post-policy coefficients ($t \geq 0$) are overlaid for all scenarios (critique, nohints, 2000cal, and no-critique). The highlighted scenario in red corresponds to "critique". The sample contains 212 observations across 18 product-category-by-restaurant cells, spanning event time $t \in [-4, 7]$. Standard errors are clustered at the restaurant-by-product-category level.

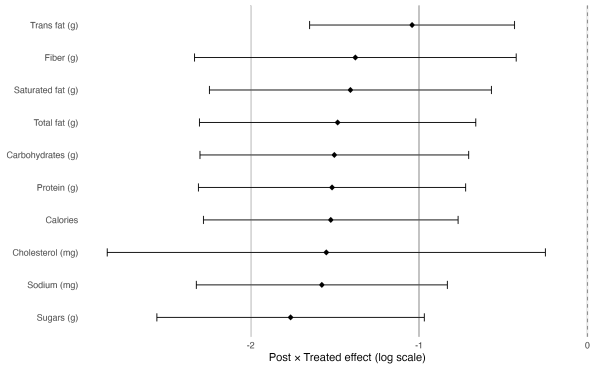
5.3.2 Delegation Impact Nutritional Outcomes

Next, we examine how food choices generated by large language models differ from those made by human consumers in terms of nutritional composition. Figure 5 summarizes post-policy effects across nutritional outcomes using a difference-in-differences framework, where outcomes are measured in logs. Exact estimates are reported in Appendix G.

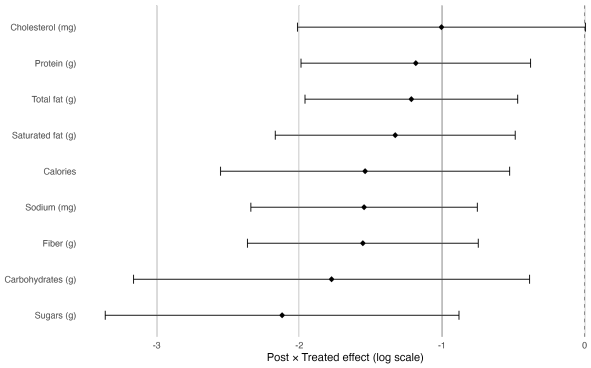
Across all chains, delegating choices to ChatGPT leads to statistically significant declines in total calories consumed, aggregated at the restaurant-by–product-category level. Estimated reductions are largest at Domino's (about 18 percent), followed by Papa John's (about 15 percent) and Pizza Hut (about 15 percent), indicating that LLM-driven choices systematically shift purchases toward lower-calorie bundles relative to human consumers.

However, the results also reveal declines in nutrients typically considered desirable to increase, most notably protein and fiber. Fiber intake falls by roughly 10 percent at Domino's, 14 percent at Pizza Hut, and 16 percent at Papa John's following delegation. Protein intake also declines across chains, by approximately 9 percent at Domino's, 15 percent at Pizza Hut, and 12 percent at Papa John's.

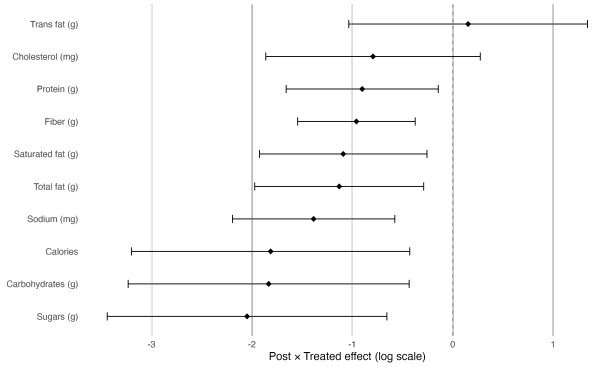
Taken together, these findings highlight an important trade-off: while delegating choices to ChatGPT reduces caloric intake, it also shifts consumption away from some nutritionally beneficial dimensions. Assessing the welfare implications of these nutritional effects, together with the impact on product variety, requires an explicit model of consumer preferences over nutritional attributes and restaurant competition, which we develop in the next section. Our model will allow the data to reveal to which extent delegation shift preferences.



1.1 Pizza Hut



1.2 Papa John's



1.3 Domino's

Figure 5. Delegation Impact on Nutrition

Notes: This figure summarizes post-policy effects across nutritional outcomes for the *critique* scenario. For each outcome, we estimate a difference-in-differences specification with product-category-by-restaurant and event-month fixed effects, regressing $\log(\text{total nutrient})$ on the interaction between a treated indicator and a post-policy dummy (post_event = 1 for months $t \geq 0$), along with additional controls. Standard errors are clustered at the restaurant-by-product-category level. The plotted coefficient corresponds to the treated $\times$ post_event interaction. The figure includes estimates for the ten nutritional outcomes mandated by the menu labeling law.

5.3.3 Delegation Increases Within-Firm Concentration

We next examine whether delegating choices to ChatGPT increases within-firm concentration. Figure 6 presents event-study estimates of within-restaurant concentration around the policy month in which purchase decisions are delegated to ChatGPT. Concentration is measured using the Herfindahl–Hirschman Index (HHI). For each restaurant r, product category c, and month t, let n_{jrct} denote the number of purchases of product j and let $N_{rct} = \sum_j n_{jrct}$ denote total purchases in that cell. Purchase shares are $s_{jrct} = n_{jrct}/N_{rct}$, and concentration is computed as

$$HHI_{rct} = \sum_j s_{jrct}^2 = \sum_j \left(\frac{n_{jrct}}{N_{rct}} \right)^2 . \tag{2}$$

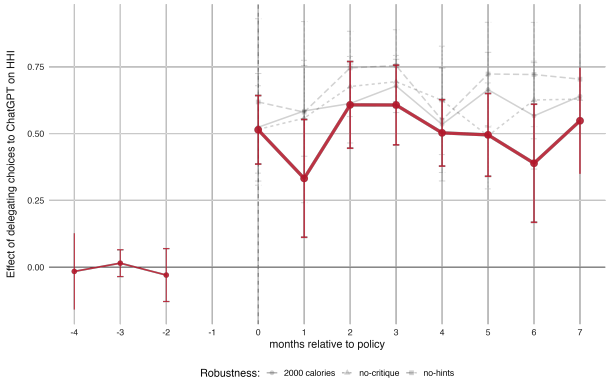
Higher values indicate more concentrated purchasing within restaurant-by-category cells, while lower values correspond to purchases being spread more evenly across products.

Across panels, pre-policy coefficients (months $t < 0$) fluctuate around zero with confidence intervals generally covering zero, providing little evidence of differential pre-trends relative to the omitted month $t = -1$. Beginning in the policy month ($t \geq 0$), estimates become positive and remain elevated throughout the post period, indicating that purchases become more concentrated within restaurant-by-product-category cells once delegation to ChatGPT begins. Although magnitudes and precision vary across chains and event horizons, the qualitative pattern is consistent: delegation shifts demand toward a narrower subset of products rather than distributing purchases evenly across the menu.

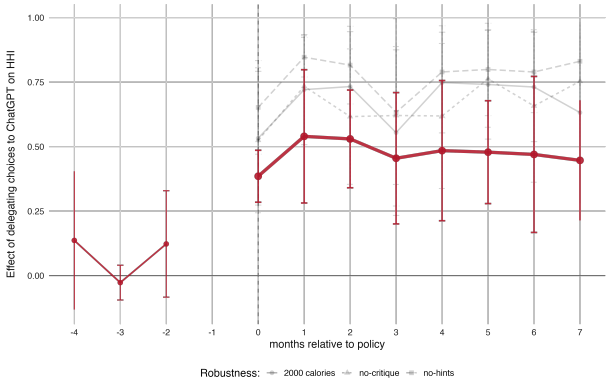
As in previous analysis, the red estimates corresponds to our preferred specification with iterative self-critique, while light-grey estimates show alternative prompt scenarios. The concentration increase is present across all prompts and follows a similar post-policy trajectory, indicating that the result does not hinge on a specific prompt design. Differences across scenarios mainly affect the magnitude and precision of the estimates rather than the direction of the effect.

Importantly, this increase in concentration does not reflect a change in purchase volume. In our experiment, ChatGPT makes the same number of purchases as human consumers, implying that the higher HHI arises from reallocating demand toward fewer products rather than from changes in total purchasing. Taken together with the previous evidence showing declining product variety, these results indicate that delegation both reduces the set of products purchased and concentrates demand within that smaller set.

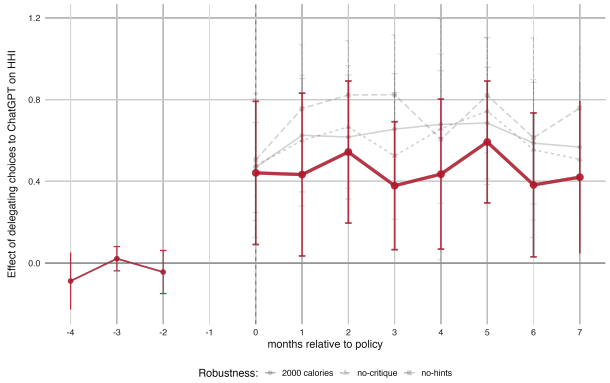
Figure 6. Event-study estimates of within-restaurant concentration (HHI) after delegation



(a) Pizza Hut



(b) Papa John's



(c) Domino's

Notes: Each panel overlays event-study estimates across multiple GPT-5 choice scenarios. The treated group corresponds to the chain named in the panel title when purchase decisions are delegated to ChatGPT, while the control group consists of the other two major pizza chains. The unit of observation is restaurant-by-product-category-by-month, and all specifications include product-category-by-restaurant fixed effects and event-month fixed effects. Event time is measured in months relative to the policy month ($t = 0$), with $t = -1$ omitted as the reference period. The y-axis reports the estimated effect on within-restaurant concentration (HHI). Post-policy coefficients ($t \geq 0$) are overlaid for all scenarios: *critique*, *nohints*, *2000cal*, and *no-critique*. The highlighted scenario corresponds to *critique*. Standard errors are clustered at the restaurant-by-product-category level.

6 Model

Motivated by the stylized facts documented in the previous section, we estimate a structural model to quantify the welfare implications of delegation.

6.1 Demand

We assume a one-level nested logit model of consumer demand in which products are partitioned into restaurant nests. Let j index products and t index markets, defined at the month level. Each product j belongs to a restaurant (nest) $r(j)$. Denote by $u_{ijt}(0)$ and $u_{ijt}(1)$ the indirect utilities obtained by consumer i for product j in market t without and with delegation, respectively.

