UNIVERSITÉ DU QUÉBEC EN OUTAOUAIS

DÉPARTEMENT DES SCIENCES ADMINISTRATIVES

WORKING PAPER NO. 37

The Hidden Architecture of Grocery Prices: Promotions, Retail Competition, and the Cost of Living in Quebec

Evidence from 300,613 posted-price events across 54 Quebec grocery feeds, August–November 2026

Simon-Pierre Boucher

Département des sciences administratives
Université du Québec en Outaouais

simon-pierre.boucher@uqo.ca

Gatineau – Pavillon Alexandre-Taché
283, boulevard Alexandre-Taché
Gatineau, Québec, Canada J9A 1L8

November 2026

Version 1.0

Abstract

Grocery prices are not a single inflation process. Using a read-only research copy of Food-Ka, a Quebec grocery price aggregator, we assemble an auditable panel of 124,579 listings from 54 active source feeds (26 retail companies, two channels: online storefronts and weekly circulars) observed every seven hours between 12 August and 1 November 2026 (300,613 logged price events, 723,140 listing-weeks). We first document a measurement finding relevant to any study of scraped prices: storefronts that serve store-specific prices to anonymous sessions produce intra-day oscillation among three to five price levels, so sub-weekly "volatility" in such feeds is a mixture over unidentified stores, not repricing. A calibrated matching procedure (exact UPC within Metro Inc.; brand, package and token-identical names elsewhere; precision 0.93 and recall 0.98 on 4,413 UPC-identified pairs, plus a 240-pair reviewed sample) yields 5,909 cross-banner groups. On identical products in the same week, the median max–min log price gap across banners is 0.154 (79% of group-weeks exceed 5%), banner-pair gaps are persistent (AR(1) 0.66 against a 0.00 permutation placebo), and banner fixed effects range from -0.20 (Walmart, Giant Tiger, Maxi) to $+0.12$ (specialty shops) relative to Metro. Inferred promotions cover 4.6% of storefront listing-weeks with a median depth of 22%; half of all weekly price changes are promotion entries or exits, doubling measured volatility (ratio 1.92, CI 1.84–2.02). Temporary sales return exactly to their pre-sale reference (mean change -0.0003 log points): no reference-price drift within the window; a pre-sale price elevation is mostly a store-mixture artefact and shrinks to about two log points in single-store shops. Lead–lag price associations between rivals are small and fail a lead placebo for cross-company pairs, so no causal response is claimed. Matched-model and naive indices agree within one point at the market level over eleven weeks, but naive averages are off by 10–30 points for banners with assortment turnover. A fixed-basket mixed-integer program shows that unconstrained cheapest-banner shopping saves 15% relative to the best single banner, 85% of which two banners capture; store geography and travel costs are unobserved. We publish a five-component Quebec Grocery Price and Affordability Index with uncertainty bands and a prospective collection plan, and report null and failed tests alongside positive ones.

Keywords: grocery prices; price dispersion; promotions; high–low pricing; web-scraped prices; price indices; record linkage; retail competition; Quebec

JEL Classification: D22, D40, E31, L11, L81, C81

1. Introduction

Food inflation in Canada has become a political object. Between 2022 and 2024 food purchased from stores rose faster than the all-items Consumer Price Index, parliamentary committees summoned grocery executives, and the Competition Bureau concluded that the sector "needs more competition" (Competition Bureau Canada, 2023). Yet the public debate rests on a thin empirical layer: a monthly official index, a handful of average retail prices for selected products, and anecdotes about "shrinkflation" and dubious promotions. What a household actually pays for a comparable basket depends on which banner it visits, whether the item is on promotion that week, which package it buys, and how far it is willing to travel. None of these margins is visible in an aggregate index.

This paper asks a narrow, falsifiable set of questions about the *architecture* of posted grocery prices in Quebec, using a data source that was not built for research but turns out to be unusually informative: Food-Ka, an independent price-comparison service that re-reads the public catalogues and weekly circulars of 57 grocery feeds every few hours and logs every price change it observes. We obtained a read-only copy of its production database on 1 November 2026. The copy contains 124,579 listings, 300,613 logged price events and 15,630 connector passes between 12 August and 1 November 2026. We treat this as what it is—eleven complete weeks of *posted* prices from *online* channels with no store identifiers, no transactions and no quantities—and we ask what can and cannot be learned from it.

Our contribution is threefold. First, we document a measurement phenomenon that we believe is new to the literature on scraped prices (Cavallo and Rigobon, 2016; Cavallo, 2016, 2018; Gorodnichenko and Talavera, 2017): storefronts of the large chains answer anonymous sessions with the price of a store they assign to the session, so a high-frequency scraper that does not fix a store records a mixture of store prices. The signature is unmistakable—a Metro listing of mayonnaise cycles between \$5.77, \$6.99, \$7.49, \$7.79 and \$7.99 within a single day, more than half of logged "changes" revert to the previous-but-one value, and the phenomenon is absent from single-store online shops (Section 4). We show how to extract a defensible weekly posted price from such data and how the oscillation amplitude itself bounds within-banner, across-store dispersion.

Second, we build and validate a cross-banner matching procedure. Only the Metro Inc. feeds expose GS1-valid UPCs; everything else requires probabilistic linkage. We calibrate the text rules on 37,629 candidate pairs between Metro and Super C whose shared UPC provides ground truth, obtaining a rule (identical brand, package and cleaned token set) with precision 0.93 and recall 0.98 on that labelled set, and we validate it on a 240-pair stratified sample reviewed pair by pair. A looser rule that a first version of the pipeline used (token-*set* similarity) was rejected after validation; we report it. The result is 5,909 cross-banner product groups—a few per cent of all listings, which is the honest denominator for any "price comparison" claim about this market.

Third, on this matched panel we estimate the preregistered hypotheses H1–H6 of our registration (frozen before estimation; Section 6). On identical products in the same week the median across-banner log price gap is 0.154, dispersion is persistent (banner-pair gaps have an AR(1) coefficient of 0.66 against a permutation placebo of 0.00), and banner fixed effects span 32 log points. Inferred promotions are rare in storefront catalogues (4.6% of listing-weeks) but deep (median 22%), account for half of all weekly price changes and double measured volatility. Temporary sales return exactly to their pre-sale reference price—we find no reference-price drift within the window—while a pre-sale elevation of the posted price is largely a store-mixture artefact. Lead–lag associations between rivals' price changes are small, mostly within company, and fail a lead placebo for cross-company pairs; we therefore make no causal claim about competitive responses. Matched-model and naive indices agree closely at the market level over eleven weeks, but diverge by 10–30 points for individual banners with assortment turnover. A fixed-basket optimisation shows that unconstrained cheapest-banner shopping would save about 15% relative to the best single banner and that two banners capture 85% of that; travel costs and store access are unobserved and enter only as labelled assumptions. We assemble these pieces into a transparent five-component Quebec Grocery Price and Affordability Index (QGPAI) whose fifth component—access and affordability—is deliberately left empty.

Throughout we separate what the data identify from what they suggest. Posted prices are not paid prices; a listing that disappears has not been sold; a "sale" tag is not a discount. The eleven-week window is too short for several of the registered tests (notably H3 on reference-price adjustment and the validation of QGPAI against official indices), and we say so rather than stretch the estimand. Section 10 and the prospective collection plan in Section 12 describe what a longer, store-identified panel would add.

The remainder of the paper is organised as follows. Section 2 reviews the literature on price dispersion, sales and scraped prices; Section 3 describes the Quebec grocery sector and the Food-Ka source; Section 4 audits the data and constructs the panel and the matches; Section 5 sets out a simple model whose observable implications we test; Section 6 states the preregistered designs; Sections 7 and 8 report results and robustness; Sections 9–11 discuss implications, limitations and conclusions; Section 12 documents reproducibility and data access.

2. Related literature

Price dispersion and search. That identical goods sell at different prices in the same market is the oldest puzzle of the economics of information (Stigler, 1961). Equilibrium dispersion arises when some consumers are informed and others are not (Salop and Stiglitz, 1977), or when consumers sample a small number of prices (Burdett and Judd, 1983); Baye et al. (2006) survey the theory

and the evidence. Empirically, dispersion is large and persistent even for homogeneous products (Sorensen, 2000; Lach, 2002), and most of it is *relative* dispersion—the same store is cheap for some goods and expensive for others—rather than a stable store-level ranking (Kaplan and Menzio, 2014; Kaplan et al., 2019). Our H1 tests whether the Quebec banner-level posted prices display economically meaningful and persistent gaps on exactly matched items; the within-banner store mixture we uncover connects directly to the store-level dispersion these papers measure with scanner data.

Sales, high–low pricing and reference prices. Varian (1980) rationalises temporary sales as a mixed strategy that price-discriminates between informed shoppers and loyal customers; Pesendorfer (2002) and Hendel and Nevo (2005, 2013) show how storable-goods demand accumulation generates sale cycles; Hosken and Reiffen (2004) document the empirical regularities of retail price variation; Chevalier et al. (2003) and MacDonald (2000) show prices falling at demand peaks; Chevalier and Kashyap (2019) measure how consumers exploit "best prices". The macro literature treats sales as distinct from regular-price changes: regular prices are sticky, sales are frequent, temporary and return to the reference (Bils and Klenow, 2004; Nakamura and Steinsson, 2008; Klenow and Kryvtsov, 2008; Eichenbaum et al., 2011; Kehoe and Midrigan, 2015), sales respond weakly to aggregate shocks (Guimaraes and Sheedy, 2011; Coibion et al., 2015; Anderson et al., 2017; Gagnon and López-Salido, 2019), and menu-cost models with multi-product firms reproduce the size distribution of changes (Golosov and Lucas Jr., 2007; nan, 2011; Nakamura and Steinsson, 2013; Anderson et al., 2015). Our H2 and H3 are the posted-price implications of this literature: volatility with and without sales, and whether the reference price moves after a sale.

Retail competition and chain pricing. Retailers choose between every-day-low-price and high–low formats (Hoch et al., 1994; Lal and Rao, 1997; Ellickson and Misra, 2008); large chains price nearly uniformly across their stores (DellaVigna and Gentzkow, 2019; Hitsch et al., 2019, 2021), with zone pricing as the main exception (Adams and Williams, 2019); entry of mass merchants lowers measured prices (Hausman and Leibtag, 2006). The Canadian retail grocery market is concentrated among three national chains plus Walmart and Costco (Competition Bureau Canada, 2023). We map Food-Ka's 54 feeds to 26 companies and show that most exactly matched comparisons are *within* company (Metro vs. Super C; Maxi vs. Provigo vs. Club Entrepôt)—format differentiation inside a chain—rather than across chains.

Scraped and online prices. The Billion Prices Project established that online prices are a valid high-frequency source for inflation measurement and that online and offline prices coincide for most items of multi-channel retailers (Cavallo and Rigobon, 2016; Cavallo, 2016); scraped data

reveal more frequent price changes and stockouts than official collections (Cavallo, 2018; Cavallo and Kryvtsov, 2023; Gorodnichenko and Talavera, 2017). Our contribution to this literature is cautionary: when a retailer's site serves store-specific prices and the scraper does not pin the store, the "online price" is a draw from the chain's store-price distribution.

Price indices, substitution and shrinkflation. Matched-model and multilateral methods address chain drift and assortment change (Feenstra, 1994; Ivancic et al., 2011; de Haan and van der Grient, 2011; Melser, 2017; Diewert and Fox, 2020b; Redding and Weinstein, 2019); price measurement during the pandemic highlighted the role of changing baskets (Diewert and Fox, 2020a); household-level differences in inflation arise from product-level heterogeneity (Jaravel, 2018; Argente and Lee, 2020; Beck and Jaravel, 2020; Hottman et al., 2016). Package downsizing is a hidden price increase to which demand is less sensitive than to list-price increases (Gourville and Koehler, 2004; Çakır and Balagtas, 2014; Çakır, 2021; Melmiès, 2024). Our H5 compares naive average posted prices with matched-model indices on the same data; we cannot observe within-listing package changes and say so.

Shopping effort, store choice and access. Paid prices fall with shopping time and store switching (Aguiar and Hurst, 2007; Griffith et al., 2009; Kaplan and Menzio, 2016; Nevo and Wong, 2018); private labels and brand premia reflect information as much as quality (Bronnenberg et al., 2014; Dubé et al., 2018); spatial access and assortment shape effective prices (Handbury and Weinstein, 2014; Allcott et al., 2019; Atkin et al., 2018). Our H6 computes the posted-price savings frontier under a banner-count constraint; because store geography is unobserved, travel costs enter only as an illustrative sensitivity.

Record linkage. Probabilistic matching of product listings is a record-linkage problem (Fellegi and Sunter, 1969; Christen, 2012); recording error in scanner panels is itself a research topic (Einav et al., 2009). We treat matching as a measured component of the design, with tier-specific precision estimates propagated to the results.

The evidence matrix in the replication package (`literature/evidence_matrix.csv`) links each theoretical claim used in this paper to these sources, states the limitation for our setting, and identifies where WP37 adds something.

3. Institutional context: Quebec grocery retail and the Food-Ka source

Market structure. Grocery retail in Quebec is dominated by three vertically integrated companies operating several banners each: Metro Inc. (Metro, Super C, Marché Adonis, the affiliated Richelieu,

Marché Ami and Axep banners, and the Jean Coutu and Brunet pharmacies), Loblaw Companies (Maxi, Provigo, Club Entrepôt, Pharmaprix, T&T Supermarket, and affiliated Inter-Marché stores) and Sobeys/Empire (IGA, Les Marchés Tradition, Marché Bonichoix, Rachelle-Béry). Walmart Canada and Costco add mass-merchant and warehouse-club formats; Giant Tiger is a discount general merchandiser with a large food assortment; a long tail of independent, ethnic, health-food and online-only grocers completes the market. The Société des alcools du Québec (SAQ) is a provincial monopoly with uniform posted prices and is excluded from all grocery analyses. Company affiliations used in the paper are recorded feed by feed in the `retailers` table with the basis for each attribution (public corporate information or the Food-Ka registry); they are descriptive and do not affect any estimate except the same-company/cross-company split.

