The Hidden Transaction Structure of Price Distributions

Inferring Unit-Sales Shares from Retail Price Frequencies

Iván Franco
Triplethree International
Working paper | October 2026

Abstract

Price dispersion records how frequently stores charge each price, not how many units sell at those prices. We test whether the observed distribution of selling stores reveals the hidden distribution of realized unit sales. Across 167,216 UPC-week systems in Dominick's scanner data, median overlap between store-price and unit-sales shares is 94.97%; in 38,553 later systems, raw frequencies recover 95.40% overlap and an independently reconstructed boosting correction reaches 95.70%. A separate retailer check using Walmart M5 data yields 91.35% median overlap on 115,931 product-weeks when conditioning on selling stores, and 88.51% on 176,704 product-weeks when including all stores with a recorded weekly price. This second data source has only ten stores and weekly rather than transaction-receipt prices. The results establish a measurable correspondence, not a universal 95% accuracy rate: the mapping weakens under heterogeneous sales intensities, promotions, and more complex price distributions.

Keywords: price dispersion; realized transaction distributions; unit-sales shares; seller-price frequency; missing quantities; scanner data
JEL classifications: C81, D40, L81

1. Introduction

A retail price distribution records how frequently sellers charge each observed price. It ordinarily does not record how many units consumers actually buy at each price. This distinction raises a fundamental question: when sales quantities are unavailable, can the observable distribution of selling stores across prices reveal the hidden distribution of realized transactions? The question concerns the allocation of observed market activity across coexisting prices, not inflation measurement or the structural demand curve. Price dispersion is a natural setting because identical goods sell simultaneously at several prices (Stigler, 1961; Kaplan and Menzio, 2015). Our central proposition is: Price dispersion is not merely a distribution of prices. Under identifiable conditions, it also reveals how realized market transactions are distributed across those prices.

Price dispersion has a systematic empirical morphology (Kaplan and Menzio, 2015). The present paper asks whether that morphology contains information about realized unit sales. Specifically, we compare the share of selling stores charging each exact price of a fixed UPC in a fixed week with the share of units sold at that price. This within-UPC-week objective differs from the across-product expenditure-weight imputation of Antoniades, Feenstra, and Xu (2022) and builds on evidence relating within-product prices and quantities in Eden, Eden, and Yuen (2016).

That missing quantity dimension is the problem addressed here. A region can be prominent in a price distribution because many sellers operate there, yet it need not account for a comparable share of transactions. Stores charging less frequent prices may sell much larger quantities, promotions may shift disproportionate volume toward particular price points, and store heterogeneity may separate seller concentration from transaction concentration. The theoretical interpretation of price-system morphology therefore requires an empirical bridge between the distribution of sellers across prices and the distribution of realized quantities across those same prices.

This paper asks whether that bridge exists. When quantity weights are unavailable, how closely does the observed organization of sellers across prices reveal where transactions actually occurred? For the same product observed across many stores in the same week, we distinguish two distributions over the same price support. The first records the share of transacting stores at each price. The second records the share of units actually sold at each price. The first is observable from the price system; the second is the realized transaction distribution that the price system is hypothesized to reveal.

Scanner data allow the proposition to be tested directly. Across 167,216 UPC-week systems in cereals, laundry detergents, paper towels, shampoo, and canned tuna, the two distributions are strikingly close. The most frequent store price is also a transaction mode in 94.5% of systems. Median total variation distance between the complete distributions is 5.03%, equivalent to 94.97% median distributional overlap. Thus the high-density regions visible in price morphology generally coincide with the regions in which realized transactions are concentrated.

A close match does not by itself prove an equilibrium mechanism, and the paper does not interpret it that way. A simple accounting identity shows that transaction shares depend on both the number of stores at each price and average sales per store at that price. If sales per store are similar across price points, store frequency will mechanically be a strong predictor of transaction frequency. We isolate this component with a quantity-permutation benchmark and then study the economically informative departures from it. Those departures become larger when sales intensity differs sharply across price points, particularly under localized promotions and greater store-level heterogeneity.

The empirical test therefore does two things. First, it asks whether the observable price-frequency distribution maps onto the realized distribution of unit sales across exact price points. Second, it identifies when and why that correspondence weakens. These tests address a transaction-data validation problem left open by the posted-price Price Systems framework (Franco, 2026); they do not require that framework's conditional market-clearing interpretation to hold.

The analysis draws on three literatures. Search-based models and descriptive work establish the existence and morphology of price dispersion (Stigler, 1961; Varian, 1980; Burdett and Judd, 1983; Sorensen, 2000; Kaplan and Menzio, 2015). Scanner-data studies observe within-product prices and quantities and examine price-bin

elasticities, sales and promotion heterogeneity (Campbell and Eden, 2014; Eden, Eden, and Yuen, 2016; Eden, 2018; Hitsch, Hortaçsu, and Lin, 2021). Work on missing weights shows that retail seller presence can inform economic imputation across products (Antoniades, Feenstra, and Xu, 2022; Ayoubkhani and Thomas, 2022). Our contribution concerns the exact-price distribution of unit-sales shares within a fixed product-week when those shares are unobserved.

1.1. Eden, Eden, and Yuen (2016): precedent and incremental contribution

Eden, Eden, and Yuen (2016) are the closest within-product scanner-data antecedent. Using IRI records, they allocate stores in each UPC-week to two, three or five price-ranked bins, and study average prices and quantities across bins, their variation over time, temporary sales, own- and cross-bin price-quantity elasticities, and the probability that lower-price stores attract shoppers. Thus, they establish that prices and unit sales can be analyzed jointly within a UPC-week and that price rank, promotions, and store sales intensity matter. Their sample also explicitly excludes UPC-store records lacking continuously positive revenue, in part because those records do not distinguish an out-of-stock item from an item on the shelf with zero sales. We acknowledge these substantive precedents rather than claiming to introduce within-UPC-week price-and-quantity comparisons (Eden, Eden, and Yuen, 2016, Sections 1, 4, and 9).

Our paper uses that within-UPC-week comparison as a starting point but asks a different inverse question. We do not estimate bin elasticities or treat price-bin quantity differences as a new finding. Instead, we retain the full observed support of exact prices and construct two normalized distributions on it: the fraction of selling stores at each price, $sP(p)$, and the fraction of units sold there, $sQ(p)$. We quantify their full-distribution proximity with total-variation distance and modal correspondence, characterize failure under heterogeneous sales intensity and promotions, and then hide the unit-sale weights in later weeks to test whether they can be inferred from seller-price frequencies. Neither the bin-level elasticity estimates nor the shopper-attraction probabilities in Eden, Eden, and Yuen (2016) are inputs to our calculations. The claimed addition is evaluated distributional recovery of missing unit-sales shares, not priority over the study of price and sales quantities.

1.2. Relationship to the earlier Price Systems working paper

Franco (2026) develops the Price Systems representation of posted-price morphology using a Mexican price-monitoring panel and tests whether its structure is persistent and predictive. It does not observe quantities sold. The present analysis uses a different dataset, Dominick's store-level scanner sales, to test the additional price-to-realized-quantity link directly. The earlier paper provides motivation and a conceptual vocabulary, but no result in this paper depends on a unique market-clearing price, on unverified elasticities, or on the acceptance of the earlier paper's equilibrium hypothesis. The prior working paper is identified in the References and supplied separately for reviewer convenience; this manuscript remains self-contained.

The paper has four specific contributions. First, it measures how closely the full distribution of selling-store frequencies approximates the unobserved within-UPC-week distribution of unit sales at each exact price. Second, it identifies when the proxy fails, particularly under localized promotions and price-related differences in sales per store. Third, it tests quantity-share reconstruction on later weeks withheld from model fitting, with a directly observable frequency baseline and transparent and nonlinear corrections. Fourth, it illustrates implications for estimating a unit-weighted average price; that application is ancillary rather than the paper's scientific premise. These are descriptive and predictive findings, not identified structural demand or equilibrium effects.

We make no claim that a price distribution identifies a Marshallian demand curve or a market-clearing price. Transaction shares are an observational measure of realized allocations, not a counterfactual response function. The focus is the potential for price-only information, conditional on observing positive-sales store-price cells, to approximate missing quantity weights and reduce economic measurement errors.

We first define the store-to-transaction mapping, document its empirical accuracy and failure cases, evaluate prediction in a later-week holdout, and examine both economic price measurement and stronger out-of-sample benchmarks.