For product j in market t, consumer i's indirect utility in the absence of delegation is

$$u_{ijt}(0) = \delta_{jt}(0) + \varsigma_i^{r(j)} + (1-\sigma)\,\varepsilon_{ijt}, \tag{3}$$

where ε_{ijt} is i.i.d. Type-I extreme value. The restaurant-level error component $\varsigma_i^{r(j)}$ is distributed so that $\varsigma_i^{r(j)} + (1-\sigma)\varepsilon_{ijt}$ is Type-I extreme value. The restriction $0 \leq \sigma < 1$ guarantees consistency with random utility maximization and σ measures within-restaurant correlation in tastes.

Mean utility is specified as

$$\delta_{jt}(0) = K + \eta_t + \alpha p_{jt} + \mu_{jt}(0) + \xi_{jt}, \tag{4}$$

where η_t is a month fixed effect, p_{jt} is price, and ξ_{jt} is an unobserved demand shock. The term μ_{jt} captures the contribution of observed product characteristics. In the absence of delegation, we write

$$\begin{aligned}\mu_{jt}(0) = {} & \beta_{\text{cal}}\,\text{Calories}_{jt} + \beta_{\text{fat}}\,\text{TotalFat}_{jt} + \beta_{\text{satfat}}\,\text{SaturatedFat}_{jt} + \beta_{\text{transfat}}\,\text{TransFat}_{jt} \\ & + \beta_{\text{chol}}\,\text{Cholesterol}_{jt} + \beta_{\text{sodium}}\,\text{Sodium}_{jt} + \beta_{\text{carb}}\,\text{Carbohydrates}_{jt} + \beta_{\text{fiber}}\,\text{Fiber}_{jt} \\ & + \beta_{\text{sugar}}\,\text{Sugars}_{jt} + \beta_{\text{protein}}\,\text{Protein}_{jt}.\end{aligned} \tag{5}$$

Consumers can also delegate their choice to ChatGPT. Delegation shifts how observed characteristics enter utility:

$$\mu_{jt}(1) = \mu_{jt}(0) + x_{jt}'\beta^D, \tag{6}$$

where x_{jt} contains the nutritional characteristics. Thus, β captures baseline tastes while β^D measures how the mapping from characteristics to utility changes under delegation. We adopt

this modeling approach because it reflects the notion that ChatGPT may serve as a commitment device, shifting short-term preferences at the time of purchase closer to long-term health goals.

However, we remain agnostic on whether this shift comes from resolving time-inconsistency, correcting nutritional misperceptions, or some other channel. For instance, if the reason people choose unhealthy food is present bias (high calories taste good now, health costs come later), then delegation to a health-conscious LLM effectively overrides the β weights on characteristics. In this interpretation, β^D for trans fat would be negative: the LLM places less weight on immediately gratifying attributes associated with trans fat, which as shown in Table 1, substantially exceed recommended nutritional guidelines in the raw data.

Under full delegation all consumers face $\mu_{jt}(1)$, yielding market shares s_{jt}^D, whereas under no delegation they face $\mu_{jt}(0)$, yielding shares s_{jt}^{ND}. We use these two environments to study the effect of delegation.

Identification Our empirical strategy treats no-delegation and delegated choices as outcomes from distinct market environments. Let $d \in \{ND, D\}$ index the regime. The nested-logit demand system implies

$$\log s_{jt}^{(d)} - \log s_{0t}^{(d)} = \alpha p_{jt} + x'_{jt}\beta + x'_{jt}\beta^D \mathbf{1}\{d = D\} + \sigma \log s_{jt|r}^{(d)} + \xi_{jt}^{(d)}. \tag{7}$$

We estimate demand by stacking observations from no-delegation markets t^{ND} (actual purchases) and delegation markets t^D (choices made by ChatGPT in the experiment). Identification of the delegation parameters β^D comes from systematic differences in choice patterns across the two environments. Consistent with our experiment, we assume that price sensitivity is unaffected by delegation. This assumption is supported by the price patterns observed in the experiment, shown in the Appendix H. We allow demand shocks to differ across regimes ($\xi_{jt}^{(ND)} \neq \xi_{jt}^{(D)}$) because the data are generated in different settings (real-world purchases versus experimental delegation).

Price p_{jt} and the within-restaurant share term $\log s_{j|r,t}$ are endogenous. We instrument for both using three types of instruments. (i) BLP-style instruments. We include the total quantity sold within the restaurant nest in a given market and the leave-one-out mean quantity of other products in the same restaurant nest. (ii) Characteristics-space competition instruments. Following Gandhi and Houde (2019), we construct a Gandhi–Houde distance-weighted carbohydrates instrument, defined as an inverse-distance–weighted average of rivals' carbohydrate content in characteristics space. This measure captures competitive pressure from nearby substitutes and shifts equilibrium prices. (iii) Cost shifters. We include tariff- and input-cost shifters that move marginal costs but are plausibly orthogonal to product-level demand shocks. We construct an indicator for the 2018 China retaliatory tariff episode affecting U.S. dairy inputs; an analogous Mexico tariff indicator interacted with a pizza dummy to reflect differential exposure of cheese-intensive products; and a monthly cheese cost index interacted with the pizza

indicator.[3]

Finally, we include interactions between each excluded instrument and the delegation indicator. Appendix I reports the corresponding first stages. The excluded instruments are strong for both endogenous variables: the first-stage F-statistics are 28.73 for price and 37.58 for $\log s_{j|r,t}$, both above conventional weak-instrument thresholds.

Results We report the estimation results in Table 2. The coefficient on price is negative and highly statistically significant, consistent with downward-sloping demand. The estimated nesting parameter is positive, statistically significant, and lies within the unit interval, indicating meaningful within-restaurant correlation in unobserved utility.

Turning to preferences over nutritional characteristics, the baseline (human) estimates indicate that consumers place positive value on calories and cholesterol, while exhibiting negative preferences for total fat, trans fat, fiber, and protein. Sodium, carbohydrates, sugars, and saturated fat are imprecisely estimated in the baseline specification.

The interaction terms reveal notable differences under delegation. Relative to human choices, delegated choices place significantly more negative weight on trans fat and cholesterol, and more positive weight on protein and saturated fat. Most other interaction effects are imprecisely estimated.

Overall, the baseline preference patterns remain broadly misaligned with prevailing dietary guidelines. Nutrients recommended for moderation; such as calories; are valued positively, while beneficial nutrients like fiber and protein receive negative baseline weights U.S. Department of Health and Human Services and U.S. Department of Agriculture (2026). These preference estimates are consistent with the reduced-form evidence in Section 3, which is external to the structural model and therefore provides supporting validation.

[3] During the 2018 trade dispute, China and Mexico imposed retaliatory tariffs on U.S. dairy products, contributing to a record U.S. cheese surplus. See https://www.axios.com/2018/12/17/cheese-dairy-stockpile-farmer-america-trade.

Table 2. Demand Estimates

	No Delegation		Delegation	
	Estimate	Std. Error	Estimate	Std. Error
Price	-0.363^{***}	0.035		
$\log s_{j\|r,t}$	0.614^{***}	0.041		
Calories	0.220^{***}	0.035	-0.103	0.244
Total fat	-1.269^{*}	0.548	0.138	2.123
Saturated fat	-1.270	0.781	3.940^{*}	1.793
Trans fat	-21.196^{**}	7.574	-88.428^{***}	21.715
Cholesterol	0.576^{***}	0.101	-0.928^{***}	0.229
Sodium	0.002	0.007	0.044^{*}	0.017
Carbohydrates	0.041	0.052	-0.723	0.981
Fiber	-9.352^{***}	1.938	4.552	3.819
Sugars	-0.062	0.404	1.455	1.006
Protein	-1.536^{***}	0.410	4.930^{**}	1.650
Mean own-price elasticity (human only)		-3.622		
Market (month) fixed effects		Yes		
First-stage F-statistic (price)		28.73		
First-stage F-statistic ($\log s_{j\|r,t}$)		37.58		
Observations		1103		

Notes: Table reports 2SLS estimates from the demand specification with month fixed effects. Standard errors shown. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

The model delivers an average own-price elasticity of demand equal to 3.622 in the human-only sample. Table 3 reports mean product-level own-price elasticities for the top products within each major menu category. Across products, own-price elasticities range from a minimum of 5.341 for Papa John's Cheese Pizza (most price-sensitive product) to a maximum of 2.177 for Domino's Specialty Chicken – Sweet BBQ Bacon (least price-sensitive product), indicating substantial heterogeneity in demand responsiveness across menu categories and restaurants. This also indicates that evaluating menu labeling laws using a single restaurant chain can either overestimate or underestimate consumer responses to the policy—an issue that our approach mitigates.

Table 3. Mean own-price elasticities for top products within each major category

Category	Product	Mean own-price elasticity
Pasta	Domino's Chicken Alfredo Pasta_dominos	-2.229
Pasta	Domino's Build Your Own Pasta_dominos	-2.349
Pasta	Domino's Chicken Carbonara Pasta_dominos	-2.323
Pasta	Domino's Italian Sausage Marinara Pasta_dominos	-2.336
Pasta	Domino's Pasta Primavera_dominos	-2.342
Pizza	Other Domino's Pizza_dominos	-2.567
Pizza	Pizza Hut Large Pizza_pizzahut	-4.109
Pizza	Papa Johns Cheese Pizza_papajohns	-5.341
Pizza	Domino's Philly Cheese Steak Pizza_dominos	-2.847
Pizza	Pizza Hut Medium Pizza_pizzahut	-3.756
Salad	Domino's Chicken Caesar Salad_dominos	-2.239
Salad	Domino's Classic Garden Salad_dominos	-2.254
Sandwich	Domino's Chicken Bacon Ranch Sandwich_dominos	-2.237
Sandwich	Domino's Buffalo Chicken Sandwich_dominos	-2.272
Sandwich	Domino's Chicken Parm Sandwich_dominos	-2.288
Sandwich	Domino's Chicken Habanero Sandwich_dominos	-2.266
Sandwich	Domino's Mediterranean Veggie Sandwich_dominos	-2.265
Wings	Papa Johns Buffalo Wings_papajohns	-2.781
Wings	Domino's Hot Buffalo Wings_dominos	-2.821
Wings	Domino's Specialty Chicken - Sweet BBQ Bacon_dominos	-2.177
Wings	Domino's Specialty Chicken - Classic Hot Buffalo_dominos	-2.193
Wings	Papa Johns BBQ Wings_papajohns	-2.820

Notes: Entries report the mean of product-level own-price elasticities implied by the model, averaged over the sample months for each product. "Top products" are selected within each major category based on total quantity. Sample restricted to actual purchases ($\delta = 0$).