Channels. Each large banner is present in Food-Ka through two *feeds*: an online storefront (the e-commerce catalogue, read every ≈7 hours) and a weekly circular (read from the Flipp platform for the Montréal postal code H2X 1Y4, with validity dates, regular price and fine print). Smaller independents are present only through circulars; online-only grocers only through storefronts. Treating the 54 feeds as 54 retailers would therefore double count the large chains and conflate two very different price-setting objects: a catalogue price that applies to whatever store the site assigned to the session, and an advertised weekly promotion with an explicit regular price. We keep the two channels apart throughout.

The Food-Ka source. Food-Ka (`https://www.food-ka.com`) is an independent, non-commercial aggregator operated by the author. It maintains one connector per feed, normalises each listing to a common schema (name, brand, category, package label, price, regular price, unit price), computes a content hash and, at every pass, classifies each listing as new, modified or missing. Every change of the displayed price is appended to a log; a listing missing for two consecutive passes is deactivated. The public home page on 1 November 2026 reported 50,760 products, 54 banners and 8,631 products on promotion; we verified that these equal the number of *active listings*, feeds with active listings and listings flagged on sale in the snapshot, and we stress that they are counts of source-feed listings, not of unique products: the same Coca-Cola 12-pack appears once per feed that carries it.

What the source does and does not record. Food-Ka was built to compare current prices, not to study price dynamics. Three consequences shape everything below. (i) It logs *changes*, not observations: between two logged events a listing kept its price and was re-observed at every pass, but the pass-level presence of each listing is not stored. (ii) It historises the *current* price only: regular prices, promotion flags and stock status are overwritten at each pass, so the promotion state of a storefront listing in the past must be inferred from its price path, while circular items carry an

explicit regular price for their validity week. (iii) It has no store dimension: storefront connectors browse without selecting a store, and circulars are pulled for one postal code. These are the facts; the research design (Section 6) is built around them rather than against them.

4. Data: audit, panel reconstruction and matching

4.1. Snapshot and provenance

On 1 November 2026 at 07:57 UTC a consistent backup of the production SQLite database was taken through a read-only connection (SHA-256 recorded in the replication package; the Food-Ka code at commit `2c76523` was copied alongside to document the data-generating process). Production was not modified at any point. The snapshot holds 124,579 listings (50,760 active), 300,613 price-log rows, 15,630 connector passes over 57 feeds (median gap between passes of the same feed: 7.2 hours), 2,537 listings linked by Food-Ka's own cross-banner matcher, and auxiliary caches. Table 1 summarises coverage for the largest feeds; Figure 1 shows weekly coverage by channel and active listings by company.

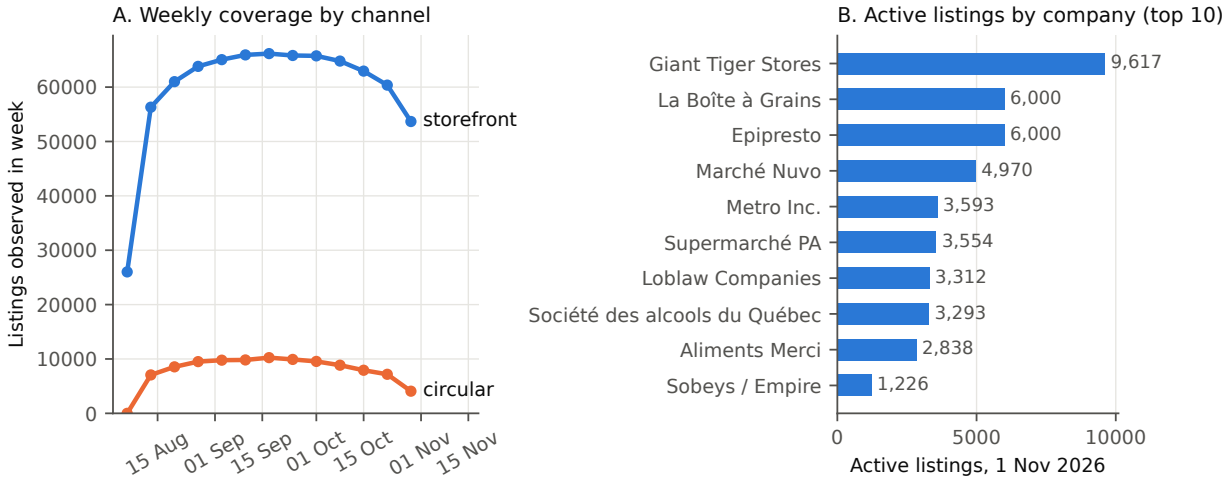


Figure 1. Data coverage. A: listings observed per circular week by channel. B: active listings on 1 November 2026 by company (top 10). Source: Food-Ka snapshot of 1 November 2026; authors' computations.

The composition is heavily skewed. Four feeds—Giant Tiger, Epipresto, La Boîte à Grains and Marché Nuvo—hold 52% of active listings because their Shopify catalogues are read exhaustively, whereas the connectors of the large chains read selected aisles (Metro 1,540 active listings, Maxi 1,086, Walmart 922, Super C 857, IGA 600, Provigo 450). The large chains nevertheless generate most of the price events (Metro alone 63,868; Club Entrepôt 68,438) and most of the deactivated listings, for reasons we turn to now.

Table 1. Coverage of the 24 largest source feeds (snapshot of 1 November 2026)

Company	Banner	Channel	Listings	Active	Median days	Price events	Unit price	Brand
Metro Inc.	Metro	online_storefront	16 382	1 540	50	63 868	0.91	0.88
Giant Tiger Stores	Giant Tiger	online_storefront	10 352	9 617	81	16 810	0.74	1.00
Société des alcools du Québec	SAQ	online_storefront	7 477	3 293	36	8 153	1.00	1.00
Metro Inc.	Super C	online_storefront	6 418	857	48	15 210	0.94	0.93
La Boîte à Grains	La Boîte à Grains	online_storefront	6 279	6 000	75	11 903	0.76	1.00
Marché Nuvo	Marché Nuvo	online_storefront	6 224	4 970	75	8 186	0.97	1.00
Epipresto	Epipresto	online_storefront	6 000	6 000	75	5 984	0.32	1.00
Walmart Canada	Walmart	online_storefront	4 601	922	19	7 872	0.56	0.28
Supermarché PA	PA Supermarché	online_storefront	3 602	3 554	81	6 864	0.19	1.00
Walmart Canada	Walmart	circular	3 330	219	8	3 501	0.09	0.66
Sobeys / Empire	IGA	online_storefront	3 083	600	34	5 848	0.95	0.94
Metro Inc.	Marché Adonis	online_storefront	2 950	447	52	5 935	0.93	0.00
Sobeys / Empire	IGA	circular	2 933	350	9	3 527	0.03	0.75
Aliments Merci	Aliments Merci	online_storefront	2 859	2 838	75	3 367	0.05	0.85
Loblaw Companies	Club Entrepôt	online_storefront	2 808	763	74	68 438	1.00	0.76
Loblaw Companies	Maxi	online_storefront	2 441	1 086	78	13 674	1.00	0.82
Metro Inc.	Metro	circular	2 256	314	9	2 871	0.01	0.00
Loblaw Companies	Provigo	online_storefront	2 138	450	77	5 582	1.00	0.80
Loblaw Companies	Maxi	circular	2 107	271	9	2 341	0.62	0.85
Metro Inc.	Super C	circular	1 954	192	9	2 448	0.00	0.30
Loblaw Companies	Provigo	circular	1 914	175	9	2 035	0.55	0.79
Metro Inc.	Marché Adonis	circular	1 882	203	9	2 131	0.01	0.62
Pasquier	Pasquier	circular	1 611	152	9	1 727	0.21	0.75
Loblaw Companies	Pharmaprix	circular	1 225	167	8	1 378	0.51	0.51

Listings = all uids ever observed in the feed (active and deactivated); Active = listed on 1 Nov 2026; Median days = median observation span; Price events = rows in the price log; Unit price / Brand = share of active listings with a parsable package size (re-parsed) / a brand field. Channel: online storefront vs weekly circular (Flipp, postal code H2X 1Y4). Full table of 54 feeds in the replication package.

4.2. Source-to-variable inventory

Table 2 classifies each variable the mandate asked for. Product name, category (18 canonical classes), package label and current price are confirmed for all feeds; brand is missing for the Walmart, Adonis, T&T, Akhavan, BocoBoco and Costco feeds (the brand sits inside the name); GS1-valid UPCs exist only for Metro and Super C (100% check-digit validity, 1,954 active listings), while Shopify, Loblaw and Walmart identifiers are internal SKUs (check-digit validity at the 10% chance level). Package size is derivable for 63% of active grocery listings after re-parsing labels and names (Food-Ka's own parser: 43%); it is missing for most circular items and for Giant Tiger and Nuvo. Regular prices, promotion flags and stock status exist only as current state. Store and geography do not exist. Quantities and expenditure weights do not exist.

Table 2. Source-to-variable inventory

Variable	Status	Notes
Product name	confirmed	all feeds; French/English mixed
Brand	confirmed / partly unavailable	missing for 23% of active listings (six feeds carry no brand field)
UPC / GTIN	confirmed for 2 feeds	Metro, Super C: GS1-valid; elsewhere internal SKUs; Open Food Facts barcodes attached by name lookup are derived and unreliable
Package size, unit	derivable	re-parsed from label or name: 63% of active grocery listings; missing mostly for circulars, Giant Tiger, Nuvo
Category	confirmed	18 canonical classes; "Autres" is a residual (19%)
Retailer / banner / company	derivable	54 feeds $\rightarrow$ 43 banners with grocery listings $\rightarrow$ 26 companies
Store	unavailable	no store identifier; storefronts answered anonymous sessions; circulars for one postal code
Current price	confirmed	event log: first observation + every change
Regular / reference price	current state only	historised nowhere; circulars carry a regular price per item-week
Unit price	derivable	recomputed; per-kg labels, multi-buy labels flagged
Promotion flag	current state only	historical state inferred from price paths (storefront) or implicit (circulars)
Observation timestamp	confirmed (feed level)	15,630 passes; listing-level presence between events inferred
Circular validity	confirmed	valid-from / valid-to per item
Availability / stock	unreliable	NULL for 23% of active listings; current only
Geography, quantities, expenditure weights	unavailable	—

Classification per the mandate: confirmed, derivable, unavailable, unreliable. Full inventory with per-feed rates in `reports/phase1_data_audit.md`.

4.3. Panel reconstruction

Because the log records changes, a listing's price between two events is the last logged value, and the listing was observed at every pass of its feed between its first and last sighting. We reconstruct, for every listing and local day (America/Toronto), the duration-weighted distribution of displayed prices—mode, minimum, maximum, time-weighted mean, open and close, number of changes and number of distinct values—and aggregate to the Quebec circular week (Thursday to Wednesday). Two limitations are inherent: a listing that leaves and re-enters a feed without a price change leaves no trace, and deactivation requires two consecutive misses, so removals are dated to within one pass. Invariant tests verify that covered hours equal the observation span, that daily bounds hold and that single-event listings have a single price. The grocery universe (feeds other than pharmacies and the SAQ; categories other than health and beauty, household, pets, baby and alcohol; at least one observed price; no unrealistic price) spans 723,140 listing-weeks for 102,835 listings in 43 banners over the 11 complete weeks from 13 August to 22 October 2026 (the partial first and last weeks are dropped).

4.4. A measurement finding: store-mixture noise

Figure 2A plots every logged price of one Metro listing. The price alternates, within the same day, among \$5.77, \$6.99, \$7.49, \$7.79 and \$7.99. This is not repricing. In the feeds of Club Entrepôt, Metro, Maxi, Super C, Provigo and Walmart, 48–68% of logged changes revert to the previous-but-one value, the median listing with six or more events cycles among four distinct values, and the set of values observed in one week overlaps the next week's set with a median Jaccard index of 0.5. Single-store online shops (Shopify and WooCommerce catalogues) show none of this. The Food-Ka connectors for the large chains reach the sites through residential proxy pools without selecting a store; the sites then serve the price of whatever store they associate with the session. Each logged "change" in these feeds is therefore a draw from the banner's distribution of store prices.

Three consequences follow and were fixed in the preregistration before any estimation. (i) The posted price of a listing in a week is defined as the duration-weighted *modal* price, with minimum, maximum and mean as sensitivity definitions. (ii) Sub-weekly variation in these feeds is not interpreted as price flexibility; H2 is tested on weekly series and the falsification that single-store shops must not display the oscillation is part of the design. (iii) The within-week range of a listing in these feeds is a lower bound on within-banner, across-store dispersion for that product—a quantity we report (Table 5 note and Figure 2B) with the caveat that the stores are unidentified: 23% of storefront listing-weeks in the anti-bot feeds display more than one price value, with a median log range of 0.15–0.24, against 4.7% in single-store shops.