2. Price-Quantity Mapping and the Realized Transaction Distribution

Consider one product observed across many stores during the same period. For a price point p, let sP(p) be the share of transacting stores charging p, and let sQ(p) be the share of total units sold at p. The first measure requires only prices. The second requires quantities.

$$sP(p) = \text{transacting stores charging } p \,/\, \text{total transacting stores}$$

$$sQ(p) = \text{units sold at } p \,/\, \text{total units sold}$$

Let qbar(p) be the average number of units sold by stores charging p. Then the two distributions are linked exactly by:

$$sQ(p) = sP(p)\, qbar(p) \,/\, \Sigma\, [sP(k)\, qbar(k)]$$

The equation contains the main intuition of the paper. Suppose 60% of stores charge $10, 30% charge $11, and 10% charge $12. If stores at those three prices sell broadly similar quantities, then approximately 60% of units will also be sold at $10, 30% at $11, and 10% at $12. The price-frequency distribution will closely reproduce the transaction distribution.

$$\text{If } qbar(p) \text{ is broadly similar across price points, then } sP(p) \approx sQ(p).$$

Average sales do not have to be identical across prices. Stores can sell different quantities and the two distributions can still look very similar. The mapping breaks only when those differences become large enough to overcome the store-frequency pattern. For example, a price used by only 10% of stores can dominate transactions if those stores sell many times more units than the rest.

This gives us two useful terms. The extensive margin is the share of stores at each price. The intensive margin is how much those stores sell on average. The terms simply name the two parts of the mechanism above.

2.1 Demand schedules and realized transaction distributions

The relation $sP(p) \approx sQ(p)$ does not mean that the price-frequency distribution is a demand curve. A Marshallian demand schedule is a behavioral and counterfactual object: it describes how quantity demanded would change across alternative prices, holding other relevant conditions fixed. The realized transaction distribution is observational: it records how the units actually sold were distributed across the prices that coexisted across stores and market states.

$$\text{Structural demand: } Q_D = D(p \mid x)$$

$$\text{Realized transaction structure: } sQ(p) = \text{share of observed units transacted at } p$$

These are different objects and answer different questions. Structural demand is the appropriate object for elasticities, counterfactual pricing, and welfare analysis. The realized transaction distribution is more direct evidence for a different question: where did market activity actually occur? The paper does not try to recover D(p) from prices. It tests whether the observed distribution of stores across prices can reveal that realized transaction distribution.

2.2 Directly comparing two empirical distributions

For a given UPC-week, the store-frequency distribution and unit-sales distribution are defined on the same observed price support. Plotting them together highlights both the predictive correspondence and economically informative deviations. No rotation or hypothetical supply-and-demand curves are required to visualize this result.

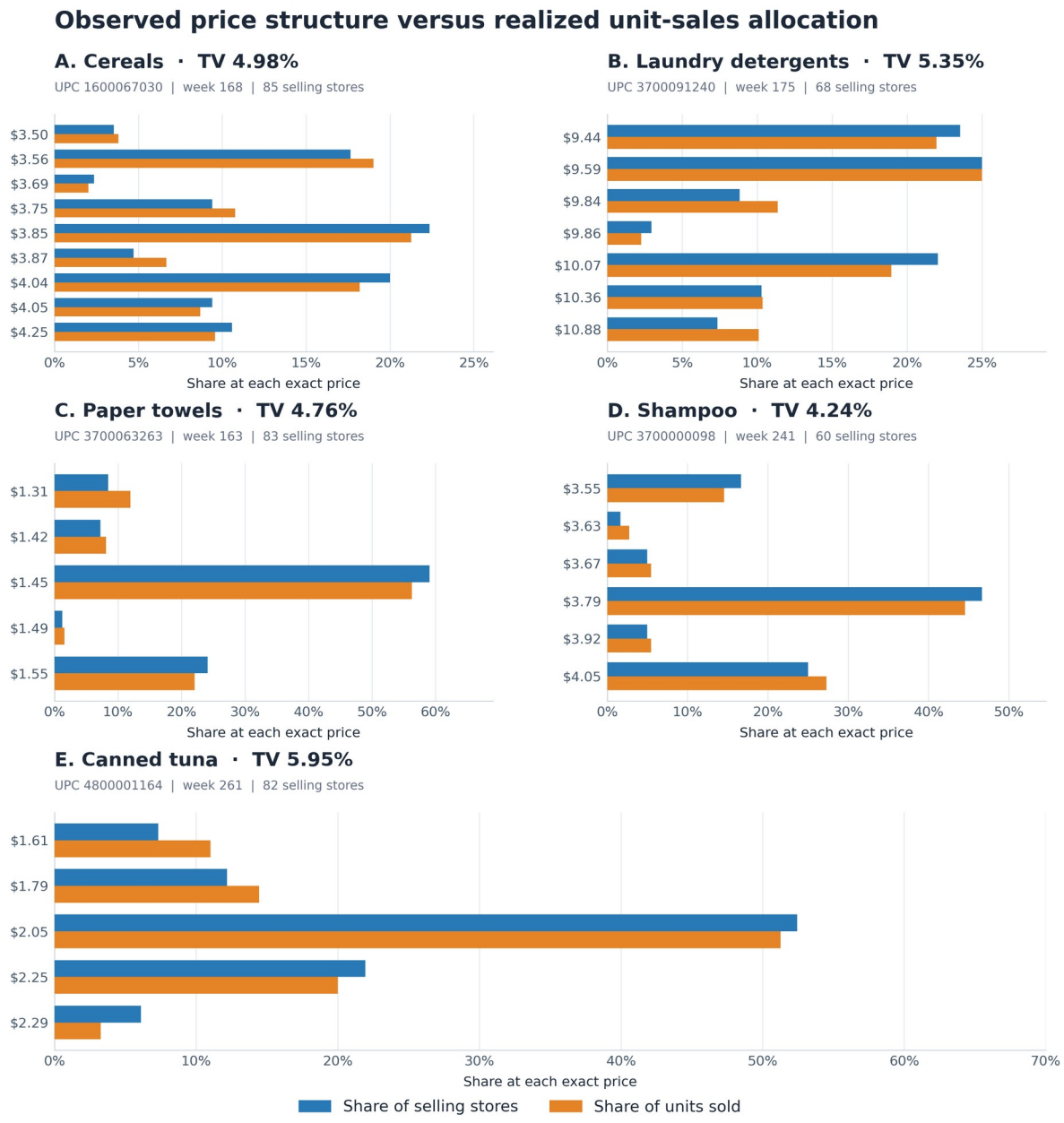


Figure 1. Selling-store shares (blue) versus unit-sales shares (orange) at each exact price in five categories. Illustrative systems are selected objectively: among systems with at least 60 active stores, 5–9 price points, and a modal-store share of 20–70%, choose the UPC-week closest to its category-wide median TV distance (ties by UPC then week). All systems, including cases that diverge substantially, appear in Figure 2. Source: author's calculations from Dominick's Finer Foods scanner data.

The five panels display typical distributional distances from every category, rather than one favorable product. Despite different levels and concentrations of exact prices, store and unit shares broadly track each other. They illustrate realized allocation—not a structural demand schedule or a unique market-clearing price. The full, unselected sample is summarized in Figure 2.

2.3 Transaction modes and the limits of economic interpretation

Let pF denote a modal store price - a price at which sP(p) is highest - and let pT denote a modal transaction price - a price at which sQ(p) is highest. When the two modal sets overlap, an observable price mode is also a price at which the largest share of units is sold. We call this the focal transaction price.

The realized transaction distribution is not a structural demand curve, and its mode cannot be interpreted as a clearing price without separately identified supply, demand, and price-selection conditions. Rather than adding such assumptions, this paper directly evaluates an economic measurement implication that can be identified in scanner data: approximation of missing transaction weights and unit-weighted average prices.

Any structural equilibrium interpretation belongs to a separate theoretical exercise. The claims tested here concern realized transaction allocation conditional on known positive-sales store-price support, not counterfactual demand, consumer surplus, or welfare.

3. Data and Empirical Design

The empirical analysis uses the Dominick's Finer Foods scanner dataset distributed by the Kilts Center for Marketing at the University of Chicago Booth School of Business (Kilts Center for Marketing, n.d.). The data report weekly store-level sales for individual UPCs, including price, units sold, bundle quantity, gross margin, sale indicators, and a data-quality flag. We analyze five categories: cereals, laundry detergents, paper towels, shampoo, and canned tuna. The categories were chosen before the cross-category results were calculated so that the sample would span food, household, and personal-care products.

The unit of analysis is a UPC-week price system. Within one UPC and week, stores provide the cross-sectional dimension. The design therefore compares prices and quantities for the same product at the same time, rather than pooling different products or periods.

The movement records are cleaned using OK=1, positive units (MOVE), a positive recorded price, and positive bundle quantity. Bundle offers are converted to a unit-equivalent price as PRICE/QTY, rounded to cents. Baseline analyses retain UPC-week systems with at least 20 positive-sales stores and at least two distinct prices. Records with zero units usually report PRICE=0 rather than a separately observed shelf price; they therefore cannot supply a valid posted-price comparison without additional shelf-price information.

These rules produce 167,216 UPC-week systems, representing approximately 11.23 million active store observations and 194.1 million units sold. Weeks range up to week 399 in the selected categories; shampoo enters later in the panel.