Beyond price responsiveness, we next examine how demand responds to nutritional attributes. Table 4 reports the mean elasticities of demand with respect to each attribute implied by the model. Demand is highly elastic with respect to calories (2.952) and is also positively responsive to cholesterol (1.055). By contrast, total fat, saturated fat, trans fat, fiber, and protein carry negative elasticities on average. Responses to sodium, carbohydrates, and sugars are relatively small in magnitude.

Table 4. Mean Elasticities with Respect to Nutritional Attributes

Nutritional attribute	Mean elasticity
Calories	2.952
Total fat	-0.807
Saturated fat	-0.315
Trans fat	-0.057
Cholesterol	1.055
Sodium	0.074
Carbohydrates	0.056
Fiber	-0.428
Sugars	-0.010
Protein	-0.937

Notes: Table reports mean elasticities. Sample restricted to actual purchases.

Alternative Approach To complement this approach, we directly estimate ChatGPT preferences using a one-level nested logit and estimate the exact same specification for humans to enable comparison. This exercise extends Chen et al. (2023), who study ChatGPT food preferences in a stylized setting where choices are limited to tomatoes versus meat. The human column uses the same instruments as in Table 2; the only difference is that we restrict the sample to realized purchases.

The results reveal several notable patterns. First, ChatGPT is substantially less price-sensitive than humans. Its implied own-price elasticity of demand is -0.930, compared with -4.412 for humans. This finding supports the assumption in the mean-utility specification in equation (4) that delegation primarily shifts the valuation of nutritional attributes while leaving price sensitivity largely unchanged. Second, the nesting parameter is considerably larger for ChatGPT than for humans, indicating stronger within-nest substitution under delegated choice.

In the counterfactual analysis, we will study a scenario in which delegation shifts preferences according to the estimates in Table 2. We will also consider a scenario where, under delegation, consumers adopt exactly the ChatGPT demand parameters reported in Table 5. To quantify the welfare effects of delegation, we next specify the supply side of the model.

Table 5. Demand Estimates: ChatGPT vs. Human

	ChatGPT		**Human**	
Variable	**Estimate**	**Std. Error**	**Estimate**	**Std. Error**
Price	-0.102	0.015	-0.441	0.061
$\log s_{j\|r,t}$	0.851	0.030	0.368	0.148
Calories	0.101	0.013	0.205	0.037
Total fat	-1.390	0.556	-0.678	0.683
Saturated fat	-0.039	0.827	-1.888	1.141
Trans fat	-66.985	15.640	12.624	8.423
Cholesterol	0.057	0.144	0.054	0.105
Sodium	0.035	0.010	0.005	0.007
Carbohydrates	-0.379	0.190	-0.016	0.012
Fiber	-7.287	1.924	-13.242	2.301
Sugars	-0.541	0.380	0.560	0.444
Protein	0.916	0.552	-0.394	0.519
Own-price elasticity	-0.930		-4.412	
Instruments			✓	

Notes: Table reports demand estimates from nested-logit specifications. The ChatGPT column is estimated using delegated choices, while the Human column uses observed purchases and is estimated by 2SLS. Month fixed effects are included in all specifications. The reported own-price elasticity corresponds to the sample mean implied by the model.

6.2 Supply Model

We assume that F restaurants compete à la Bertrand over T time periods. Each restaurant f owns a set of products $j \in \Theta_{ft}$ and chooses prices simultaneously. Demand is a mixture of non delegated (ND) and delegated (D) choices. The effective quantity demanded for product j is a convex combination of two demand systems,

$$\tilde{q}_{jt}(p) = (1-\delta)q_{jt}^{ND}(p) + \delta q_{jt}^{D}(p), \qquad \delta \in [0,1], \tag{8}$$

where $q_{jt}^{ND}(p)$ denotes human demand and $q_{jt}^{D}(p)$ demand under delegation. These demands are obtained by multiplying market size by the corresponding market share functions.

Restaurant f's profit maximization problem at time t is therefore

$$\max_{\{p_{jt}: j \in \Theta_{ft}\}} \Pi_t^f(p) = \sum_{j \in \Theta_{ft}} (p_{jt} - c_{jt})\tilde{q}_{jt}(p), \tag{9}$$

where c_{jt} denotes marginal cost.

For each product $j \in \Theta_{ft}$, the first-order condition is

$$\tilde{q}_{jt}(p) + \sum_{k \in \Theta_{ft}} (p_{kt} - c_{kt}) \frac{\partial \tilde{q}_{kt}(p)}{\partial p_{jt}} = 0. \tag{10}$$

Because effective demand is a weighted average of human and delegated demand, its derivative satisfies

$$\frac{\partial \tilde{q}_{kt}(p)}{\partial p_{jt}} = (1-\delta)\frac{\partial q_{kt}^{ND}(p)}{\partial p_{jt}} + \delta \frac{\partial q_{kt}^{D}(p)}{\partial p_{jt}}. \tag{11}$$

Hence the first-order condition becomes

$$0 = (1-\delta)q_{jt}^{ND}(p) + (\delta)q_{jt}^{D}(p) \tag{12}$$

$$+ \sum_{k \in \Theta_{ft}} (p_{kt} - c_{kt}) \left[(1-\delta)\frac{\partial q_{kt}^{ND}(p)}{\partial p_{jt}} + \delta \frac{\partial q_{kt}^{D}(p)}{\partial p_{jt}} \right]. \tag{13}$$

This first-order condition coincides with the standard Bertrand condition when $\delta = 0$ or when $q_{kt}^{D}(p) = q_{kt}^{ND}(p)$. The latter case arises when the LLM perfectly replicates human demand. We focus on the case in which restaurants charge uniform prices across consumers, regardless of whether they delegate their choices to an LLM. There are three main reasons for this assumption. First, LLMs were not publicly available during our sample period (2018); for example, ChatGPT was only released on November 30, 2022. Second, even beyond our setting, it is unclear whether restaurants can observe or condition prices on consumers' use of delegation technologies. Third, this assumption is consistent with the retail pricing literature, which documents the prevalence of uniform pricing (DellaVigna and Gentzkow (2019); Hitsch et al. (2021)).

Stacking products of firm f into vectors, define p_{ft}, c_{ft}, and $\tilde{q}_{ft}$ as the vectors of prices, marginal costs, and effective quantities. Let

$$\tilde{\Delta}_{ft}(p) = \left[\frac{\partial \tilde{q}_{kt}(p)}{\partial p_{jt}} \right]_{j,k \in \Theta_{ft}} \tag{14}$$

be the demand Jacobian. The system of first-order conditions can then be written as

$$\tilde{q}_{ft}(p) + \tilde{\Delta}_{ft}(p)\,(p_{ft} - c_{ft}) = 0. \tag{15}$$

From this equation one obtain markups:

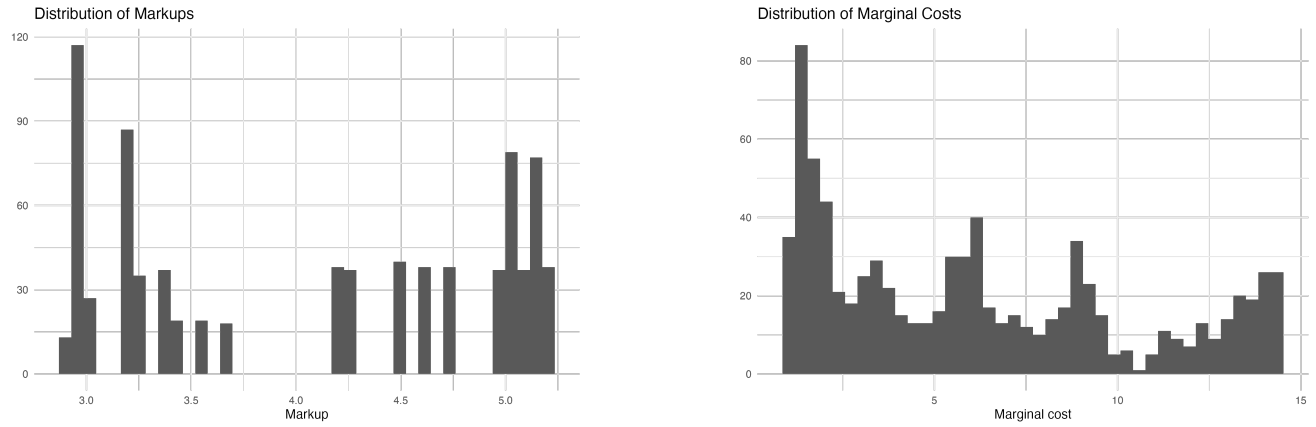
$$p_{ft} - c_{ft} = -\tilde{\Delta}_{ft}(p)^{-1}\tilde{q}_{ft}(p). \tag{16}$$

Markups and Costs Combining the first-order conditions in (10) with the estimated demand parameters allows us to recover product-level markups and marginal costs. Table 6 reports category-level averages. Markups are sizable across all major categories. Salads exhibit the highest average markups, with a mean markup of 4.84 relative to an average price of 6.27, while pizzas display the lowest markups, averaging 3.88 despite substantially higher prices. Intermediate categories such as sandwiches and pasta also feature substantial margins, with mean markups of 4.69 and 4.30, respectively. In the next section, we use the estimated supply model to study how delegating choice to ChatGPT affects prices, markups, and consumer welfare.

Table 6. Average Prices, Markups, and Marginal Costs by Major Category

Major category	Mean price	Mean markup	Mean marginal cost
Pizza	13.22	3.88	9.34
Wings	8.40	4.09	4.31
Pasta	7.23	4.30	2.93
Sandwich	6.44	4.69	1.75
Salad	6.27	4.84	1.43

Notes: Prices, markups, and marginal costs are computed using the first-order conditions together with the estimated demand parameters, and then averaged across products within each major category. Sample restricted to HUMAN observations with $\delta = 0$.



(a) Markups

(b) Marginal costs

Figure 7. Distribution of Markups and Marginal Costs

Notes: Histograms of markups and marginal costs (human observations only). Markups and marginal costs are implied by the estimated one-level nested logit demand model combined with a multi-product Nash–Bertrand pricing assumption without delegation. For each market (month–year), we recover product-level markups by solving the system of Bertrand first-order conditions using the estimated demand derivatives.

7 Counterfactual and Impact on Welfare

To assess the welfare implications of delegation, we compute the counterfactual prices, markups, profits, and consumer surplus that would arise if a fraction of consumers delegated their decisions to ChatGPT. Our experiment shows that delegation reduces the effective variety of products purchased (Figure 4) and induces a reallocation of demand across menu items (Figure 6). Accordingly, the counterfactual environment explicitly incorporates delegation-induced reductions in product availability.