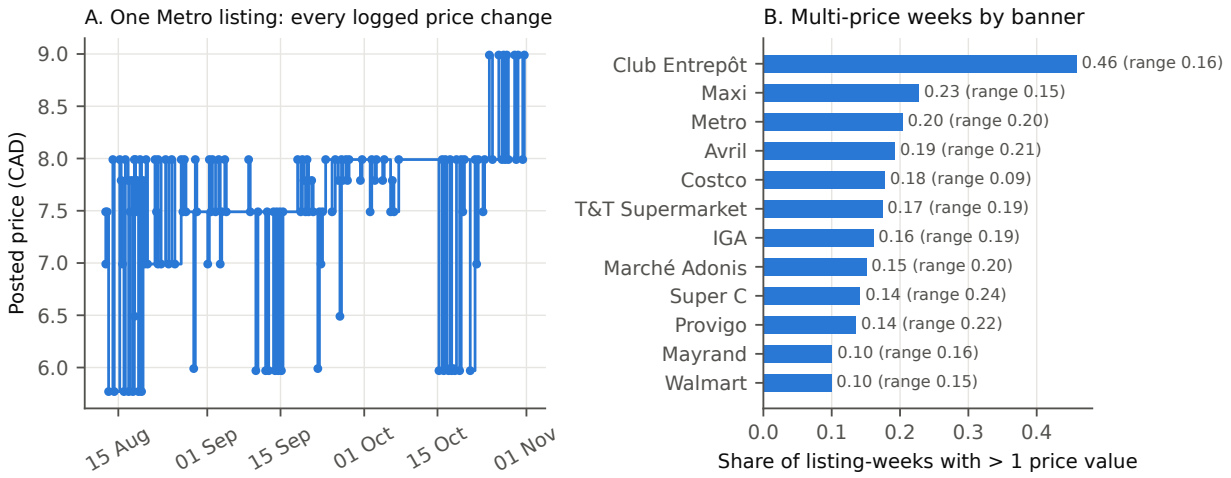


Figure 2. Store-mixture noise. A: all logged prices of one Metro listing (mayonnaise, 890 mL), 12 August – 31 October 2026. B: share of listing-weeks in which more than one price value was displayed, by banner, with the median log range (max/min) on those weeks. Feeds behind anti-bot protection that serve store-specific prices (Club Entrepôt, Maxi, Metro, Super C, Provigo, Walmart) stand apart from single-store online shops.

4.5. Attrition

Table 3 and Figure 3A track listings through the filters. Of 124,579 listings, 111,575 belong to the grocery universe, 66,340 have a parsable package size and are eligible for matching, and 13,273 end up in a cross-banner group under tiers A–C (13,088 under the primary tiers A+B). The denominator of every cross-banner statistic in this paper is therefore about 11% of listings, not the 50,760 products advertised on the home page.

Table 3. Listing-level sample attrition

Step	Listings	Share of initial
All listings in snapshot (products table)	124 579	1.000
excl. SAQ (alcohol monopoly)	117 102	0.940
excl. pharmacy feeds	115 352	0.926
excl. non-food categories (health & beauty, household, pets, baby, alcohol)	112 233	0.901
with at least one observed price	111 924	0.898
excl. unrealistic prices	111 575	0.896
… with parsable package size (unit-price / matching universe)	66 340	0.533
… in a cross-banner match group (tiers A–C)	13 273	0.107

Each row applies the stated filter to the previous row. Matching tiers A–C shown; the primary analysis uses A+B.

4.6. Cross-banner matching

We match listings across banners in tiers of decreasing confidence. Tier A is exact equality of GS1-valid UPCs (only Metro and Super C). Tier B requires the same brand key, the same package (total quantity in grams, millilitres or units *and* the same multipack count, so that 12×355 mL and 6×710 mL are different packages), identical cleaned token sets after removing size tokens, stop words and the brand, and identical numeric tokens ("2%" versus "3.25%"). Tier C relaxes token identity to a token-sort similarity of at least 0.90. Tier D handles feeds without a brand field by requiring identical full-name token sets, the same package, and the presence of the branded side's brand tokens in the brand-less name. Size-less listings and multi-product circular offers ("X ou Y", "assortis", "sélectionnés") are excluded from matching; a first version that allowed them produced chained groups of up to 141 listings.

The thresholds were not chosen by eye. All 36,700 Metro×Super C candidate pairs sharing a brand and a package were scored and labelled by whether they share a UPC (4,328 positives). Token-set similarity—the scorer used by Food-Ka's own matcher and by a first version of ours—scores a name that is a subset of another at 1.0 ("beurre végétal salé" matches "beurre végétal non salé"); its precision at threshold 0.90 is far lower (0.61 on a first calibration without multipack-aware package keys). Identical token sets achieve precision 0.932 and recall 0.982; token-sort similarity ≥ 0.90 cumulatively 0.870 and 0.987. The residual 6.8% of identical-text pairs with different UPCs are packaging variants invisible in text, so 0.93 is a lower bound on the precision of tier B relative to what a reader of the two listings could tell apart.

Figure 3B and Table 4 report the validation on a stratified random sample of 60 pairs per tier. Each pair was labelled by Claude (claude-opus-5-5) as a second annotator and then reviewed, pair by pair, by the research operator; the two agreed on all 240 pairs after review. Tiers A and B are correct in 60/60 cases each, tier D in 59/60 (one unsure), tier C in only 25/60 (five unsure): the token-sort rule admits flavour, roast, fat-content and colour variants ("douces" vs. "piquantes", "torréfaction moyenne" vs. "foncée"). The preregistration therefore fixes tiers A+B as the primary specification and uses C and D only in sensitivity analyses. The tiers yield 4,413 (A), 4,088 (B), 714 (C) and 102 (D) cross-banner pairs, and 5,909 groups under A+B covering 13,088 listings; the median group spans two banners and the largest seven. Against Food-Ka's own 2,033 linked pairs, our tiers recover 42.5%—the rest are size-less or multi-offer circular links that we exclude by design.

One structural fact emerges from the matched panel and conditions the interpretation of everything that follows: 93% of matched group-weeks involve banners of the *same* company (Metro vs. Super C; Maxi vs. Provigo vs. Club Entrepôt). Exactly matched cross-company comparisons number 2,103 group-weeks. The apparent diversity of 54 banners collapses, for the purpose of like-for-like comparison, to format differentiation within two chains plus a thin layer of cross-chain

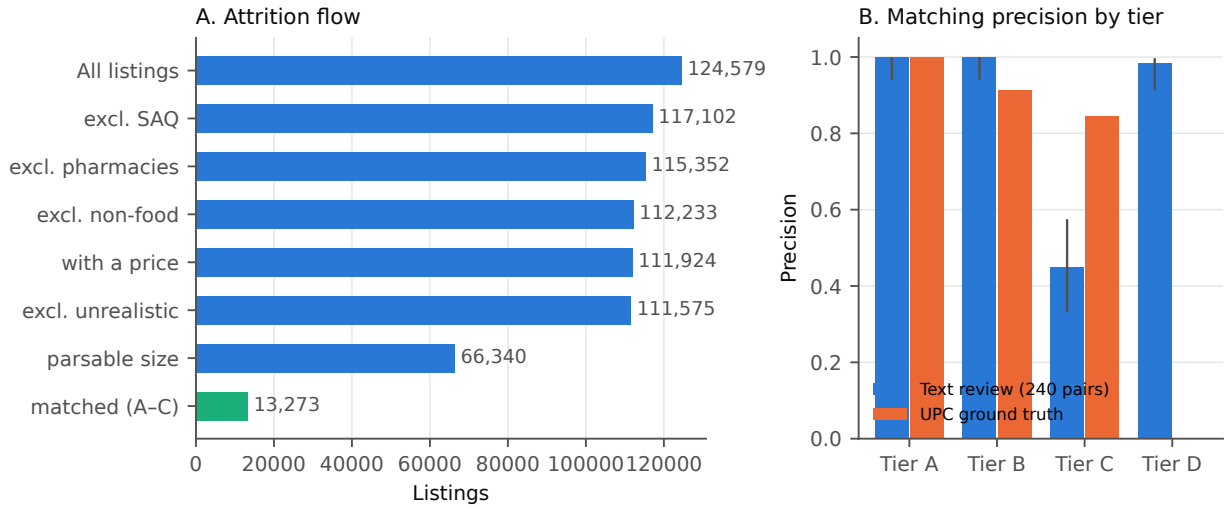


Figure 3. Matching validation and attrition. A: listing-level attrition flow. B: precision by tier from the 240-pair reviewed sample (Wilson 95% intervals) and from UPC ground truth (Metro×Super C blocks; none for tier D).

Table 4. Matching validation: stratified sample (60 pairs per tier) and UPC ground truth

Tier	n	Same	Different	Unsure	Precision	CI low	CI high	UPC precision	UPC recall
A	60	60	0	0	1.000	0.940	1.000	1.000	
B	60	60	0	0	1.000	0.940	1.000	0.932	0.982
C	60	25	30	5	0.417	0.301	0.543	0.870	0.987
D	60	59	0	1	0.983	0.911	0.997		

Precision = share labelled SAME (Wilson 95% CI); labels by Claude (claude-opus-5-5) as second annotator with full operator review of every pair; PI adjudication pending. UPC columns: precision/recall of the tier's text rule on all Metro × Super C candidate pairs within brand × package blocks, where the shared GS1 code is the truth (B: identical token sets; C: token-sort ratio ≥ 0.90). Tier D has no UPC benchmark.

pairs.

5. Theoretical framework

We use a deliberately small model to organise the hypotheses and to be explicit about what posted-price data can identify. Retailer (banner) r sells product i in week t to two consumer types in the spirit of Varian (1980) and Salop and Stiglitz (1977): a share $1-\lambda_i$ of *loyal* shoppers who buy at their habitual banner if the posted price does not exceed their reservation value v_i, and a share λ_i of *shoppers* who observe all posted prices (for instance through a comparison service) and buy from the cheapest banner. Banner r faces marginal procurement cost c_{irt}, a format-specific cost shifter κ_r (service level, store size, location rents) and chooses a posted price p_{irt} and a promotion state $s_{irt} \in \{0,1\}$. Promotions are temporary reductions below a *reference* price $\bar{p}_{irt}$ that is costly to change (a menu cost m in the sense of Golosov and Lucas Jr., 2007 and nan, 2011) and that loyal shoppers use as their norm.

Equilibrium structure. With R banners carrying the product, the standard result is a mixed-strategy equilibrium in which each banner posts its reference price $\bar{p}_{ir} = \bar{p}_{ir}(\kappa_r, c_{ir}, v_i)$ most of the time and randomises temporary reductions to attract shoppers; the probability of a promotion increases in λ_i and in the number of rival banners, and the depth of the reduction is bounded by the gap between the reference price and the shopper-competitive price. Three properties are robust across variants of the model and generate our registered hypotheses.

1. **Persistent reference-price dispersion (H1).** Reference prices differ across banners by κ_r and c_{ir}, which move slowly; hence log gaps between banners on the same product should be large relative to zero and autocorrelated across weeks, whereas a model in which banners price at a common marginal cost predicts gaps near zero and no persistence. Within a company, format differentiation ($\kappa_{\text{discount}} < \kappa_{\text{supermarket}}$) predicts a *stable sign* of the gap (the discount banner is cheaper), which is testable on the Metro/Super C and Maxi/Provigo pairs.
2. **Sales double the volatility of posted prices without moving the reference (H2, H3).** Because promotions are temporary deviations, the variance of weekly log price changes decomposes into a reference component and a promotion component; the ratio of total to reference volatility exceeds one by the promotion share and depth. The model with a menu cost on $\bar{p}$ predicts that the reference price after a sale equals the reference before it (no drift). The alternative "reference-price inflation" hypothesis—the retailer raises $\bar{p}$ before or after a sale so that the displayed discount looks larger—predicts a pre-sale elevation or a post-sale increase of the reference.
3. **Competitive spillovers are an equilibrium feature, not a causal response (H4).** In the

mixed-strategy equilibrium promotions are independent draws conditional on the state; observed co-movement of prices across banners arises from common cost shocks c_{it} and common calendar events (circular weeks), so lead–lag regressions of one banner's changes on another's identify a response only if a rival's change is exogenous. We have no such instrument; the model therefore tells us to treat contemporaneous and lagged associations as descriptive and to use *leads* as a placebo: a response cannot precede its cause.

Measured inflation and the basket (H5, H6). If the set of listings changes over time (entry of expensive items, exit of cheap ones), an average of posted prices moves even when every continuing item's price is unchanged; a matched-model index conditions on continuing items and recovers the price change of the fixed set. The difference is an identified object (the composition effect) that does not require quantities. The shopper's problem is a set-covering problem: choose at most K banners and an assignment of items to banners to minimise the cost of a fixed basket; the model predicts decreasing returns to K because the cheapest banner for one item is often the cheapest for others (a low κ_r shifts all prices).

What is not identified. λ_i, v_i, m and c_{irt} are not identified from posted prices alone; neither are store-level κ within a banner, since stores are unobserved—although the store-mixture oscillation of Section 4.4 reveals that within-banner store dispersion exists and bounds its size. We make no attempt to estimate these primitives.

6. Identification and preregistered methods

The hypotheses, samples, estimands, nulls, clustering and falsification tests were written down and hashed (`research/preregistration.md`) after the data audit and before any cross-banner gap, promotion statistic, index, response model or forecast was computed. The registration also lists what had been examined before freezing (table counts, coverage, the oscillation diagnostics, the distribution of current discount depth, matching validation). Three amendments were made during estimation and are logged with their rationale (`research/decisions.md`): a complete-block design for the basket problem (A-001), a unit-price screen on basket candidates (A-002), and a classification of low-price episodes into temporary sales and regular-price resets (A-003). None changes an estimand; all are reported with the preregistered version where it differs.

Common definitions. Posted price p_{irw} of listing i (banner r) in week w is the duration-weighted modal price. A *matched group* is a connected component of tier A+B pairs; group-weeks with fewer than two banners posting, or with more than one simultaneously posted listing from the same banner-channel ("ambiguous", 2.2%), are dropped. The storefront channel is primary; circulars are added

in sensitivity analyses. Standard errors are clustered by match group; a two-way group×banner clustering is reported for banner effects (with 14 banners carrying matched storefront listings the banner dimension is small, which we flag).

Model A (H1). We estimate $\log p_{irw} = \alpha_{iw} + \gamma_r + \varepsilon_{irw}$ with product-week and banner fixed effects on 67,732 listing-weeks. The γ_r are relative posted-price levels for identical products in the same week; the residual standard deviation measures dispersion not explained by banner levels. Dispersion statistics per group-week (max–min log gap, coefficient of variation, interquartile range) are summarised overall, by company relation and by category. Persistence is the pooled AR(1) coefficient of banner-pair gaps $g_{i,rs,w} = \log p_{irw} - \log p_{isw}$ on their own one-week lag and the share of pairs (≥ 4 weeks) whose sign never flips; the placebo permutes banner labels within group-week (20 replications), which must drive the AR(1) to zero if persistence is a property of banners and not of the construction. The registered economic threshold is a median gap of at least 5% and AR(1) of at least 0.5.