For every system, we construct sP(p), the share of transacting stores at each price, and sQ(p), the share of units sold at each price. We evaluate the mapping in two ways. First, modal correspondence records whether the set of price modes and the set of transaction modes share at least one price point. Treating modal ties as sets avoids arbitrary results caused by choosing one tied mode. Second, total variation distance measures the difference between the full distributions:

$$TV(sP, sQ) = 1/2 \Sigma |sP(p) - sQ(p)|$$

TV=0 means the two distributions are identical. TV=0.05 means that 5% of probability mass would need to be reallocated to make them identical.

We explicitly condition on observed stores recording positive sales for the UPC-week. Although transactions are hidden from the prediction model, the identity of positive-sales stores is known in the scanner file. Thus, these exercises cannot establish identical performance on an unrestricted panel of shelf prices at stores with zero sales. Appendix A.6 provides high-coverage sensitivity checks; this is a boundary on inference, not a claim that zero-sales stores are absent from retail markets.

The sale indicator also deserves caution. The data manual identifies sale codes, but the absence of a sale code does not guarantee that no promotion occurred. Promotion results should therefore be read as a useful but imperfect classification of promotional intensity.

4. Results

4.1 Price frequency closely tracks transaction allocation

The first result is straightforward: the share of stores at each price contains substantial information about the share of units sold at those same prices. Table 1 reports the results by category.

Table 1. Price-frequency and transaction-share correspondence by category.

Category	UPC-week systems	Modal correspondence	Median TV distance
Cereals	57,964	94.19%	4.98%
Laundry detergents	38,694	93.95%	5.35%
Paper towels	9,181	95.01%	4.76%
Shampoo	38,237	97.00%	4.30%
Canned tuna	23,140	91.97%	5.95%
All categories	167,216	94.51%	5.03%

Across all categories, at least one modal store price is also a modal transaction price in 94.5% of systems. The result goes beyond the mode. Median TV distance is 5.03%, equivalent to 94.97% median distributional overlap. The typical full price-frequency distribution therefore reproduces most of the actual distribution of units across prices.

A simple benchmark helps show how strong this is. Suppose an analyst knew which price points existed but gave each one equal transaction weight. Median TV distance would then be 33.21%. Using the actual share of stores at each price lowers median TV distance to 5.03% and beats the equal-weight benchmark in 97.56% of systems. Across all price points pooled together, store share and transaction share have a correlation of about 0.986.

Put simply, knowing how stores are distributed across prices gives a strong first estimate of how transactions are distributed across those prices.

Figure 2 uses the full analysis population rather than selected examples. Each line gives the cumulative distribution of total-variation errors within one category; the thick line pools all 167,216 UPC-week systems. Thus the typical correspondence and the material tail of failure cases are both directly visible.

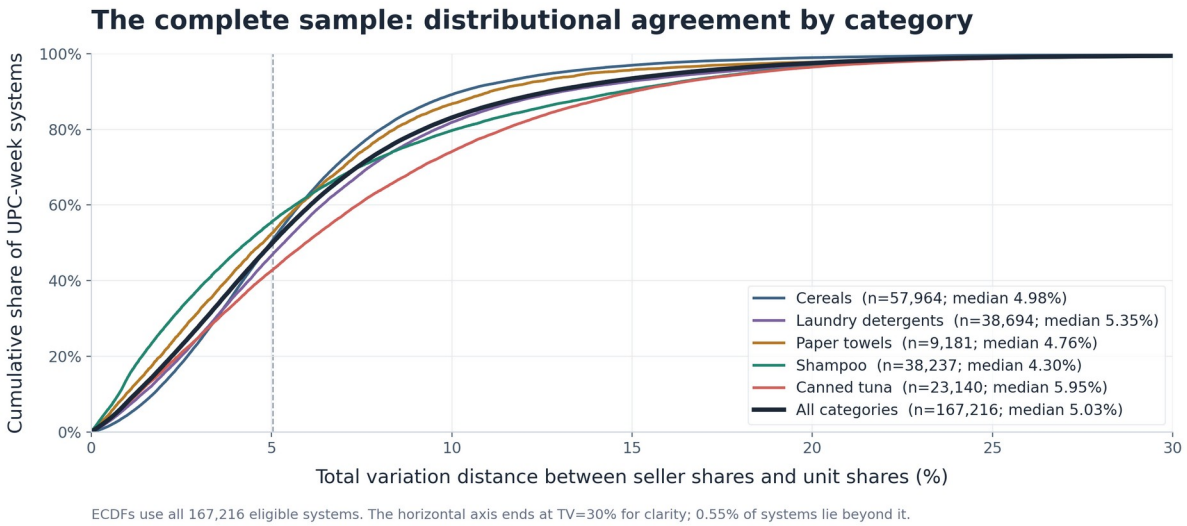


Figure 2. Empirical cumulative distributions of total-variation distance between selling-store shares and realized unit-sales shares, by category and overall. Every one of the 167,216 eligible UPC-week systems is included; category sample sizes and medians are displayed in the legend. The horizontal axis is truncated at TV = 30% for readability (0.55% of systems exceed this value). Source: author's calculations from Dominick's Finer Foods scanner data.

4.2 Part of the correspondence is mechanical

An extensive-margin baseline can be informative without representing an independently identified pricing mechanism. By definition, transaction weights equal the share of stores at a given price multiplied by their relative sales intensity. This identity should be treated as the null benchmark, not as evidence that the focal price is an equilibrium price.

We test this directly with a permutation benchmark. Within each sampled UPC-week system, we keep the observed prices, price frequencies, and store-level quantities, but randomly reassign the quantities across stores. This removes the systematic link between a store's price and its sales while keeping the same prices, store counts, and quantities. The exercise uses 1,500 systems - 300 per category - and 50 random reallocations per system.

The correspondence survives the randomization. In this sample, actual median TV distance is 5.04%, while the median randomized benchmark is even smaller at 3.20%. Actual modal correspondence is 94.8%; average randomized modal correspondence is 96.0%. In 69.4% of sampled systems, actual TV distance is larger than that system's randomized median.

Randomly reallocating quantities across stores produces closer alignment than in the observed assignments, confirming that much of the raw store-to-unit correspondence comes from the extensive margin. This test does not invalidate the practical imputation result. It clarifies its scope: the economic contribution must be established by measured performance when quantity weights are unavailable and by characterizing where heterogeneous sales break the approximation.

Transaction allocation = price-frequency baseline + intensive-margin correction

The remaining economic information appears in the departures from that baseline. Demand response, local market conditions, promotions, inventories, and store heterogeneity determine whether a price point sells more or fewer units than its store frequency alone would predict.

4.3 When does the mapping weaken?

The price-quantity identity predicts a clear pattern: the mapping should become less accurate when average sales per store differ sharply across price points. The data show exactly that. Table 2 divides systems into four groups based on how much average sales per store vary across prices, measured by the coefficient of variation of qbar(p).

Table 2. Mapping accuracy as sales intensity becomes more heterogeneous across price points.

Sales-intensity heterogeneity	Systems	Modal correspondence	Median TV distance
Q1 - lowest	41,804	99.26%	1.58%
Q2	41,804	97.20%	4.09%
Q3	41,804	94.08%	6.64%
Q4 - highest	41,804	87.50%	11.54%

The pattern is monotonic. When stores at different prices sell similar quantities, the two distributions are almost identical. In the lowest-heterogeneity quartile, modal correspondence is 99.3% and median TV distance is only 1.58%, equivalent to 98.42% median overlap. In the highest-heterogeneity quartile, modal correspondence falls to 87.5% and median TV rises to 11.54%, still equivalent to 88.46% median overlap.

This is exactly what the identity predicts. Store frequency works best when sales per store are broadly similar across prices and becomes less accurate as differences in sales intensity grow.

The result also survives substantial price dispersion. In 25,884 systems where no single price represents more than 40% of transacting stores, the median system has eight distinct price points. Even in this harder subset, modal correspondence is 72.6% and median TV distance is 7.66%. Focal pricing makes the mapping stronger, but the result is not simply an artifact of nearly uniform chain prices.

4.4 Localized promotions provide a clear failure case

Promotions provide a clear example of when the mapping can fail. A discounted price may appear in only a few stores but generate unusually high sales in those stores. Its transaction share can then be much larger than its store share.