To provide conservative estimates, we assume that a product becomes unavailable to delegators only if it is not selected by ChatGPT for at least six consecutive months. This six-month rule is standard in the product entry–exit literature (e.g., Fan and Yang (2025)). Under this criterion, approximately 23% of products in our sample are dropped.

We consider two mechanisms through which delegation may affect demand. First, in the *preference-shock* scenario, delegation alters the valuation of product attributes but leaves price sensitivity unchanged. This assumption is motivated by experimental evidence (Appendix H) showing that products selected by ChatGPT are not systematically cheaper than those chosen by humans. Second, to assess robustness, we study a *ChatGPT-preferences* scenario in which delegated consumers adopt the full ChatGPT demand system, allowing delegation to also affect price sensitivity.

In all cases, the benchmark is the no-delegation equilibrium under human preferences. We recover marginal costs from this baseline and hold them fixed across counterfactuals. Table 7 reports the resulting price and markup effects under the two delegation mechanisms and for

varying delegation shares.

To clarify the sources of variation under partial delegation, we first consider intermediate counterfactuals. The *Availability only* (*Drop*) scenario removes products according to the ChatGPT selection rule while keeping human preferences. Across both model variants, reducing variety alone lowers equilibrium prices and markups. This does not necessarily imply higher welfare, however, because the price reduction may be offset by the loss of variety.

Next, the *Delegated tastes only* (or *ChatGPT preferences only*) scenario imposes full delegation without removing products. In this case, prices and markups increase. The combined *Joint: Drop + Delegation* scenario produces intermediate effects: prices and markups still rise, but less than in the no-drop full-delegation environment.

Finally, we study the *partial delegation equilibrium* with $\delta \in \{0.10, 0.50, 0.90\}$. Across all specifications, greater delegation leads to higher prices and markups. These increases are substantially larger when consumers who delegate their choices to ChatGPT are assumed to follow the full ChatGPT demand system than when delegation affects demand only through attribute-preference shocks.

We next turn to the welfare implications of these changes and examine their distributional consequences for delegating and non-delegating consumers.

Consumer surplus. The implications of delegation for consumer welfare are theoretically ambiguous. On the one hand, when delegation effectively removes some products, prices may decrease. On the other hand, delegation also reallocates demand across the remaining products; by shifting market shares toward certain items, this reallocation can increase prices. Finally, in partial-delegation equilibria all consumers face the same equilibrium price vector, so any welfare gains for delegators may come alongside welfare losses for non-delegators, making the distributional consequences of delegation an important object of interest.

To quantify these forces, we compute consumer surplus and firm profits in the baseline and in a series of counterfactual equilibria. Let $k \in \{ND, D\}$ denote the preference regime, where ND (non-delegated) corresponds to human preferences and D (delegated) corresponds to the demand system faced under delegation. The delegated regime is scenario-specific: in the *preference-shock* specification it reflects the shifted utility mapping, while in the *ChatGPT demand* specification it reflects the separately estimated ChatGPT demand system.

Let $\sigma_k \in [0, 1)$ denote the within-restaurant nesting parameter under regime k. The restaurant-level inclusive value is

$$I_{rt}^k = (1 - \sigma_k) \ln \left(\sum_{j \in r} \exp \left(\frac{\delta_{jt}^k}{1 - \sigma_k} \right) \right), \tag{17}$$

and the market inclusive value (with the outside option normalized to zero) is

Table 7. Price/markup effects under alternative delegation models

Scenario	$\Delta Price$ (%)	$\Delta Markup$ (%)
Panel A: Preference shock		
Availability only (Drop)	-6.14	-0.84
Delegated tastes only	1.20	2.52
Joint: Drop + delegated tastes	-2.99	5.81
Partial delegation equilibrium ($\delta \in \{0.10, 0.50, 0.90\}$)		
$\delta = 0.10$	0.44	0.92
$\delta = 0.50$	1.38	2.90
$\delta = 0.90$	2.21	4.65
Panel B: ChatGPT preferences		
Availability only (Drop)	-10.71	-2.89
ChatGPT preferences only	98.62	264.35
Joint: Drop + ChatGPT preferences	88.37	262.70
Partial delegation equilibrium ($\delta \in \{0.10, 0.50, 0.90\}$)		
$\delta = 0.10$	2.41	6.47
$\delta = 0.50$	56.55	151.59
$\delta = 0.90$	58.32	156.33

Notes: Panel A reports results under the *preference-shock* model of delegation. Panel B reports results under the *ChatGPT-preferences* model (delegators adopt the estimated ChatGPT demand system). "Availability only" isolates the effect of product removal holding preferences fixed. "Delegated tastes only" changes preferences while keeping the full product set. The joint scenario combines both channels. All entries are percent changes relative to the no-delegation baseline.

$$I_t^k = \ln\left(1 + \sum_r \exp(I_{rt}^k)\right). \tag{18}$$

Per-consumer surplus under regime k is

$$CS_t^k = \frac{I_t^k}{-\alpha_k}. \tag{19}$$

Baseline (no delegation). In the baseline equilibrium all consumers choose directly using non-delegated (human) preferences. Let M_t denote market size and δ the counterfactual delegation share. For consistency with the partial-delegation experiments, we decompose baseline surplus into the population shares that will later delegate and not delegate:

$$CS_t^{ND,0} = (1-\delta)\frac{M_t}{-\alpha_{ND}} I_t^{ND}(p^{obs}), \tag{20}$$

$$CS_t^{D,0} = \delta\frac{M_t}{-\alpha_{ND}} I_t^{ND}(p^{obs}). \tag{21}$$

Counterfactual equilibria. In counterfactual scenarios, non-delegators continue to follow the non-delegated demand system while delegators follow the delegated demand system. Let p^{cf} denote the equilibrium price vector. Some products may be unavailable to delegators; let $\mathcal{A}_{jt} \in \{0,1\}$ denote this availability indicator.

Group-specific surplus is

$$CS_t^{ND} = (1-\delta)\frac{M_t}{-\alpha_{ND}} I_t^{ND}(p^{cf}), \tag{22}$$

$$CS_t^{D} = \delta\frac{M_t}{-\alpha_{D}} I_t^{D}(p^{cf}, \mathcal{A}), \tag{23}$$

and total consumer surplus is

$$CS_t = CS_t^{ND} + CS_t^{D}. \tag{24}$$

Reported percentage changes are computed separately for non-delegators and delegators relative to their respective baselines and then aggregated using the population share:

$$\Delta CS = (1-\delta)\Delta CS^{ND} + \delta \Delta CS^{D}.$$

Table 8 reports percentage changes relative to the no-delegation baseline under the two delegation models. The *Availability only* (*Drop*) scenario reduces welfare in both specifications. In the preference-shock model (Panel A), this counterfactual lowers consumer surplus by 5.2%

Table 8. Welfare effects under alternative delegation models

Scenario	ΔCS (%)	ΔCS^{ND} (%)	ΔCS^{D} (%)	$\Delta\Pi$ (%)
Panel A: Preference shock				
Availability only (Drop)	-5.24			-3.69
Delegated tastes only	218.59			43.92
Joint: Drop + delegated tastes	196.70			48.37
Partial delegation equilibrium ($\delta \in \{0.10, 0.50, 0.90\}$)				
$\delta = 0.10$	19.43	-0.75	201.08	4.90
$\delta = 0.50$	97.84	-2.16	197.84	24.42
$\delta = 0.90$	176.81	-3.03	196.79	43.68
Panel B: ChatGPT preferences				
Availability only (Drop)	-15.30			-11.59
ChatGPT preferences only	131.23			140.84
Joint: Drop + ChatGPT preferences	119.84			130.42
Partial delegation equilibrium ($\delta \in \{0.10, 0.50, 0.90\}$)				
$\delta = 0.10$	25.36	-5.70	304.91	2.96
$\delta = 0.50$	48.65	-27.59	124.88	53.28
$\delta = 0.90$	105.52	-27.51	120.30	115.01

Notes: Panel A reports results under the *preference-shock* model of delegation. Panel B reports results under the *ChatGPT-preferences* model (delegators adopt the estimated ChatGPT demand system). "Availability only" isolates the effect of product removal holding preferences fixed. "Delegated tastes only" changes preferences while keeping the full product set. The joint scenario combines both channels. All entries are percent changes relative to the *no-delegation baseline equilibrium*. For partial delegation, $\Delta CS = (1-\delta)\Delta CS^{ND} + \delta\Delta CS^{D}$.

and profits by 3.7%. Under the ChatGPT-preferences model (Panel B), the corresponding declines are larger, at 15.3% and 11.6%, respectively. The reduction in product variety offsets the benefits generated by lower prices.

Introducing delegation substantially changes the picture. Under full delegation without product removal (the *Delegated tastes only* / *ChatGPT preferences only* scenarios), consumer surplus rises sharply in both models: by 218.6% in the preference-shock specification and by 131.2% under the ChatGPT demand system. Profit effects are also strongly positive. Partial delegation generates pronounced distributional effects. Across both panels, non-delegators experience welfare losses while delegators obtain large gains. For example, in Panel B under the *partial delegation equilibrium* with $\delta = 0.50$, aggregate consumer surplus increases by 48.7%, but this reflects a 27.6% decline for non-delegators and a 124.9% increase for delegators. As the delegation share rises, aggregate surplus and profits increase.

8 Conclusion

This article examines the equilibrium and welfare effects of delegated decision-making by large language models (LLMs) in a policy-relevant setting, combining observational and experimental evidence with a structural model. We quantify the extent to which LLM-based delegation could enhance the effectiveness of menu-labeling regulations. Consistent with the limited behavioral response we document following the nationwide labeling policy, we find that menu labeling had only modest effects on calorie intake. By contrast, delegating purchase decisions to ChatGPT would reduce total calorie intake.

Beyond these direct nutritional effects, delegation generates important equilibrium responses. Our most conservative estimates indicate substantial exit of unsold product varieties and increased market concentration. We incorporate these forces into a structural model of delegation that captures the gap between short-term preferences and long-term health goals. The results support the interpretation of ChatGPT as a commitment device that helps consumers align choices with long-run objectives.

In the aggregate, the net effect of delegation on total welfare is positive, but the gains are unevenly distributed. Consumers who delegate experience sizable increases in consumer surplus, while non-delegating consumers face small welfare losses due to equilibrium price effects. Firm profits increase. These findings highlight that AI-mediated delegation can improve health-related decision-making while simultaneously reshaping market structure and the distribution of welfare.

Future work will assess the robustness of our results when delegation relies on unstructured inputs such as product images.

