Path Dependence in the Price Level: Testing Pricing Models with a Fully Reversed VAT Cut

Youssef Benzarti, Santiago Garriga, Darío Tortarolo

October 2026

Abstract: This paper uses a large, unanticipated, and temporary consumption tax change to test workhorse models of price adjustment. In August 2019, Argentina eliminated the 21 percent VAT on selected food staples. The repeal, announced at the outset, took effect in January 2020. Using scanner data on identical products across supermarkets, with closely related untreated products as controls, we document six facts. Adjustment to the cut is immediate, there is no anticipatory response to the announced repeal, the frequency of price changes responds to both shocks, and cross-store price dispersion compresses at the cut. The first rejects Calvo pricing, and the other three fit state-dependent pricing better than Calvo. Moreover, pass-through is roughly twice as large upward as downward, and the resulting overshoot does not decay. An estimated

Benzarti: University of California, Santa Barbara and National Bureau of Economic Research. Garriga: *Centro de Estudios en Finanzas Públicas*, IIE-FCE-UNLP. Tortarolo: World Bank DECRG. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

menu cost model that includes Calvo as a special case reproduces the first four facts in sign, misses the fifth by passing through more than the whole cut, and cannot produce the sixth. Something outside these models must carry the history of the episode. We find that it works product by product, not store by store, and that the increase on the same good depended on who sold it. These facts fit sellers coordinating on public numbers but do not rule out a reference price that consumers attach to the good.

JEL codes: E31, H22, L11, L81. **Keywords**: price adjustment, menu costs, Calvo pricing, value-added tax, pass-through, hysteresis.

1. Introduction

There is a large literature in macroeconomics documenting that demand shocks can have significant effects on real output. Whether they do, and for how long, depends on how quickly firms adjust nominal prices. Two workhorse models answer that question differently. In the Calvo (1983) model a firm reprices when it draws the opportunity, so adjustment to any shock is staggered across firms and spread over time. In the menu cost model a firm reprices when the gain exceeds a fixed cost, so the firms that move are selected to be the ones furthest from their target and aggregate adjustment is faster. Distinguishing the two requires more than a time series of prices. It requires a shock large enough to move prices out of any inaction region, exogenous to the conditions firms face, and observed alongside a control group of goods that the shock did not touch, with prices measured for identical items rather than for aggregates in which composition and quality move together with price. There is a fifth requirement. Both models predict that prices return to the level that fundamentals imply after a shock, and disagree only about how fast. Testing that prediction needs the fundamentals to go back to where they were, which means the shock has to be undone: a shock that is never reversed cannot tell a slow return from no return. Few settings supply all five requirements together, and the fifth is rarer still.

We use one such setting: the Argentine value-added tax holiday of 2019. On August 16 the government removed the 21 percent VAT from thirteen staple food categories accounting for 26 percent of food expenditure. The cut was unanticipated, it was large, and its expiry date was announced on the day it began, which makes the repeal on January 1, 2020, a pre-announced shock of the same size in the opposite direction. We compare treated with untreated goods sold in the same aisles. We estimate the difference between the treated and control groups period by period in a dynamic difference-in-differences with item and time fixed effects, so that composition and quality are held

fixed and the coefficients trace the price of the same barcode through the cut and the repeal. The data are retail scanner data for 2019 and 2020: a weekly panel of the national chains, a monthly panel of individual stores in Córdoba that follows the same barcode across stores, and a national monthly panel that distinguishes the large chains from smaller sellers.

We report six facts. First, adjustment to the cut is immediate rather than staggered: the difference between the treated and control groups reaches −13.2 index points within two weeks, with prices indexed to 100 in the week before the cut, and then sits flat for nineteen. Second, there is no anticipatory response to the pre-announced repeal. In the four weeks before it, treated prices rise more slowly than control prices, not faster. Third, the frequency of price changes responds to both shocks. Around the cut, the price of a treated product in a given store falls by more than five percent in 50 percent of cases, against 15 percent for control products. In the month of the repeal, it rises by more than five percent in 75 percent of cases, against 29 percent. Fourth, at the cut the spread of prices of identical products across stores narrows relative to control goods in the same category, by 1.8 log points, about a fifth of its level before the cut. Fifth, prices respond roughly twice as much to the VAT increase as to the VAT cut. On goods the price cap did not cover, pass-through is 65 percent for the increase and 28 percent for the cut. Sixth, the resulting overshoot does not decay: a year after the tax was restored, the wedge is at least as wide as it was in the first month. Benzarti, Garriga, and Tortarolo (2024) document the first fact and the last two for this episode, and we measure them again here because the tests that follow depend on them.

The setting has three shortcomings. First, our price is a unit value, revenue divided by quantity sold, rather than a posted price, so a product whose posted price never changed can register a change when the mix of transactions behind the average moves, which inflates any measure of the frequency of price adjustment. Second, the repeal was accompanied by a negotiated seven percent cap on the price increases of the national

chains, so for anything measured at the repeal we rely on the categories the cap did not cover. Third, the post-repeal window is truncated by the onset of the pandemic and by a general price freeze in March 2020, which leaves ten clean weeks for high-frequency work and requires the longer-horizon results to be read against those events. We address the first by matching the menu cost model to how the frequency of price changes moves with inflation rather than to its level. We also re-estimate the central quantities on a sale-filtered series, on which the overshoot is unchanged and the asymmetry sharpens (Appendix B). We address the second by restricting repeal tests to uncapped goods and reporting capped goods separately throughout. The third does not weaken the central test, because the menu cost model we estimate predicts that the wedge closes within weeks of the repeal, well before the pandemic, and in the data it does not close in those ten weeks or over the rest of 2020.

We then estimate a flexible menu cost model similar to Karadi and Reiff (2019), with firm-specific shocks that are occasionally large, trend inflation, and a Calvo weight left free so that Calvo is nested. We find that the menu cost model reproduces the first four facts in sign, overstates pass-through at the cut, and cannot produce the sixth at all: its post-repeal wedge has a half-life of under three weeks against a measured wedge that has not decayed after a year. That failure comes from the structure of the model, not from our calibration of it. In this model, and in any model with a single long-run equilibrium and no slow-moving state such as customer capital, once the tax is back every firm wants to charge what it charged before the cut, so the wedge stays open only while some firms have not yet repriced, and every month some of them do. Changing the parameters makes the wedge close faster or more slowly, but it always closes.

Our first contribution follows from that. A fully reversed shock left a lasting trace in these prices that the menu cost and Calvo models cannot produce at any calibration, so something these models leave out must carry the history of the episode. Benzarti, Garriga, and Tortarolo (2024) conjecture that it is multiple equilibria in the price level,

the premise of models of coordination failure in price setting (Cooper and John 1988; Ball and Romer 1991). A reference price that consumers attach to the good is the other candidate.

Our second contribution is to ask what carries that history, and we find three things. First, it works product by product, not store by store: the episode widened the spread of prices across products by about a third and left none across the stores selling the same product. Second, the amount depended on the seller, not the product. On goods covered by the seven percent cap, the chains that had signed it raised the price of the same good 4.3 log points less than the stores that had not, and on goods the cap did not cover, 5.4 points more. Third, the increase arrived within two weeks of the repeal. These facts fit sellers coordinating on public numbers, and the second is hard for a consumer-side explanation to match. Two other results do not favor coordination. The increase does not look like a common change added to every store's price, and the overshoot does not fall where a good has more sellers. We do not think data of this kind can separate coordination among sellers from a reference price that consumers attach to the good (Section 7.7).

Our paper is related to three literatures. The first tests models of price adjustment on micro data (Nakamura and Steinsson 2008; Gagnon 2009; Alvarez et al. 2019; Gagnon, López-Salido, and Vincent 2013; Karadi and Reiff 2019). Alvarez et al. (2026) use the euro cash changeover, a nominal reform on a date known in advance, to distinguish among theories of price rigidity, and find that around the changeover prices change more often and by smaller amounts, as their menu cost model with an anticipated date predicts. Like our first four facts, their evidence favors state-dependent pricing, in which firms change prices when conditions make it worthwhile. This episode adds a shock that is large, unanticipated, and above all fully reversed, together with a control group of near-identical goods and prices observed for the same barcode in many stores at once. The reversal turns the sixth fact into a test of these models, because without it there is

no way to tell whether prices are returning slowly to their old level or have settled at a new one.

The second is the literature on value-added tax pass-through (Benzarti and Carloni 2019; Benzarti et al. 2020; Benzarti and Tazhitdinova 2021). We use the same variation as Benzarti, Garriga, and Tortarolo (2024), which asks whether VAT cuts can dampen food price inflation and what the accompanying price controls did to their incidence. That paper documents the asymmetry of the pass-through and its hysteresis, which are Facts 5 and 6 here. There the caps are the policy under evaluation. Here they matter as a confound at the repeal and as one of the public numbers sellers moved to. The tax incidence literature has found the same hysteresis elsewhere but has struggled to explain it, since it is inconsistent with standard static incidence models (Benzarti 2025). This paper narrows the explanation down. The standard explanations of the asymmetry rely on price rigidities, but prices in this economy move every few weeks, and no calibration of the estimated menu cost model can produce Fact 6. What remains works product by product, and the comparison of the same barcode at sellers bound and not bound by the cap, with January increases bunching at the two numbers the policy made public, shows that sellers moved to the public numbers they faced. Benzarti, Garriga, and Tortarolo (2024) also show that prices adjusted immediately to both tax changes, which Calvo pricing cannot produce, and that January price changes in the chains bunch at the cap and at full pass-through. Facts 2 to 4 are new here, as is the comparison of the same barcode across sellers bound and not bound by the cap.

The third literature studies multiplicity and coordination in price setting. Models with strategic complementarities, in which a firm wants to charge more when its rivals do, generate multiple equilibria in the price level (Ball and Romer 1991; Blanchard and Kiyotaki 1987), and Cooper and John (1988) set out the class of games in which such coordination failures arise. Firms themselves most often cite coordination failure as the reason their prices are sticky (Blinder et al. 1998). When there is more than one

equilibrium, something must decide which one prevails, and Krugman (1991) frames the choice as one between history and expectations. This literature is largely theoretical, and this episode lets us check three of its predictions against market prices (Section 7.2). The closest empirical work studies sellers coordinating on public focal points, numbers that every seller sees and expects the others to see: price-leadership cycles in retail gasoline (Byrne and de Roos 2019), and post-merger price jumps that competition without coordination cannot explain (Miller and Weinberg 2017). Our episode differs in that the focal point was set by the state, named a number, and was attached to a shock that was later undone in full.

2. Institutional Setting

2.1. A 21 Percentage Point VAT Holiday

On August 11, 2019, President Macri lost the presidential primaries to the left-wing candidate Fernández by 15.5 percentage points, a much wider margin than expected. Markets fell the following day, and the dollar rose 24 percent against the peso. On August 15 the government responded with a 4.5-month VAT holiday on basic food, with the stated aim of limiting the effect of the depreciation on prices (Decree 567/2019). The decree was published and took effect the following day, August 16, so the VAT cut was fully unexpected. The government announced on the same day that the cut would be temporary, with an expiration date of December 31, 2019.[1] Figure 1 collects the dates of the episode.

The rate fell to 0 percent on thirteen goods in the "Basic Food Basket". Most had been taxed at the standard 21 percent; wheat flour and unpackaged bakery bread had been taxed at the reduced 10.5 percent, and fluid milk only when sold with additives, since

[1]This is the same variation we use in Benzarti, Garriga, and Tortarolo (2024), where we focus on a different question (whether governments can use VAT cuts to dampen the effect of inflation and what welfare effects those VAT cuts have on households).

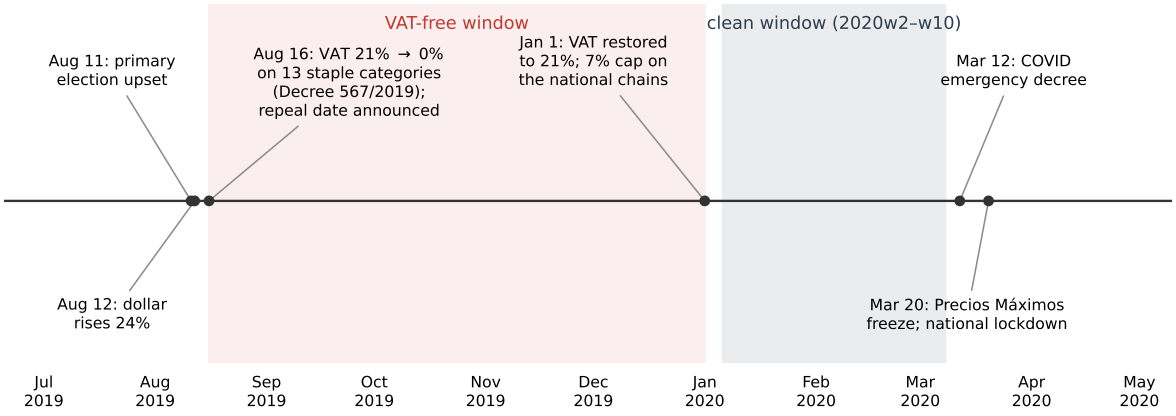


FIGURE 1. Timeline of the episode

Dates of the events described in this section. The VAT was removed by Decree 567/2019, signed on August 15, 2019, and effective on its publication the following day, with the expiry date of December 31 announced the same day; it was restored on January 1, 2020, accompanied by the negotiated 7 percent cap on the price increases of the national chains. The clean window is the post-repeal period free of both the pandemic and the Precios Máximos freeze (Section 2.3).

plain milk was already exempt.[2] The other basic foods kept their usual rates. According to the National Statistical Institute, the categories with a temporary zero rate accounted for 26 percent of total food expenditure. Importantly, the cut applied only at the register: supermarkets could still recover the VAT they paid their suppliers, either as a refund or as a credit against other taxes. The two main Argentine newspapers carried the cut on their front pages the day after it was announced and the repeal the day after it took effect, and supermarkets advertised the cut in their aisles with flyers and price tags, some of them required by the government.

2.2. Treatment and Control Groups

The left column of Table 1 lists the goods the policy covered and the right column lists the untreated goods that serve as their controls. Sunflower, corn and mixed oils had their VAT cut but olive, soy and canola oils did not, and tea and yerba mate were treated

[2]Argentina computes its extreme poverty line from the cost of the Basic Food Basket, which also enters the Consumer Price Index.

TABLE 1. Classification of data into treatment and control

Treatment (temporary 0% VAT)	Control (standard 21% VAT)
Cooking oils (sunflower, corn, mix)	Other cooking oils (olive, soy, canola)
Rice	Rice-based meals
Dried pasta	Breakfast cereal
Tea, yerba mate, and mate cocido	Coffee
Sugar	Salt
Canned vegetables and beans	Herbs, spices, and seasonings
Canned fruits	Dulce de leche
Corn flour (*polenta*)	Jam and jelly
Wheat flour	Other flours
Fluid milk (whole or skim)	Crackers, biscuits, toasts, puddings
Yogurt (whole or skim)	Chocolate
Eggs	Mayonnaise
Bread	Vinegar
Breadcrumbs and batter	Dried legumes and beans

The split of the data into treatment and control categories. Treatment categories are determined by the Annex to Decree 567/2019; control products are the remaining food categories in the data. The decree's single item for canned fruit, vegetables and legumes appears as two rows because the data classify canned fruit and canned vegetables as separate categories, which is why the column lists fourteen rows for thirteen items. The non-food control group used for the SUTVA checks (Appendix B) is not listed.

but coffee was not. Comparing treated goods with close substitutes that were not treated removes shocks other than the VAT change.

A second control group of non-food items lets us check that the control group itself is unaffected by the policy, the stable unit treatment value assumption (SUTVA). The assumption fails if consumers substitute from control goods toward treated goods: if tea becomes cheaper after the VAT cut, some coffee drinkers may switch to it. Demand for the treated goods would then rise, and with it their prices, biasing the estimated price decrease toward zero. Non-food items are not substitutes for the treated goods, so substitution from that control group is not a concern.

2.3. The Repeal and the Post-Repeal Window

The new administration did not extend the VAT holiday, but it regulated the return of the 21 percent rate on the goods the cut had covered. The rate went back to 21 percent, and

the government limited the accompanying price increase with a negotiated cap on some categories.[3] Goods under the cap could rise by at most 7 percent when the VAT returned, and goods outside it, canned fruit for example, could rise freely. Importantly, the cap applied only to the national supermarket chains, so regional chains and single-store retailers could adjust their prices freely, and there were no caps or ceilings on prices when the VAT cut was implemented. For this reason, tests that use the VAT increase as a shock restrict attention to the uncapped categories, and we report the capped categories separately throughout.

Three features shorten the window over which prices after the repeal can be read: the price caps just described (which we exclude by construction), the onset of the COVID-19 pandemic, and the Precios Máximos price freeze on essential goods introduced in March 2020. Precios Máximos (Resolution SCI 100/2020, issued March 19, 2020, and published March 20) rolled supermarket prices of essential goods back to their March 6 levels for thirty days, later extended. The national lockdown began the same week, and the COVID emergency decree is dated March 12. The post-repeal window free of the pandemic and the freeze is therefore the first ten weeks of 2020: weekly analyses use 2020w1–w10, with the transition week 2020w1 excluded from level comparisons, and monthly store-level analyses end in March 2020. We call this the clean window and use it in Sections 5 and 6.

2.4. External Validity

Our results come from an economy with high inflation. Consumer prices in Argentina rose 53.8 percent over 2019 and food prices 56.8 percent, the highest rates since 1991 according to the National Statistical Institute. Prices in our own control basket rose

[3]The caps were negotiated, not legislated: they were set at a meeting of the government, producers and the supermarket chains on December 31, 2019, announced the same day, and enforced through the government's price-monitoring system. Benzarti, Garriga, and Tortarolo (2024) document the agreement and its category-level detail. That the number was agreed rather than imposed matters for Section 7.4, where the cap serves as a public number the signatories, and only the signatories, had committed to.

about 19 log points between June–July and September–October 2019 alone. Firms in this economy reprice often, and a store could pass the whole VAT cut through in real terms by leaving its nominal price unchanged for three months.

The two closest studies of price setting under high inflation, Gagnon (2009) on Mexico and Alvarez et al. (2019) on Argentina, find that their estimates for the low-inflation years of their samples are broadly consistent with studies of the United States and the euro area. In low-inflation economies, Benzarti et al. (2020) document that prices respond roughly twice as much to VAT increases as to VAT decreases, that the resulting gap persists for several years, and that a similar asymmetry holds across all VAT changes in the European Union between 1996 and 2015. Facts 5 and 6 (Sections 5.6 and 5.7), which Benzarti, Garriga, and Tortarolo (2024) document for this episode, show the same asymmetry and hysteresis in one large reform. Beyond reproducing these two facts, this setting lets us tell their explanations apart. The control goods are nearly identical to the treated goods, the same barcode is observed in many stores at once, and the shock is large enough that the response shows up within weeks rather than having to be inferred from a slow-moving average.

Two features of the setting may make it easier for sellers to coordinate on a published number here than elsewhere, though neither bears on whether the shock left a trace. First, when prices move every few weeks, a public announcement reaches firms that are about to set a new price anyway. A published number is then plausibly a more useful focal point than in an economy where the typical price is a year old and firms pay little attention to inflation (Coibion, Gorodnichenko, and Kumar 2018). Second, Argentina has a long history of price agreements negotiated between the government and large retailers, so its sellers are used to coordinating on published numbers, and Section 7.4 uses one such agreement.