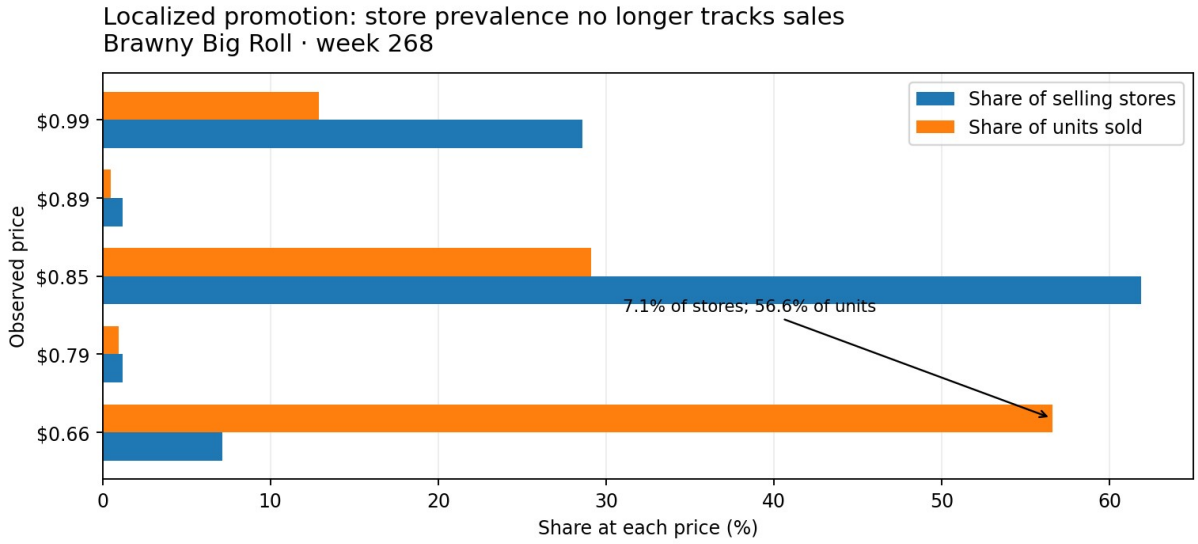


Figure 3. A localized promotion creates a systematic imputation failure: in Brawny Big Roll, week 268, $0.66 accounts for 7.14% of positive-sales stores but 56.59% of units. A discrete grouped bar chart shows the two observed distributions without connecting unrelated price points. Source: author's calculations from Dominick's data.

The aggregate data show the same pattern. With no flagged promotion, median TV distance is 4.72%. When flagged promotions appear in 10-50% of stores, median TV rises to 9.87%. When the promotion becomes widespread and covers more than half of transacting stores, median TV falls back to 6.10%. Once a promotion is widespread, it also appears strongly in the store-frequency distribution, so that distribution again captures more of the transaction shift.

The practical failure condition is simple: the mapping is weakest when a price used by relatively few stores generates unusually high sales per store.

5. Recovering Hidden Transaction Distributions from Prices Alone

5.1 Out-of-sample design

The preceding results show that price-frequency and transaction-share distributions are closely related when both are observed. The more demanding test is predictive: if quantities are not observed, can the price system recover where transactions are concentrated? We answer this with a time-based holdout experiment. Weeks 1-300 are used for estimation. All later weeks are held out. The test sample contains 38,553 UPC-week systems: 14,653 cereals, 9,596 laundry detergents, 2,234 paper towels, 7,309 shampoo, and 4,761 canned tuna. Quantities from these holdout systems are never used to generate their predictions.

The simplest prediction uses the store share at each price directly as the predicted transaction share. We then estimate a nonlinear correction from the earlier weeks using price-system information only: store share, position relative to the modal and mean price, price rank, focalization, number of stores and price points, entropy, relative dispersion, and the mass above and below the mode. To make the results intuitive, we report both total variation (TV) distance and its complement, distributional overlap. Overlap equals 1 - TV distance, or equivalently the probability mass shared by the predicted and actual distributions. A value of 100% means that the two distributions are identical. The morphology adjustment below is an independent reimplementation based on the documented specification rather than a recovery of the original historical training program; the model and its exact features are supplied in the associated replication files.

Table 3. Out-of-sample recovery of hidden transaction distributions using price data only.

Category	Holdout systems	Baseline TV	Baseline overlap	Corrected TV	Corrected overlap	Error reduction
Cereals	14,653	4.16%	95.84%	4.05%	95.95%	2.6%
Laundry detergents	9,596	4.93%	95.07%	4.71%	95.29%	4.5%
Paper towels	2,234	3.85%	96.15%	3.37%	96.63%	12.6%
Shampoo	7,309	4.68%	95.32%	3.66%	96.34%	21.7%
Canned tuna	4,761	6.16%	93.84%	5.69%	94.31%	7.6%
All categories	38,553	4.60%	95.40%	4.30%	95.70%	6.5%

5.2 How much of the hidden transaction distribution is recovered?

The raw selling-store-frequency baseline closely approximates the hidden distribution of unit sales. Across 38,553 later holdout systems, median TV distance is 4.60%, or 95.40% median distributional overlap. An independently trained LightGBM model using price morphology features reduces median TV to 4.30%, increasing median overlap to 95.70%. Mean TV declines from 6.50% to 5.74%, and the correction improves upon the raw baseline in 63.1% of holdout systems. These numbers come from a newly available independent implementation of the documented procedure, not the original historical training script. The incremental gain is modest in absolute terms because the uncorrected price distribution is a strong predictor. The high baseline performance itself is a central result.

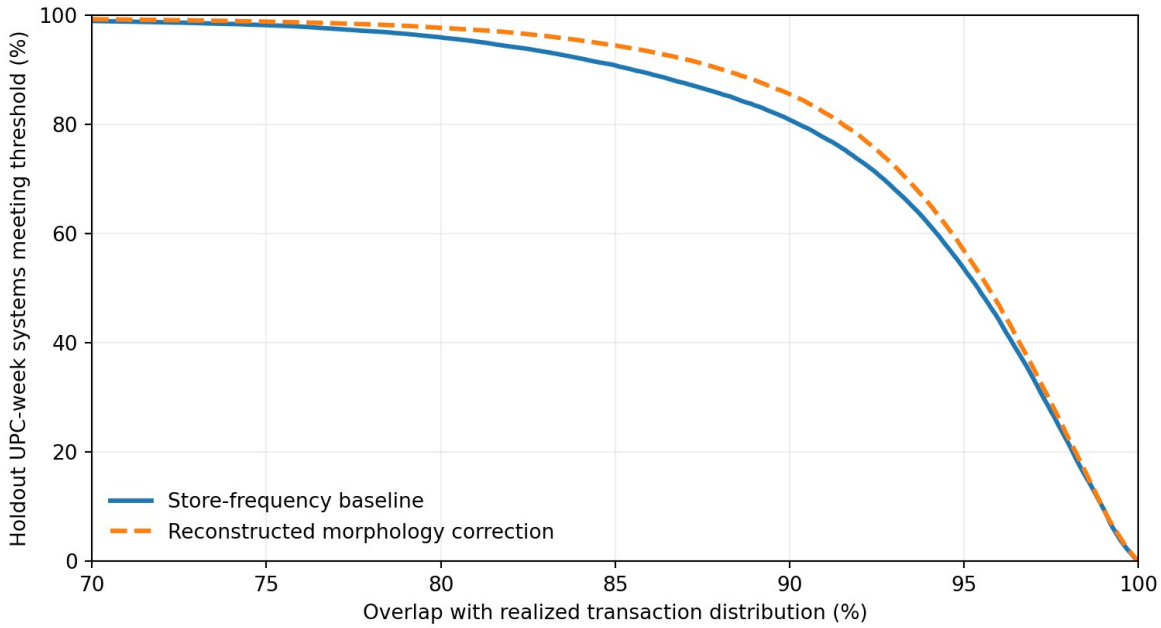


Figure 4. Out-of-sample distributional recovery in 38,553 UPC-week systems from the independent LightGBM reconstruction. For each overlap threshold, lines show the proportion of holdout systems meeting that threshold. The solid line uses raw seller frequency; the dashed line uses a model trained only on weeks 1–300. Both rely on prices and positive-sales store support; the true quantities are used exclusively for evaluation. Source: independent reconstruction from Dominick's scanner records.

The effect is not restricted to a handful of unusually easy systems. At least 90% overlap is achieved by 31,171 of 38,553 holdout systems (80.9%) using raw seller frequency and 32,975 systems (85.5%) using the reconstructed model. At the stricter 95% threshold, the respective counts are 20,646 (53.6%) and 21,919 (56.9%). These counts belong to the independently reconstructed model and differ slightly from the historical figures reported in prior manuscript versions. The distributional results remain conditional on observed positive-sales store-price support.

5.3 What the holdout test establishes

This later-week test shows that positive-sales store-price frequencies retain predictive value for the distribution of realized activity after the quantity weights are withheld. It does not identify causal price responses, total sales volume, or revenue, and it remains conditional on the active-store price support. The next section asks a separate economic question: whether these inferred weights improve calculation of the mean price actually paid.

5.4 Economic consequences: measuring the average transaction price

For each UPC-week, let the true unit-weighted transaction price equal the sum of $p \times sQ(p)$ across observed price points. When quantities are hidden, we estimate the same object using the store-frequency weights $sP(p)$; this is equivalent to the mean price across positive-sales stores. Benchmarks assign equal weights to distinct observed prices or select the modal store price. Error is measured as absolute percentage deviation from the scanner-data unit-weighted price.