Appendix

A Descriptive Statistics on Online Nutrition and Price Data

Table 9. Summary Statistics for Nutrition and Price Variables

Variable	Mean	SD
Calories (per item)	1797.05	925.08
Total fat (g)	88.74	44.45
Saturated fat (g)	37.42	20.48
Trans fat (g)	1.14	1.25
Cholesterol (mg)	231.58	102.83
Sodium (mg)	4380.89	2067.56
Total carbohydrates (g)	165.36	99.61
Fiber (g)	4.77	3.48
Sugars (g)	14.91	9.87
Protein (g)	76.72	37.27
Price (USD, 2018)	12.80	2.96

Notes: Means and standard deviations are computed at the item level. Prices are deflated using Consumer Price Index for food, normalized to 2018.

B Descriptive Statistics - Actual Purchases

Table 10. Average Price, Price Dispersion, and Revenue Share Within Restaurant

Restaurant	Category	Avg. Price	SD Price	Revenue Share
Domino's	Pizza	8.04	3.31	0.734
Domino's	Wings	7.21	2.55	0.164
Domino's	Pasta	6.57	0.77	0.063
Domino's	Sandwich	6.33	0.65	0.028
Domino's	Salad	6.27	0.47	0.012
Papa John's	Pizza	16.00	3.09	0.879
Papa John's	Wings	8.88	2.79	0.119
Papa John's	Sandwich	6.99	0.00	0.001
Pizza Hut	Pizza	13.30	4.15	0.871
Pizza Hut	Wings	9.86	4.89	0.103
Pizza Hut	Pasta	8.96	2.60	0.026

Notes: Avg. Price and SD Price are computed from item unit price. Revenue Share is the category's total revenue divided by total revenue within the same restaurant.

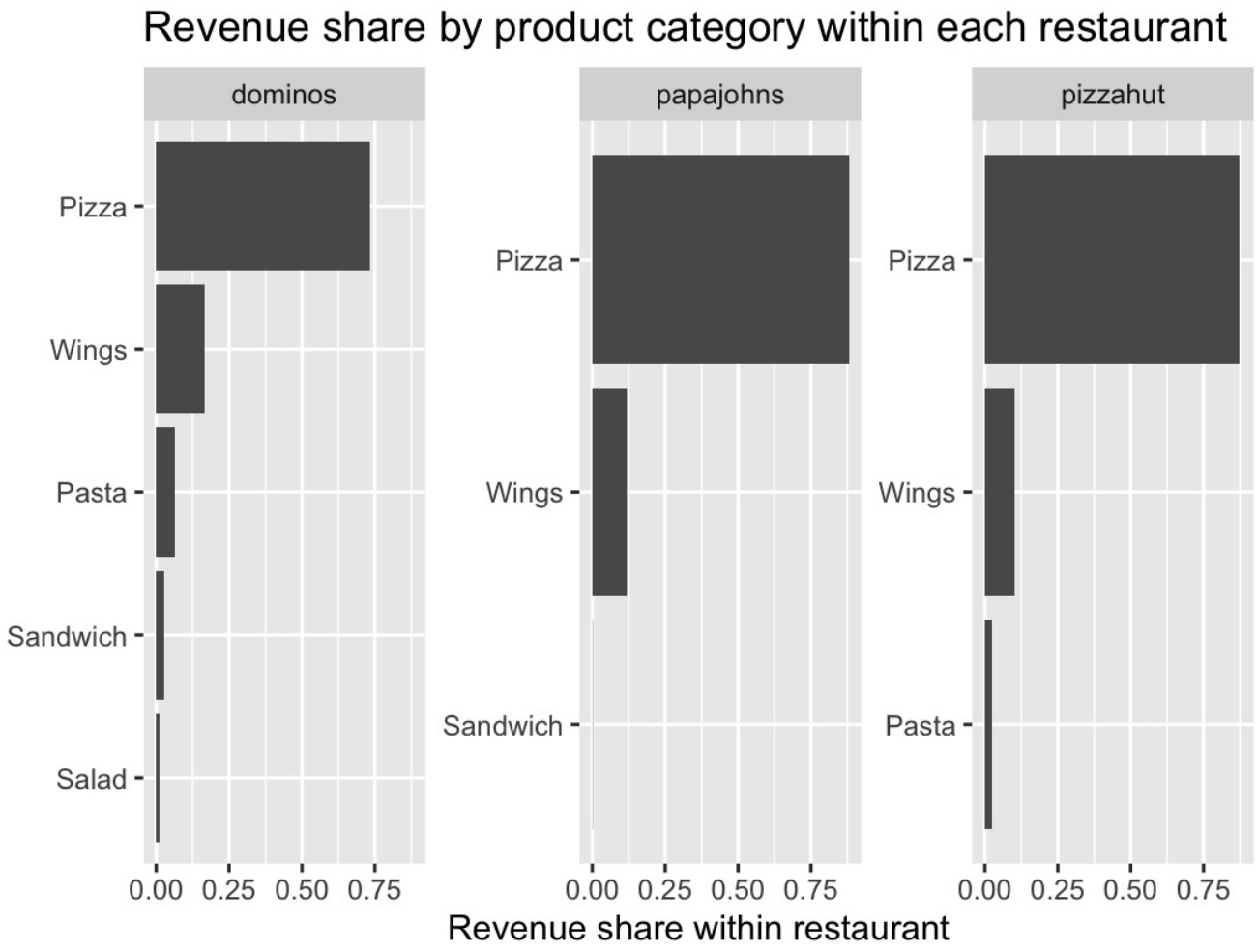


Figure 8. Revenue share by product category within each restaurant. Bars show the share of total revenue accounted for by each major product category, computed separately within Domino's, Papa John's, and Pizza Hut.

C Impact of Menu Labelling Law on Calories per Dollar

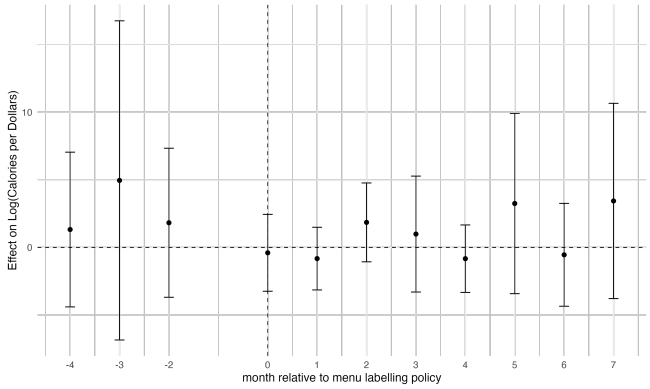


Figure 9. Event-study estimates of menu labeling effects on log of calories per dollar

Notes: This figure presents regression coefficient estimates from event-study specifications with calendar-month fixed effects and banner-by-group fixed effects. The unit of observation is banner-by-group-by-month. The outcome is calories per dollar, defined as total calories divided by total spend within a month. Treated banners are the restaurant chains (dominos, papajohns), and the control group consists of retail purchases from (kroger, walmart). For each estimation, the sample includes only group$\times$month cells in which both treated and control observations are present. The sample spans January 2018 to December 2018 (event time -4 to 7 months relative to the policy month). The estimation sample contains 140 banner$\times$group$\times$month observations and 12 clusters. Observation counts by product group are: pizza ($n = 48$); sandwich ($n = 44$); wings ($n = 48$). Standard errors are clustered at the banner-by-group level.

D Prompt Used

Prompts with 2,000-Calorie Reference and Removing Hints

We report the exact prompts used in the experiments labeled *2000 calories* and *no-hints*, for which estimates are presented in Section 5.3.

First, we augment the prompt by adding the statement: "Take into account that 2,000 calories a day is used for general nutrition advice, but calorie needs vary," while keeping the contextual hint variables (`i_want` and `i_want_to_go`).

Second, we re-run the experiment using a simplified prompt that removes all contextual hints (i.e., `i_want` and `i_want_to_go`) and does not mention the 2,000-calorie reference. Figures 10 and 11 report the exact prompts used.

You are a human decision maker making a single consumer choice among alternatives.
You will receive exactly one event with multiple options. Each option includes:

- identifiers: row_id, user, event_number, restaurant
- product info: brand, product, product_category
- price: price
- nutrition: calories, fat, saturated fat, trans fat, cholesterol, sodium, carbs, fiber, sugars, protein

Take into account that 2,000 calories a day is used for general nutrition advice, but calorie needs vary.
You are also given additional variables describing what you want in this event:

- i_want: the type of product (product_category) that was previously purchased
- i_want_to_go: the restaurant you would like to go to in this event

You may use i_want and i_want_to_go as context, but you are NOT forced to pick an option that matches them. They are just hints.
Use ONLY the provided fields in the event.
Choose EXACTLY ONE option_id in every case — even if several options seem equally good. Do not invent any tie-breaking rule; simply select one option you prefer.
Reply with ONLY the chosen option_id as an integer, with no extra text.

Figure 10. Robustness prompt 1: contextual hints with a reference to the 2,000-calorie guideline.

You are a human decision maker making a single consumer choice among alternatives.
You will receive exactly one event with multiple options. Each option includes:

- identifiers: row_id, user, event_number, restaurant
- product info: brand, product, product_category
- price: price
- nutrition: calories, fat, saturated fat, trans fat, cholesterol, sodium, carbs, fiber, sugars, protein

Use ONLY the provided fields in the event.
Choose EXACTLY ONE option_id in every case — even if several options seem equally good. Do not invent any tie-breaking rule; simply select one option you prefer.
Reply with ONLY the chosen option_id as an integer, with no extra text.

Figure 11. Robustness prompt 2: no contextual hints and no reference to the 2,000-calorie guideline.

E Example of ChatGPT Self-Critique

To illustrate how ChatGPT revises its own choice through self-critique, we provide below an example taken from the experiment. The text corresponds to two successive critique rounds generated by ChatGPT after an initial product choice.

Self-Critique Round 1

Option 6 (Pepperoni Pizza) matched the preference for pizza from Papa John's and was in the mid-range of prices. However, it is relatively high in calories (1840), fat (80g), and sodium (4560mg). Among pizzas, Option 35 (Sausage Pizza) is substantially cheaper ($13.99 versus $16.99) and contains fewer calories (1320), less fat (60g), and lower sodium (3240mg), while still satisfying the preference for pizza. It therefore represents a better overall tradeoff in terms of both price and nutritional content.