3. Data

Our data are retail scanner data from the marketing consultancy Scentia LLC, which collects the prices and quantities recorded at the till in chain and independent supermarkets. Scanner data are better suited to our question than survey data because they are higher-frequency, more precise, and less prone to measurement error.[4] The supermarkets in our data typically offer a wide variety of food and household products organized into aisles.

For every barcode the database records the week, the unit price paid at the register net of discounts, the quantity and volume sold, a detailed product description, the brand, the manufacturer and the supermarket's name. Store addresses exist for the chain stores only. All products in the dataset are classified into broad categories, such as oil, coffee or rice, which are themselves subdivided into subcategories, such as sunflower oil, corn oil, ground coffee or coffee pods, and come with detailed descriptions, such as Nescafe Gold Intense Instant Coffee Jar 200g. These variables allow us to classify products into treatment and control groups precisely, which matters because some treatments are at the barcode level, as shown in Table 1.

The store-level dataset contains 4.4 million observations. It starts in January 2019 and ends in December 2020, which covers the period when the VAT changes were implemented (August 16, 2019, to January 1, 2020). We use it through March 2020, the month of the price freeze (Section 2.3). We observe the price of 11,574 individual items ("Nescafe Gold Intense Instant Coffee Jar 200g") spanning 23 categories ("Coffee"), 1,061 brands ("Nescafe Dolce Gusto") and 455 manufacturers ("Nestle") in 173 distinct supermarkets in the province of Córdoba. These counts describe the raw file; Table 2 gives the samples we use.

[4] For example, these data include both sales made with and without receipts, the latter being a relatively common practice in independent supermarkets.

The paper draws on three panels built from this source. The first is a weekly panel of the national supermarket chains, aggregated to the barcode-by-area level across ten areas and running from the first week of 2019 through the end of 2020. It does not identify stores, and it is the panel behind the weekly event studies of Sections 5 and 6. The second is the monthly store-level panel for Córdoba just described, which follows the same barcode across individual stores and classifies each store as a national chain, a regional chain or a single-store retailer. It is the only panel in which we observe stores one by one, and it is the panel behind Facts 3 and 4, the store-type results of Fact 6 and the store-level tests of Section 7. We use it in a balanced and an unbalanced version, and Appendix C lists which results rest on which.

The third is the national monthly panel of Benzarti, Garriga, and Tortarolo (2024), built from the full national collection of more than 3,000 supermarkets and aggregated to cells of barcode by region by store type, with 8,815 barcodes over 2019–2020. It distinguishes the large chains from smaller stores but does not identify stores, and Fact 5 uses it because it is the only panel with store type at the national level. Store-level detail is available for Córdoba only, and the national data enter in aggregated form. Table 2 summarizes the first two, and treated goods account for a little over half of pre-cut food spending in both.

4. Conceptual Framework and Predictions

This section describes the Calvo and menu cost models in their simplest versions, following the notation and exposition of Nakamura and Steinsson (2013). Benzarti, Garriga, and Tortarolo (2024, Appendix B.2) use the same two models to ask whether price rigidities can explain the asymmetric pass-through in this episode, and the predictions below build on that discussion. Their simplicity makes them extreme cases of how prices adjust: in the Calvo model aggregate conditions have no effect on which firms change

TABLE 2. The weekly and store panels

	Weekly chain panel	Córdoba store panel	
		Balanced	Unbalanced
Frequency	Weekly	Monthly	Monthly
Coverage	2019w1–2020w52	2019m1–2020m3	2019m1–2020m3
Unit of observation	Barcode × area	Barcode × store	Barcode × store
Stores	—	172	173
National chains	—	90	90
Regional chains	—	56	56
Single-store retailers	—	26	27
Areas	10	1 (Córdoba)	1 (Córdoba)
Food barcodes	4,530	1,687	7,668
Treated	2,032	797	3,719
Food categories	24	16	16
Cells	24,970	53,634	197,365
Balanced across all weeks	16,670	—	—
Treated share of pre-cut food spending	53%	53%	59%

The weekly chain panel follows barcodes in the national supermarket chains at the barcode-by-area level and is the panel behind the event studies of Sections 5 and 6; it does not identify stores. The Córdoba store panel follows the same barcodes store by store and classifies each store as a national chain, a regional chain or a single-store retailer, the split used in Facts 3, 4 and 6 and Section 7; it also contains a non-food control group of 636 barcodes used for the SUTVA checks. Spending shares are computed over the pre-cut months of 2019 (through week 31 and month 7 respectively); within treated spending on the weekly panel, 79 percent is on goods later covered by the negotiated cap. Fact 5 additionally uses the national monthly panel of Benzarti, Garriga, and Tortarolo (2024), which has store type but does not identify stores. The Córdoba store panel is used in two versions (Appendix C): the balanced one keeps store–barcode cells with a usable price in every month, and the unbalanced one keeps every store and every cell with a usable price in at least one month.

prices or how many do, and in the menu cost model aggregate shocks are the only thing that determines both. Table 3 at the end of the section collects the predictions of each model for the six facts of Section 5.

4.1. The Calvo (1983) Model

Each firm gets to change its price with probability α in every period, so the log aggregate price level p_t evolves as

$$p_t = \alpha p_{it}^{*} + (1 - \alpha) p_{t-1} \tag{1}$$

where p_{it}^* is the log price set by the firms that reprice in period t. Under standard assumptions (labor as the only variable input with a linear technology, perfectly elastic labor supply, iso-elastic demand, and nominal output following a random walk), the optimal reset price equals current nominal output, $p_{it}^* = m_t$ (Appendix A gives the derivation). Therefore:

$$p_t = \alpha m_t + (1 - \alpha) p_{t-1} \tag{2}$$

This equation describes how log aggregate nominal prices evolve. They respond sluggishly to aggregate shocks because the firms that do not get to change their price keep last period's price, and how sluggish they are depends on α, the share of firms that change prices in every period.

4.2. The Menu Cost Model

The menu cost model follows Caplin and Spulber (1987). With the same notation, assume that nominal output $m_t = y_t + p_t$ always rises, that is, inflation is relatively high. Unlike in the Calvo model, firms can change prices in every period but pay a fixed cost whenever they do. Firms then follow an sS policy: they let the relative price drift down to a trigger s and raise it by $S - s$. The firms that change prices are not drawn at random, as in the Calvo model, but are those furthest from their target price, so they change prices by more than Calvo firms and the aggregate price level responds more. This is the "selection effect" of Golosov and Lucas (2007).

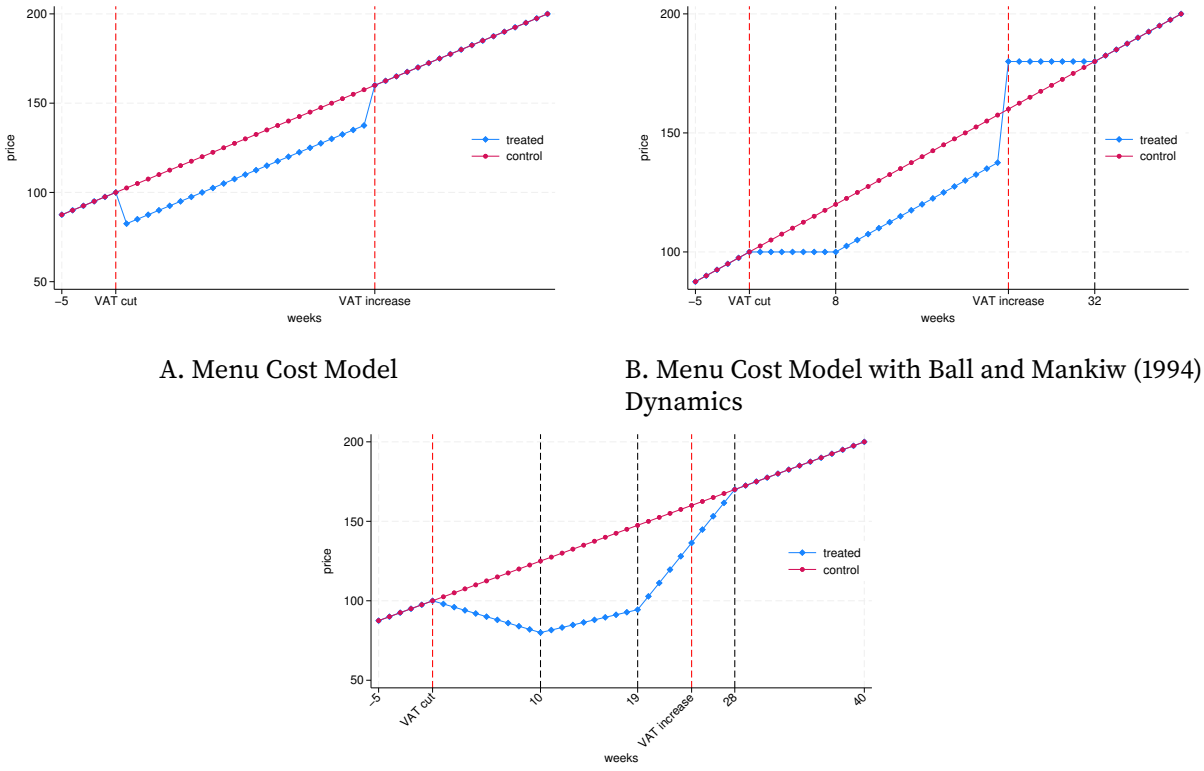
4.3. Predictions

Adjustment path. The two models predict different magnitudes for the response of prices to the VAT changes, but we do not test them on magnitudes, because the size of the menu cost and the share α of firms that can change prices in a given period

would both move the predicted size of the response. We test instead the patterns prices follow after each change, which do not depend on the parameters. The clearest is the path of adjustment: aggregate prices adjust in steps in the Calvo model and at once in the menu cost model. In the Calvo model only a share α of firms adjust prices in each period, so aggregate prices move α of the way to the new optimal price in the first period, $1-(1-\alpha)^2$ of the way in the second, and so on, so prices fall in steps after the VAT cut and rise in steps after the VAT increase. Firms may also begin responding before the increase, because it was pre-announced, as we discuss below.

The menu cost model allows two paths, depending on how the inflation rate, the menu cost and the size of the shock compare. Firms may respond to the VAT cut by doing nothing and letting inflation lower real prices until they reach the new optimal price, as in Ball and Mankiw (1994), which saves the menu cost. Once that point is reached, treated prices rise at the same rate as control prices. At the VAT increase, by the same logic, firms may raise prices above the control group's, because they expect inflation to lower real prices afterwards and bundling several price changes into one saves menu costs. Treated nominal prices then overshoot control prices and stay put until the control group catches up, which takes a number of periods that depends on the menu cost and the inflation rate, after which the two groups have similar price levels.

For these dynamics to hold, there needs to be trend inflation, which is the case in our setting, and incurring the menu cost should be more costly than being mispriced while inflation closes the gap. The menu cost is paid once. It has to be weighed not against the size of the shock, which is in different units, but against the profit a firm loses by staying mispriced while inflation at the prevailing rate erodes a gap of 19 log points, a loss that builds up over the several months this takes. For a shock of this size that cumulative loss is large relative to any ordinary menu cost. In the estimated model of Section 6, the inaction band (the range of price gaps a firm leaves uncorrected) is wide enough to contain the tax shock only at menu costs about four times the one we



A. Menu Cost Model

B. Menu Cost Model with Ball and Mankiw (1994) Dynamics

C. Calvo Model

FIGURE 2. Predicted price paths under menu cost and Calvo pricing

This figure plots the predictions of the menu cost model without the Ball and Mankiw (1994) dynamics (Panel A), with the Ball and Mankiw (1994) dynamics (Panel B), and of the Calvo model (Panel C) for the treatment and control groups.

estimate. The Ball and Mankiw (1994) dynamics are therefore unlikely, though possible, so we consider the predictions of menu costs with and without them and test which hold.

Figure 2 plots hypothetical weekly price paths for the treatment and control groups under each model. Under the menu cost model without the Ball and Mankiw (1994) dynamics (Figure 2A), the two groups follow parallel trends before the cut, which is the identifying assumption of our empirical approach. Treated prices fall immediately and fully at the cut, follow the control trend through the holiday, and return to the control level in a single period at the repeal.

With the Ball and Mankiw (1994) dynamics (Figure 2B), the cut can be passed through

in two ways: firms cut nominal prices, or they leave nominal prices fixed until inflation lowers real prices, which saves the menu cost. Which prevails depends on the size of the menu cost. The figure assumes a menu cost large enough for nominal prices to stay fixed for eight periods, after which treated prices follow the control trend at a lower level. At the repeal (period 24) treated prices rise immediately and may overshoot the control group, because firms bundle several price changes into one, until inflation erodes the overshoot (periods 25 to 32) and the two groups converge. Prices do not respond before the repeal.

Under the Calvo model (Figure 2C) aggregate prices fall in steps after the cut, over periods 1 to 10 in the figure, at a pace set by α, and then follow the control trend. Because the repeal is anticipated, some firms raise prices before it takes effect if they expect no other chance to do so in time, which is more likely when α is small, so the increase begins early (period 19) and is itself staggered, taking several periods (24 to 28) to reach the control level.

Telling the menu cost model without the Ball and Mankiw (1994) dynamics from the Calvo model is straightforward: the first predicts immediate adjustment, the second staggered adjustment with some anticipatory response. If the Ball and Mankiw (1994) dynamics hold, three differences remain. First, nominal prices fall under the Calvo model but stay constant under the menu cost model for several periods after the VAT cut. Second, some firms raise prices before the VAT increase under the Calvo model but not under the menu cost model. Third, prices overshoot after the VAT increase under the menu cost model but not under the Calvo model. The staggered decrease is a defining feature of the Calvo model. In the menu cost model, with or without the Ball and Mankiw (1994) dynamics, any decrease in prices at the VAT cut is immediate, since firms have no reason to wait, and otherwise prices are held fixed.

The underlying price distribution. A second test uses the distribution of prices across sellers instead of the time series. The intuition is easiest with no inflation, an assumption that is easily relaxed and that sets the Ball and Mankiw (1994) dynamics aside. Section 5.2 plots the distribution of treated and control prices in the periods around the VAT changes, and two features tell the models apart. In a menu cost model the whole treated distribution shifts left after the cut and right after the increase, relative to control, stays there, and keeps a single peak. Under Calvo pricing a second peak appears to the left of the main peak after the cut and to the right of it after the increase, small at first, with mass α, the share of firms able to change prices, and growing period after period as more firms adjust. The first feature is therefore the shape of the distribution, single peaked under menu costs and double peaked under Calvo. The second is how it evolves: under menu costs the bulk of the change comes in the first period, so later distributions look alike, and under Calvo the shape keeps changing for several periods after each VAT change.

Frequency of price changes. The Calvo model assumes that the timing of price changes is random. The frequency of price changes is then constant over time and does not respond to economic shocks, although the magnitude of the changes does. Treated and control frequencies should therefore stay similar through both VAT changes. A menu cost model predicts instead that the frequency itself responds: the frequency of decreases spikes at the cut, the frequency of increases spikes at the repeal, and both revert to baseline shortly after. The Ball and Mankiw (1994) dynamics change the prediction at the cut. Nominal prices are held fixed while inflation erodes real prices, so the frequency of *increases* in the treated group falls after the cut. The repeal still brings a spike of increases, followed by several periods of frozen nominal prices. Appendix A lists the predictions period by period.

Price dispersion. The Calvo model predicts more price dispersion than the menu cost model. In the Calvo model a price can drift far from its optimal level if the firm has not had a chance to change it for some time. In the menu cost model the gap between the optimal and the posted price cannot grow past a limit set by the menu cost, because beyond that limit the firm finds it worth paying to reprice. A large literature studies this question by estimating how price dispersion across products, within a narrow category or across categories, changes with inflation.[5] That approach assumes that differences in optimal prices across products are unrelated to inflation. Cheerios and Froot Loops are both breakfast cereals, for example, and the approach assumes that the difference in their optimal prices stays constant as inflation changes, although General Mills and Kellogg's may price and respond to inflation differently. We instead measure the dispersion of prices of *identical* products across supermarkets, the standard deviation of log prices across stores for the same barcode.

Under the Calvo model, dispersion in the treatment group rises after each VAT change, because a random subset of firms has adjusted and the rest have not. The rise is hump-shaped. Dispersion widens while adjustment spreads through the population and converges back to that of the control group once most firms have moved. Because the repeal is anticipated, the pattern may start before it takes effect, which weakens the repeal-based version of the test.

The menu cost model predicts the opposite. A shock prompts more firms to adjust because more posted prices are out of line, and by the Golosov and Lucas (2007) selection effect the firms most likely to adjust are those whose prices are furthest out of line, which are also the ones that contribute most to dispersion. Dispersion should therefore fall substantially after a VAT change and then return gradually to that of the control group. Under the Ball and Mankiw (1994) dynamics, the fall at the cut is delayed by the

[5] In practice, and as pointed out by Nakamura and Steinsson (2013), researchers do not estimate price dispersion, but instead the dispersion of inflation rates, because of lack of data.

TABLE 3. Predictions of the two models for the six facts

	Calvo	Menu cost	Menu cost + Ball and Mankiw (1994)
1. Adjustment to the cut	Staggered, step-wise	Immediate and full	None for several periods
2. Early increase before the repeal	Yes (if α small)	None	None
3. Frequency of price changes	Unresponsive to shocks	Spikes at both shocks	Increases fall after cut; spike at repeal
4. Dispersion, identical goods	Rises, hump-shaped	Falls (selection)	Flat, then falls
5. Cut vs. repeal symmetry	Symmetric	Near-symmetric (asymmetric band under trend inflation, Section 6)	Asymmetric: overshoot, then convergence
6. Long-run treated–control wedge	None	None	None

Predictions of each model for the six facts documented in Section 5. Each entry follows from the exposition in Section 4.3: rows 1 and 2 from the three panels of Figure 2, row 3 from the enumerated frequency predictions, row 4 from the price-dispersion predictions and the Golosov and Lucas (2007) selection effect, and rows 5 and 6 from the overshoot-then-convergence path under Ball and Mankiw (1994) dynamics. In row 5, the Calvo entry describes firms that do not anticipate the repeal. Row 3 refers to the frequency of all price changes; the share of changes larger than five percent, which we measure, also responds under Calvo, because a shock makes each change larger (Section 5.4). How much dispersion rises in row 4 depends on how often Calvo firms reprice. Table 4 repeats it in short form, with the data column added.

spell of frozen nominal prices, and the fall at the repeal is not. Appendix A lists the predictions period by period.

4.4. The Predictions Table

Table 3 collects the predictions of the two models for each of the six facts.