Table 4. Economic price-measurement accuracy: mean absolute percentage error (MAPE) across UPC-week systems. New reanalysis; source code and aggregate results supplied separately.

Sample	UPC-weeks	Distinct-price mean	Modal price	Store-frequency mean	Error reduction vs distinct prices
All systems	167,216	2.39%	2.13%	0.57%	76.2%
Later-week holdout	38,553	3.34%	1.97%	0.71%	78.6%

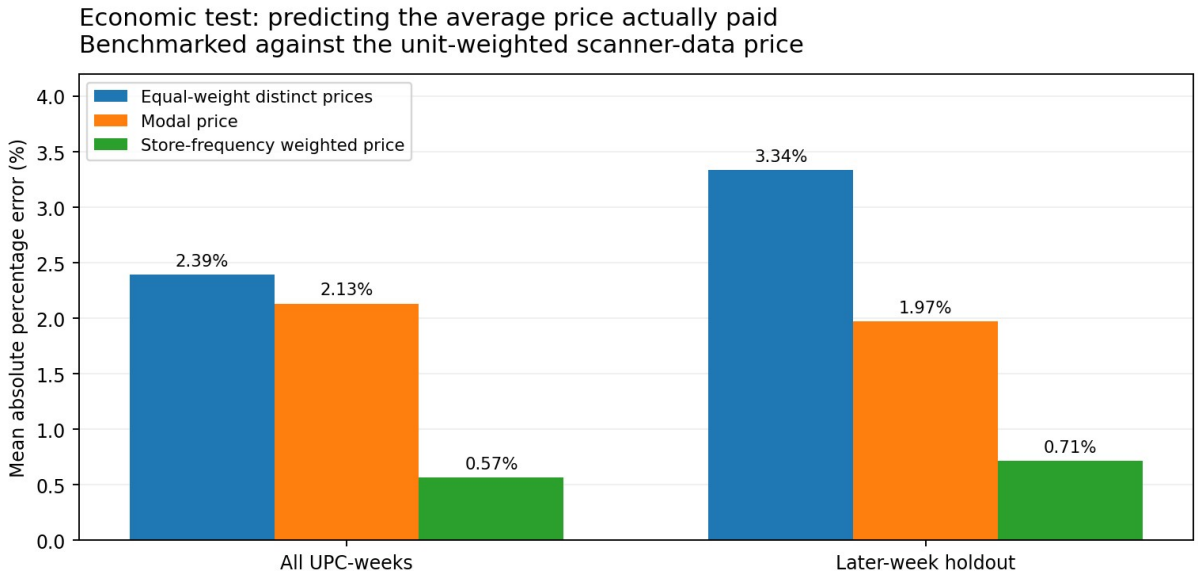


Figure 5. Relative accuracy of three observable price-level estimators in the full sample and later-week holdout. The reference price uses realized unit sales as weights. Source: independent reconstruction using Dominick's movement files.

In the reconciled sample of 167,216 UPC-week systems, the raw selling-store mean has a mean absolute percentage error (MAPE) of 0.5690%, versus 2.3916% when distinct price points receive equal weight and 2.1336% when the modal price is used. In the 38,553-system holdout, those errors are 0.7142%, 3.3381%, and 1.9725%. Store weighting therefore reduces holdout MAPE by 78.6% relative to equal weighting of distinct price points, but this comparison alone is not sufficient to establish the incremental value of a novel algorithm: a seller-weighted mean is already an obvious baseline when store-level prices exist. We next compare that baseline against a transparent regression correction and a more flexible gradient-boosting approach.

Inference accounts for repeated observations of the same UPC. A stratified nonparametric bootstrap (1,000 replicates, resampling UPC clusters within category) gives a 95% confidence interval of [1.76, 1.89] percentage points for the full-sample MAPE advantage of store-frequency weighting over the distinct-price benchmark. The corresponding later-week interval is [2.51, 2.75] points; both exclude zero. These intervals quantify imputation-error differences conditional on the analyzed chain and store-price sample, not population representativeness.

Replication and reconciliation. We independently reconstructed all five categories from raw movement files and exactly reproduced the original eligibility counts (57,964 cereals; 38,694 laundry detergents; 9,181 paper towels; 38,237 shampoo; 23,140 canned tuna). The previous five-cell mismatch was caused entirely by applying half-up rather than half-to-even price rounding at the cent level. We use IEEE ties-to-even rounding throughout, and both analyses now contain 167,216 systems; the later-period test contains 38,553. A second independent program also reproduces the original descriptive correspondence results, including 94.5125% modal correspondence and 5.0308% median total variation distance. We additionally rebuilt and trained a distinct LightGBM morphology adjustment on weeks 1–300 and evaluated it on weeks above 300, obtaining 95.70% median corrected overlap versus the historical reported 95.71%. The original training script itself has not been recovered.

5.5 Beyond a simple baseline: transparent versus nonlinear morphology corrections

The relevant incremental comparison is not only against equally weighting distinct price levels. We also estimate a deliberately simple correction: pooled ridge regression predicts the log transaction-to-seller share ratio from five observed morphological features (relative signed and absolute distance to the seller-weighted price mean, the modal-price indicator, log seller share, and relative dispersion). Predictors are standardized using weeks 1–300, a fixed ridge penalty of 10 is applied, and fitted share multipliers are renormalized to one within each product-week. The later weeks are used only for evaluation. LightGBM uses the richer sixteen-feature specification reported above. Both are evaluated on exactly the same 38,553 holdout systems.

Table 5. Stronger out-of-sample comparisons: full-distribution accuracy and mean price actually paid (38,553 product-weeks).

Method	Mean TV distance	Median TV distance	Price MAPE
Raw store frequency	6.504%	4.598%	0.714%
Ridge correction	6.151%	4.357%	0.666%
LightGBM correction	5.738%	4.297%	0.618%

The nonlinear correction improves mean TV by 0.413 percentage points relative to ridge (95% stratified UPC-cluster bootstrap interval: 0.371–0.455 points) and reduces price MAPE by an additional 0.0485 points (95% interval: 0.0408–0.0564). Relative to raw seller frequency, LightGBM reduces the holdout price MAPE from 0.7142% to 0.6176% (paired mean difference 0.0967 points; 95% UPC-cluster interval: 0.0812–0.1147). These are estimated conditional improvements within this dataset. Only 54.0% of systems have lower price-level error than the raw seller-frequency estimator after the nonlinear correction; therefore aggregate gains should not be mistaken for uniform dominance.

Heterogeneity matters. LightGBM reduces both mean TV and mean price MAPE relative to ridge in cereals, detergents, canned tuna, and shampoo. In paper towels, the simpler ridge correction is better (mean TV 4.762% versus 4.880%; price MAPE 0.413% versus 0.430%). This limits a universal claim for the more flexible model and warrants analysis of the conditions under which nonlinear corrections generalize.

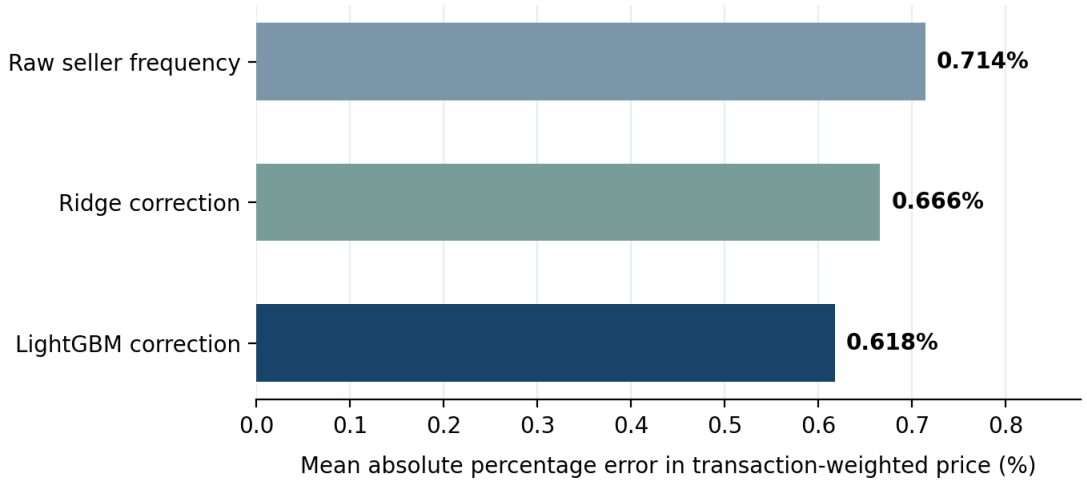


Figure 6. Mean absolute percentage error of estimated unit-weighted prices in the holdout sample using raw store-frequency, ridge-adjusted, and LightGBM-adjusted weights. Training stops at week 300. Source: independent implementations based on Dominick's Finer Foods scanner data; all errors evaluate predictions on weeks >300.