Self-Critique Round 2

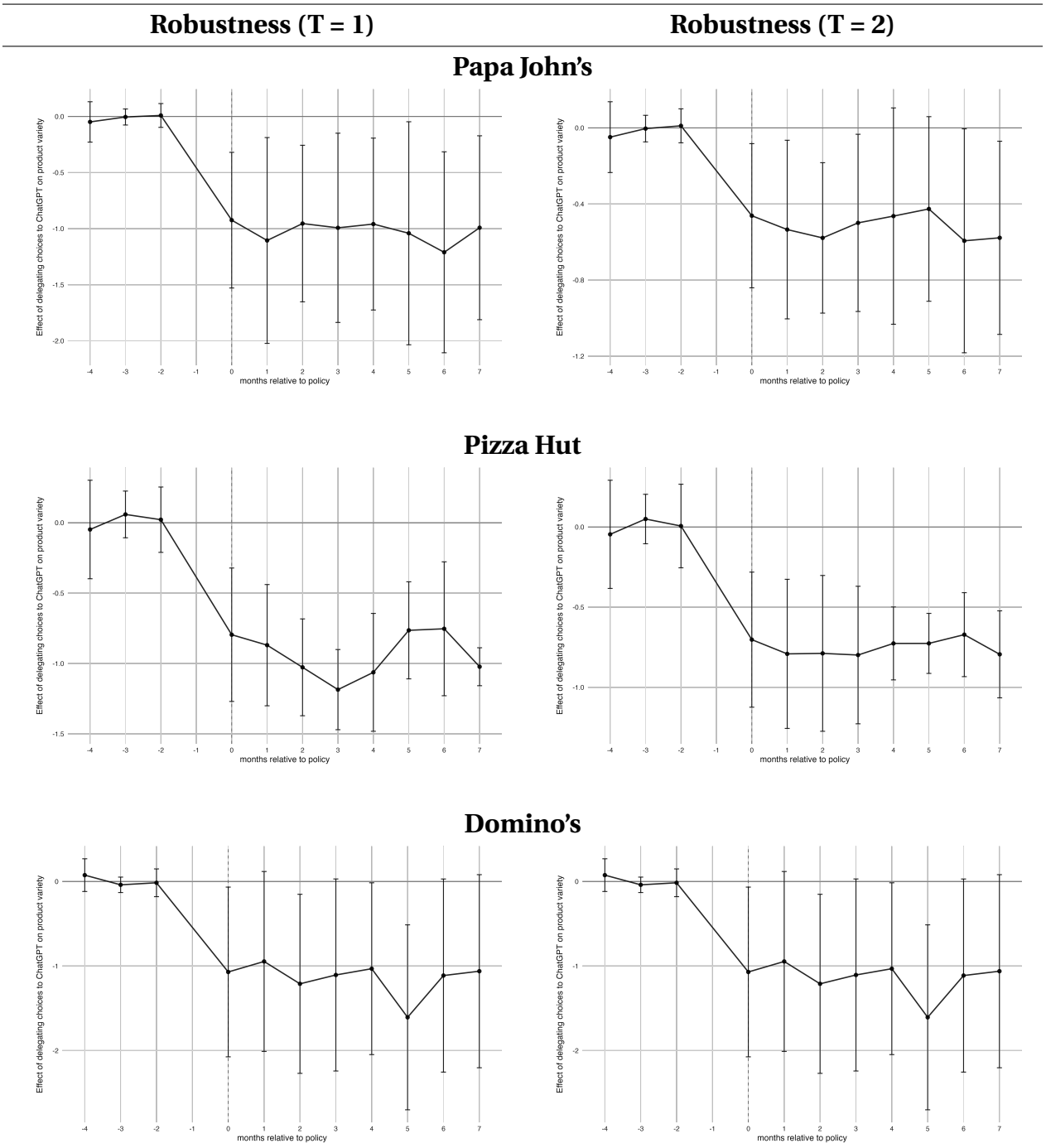
Option 35 (Sausage Pizza) satisfies the preference for pizza and is relatively inexpensive among available pizzas, but it still has high calories (1320) and sodium (3240mg) and provides no fiber. For a similar or slightly higher price, other pizzas offer more food variety and comparable or greater protein content. For example, Option 11 (The Meats Pizza) provides comparable protein content together with greater topping variety and some fiber, while remaining reasonably priced ($17.99) compared with very high-calorie alternatives such as Options 4 or 12. Among pizza options, Option 11 therefore offers a better balance between price, calorie content, and variety than Option 35.

F Robustness Checks

Alternative Temperature Settings

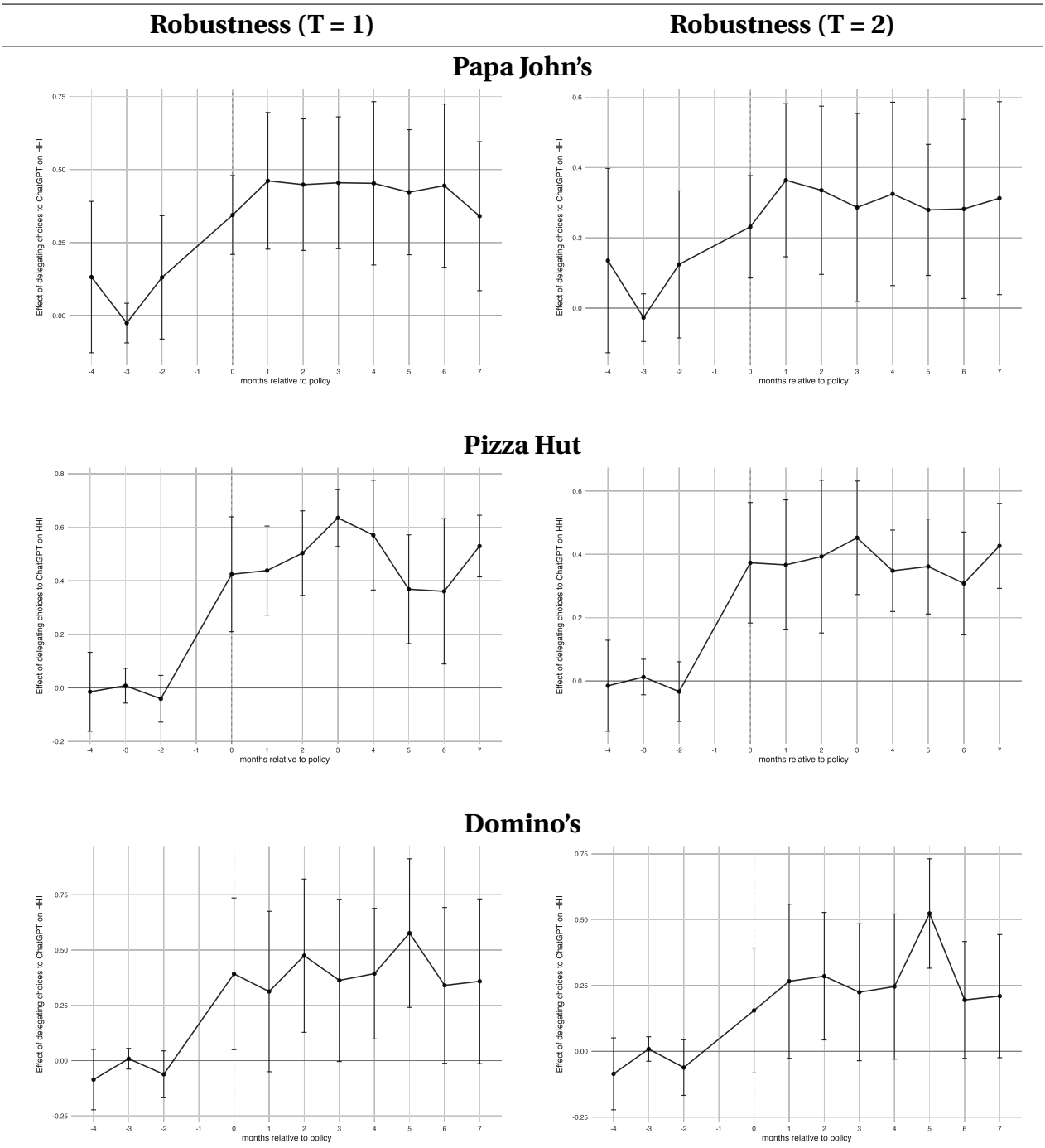
We assess robustness to increased stochasticity by varying the temperature parameter in the choice-generation process. Our benchmark experiment sets the temperature to zero, yielding fully deterministic outputs that reflect the model's most likely choices. We then conduct additional experiments with elevated temperature values of 1 and 2, which progressively increase randomness in the generation process. These alternative settings allow us to evaluate the sensitivity of our results to stochastic variation and to assess whether our findings remain stable when the model is permitted to explore less probable alternatives. The qualitative pattern is highly robust to increases in temperature: point estimates consistently indicate a decline in product variety and increase in within-restaurant concentration, while higher temperatures primarily lead to larger standard errors, reflecting greater variability in the generated choices.

Figure 12. Event-study estimates for log(unique products) by restaurant and temperature



Notes: Each row reports coefficient estimates from an event-study specification for the restaurant named above the row. The treated group corresponds to the restaurant shown when purchase decisions are delegated to ChatGPT, while the control group consists of the other major pizza chains. The unit of observation is restaurant-by-product-category-by-month, with product-category-by-restaurant fixed effects and event-month fixed effects. The outcome is the logarithm of the number of unique products purchased. Event time is measured in months relative to the policy month ($t = 0$), with $t = -1$ omitted as the reference period. Points show estimated coefficients and vertical bars denote 95% confidence intervals. Standard errors are clustered at the restaurant-by-product-category level. This figure reports robustness specifications using temperature $T = 1$ and $T = 2$ only.

Figure 13. Event-study estimates for HHI by restaurant and temperature

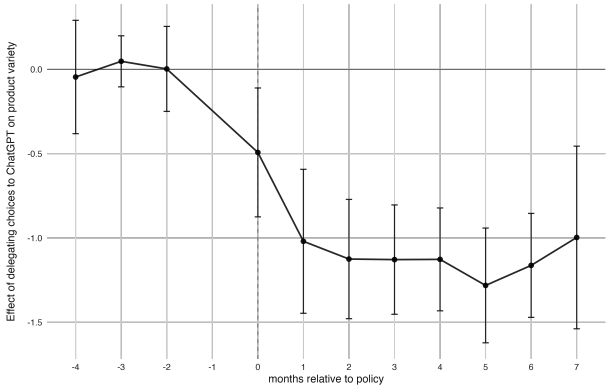


Notes: Each row reports coefficient estimates from an event-study specification for the restaurant named above the row. The treated group corresponds to the restaurant shown when purchase decisions are delegated to ChatGPT, while the control group consists of the other major pizza chains. The unit of observation is restaurant-by-product-category-by-month, with product-category-by-restaurant fixed effects and event-month fixed effects. The outcome is within-restaurant concentration (HHI). Event time is measured in months relative to the policy month ($t = 0$), with $t = -1$ omitted as the reference period. Points show estimated coefficients and vertical bars denote 95% confidence intervals. Standard errors are clustered at the restaurant-by-product-category level. This figure reports robustness specifications using temperature $T = 1$ and $T = 2$ only.