5. Six Facts About Price Adjustment

5.1. Empirical Strategy

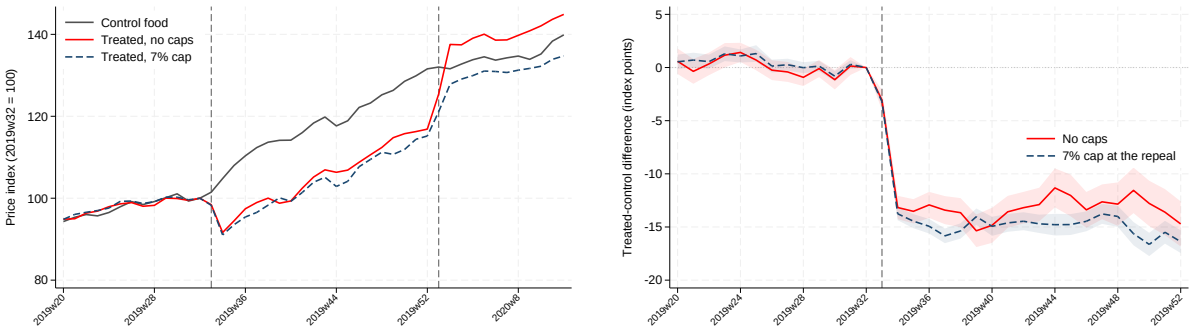
Our main specification is the following dynamic difference-in-differences model:

$$Y_{it} = \alpha_i + \gamma_t + \sum_{t \neq 2019\text{w}32}^{2020\text{w}10} \beta_t D_{it} + \epsilon_{it} \tag{3}$$

where Y_{it}, our outcome, is the price of barcode i at time t. Y_{it} is normalized to 100 for each barcode i in week 32 of year 2019. D_{it} is equal to one if barcode i is treated in week t and zero otherwise. The main coefficient of interest is β_t, which estimates the average price difference between the treatment and control groups across all barcodes at time t, relative to week 32 of year 2019.[6] The specification runs at the barcode–area level because the weekly panel does not identify stores: a cell is a barcode in one of ten areas, each cell is normalized to 100 in 2019w32, and α_i stands for barcode and area fixed effects. The monthly Córdoba store panel does identify stores, and the same specification at the store–barcode level reproduces the path on uncapped goods, with a difference between the treated and control groups of -9.2 log points at the trough in October 2019 and +9.8 in February 2020.[7] Table 5 extends the event window through the end of 2020.

[6] Treatment was assigned at the level of product types, so we also report inference that relies only on how treatment was assigned. We reassign treated status at random across the pure subcategories of the weekly panel, those whose goods are all treated or all control, and the observed VAT-free wedge is more extreme than in all of 20,000 reassignments. Appendix B reports the full exercise and the leave-one-out results, and explains why this design cannot test the uncapped post-repeal overshoot.

[7] With store–barcode cell and category-by-month fixed effects; standard errors 2.1 and 3.3, and neither figure moves under two-way clustering on barcode and store. In the weekly panel the clustering choice matters little: adding area as a second cluster dimension raises the standard error of the VAT-free wedge from 0.66 to 0.79 and leaves that of the clean-window wedge essentially unchanged. Weighting cells by pre-cut spending deepens the cut leg to -15.3 index points and reduces the clean-window overshoot to +3.4 (s.e. 1.6), so the overshoot is smaller on high-turnover cells, not absent.



A. Price levels

B. Difference in differences

FIGURE 3. The response to the VAT cut is immediate

Weekly chain panel. The left panel plots the mean weekly price index for control food, treated goods the repeal cap did not cover, and treated goods it covered, with every barcode–area cell normalized to 100 in 2019w32, the last full week before the cut. The right panel plots the weekly event-study coefficients from specification (3) over 2019w20–w52, with 95 percent confidence bands and standard errors clustered on barcode. Vertical lines mark the VAT cut and, in the left panel, its reinstatement; the cap distinction matters only at the repeal, since there were no caps at the cut.

5.2. Fact 1: Adjustment to the Cut Is Immediate

We find that the response to the VAT cut is immediate: prices in the treatment group break sharply in the week the cut is implemented, with no step-wise convergence to the new level (Figures 3A and 3B; Figure 9 extends the window to the end of 2020). Measured from the last full week before the cut, the uncapped treated–control wedge is −3.1 index points in the week the cut takes effect and −13.2 one week later. It then sits flat at about −13.2 for the following nineteen weeks. The entire response is contained in the first two weeks, where under Calvo pricing it would build up step by step over many more.

The whole distribution of prices moves at once. Measured relative to the control-food index, treated uncapped prices form a single peak at the control level two weeks before the cut, and the whole of it shifts between the week of the cut and the week after (Figure 4). The share of cells more than 5 percent below the control index is 14 percent two weeks before the cut, 39 percent in the week the cut takes effect and 83 percent one

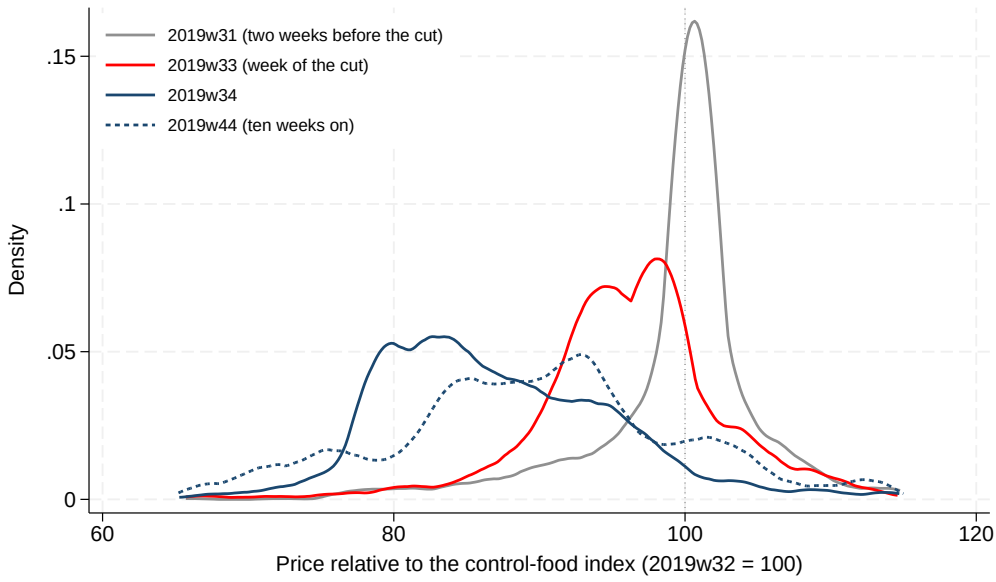


FIGURE 4. The whole distribution moves within two weeks

Kernel densities of the price of treated uncapped cells relative to the mean control-food index in the same week, scaled so that 2019w32 = 100, on the weekly chain panel; a cell is a barcode–area pair, and the same 1,304 cells appear in every week shown. Staggered adjustment would move mass from the old peak to a new one over many weeks; instead the entire distribution shifts between the week of the cut and the week after, and sits there ten weeks later. The rightward drift of the median between 2019w34 and 2019w44 is the erosion of the log wedge as the price level rises, discussed in Section 5.6.

week later, and it is still 72 percent ten weeks on. Under staggered adjustment a second peak would form and grow week by week as more firms repriced (Section 4.3), and no such peak forms.

5.3. Fact 2: No Anticipatory Response to the Pre-Announced Repeal

We find no anticipatory response in the treatment group. The repeal date was announced when the holiday began, and both events were highly salient (Section 2). Under Calvo pricing, a firm that gets its chance to reprice shortly before a known tax increase should therefore raise its price early, since its next chance may come only after the tax is back. Over the four weeks preceding the repeal the uncapped wedge runs −11.6, −12.8, −13.6 and −14.7, moving away from zero, not toward it. In levels the treated index rises 2.1

points across those weeks against the control group's 5.2, so treated prices are inflating more slowly than control right up to the day the tax returns. The drift away from zero is itself no larger than the ordinary movement of the wedge: it falls 1.9 points from 2019w48 to 2019w52, and across the VAT-free weeks before 2019w48, four-week movements of the same wedge have a standard deviation of 1.7. Menu cost pricing predicts no response before implementation, and none appears.

5.4. Fact 3: The Frequency of Price Changes Responds to Both Shocks

We find that both shocks raise the share of products whose price changes, the cut through large falls and the repeal through large rises. We measure participation, the share of cells whose price moves by more than five percent, on the Córdoba monthly store panel. At the cut we pool August and September, because the cut fell on August 16, in the middle of a month, and split a single synchronized move across two months, whereas the repeal fell on January 1 and sits entirely inside one. Comparing single months would therefore understate the cut by construction. Over August and September together, 50.3 percent of treated cells posted a fall larger than five percent against 14.6 percent of control cells. In January the direction reverses: 75.3 percent of treated cells posted a rise larger than five percent against 28.6 percent of control. In a barcode-level event study measured against the average of the months before the cut, the share of treated cells posting a large fall exceeds the control share by 21 percentage points in August and 24 in September. The share posting a large rise exceeds it by 49 percentage points in January.[8] These pooled figures are dominated by the chains: at the cut 59.4 percent of treated cells in the national chains posted a large fall against 8.0 percent among single-store retailers, a gap that is smaller at the repeal but still there.[9]

[8] Barcode and month fixed effects, standard errors clustered on barcode: 0.210 (0.008), 0.240 (0.009) and 0.488 (0.011).

[9] Seller type is classified differently in the two panels and the groups are not identical. The national monthly panel separates large chains from smaller stores by data source, which is the classification used

The same measure also responds to inflation away from the two events, which checks the measure and gives the model of Section 6 the statistic it is calibrated to match. We regress monthly participation on monthly inflation for control food, whose tax never changed, across the thirteen months for which both are available. The frequency of large increases rises with inflation with a slope of 5.68, and the frequency of large decreases falls with a slope of -0.64.[10] Each extra percentage point of monthly inflation raises the share of cells posting a large increase by about 5.7 percentage points and lowers the share posting a large decrease by about 0.6. Stores in this economy reprice more often when inflation is higher, with more increases and fewer decreases. A state-dependent model predicts this. A constant hazard, under which a firm's chance of repricing is the same every month whatever inflation does, produces part of it, because higher inflation makes each price change larger and pushes more of them past five percent. In the model of Section 6, pure Calvo pricing gives a slope of 2.7 to 4.2 for large increases, depending on how often firms reprice, against 5.4 in the estimated menu cost model and 5.7 in the data.

The weekly panel gives the same measure week by week and sharpens the timing (Figure 5). In an ordinary week about 10 percent of cells post a fall larger than five percent and 15 percent a rise, a floor that consists mostly of temporary sales (Appendix B). In the week the cut takes effect and the week after, 30 and then 69 percent of treated uncapped cells post a large fall, and in the first two weeks of January 62 and then 74 percent post a large rise. Control food spikes at neither event, and its one visible wave, elevated large increases in the weeks after the August depreciation, is the control basket absorbing the cost shock, the same movement that appears in its price level in Figure 3A.

in Benzarti, Garriga, and Tortarolo (2024); the Córdoba store panel separates national chains, regional chains and single-store retailers. We name the panel wherever the distinction matters. We use the finer split here because the group usually pooled as "independents" is 89 percent regional chains by cells, and the single-store retailers behave quite differently from them. The two papers therefore split the same sellers for different purposes. Incidence turns on which sellers were monitored, price adjustment on how large their adjustment frictions are.

[10] Standard errors 0.14 and 0.07.

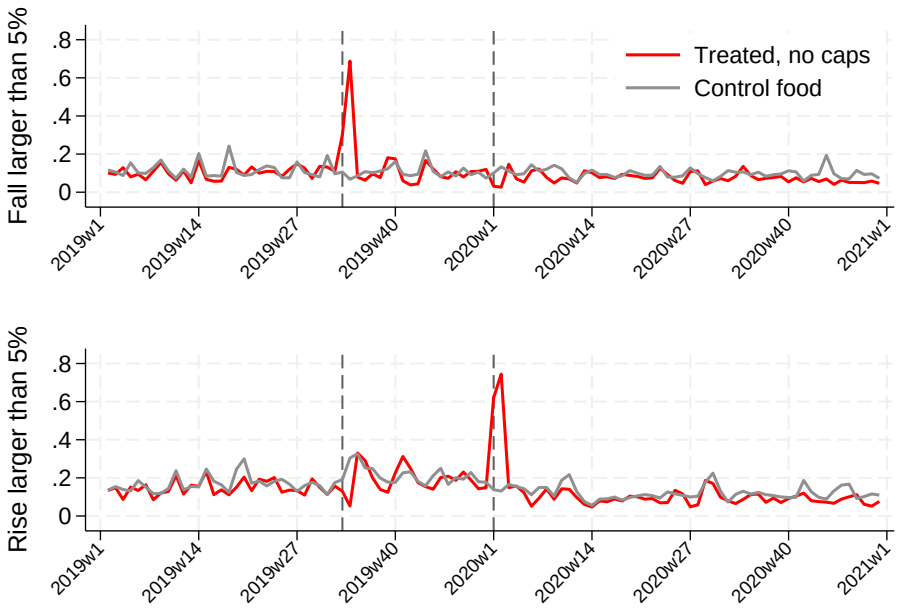


FIGURE 5. The frequency of large price changes spikes at the two tax events and nowhere else

Share of barcode–area cells whose week-over-week log price change is below −5 percent (top panel) or above +5 percent (bottom panel), on the weekly chain panel, for treated cells on goods the repeal cap did not cover and for control food. Vertical lines mark the VAT cut and its reinstatement.

5.5. Fact 4: The Dispersion of Identical-Product Prices Compresses Relative to Control

We find that when the cut hits, treated dispersion stops tracking control. In the Córdoba store-level panel we measure, for each barcode and month, the standard deviation of log prices across the stores selling that exact barcode (balanced store–barcode panel, barcodes sold in at least five stores). Over the VAT-free months, September to December 2019, treated dispersion is essentially flat while control-food dispersion rises nine percent. Relative to control, treated dispersion falls by 0.75 log points, and by 1.8 log points with category-by-month fixed effects, which compare treated and control goods in the same category and month. We use the within-category specification throughout, because biscuits alone are 59 percent of control cells and relative prices across categories

move a great deal in this period.[11] The within-category estimate is about a fifth of the treated pre-period standard deviation, and the post-repeal window shows no significant dispersion response.

Before the cut, treated and control dispersion move together. Without category fixed effects and with July 2019 as the base month, the pre-period placebo coefficients are within ±0.4 log points and insignificant. Figure 6 plots the same contrast month by month, measured against the average of the pre-cut months so that each of them remains visible, and the months before the cut are small and mostly insignificant. The exceptions are July 2019, at −1.3 log points, whose confidence interval excludes zero, and January 2019, at +0.7, whose interval only just does. Our robustness checks flag July as atypical for the control basket, which is why the figure uses the pre-cut average as its base and not any single month. The July value is a caution about that month, not a trend, and excluding July from the pre-period strengthens the estimate, because July's deviation runs in the same direction as the treatment effect and is part of the baseline the effect is measured from.

The compression comes from the chains, and splitting by store type shows why. Measuring dispersion among stores of the same kind, the difference-in-differences is −0.4 log points among the national chains, −0.1 among regional chains, and +1.1 among single-store retailers, where dispersion widens.[12] At first sight this looks like the menu cost prediction holding for chains and the Calvo prediction holding for single-store retailers. But selection, in which the stores that reprice are the ones furthest from their target, compresses dispersion only when adjustment is widespread. Where the menu cost is large, a store tolerates a wide gap from its target before repricing (a wide inaction band), so a common shock pushes only a minority of stores to reprice. Those

[11] The treated standard deviation moves from 0.0912 to 0.0926 and the control one from 0.0976 to 0.1065. The difference-in-differences is −0.0075 (s.e. 0.0019), and −0.0178 (s.e. 0.0060) within category; the post-repeal window shows no significant dispersion response, at −0.0067 (s.e. 0.0065) within category. Excluding July 2019 from the pre-period strengthens the estimate to −0.0196 (s.e. 0.0060).

[12] In logs: −0.0042 (s.e. 0.0019), −0.0010 (s.e. 0.0018) and +0.0105 (s.e. 0.0035).

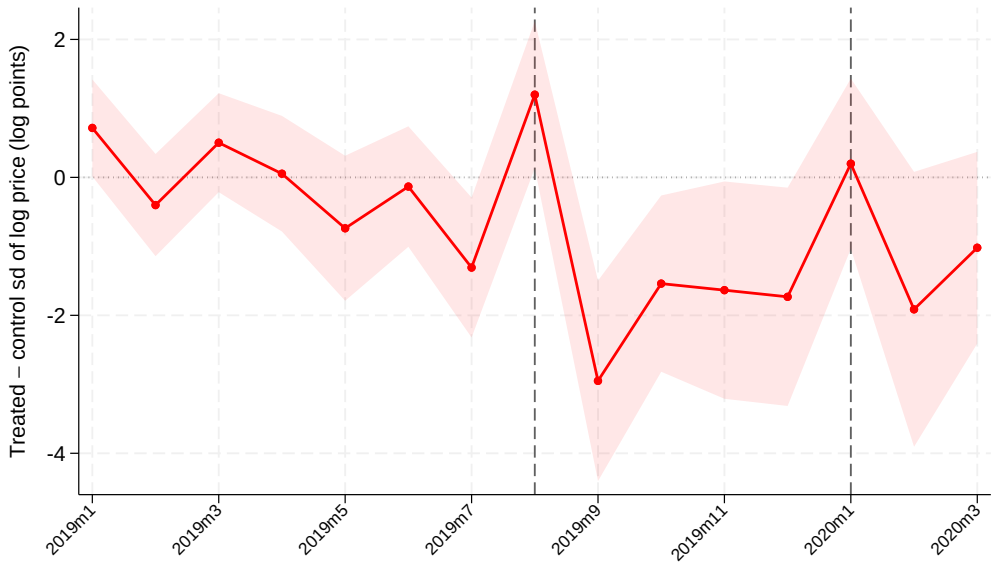


FIGURE 6. The dispersion of identical-product prices stops tracking control at the cut

Monthly event-study coefficients on the treated–control difference in the standard deviation of log price across the stores selling a given barcode, on the Córdoba store panel, with category-by-month fixed effects; the shaded band is the 95 percent confidence interval. The unit of observation is a barcode–month and the sample is the balanced panel of barcodes sold in at least five stores. Vertical lines mark the VAT cut and its reinstatement. The omitted period is the pooled pre-period rather than a single month, so the coefficients are relative to the average of January–June 2019, in log points.

stores bunch at the new price while the rest keep their old gaps, and the standard deviation rises. Single-store retailers fit this profile. Their baseline dispersion for the same barcode is 0.140, against 0.106 for the national chains and 0.066 for the regional ones. Only 8 percent of them post a large decrease at the cut, against 60 percent of national chains. When they do move, they move 11.5 log points, almost as far as the chains' 13.1.

Which stores adjust is also informative: under a menu cost they are the ones furthest from target, and under Calvo they are drawn at random. We find that in both groups the stores that cut are the expensive ones. Regressing an indicator for posting a large decrease at the cut on the store's average pre-cut gap between its own price and the mean price of the same barcode across stores gives coefficients of 0.50 for the national

chains and 0.29 for the single-store retailers.[13] A national chain store priced 10 log points above the mean is about 5 percentage points more likely to post a large decrease, and a single-store retailer about 3 points more likely. At the repeal the sign reverses among the retailers, so the stores that raise are the cheap ones. The coefficients would not be zero under a constant hazard either, because a store that drew a high price is more likely to post a large decrease whenever it does reprice, so the pattern favors selection without settling it.

Calvo predicts that dispersion rises after a shock, because a random subset of firms has adjusted and the rest have not, and the selection effect in a menu cost model predicts that it falls, because the firms that adjust are the ones furthest from target. The compression among chains counts against the constant hazard. The widening among single-store retailers does not count for it, because the stores that adjust there are selected too, and the same selection that compresses dispersion when the inaction band is narrow widens it when the band is wide. The two groups share selection and differ in what share of their stores a shock of this size moves.

5.6. Fact 5: Pass-Through Is Roughly Twice as Large Upward as Downward

We find, as Benzarti, Garriga, and Tortarolo (2024) do for this episode, that prices respond roughly twice as much to the VAT increase as to the VAT cut. On goods the negotiated cap did not cover, pass-through is 28 percent at the cut and 65 percent at the repeal.[14] We measure the two legs on the national monthly panel, comparing plateaus, periods in which prices had settled, so that neither leg is read while prices are still adjusting. The pre-cut plateau is June and July 2019, the VAT-free plateau is October to December, and the post-repeal plateau is February 2020, before the price

[13] Within category, clustering on barcode: +0.499 (s.e. 0.066) and +0.294 (s.e. 0.072) at the cut, and −0.421 (s.e. 0.124) for the retailers at the repeal.