Model B (H2, H3). For storefront listings observed in at least 8 of 11 weeks, the reference price in week w is the modal price over the trailing eight weeks (ties resolved to the higher price); a week is *promoted* if $p_w \leq 0.90\,\bar{p}_w$. This is a posted-price version of the sales filters of Nakamura and Steinsson (2008) and Kehoe and Midrigan (2015); it is validated against the current on-sale flag in the last week (agreement 0.90). H2 compares the standard deviation of weekly log changes of the posted price with that of the reference series, overall, by banner and by category, with a listing-level bootstrap (999 draws) for the ratio. Episodes are maximal runs of promoted weeks; censored episodes (touching the panel edges) are excluded from duration and post-episode statistics. Under amendment A-003 an episode is a *return* if the first post-episode price is within $\pm 2\%$ of the pre-episode reference, a *reset down* if it is lower, a *reset up* if higher. H3 asks whether the reference after a return episode differs from the reference before it; the event study regresses $\log p$ on week-relative-to-start dummies ($-4 \ldots +4$, reference -1) with listing and week fixed effects and tests the joint nullity of the pre-period dummies; a placebo draws pseudo-episode dates in never-promoted listings. The falsification registered for the store-mixture interpretation—single-store shops must not display multi-price weeks—is reported with the H2 results, and the event study is re-estimated separately for single-store and anti-bot feeds.

Model C (H4). Store geography does not exist, so the "nearby stores" part of H4 is untestable and is not tested. For the banner-level part we regress $\Delta \log p_{irw}$ on the mean change of the other banners in the group at lags 0–3 and leads 1–2, with banner and week fixed effects, separately for all rivals, same-company rivals and other-company rivals. Contemporaneous coefficients absorb common

shocks and are not causal; lags are lead–lag associations; leads are a placebo. No instrument or quasi-experiment is available and none is used.

Model D (H5). On the grocery universe we compute, weekly, (a) the naive mean and median of posted prices, (b) a chained Jevons matched-model index (geometric mean of week-on-week ratios over listings present in both weeks), (c) fixed-basket Dutot, Carli and Jevons indices over listings present in all eleven weeks, with equal item weights, and (d) unit-price versions. No expenditure weights exist, so none of these is a consumer price index and we do not call them one. The gap between (a) and (b) is bootstrapped over listings. Because the package label is stored as current state only, within-listing size changes cannot be observed; the H5 test is therefore restricted to the composition (assortment) channel. Statistics Canada's CPI for food purchased from stores in Quebec (table 18-10-0004-01) is compared over the two months of overlap, August to September 2026, as a check of coherence, not of superiority.

Model E (H6). A 30-item basket with reference quantities was fixed before any basket cost was computed (`config/basket.yaml`). Items are mapped to listings by category and name rule; listings with a unit price outside one quarter to four times the item's cross-banner median are dropped (A-002). For each week, the item cost at a banner is the cheapest unit price times the reference quantity. Under A-001 the banner set is the B widest-coverage storefront banners and the item set the items all of them carry that week ($B = 5$ primary: Super C, Metro, Giant Tiger, Walmart, Marché Nuvo or IGA depending on the week; $B = 4, 6, 7$ in sensitivity), so that the single-banner problem is feasible without imputation. The mixed-integer program assigns each item to one banner, opens at most K banners, and minimises cost; $K = 1$ is the best single banner, $K = B$ the unconstrained optimum. Travel costs, store access, loyalty programmes and stock are unobserved; a fixed cost per banner visited of 5, 10 and 20 CAD is added as a labelled illustration. For the QGPAI trajectory a block that is constant over weeks (five banners, ten items) is used.

Model F. Next-week log posted price and next-week promotion state of storefront listings are forecast with strict rolling origins ($t = 6, \ldots, 9$; train on weeks $\leq t - 1$, test on week t), lag-only features, and four model classes: last value, banner$\times$category drift, ridge regression, gradient-boosted trees; a logistic model and a persistence rule for the promotion target. Metrics are MAE/RMSE and Brier/AUC.

Multiple testing. The six families H1–H6 each carry one or two primary tests; Holm's correction is applied across them (Table 15).

7. Empirical results

7.1. H1: Matched-item dispersion is large and persistent

Table 5 and Figure 4A report the distribution of within-product, cross-banner log price gaps on 30,042 group-weeks. The median max–min gap is 0.154 (interquartile range 0.059–0.279, mean 0.189); 79% of group-weeks exceed the registered 5% threshold and 63% exceed 10%. The median coefficient of variation is 0.073. Gaps are somewhat larger across companies (median 0.183 on 2,103 group-weeks) than within a company (0.151), but the within-company gaps—the discount banner against the conventional supermarket of the same chain on the same UPC—are the bulk of the evidence and are far from zero.

Table 5. Within-product cross-banner log price gaps (max−min), tiers A+B, storefront channel

Sample	Group-weeks	Median	P25	P75	Mean	Share ≥ 5%	Share ≥ 10%
All group-weeks	30 042	0.154	0.059	0.279	0.189	0.785	0.633
Same company (two banners)	27 939	0.151	0.058	0.268	0.186	0.782	0.627
Different companies	2 103	0.183	0.080	0.337	0.237	0.825	0.701

Weekly posted price = duration-weighted modal price. Residual s.d. of the product-week/banner two-way fixed-effects model: 0.115 (N = 67,732). AR(1) of banner-pair gaps: 0.663 (s.e. 0.014); placebo with banner labels permuted within group-week: 0.004 (s.d. 0.011). Share of banner pairs (≥ 4 weeks) whose sign never flips: 0.46.

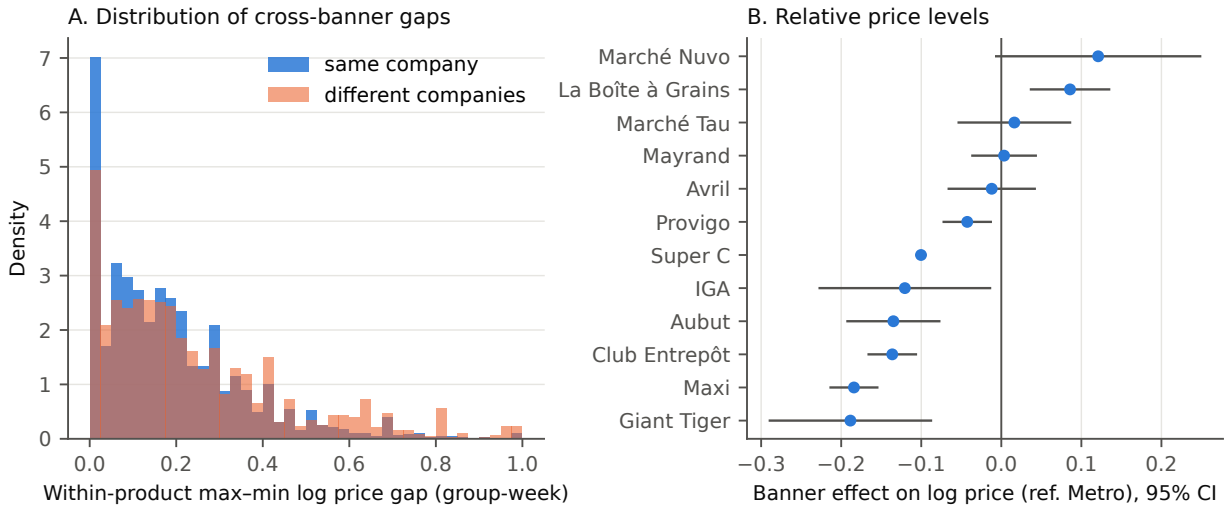


Figure 4. Cross-banner dispersion on identical products (tiers A+B, storefront). A: distribution of the group-week max–min log gap by company relation. B: banner fixed effects relative to Metro with 95% intervals (banners with at least 50 matched listing-weeks; standard errors clustered by match group).

The banner fixed effects (Table 6, Figure 4B) are the price-level architecture of the market as seen through identical products: relative to Metro, Maxi posts 18.4 log points lower (s.e. 0.016),

Club Entrepôt 13.6 lower, Super C 10.0 lower (s.e. 0.003, on 17,329 listing-weeks), Provigo 4.3 lower; Walmart and Giant Tiger are 20 and 19 points lower but on few matched listing-weeks; specialty and online shops (La Boîte à Grains, Marché Nuvo, Maturin) post 8–12 points higher. The residual standard deviation after product-week and banner effects is 0.115: knowing the banner explains the level, but a product-specific component of comparable size remains—consistent with the "relative dispersion" emphasised by Kaplan et al. (2019).

Table 6. Banner fixed effects on log posted price within product-week (reference: Metro)

Banner	Coefficient	s.e. (group)	s.e. (group+banner)	Listing-weeks
Walmart	-0.202	0.037	0.032	34
Giant Tiger	-0.189	0.052	0.039	265
Maxi	-0.184	0.016	0.027	10 844
Club Entrepôt	-0.136	0.016	0.027	9 214
Aubut	-0.135	0.030	0.023	293
IGA	-0.120	0.055	0.041	226
Super C	-0.100	0.003	0.003	17 329
Provigo	-0.043	0.016	0.027	10 405
Avril	-0.012	0.028	0.021	346
Mayrand	0.003	0.021	0.018	100
Marché Tau	0.016	0.036	0.031	58
Maturin	0.082	0.028	0.021	4
La Boîte à Grains	0.086	0.026	0.019	601
Marché Nuvo	0.121	0.066	0.054	92

Model A: $\log p_{irw} = \alpha_{iw} + \gamma_r + \varepsilon_{irw}$ on matched groups (tiers A+B), storefront channel, non-ambiguous group-weeks. Standard errors clustered by match group, and two-way by group and banner. Coefficients are relative posted-price levels for the same product in the same week; banners with few listing-weeks are imprecisely estimated.

Persistence is strong. The pooled AR(1) coefficient of banner-pair gaps is 0.663 (s.e. 0.014); permuting banner labels within group-week gives 0.004 (s.d. 0.011 over 20 permutations). The coefficient is similar within (0.668) and across (0.630) companies. Among 4,884 banner pairs observed in at least four weeks, 46% keep the same sign throughout. Part of the persistence reflects the stickiness of posted prices themselves—a gap that does not change because neither price changed—but the placebo shows that it is a property of banners, not of the construction, and the sign stability indicates that for close to half of the pairs one banner is systematically cheaper. Figure 5 displays the median pairwise gaps for banner pairs with at least 30 common group-weeks: the Loblaw discount banners (Maxi, Club Entrepôt) and Super C are cheaper than Metro and Provigo by 10–26 log points, while Mayrand and the health-food shops are dearer.

Category heterogeneity is wide (Table 14 and `robustness_category_gaps.csv`): fish and seafood (median gap 0.183, 9,577 group-weeks, dominated by Metro/Super C private-label frozen seafood), snacks (0.154) and bakery (0.136) show the largest gaps; fresh produce (0.002) and

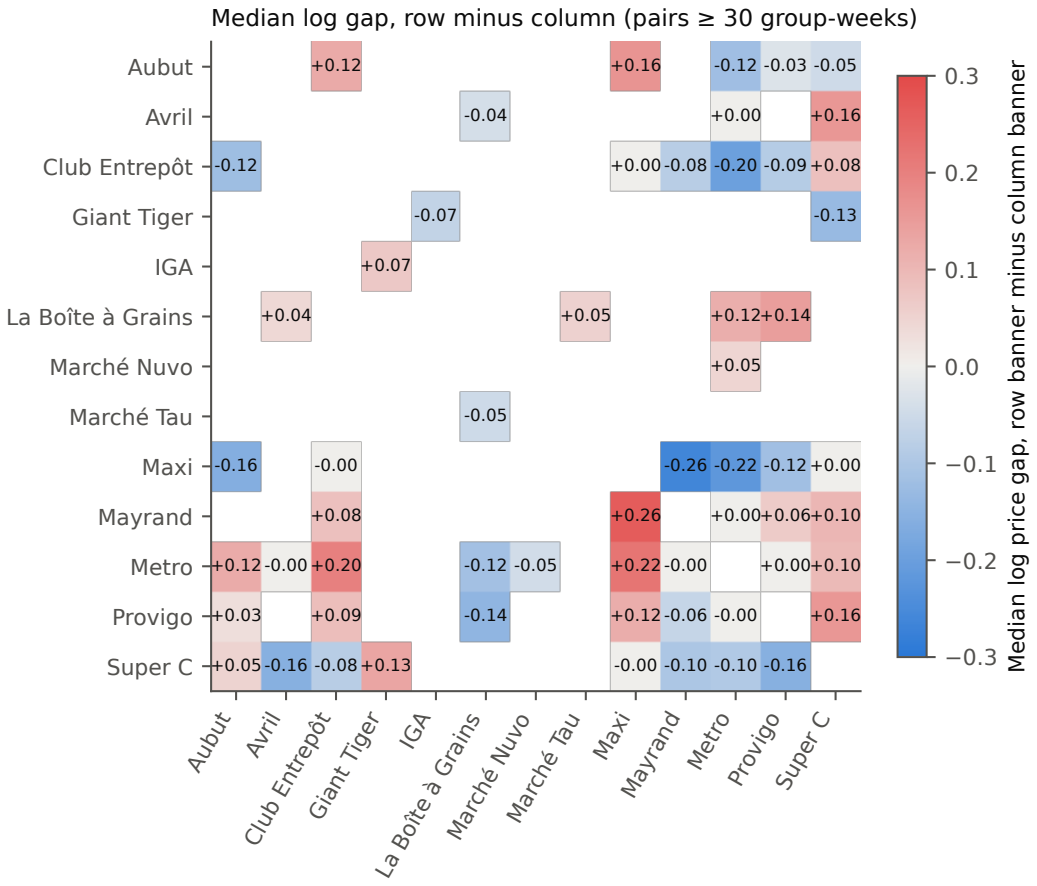


Figure 5. Pairwise median log price gaps on identical products (row banner minus column banner), tiers A+B, storefront channel, pairs with at least 30 group-weeks. Blue: row banner cheaper.

deli/cheese (0.000) show essentially none—for these categories the matched listings within a chain are priced identically across its banners. H1 is supported on both registered criteria (median gap $\geq 5\%$; AR(1) ≥ 0.5), with the qualification that most of the evidence is within-chain format differentiation.