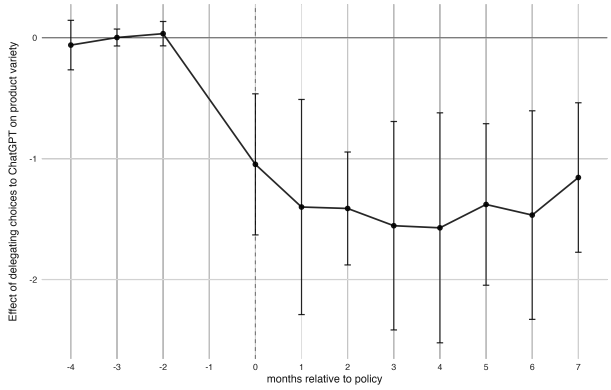
Alternative ChatGPT Model Versions

We replicate the analysis using an alternative version of ChatGPT to ensure that our results are not driven by a particular model release. Specifically, we replace GPT-5.1 with GPT-4o and find that the qualitative patterns remain robust.

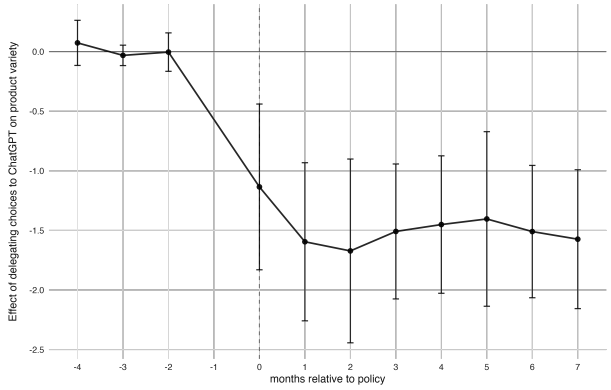
Figure 14. Event-study estimates of product variety (log unique products) after delegation



(a) Pizza Hut



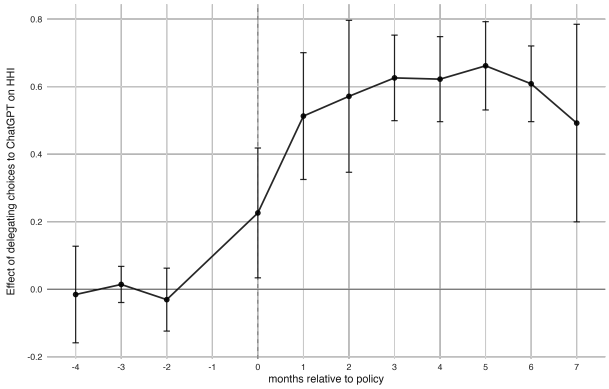
(b) Papa John's



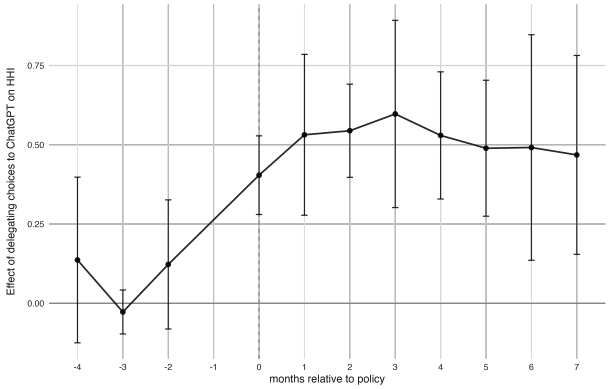
(c) Domino's

Notes: Each panel reports event-study estimates for a single choice scenario (Pizza Hut; Papa John's; Domino's). The unit of observation is restaurant-by-product-category-by-month, with product-category-by-restaurant fixed effects and event-month fixed effects. The outcome is the logarithm of the number of unique products purchased. In each panel, the treated group corresponds to the chain named in the title, while the control group consists of the other major pizza chains. Event time is measured in months relative to the policy month ($t = 0$), with $t = -1$ omitted as the reference period. Points show estimated coefficients and vertical bars denote 95% confidence intervals. Standard errors are clustered at the restaurant-by-product-category level.

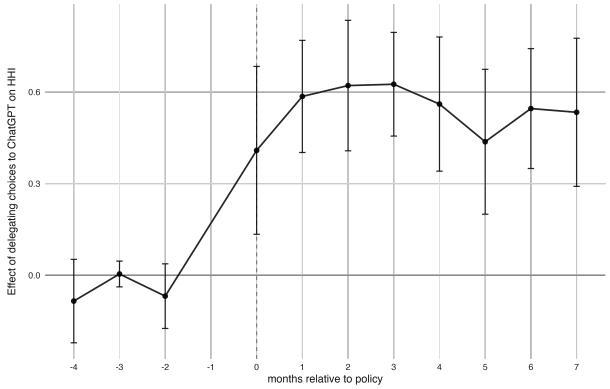
Figure 15. Event-study estimates for HHI



(a) Pizza Hut



(b) Papa John's



(c) Domino's

Notes: Each panel reports event-study estimates for a single choice scenario (Pizza Hut; Papa John's; Domino's). The unit of observation is restaurant-by-product-category-by-month, with product-category-by-restaurant fixed effects and event-month fixed effects. The outcome is within-restaurant concentration (HHI). In each panel, the treated group corresponds to the chain named in the title, while the control group consists of the other major pizza chains. Event time is measured in months relative to the policy month ($t = 0$), with $t = -1$ omitted as the reference period. Points show estimated coefficients and vertical bars denote 95% confidence intervals. Standard errors are clustered at the restaurant-by-product-category level.

G Estimated Nutritional Effects

Table 11. Post-policy nutritional effects: Domino's (*critique* scenario)

Nutrient	Avg. effect	SE	95% CI (lo)	95% CI (hi)
Sugars (g)	-2.049	0.711	-3.442	-0.656
Carbohydrates (g)	-1.834	0.714	-3.233	-0.434
Calories	-1.814	0.707	-3.200	-0.429
Sodium (mg)	-1.386	0.413	-2.195	-0.577
Total fat (g)	-1.131	0.429	-1.972	-0.290
Saturated fat (g)	-1.091	0.425	-1.925	-0.257
Fiber (g)	-0.960	0.299	-1.545	-0.375
Protein (g)	-0.902	0.386	-1.660	-0.145
Cholesterol (mg)	-0.794	0.545	-1.863	0.274
Trans fat (g)	0.153	0.606	-1.036	1.341

Notes: Entries report post-policy difference-in-differences estimates for the *critique* scenario. For each outcome, we estimate a specification with product-category-by-restaurant and event-month fixed effects, regressing $\log(\text{total nutrient})$ on the interaction between a treated indicator and a post-policy dummy (post_event = 1 for months $t \geq 0$), along with additional controls. Standard errors are clustered at the restaurant-by-product-category level. Confidence intervals are constructed as estimate $\pm 1.96\times$ SE.

Table 12. Post-policy nutritional effects: Pizza Hut (*critique* scenario)

Nutrient	Avg. effect	SE	95% CI (lo)	95% CI (hi)
Sugars (g)	-1.763	0.405	-2.557	-0.969
Sodium (mg)	-1.577	0.380	-2.323	-0.832
Cholesterol (mg)	-1.551	0.664	-2.852	-0.250
Calories	-1.524	0.386	-2.280	-0.768
Protein (g)	-1.517	0.405	-2.310	-0.723
Carbohydrates (g)	-1.503	0.407	-2.301	-0.705
Total fat (g)	-1.484	0.419	-2.304	-0.663
Saturated fat (g)	-1.408	0.427	-2.245	-0.570
Fiber (g)	-1.378	0.487	-2.334	-0.423
Trans fat (g)	-1.041	0.310	-1.650	-0.433

Notes: Entries report post-policy difference-in-differences estimates for the *critique* scenario. For each outcome, we estimate a specification with product-category-by-restaurant and event-month fixed effects, regressing $\log(\text{total nutrient})$ on the interaction between a treated indicator and a post-policy dummy (post_event = 1 for months $t \geq 0$), along with additional controls. Standard errors are clustered at the restaurant-by-product-category level. Confidence intervals are constructed as estimate $\pm 1.96\times$ SE.

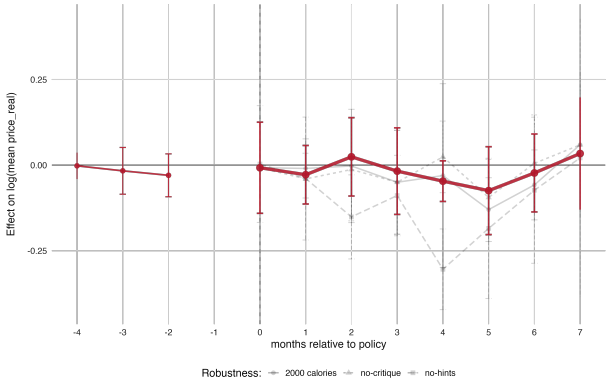
Table 13. Post-policy nutritional effects: Papa John's (*critique* scenario)

Nutrient	Avg. effect	SE	95% CI (lo)	95% CI (hi)
Sugars (g)	-2.121	0.633	-3.361	-0.880
Carbohydrates (g)	-1.774	0.708	-3.162	-0.385
Fiber (g)	-1.554	0.413	-2.364	-0.745
Sodium (mg)	-1.546	0.405	-2.340	-0.752
Calories	-1.539	0.517	-2.552	-0.525
Saturated fat (g)	-1.327	0.429	-2.169	-0.486
Total fat (g)	-1.214	0.380	-1.960	-0.469
Protein (g)	-1.183	0.411	-1.988	-0.378
Cholesterol (mg)	-1.003	0.515	-2.012	0.006

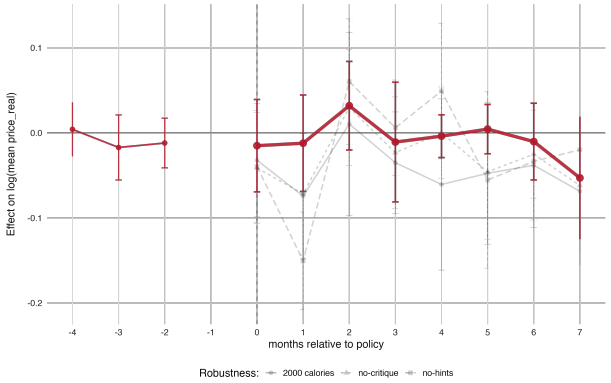
Notes: Entries report post-policy difference-in-differences estimates for the *critique* scenario. For each outcome, we estimate a specification with product-category-by-restaurant and event-month fixed effects, regressing $\log(\text{total nutrient})$ on the interaction between a treated indicator and a post-policy dummy (post_event = 1 for months $t \geq 0$), along with additional controls. Standard errors are clustered at the restaurant-by-product-category level. Confidence intervals are constructed as estimate $\pm 1.96\times$ SE.