[14] Standard errors 8.0 and 6.9, clustered on barcode. Three pass-through figures appear in the paper and measure different things; Appendix B reconciles them.

freeze of March. Removing and restoring a 21 percent tax moves the desired price by log(1.21) = 19.1 log points in each direction, so pass-through is the treated–control move as a share of that. A cell is a barcode in a region in a store type, and we control for category throughout. Most categories are either all treated or all control, so the treated–control contrast is identified by the five categories containing both treated and untreated barcodes (rice, oils, canned fruit, dried pasta and flour products), and on uncapped goods by the first three.[15]

Prices rose more at the repeal than they fell at the cut. We state the asymmetry as the sum of the two legs rather than their ratio. Because the cut leg is negative and the repeal leg positive, the sum is the price level at the end of the episode relative to its start, and on uncapped goods it is +7.1 log points.[16] Across all treated goods, including capped ones, the same statistic is +0.5, and the asymmetry disappears because the cap removes it.

Splitting by store type shows what the cap did, and it removes the need to set capped goods aside. The seven percent cap bound the national chains. Smaller stores selling the same goods were free to raise prices by the full amount of the tax. Among the large chains, as the national panel classifies them, uncapped goods pass through 44 percent at the cut and 92 percent at the repeal, an asymmetry of +9.1 log points. Capped goods pass through 53 percent at the cut and only 37 percent at the repeal, an asymmetry of -2.9 that runs the other way. Among the smaller stores, uncapped goods pass through 17 percent at the cut and 38 percent at the repeal, an asymmetry of +4.0, and capped goods pass through 15 percent at the cut and 24 percent at the repeal, an asymmetry of +1.8.[17] The same capped good, in the same month, ends the episode 4.7 log points

[15] These five categories hold 6,133 of the 27,798 cells, and the estimates for all treated, uncapped and capped goods are the same to two decimals when run on them alone. Uncapped treated goods appear in only three of them (778 cells).

[16] Standard error 2.2; across all treated goods including capped ones, +0.5 (s.e. 0.9).

[17] Standard errors 2.9, 1.4, 2.3 and 1.0. Net of the August 2019 depreciation, the smaller stores' capped asymmetry is -0.4 (s.e. 1.1) and their difference from the large chains 4.1 (Appendix B).

higher where its seller did not sign the agreement than where it did. The cap suspends the asymmetry where it binds.

This asymmetry differs from the one in Ball and Mankiw (1994). Under those dynamics firms pass a tax cut through by holding nominal prices fixed and letting trend inflation erode the real price, so nominal prices should be flat after the cut, and instead they fell. On the weekly chain panel treated uncapped prices go from 100.0 to 98.4 to 91.6 across 2019w32–w34, a nominal decline of 8.4 index points in two weeks, a slope of -4.2 points per week while control prices were rising at +2.4.[18] Some sellers did hold their prices. They are a minority in the aggregate, but on the Córdoba store panel only 11.5 percent of treated cells outside the national chains cut the nominal price by more than five percent between the June–July and September–October averages. For the rest of those sellers, pass-through ran through frozen nominal prices and trend inflation instead (Section 7.4).

Three cautions apply. First, the ratio of the two legs is close to two on uncapped goods, the magnitude Benzarti et al. (2020) report across European Union VAT changes. But a ratio of two estimates is fragile, since a small error in the smaller one moves it a lot, so we base the claim on the sum of the two legs instead. Second, the figures here are log points, whereas the weekly event study of Section 5.2 and Table 5 are index points relative to a base week. Because the price level roughly doubled over the sample, a constant wedge in index points is a shrinking wedge in logs, and the two units should not be divided into one another. The post-repeal wedge of Section 5.7 rises in index points, so the conversion does not affect its conclusion. On the weekly chain panel the asymmetry also holds when V-shaped temporary price cuts are removed, and it sharpens, because part of the downward leg but none of the overshoot runs through promotions (Appendix B).

Third, the repeal took effect on January 1, 2020, so the repeal leg also picks up any

[18] Slopes -4.18 (s.e. 0.26) and +2.40 (s.e. 0.10).

change that comes with a new year. On the weekly chain panel uncapped goods rise against control goods by 1 to 2 log points at the New Year of 2019 and of 2021, when no tax changed, and capped goods do not. Removing a step of that size from the repeal leg would put pass-through at the repeal between 55 and 60 percent and the sum of the two legs between 5 and 6 log points. Prices would still respond roughly twice as much to the increase as to the cut.[19]

5.7. Fact 6: The Overshoot Does Not Decay

We find, as Benzarti, Garriga, and Tortarolo (2024) do for this episode, that the overshoot does not decay. Under the Ball and Mankiw (1994) dynamics, the overshoot after the repeal should be followed by nominal prices held fixed while inflation erodes real prices, leading to convergence between treatment and control. In the clean window (Section 2.3), the treated–control wedge for uncapped goods averages +5.5 index points over 2020w2–w10 and is flat to rising (Figure 7). The wedge moves +0.15 points per week, where our benchmark, the slope it would have if treated prices stayed fixed while inflation eroded the gap, is -0.32, and the difference between the two is significant at $p < 10^{-5}$. In the last three clean weeks the wedge exceeds the first three by about one point, so it rises where decay predicts a fall. The window average stays between +5.1 and +5.6 however we truncate the clean window (Appendix B).[20] Within a single area the pattern holds as well, since Córdoba alone gives +4.2. The independent Córdoba store-level monthly panel confirms it when treated goods are compared with untreated goods in the same category. In the product categories containing both treated and untreated barcodes (rice, oils, canned fruit), the uncapped treated–control wedge at

[19] The step compares weeks 2 to 4 of January with the last three weeks of December, against control goods: +1.9 (s.e. 0.4) in 2019 and +1.1 (s.e. 0.3) in 2021.

[20] Standard errors, clustered on barcode: 1.2 on the window average, 0.10 on the slope, and 0.62 on the last-three-minus-first-three difference of +1.00. The benchmark is control-group food inflation in the window, about 1.1 percent a month; the raw control index gives 1.5, and the national food index rose faster still in the same months, so the benchmark is conservative.

March 2020, measured from the last pre-cut month, is +7.4 log points, and about the same whichever pre-cut month serves as the base.[21]

The uncapped set behind the weekly estimates is thin, 265 of the 1,934 classified treated barcodes on the chain panel, and the capped categories give a second test, because the cap bound only the sellers who signed it. For the stores that were never party to the agreement the capped categories are as informative as the uncapped ones, and on the store panel they hold nine in ten of the classified treated barcodes. On capped goods the full-episode wedge, the treated–control change from the two months before the cut to the two months after the repeal, is +6.4 log points among the single-store retailers and −1.4 among the national chains the cap bound. On uncapped goods the two show +11.2 and +7.5.[22] Among the stores the cap did not bind, the overshoot is concentrated in the single-store retailers. The regional chains, which hold most of the unbound cells, end close to control, at +1.0 log points on capped goods and +3.5 on uncapped ones. The only sellers in whom the overshoot disappears entirely are the ones a price control was written to restrain. The contrast also holds against a control group that shares no category with the treated goods, the non-food basket sold in the same stores. Against that basket, the capped-goods wedge is +3.1 log points at the single-store retailers and −6.9 at the chains the cap bound (Appendix B).

Three points about measurement follow. First, promotions do not account for the wedge. Our price is a unit value, and treated goods do promote less after the repeal than before the cut, but on a series from which V-shaped temporary price cuts are removed the clean-window and late-2020 wedges are unchanged (Appendix B). Second,

[21] Standard errors 1.9 and 2.6. The raw store-level contrast is uninformative about the overshoot because the control basket's category composition dominates it.

[22] Standard errors 1.9 and 1.7 on the capped-goods wedges and 4.9 and 4.1 on the uncapped ones; 1.1 and 0.9 against the non-food basket; 2.1 and 2.1 on the regional chains' two wedges. The regional chains' legs show the same asymmetry as every other group's, −4.6 log points (s.e. 1.6) at the cut and +9.6 (1.2) at the repeal on uncapped goods, but at half the national chains' size. The single-store retailers' capped-goods wedge owes much to the cut window, when their treated prices rose 5.6 log points (s.e. 1.3) relative to control goods whose prices barely moved.

the price freeze of March 2020 does not contaminate the comparison. Precios Máximos has no product list of its own. Resolution SCI 100/2020 applies to the goods in Annex I of Disposition 55/16, the categorization built for the Precios Claros reporting system. Its 304 obligatory categories cover essentially the whole supermarket basket and include the control categories (biscuits, coffee, confectionery, soft drinks) alongside the treated ones. A ceiling that binds treated and control goods alike largely cancels out of the comparison, and in any case a freeze cannot make the wedge widen. Third, the depreciation of the peso four days before the cut does not account for the wedge. If treated goods respond more to the exchange rate than control goods, part of any wedge measured from before the cut is the depreciation and not the tax. In the 2018 depreciation, uncapped goods barely moved relative to control. Applying their 2018 response to the August 2019 depreciation lowers the clean-window wedge from +5.5 to +5.1 index points and the October–December 2020 wedge from +8.3 to +7.8. The larger correction that Benzarti, Garriga, and Tortarolo (2024) estimate for all treated goods together, which comes from the capped goods, leaves +4.1 and +6.9 (Appendix B).

5.8. Taking Stock

The first four facts establish that prices in this market behave as the state-dependent literature says prices behave. Adjustment is concentrated at the shock rather than spread over it, there is no anticipation of a known future change, the repricing frequency moves with the size of the shock and with inflation, and the cross-store price distribution does not widen in the way a constant hazard implies. Their role in the paper is to show that this episode is a reasonable laboratory for testing models of price setting, and the contribution lies in what follows.

The test uses the last two facts, which Benzarti, Garriga, and Tortarolo (2024) document for this episode, and it depends on the shock being fully reversed. A tax cut that is

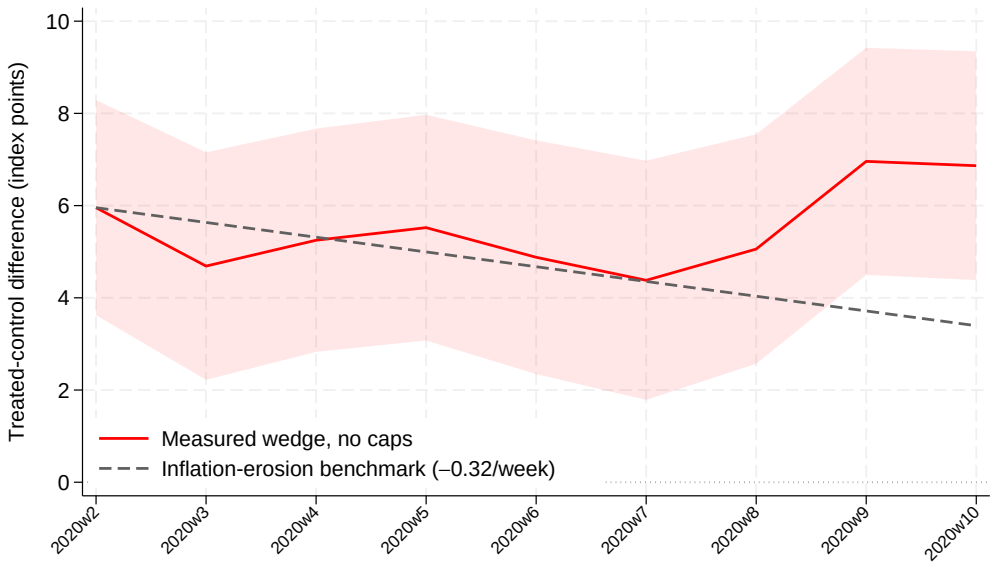


FIGURE 7. The wedge does not follow the erosion path

Weekly event-study coefficients for uncapped goods over the clean window 2020w2–w10, from specification (3) on the weekly chain panel, base week 2019w32; the shaded band is the 95 percent confidence interval, with standard errors clustered on barcode. The dashed line starts at the first clean week's coefficient and declines at 0.32 index points per week, the measured control-food inflation over the window: the path the wedge would follow if treated nominal prices froze and inflation eroded the gap.

later undone returns the fundamentals to exactly where they started, and this is the only way to tell a slow return apart from no return. Observing a wedge that is still open after a year tells us nothing if the fundamentals moved in the meantime, but a great deal if they did not. Every friction in the literature changes how long the return takes, and none of them changes whether it happens. The rest of the paper is about what the sixth fact implies for these models.

6. A Quantitative Menu Cost Model

6.1. Setup

A firm sells one good and posts a nominal price p against a desired price p^* that drifts up with trend inflation π and is hit by firm-specific shocks. Write $x = p - p^*$ for the

TABLE 4. Scoreboard: predictions and data

	Calvo	Menu cost	Menu cost + BM	Data
1. Adjustment to the cut	Staggered	Immediate	Delayed	**Immediate**
2. Increase before repeal	Yes	None	None	**None**
3. Frequency response	None	Spikes	Falls, then spike	**Spikes, both shocks**
4. Dispersion response	Rises	Falls	Flat, falls	**Falls vs. control**
5. Cut/repeal symmetry	Symmetric	Near-symmetric	Asymmetric	**28% down, 65% up**
6. Long-run wedge	None	None	None	**Persistent**

Predictions of each model for the six facts, against what we measure. The first four facts are the ones the state-dependent literature has established elsewhere and which this episode reproduces; the last two are the ones it has not been able to test. Benzarti, Garriga, and Tortarolo (2024) document the first and the last two for this episode. The Calvo entries in rows 3 and 4 hold for the frequency of all price changes and for firms that reprice rarely; for the measures we use, these two facts separate the models less sharply (Sections 5.4 and 5.5). BM denotes the Ball and Mankiw (1994) dynamics.

log price gap. A firm that leaves its price alone enters the next month with a gap of $x' = x - \pi - \varepsilon'$, so inflation erodes the gap without any action by the firm, and a firm that adjusts pays a menu cost and resets to a gap of its choosing. Profit loss rises with the square of the gap, and we set its coefficient B to one, so the menu cost is measured in units of one month's profit loss at a log price gap of one, that is, 100 log points. The monthly discount factor is $\beta = 0.96^{1/12}$.

Two features matter for what follows. First, following the CalvoPlus model of Nakamura and Steinsson (2010), the menu cost is κ_L with probability α and κ_H otherwise, with $\kappa_L = 0$, so a draw of the low cost is a free adjustment opportunity. At $\alpha = 0$ the model is a pure menu cost model. As $\kappa_H \to \infty$ it becomes Calvo, because a firm then changes its price only on a free draw, which comes with the same probability α every month whatever its price. We estimate both parameters rather than impose them, so Calvo is a special case that the data can accept or reject. Second, following Karadi and Reiff (2019), ε is drawn from a mixture of two normal distributions, which produces the occasional very large changes, the fat tails, that distributions of price changes display and that a single normal cannot. Writing V for the value function and W for the value

of resetting,

$$
\begin{aligned}
V(x) &= \alpha \max\{c(x),\ -\kappa_L + W\} \ +\ (1-\alpha)\max\{c(x),\ -\kappa_H + W\}, \\
c(x) &= -Bx^2 + \beta\, \mathbb{E}_\varepsilon V(x - \pi - \varepsilon),
\end{aligned}
\tag{4}
$$

with $W = \max_x c(x)$. We solve (4) by value function iteration on a grid. The resulting pricing rule is an inaction band with a reset point: unless it draws a free adjustment, a firm leaves its price alone while the gap stays inside the band, and any firm that adjusts moves its gap to the reset point. Under positive trend inflation the band is asymmetric. At our estimates it runs $[-0.087,\ 0.105]$, so a firm tolerates a price further above its target than below it. The reset gap is +0.021, so a firm that adjusts sets its price *above* its target and lets inflation bring it down. Nothing imposes that asymmetry. It comes out of the solution, and it is what makes the model's response to a tax cut differ from its response to a tax rise.

6.2. Calibration and Estimation

We calibrate on the control group, whose tax never changed, so the targets are untouched by the episode. Trend inflation is set to 2.94 percent a month, the mean monthly inflation of control food on the store panel over the pre-shock months of 2019. We choose the remaining four parameters, the menu cost, the Calvo weight, and the two shock standard deviations, to match four statistics from the data. Two are the slopes of the frequency of large price increases and of large price decreases on monthly inflation, which measure how much each frequency moves when inflation rises. The other two are the mean size of increases and of decreases.

We match slopes rather than levels because our price is a monthly unit value and not a posted price. A cell whose price never changed can still register a change when the mix of transactions behind the average moves, so some cells that did not reprice cross

any threshold we set. That classification error raises the *level* of measured frequency, but because it is roughly unrelated to inflation, it can only flatten the *slope*. The slope is therefore the credible target, and the true slope is at least as steep as the measured one. Section 5.4 reports the underlying relationship: across thirteen months the frequency of large increases rises with inflation with a slope of 5.68, and the frequency of large decreases falls with a slope of -0.64.[23]

We estimate a menu cost of 0.013, which equals the profit a firm loses in one month when its price is about eleven log points off target. The shocks have standard deviations of 0.035 and 0.178, with one shock in ten drawn from the wider distribution, and the Calvo weight is $\alpha = 0.34$, well away from zero. Roughly a third of firms draw a free adjustment opportunity each month, and since at these parameters the probability that a firm changes its price in a given month is 0.45, three quarters of all price changes come on a free draw and cost the firm nothing. The Calvo weight is not tightly pinned down: a calibration with a weight of 0.22 fits the four targets almost as well. The model that follows is more flexible than a pure menu cost model, and correspondingly a harder benchmark to defeat.

The fit is close on all four targets. The increase slope is 5.42 against a measured 5.68 and the decrease slope -0.64 against -0.64, and the simulated size of increases is +0.132 against +0.140 in the data and the size of decreases -0.157 against -0.156. The largest miss is the size of increases, which the model puts six percent too low. Since classification error can only flatten the measured slope, the true increase slope is at least 5.68, so in the model how often firms reprice may depend somewhat less on inflation than it does in the data.

Two statistics were held out of the calibration and provide a check on it. At a monthly inflation rate of three percent, close to the pre-shock rate, the model produces large increases in 26.2 percent of cells and large decreases in 3.1 percent, against 31.1 and 10.5

[23] Standard errors 0.14 and 0.07.

percent in the data. The model lies below the data on both, by 4.9 and 7.4 percentage points, which is the direction classification error predicts.

6.3. The Simulated Episode

We simulate the episode as it happened. The tax is removed, which lowers the desired tax-inclusive price by log(1.21) = 19.1 log points, and restored five months later. The restoration date was announced at the outset, so a firm inside the window knows the tax is coming back and does not follow the band it uses in normal times. We solve for its pricing rule by working backward from the repeal date. In the final month of the window the reset gap falls from +0.021 to zero and the inaction band widens, because a firm that expects to reset next month in any case stops pricing above target to ride out inflation and simply minimizes its current loss.

The model reproduces the first two facts without being asked to. Adjustment to the cut is immediate: the treated–control wedge reaches −16.7 log points in the first month, so almost the whole response comes at once rather than building up step by step. The model also shows no response ahead of the repeal. Through the VAT-free window the wedge deepens from −16.7 to −20.7, moving away from zero, not toward it, in the months before the tax returns, which is the pattern Section 5.3 documents in the data.