7.2. H2: Promotions double posted-price volatility

Table 7 and Figure 6 summarise Model B on 102,835 storefront listings observed in at least eight weeks. Inferred promotions cover 4.6% of listing-weeks (4.4% in the last week, against 9.5% of active listings carrying an on-sale flag—the flag also captures small reductions and loyalty prices that our 10% rule excludes; agreement between the two is 90%). Promotions are deep: the median depth is 0.224 log points (mean 0.29). The weekly change frequency of posted prices is 8.5%, and *half* of all weekly changes are entries into or exits from a promotion. The standard deviation of weekly log changes is 0.106 for posted prices and 0.055 for the reference series, a ratio of 1.92 (bootstrap 95% CI 1.84–2.02). The ratio is 2.1–2.8 for the large chains' storefronts (Maxi 2.84, Giant Tiger 2.30, Metro 2.12, Super C 2.03) and close to one for health-food and online-only shops that rarely promote (La Boîte à Grains 1.01). The registered falsification holds: 4.7% of listing-weeks display more than one price value in single-store shops against 22.8% in anti-bot storefronts, so the weekly aggregation was necessary and the store-mixture reading of the oscillation stands. H2 is supported: promotion-adjusted (reference) volatility is about half of raw volatility.

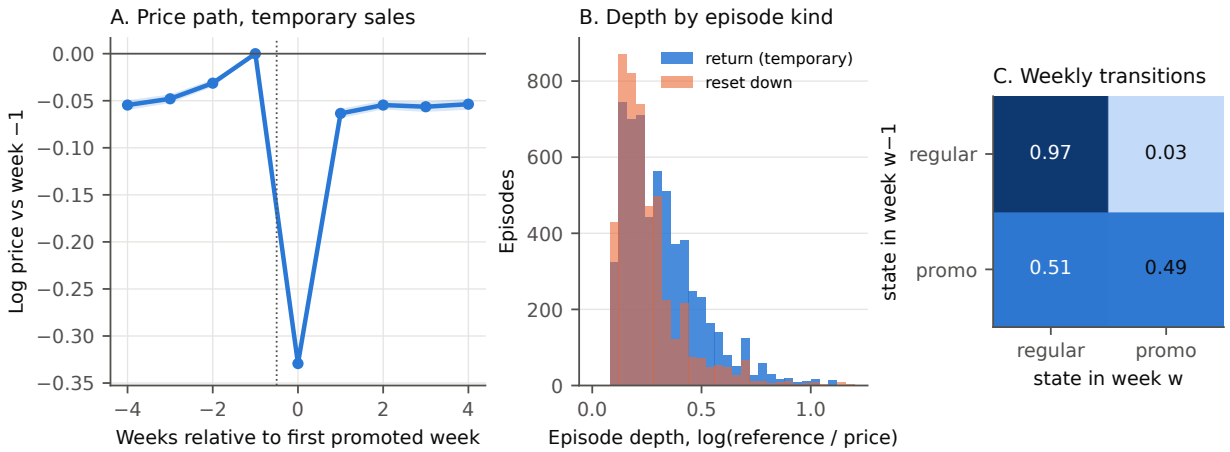


Figure 6. Promotion dynamics (Model B, storefront listings). A: event-study coefficients of $\log p$ around the first promoted week of return episodes (listing and week fixed effects, reference week -1, 95% bands clustered by listing). B: depth of uncensored episodes by kind. C: weekly transition matrix between regular and promoted states.

Table 7. Promotion dynamics in storefront feeds (Model B)

Statistic	Value
Listing-weeks promoted (inferred, $p \leq 0.9\times$ trailing 8-week mode)	0.046
Median depth (log ref/p), promoted weeks	0.224
Agreement with current on-sale flag (last week)	0.902
S.d. weekly $\Delta \log p$, raw	0.106
S.d. weekly Δlog reference	0.055
Ratio raw / reference (bootstrap 95	
Share of weekly price changes that are promotion transitions	0.504
Episodes (uncensored)	14,774 (11,705)
Episode kinds: return / reset down / reset up	0.52 / 0.42 / 0.06
Return episodes: median duration (weeks), share 1 week	1, 0.80
Return episodes: median depth	0.288
Median recurrence interval (weeks), share of listings with $\geq$ 2 episodes	3, 0.23
H3 (return episodes): mean Δlog reference after vs before (s.e.)	-0.0003 (0.0001), n = 6,049
H3 (all low-price episodes): mean Δlog reference (s.e.)	-0.1018 (0.0018)
Placebo (random weeks, non-promoted listings)	0.0003 (0.0007)
Event-study pretrend test (weeks $-4..-2$), p	0.000
Circular items with observed regular price; median observed depth	0.23; 0.224
Multi-price weeks: single-store shops vs anti-bot storefronts	0.047 vs 0.228

Sample: grocery storefront listings observed in $\geq$ 8 of 11 complete weeks. Promotion state is inferred from the weekly modal price; regular prices are not historised in the source. Return episodes end with a price within $\pm$2% of the pre-episode reference; reset episodes are regular-price changes that the 10% rule also flags.

7.3. H3: Temporary sales return exactly to their reference; no drift within the window

The transition matrix (Figure 6C) has $P(\text{promo} \to \text{promo}) = 0.49$ and $P(\text{regular} \to \text{promo}) = 0.03$. Of 11,705 uncensored low-price episodes, 52% are *return* episodes (the price comes back to within 2% of the pre-episode reference), 42% are regular-price resets downward and 6% resets upward; the 10% rule, applied to a short panel, flags many permanent price cuts, which is why amendment A-003 was needed. Return episodes—the promotions of the theory—last one week in 80% of cases (median 1, mean 1.3 weeks) with a median depth of 0.288 log points; the median interval between two episodes of the same listing is three weeks and 23% of listings have at least two episodes. For return episodes the mean change of the reference price from before to after the sale is -0.0003 log points (s.e. 0.0001; 0.5% of references move up, 0.5% move down), against a placebo of $+0.0003$ (s.e. 0.0007) at random dates in never-promoted listings. The point estimate is statistically distinguishable from zero because of the sample size, but it is economically zero: within eleven weeks, Quebec storefront promotions neither precede nor follow a reference-price adjustment. For all low-price episodes the reference falls by 10 log points on average, which is simply the reset-down class.

The event study (Figure 6A) shows the price 33 log points below the pre-week during the first promoted week and back to 5–6 points below it afterwards; the pre-period dummies are jointly non-zero ($p < 0.001$) and *negative*: the week immediately before a sale is about 5 log points higher than the weeks -4 to -2 and than the weeks after. Taken at face value this is the "pre-sale price elevation" predicted by the reference-price-inflation hypothesis. Splitting by feed type deflates it: in anti-bot storefronts the elevation of week -1 relative to weeks $-4 \ldots -2$ and $+1 \ldots +4$ is 5–9 log points, in single-store shops it is 1–3 log points. The larger part is therefore a store-mixture artefact—the week before an episode is, by construction, a week in which the modal price was the higher value of the store distribution—and the residual two-point elevation in single-store shops is a small, exploratory indication of pre-sale repricing that a longer and store-identified panel should re-examine. We register the direction found for H3 as "no reference-price drift after temporary sales; a small pre-sale elevation in clean feeds".

Circulars provide the complementary, observed view: 33,006 circular items, 22.7% with an explicit regular price, a median observed depth of 0.224 log points (mean 0.287) that matches the inferred storefront depth, and 60% of items recurring in more than one weekly circular within the window.

7.4. H4: Small lead–lag associations; cross-company placebo fails

Table 8 and Figure 7 report Model C. A banner's weekly log change is associated with the mean change of its rivals in the same group at lag 0 (0.037, s.e. 0.018), lag 1 (0.037, 0.012), lag 2 (0.015,

0.013) and lag 3 (0.043, 0.015), with leads at 0.009 (0.012) and 0.013 (0.015). The probability that a banner changes its price in a week is 25% if a rival changed its price the week before and 17% if not. These associations come almost entirely from same-company pairs (lag 1: 0.042, s.e. 0.012, on 21,192 observations), i.e. from Metro following Super C or Maxi following Provigo on the same UPC—coordination within a chain's pricing calendar rather than competitive response. For other-company rivals the sample is small (1,784 observations), the lags are insignificant or negative, and the *lead* coefficient is −0.108 (s.e. 0.048): rivals' future changes "predict" current changes, which a causal response cannot do. The placebo fails, and we draw no causal conclusion. Promotions co-move only weakly: two or more banners of a group are simultaneously at a low in 1.7% of group-weeks against 0.6% expected under independence.

Table 8. Descriptive lead–lag associations between a banner's weekly log price change and rivals' changes (Model C)

Rival set	N	$k=0$	$k=1$	$k=2$	$k=3$	lead 1	lead 2
All rivals	22 213	0.037 (0.018)	0.037 (0.012)	0.015 (0.013)	0.043 (0.015)	0.009 (0.012)	0.013 (0.015)
Same-company rivals	21 192	0.028 (0.018)	0.042 (0.012)	0.011 (0.013)	0.034 (0.015)	0.013 (0.012)	0.015 (0.014)
Other-company rivals	1 784	0.081 (0.066)	-0.001 (0.048)	-0.057 (0.029)	0.047 (0.060)	-0.108 (0.048)	-0.064 (0.049)

Dependent variable: $\Delta \log p$ of banner r for product i in week w; regressors: mean $\Delta \log p$ of the other banners in the match group at $w-k$ (lags) and $w+k$ (leads, placebo). Banner and week fixed effects; s.e. clustered by match group. Contemporaneous terms absorb common shocks and are not causal; a non-zero lead coefficient indicates timing noise or anticipation and is reported as a failed placebo.

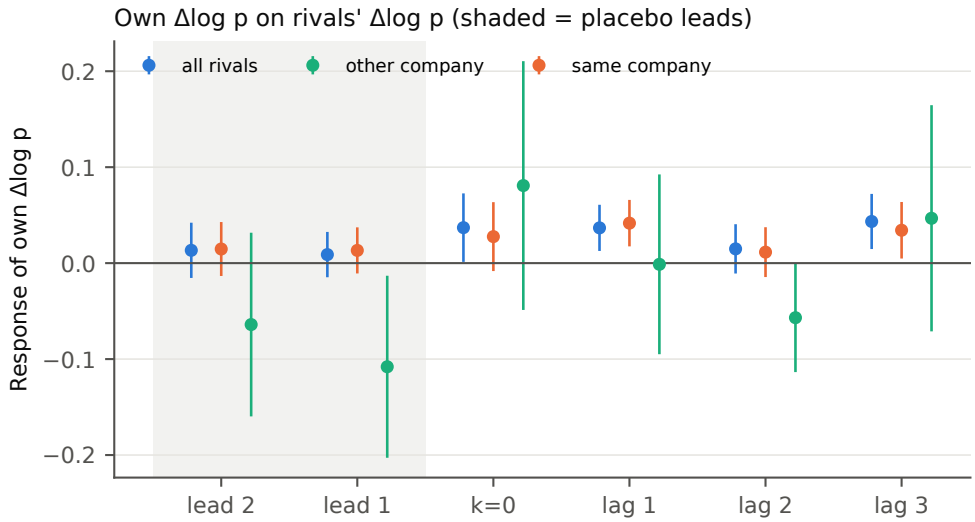


Figure 7. Lead–lag associations between a banner's weekly log price change and its rivals' (Model C), 95% intervals clustered by match group. Shaded region: leads (placebo).

7.5. H5: Composition, not price, drives naive averages

Over eleven weeks storefront posted prices rose by 0.28% on the chained Jevons matched-model index (bootstrap 95% CI 0.15–0.42%), 0.10% on the fixed-basket Jevons over the 44,423 listings present in every week, −0.05% on the fixed-basket Dutot and 2.7% on the Carli (which is upward-biased by construction); the naive mean of posted prices rose by 0.37% and the naive median not at all (Table 9, Figure 8). At the market level the registered gap between the naive mean and the matched-model index is 0.06 percentage points with a bootstrap interval of −0.78 to +0.84: H5 is *not* supported over this window for the market as a whole. It is strongly supported for individual banners whose assortment turned over: the naive mean rose 28.8% for Walmart, 14.6% for Maxi and 11.0% for Provigo while their matched-model indices moved by 0.2%, 0.9% and 0.6%; across all channels the naive mean *fell* 2.9% because circular items—cheap by construction—entered the catalogue, while the matched-model index rose 0.24%. Assortment turnover is substantial: of 71,694 storefront listings, 20,260 entered and 17,295 exited during the window (weekly entry rate 3.6%). Package-size changes within a listing are unobservable in this source. The Statistics Canada index for food purchased from stores in Quebec fell 0.67% from August to September 2026 while our storefront matched-model index rose 0.10% over the same two months; with two monthly observations, different coverage (online catalogues vs. store collection) and no weights, this is a coherence check, not a validation.

Table 9. Cumulative price change over 11 weeks (13 Aug → 22 Oct 2026) under alternative unweighted index formulas

Sample	Naive mean	Naive median	Chained Jevons	Fixed Jevons	Fixed Dutot	Fixed Carli	Balanced n
Storefront, posted price	0.004	0.000	0.003	0.001	-0.001	0.027	44 423
Storefront, unit price	-0.012	-0.007	0.003	0.001	-0.004	0.036	29 671
All channels, posted price	-0.029	0.000	0.002	0.001	-0.001	0.028	45 214

No expenditure weights exist: none of these is a consumer price index. Bootstrap (999 listing resamples) 95% CI for the storefront chained Jevons: [0.002, 0.004]; naive-mean minus Jevons gap: 0.001 [-0.008, 0.008]. Statistics Canada CPI, Quebec, food purchased from stores: Aug→Sep 2026 -0.007 (our chained Jevons, monthly means, +0.001); 12-month CPI change to Sep 2026 +0.016.