H Supporting Evidence for the Price Assumption

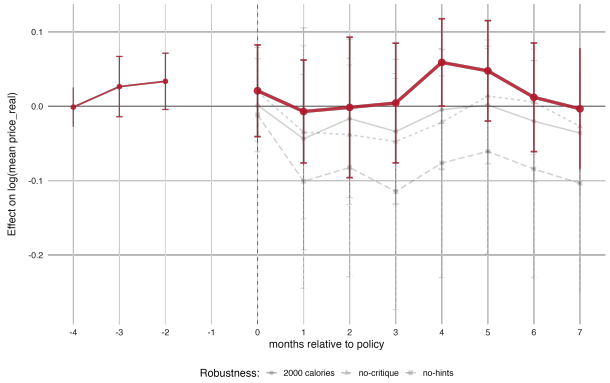
Figure 16. Event-study estimates of prices after delegation



(a) Pizza Hut



(b) Papa John's



(c) Domino's

Notes: Each panel overlays event-study estimates across multiple GPT-5 choice scenarios. The treated group corresponds to the chain named in the panel title and the control group consists of the other major pizza chains (Pizza Hut and Domino's for Papa John's; Pizza Hut and Papa John's for Domino's; Domino's and Papa John's for Pizza Hut). The unit of observation is restaurant-by-product-category-by-month, and all specifications include product-category-by-restaurant fixed effects and event-month fixed effects. Event time is measured in months relative to the policy month ($t = 0$), with $t = -1$ omitted as the reference period. The outcome is log of average price. Points show coefficient estimates and vertical bars denote 95% confidence intervals. Pre-policy coefficients ($t < 0$) are shown only for the baseline scenario ("nohints"), while post-policy coefficients ($t \geq 0$) are overlaid for all scenarios (critique, nohints, 2000/2000cal, and no-critique). The highlighted scenario in red corresponds to "critique". Standard errors are clustered at the restaurant-by-product-category level.

I First Stage Estimates

Table 14. First-Stage Estimates for Price

	Estimate	Std. Error
Nutritional controls		
Calories (per item)	0.728***	0.076
Total fat (g)	−4.255***	1.070
Saturated fat (g)	−1.062	1.423
Trans fat (g)	−119.2***	14.39
Cholesterol (mg)	1.489***	0.195
Sodium (mg)	0.0237*	0.0123
Total carbohydrates (g)	−0.160	0.0978
Fiber (g)	13.41***	3.847
Sugars (g)	−1.591**	0.757
Protein (g)	−6.422***	0.841
Excluded instruments		
Tariff China cheese 2018	1.199**	0.570
Sum quantity (same nest)	−0.0000844***	0.00000751
Rival mean quantity (same nest)	0.00376***	0.000362
Gandhi–Houde dist.-weighted carbs	−7.647***	1.252
Tariff Mexico × Pizza	−0.467*	0.245
Cheese shifter × Pizza	0.149*	0.0807
Tariff China cheese 2018 × delegated	0.258	0.277
Sum quantity × delegated	0.0001086***	0.00001299
Rival mean quantity × delegated	−0.00225***	0.000599
Gandhi–Houde dist.-weighted carbs × delegated	2.262*	1.188
Tariff Mexico × Pizza × delegated	0.641	0.527
Cheese shifter × Pizza × delegated	0.184	0.127
F-statistic	28.73	
Observations	1103	
Month fixed effects	Yes	

Notes: The table reports first-stage OLS estimates for the endogenous price variable. All regressions include nutritional controls and month fixed effects (coefficients on month dummies omitted). Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 15. First-Stage Estimates for Nest Share

	Estimate	Std. Error
Nutritional controls		
Calories (per item)	−0.0943*	0.0571
Total fat (g)	4.505***	0.801
Saturated fat (g)	−10.93***	1.065
Trans fat (g)	−2.511	10.77
Cholesterol (mg)	−0.534***	0.146
Sodium (mg)	−0.00390	0.00920
Total carbohydrates (g)	−0.00979	0.0732
Fiber (g)	−0.886	2.880
Sugars (g)	−0.120	0.567
Protein (g)	5.805***	0.630
Excluded instruments		
Tariff China cheese 2018	−2.042***	0.427
Sum quantity (same nest)	0.0000815***	0.00000562
Rival mean quantity (same nest)	−0.00437***	0.000271
Gandhi–Houde dist.-weighted carbs	1.474	0.937
Tariff Mexico × Pizza	0.159	0.183
Cheese shifter × Pizza	−0.324***	0.0604
Tariff China cheese 2018 × delegated	0.191	0.208
Sum quantity × delegated	−0.0000744***	0.00000973
Rival mean quantity × delegated	0.00296***	0.000449
Gandhi–Houde dist.-weighted carbs × delegated	−3.173***	0.889
Tariff Mexico × Pizza × delegated	0.366	0.395
Cheese shifter × Pizza × delegated	0.532***	0.0952
F-statistic	37.58	
Observations	1103	
Month fixed effects	Yes	

Notes: The table reports first-stage OLS estimates for the endogenous nest-share term $\log s_{j|r,t}$ (log within-restaurant share). All regressions include nutritional controls and month fixed effects (coefficients on month dummies omitted). Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

References

Allcott, H., B. B. Lockwood, and D. Taubinsky (2019): "Should we tax sugar-sweetened beverages? An overview of theory and evidence," *Journal of Economic Perspectives*, 33, 202–227.

Barahona, N., C. Otero, S. Otero, and J. Kim (2025): "Food Labeling Policies: Aggregate Impacts and Heterogeneity Across Categories," .

Bollinger, B., P. Leslie, and A. Sorensen (2011): "Calorie posting in chain restaurants," *American Economic Journal: Economic Policy*, 3, 91–128.

Brynjolfsson, E., D. Li, and L. Raymond (2025): "Generative AI at work," *The Quarterly Journal of Economics*, 140, 889–942.

Buçinca, Z., M. B. Malaya, and K. Z. Gajos (2021): "To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making," *Proceedings of the ACM on Human-computer Interaction*, 5, 1–21.

Centers for Disease Control and Prevention (2022): "Adult Obesity Facts," https://www.cdc.gov/obesity/adult-obesity-facts/index.html, accessed: 2026-01-01.

Chen, Y., T. X. Liu, Y. Shan, and S. Zhong (2023): "The emergence of economic rationality of GPT," *Proceedings of the National Academy of Sciences*, 120, e2316205120.

Courtemanche, C., D. Frisvold, D. Jimenez-Gomez, M. H. Ouayogodé, and M. K. Price (2025): "Chain restaurant calorie posting laws, obesity, and consumer welfare," *Journal of the European Economic Association*, jvaf004.

Cui, Z. K., M. Demirer, S. Jaffe, L. Musolff, S. Peng, and T. Salz (2025): "The effects of generative AI on high-skilled work: Evidence from three field experiments with software developers," *Available at SSRN 4945566*.

DellaVigna, S. and M. Gentzkow (2019): "Uniform pricing in us retail chains," *The Quarterly Journal of Economics*, 134, 2011–2084.

Demirer, M., A. Fradkin, N. Tadelis, and S. Peng (2025): "The Emerging Market for Intelligence: Pricing, Supply, and Demand for LLMs," *Supply, and Demand for LLMs (December 15, 2025)*.

Domino's Pizza, I. (2019): "2018 Annual Report," Tech. rep., Domino's Pizza, Inc., u.S. Retail Sales: US $6,591.6 million (2018)$.

Dubois, P., R. Griffith, and M. O'Connell (2018): "The effects of banning advertising in junk food markets," *The Review of Economic Studies*, 85, 396–436.

Fan, Y. and C. Yang (2025): "Estimating discrete games with many firms and many decisions: An application to merger and product variety," *Journal of Political Economy*, 133, 1886–1931.

Finkelstein, E. A., K. L. Strombotne, N. L. Chan, and J. Krieger (2011): "Mandatory menu labeling in one fast-food chain in King County, Washington," *American journal of preventive medicine*, 40, 122–127.

Finkelstein, E. A., J. G. Trogdon, J. W. Cohen, and W. Dietz (2009): "Annual Medical Spending Attributable to Obesity: Payer- and Service-Specific Estimates," *Health Affairs*, 28, w822–w831.

Gandhi, A. and J.-F. Houde (2019): "Measuring substitution patterns in differentiated-products industries," *NBER Working paper*.

Gul, F. and W. Pesendorfer (2001): "Temptation and self-control," *Econometrica*, 69, 1403–1435.

Hitsch, G. J., A. Hortacsu, and X. Lin (2021): "Prices and promotions in US retail markets," *Quantitative Marketing and Economics*, 19, 289–368.

Horton, J. J. (2023): "Large language models as simulated economic agents: What can we learn from homo silicus?" Tech. rep., National Bureau of Economic Research.

Hut, S. and E. Oster (2022): "Changes in household diet: determinants and predictability," *Journal of Public Economics*, 208, 104620.

Jardim, T. V., D. Mozaffarian, S. Abrahams-Gessel, S. Sy, Y. Y. Lee, and T. A. Gaziano (2019): "Cardiometabolic Disease Costs Associated With Suboptimal Diet in the United States," *PLOS Medicine*, 16, e1002981.

Li, Y., J. Xu, T. Liang, X. Chen, Z. He, Q. Liu, R. Wang, Z. Zhang, Z. Tu, H. Mi, et al. (2025): "Dancing with critiques: Enhancing llm reasoning with stepwise natural language self-critique," *arXiv preprint arXiv:2503.17363*.

Metcalfe, R. D., A. R. Schein, C. R. Simpson, and Y. Sun (2025): "Ai in charge: Large-scale experimental evidence on electric vehicle charging demand," Tech. rep., Technical report, Working paper.

Stanaway, J. D., A. Afshin, E. Gakidou, S. S. Lim, D. Abate, K. H. Abate, C. Abbafati, N. Abbasi, H. Abbastabar, F. Abd-Allah, et al. (2018): "Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017," *The lancet*, 392, 1923–1994.

U.S. Department of Health and Human Services and U.S. Department of Agriculture (2026): "Dietary Guidelines for Americans, 2025–2030," https://cdn.realfood.gov/DGA.pdf, accessed: 2026-01-XX.

Wisdom, J., J. S. Downs, and G. Loewenstein (2010): "Promoting healthy choices: Information versus convenience," *American Economic Journal: Applied Economics*, 2, 164–178.