The model fails on Fact 5. In the VAT-free window the wedge reaches −20.7 log points, against −8.4 for the large chains on the national monthly panel (Section 5.6) and −11.5 on the weekly chain panel measured plateau to plateau (Appendix B). The model passes the cut through more than completely where the data pass through between 44 and 60 percent. All three figures are in log points, not the index points of Table 5. The reason is the anticipation just described: firms that expect to reprice at the repeal set prices at target, not above it, which pushes the treated group further below control

than a pure level shift would. Trend inflation and fat tails make the model pass through more than the full cut, not less. The menu cost model therefore gets the downward leg of the asymmetry in Section 5.6 wrong.[24]

The model gets the direction of Facts 3 and 4 right and their size wrong, overstating the first and understating the second. The frequency of price changes spikes at both shocks, as it must in a state-dependent model. At the cut 87.7 percent of treated firms post a fall larger than five percent against 3.1 percent of control, a gap of 84.7 percentage points where the data show 35.7. At the repeal 98.1 percent post a large rise against 25.8 percent of control, a gap of 72.3 where the data show 46.7. Too many firms in the model reprice at both events, the same overshoot that appears in its pass-through.

Cross-store dispersion compresses relative to control, and here the model does too little, not too much. The difference-in-differences in the dispersion of price gaps is -0.004 against a measured -0.018, which is twelve percent of the model's own baseline dispersion against twenty percent of the data's. With trend inflation held constant, the model also compresses dispersion through the wrong channel. In the model treated dispersion falls while control stays flat. In the data treated dispersion holds flat while control rises, because the control basket was absorbing the August depreciation. Giving the model the economy-wide cost shock it lacks corrects this. Simulated under the measured monthly path of control-food inflation, treated dispersion holds flat while control dispersion rises, as in the data, and the difference-in-differences is -0.003. The model misses the magnitude, not the mechanism. The measured path changes nothing else of consequence: the post-repeal wedge's half-life lengthens from under three weeks to about three and a half, because measured inflation nearly stops in February 2020, and the conclusion of the next subsection does not depend on holding trend inflation

[24] Benzarti, Garriga, and Tortarolo (2024) show in a stylized menu cost model that firms facing inflation can fold future increases into the cut and so pass less of it through. That model has no announced repeal. Here firms know the tax is coming back and expect to reprice when it does, so at the cut they set prices at target, which removes the reason to fold in future increases.

constant.

6.4. The Post-Repeal Wedge: Fact 6

Whatever the model does at the cut, it cannot produce what the data show after the repeal. Figure 8 puts the two paths on the same axes. In the calibrated model the wedge is +1.9 log points in the first month after the tax returns (+2.6 index points at the control level of that month), falls to +0.3 in the second, and is indistinguishable from zero by the third. Its half-life is 0.6 months, under three weeks. Most of that decay falls inside the ten clean weeks, so the model fails before the pandemic begins. In the data the wedge is +6.0 index points in the first week after the transition week, +5.3 nine weeks later, +4.7 at six months and +9.8 at the end of the year, part of it from goods released from the price freeze in November (Section 7). It does not decay at all.

No choice of parameters gives this class of model a wedge that is both as large as the measured one and as durable. Setting the Calvo weight to zero leaves the half-life under three weeks. Doubling the menu cost gives three weeks, and at five times the estimated value the wedge already has the wrong sign. Even thirty weeks requires a menu cost fifty times the estimated value together with a Calvo weight of zero. At those values the post-repeal wedge is 1.0 log points, about 1.4 index points at the first-month control level used above, where the data show close to six index points, so the model makes the wedge last longer only by making it smaller. At the estimated Calvo weight, menu costs large enough to slow the decay appreciably also give the wedge the wrong sign.[25]

The sweep confirms the result, but the result does not depend on it. However the model is calibrated, it cannot hold the wedge open, because of how it is built. Once the tax is restored a treated good faces exactly the fundamentals it faced before the cut, so

[25]The Calvo weight is not tightly pinned down. At the calibration with a weight of 0.22, which fits the targets almost as well, five and ten times the menu cost give a wedge of the right sign with half-lives of 4.5 and 8.5 weeks, and only fifty times gives the wrong sign. Even there the wedge is gone within a few months, far from the year without decay in the data.

every firm wants to charge what it charged then, and the wedge stays open only while some firms have not yet repriced. Every month some of them do. A share set by the Calvo weight draws a free adjustment, and inflation keeps lowering the others' prices relative to target until paying the menu cost is worthwhile. Changing the parameters makes the wedge close faster or more slowly, but as long as the menu cost is finite, it always closes.

The sixth fact therefore differs in kind from the five before it. On Facts 1 to 5 the question is how close the model comes, and any calibration, ours included, fits some of them better at the cost of others. On Fact 6 the model fails at every calibration. A wedge that stays open for a year can come from three sources. The first is a friction so large that prices essentially never move, which Facts 1 and 3 refute directly and which this calibration rejects at any plausible menu cost. The second is a state that changes more slowly than a year, such as customer capital or consumer habits, which Section 7 discusses and our data cannot rule out. The third is a model with more than one long-run equilibrium, in which prices need not return to where they started once the tax is back.

7. Interpretation: What Carries the History of the Episode

7.1. Path Dependence and Multiplicity

Section 6 shows that in the menu cost model, and in Calvo, which it nests, a fully reversed shock leaves no trace: the friction sets how long the return takes but not whether it happens. A treated–control wedge that survives the shock being undone is therefore evidence against this class of models, not against particular parameter values.

We find that the wedge is still open a year after the shock was undone. Figure 9 follows the uncapped and capped treated–control differences for a full year after the

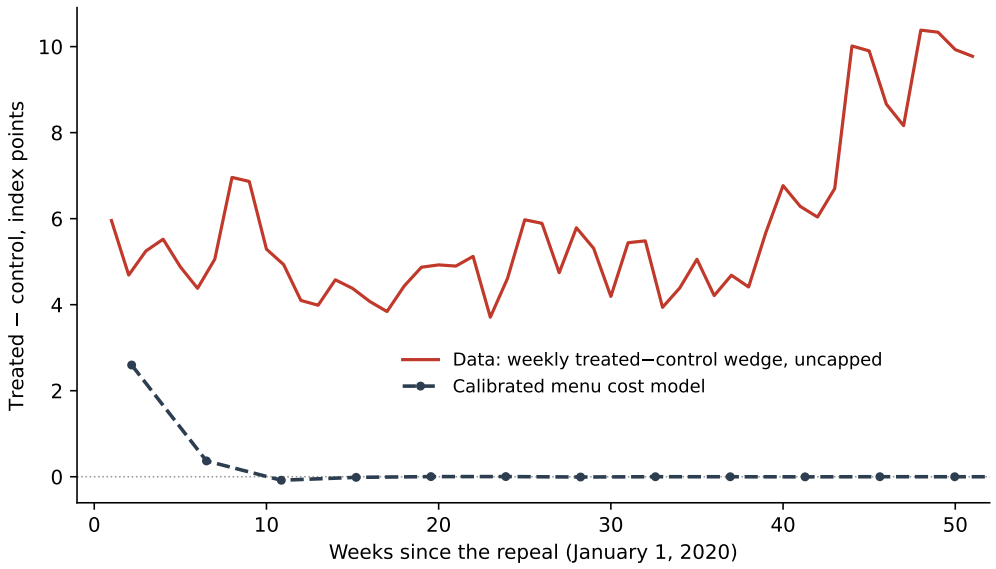


FIGURE 8. The model's post-repeal wedge has a half-life under three weeks; the measured one does not decay

The solid line is the weekly treated–control wedge for uncapped goods from the weekly chain panel, in index points, plotted against weeks since the reinstatement of the VAT on January 1, 2020; the transition week 2020w1 is omitted, as in Section 2.3. The dashed line is the calibrated CalvoPlus menu cost model of this section, simulated through the same episode with the repeal anticipated, plotted at the midpoint of each model month and converted from log points to index points using the control group's mean index level in the corresponding week, so the two series share units. Model parameters are a menu cost of 0.013, a Calvo weight of 0.34, shock standard deviations of 0.035 and 0.178 with a ten percent mixture weight, and trend inflation of 2.94 percent a month; the coefficient on the squared gap in the profit loss is set to one and the monthly discount factor is $0.96^{1/12}$.

repeal, and Table 5 reports them by period. On a sample balanced across every week of 2019 and 2020, the uncapped wedge is +0.2 index points before the cut, −12.7 through the VAT-free months and +7.0 in the clean window of early 2020. It is higher in October–December 2020, at +10.1, mostly because of goods released from the price freeze in November 2020, when the government lifted the ceilings on a list of goods that includes three of the uncapped ones: canned fruit other than peaches, corn oil and brown rice. They make up about a quarter of the uncapped cells. Without them, the wedge is +5.5 index points in the clean window and +6.4 in October–December.[26] A year after a tax

[26] Resolution 552/2020 of the Secretariat of Domestic Trade, published on November 12, 2020, suspended the ceilings of the March freeze until January 31, 2021 for the goods in its annexes. We drop the three goods from both groups. Standard errors 1.3 and 1.6; on the full sample the two figures are +4.4 (s.e. 1.2)



FIGURE 9. The treated–control price difference does not decay

Weekly event-study coefficients on the weekly chain panel, balanced across every week of 2019 and 2020, with barcode, area and week fixed effects and standard errors clustered on barcode; shaded bands are 95 percent confidence intervals. The omitted week is 2019w32. Vertical lines mark the VAT cut (August 16, 2019) and its reinstatement (January 1, 2020).

change that was completely undone, treated prices still sit above where they began relative to the control group, so the gap does not close. Nor can we say that it widens, since the path falls back to +5.1 over April–June and the confidence intervals in late 2020 are wide.

Two caveats bound what this establishes. First, a model with a single long-run equilibrium can still fit this evidence if it adds something that changes slowly enough, such as customer capital, consumer habits or long-term supply arrangements, and we cannot exclude such a model. It needs a mechanism whose half-life exceeds a year, and it has to explain why the wedge shows no sign of heading anywhere within the sample. Quantities offer a test of the customer-capital version, because a seller holding prices high by running down a stock of customer goodwill should be selling less after the repeal. We compare quantities with the same month a year earlier to remove seasonality,

and +5.5 (s.e. 1.5).

TABLE 5. The treated–control price difference over the full post-repeal period

	No caps		7% cap	
	Full sample	Balanced	Full sample	Balanced
Pre-cut (2019w20–w31)	0.10	0.24	0.46	−0.03
	[−0.58, 0.78]	[−0.48, 0.97]	[−0.01, 0.92]	[−0.57, 0.50]
VAT-free (2019w34–w52)	−13.23	−12.74	−14.90	−14.94
	[−14.53, −11.94]	[−14.26, −11.22]	[−15.60, −14.19]	[−15.75, −14.13]
Clean window (2020w2–w10)	5.51	6.97	−3.31	−2.32
	[3.18, 7.84]	[4.18, 9.76]	[−4.45, −2.17]	[−3.68, −0.96]
April–June 2020	4.50	5.11	−4.38	−2.94
	[2.06, 6.94]	[2.31, 7.91]	[−5.59, −3.17]	[−4.36, −1.51]
July–September 2020	4.91	5.98	−3.28	−1.97
	[2.42, 7.39]	[3.14, 8.82]	[−4.64, −1.92]	[−3.55, −0.39]
October–December 2020	8.30	10.10	−1.61	−0.82
	[5.23, 11.37]	[6.60, 13.60]	[−3.24, 0.03]	[−2.66, 1.03]

Treated–control differences in the weekly price index, in index points, estimated on the weekly chain panel with barcode, area and week fixed effects and standard errors clustered on barcode; 95 percent confidence intervals in brackets. The omitted week is 2019w32, the last full week before the VAT cut, so coefficients are relative to that week and the pre-cut row is a test of parallel trends rather than a normalization. "Balanced" restricts to the 16,670 barcode–area cells observed in every week of 2019 and 2020; the full sample also keeps cells that drop out or have gaps after 2020w10. The VAT was removed on August 16, 2019, and reinstated on January 1, 2020, the latter accompanied by a negotiated 7 percent cap on the national chains for the goods in the right-hand panel.

and we find no lost sales where the test is clean. At the smaller stores, the sellers with the largest lasting overshoot, treated quantities in the two usable post-repeal months are indistinguishable from their year-earlier level, at +2.6 and −0.4 percent, with no detectable pre-trend.[27]

Second, and more important for what follows, path dependence establishes only that something in the environment carries the history of the episode, not what that something is. There are two candidates. First, multiple equilibria among the sellers of a good: when each seller wants to charge more because its rivals do, the market can settle at more than one price (Ball and Romer 1991; Blanchard and Kiyotaki 1987). Benzarti,

[27] Standard errors about 4; joint pre-trend $p = 0.75$. At the large chains the design carries a pre-trend and the post-repeal readings are specification-dependent, bracketed between −8 and +6 percent, so we do not lean on them. March 2020 is excluded from the window because pandemic stockpiling dominates it.

Garriga, and Tortarolo (2024) already note that a year-long overshoot would require implausibly large menu costs, and conjecture that it calls for models of coordination failure (their Appendix B.2), which firms themselves most often cite as the reason their prices are sticky (Blinder et al. 1998). Second, a reference price that consumers attach to the good. Nothing established so far separates coordination among sellers from a reference point on the consumer side.

7.2. Coordination Among Sellers

Multiple equilibria arise when the pull of rivals' prices is strong. Each seller of a good charges more when its rivals charge more, because a seller whose rivals are expensive loses fewer customers by being expensive itself. If over some range a seller's best markup rises more than one for one with its rivals' markup, the market can settle at a low markup or at a high one, and at either no seller gains by changing its price on its own (Cooper and John 1988). Such a range does not arise in the standard case, in which buyers divide their purchases among sellers by relative prices alone and demand for the good has a constant elasticity. It requires that how readily buyers switch between sellers depend on how high prices are. The high markup is not collusion. In models of collusion the high price exceeds what each firm would charge given its rivals' prices, so it has to be sustained by monitoring and the threat of a price war (Green and Porter 1984; Rotemberg and Saloner 1986). At a high-markup equilibrium each seller already charges its best price given its rivals' prices, and nothing has to be enforced.

Sellers prefer the high markup only if each seller gains when its rivals raise their prices. That holds when enough of the customers a rival loses come to the other sellers of the good rather than leaving it altogether.[28] When it holds, an announcement that points upward describes a price no seller wants to leave once the others have moved to

[28]With N sellers, an elasticity of substitution σ between the sellers of the good and an elasticity of demand for the good η, the condition is $\sigma > \eta(N-1)/N$.

it, and an announcement that points downward does not, which gives the asymmetry between the cut and the repeal a direction.

Fundamentals do not decide which of two equilibria prevails, so something else has to (Krugman 1991). A public announcement can, if everyone observes it, everyone knows that everyone observes it, and it is specific enough to name an action. A statutory tax change meets all three conditions, and so does the negotiated cap at the repeal: each is published, dated, and names a number. The number the tax names, however, is a *change*, "the value-added tax returns to 21 percent", and not a price. Sellers can follow it in two ways. Each seller can return to the price it charged before the cut, or every seller can add the announced change to whatever it is charging. Only the second leaves prices higher than before the cut wherever the cut was not passed through in full. If the explanation is multiplicity, Fact 6 measures how far apart the two markups are: about four to five log points on the weekly panel, 7.1 on the national monthly panel (Section 5.6) and 7.4 on the store panel (Section 5.7).

This reading makes three predictions that the data can test. First, sellers coordinate on the price of a particular good, so the trace should attach to products and not to stores. Second, sellers that face different public numbers should move to different numbers, even on the same good. Third, the increase should arrive when the announcement does, with sellers moving together. The reading says nothing about why stores passed through only part of the cut, which Fact 5 traces to frozen nominal prices eroded by inflation.

7.3. Evidence I: Product-Level, Not Store-Level

We find that the episode changed prices product by product and not store by store, which is the first prediction of Section 7.2. Sellers coordinate on the price of a particular good, so whatever they coordinate on attaches to the product and not to the individual store, and any lasting widening of dispersion should appear on the product margin, as

differences between products. For each response window we split the price response of each store–barcode cell into a store effect and a barcode effect, and measure how much each varies. The measure is the split-half signal standard deviation, the spread of the group means scaled by how closely the means of two random halves of each group agree. This removes the sampling noise that would otherwise make a margin with many small groups look more variable than it is. The exercise needs no store locations, so it runs on all 172 stores and 797 treated barcodes.

We find that the lasting increase sits on the product margin and that the episode widened it (Table 6). Over the full episode the treated barcode margin is 0.135 against 0.099 in control food, a 36 percent increase. The treated store margin is 0.026 against 0.029, which is to say the episode added nothing to it and if anything treated stores vary slightly less than control stores. Adjusted R^2, twenty random re-splits and the non-food control group all give the same picture (note to Table 6). Whatever the VAT episode did to the cross-section of prices, it did it product by product.

Store-level differences open up at the cut and are gone after the repeal. In the cut window the treated store margin is 0.056 against a control value of 0.028, a doubling, and against 0.017 for treated cells in a placebo window. In the repeal window it is 0.036 against a control value of 0.026, and by the end of the episode, as above, the excess is gone entirely. Stores differed in whether they passed the cut through, and after the repeal those differences have disappeared, which is what a common announcement produces when sellers that set their own prices at the cut follow the same number at the repeal.

The zero on the store margin over the full episode rules out explanations that leave a lasting store-specific price, such as local market power or a store's own pricing policy. It does not rule out store-level differences that the repeal itself removed, and the store margin is elevated in the cut and repeal windows.

The decomposition does not, however, separate coordination from the consumer

TABLE 6. The episode added dispersion across products and none across stores

	Treated		Control food	
	Store	Barcode	Store	Barcode
Placebo window	0.017	0.088	0.030	0.078
Cut window	0.056	0.088	0.028	0.085
Repeal window	0.036	0.086	0.026	0.067
Full episode	0.026	0.135	0.029	0.099

Split-half signal standard deviations of the store and barcode effects in the cell-level change in log price (in logs, so 0.10 is ten log points), on the Córdoba store panel; a cell is a store–barcode pair. Within each group the cells are randomly halved, the two half-means are correlated across groups, and the signal standard deviation is the square root of the reliability times the variance of the group means. This removes the sampling noise that would otherwise favor whichever margin has more and smaller groups, which matters here because there are 172 stores and 797 treated barcodes. Treated columns use 172 stores and 797 barcodes, control food 171 stores and 890 barcodes. Each window compares two-month averages: placebo 2019m2–m3 against m5–m6; cut 2019m6–m7 against m9–m10; repeal 2019m11–m12 against 2020m1–m2; full episode 2019m6–m7 against 2020m1–m2. On the non-food control group the full-episode figures are 0.030 on the store margin and 0.118 on the barcode margin. Adjusted R^2 for the full episode is 0.030 on store effects against 0.600 on barcode effects for treated cells, and 0.049 against 0.396 in control food. The split is seeded; twenty draws put the treated full-episode store figure between 0.025 and 0.028 and the barcode figure between 0.134 and 0.137.

side. It is tempting to say that consumer-side explanations are about products while coordination is about sellers, but that is too clean. Firms coordinate on the price of a particular good, so coordination is itself a product-level phenomenon, and differences across products fit coordination among sellers as well as they fit a reference price attached to that good.

7.4. Evidence II: The Seller's Institutional License

We find that sellers who faced different public numbers raised prices by different amounts, which is the second prediction of Section 7.2. This is also the sharpest test these data allow between an explanation that works through sellers and one that works through consumers. Anchoring on a past price, fairness about the tax and the salience of the tax change all operate at the level of the product, because a shopper's tolerance for an increase on a given barcode does not depend on whether a national chain or a single-

store retailer is selling it. If the January increase on an identical good instead depends on who is selling it, no explanation that works through the consumer's reference price for that good can produce it.