7.6. H6: Two banners capture most of the posted-price savings

Table 10 and Figure 9 report Model E. On the primary block (five banners, 15–18 items depending on the week, weekly mean cost), the best single banner costs 67.3 CAD, two banners 58.5, three 57.2 and the unconstrained optimum 56.9: unconstrained cheapest-banner shopping saves 15.4% relative to the best single banner, and two banners capture 85% of that saving. The pattern is stable across block sizes ($B = 4$: 15.1%, 82%; $B = 6$: 11.4%, 75%; $B = 7$: 12.0%, 72%). The spread between the cheapest and the most expensive single banner for the same items is 91% of the

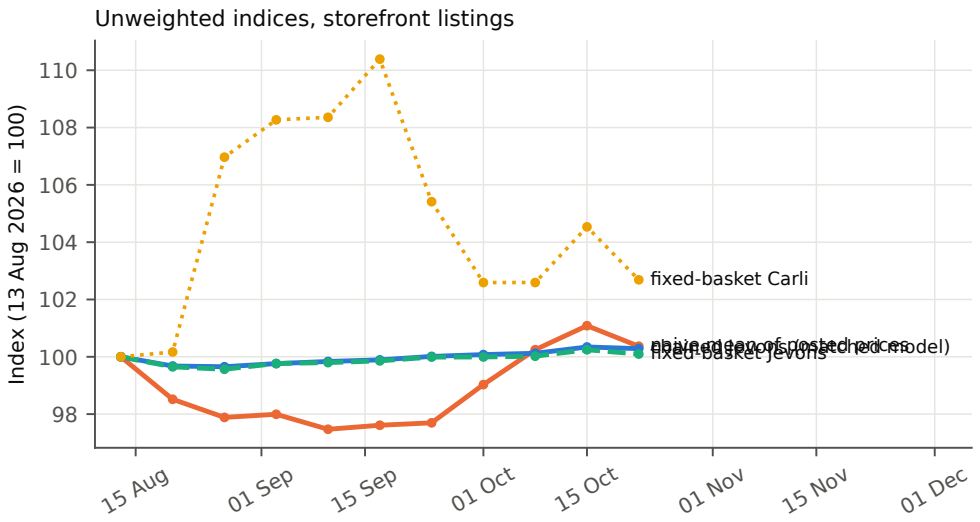


Figure 8. Unweighted price indices on storefront listings, 13 August 2026 = 100: naive mean of posted prices, chained Jevons (matched model), fixed-basket Jevons and Carli. No expenditure weights exist; these are not consumer price indices.

cheapest—banner choice matters far more than multi-banner optimisation. With an illustrative fixed cost of 5 CAD per banner visited the optimum uses 1.8 banners on average, with 10 CAD 1.3 and with 20 CAD one banner only. H6 is supported in its posted-price form; whether the savings are *feasible* depends on store access and travel costs that this source does not observe, so the 15% is an upper bound on achievable savings from banner choice within the matched catalogues, not an estimate of what a household can realise.

Table 10. Fixed-basket cost under banner-count constraints (Model E, weekly means over 11 weeks, CAD)

Banners (B)	Items	Cost K=1	Cost K=2	Cost unconstrained	Savings share	Share captured at K=2
4	18	78.72	68.97	66.86	0.15	0.82
5	15	67.28	58.50	56.92	0.15	0.85
6	11	49.27	45.05	43.66	0.11	0.75
7	9	31.01	28.33	27.28	0.12	0.72

Complete-block design (amendment A-001): the B widest-coverage storefront banners and the basket items all of them carry; item cost = cheapest unit price × reference quantity. K = 1: best single banner. Travel costs are unobserved; with an illustrative fixed cost per banner of 5 / 10 / 20 CAD the optimum uses 1.82 / 1.27 / 1.00 banners on average. Spread between the cheapest and most expensive single banner for the same items: 91% of the cheapest.

7.7. Forecasting

Table 11, Table 12 and Figure 10 report Model F on 215,264 test observations over four origins. For next-week log posted price the last value is the best model on mean absolute error (0.0226,

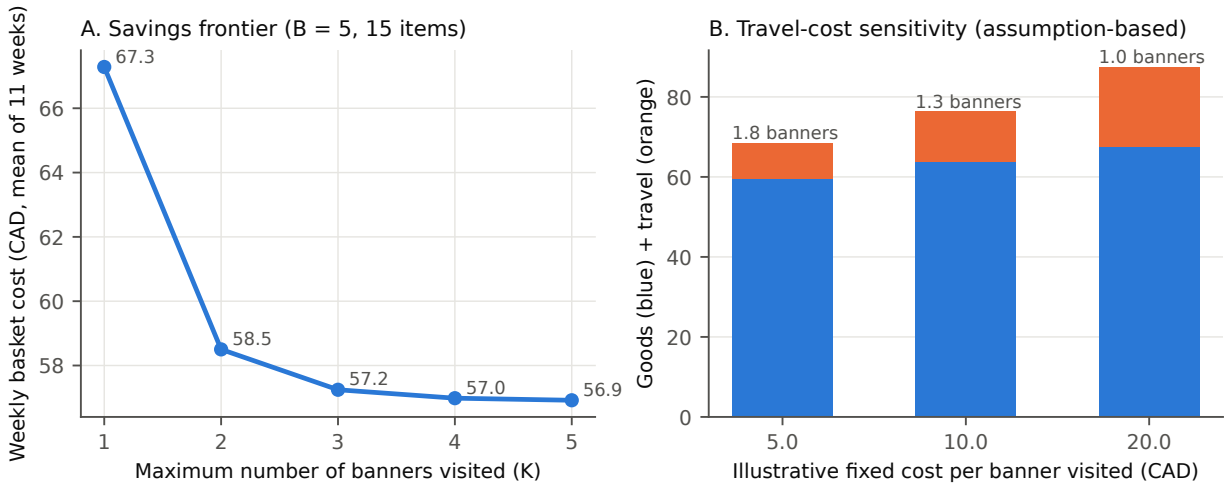


Figure 9. Fixed-basket optimisation (Model E, $B = 5$). A: mean weekly basket cost against the maximum number of banners visited. B: sensitivity to an illustrative fixed cost per banner visited (travel costs are unobserved).

against 0.0248 for banner–category drift, 0.0299 ridge and 0.0301 gradient boosting), while ridge and boosting have the lowest root mean squared error (0.099–0.100 against 0.108): the learned models trade frequent small errors for fewer large ones, which is what a mixture of sticky prices and occasional deep promotions implies. For next-week promotion state gradient boosting reaches an AUC of 0.936 and a Brier score of 0.031 against a base rate of 5.7% (Brier 0.057) and a persistence rule (AUC 0.75, Brier 0.051): promotions are predictable from a listing's own recent history (depth, recurrence, banner, category). These are honest out-of-time results on a short panel; they do not speak to horizons beyond one week.

Table 11. Rolling-origin forecasts of next-week log posted price (storefront listings)

Model	Origins	N	MAE	RMSE
banner_cat_drift	4	215 264	0.0248	0.1084
lightgbm	4	215 264	0.0301	0.1000
mean_reversion_ref	4	215 264	0.0284	0.1025
naive_last	4	215 264	0.0226	0.1083
ridge	4	215 264	0.0299	0.0992

Strict time-based holdouts: train on weeks $\leq t-1$, test on week t, origins $t = 6, \ldots, 9$; features are lags only. Lower is better.

Table 12. Rolling-origin forecasts of next-week promotion state

Model	Origins	N	Brier	AUC
base_rate	4	215 264	0.0568	0.5000
hist_rate	4	215 264	0.0572	0.6778
lightgbm	4	215 264	0.0313	0.9359
logit	4	215 264	0.0439	0.8940
persist	4	215 264	0.0510	0.7519

Promotion = modal price $\leq 0.9\times$ trailing 8-week mode. Brier: lower is better; AUC: higher is better. Base rate $\approx$ 5%.

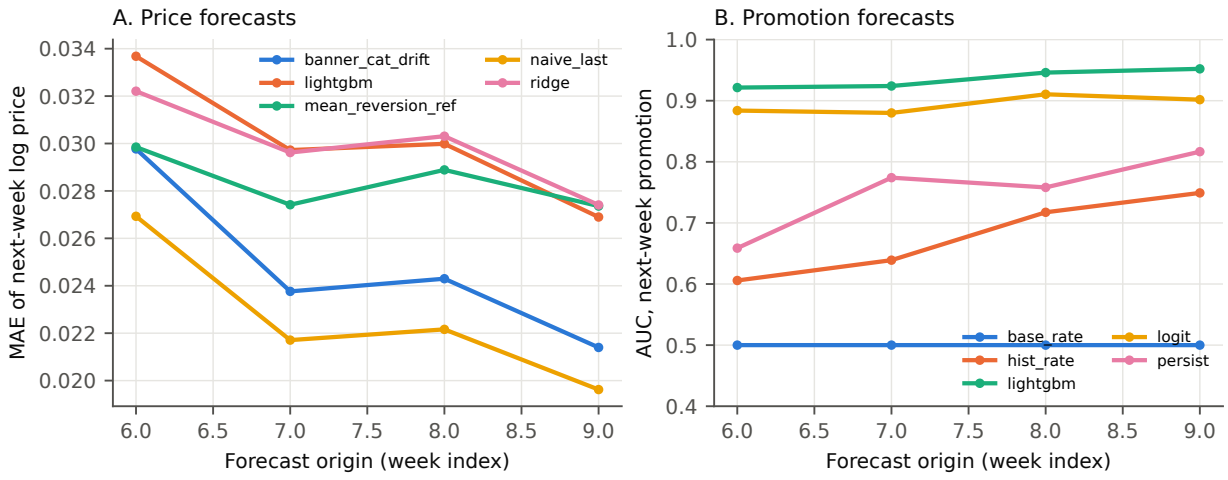


Figure 10. Rolling-origin forecast errors by origin week (Model F). A: MAE of next-week log posted price. B: AUC for next-week promotion state.

7.8. The QGPAI components

Figure 11 and Table 13 present the five components over the eleven weeks. The comparable price index on matched groups dips in late August (98.7 on 27 August) and ends at 100.7 on 22 October, with a bootstrap band over groups of about ±0.5; the promotion share rises from 3–4% in August to 6–7% in October while the observed circular depth is flat at 0.21–0.25; the dispersion component is remarkably stable at a median gap of 0.15; the fixed-block basket cost falls from 34 to 31 CAD for the best single banner and from 30 to 28 CAD unconstrained. The fifth component is not computed. Two illustrative composites with prespecified weights are in the replication package; we do not interpret them, because any weighting of these components is a normative choice, not an estimate.

Table 13. QGPAI components by circular week (Thursday start), 2026

Week	Comparable index	CI low	CI high	Promo share	Dispersion	Basket K=1	Basket K=2	Basket uncon.	Circular depth
13 Aug	100.000	100.000	100.000		0.155	34.166	31.429	29.523	0.224
20 Aug	99.654	99.234	100.089	0.043	0.154	32.404	29.735	28.951	0.224
27 Aug	98.670	98.188	99.187	0.029	0.154	34.867	31.067	29.568	0.246
03 Sep	99.498	99.049	99.905	0.044	0.154	33.089	29.240	27.933	0.230
10 Sep	100.065	99.544	100.498	0.042	0.154	31.877	29.613	28.120	0.237
17 Sep	100.246	99.676	100.758	0.046	0.153	32.166	29.616	28.281	0.224
24 Sep	100.367	99.784	100.857	0.052	0.153	32.322	28.410	27.176	0.224
01 Oct	100.802	100.244	101.322	0.058	0.151	31.091	27.831	26.973	0.209
08 Oct	100.813	100.207	101.325	0.068	0.151	31.165	27.637	26.687	0.224
15 Oct	101.308	100.789	101.829	0.064	0.141	31.165	27.108	27.108	0.237
22 Oct	100.688	100.106	101.212	0.054	0.151	31.165	27.907	27.907	0.229

Component 5 (access and affordability) is not computed: no store geography or income data are joined. Comparable index: chained Jevons over matched groups (A+B), base 13 Aug = 100, bootstrap band over groups; Dispersion: median within-group max−min log gap; Basket: Model E (B = 5). Composites are illustrative only and reported in the replication package.

8. Robustness, falsification and failed tests

Table 14 collects the registered robustness checks on the H1 statistic. The median gap is 0.163 on the balanced panel of 1,300 groups present in all eleven weeks, 0.154 with unit-price normalisation, 0.151 after trimming the 1st and 99th percentiles (mean 0.181 against 0.189 untrimmed), 0.183 on cross-company group-weeks only, 0.135 when only non-promoted weeks are kept and 0.237 in group-weeks where at least one banner is promoted; across matching tiers (A, A+B, A+B+C, A+B+C+D) and posted-price definitions (mode, minimum, maximum, time-weighted mean) it ranges from 0.138 to 0.154, and adding circular listings leaves it at 0.154. Dispersion is therefore not an artefact of a tier, a price definition, outliers or promotions, although promotions widen it.