5.6 External validation across a second retailer: Walmart M5

We evaluate whether the price-to-quantity correspondence is confined to Dominick's using independently released historical Walmart data from the M5 Forecasting competition (Makridakis, Spiliotis, and Assimakopoulos, 2022; Seaman and Bowman, 2022). The M5 validation dataset covers 3,049 products in ten Walmart stores across three U.S. states from 2011 to 2016. We join daily units to store-product weekly selling prices and retain 273 complete

retail weeks. For each fixed item-week, units are summed across the seven days and allocated to the recorded weekly store price. This is an approximation to the distribution of units at weekly quoted prices, not a transaction-receipt-level observation of prices paid.

Because M5 follows at most ten stores per product, the Dominick's requirement of at least 20 selling stores cannot be applied. We instead require at least six stores in the relevant support and at least two exact price points. The first specification counts only stores with positive unit sales, matching Dominick's support convention. The second, which is closer to the intended setting with quantities unavailable, uses all stores with a recorded weekly selling price, including those with zero weekly units. The price records include 1,581,733 priced product-store-week cells with zero units, permitting this additional test.

Table 6. Walmart M5 correspondence under alternative observable-price supports.

Price support	Eligible item-weeks	Median overlap	95% CI / note
Positive-sales stores	115,931	91.35%	91.11–91.60%
All stores with weekly price	176,704	88.51%	88.05–88.89%
All priced, same 115,931 item-weeks	115,931	90.00%	Matched comparison

The positive-sales specification contains 115,931 eligible item-weeks and yields 91.35% median distributional overlap (95% product-cluster bootstrap interval: 91.11–91.60%). The all-priced specification covers 176,704 item-weeks and yields 88.51% overlap (88.05–88.89%). These two figures do not use the same sample. Holding the sample fixed at the original 115,931 item-weeks, inclusion of all price-reporting stores lowers the median from 91.35% to 90.00% and increases mean total-variation distance by 1.74 percentage points. Raw store frequency improves on equal weighting of distinct prices in 82.8% of the positive-sales sample and 77.0% of the all-priced sample, a comparison that remains an informative but undemanding benchmark.

The second-retailer result has important limits. Approximately 21.7% of M5 item-weeks with at least six selling stores have two or more distinct prices and qualify for the main exercise; most eligible systems have only two prices. Requiring at least three distinct prices reduces median overlap to 86.11% on selling-store support and 80.00% on all-priced support; requiring at least four reduces it to 82.52% and 71.53%. As in Dominick's, a store-quantity permutation placebo shows that much of the raw alignment arises from the extensive-margin arithmetic rather than from a newly identified equilibrium mechanism. Figure 7 documents both the full overlap distributions and their dependence on price-system complexity.

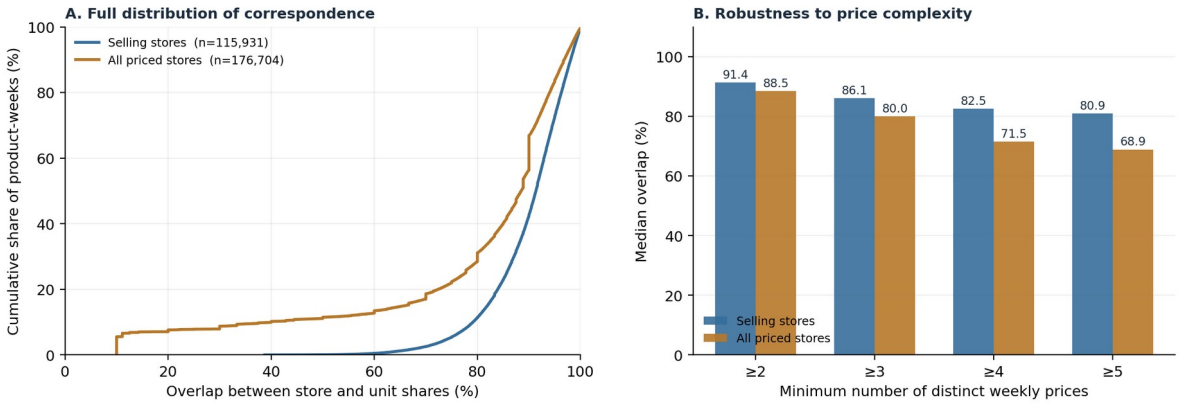


Figure 7. External descriptive replication with Walmart M5. Panel A compares the cumulative distributions of store-to-unit overlap for eligible item-weeks; panel B shows sensitivity to the minimum number of distinct observed prices. Selling-store and all-priced analyses are defined on different populations unless explicitly described as matched. Store prices are recorded weekly, not at the receipt level. Source: author's calculations using publicly released M5 Accuracy data.

Walmart provides complementary cross-retailer evidence, but it does not establish universally high accuracy. Ten-store coverage, frequent two-price systems, and weekly rather than receipt-level price observations limit direct comparison. The original Dominick's estimand, numerical results, and holdout design remain unchanged.

6. Implications and Conclusion

6.1 A distinct empirical object: the realized transaction distribution

Traditional demand analysis and the transaction distribution answer different questions. A Marshallian demand schedule describes how quantity demanded would change across alternative prices, holding other relevant conditions fixed. It is therefore a behavioral and counterfactual object (Marshall, 1920). The object studied here is observational: the realized transaction distribution, sQ(p), records how the units actually sold were distributed across the prices that coexisted in the market.

This distinction matters because the two objects have different strengths. A structural demand curve is needed to estimate elasticities, evaluate counterfactual price changes, and calculate welfare. The realized transaction distribution cannot do those jobs. But for the question of where market activity actually occurred, it is more direct evidence: the quantities are observed rather than inferred from a hypothetical demand schedule. Scanner data make that distinction visible because they jointly record prices and realized quantities at highly disaggregated levels (Dubois, Griffith, and O'Connell, 2022).

The contribution is therefore not to replace the Marshallian demand curve with a price distribution. It is to add an empirical object that the conventional demand diagram does not directly display: the realized distribution of market activity across coexisting prices.

6.2 What price distributions reveal about realized market activity

The empirical connection between that transaction distribution and the price system is strong. Across 167,216 UPC-week systems, the most frequent store price is also a transaction mode in 94.51% of cases. The median distance between the full store-price and transaction distributions is 5.03%, equivalent to 94.97% median overlap. The result is not confined to highly concentrated price systems and weakens in exactly the cases predicted by the accounting identity: when sales per store differ sharply across price points, especially under localized promotions.

The holdout experiment makes the result operational rather than merely descriptive. Quantity weights are deliberately hidden for 38,553 later UPC-week systems. Conditional on the observed active store-price support, the price-frequency distribution alone achieves 95.40% median overlap with the hidden transaction distribution. A correction based only on observable price-system morphology raises median overlap to 95.70% and outperforms the raw baseline in 63.1% of systems. The central result is not the incremental improvement alone. Before any correction is applied, prices already recover most of the realized transaction structure.

This result sits between two established antecedents. Eden, Eden, and Yuen (2016) use observed quantities to analyze price-ranked bins within UPC-week cells; our question is whether quantities can be recovered after they are withheld, on the complete support of exact prices. Antoniades, Feenstra, and Xu (2022) infer expenditure shares across different goods from the retail distribution of sellers; our target is the fraction of units sold at each price for one fixed UPC-week. The distinction is one of estimand and validation design, not a claim that prices, quantities, or seller presence were first studied here.

An independent Walmart M5 exercise additionally finds median overlap of 91.35% when using price frequencies among selling stores and 88.51% when all weekly priced stores are included. The Walmart results strengthen evidence that the mapping is not unique to the Dominick's retail chain, while indicating materially lower correspondence in systems with more distinct prices and when zero-sales stores are included. They do not verify the same transaction-price estimand, because M5 prices are store-week observations rather than receipt-level prices.

6.3 Economic interpretation and identification limits

Our empirical object is the allocation of units sold across simultaneously observed price points. Its mode is the most common realized transaction price, not necessarily the unique clearing price of a market. The original paper proposed a conditional link to Marshallian equilibrium regions; this version does not rely on that interpretation to establish the measurement result.

Without identifying separate supply and demand schedules, the realized transaction distribution does not identify marginal willingness to pay, elasticity, welfare, or structural equilibrium. Those questions require additional assumptions or data. The direct contribution here is narrower and more testable: price prevalence among selling stores can be used as a proxy for otherwise missing transaction weights in well-defined settings.

The approach can be useful even if the correspondence is partly mechanical, as long as its errors, alternatives, and failure cases are explicitly measured. Promotions and systematic volume differences between stores remain important practical limitations.

6.4 Implications for market measurement and decision-making

Price information is routinely available even when transaction weights are missing, making price-level measurement a practical use of this research. In the additional test in Section 5.4, equal seller weights reduce mean absolute percentage error in the unit-weighted average price substantially relative to two familiar, but imperfect, reference rules: equal weighting of distinct observed price levels and selection of the modal price. These results apply to the active-store support in a single retail chain; they do not automatically generalize to all price audits.