The repeal gives us the two numbers. The national chains signed the negotiated 7 percent cap, and the other stores in the panel were never party to it and were free to restore the statutory 21 percent. In the same month and the same city, the same barcode faced a different public number depending on who was selling it. Writing Δ_{js} for the December-to-January log price change of barcode j at store s, we estimate

$$\Delta_{js} \;=\; a\,\text{signatory}_s \;+\; b\,(\text{signatory}_s \times \text{capped}_j) \;+\; \phi_j \;+\; \varepsilon_{js}, \tag{5}$$

where ϕ_j is a barcode fixed effect and standard errors are clustered on barcode. The comparison is between sellers of the same barcode, so a barcode that only one type of store stocks contributes nothing to it. We find that the gap between sellers of the same good changes sign with the kind of good. On goods the cap did not cover, signatories raised prices 5.4 log points more than non-signatories. On goods it did cover, they raised prices 4.3 log points less. The difference-in-differences is -9.6 log points, on 22,559 store–barcode pairs covering 716 barcodes in 172 stores.[29]

A consumer-side explanation cannot produce this reversal. A consumer's reference price attaches to the product, so it can make January increases small or large, or larger in some categories than in others. It cannot make the same barcode rise by different amounts at different sellers, and it cannot reverse which seller raises more according to whether the chains' agreement with the government covers the good. The difference-in-differences is 2.3 times the largest estimate the same regression gives in six pre-period placebo months, and it survives volume weighting and two-way clustering on barcode

[29] $b = -0.096$ (s.e. 0.029, $p = 0.001$); the two level gaps carry standard errors of 2.8 and 0.7.

and store.[30]

The distribution of January price changes shows the same thing more directly (Figure 10). The density of December-to-January log price changes on treated goods has two peaks, one at the cap and one at full restoration of the tax. To read shares off that density we need a benchmark, because a window six or seven log points wide also collects ordinary price changes. Control food in the same month, which faced neither number, is the benchmark: 19.5 percent of control cells fall in the cap window $[0.04,\ 0.10]$ and 4.6 percent in the restoration window $[0.16,\ 0.23]$ around $\log(1.21) = 0.191$.

Against that benchmark, we find that full restoration of the tax was a focal number for every seller, and that the cap was a focal number only for the signatories on the goods it covered. First, in all four seller-by-good groups, far more increases fall at full restoration than in control food. Among signatories the excess is 14 percentage points on capped goods and 31 points on goods the cap did not cover, where the share is 36.0 percent, nearly eight times the control rate. Second, only signatories selling capped goods, the one group the agreement binds, bunch at the cap, at 38.5 percent against 19.5 in control food, while the other three groups sit slightly below the benchmark, at 16.0, 16.1 and 14.4 percent. Third, the same holds within barcode: on the same good, signatories are 23.0 points more likely than non-signatories to land at the cap and 10.3 points less likely to land at full restoration.[31] Compliance was partial, however. On capped goods, signatory increases exceed control food at both numbers, by 19 points at the cap and 14 at restoration, and their median increase of 10.2 log points sits between the two. The agreement therefore moved a large minority of signatory increases, not all of them.

The comparison is between two kinds of chain, not between chains and independent

[30] -0.100 under volume weights and -0.096 under two-way clustering. We report b rather than either level gap because the capped-goods gap on its own is only 1.8 times the largest placebo, too small to tell apart from the noise in a monthly unit value.

[31] Barcode fixed effects, clustering on barcode; standard errors 2.5 and 2.4.

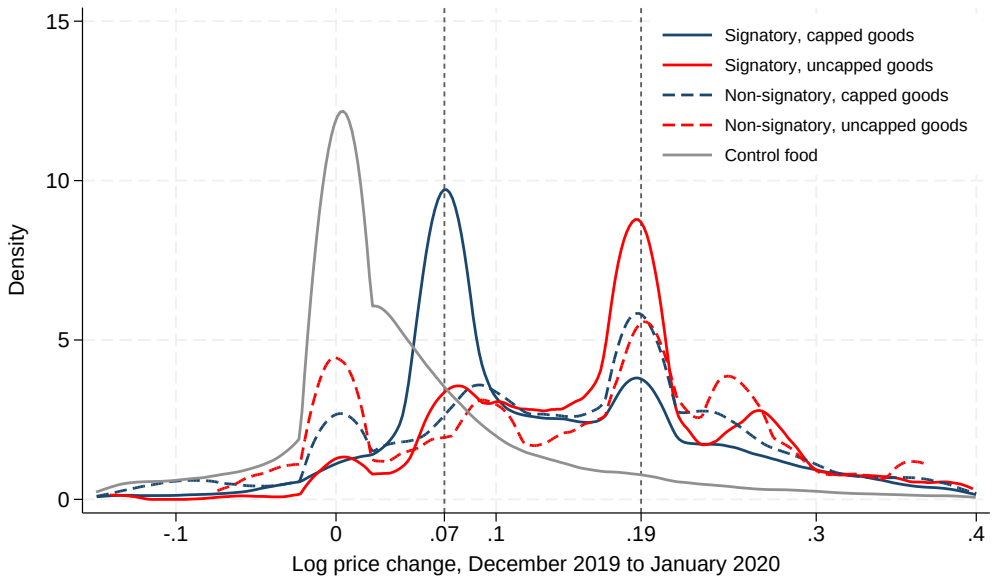


FIGURE 10. The January increase bunches on each seller's institutional number

Kernel densities of the December 2019 to January 2020 log price change on the Córdoba store panel; a cell is a store–barcode pair. Treated cells are the sample of this section, with fluid milk and goods of unclassified cap status excluded; control food pools all stores. Solid lines are signatory (national chain) stores and dashed lines non-signatory stores; navy marks goods covered by the negotiated cap and red goods it did not cover. Vertical lines at log(1.07) = 0.07, the negotiated cap, and log(1.21) = 0.19, full statutory restoration. The shares reported in the text are computed on all 22,559 treated cells; restricting them to the 15,821 cells on barcodes that both kinds of store stock leaves them nearly unchanged, at 38.8 percent for signatories on capped goods against 16.0 for non-signatories. Narrowing the cap window to one log point either side of log(1.07) widens the same contrast, to 24.6 against 4.9 percent. Six pre-period placebo months put between 9 and 38 percent of cells in the cap window depending on month and group, averaging 18 to 21 percent by group, and between 0 and 8 percent in the restoration window; the identification is the contrast across groups within January, not the level of any one share.

stores. Of the 82 non-signatory stores, 56 are regional chains and 26 single-store retailers, and by cells the group is 89 percent regional chain. The three types of store behave very differently: on capped goods the median January increase is 10.2 log points in the national chains, 17.4 in the regional chains and 3.7 in the single-store retailers. The result therefore comes from comparing signatory and non-signatory chains, both multi-store food retailers that differ only in whether they signed. The single-store retailers could not drive it in any case. About half of their prices do not change in any given month, January included, so using them as the unbound group would push the measured gap toward zero for that reason alone. They do move over the episode as a whole, by 17.5

log points from November to March against 22.0 for the regional chains and 22.9 for the national ones, but they move less and later.

Relative to their own control goods, however, the single-store retailers end the episode highest of the three, because their control goods rose slowly too. Over the full episode, from the two months before the cut to the two months after the repeal, the difference between the treated and control groups is +6.4 log points among the single-store retailers, +0.7 among the regional chains and -0.1 among the national chains. On goods the cap did not cover the single-store figure is +11.2; on goods it did cover it is unchanged at +6.4, where the national chains show -1.4.[32] These retailers raise prices least in the month of the repeal and finish the episode highest relative to control. This is a finer split than in Benzarti, Garriga, and Tortarolo (2024), which separates monitored chains from less-monitored smaller sellers. The store-type classification here separates the 26 single-store retailers from the regional chains they are usually pooled with, and the two behave differently. The cap therefore removed the lasting increase for the sellers it bound. Among the sellers it did not bind, only the single-store retailers end clearly above control, and part of that comes from control goods whose prices barely moved.

The two shocks also differ by store type in the direction Section 7.2 gives the asymmetry. At the cut, the national chains cut nominal prices and the other stores mostly did not. Between the June–July and September–October averages, 32.0 percent of the national chains' treated cells cut the nominal price by more than five percent, against 11.5 percent at the other stores, nine in ten of them regional chains by cells. For those stores pass-through came instead from frozen nominal prices and trend inflation. At the repeal the two groups are nearly identical, with 89.6 and 85.4 percent raising the nominal price by more than five percent between the November–December and January–February

[32] Standard errors 1.8, 1.9 and 1.7 on the three full-episode wedges, and 4.9, 1.9 and 1.7 on the split by cap status.

averages.[33] At the cut, each store did what suited its own circumstances. At the repeal, every store moved, and only the number it moved to differed.

This evidence has three limits. First, the uncapped set is thin: 66 of the 716 barcodes and 1,728 of the 22,559 cells, with only 30 uncapped barcodes stocked by both types of store, and b is identified from them. Second, the estimate is smaller when two-month averages replace single months in (5): the difference-in-differences falls to -7.2 log points and the capped-goods gap to -1.5, which loses significance under volume weights.[34] The two kinds of chain are not converging on a price, however. Figure 11 follows the same barcode at both kinds of chain, 295 capped and 468 control-food barcodes that both sell in every month. Net of the same gap in control food, the regional chains' premium over the signatory chains on capped goods is +6.0 log points in January, +6.3 in February and +7.7 in March, so it widens rather than closes.[35] The pooled estimate narrows because the single-store retailers complete their increase late and from below, and because the regional chains inflated more slowly than the national ones in February on all goods. Each kind of chain moves to its number and stays there for as long as the store panel runs, with March read against the price freeze of Section 2.3. Third, we infer signatory status from the store's type and not from the text of the agreement, and our cap classification marks the goods the agreement covered, not stores that signed it.

Finally, this evidence supports a narrower claim than tacit coordination. The cap was an explicit, published agreement, negotiated with the chains on December 31 and never legislated (Section 2.3). In the data, sellers comply with a focal point that the government brokered. The announcement was common knowledge within one group of sellers and did not bind the other, and the increases bunch at the cap in exactly the group it bound. Because the cap was enforced, bunching among its signatories

[33] These are moves between two-month averages, not the month-on-month moves of Section 5.4, so the two sets of shares are not comparable.

[34] -0.072 (s.e. 0.028) and -0.015 (s.e. 0.007).

[35] Standard errors 0.8, 0.9 and 1.0.

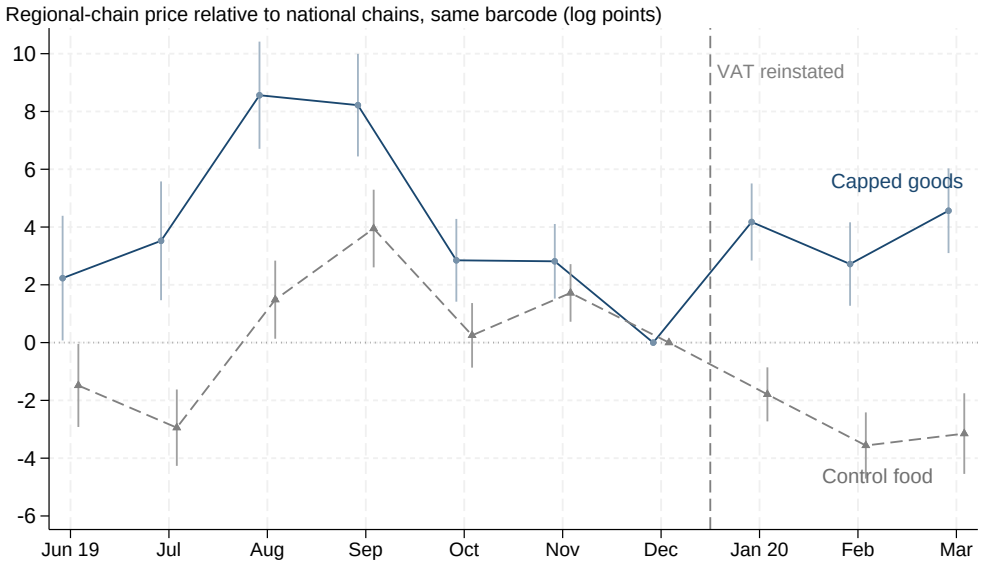


FIGURE 11. The regional chains' premium over the signatory chains opens at the repeal and does not close

Monthly event-study coefficients on the within-barcode gap between the mean log price at regional chains and at national chains, for the 295 capped treated barcodes (solid) and 468 control-food barcodes (dashed) that both kinds of chain sell in every month from June 2019 through March 2020, on the Córdoba store panel; the base month is December 2019, with barcode fixed effects, 95 percent confidence intervals, and standard errors clustered on barcode. The August–September rise on capped goods is the cut leg: the national chains passed the cut through immediately, the regional chains later, and that gap closes by December. After the repeal the capped-goods gap opens and stays open while the control-food gap drifts down; net of that drift the premium is +6.0, +6.3 and +7.7 log points in January, February and March. March is read against the Precios Máximos freeze (Section 2.3).

shows that it bound, not by itself that sellers coordinated on it, and the comparison that bears on coordination is on the goods the cap did not cover. This is the kind of public signal that multiplicity needs for sellers to coordinate on one equilibrium (Section 7.2), and it is direct evidence that sellers chose their January increase according to whether they had signed the agreement. It does not show that sellers coordinate without such a signal, and on its own it does not settle whether the explanation lies with sellers or with consumers. A consumer-side explanation can match it, but only by assuming that consumers grant different entitlements to different types of seller (Section 7.7).

7.5. Evidence III: The Increase Arrived Within Two Weeks

We find that the January increase arrived within two weeks, which is the third prediction of Section 7.2. Testing this requires weekly data, so we turn to the weekly chain panel behind Figure 9. The repeal fell on January 1, 2020, a week boundary, so there is no split week at the event. Figure 12 shows, for each cell, which of the eight weeks beginning January 1 contains its largest single-week price rise. Control food is flat: between 9.6 and 15.6 percent of control cells in each of the eight weeks, against the 12.5 percent that an even spread over the eight weeks would give. Treated cells on goods outside the cap put 32.0 percent of their largest rises in the first week and 45.9 percent in the second, or 77.9 percent inside two weeks, and sit below that benchmark in every week thereafter. Capped goods put 64.3 percent in the first two weeks. Uncapped treated prices complete 91.5 percent of their eight-week rise within two weeks and then stop, while control food drifts up at about half a log point a week throughout.

Two conclusions follow and one does not. First, the concentration rules out inattention and staggered review: a mechanism in which each store reprices when its own review comes due has stores moving over weeks or months, not 78 percent inside two weeks against a flat control. Second, the wave is not the national chains executing their negotiated agreement on schedule, because it is strongest on uncapped goods, which had no agreement in force, and if anything the cap made capped goods less synchronized (64 against 78 percent inside two weeks). The concentration cannot, however, separate a coordinated move from an ordinary repricing wave. In a menu cost model, a large common shock pushes most stores out of their inaction bands at once, so that model also predicts a wave lasting two weeks.

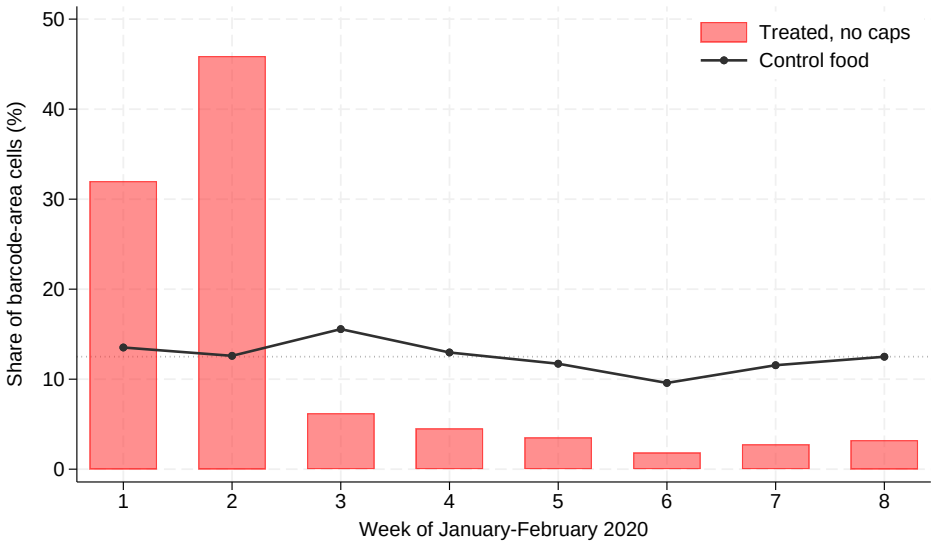


FIGURE 12. The repeal increase is complete within two weeks, and control food is flat

Share of cells whose largest single-week price rise falls in each of the eight weeks beginning January 1, 2020, on the weekly chain panel. Bars are treated cells on goods not covered by the negotiated 7 percent cap; the line is within-category control food. The dotted reference line at 12.5 percent is the share a uniform distribution would place in each week. Weeks begin on January 1, so the repeal falls on the boundary of the first week. Cells whose largest weekly rise is tied across two or more weeks are excluded.

7.6. Two Results That Do Not Favor Coordination

Two results that would have favored coordination among sellers are not in the data. First, we do not see sellers adding the announced change to whatever they were charging. On capped goods the national chains, the regional chains and the single-store retailers had passed through different shares of the cut by the time the tax returned. A store type that had passed through one more point raised its price at the repeal by 0.88 more points, with a 95 percent confidence set of 0.66 to 1.27. Adding the announced change at once would give a weight near zero, and a return to the old price a weight near one.[36] The

[36]The weight is minus the slope, across store types, of the repeal change in log price on the change from June–July to November–December 2019, each measured relative to the barcode's mean across stores and net of the same store type's control food, with the second change instrumented by the same quantity computed on a disjoint random half of the barcodes; inference resamples stores within type and barcodes. Comparing only the regional chains and the single-store retailers gives 0.80 (0.59 to 1.14). On data simulated from the menu cost model of Section 6 with a true weight of zero, the same estimator returns between −0.3 and 0.3 in nine replications out of ten. The panel does not identify chains, so stores of one chain are treated as independent.

result rules out only an immediate common change. The single-store retailers reprice about seven-tenths as often as the regional chains, and a common change applied at their slower pace produces a similar weight. February and March fit that slower path better than a return to the old price, although January fits neither. Across individual store–barcode cells the question cannot be answered. The variance of treated cells' price changes over the cut is 0.97 times that of control cells, inside the range of 0.94 to 1.03 that the same comparison gives in three windows with no tax change, so pass-through differs across cells by little more than ordinary noise.[37]

Second, the overshoot does not fall where a good has more sellers. In the usual models of multiplicity, more sellers weaken the pull of rivals' prices, and enough of them remove the high-markup equilibrium, although this need not hold in every such model. On the store panel we count, for each store and barcode, the other stores in the same locality that sold the barcode before the cut. The full-episode overshoot on uncapped goods does not fall with that count. Its slope on the log of one plus the count is +0.005 (s.e. 0.003), which rules out a fall of more than about 0.7 log points between a good with no rival seller and one with fifty.[38] A second version of the test, on the 90 national-chain stores with counts of competing stores within 300 meters, also finds no slope once store format is controlled for (-0.003, s.e. 0.005).