The category-level sign tests that the median gap exceeds 5% reject for eight of eleven categories after Holm correction and do not reject for fresh produce, deli/cheese and ready meals. Table 15 applies Holm's correction across the primary tests: H1 (both criteria), H2, the H3 reference-change

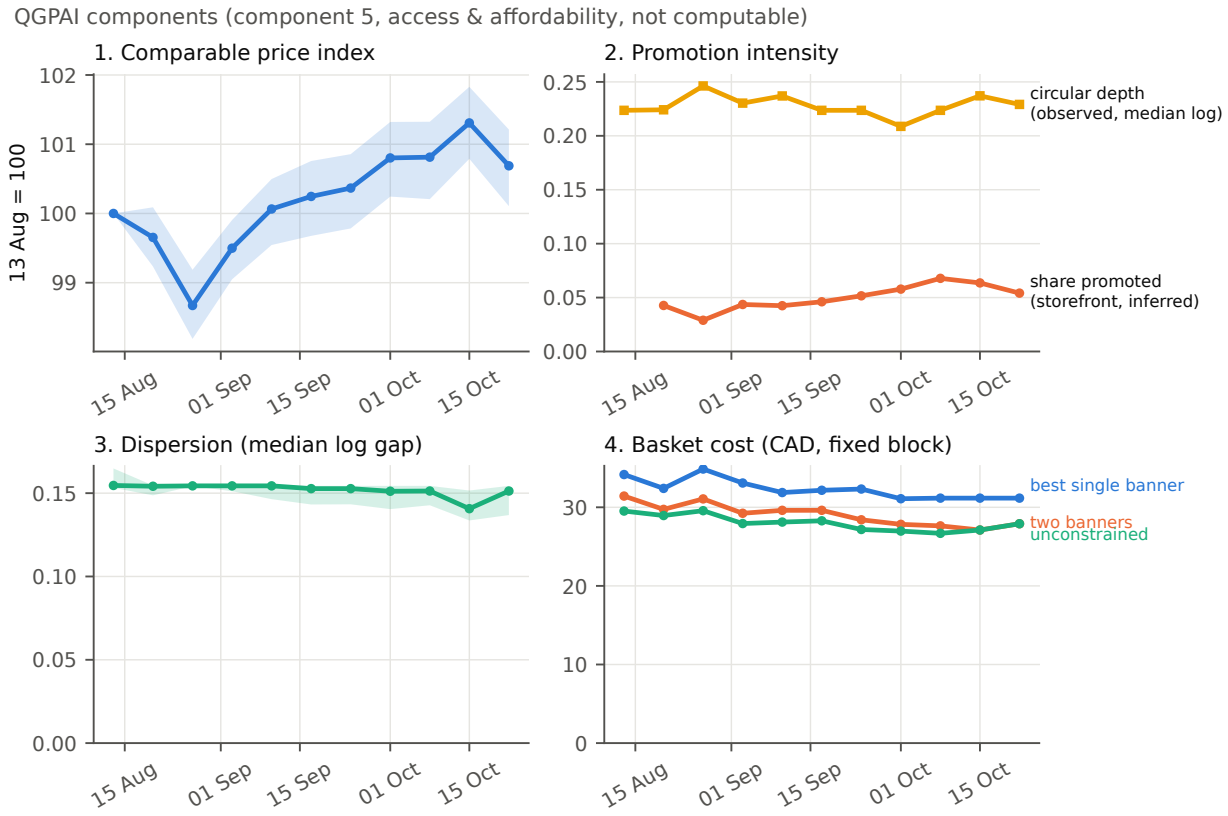


Figure 11. QGPAI components by circular week, 2026. Bands: bootstrap over match groups (components 1 and 3). Component 2 is undefined in the first week (no trailing history). Component 5 (access and affordability) is not computable from this source.

Table 14. Robustness of the dispersion statistic (H1)

Specification	Median log gap
Baseline median gap (A+B, storefront)	0.154
Balanced groups present in all 11 weeks (n groups 1300)	0.163
Unit-price normalisation	0.154
Trimmed 1/99	
Cross-company group-weeks only	0.183
Non-promoted weeks only	0.135
Group-weeks with a promoted listing	0.237
Tiers A, price = p_mode	0.144
Tiers A, price = p_min	0.143
Tiers A, price = p_max	0.140
Tiers A, price = p_twmean	0.138
Tiers AB, price = p_mode	0.154
Tiers AB, price = p_min	0.154
Tiers AB, price = p_max	0.145
Tiers AB, price = p_twmean	0.145
Tiers ABC, price = p_mode	0.154
Tiers ABC, price = p_min	0.154
Tiers ABC, price = p_max	0.147
Tiers ABC, price = p_twmean	0.145
Tiers ABCD, price = p_mode	0.154
Tiers ABCD, price = p_min	0.154
Tiers ABCD, price = p_max	0.149
Tiers ABCD, price = p_twmean	0.146
Tiers AB, price = p_mode+circulars	0.154

All rows use storefront listings in non-ambiguous group-weeks with ≥ 2 banners.

tests and the H4 lag-1 association remain significant; the market-level H5 gap does not.

Table 15. Preregistered primary tests with Holm correction

Preregistered test	p	Holm-adjusted p
H1_median_gap_gt_5pct	0.0000	0.0000
H1_AR1_gt_0.5	0.0000	0.0000
H2_ratio_ne_1	0.0000	0.0000
H3_all_low_price_episodes_dlog_ref_ne_0	0.0000	0.0000
H4_lag1_ne_0	0.0028	0.0083
H3_return_episodes_dlog_ref_ne_0	0.0091	0.0183
H5_naive_vs_jevons_gap_ne_0	1.0000	1.0000

CI-based entries take p = 0 when the bootstrap 95% CI excludes the null value and 1 otherwise (conservative coding).

Falsification tests that passed. The AR(1) placebo (permuted banner labels) returns 0.004; the H3 placebo at random dates returns $+0.0003$; single-store shops show 4.7% multi-price weeks against 22.8% in anti-bot feeds; the H4 leads for all-rival and same-company rivals are indistinguishable from zero.

Tests that failed or were inconclusive. (i) The cross-company lead placebo in Model C fails (-0.108, s.e. 0.048): with 1,784 observations and common weekly shocks, the descriptive lag coefficients for other-company rivals cannot be read as responses. (ii) The event-study pretrend test rejects; the decomposition by feed type attributes most of the pre-sale elevation to store-mixture noise, leaving a small exploratory residual. (iii) The market-level H5 gap is not different from zero over eleven weeks; the hypothesis holds only at the banner level. (iv) Tier C matching failed validation (precision 0.42) and was removed from the primary specification; a first version of the matcher based on token-set similarity was rejected outright. (v) The preregistered basket rule left only seven items common to all wide-coverage banners and had to be amended (A-001). (vi) The comparison with Statistics Canada rests on two monthly observations and is inconclusive. We regard reporting these as part of the result.

Out-of-time validation. Model F's rolling-origin design is the only strictly out-of-time evidence in the paper; the promotion-state forecasts (AUC 0.94) are the component of the architecture that generalises forward within the window.

9. Policy and practical implications

The results are descriptive and the window is short, so implications are stated as what the evidence supports and what it would take to go further.

For consumers and comparison services. On identical products the choice of banner moves the posted price by 10–25 log points, and a shopper restricted to two banners captures most of the attainable posted-price saving. A comparison service that reports the *weekly modal* price, flags multi-price weeks and labels same-company pairs would convey this architecture more faithfully than one that reports the last scraped price.

For statistical agencies and researchers using web-scraped prices. Store-selected storefronts are a mixture of store prices unless the collector pins the store. The Billion-Prices-style assumption that one online price represents the retailer does not hold for Quebec's large chains; high-frequency "changes" in such feeds overstate price flexibility by an order of magnitude, and the correct unit of observation is a store-identified session or, failing that, a weekly modal price with the range reported.

For competition policy. The like-for-like evidence is overwhelmingly *within* company: the Loblaw and Metro discount banners price 10–18 log points below their conventional banners on the same UPCs. This is consistent with deliberate format segmentation inside concentrated groups, which may be efficient (serving heterogeneous consumers) or may be a substitute for cross-chain competition; posted prices cannot tell these apart. Cross-chain exactly matched comparisons are rare in this source because chains differentiate assortments and private labels, which is itself a finding about how competition operates.

For the measurement of food inflation. Over eleven weeks the matched-model storefront index and the naive average agreed at the market level, but naive averages for individual banners were off by 10–30 points because of assortment turnover. Any index built on catalogue data must be a matched-model index; any claim about "shrinkflation" requires storing package labels over time, which this source did not do and which is a one-line change in the collection design.

10. Limitations

1. **Posted, not paid, prices.** We observe catalogue and circular prices, not transactions, quantities or expenditure weights. Nothing here is a welfare measure or a consumer price index.
2. **Eleven weeks.** The window is too short for the registered minimum durations of H3 (26 weeks) and of the index validation against official statistics (52 weeks); seasonal patterns are invisible; the promotion recurrence interval is right-censored.
3. **No store identifiers.** Storefront prices are mixtures over unknown stores; circulars are for one

Montréal postal code; the "nearby stores" part of H4, the travel-cost part of H6 and QGPAI component 5 cannot be computed. The prospective design in Section 12 addresses this.

4. **Inferred promotion states.** Regular prices are not historised for storefronts, so promotion states rest on a 10%-below-trailing-mode rule validated only against the current on-sale flag (agreement 0.90) and the circular depth distribution. Small reductions and loyalty prices are classified as regular.
5. **Matching coverage and error.** Exactly matched products are 11% of listings and 93% of matched group-weeks are within-company pairs. Tier B carries an estimated 7% rate of hidden-variant false matches on UPC ground truth; the reviewed sample is small (60 pairs per tier) and its final labels were set by the research operator rather than by independent human coders; PI adjudication is pending.
6. **Assortment coverage.** The connectors of the large chains read a subset of aisles; Giant Tiger, Epipresto, La Boîte à Grains and Nuvo are read exhaustively. Market-level statistics are therefore not representative of the grocery basket's expenditure shares.
7. **Package changes unobserved.** Size labels are stored as current state; within-listing downsizing cannot be detected.
8. **Weak external benchmark.** Two monthly CPI observations overlap the window.
9. **Analytic amendments.** Three post-registration amendments (A-001 to A-003) were required; each is logged with the preregistered alternative.

11. Conclusion

Taking a price-comparison service's database as it is—posted prices, online channels, eleven weeks, no stores—we have tried to extract what it identifies and to say clearly what it does not. The architecture that emerges has four load-bearing elements. Banners differ by 10–25 log points on identical products, persistently, and mostly inside the two chains whose discount and conventional formats carry the same UPCs. Promotions are rare in catalogues but deep, recur every few weeks, account for half of all posted-price changes, double measured volatility, and return exactly to their reference price within the window. Cross-banner price co-movement is small and, where it is visible, is the calendar of a chain rather than a response to a rival. Measured "inflation" on catalogue data is driven by what enters and leaves the catalogue unless the index is matched-model. Underneath these sits a measurement fact that any user of scraped grocery prices should check for: a storefront that serves store-specific prices to anonymous sessions produces an oscillating series whose amplitude is cross-store dispersion, not price change.

A longer, store-identified, regular-price-logging collection—three changes to an existing pipeline—would turn several of this paper's descriptive statements into tests: whether reference prices drift

after sales over a year, whether the two-point pre-sale elevation in clean feeds survives, how dispersion across stores of the same banner compares with dispersion across banners, and what a Quebec household at a given address can actually save. We have set the index, the matching procedure and the registration up so that those tests can be run on the same code.

12. Reproducibility, data access and prospective design

Code and outputs. The pipeline (Python 3.12; pinned dependencies) runs from the raw snapshot to the compiled paper with `make`: ingestion to Parquet with checksum verification, unit normalisation, retailer hierarchy, daily and weekly panel reconstruction (DuckDB), tiered matching with UPC calibration and LLM second annotation, Models A–F, QGPAI, robustness, LaTeX tables and vector figures. Every stochastic step uses seed 20261101. Each stage writes a manifest with SHA-256 hashes of its outputs and a line in a structured run ledger; model scripts write JSON summaries from which all numbers in the text and tables are generated. Fifteen tests (unit parsing, panel invariants, matching rules, output consistency) pass. The preregistration and basket were hashed before estimation; the decision log records three amendments. The LLM second annotation (claude-opus-5-5, four runs, 1.58 USD, model and tokens logged) is auxiliary: its labels are stored, so the pipeline runs without API access.

Data access. The raw snapshot is a read-only export of a database owned by the author and is not redistributed in the repository; it can be made available to referees under a data-use agreement, together with the SHA-256 of the file used here. Food-Ka's `robots.txt` disallows automated access to its API; the public API was queried once to cross-check headline counts and is not a data source. Statistics Canada series were obtained from the public Web Data Service (tables 18-10-0004-01 and 18-10-0245-01; raw JSON kept with download timestamps).

Prospective collection design (not implemented here; no production change was made). (i) Store selection: pin a fixed panel of about twenty stores across Quebec regions in the Loblaw, Metro and Walmart connectors and pull Flipp circulars for at least six postal codes (Montréal, Québec, Gatineau, Sherbrooke, Saguenay, Trois-Rivières), one pass per store per day. (ii) Regular-price logging: add regular price and on-sale flag to the price log. (iii) Package-label logging: keep a history of size labels to detect downsizing. (iv) Duration: 26 weeks for promotion recurrence and reference-price tests, 52 weeks for seasonality and validation against official indices. The authorisation is the owner's; the changes are reversible and polite.

Use of AI assistance. The pipeline, the analyses and the drafting of this manuscript were produced by the author with an AI research assistant (Claude Code) operating under the written mandate

reproduced in the repository; all design decisions, amendments and the validation review are documented and the author is responsible for the content. No coauthor or institutional approval is implied.

References

Adams, Brian and Kevin R. Williams, "Zone Pricing in Retail Oligopoly," *American Economic Journal: Microeconomics*, 2019, *11* (1), 124–156.

Aguiar, Mark and Erik Hurst, "Life-Cycle Prices and Production," *American Economic Review*, 2007, *97* (5), 1533–1559.

Allcott, Hunt, Rebecca Diamond, Jean-Pierre Dubé, Jessie Handbury, Ilya Rahkovsky, and Molly Schnell, "Food Deserts and the Causes of Nutritional Inequality*," *The Quarterly Journal of Economics*, 2019, *134* (4), 1793–1844.

Anderson, Eric, Benjamin A. Malin, Emi Nakamura, Duncan Simester, and Jón Steinsson, "Informational rigidities and the stickiness of temporary Sales," *Journal of Monetary Economics*, 2017, *90*, 64–83.