The implication is not that sales data or structural demand analysis become unnecessary. Rather, observable cross-sectional store-price frequencies can provide an empirically testable proxy when true unit-based weights are unavailable. Local promotions, differences in store size and assortment, and the absence of observed shelf prices at zero-sales stores define important boundaries on that practical use.

6.5 Research agenda and conclusion

The independently released Walmart data provide an initial second-retailer check, including an observable set of priced but zero-sales store-weeks. The next step is to test the mapping in further retailers and countries with sufficiently many stores and receipt-level product prices, under genuinely different promotional regimes and store-size distributions. Comparable product-by-store-by-time records are demanding to obtain; we invite research institutions, statistical agencies, and retail organizations with such data to replicate, challenge, and refine the conditions under which retail price frequencies contain transaction-share information.

In the settings examined, price dispersion reveals more than the location and shape of observed prices. The Dominick's data show a high-overlap proxy for realized allocations on positive-sales price support; the distinct Walmart data show a related, weaker pattern under both active-store and all-priced definitions. The main contribution remains the tested price-to-transaction mapping and an explicit account of its limits. Neither an inflation application nor a Marshallian equilibrium interpretation is necessary for that conclusion.

Data acknowledgment

This study uses the Dominick's Finer Foods scanner dataset made available by the Kilts Center for Marketing at the University of Chicago Booth School of Business, together with the publicly released Walmart M5 Forecasting competition data for a separate external robustness exercise. The author acknowledges both data providers. All transformations, calculations, interpretations, and conclusions in this paper are the author's own.

Appendix A. Robustness and Implementation Details

Visualization sampling and full-population figure

Figure 1 displays one UPC-week system from each of the five categories. The selection algorithm first calculates the category-wide median TV distance from the complete baseline dataset, then restricts illustration candidates to UPC-week systems with at least 60 positive-sales stores, 5–9 exact observed prices, and store focalization from 20% to 70%. It selects the candidate closest in TV to the unconditional category median; exact ties are broken by UPC and then week. The five selected UPC-week pairs are: cereals 1600067030/week 168; laundry detergents 3700091240/week 175; paper towels 3700063263/week 163; shampoo 3700000098/week 241; and canned tuna 4800001164/week 261. These display filters do not apply to Figure 2 or any table. Selection uses realized quantities to construct illustrative examples; therefore the examples are not a predictive holdout or a random sample. Their purpose is visualization only. Figure 2 uses the entire 167,216-system analysis sample; the rightmost 0.55% of TV distances exceed its displayed horizontal axis. The figure construction code and derived cells are provided with the replication materials.

A.1 Focalization

The baseline correspondence strengthens as the store-frequency distribution becomes more concentrated, but remains meaningful in strongly dispersed systems. When no price exceeds 40% of active stores, there are 25,884 systems, the median number of price points is eight, modal correspondence is 72.55%, and median TV distance is 7.66%. For systems with focalization of 40-60%, 60-80%, and at least 80%, modal correspondence is 95.06%, 99.58%, and 99.73%, respectively; median TV distance is 6.85%, 4.45%, and 2.14%.

A.2 Promotion bins

The Dominick's sale flag is used as a conservative indicator of promotional intensity. Median TV distance is 4.72% in systems with no flagged promotion, 6.91% when 0-10% of transacting stores are flagged, 9.87% when 10-50% are flagged, and 6.10% when more than half of transacting stores are flagged. Because an unflagged observation does not prove the absence of a promotion, these bins should be interpreted as descriptive rather than exhaustive.

A.3 Permutation benchmark

The permutation exercise samples 300 baseline systems from each of the five categories. Within each system, store-level quantities are randomly reassigned across the observed store-price configuration 50 times. Prices, store counts, and the empirical quantity distribution are preserved; only the association between price and store quantity is destroyed. The median actual TV distance is 5.039%, compared with a median randomized TV distance of 3.197%. Actual modal correspondence is 94.8%; mean randomized modal correspondence is 96.04%. Actual TV exceeds the system-specific randomized median in 69.4% of sampled systems.

A.4 Hidden-quantity prediction

The independent reconstructed out-of-sample analysis uses weeks 1–300 for estimation and weeks above 300 for testing. Conditional on the observed positive-sales store-price cells, a LightGBM regression with squared-error objective predicts log[sQ(p)/sP(p)] from store share, log share, modal-price indicator, signed and absolute relative distances to the modal and store-frequency mean prices, normalized price rank, cumulative store mass below the price, focalization, counts of stores and distinct prices, Shannon entropy, relative price dispersion, and the store mass below and above the modal price. The implementation uses 100 boosting trees, learning rate 0.1, 31 leaves per tree, no L2 penalty, random seed 42, and the same settings in all categories; no holdout-based tuning is performed. The predicted correction factor is exp(predicted log ratio); factors multiply sP(p) and are normalized within each UPC-week. In 38,553 holdout systems median TV declines from 4.60% to 4.30% and mean TV from 6.50% to 5.74%; the correction improves on the raw baseline in 63.1%. A stratified UPC-cluster bootstrap with 500 replicates gives a 95% percentile interval of [0.709, 0.829] percentage points for the paired mean TV

reduction. This code independently reconstructs the documented model class but is not the unrecovered original implementation. The procedure is predictive, not causal.

A.5 Modal ties

Because prices are discrete, two or more price points can occasionally tie for the maximum store count or maximum transaction share. Baseline modal correspondence is defined as overlap between the two modal sets rather than by selecting one arbitrary tied price. A deterministic single-mode rule produces slightly lower correspondence and does not change the substantive conclusions.

A.6 Positive-sales coverage

Because zero-sales store-weeks lack a usable price in the Dominick's movement files, we calculate, for each UPC-week, the share of all valid (OK=1) store records with positive sales. The median positive-sales coverage among baseline systems is 92.31%. Requiring at least 90% coverage leaves 91,938 systems: modal correspondence is 93.85% and median TV distance is 5.07%. Requiring at least 95% coverage leaves 69,217 systems: modal correspondence is 93.39% and median TV distance is 5.12%. The main mapping is therefore not driven by UPC-weeks with low positive-sales coverage, although complete shelf-price data remain necessary to observe prices in zero-sale stores directly.

A.7 Economic-price measurement, uncertainty, and replication

We compute, for each UPC-week, $P_{actual} = \Sigma_p p \times units_p / \Sigma_p units_p$, $P_{store} = \Sigma_p p \times stores_p / \Sigma_p stores_p$, $P_{levels} = (1/K)\Sigma_p p$, and $P_{mode} = argmax_p(stores_p)$, choosing the lowest tied price. The relative error is $APE = 100 \times |P_{estimate} - P_{actual}| / P_{actual}$. We report the mean APE across eligible UPC-weeks. Aggregation weights systems equally; results do not pretend to estimate the chain-wide revenue-weighted effect.

Uncertainty in the price-measurement exercise is evaluated with 1,000 category-stratified UPC-cluster bootstrap replicates. For the independent morphology reconstruction, 500 UPC-cluster replicates yield a 95% interval of [0.709, 0.829] percentage points for the paired mean reduction in total-variation distance. IEEE half-to-even rounding reconciles 167,216 original systems and 38,553 later-period holdouts. The original gradient-boosting training file remains unavailable, but an independently written and executed model based on the documented specification reproduces a closely matching improvement; the reimplementation, trained model, holdout predictions, and feature definitions are archived.

To reproduce the reconciled economic measurement and descriptive modal/TV results, use reanalyse_paper2.py and audit_original_results.py alongside the original five Dominick's category files. For the separately reconstructed morphology adjustment, run reconstruct_paper2_morphology.py with the same original files. Derived model outputs, the fitted LightGBM model, specifications, and the Figure 3 regeneration are included in the private replication package. The underlying Dominick's records are not redistributed. This procedure replaces an earlier unverified corrected figure with independently generated, closely matching results, but does not establish that the historical original software is identical.

A.8 Independent reproduction of the principal descriptive results

A second independent script groups the original Dominick's movement data at the UPC-week-price level and computes seller shares and unit shares, without using any of the reanalysis output tables. It reproduces 167,216 eligible systems, aggregate modal correspondence of 94.5125%, median total-variation distance of 5.0308%, and median total-variation distance of 4.5980% for 38,553 systems after week 300. For 25,884 systems where no store price exceeds 40% of selling-store observations, the modal correspondence is 72.5506% and median TV is 7.6603%. Small differences from earlier printed results are at the rounding precision only.

The economic-price experiment is descriptive, not a causal estimate, and does not establish that price presence predicts absolute sales, revenue, or demand elasticities. The cross-sectional store average is a natural descriptive

baseline, and the estimated gains relative to modal and equally weighted distinct prices must not be misread as gains over a sophisticated transaction model. The separate Walmart exercise in Section 5.6 supplies additional cross-retailer evidence but is constrained by weekly price recording and ten-store coverage; more demanding external tests remain necessary.