7.7. Alternative Explanations

The evidence in Sections 7.3 to 7.5 rules out two further classes of explanation. It rules out inattention and staggered review, because if the tax change set off each store's own

[37] Same cells in all four windows (41,518 treated and 47,400 control store–barcode pairs), price changes between two-month averages three months apart, demeaned by category; the no-tax windows are January–February to April–May, February–March to May–June and March–April to June–July 2019.

[38] Store-by-category and barcode fixed effects, standard errors clustered on barcode; 63,587 store–barcode pairs in 173 stores, 185 of the barcodes treated and uncapped. A seller is a store with a price for the barcode in at least four of the seven months before the cut, and all store types count. With fixed effects for store by category by treated status the slope is +0.003 (s.e. 0.008), too imprecise to rule out a fall of five log points. The geography is coarse: the city of Córdoba is a single market of 99 stores, and stores outside the panel are not counted.

review with a delay, increases would arrive over weeks or months, and the data show a wave complete within two weeks. It rules out explanations that leave a lasting store-specific price, whether a store's own pricing policy or local market power, because the episode added no lasting variation across stores (Section 7.3). There are two remaining explanations. First, coordination among the sellers of a good. Second, a reference price that consumers attach to the good. Both work product by product, and the data cannot really discriminate between the two.

When firms choose how much attention to pay (Maćkowiak and Wiederholt 2009), a shock as large and as publicized as this one requires every firm's attention at once, so the model predicts the synchronized increase of the first two weeks of January. Firms otherwise pay limited attention to aggregate conditions (Coibion, Gorodnichenko, and Kumar 2018). Attention models keep a single long-run equilibrium, however, so whatever attention does at the repeal, the wedge it leaves closes as firms start paying attention again, and a year without decay would require firms that never look again. By Fact 3, firms reprice every few weeks, and more often when inflation is higher. Attention that firms choose explains the timing of the increase but not the wedge a year later.

The consumer-side explanation in its strongest form is the dual entitlement of Kahneman, Knetsch, and Thaler (1986): a seller is entitled to its reference profit, a consumer to its reference price, and an increase is acceptable when a cost the consumer can see justifies it (see also Okun 1981; Blinder et al. 1998). Eyster, Madarász, and Michaillat (2021) give the theory its modern form. Customers dislike a high markup but cannot observe costs, so they read part of any cost-driven increase as a higher markup, and firms hold back part of their increases in response. This version does not fit our episode, for three reasons. First, a statutory tax change switches the friction off, because the cost change is public and its size is known, so there is nothing left to infer. Second, the theory predicts muted increases, and in this episode the increase is the change that overshoots. Third, their consumer judges the good, not the seller, so the seller's license

of Section 7.4 has no place in the theory.

The entitlement logic itself, without the inference step, matches the timing of the increase and its incidence by product as well as coordination does and needs no further assumption. Every seller becomes entitled to the increase on the same day, January 1, and the entitlement attaches to the good, not to the store, so the decomposition of Section 7.3 favors fairness as much as coordination. It is also consistent with the weight near one in Section 7.6, since an entitled seller restores what it gave up and no more. These facts do not discriminate between the two explanations.

The two explanations differ in what each has to assume for the remaining three facts. The first is the overshoot, which dual entitlement cannot produce on its own. It entitles the seller to restore the pre-cut relative price, and if the reference profit stays at its pre-cut level it predicts nothing more: the treated–control wedge should return to zero, and we find that it does not. The wedge is +0.3 index points before the cut and +6.4 a year after the repeal on the goods that stayed under the price freeze. To match this, the seller's reference profit must have moved up to the windfall it earned while the tax was off. Nothing in the model implies that it did, so this has to be assumed. The second is the result of Section 7.4. To produce a difference of 9.6 log points on the same barcode between the stores that signed the agreement and the stores that did not, consumers must grant different entitlements to different types of seller. They must know the published 7 percent agreement and apply it, on the same good, only to the stores that signed it, an unattractive assumption.

The third is the cut. Dual entitlement is explicit that when costs fall the seller may keep the windfall: the consumer's reference price is unchanged, so a seller that does not cut its price treats nobody unfairly. Fairness therefore predicts low pass-through of a tax cut, and we find the opposite. On the weekly panel, about 70 percent of the cut was passed through within two weeks,[39] and sellers did it by cutting nominal prices rather than by

[39] This exceeds the plateau estimates of Section 5.6, for the reasons given in Appendix B.

holding them flat under inflation. Treated nominal prices fell 4.2 index points a week over the cut while control prices rose 2.4 (Section 5.6). The theory can accommodate this if the cut, being publicized, lowered consumers' reference prices, so that not passing it through would have violated the entitlement after all. But that requires the consumer's reference to be flexible downward in August and sticky upward in January, two opposite asymmetries, each assumed separately. Coordination obtains both from one condition, that each seller gains when its rivals raise their prices, which makes an announcement that points upward easy to follow and one that points downward not (Section 7.2).

In a model of search with reference prices similar to Bénabou and Gertner (1993), a publicly announced tax rise makes a high posted price less informative about a seller's markup, since consumers can now attribute it to the tax. Consumers then search less, sellers can sustain more varied markups, and the dispersion of identical barcodes' prices across stores should rise after the repeal and stay high, whereas coordination predicts the opposite. We test this prediction and find that the sign of the estimate depends on the specification. On a 24-month balanced panel, dispersion in late 2020 is +0.029 under the category-by-month fixed effects we impose elsewhere and -0.004 under barcode and month fixed effects alone. But requiring a balanced panel drops 40 percent of the barcodes, and under the same specification that panel does not find the compression during the holiday that the fuller sample shows, returning -0.010 against -0.018.[40] We do not read a significant estimate for late 2020 from a panel that misses an effect we know is there as evidence of a rise. One confounder remains, and it works toward finding a rise: stockouts and gaps in assortment during COVID plausibly widen dispersion for reasons that have nothing to do with search. The price freeze is less of a problem than it appears, since Precios Máximos covers the control categories as well as the treated ones (Section 5.7), so its compression of the upper tail affects treated and control alike. The test neither confirms nor excludes search.

[40] Standard errors 0.009, 0.003, 0.013 and 0.006.

There are two other explanations. The rockets-and-feathers literature (Borenstein, Cameron, and Gilbert 1997) documents the asymmetry across many markets but is a description, not a mechanism, and each of the mechanisms proposed for it reduces to one of the explanations above. A cost shock at the repeal that raised what sellers paid for the treated goods but not for the control goods beside them would also leave a lasting wedge, and we cannot rule it out. The flat pre-period of Table 5 shows only that no such shock arrived before the repeal.

No test in these data separates coordination among sellers from a consumer-side explanation given enough extra assumptions, and we do not believe data of this kind offer one. The evidence points in both directions. The seller's license favors coordination, because a consumer-side explanation can match it only by assuming that consumers grant different entitlements to different types of seller. The weight near one and the flat response to local competition do not favor coordination, and both fit a reference price at least as well. Fairness, in turn, needs a separate assumption about the consumer's reference for the overshoot and for the fast pass-through of the cut.

A final point makes the choice matter less than it appears. The assumption fairness needs for the overshoot, a reference profit that moved up to the windfall earned during the VAT-free window, makes the reference depend on the history of the episode rather than on fundamentals alone, as multiple equilibria do. A fairness model with enough added assumptions to match the overshoot shares that property with coordination, with consumer memory playing the role that coordination among sellers plays. Eyster, Madarász, and Michaillat (2021) call this acclimation: the customer's reference adjusts until the prevailing markup is judged acceptable. The paper's central claim, that a fully reversed shock left a product-level trace in prices that the menu cost and Calvo models cannot produce, holds whether sellers or consumers carry the history, and the open question is which of the two does.

8. Conclusion

This paper tests workhorse models of price adjustment on a shock with a property those models are never confronted with: it was completely undone. Four of the six facts we document (immediate adjustment, no anticipation of the pre-announced repeal, a repricing frequency that moves with the shocks and with inflation, and cross-store compression of identical-product prices) establish that prices in this market behave as the state-dependent literature says prices behave. The last two, which Benzarti, Garriga, and Tortarolo (2024) document for this episode, are the ones that literature has not been able to test. Prices respond roughly twice as much to the VAT increase as to the VAT cut. The overshoot this leaves does not decay. A year after a 21 percent tax had been removed and restored in full, prices of uncapped treated goods still stood about six index points above the control group on the goods that stayed under the price freeze.

The menu cost model we estimate, with occasionally large firm-specific shocks, trend inflation and a free Calvo weight, reproduces the first four facts in sign and cannot produce the last one. The failure comes from the structure of the model, not from its calibration. With a single long-run equilibrium, the post-repeal wedge stays open only while some firms have not yet repriced, and every month some of them do. Changing the parameters makes the wedge close faster or more slowly, but it always closes. Something these models leave out must carry the history of the episode. It works product by product, not store by store, and how much a seller raised its price in January tracked its own institutional number, the negotiated cap or full restoration of the tax, rather than the product it was selling. These facts fit sellers coordinating on public numbers, but the increase does not look like a common change added to every price, and the overshoot does not fall where a good has more sellers. These data cannot settle whether sellers or consumers carry the history, and the claim that something must carry it holds either way.

The policy reading is direct, because governments have begun using temporary consumption tax cuts as instruments against inflation on the presumption that the repeal undoes the cut. In this episode it did not. On goods the cap did not cover, the holiday held prices about thirteen index points below the control group for the four and a half months it lasted, and the repeal left them about five points above. Within roughly a year of the repeal, the lasting increase had cost consumers as much as the holiday saved them, and the wedge was still open when our data end. The overshoot is absent where the government negotiated a cap on the repeal increase, which is where the mandate bound (Benzarti, Garriga, and Tortarolo 2024). A temporary consumption tax cut without a plan for its repeal is not a lasting transfer to consumers. On these goods it was a transfer to consumers for about a year after the repeal, and, with the wedge still open when our data end, a transfer in the other direction after that.

References

Alvarez, Fernando, Martin Beraja, Martin Gonzalez-Rozada, and Pablo Andrés Neumeyer. 2019. "From Hyperinflation to Stable Prices: Argentina's Evidence on Menu Cost Models." *Quarterly Journal of Economics* 134 (1): 451–505.

Alvarez, Fernando E., David Argente, Alberto Cavallo, and Francesco Lippi. 2026. "Price Setting During a Currency Changeover." NBER Working Paper 35790, National Bureau of Economic Research.

Ball, Laurence, and N. Gregory Mankiw. 1994. "Asymmetric Price Adjustment and Economic Fluctuations." *Economic Journal* 104 (423): 247–261.

Ball, Laurence, and David Romer. 1991. "Sticky Prices as Coordination Failure." *American Economic Review* 81 (3): 539–552.

Bénabou, Roland, and Robert Gertner. 1993. "Search With Learning From Prices: Does Increased Inflationary Uncertainty Lead to Higher Markups?" *Review of Economic Studies* 60 (1): 69–93.

Benzarti, Youssef. 2025. "Tax Incidence Anomalies." *Annual Review of Economics* 17: 615–634.

Benzarti, Youssef, and Dorian Carloni. 2019. "Who Really Benefits From Consumption Tax Cuts? Evidence From a Large VAT Reform in France." *American Economic Journal: Economic Policy* 11 (1): 38–63.

Benzarti, Youssef, Dorian Carloni, Jarkko Harju, and Tuomas Kosonen. 2020. "What Goes Up May Not Come Down: Asymmetric Incidence of Value-Added Taxes." *Journal of Political Economy* 128 (12): 4438–4474.

Benzarti, Youssef, Santiago Garriga, and Darío Tortarolo. 2024. "Can VAT Cuts and Anti-Profiteering Measures Dampen the Effects of Food Price Inflation?" NBER Working Paper 32241, National Bureau of Economic Research.

Benzarti, Youssef, and Alisa Tazhitdinova. 2021. "Do Value-Added Taxes Affect International Trade Flows? Evidence from 30 Years of Tax Reforms." *American Economic Journal: Economic Policy* 13 (4): 469–89.

Blanchard, Olivier Jean, and Nobuhiro Kiyotaki. 1987. "Monopolistic Competition and the Effects of Aggregate Demand." *American Economic Review* 77 (4): 647–666.

Blinder, Alan, Elie R. D. Canetti, David E. Lebow, and Jeremy B. Rudd. 1998. *Asking About Prices: A New Approach to Understanding Price Stickiness*. New York: Russell Sage Foundation.

Borenstein, Severin, A. Colin Cameron, and Richard Gilbert. 1997. "Do Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes?" *Quarterly Journal of Economics* 112 (1): 305–339.

Byrne, David P., and Nicolas de Roos. 2019. "Learning to Coordinate: A Study in Retail Gasoline." *American Economic Review* 109 (2): 591–619.

Calvo, Guillermo A. 1983. "Staggered Prices in a Utility-Maximizing Framework." *Journal of Monetary Economics* 12 (3): 383–398.

Caplin, Andrew S., and Daniel F. Spulber. 1987. "Menu Costs and the Neutrality of Money." *Quarterly Journal of Economics* 102 (4): 703–725.

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

Coibion, Olivier, Yuriy Gorodnichenko, and Saten Kumar. 2018. "How Do Firms Form Their Expectations? New Survey Evidence." *American Economic Review* 108 (9): 2671–2713.

Cooper, Russell, and Andrew John. 1988. "Coordinating Coordination Failures in Keynesian Models." *Quarterly Journal of Economics* 103 (3): 441–463.

Eyster, Erik, Kristóf Madarász, and Pascal Michaillat. 2021. "Pricing under Fairness Concerns." *Journal of the European Economic Association* 19 (3): 1853–1898.

Gagnon, Etienne. 2009. "Price Setting during Low and High Inflation: Evidence from Mexico." *Quarterly Journal of Economics* 124 (3): 1221–1263.

Gagnon, Etienne, David López-Salido, and Nicolas Vincent. 2013. "Individual Price Adjustment along the Extensive Margin." *NBER Macroeconomics Annual* 27 (1): 235–281.

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

Green, Edward J., and Robert H. Porter. 1984. "Noncooperative Collusion under Imperfect Price Information." *Econometrica* 52 (1): 87–100.

Kahneman, Daniel, Jack L. Knetsch, and Richard Thaler. 1986. "Fairness as a Constraint on Profit Seeking: Entitlements in the Market." *American Economic Review* 76 (4): 728–741.

Karadi, Peter, and Adam Reiff. 2019. "Menu Costs, Aggregate Fluctuations, and Large Shocks." *American Economic Journal: Macroeconomics* 11 (3): 111–146.

Krugman, Paul. 1991. "History versus Expectations." *Quarterly Journal of Economics* 106 (2): 651–667.

Maćkowiak, Bartosz, and Mirko Wiederholt. 2009. "Optimal Sticky Prices under Rational Inattention." *American Economic Review* 99 (3): 769–803.

Miller, Nathan H., and Matthew C. Weinberg. 2017. "Understanding the Price Effects of the MillerCoors Joint Venture." *Econometrica* 85 (6): 1763–1791.

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

Nakamura, Emi, and Jón Steinsson. 2010. "Monetary Non-Neutrality in a Multisector Menu Cost Model." *Quarterly Journal of Economics* 125 (3): 961–1013.

Nakamura, Emi, and Jón Steinsson. 2013. "Price Rigidity: Microeconomic Evidence and Macroeconomic Implications." *Annual Review of Economics* 5 (1): 133–163.

Okun, Arthur. 1981. *Prices and Quantities: A Macroeconomic Analysis*. Washington, DC: Brookings Institution Press.

Rotemberg, Julio J., and Garth Saloner. 1986. "A Supergame-Theoretic Model of Price Wars during Booms." *American Economic Review* 76 (3): 390–407.

Appendix A. Model Derivations

Derivation of the Calvo pricing equation

We derive the reset price in Section 4 from four assumptions, following Nakamura and Steinsson (2013) and Benzarti, Garriga, and Tortarolo (2024, Appendix B.2). First, labor is the only variable input and technology is linear, so log nominal marginal cost equals the log nominal wage, $mc_t = w_t$. Second, demand for product i is $y_{it} - y_t = -\theta(p_{it} - p_t)$, with y_{it} log demand for product i, p_{it} its log nominal price and y_t log real aggregate output. With β the discount factor per period, the reset price is

$$p^*_{it} = \big(1 - (1-\alpha)\beta\big) \sum_{j=0}^{\infty} \big((1-\alpha)\beta\big)^j E_t mc_{t+j}, \tag{A1}$$

a discounted average of marginal cost over the periods the price is expected to last.

Third, log nominal output $m_t = y_t + p_t$ is a random walk, $m_t = \mu + m_{t-1} + \epsilon_t$, and we set the drift μ to zero for simplicity. Fourth, household utility is $c_t - L_t$, with c_t log consumption and L_t labor, which makes labor supply perfectly elastic, so that $c_t = w_t - p_t$. Nominal output then equals marginal cost, $m_t = y_t + p_t = c_t + p_t = (w_t - p_t) + p_t = w_t = mc_t$, and since $m_t = m_{t-1} + \epsilon_t$, expected marginal cost in every future period is today's nominal output, $E_t mc_{t+j} = E_t m_{t+j} = m_t$. Substituting gives $p^*_{it} = \big(1-(1-\alpha)\beta\big) m_t \sum_{j=0}^{\infty} \big((1-\alpha)\beta\big)^j = m_t$, the reset price used in the text.

Computation of the quantitative model

The model of Section 6 is solved by value function iteration on a uniform grid of 1,201 points for the log price gap on $[-0.9,\ 0.9]$, with the shock mixture discretized on 41 nodes spanning five standard deviations of the wider component; iteration continues to a sup-norm tolerance of 10^{-12}, and the policy (the inaction band and the reset point) is

stable long before the value function converges. The episode of Section 6.3 is solved by backward induction: the five VAT-free months are computed from the stationary post-repeal solution backwards, so that each month's policy anticipates the announced restoration date. Simulations use 200,000 firms with a 120-month burn-in and common random numbers across the treated and control groups, so that treated–control differences are free of simulation noise; all seeds are fixed. The calibration minimizes the distance to the four moments of Section 6.2 by Nelder–Mead, with each candidate parameter vector solved on a grid of 601 points with 21 shock nodes and evaluated on common-random-number simulations of 20,000 firms over 140 months. The distance has more than one local minimum, so we start Nelder–Mead from the best points of a Latin-hypercube screen of the parameter bounds and from 32 random starting points, check the result with differential evolution, and report the lowest minimum.

Period-by-period predictions

Section 4.3 summarizes the frequency and dispersion predictions of the two models; this subsection records the frequency predictions of the menu cost model and the dispersion predictions of both models period by period, keyed to the panels of Figure 2.

Frequency of price changes, menu cost model. If the Ball and Mankiw (1994) dynamics hold, we should observe:

a. A similar frequency of price increases and decreases prior to the VAT cut (this is the parallel trend assumption).
b. A decrease in the frequency of price increases after the VAT cut. This corresponds to periods 0 to 8 in Figure 2B, when nominal prices are held fixed and real prices decrease because of inflation.
c. A similar frequency of price increases and decreases in the treatment and control groups in periods 9 to 23 in Figure 2B.

d. An increase in the frequency of price increases in the treatment group relative to the control group when the VAT is increased (period 24).
e. A decrease in the frequency of price increases for a few periods after the VAT increase (periods 24 to 32).
f. A similar frequency of price increases and decreases in the control and treatment groups after that.

If the Ball and Mankiw (1994) predictions do not hold, we should observe:

a. A similar frequency of price changes outside of the periods when the VAT rate changes.
b. An increase in the frequency of price decreases and a decrease in the frequency of price increases when the VAT is cut (period 0 in Figure 2B).
c. An increase in the frequency of price increases and a decrease in the frequency of price decreases when the VAT is increased (period 24 in Figure 2B).