_ , Nir Jaimovich, and Duncan Simester, "Price Stickiness: Empirical Evidence of the Menu Cost Channel," *Review of Economics and Statistics*, 2015, *97* (4), 813–826.

Argente, David and Munseob Lee, "Cost of Living Inequality During the Great Recession," *Journal of the European Economic Association*, 2020, *19* (2), 913–952.

Atkin, David, Benjamin Faber, and Marco Gonzalez-Navarro, "Retail Globalization and Household Welfare: Evidence from Mexico," *Journal of Political Economy*, 2018, *126* (1), 1–73.

Baye, Michael R., John Morgan, and Patrick Scholten, "Chapter 6 Information, Search, and Price Dispersion," *Handbooks in Information Systems Economics and Information Systems*, 2006, pp. 323–375.

Beck, Günter W. and Xavier Jaravel, "Prices and Global Inequality: New Evidence from Worldwide Scanner Data," *SSRN Electronic Journal*, 2020.

Bils, Mark and Peter J. Klenow, "Some Evidence on the Importance of Sticky Prices," *Journal of Political Economy*, 2004, *112* (5), 947–985.

Bronnenberg, Bart, Jean-Pierre Dubé, Matthew Gentzkow, and Jesse Shapiro, "Do Pharmacists Buy Bayer? Informed Shoppers and the Brand Premium," Technical Report, Competition Bureau Canada 2014.

Burdett, Kenneth and Kenneth L. Judd, "Equilibrium Price Dispersion," *Econometrica*, 1983, *51* (4), 955.

Cavallo, Alberto, "Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers," Technical Report, Competition Bureau Canada 2016.

_ , "Scraped Data and Sticky Prices," *The Review of Economics and Statistics*, 2018, *100* (1), 105–119.

_ **and Oleksiy Kryvtsov**, "What can stockouts tell us about inflation? Evidence from online micro data," *Journal of International Economics*, 2023, *146*, 103769.

_ **and Roberto Rigobon**, "The Billion Prices Project: Using Online Prices for Measurement and Research," Technical Report, Competition Bureau Canada 2016.

Chevalier, Judith A. and Anil K Kashyap, "Best Prices: Price Discrimination and Consumer Substitution," *American Economic Journal: Economic Policy*, 2019, *11* (1), 126–159.

Chevalier, Judith A, Anil K Kashyap, and Peter E Rossi, "Why Don't Prices Rise During Periods of Peak Demand? Evidence from Scanner Data," *American Economic Review*, 2003, *93* (1), 15–37.

Christen, Peter, *Data Matching*, Springer Berlin Heidelberg, 2012.

Coibion, Olivier, Yuriy Gorodnichenko, and Gee Hee Hong, "The Cyclicality of Sales, Regular and Effective Prices: Business Cycle and Policy Implications," *American Economic Review*, 2015, *105* (3), 993–1029.

Competition Bureau Canada, "Canada Needs More Grocery Competition: Retail Grocery Market Study Report," Technical Report, Competition Bureau Canada 2023.

de Haan, Jan and Heymerik A. van der Grient, "Eliminating chain drift in price indexes based on scanner data," *Journal of Econometrics*, 2011, *161* (1), 36–46.

DellaVigna, Stefano and Matthew Gentzkow, "Uniform Pricing in U.S. Retail Chains*," *The Quarterly Journal of Economics*, 2019, *134* (4), 2011–2084.

Diewert, W. Erwin and Kevin Fox, "Measuring Real Consumption and CPI Bias under Lockdown Conditions," Technical Report, Competition Bureau Canada 2020.

_ **and Kevin J. Fox**, "Substitution Bias in Multilateral Methods for CPI Construction," *Journal of Business & Economic Statistics*, 2020, *40* (1), 355–369.

Dubé, Jean-Pierre, Günter J. Hitsch, and Peter E. Rossi, "Income and Wealth Effects on Private-Label Demand: Evidence from the Great Recession," *Marketing Science*, 2018, *37* (1), 22–53.

Eichenbaum, Martin, Nir Jaimovich, and Sergio Rebelo, "Reference Prices, Costs, and Nominal Rigidities," *American Economic Review*, 2011, *101* (1), 234–262.

Einav, Liran, Ephraim Leibtag, and Aviv Nevo, "Recording discrepancies in Nielsen Homescan data: Are they present and do they matter?," *Quantitative Marketing and Economics*, 2009, *8* (2), 207–239.

Ellickson, Paul B. and Sanjog Misra, "Supermarket Pricing Strategies," *Marketing Science*, 2008, *27* (5), 811–828.

Feenstra, Robert C., "New Product Varieties and the Measurement of International Prices," *American Economic Review*, 1994, *84* (1), 157–177. https://www.jstor.org/stable/2117971.

Fellegi, Ivan P. and Alan B. Sunter, "A Theory for Record Linkage," *Journal of the American Statistical Association*, 1969, *64* (328), 1183–1210.

Gagnon, Etienne and David López-Salido, "Small Price Responses to Large Demand Shocks," *Journal of the European Economic Association*, 2019, *18* (2), 792–828.

Golosov, Mikhail and Robert E. Lucas Jr., "Menu Costs and Phillips Curves," *Journal of Political Economy*, 2007, *115* (2), 171–199.

Gorodnichenko, Yuriy and Oleksandr Talavera, "Price Setting in Online Markets: Basic Facts, International Comparisons, and Cross-Border Integration," *American Economic Review*, 2017, *107* (1), 249–282.

Gourville, John T. and Jonathan J. Koehler, "Downsizing Price Increases: A Greater Sensitivity to Price than Quantity in Consumer Markets," 2004. Working paper, SSRN.

Griffith, Rachel, Ephraim Leibtag, Andrew Leicester, and Aviv Nevo, "Consumer Shopping Behavior: How Much Do Consumers Save?," *Journal of Economic Perspectives*, 2009, *23* (2), 99–120.

Guimaraes, Bernardo and Kevin D Sheedy, "Sales and Monetary Policy," *American Economic Review*, 2011, *101* (2), 844–876.

Handbury, J. and D. E. Weinstein, "Goods Prices and Availability in Cities," *The Review of Economic Studies*, 2014, *82* (1), 258–296.

Hausman, Jerry and Ephraim Leibtag, "Consumer Benefits from Increased Competition in Shopping Outlets: Measuring the Effect of Wal-Mart," Technical Report, Competition Bureau Canada 2006.

Hendel, Igal and Aviv Nevo, "Measuring the Implications of Sales and Consumer Inventory Behavior," Technical Report, Competition Bureau Canada 2005.

_ and _ , "Intertemporal Price Discrimination in Storable Goods Markets," *American Economic Review*, 2013, *103* (7), 2722–2751.

Hitsch, Günter, Ali Hortaçsu, and Xiliang Lin, "Prices and Promotions in U.S. Retail Markets: Evidence from Big Data," Technical Report, Competition Bureau Canada 2019.

Hitsch, Günter J., Ali Hortaçsu, and Xiliang Lin, "Prices and promotions in U.S. retail markets," *Quantitative Marketing and Economics*, 2021, *19* (3-4), 289–368.

Hoch, Stephen J., Xavier Drèze, and Mary E. Purk, "EDLP, Hi-Lo, and Margin Arithmetic," *Journal of Marketing*, 1994, *58* (4), 16–27.

Hosken, Daniel and David Reiffen, "Patterns of Retail Price Variation," *The RAND Journal of Economics*, 2004, *35* (1), 128.

Hottman, Colin J., Stephen J. Redding, and David E. Weinstein, "Quantifying the Sources of Firm Heterogeneity *," *The Quarterly Journal of Economics*, 2016, *131* (3), 1291–1364.

Ivancic, Lorraine, W. Erwin Diewert, and Kevin J. Fox, "Scanner data, time aggregation and the construction of price indexes," *Journal of Econometrics*, 2011, *161* (1), 24–35.

Jaravel, Xavier, "The Unequal Gains from Product Innovations: Evidence from the U.S. Retail Sector*," *The Quarterly Journal of Economics*, 2018, *134* (2), 715–783.

Kaplan, Greg and Guido Menzio, "The Morphology of Price Dispersion," Technical Report, Competition Bureau Canada 2014.

_ and _ , "Shopping Externalities and Self-Fulfilling Unemployment Fluctuations," *Journal of Political Economy*, 2016, *124* (3), 771–825.

_ , _ , Leena Rudanko, and Nicholas Trachter, "Relative Price Dispersion: Evidence and Theory," *American Economic Journal: Microeconomics*, 2019, *11* (3), 68–124.

Kehoe, Patrick and Virgiliu Midrigan, "Prices are sticky after all," *Journal of Monetary Economics*, 2015, *75*, 35–53.

Klenow, Peter J. and Oleksiy Kryvtsov, "State-Dependent or Time-Dependent Pricing: Does It Matter for Recent U.S. Inflation? [*]," *Quarterly Journal of Economics*, 2008, *123* (3), 863–904.

Lach, Saul, "Existence and Persistence of Price Dispersion: an Empirical Analysis," Technical Report, Competition Bureau Canada 2002.

Lal, Rajiv and Ram Rao, "Supermarket Competition: The Case of Every Day Low Pricing," *Marketing Science*, 1997, *16* (1), 60–80.

MacDonald, James M., "Demand, Information, and Competition: Why Do Food Prices Fall at Seasonal Demand Peaks?," *The Journal of Industrial Economics*, 2000, *48* (1), 27–45.

Melmiès, Jordan, "Hidden price increases, product downsizing, and shrinkflation: a post-Keynesian macroeconomic perspective," *Industrial and Corporate Change*, 2024, *34* (2), 318–341.

Melser, Daniel, "Scanner Data Price Indexes: Addressing Some Unresolved Issues," *Journal of Business & Economic Statistics*, 2017, *36* (3), 516–522.

Nakamura, Emi and Jón Steinsson, "Five Facts about Prices: A Reevaluation of Menu Cost Models [*]," *Quarterly Journal of Economics*, 2008, *123* (4), 1415–1464.

_ **and** _ , "Price Rigidity: Microeconomic Evidence and Macroeconomic Implications," Technical Report, Competition Bureau Canada 2013.

nan, "Menu Costs, Multiproduct Firms, and Aggregate Fluctuations," *Econometrica*, 2011, *79* (4), 1139–1180.

Nevo, Aviv and Arlene Wong, "THE ELASTICITY OF SUBSTITUTION BETWEEN TIME AND MARKET GOODS: EVIDENCE FROM THE GREAT RECESSION," *International Economic Review*, 2018, *60* (1), 25–51.

Pesendorfer, Martin, "Retail Sales: A Study of Pricing Behavior in Supermarkets," *The Journal of Business*, 2002, *75* (1), 33–66.

Redding, Stephen J and David E Weinstein, "Measuring Aggregate Price Indices with Taste Shocks: Theory and Evidence for CES Preferences*," *The Quarterly Journal of Economics*, 2019, *135* (1), 503–560.

Salop, Steven and Joseph Stiglitz, "Bargains and Ripoffs: A Model of Monopolistically Competitive Price Dispersion," *The Review of Economic Studies*, 1977, *44* (3), 493.

Sorensen, Alan T., "Equilibrium Price Dispersion in Retail Markets for Prescription Drugs," *Journal of Political Economy*, 2000, *108* (4), 833–850.

Stigler, George J., "The Economics of Information," *Journal of Political Economy*, 1961, *69* (3), 213–225.

Varian, Hal R., "A Model of Sales," *American Economic Review*, 1980, *70* (4), 651–659. https://www.jstor.org/stable/1805363.

Çakır, Metin, "Retail pass-through of package downsizing," *Agribusiness*, 2021, *38* (2), 259–278.

_ **and Joseph V. Balagtas**, "Consumer Response to Package Downsizing: Evidence from the Chicago Ice Cream Market," *Journal of Retailing*, 2014, *90* (1), 1–12.

A. Supplementary material

A.1. Matching rules in detail

Cleaned name key: strip accents, lower-case, remove size tokens (a quantity, an optional multiplier and a unit among kg, g, mg, lb, oz, ml, cl, l, un, ea, pk, ch), punctuation, stop words (de, du, des, la, le, les, et, à, pour, the, of, and, with, un, une, sans, avec, ou, or) and the brand's own tokens; append the residual of the package label (e.g. "calibre 21-25") after removing deposit and unit-price fragments; sort tokens. Numeric tokens must be identical. Package key: total quantity in base unit rounded to 0.1 and multipack count. Tier B: identical token sets; tier C: rapidfuzz `token_sort_ratio` ≥ 0.90; tier D: identical full-name token sets with brand tokens present. The calibration table over scorers and thresholds, the recall analysis against UPC truth, the three validation samples (the rejected token-set version, round 1 and round 2) and the LLM prompts are in the replication package.

A.2. Preregistered hypotheses (abridged)

H1: median within-group max–min log gap ≥ 0.05 and AR(1) of banner-pair gaps ≥ 0.5, placebo by label permutation. H2: ratio of raw to reference weekly volatility $\neq 1$, bootstrap over listings; falsification on single-store shops. H3: mean change of the reference price after promotion episodes $= 0$ (null), direction left to the data; event study with pretrend test, censoring and overlap rules; placebo at random dates. H4: distributed lags 0–3 with leads as placebo; geography untestable. H5: naive mean vs. matched-model gap $= 0$ (null), bootstrap; StatCan coherence check. H6: $S(\infty) - S(2) = 0$ (null) on a 30-item basket fixed before computation. Full text, estimands, clustering and missingness rules: `research/preregistration.md` (SHA-256 in `research/decisions.md`).

A.3. Additional tables

Per-feed coverage for all 54 feeds, category-level gaps with Holm-corrected sign tests, volatility ratios by banner and category, index changes by banner and category, basket item coverage,

block sensitivity, the travel-cost grid, and per-origin forecast errors are provided as CSV files in `data/processed/` and as JSON summaries in `experiments/`.