A.9 Independent morphology reconstruction and product holdout

Independent reconstruction matches 167,216 eligible systems and 38,553 later holdouts. The corrected median overlap is 95.7026%, versus 95.71% originally reported; 63.10% of systems improve, versus 62.8% historically. Source code, features, trained model and predictions are archived. No holdout tuning was performed. Among 8,460 systems involving UPCs unseen in training, corrected median overlap is 95.97%, and 62.0% improve. This is a within-chain robustness check, not an external validation. The historical training script was not recovered.

A.10 Stronger-benchmark specification and statistical uncertainty

The ridge comparison uses standardized observed price-morphology features, a fixed L2 penalty of 10, and the same UPC-week-price training and holdout definitions as the reconstructed LightGBM model. The regression target is log[sQ(p)/sP(p)]; predicted ratios are exponentiated, multiplied by sP(p), and renormalized within each UPC-week. These models are descriptive and predictive, not structural. The test-set evaluation computes mean TV, median TV, and mean absolute percentage error in the scanner-observed unit-weighted transaction price. The baseline, ridge and boosted estimators are evaluated on identical price supports.

For precision, we resample UPC clusters separately within each product category, keeping all their holdout weeks together, over 2,500 bootstrap replicates. For the LightGBM improvement over ridge, the 95% percentile interval for mean TV is [0.371, 0.455] percentage points and for mean price MAPE is [0.0408, 0.0564] points. For the advantage over the raw-frequency price estimator, the bootstrap interval is [0.0812, 0.1147] percentage points. These intervals measure sampling variation in the evaluated panel, not performance on independent retailers. No cross-retailer external test is available.

A supplementary test of one-week changes in the transaction-weighted price for 32,356 consecutive product-week pairs yields mean absolute price-change errors of 0.789 percentage points for raw store-frequency weights and 0.722 points for LightGBM. However, the nonlinear correction is better for only 42.3% of those weekly comparisons and slightly worsens the median weekly-change error. The substantial asymmetry cautions against claims of uniformly better inflation dynamics. This test is exploratory and should not be interpreted as a validated aggregate consumer price index.

A separate four-week circular moving-block bootstrap across the 99 test weeks (4,000 replicates) retains positive mean advantages of LightGBM versus ridge: [0.310, 0.514] percentage points for TV distance and [0.0309, 0.0689] points for mean price MAPE. This check allows the observations in adjacent weeks to move jointly; it complements, but does not replace, inference across independent retail chains.

The independent LightGBM implementation was reconstructed with knowledge of the aggregate result printed in an earlier paper version. Although weeks after 300 were excluded from fitting and its hyperparameters match the documented specification, the exercise is not a blinded prospective validation. The Walmart M5 analysis is independent of this fitted LightGBM model: no correction is trained or transferred to Walmart. Additional validation using independent receipt-level data with broader store coverage remains necessary.

A.11 Walmart M5 external-validation design and replication

Source and preprocessing. We use the official M5 Accuracy files sales_train_validation.csv, sell_prices.csv, and calendar.csv (https://www.kaggle.com/competitions/m5-forecasting-accuracy/data). Original daily sales cover days d_1 through d_1913; we retain 273 complete seven-day Walmart retail weeks, matching each store's item-week units to the M5 store-item weekly sell_price. The raw files represent 3,049 distinct products and ten stores in California, Texas, and Wisconsin. Item-week prices are grouped in cents. All 4,984,978 positive-unit item-store-week cells have a corresponding recorded weekly price. A further 1,581,733 priced cells register zero weekly unit sales. No receipts, within-week price changes, exact stock-on-shelf indicators or shopper identities are available.

Definitions, eligibility, and precision. For each fixed item-week with positive aggregate units, let sP(p) be the share of stores in the selected support charging price p, and sQ(p) the share of total recorded weekly unit sales assigned to stores at that price. We compute TV = 0.5 Σ|sP(p) – sQ(p)| and overlap = 1 – TV; modal correspondence allows ties. Analyses require at least six stores and two or more distinct prices; both conditions are evaluated separately for positive-sales and all-priced supports. An independent item-cluster bootstrap stratified by broad product category with 400 replicates yields 95% percentile intervals for the sample medians. All five categories in Dominick's use a distinct source and retain their original eligibility criteria.

Robustness and scope. For at least three price points, the M5 median overlaps are 86.11% on positive-sales support (25,491 systems) and 80.00% on all-priced support (48,221 systems); for at least four points they are 82.52% (3,258) and 71.53% (8,382). A category-stratified permutation check shuffles store-level weekly unit totals within 1,500 eligible item-weeks, with 100 permutations per system. The median actual overlap is 91.59% and the median system-level shuffled benchmark is 92.68%; this indicates a substantial arithmetic component, not causal identification of demand. The released M5 prices are weekly store-item prices, not necessarily the actual receipt prices for every unit transacted during the week. Zero-unit stores with listed prices are observable in these files, but a listed price does not establish physical availability on the shelf. The replication package provides exact transformations and aggregated analyses, without redistributing Walmart's original files.

Declaration of generative AI and AI-assisted technologies

During the preparation of this work, the author used OpenAI ChatGPT to assist with drafting and editing text, developing and reviewing analysis code, checking statistical computations, and preparing figures. The author reviewed and revised the outputs and takes full responsibility for the methods, results, interpretations, and content of the manuscript.

References

Antoniades, Alexis, Robert C. Feenstra, and Mingzhi (Jimmy) Xu. 2022. "Using the Retail Distribution of Sellers to Impute Expenditure Shares." American Economic Review 112(7): 2213-2236. https://doi.org/10.1257/aer.20190896.

Ayoubkhani, Daniel, and Heledd Thomas. 2022. "Estimating Weights for Web-Scraped Data in Consumer Price Indices." Journal of Official Statistics 38(1): 5-21. https://doi.org/10.2478/JOS-2022-0002.

Burdett, Kenneth, and Kenneth L. Judd. 1983. "Equilibrium Price Dispersion." Econometrica 51(4): 955-969. https://doi.org/10.2307/1912045.

Campbell, Jeffrey R., and Benjamin Eden. 2014. "Rigid Prices: Evidence from U.S. Scanner Data." International Economic Review 55(2): 423-442. https://doi.org/10.1111/iere.12055.

Dubois, Pierre, Rachel Griffith, and Martin O'Connell. 2022. "The Use of Scanner Data for Economics Research." Annual Review of Economics 14: 723-745. https://doi.org/10.1146/annurev-economics-051520-024949.

Eden, Benjamin. 2018. "Price Dispersion and Demand Uncertainty: Evidence from U.S. Scanner Data." International Economic Review 59(3): 1035-1075. https://doi.org/10.1111/iere.12296.

Eden, Benjamin, Maya Eden, and Jonah Yuen. 2016. "Inside the Price Dispersion Box: Evidence from U.S. Scanner Data." Vanderbilt University Department of Economics Working Paper 16-00017. https://www.accessecon.com/pubs/VUECON/VUECON-16-00017.pdf.

Franco, Iván. 2026. "Price Systems: Market-Clearing Focal Prices and the Predictive Structure of Retail Price Dispersion." Working paper, Triplethree International.

Hitsch, Günter J., Ali Hortaçsu, and Xiliang Lin. 2021. "Prices and Promotions in U.S. Retail Markets." Quantitative Marketing and Economics 19: 289-368. https://doi.org/10.1007/s11129-021-09238-x.

Kaplan, Greg, and Guido Menzio. 2015. "The Morphology of Price Dispersion." International Economic Review 56(4): 1165-1206. https://doi.org/10.1111/iere.12134.

Kilts Center for Marketing. n.d. 'Dominick's Dataset.' University of Chicago Booth School of Business. Accessed October 12, 2026. https://www.chicagobooth.edu/research/kilts/research-data/dominicks.

Makridakis, Spyros, Evangelos Spiliotis, and Vassilios Assimakopoulos. 2022. "The M5 Competition: Background, Organization, and Implementation." International Journal of Forecasting 38(4): 1325–1336. https://doi.org/10.1016/j.ijforecast.2021.07.007.

Marshall, Alfred. 1920. Principles of Economics. 8th ed. London: Macmillan.

Seaman, Brian, and John Bowman. 2022. "Applicability of the M5 to Forecasting at Walmart." International Journal of Forecasting 38(4): 1468–1472. https://doi.org/10.1016/j.ijforecast.2021.06.002.

Sorensen, Alan T. 2000. "Equilibrium Price Dispersion in Retail Markets for Prescription Drugs." Journal of Political Economy 108(4): 833-850. https://doi.org/10.1086/316103.

Stigler, George J. 1961. "The Economics of Information." Journal of Political Economy 69(3): 213-225. https://doi.org/10.1086/258464.

Varian, Hal R. 1980. "A Model of Sales." American Economic Review 70(4): 651-659.