Price dispersion, Calvo model. If the Calvo model prevails, we should observe the following patterns:

a. Price dispersion should be similar in the control and treatment groups prior to the VAT cut.
b. The VAT cut will prompt a proportion α of firms to cut prices, while a proportion $(1 - \alpha)$ will keep prices fixed, thus leading to higher price dispersion.
c. Since the price adjustment process occurs in a staggered fashion, price dispersion will progressively increase, until the majority of firms have adjusted their prices to account for the lower VAT. After this point, price dispersion will start decreasing until it converges to that of the control group.
d. We should observe similar patterns for the VAT increase: progressively increasing price dispersion in the treatment group, which will start decreasing after it reaches a peak and converging to the levels of the control group. The difficulty with the VAT increase is that it is anticipated and so these patterns might start before the VAT

increase takes place, making the VAT-increase-based test weaker.

Price dispersion, menu cost model. Assuming that the Ball and Mankiw (1994) dynamics do not hold, we should observe the following price dispersion patterns under the menu cost model assumptions:

a. Price dispersion will be similar in the control and treatment groups when there are no VAT changes.
b. When the VAT is cut or increased, we should observe a drop in price dispersion, which will progressively revert back to that of the control group after a certain number of periods.

If the Ball and Mankiw (1994) dynamics hold, we should observe:

a. Similar levels of price dispersion in the treatment and control groups prior to the VAT cut.
b. Constant price dispersion in the treatment group in periods 0 to 8 in Figure 2B, since firms are keeping nominal prices fixed and waiting for inflation to adjust real prices.
c. A decrease in price dispersion starting in period 9, which should follow the trends of the control group (but not necessarily have the same levels).
d. A substantial decrease in price dispersion when the VAT is increased (period 24).
e. Constant price dispersion in periods 25 to 32, as nominal prices are held fixed.
f. Price dispersion will decrease in period 32, and will follow the trends of the control group.

Appendix B. Additional Results and Robustness

The non-food control group

Section 2 raises the possibility that consumers substitute from control food toward the cheaper treated goods, which would contaminate the food control group. The non-food

control group addresses it, since substitution from detergent to rice is not a concern. On the Córdoba store panel, comparing uncapped treated goods against the 636 non-food barcodes with store–barcode cell and month fixed effects, the difference between the treated and control groups is −12.1 log points (s.e. 1.5) in October 2019 and −11.8 (s.e. 1.5) in December. The cut leg, which is what demand substitution would bias, is thus confirmed against a control group that rules substitution out. The post-repeal comparison is weaker by construction. No product category contains both food and non-food goods, so this contrast cannot use category-by-month fixed effects, and the composition movements they absorb in the main text remain in it. The post-repeal gap is positive but imprecise, +1.9 (s.e. 1.8) in February 2020 and +2.7 (s.e. 2.3) in March, and rising rather than falling. The overshoot estimates in the text therefore rest on the within-category food comparisons, and the non-food group's job is the cut leg and the SUTVA concern. The decomposition of Section 7.3 reports its store- and barcode-margin results against the non-food group directly in the note to Table 6, with the same conclusion.

The group also supplies a check on the capped-goods arm of Fact 6, and the check matters because the within-food identification there is narrow. On the store panel the treated–control contrast is identified by the categories containing both treated and control barcodes, which on capped goods reduces to oils and dried pasta (rice's control side at the single-store retailers is a single cell), so the by-seller capped estimates of Section 5.7 rest on about fifty-five control barcodes. Compared instead against the non-food basket in the same stores, the capped-goods full-episode wedge is +3.1 log points (s.e. 1.1) among the single-store retailers, −5.4 (s.e. 1.2) among the regional chains and −6.9 (s.e. 0.9) among the national chains. Non-food rose faster than control food in these months, so every level shifts down; the cross-seller ordering that carries the argument (the sellers the cap never bound end highest, the sellers it bound end lowest) survives at a spread of ten log points. A related problem leads us to set one estimate aside. The

within-category *cut-leg* contrast at the single-store retailers is identified off control cells that are largely frozen over that window, and it reverses sign accordingly.

The August 2019 depreciation

The peso lost value against the US dollar four days before the cut (Section 2). If treated goods respond more to the exchange rate than control goods, part of every wedge measured from before the cut comes from the depreciation and not from the tax. We find that Facts 5 and 6 hold once we take that part out (Table A1). The uncapped clean-window wedge falls from +5.5 to +5.1 index points and the October–December 2020 wedge from +8.3 to +7.8. In Fact 5, uncapped pass-through at the cut rises from 28 to 34 percent while the repeal leg stays at 65 percent, so prices still respond almost twice as much to the increase as to the cut.

We measure how strongly each group of goods responds to the exchange rate from the depreciation of 2018, as Benzarti, Garriga, and Tortarolo (2024) do. Between the first half of 2018 and the end of the year the dollar roughly doubled against the peso and then held. Over the same months capped goods rose about ten log points relative to control, while uncapped goods barely moved. If both groups responded to the 24 percent rise of the dollar in August 2019 as they did in 2018, the depreciation raised capped prices by about three percent relative to control and uncapped prices by less than half a percent. We subtract these amounts from every wedge measured from before the cut. The repeal leg does not change, because the depreciation came before both of its windows.

We depart from the companion paper in two ways. First, we measure the response separately for capped and uncapped goods. Second, we measure it against the same control goods as the estimate it corrects: all control goods for the weekly panel, and control goods in the same category for the monthly and store panels. Within category the uncapped goods responded more in 2018, so the uncapped estimates on those two

panels receive a larger correction than on the weekly panel.

Two results weaken. First, among the smaller stores of the national panel, capped goods no longer show an asymmetry, as the sum of their two legs goes from +1.8 to −0.4 log points. Their gap from the same goods at the large chains, 4.7 log points in Section 5.6, narrows to 4.1. Second, the capped-goods wedge of the single-store retailers on the store panel falls from +6.4 to +4.7 log points.

The correction in Benzarti, Garriga, and Tortarolo (2024) is larger for uncapped goods, because it pools all treated goods and the pooled response comes from the capped ones. Subtracting its 1.44 index points from every post-cut window still leaves +4.1 in the clean window, +6.9 in October–December and +5.6 log points for the uncapped sum in Fact 5.

Window-endpoint robustness for Fact 6

The clean-window average of +5.5 index points does not depend on the window's endpoints. Truncating the window from the right, the mean over the first five to nine clean weeks runs from +5.10 to +5.51; truncating from the left, the mean from the second, third, fourth or fifth clean week onward runs from +5.45 to +5.63. Every truncation lies between +5.1 and +5.6, and the left-truncated means exceed the right-truncated ones because the wedge rises over the window, which is the opposite of decay.

Three pass-through figures

Three pass-through figures appear in this paper and they measure different things, so they should not be differenced against one another. The figures in Section 5.6 are plateau to plateau on the national monthly panel, averaging the whole VAT-free window against the whole pre-period. The 70 percent quoted in Section 7.7 is the weekly chain panel measured two weeks after the cut, from the last full week before the tax changed, where

TABLE A1. Main estimates net of the August 2019 depreciation

	Main text	Corrected by group	Corrected as in Benzarti, Garriga, and Tortarolo (2024)
Weekly chain panel, index points			
Uncapped, clean window	5.51	5.06	4.07
	(1.19)	(1.28)	(1.19)
Uncapped, October–December 2020	8.30	7.79	6.86
	(1.56)	(1.65)	(1.56)
Capped, clean window	−3.31	−6.84	−4.75
	(0.58)	(0.61)	(0.58)
Capped, October–December 2020	−1.61	−5.63	−3.05
	(0.83)	(0.86)	(0.83)
National monthly panel (Fact 5), log points			
Uncapped, pass-through at the cut (%)	28	34	36
Uncapped, pass-through at the repeal (%)	65	65	65
Uncapped, sum of the two legs	7.09	5.89	5.65
	(2.19)	(2.33)	(2.19)
Smaller stores, capped, sum of the two legs	1.82	−0.43	0.38
	(1.01)	(1.05)	(1.01)
Córdoba store panel, log points			
Uncapped, within category, March 2020	7.4	5.6	6.0
	(1.9)	(2.1)	(1.9)
Single-store retailers, capped, full episode	6.4	4.7	5.0
	(1.9)	(1.9)	(1.9)

Standard errors in parentheses; in the middle column they include the uncertainty of the 2018 response. The 2018 response is the treated–control change in log prices between the first 24 weeks of 2018 and the weeks 2018w45 to 2019w30, a period over which the dollar rose 93 percent against the peso. On the weekly chain panel it is estimated with the fixed effects of specification (3) on the cells observed in every week from 2018w1 to 2020w10, and is 10.5 log points (s.e. 0.5) for capped goods and 1.2 (1.3) for uncapped ones against all control goods, and 6.7 (1.3) and 7.1 (3.4) against control goods in the same category. On the national monthly panel, with the estimator of Fact 5 and January–April against October–December 2018, it is 7.9 (0.9) and 4.4 (2.8). Each response is scaled by the ratio of the 24 percent rise of the dollar in August 2019 to the 2018 rise. The Córdoba store panel has no 2018 data and takes the weekly within-category responses. The last column subtracts 1.44 points from every estimate measured from before the cut. The exchange rate is the Central Bank's official wholesale rate (Communication A 3500).

the treated–control log wedge reaches −13.3 log points. Part of the gap is the window, and it is visible within a single dataset: on the weekly panel the same plateau-to-plateau estimator returns 60 percent at the cut, a log wedge of −11.5, and not 70, because the log wedge erodes on average as the nominal price level rises, reaching −11.9 log points by the last week of December. The larger part, 60 percent there against 44 among the

chains on the monthly panel, is the two panels covering different sellers.

Capped categories

Table 5 reports the capped columns alongside the uncapped ones in every period. The capped wedge is -3.3 index points in the clean window and -1.6 by the end of 2020, against +5.5 and +8.3 uncapped. For the sellers the negotiated cap binds, it eliminates the overshoot rather than dampening it. Section 7.4 develops this contrast, and the pooled permutation result below reflects it.

Coarse clustering

Treatment varies at the level of product categories, so barcode-level clustering may understate the standard errors. Clustering instead on category-by-area-by-treatment cells (100 clusters) leaves the point estimates unchanged and the conclusions intact: the VAT-free wedge is -13.2 with a wild cluster bootstrap 95 percent interval of $[-14.6,\ -11.8]$, and the clean-window wedge is +5.5 with an interval of $[3.1,\ 7.8]$; both p-values are below 10^{-3} (Webb weights, 9,999 draws).

Design-based inference

Treatment was assigned at the level of product types rather than barcodes, so alongside the barcode-clustered standard errors of the main text we report inference that uses only the assignment itself. The unit is the subcategory of the weekly chain panel. Thirty-eight subcategories are pure, thirteen treated and twenty-five control; four are mixed, containing treated and control barcodes together, and are set aside here, though they carry about fifteen percent of cells. For each window we compute the difference between the mean subcategory index change from the pre-cut period, treated less control, weighting subcategories equally, and compare the observed value with its distribution

over 20,000 random reassignments of the thirteen treated labels.

The cut is beyond doubt on this metric: the observed VAT-free wedge of −14.6 index points is more extreme than every one of the 20,000 permutations. The post-repeal windows require more care, in two respects. First, the overshoot of Fact 6 is a property of the goods the cap did not cover. About seven in ten of their cells sit in subcategories with no control goods (sliced bread, rice and two yogurt subcategories), where capped and uncapped goods are mixed, and the rest sit in the two subcategories that contain control goods (canned fruit and oil). The uncapped overshoot therefore cannot be permutation-tested at this level. The weekly estimates in the main text compare it with all control food, with barcode, area and week fixed effects and clustering on barcode. Second, the statistic that can be permutation-tested pools capped and uncapped goods, and there the observed values are −2.7 in the clean window and −1.1 over the last nine weeks of 2020, with permutation p-values of 0.32 and 0.78. The pooled post-repeal wedge is thus indistinguishable from zero. That is the seven percent cap showing through, as in the capped column of Table 5, and not evidence against the uncapped overshoot. Leave-one-out confirms that no single subcategory carries the pooled statistics: dropping any one control subcategory, including the largest, moves them by at most 1.2 index points, and dropping any one treated subcategory by at most 2.5.

The store panel permits an exercise the weekly panel does not. We report it although it fails, because the failure shows what limits the test. The unbound sellers, the regional chains and the single-store retailers, were never party to the agreement, so for them the capped categories are informative and the treated set is no longer confined to mixed subcategories: of the thirty subcategories there, twenty-seven are pure, ten treated and seventeen control, with three mixed set aside. On the same design (subcategory means of the full-episode response, June–July 2019 to January–February 2020, subcategories weighted equally, 20,000 reassignments of the ten treated labels), the cut leg is confirmed wherever prices move: the cut-window difference is −13.1 log points among the regional

chains and -12.2 among the national ones, each more extreme than all but a handful of reassignments ($p = 10^{-4}$), and -11.2 pooling the unbound sellers ($p = 3 \times 10^{-4}$). The full-episode statistic clears significance nowhere: -3.1 among the regional chains ($p = 0.30$), +2.5 among the single-store retailers ($p = 0.55$), and +4.2 ($p = 0.46$) restricting their treated set to the five fully capped subcategories. Two constraints bind, and neither is the cap. Among the single-store retailers the design has no power at all, since the cut leg itself is -0.9 ($p = 0.69$), because half of their cells are frozen in any month and their pass-through arrives after the September–October window. Everywhere else the statistic is dominated by relative price movement across subcategories. Over the seven months of the full episode the control subcategory means alone range from 17 to 41 log points within a store type, and twenty-seven units cannot separate a five-to-seven point effect from movement of that size, though they can separate a thirteen-point one. The conclusion of the previous paragraph therefore stands on the store panel as on the weekly one: the overshoot is a within-category object, and within-category estimation with clustering on barcode is the inference the object admits.

Sales and regular prices

Our price is a unit value including discounts, and Coibion, Gorodnichenko, and Hong (2015) document that regular, sale and effective prices behave differently in retail scanner data. The overshoot could therefore come, in principle, from fewer promotions on treated goods after the repeal and not from a higher regular price. We therefore re-estimate the central quantities on a sale-filtered series. A week begins a sale if the log unit value falls at least 5 log points below the previous week and returns to within one log point of the pre-dip level within four weeks; sale weeks are replaced by the pre-dip price. A permanent decrease never returns and is never filtered, so the VAT cut itself passes through the filter untouched. A coarser filter, a 10 log point dip with eight weeks

to return, gives materially identical results throughout.

The filter finds real sales, and their frequency moved with the tax. About 15 percent of control cell-weeks sit inside a filtered sale, a share that is flat across the whole sample; treated incidence is 14 percent before the cut, rises to 18–20 percent during the VAT-free window, and falls to 9–12 percent after the repeal. Promotions amplified the cut and were withdrawn at the repeal, which is the direction that would produce a spurious overshoot in unit values. Four results say it does not. First, the overshoot is unchanged: the clean-window wedge is +6.4 index points (s.e. 1.2) filtered against +5.5 (s.e. 1.2) raw, and October–December 2020 is +8.9 (s.e. 1.6) against +8.3 (s.e. 1.6); the filtered series carries a pre-cut placebo of +1.3 (s.e. 0.3), and measured against it the filtered wedges are +5.1 and +7.6, against +5.4 and +8.2 raw measured the same way. The capped-goods null is also unchanged, at −0.2 filtered against −1.6 raw over October–December. Second, the cut leg is smaller on regular prices: the VAT-free plateau is −11.6 against −13.2, and the wedge two weeks after the cut is −8.0 against −13.2. Part of the immediate response therefore ran through promotions (deep temporary cuts, later re-set), and the asymmetry of Section 5.6 sharpens on regular prices rather than weakening. Third, the ordinary-week participation floor of Section 5.4 is mostly sales churn: over 2019w20–w31, filtering cuts it from 11 to 3 percent for large decreases and from 16 to 8 percent for large increases. The event spikes survive (29 percent of treated uncapped cells post a filtered large fall in the second week of the cut against a 3 percent floor, and 72 percent post a filtered large rise in the second week of the repeal against an 8 percent floor), so the repeal wave in particular is almost entirely a regular-price wave. Fourth, the unfiltered series reproduces Table 5 to the second decimal, because both exclude the same observations: the 93 with a nonpositive unit value and the other 102 weeks of the one barcode–area cell whose base-week unit value is negative. One limit belongs with the result. The filter operates at the frequency and aggregation of the weekly panel, a unit value across the chain stores of an area, so it removes area-wide temporary cuts and

cannot see store-level promotions that average out; the store panel is monthly, where no comparable filter is available.

Appendix C. Data Appendix

Treatment classification

Treated categories are those in the Annex to Decree 567/2019, listed in Table 1; a small number of treatments apply at the barcode rather than the category level, and the detailed product descriptions in the data allow those cases to be classified directly. Control food comprises the remaining food categories in the data, and the non-food control group used in Appendix B comprises 636 barcodes of household and personal-care products.

The weekly chain panel

The weekly panel covers the national supermarket chains at the barcode-by-area level, ten areas, from the first week of 2019 through the end of 2020. The analysis sample keeps barcode–area cells observed in every week from 2019w1 through 2020w10, excludes the barcodes of the Precios Esenciales and Precios Cuidados programs (whose prices were subject to their own agreements), and excludes the toasts and flours categories; cap status at the repeal is assigned by the category-level classification of the negotiated agreement. The outcome is the unit value normalized to 100 in 2019w32 for each cell. We drop the 93 observations whose unit value is zero or negative, most of them weeks with missing revenue, and the one cell whose base-week unit value is negative, which would otherwise turn its whole series negative. The "balanced" sample of Table 5 additionally requires a cell to appear in every week of 2019 and 2020; the final calendar week of 2020 cannot be dated in the 52-week calendar and is excluded from estimation throughout.

The Córdoba store panel

The store panel is built from the raw store-level file, which carries 173 supermarkets. Nonpositive unit values, which arise from returns and reporting errors, are treated as missing, and the Precios Esenciales and Precios Cuidados barcodes are excluded. We use the panel in two versions, both running from January 2019 through March 2020, and Table 2 gives the counts for each. The balanced version keeps store–barcode cells with a usable price in every month. The balance requirement is why it has 172 stores against the raw file's 173: the remaining store, a single-store retailer in Las Higueras, reports a usable unit value in only 27 percent of its rows, so none of its barcodes has a price in all fifteen months. Fact 4's dispersion measure further restricts to barcodes sold in at least five stores in a given month. The unbalanced version keeps every store and every cell with a usable price in at least one month, and takes the cap status of each treated barcode from the code of Benzarti, Garriga, and Tortarolo (2024), which agrees with our own classification on every barcode the two share.

The balanced version carries the store-level event study of Section 5, Fact 4's dispersion estimate and selection regressions, the license regression, bunching shares, seller-type medians and nominal-move shares of Section 7.4, the decomposition of Section 7.3, and the non-food comparisons by store type. The unbalanced version carries Fact 3 and the calibration moments of Section 6, the store-type splits of Fact 4, Fact 6's within-category check and full-episode wedges by store type, the store-type weight, the cell-level comparison and the local-competition test of Section 7.6, the full-episode wedges and Figure 11 in Section 7.4, and the store-panel permutation test of Appendix B.

The national monthly panel

The national monthly panel is the companion paper's, built from the full national collection of more than 3,000 supermarkets and aggregated to barcode-by-region-by-

store-type cells; Benzarti, Garriga, and Tortarolo (2024) document its construction. We use it as delivered, with the plateau windows and cell definitions described in Section 5.6.