–APPENDIX–

Scraped Data and Sticky Prices

Alberto Cavallo
MIT Sloan

A Appendix

A.1 Treatment: Missing Values and Sales

Missing values and sales are treated using standard methods in the literature. Table A1 illustrates these treatment decisions using a hypothetical individual price series.

Table A1: Treatment of Price Spells - Sales and Missing Values

An Example

Price Recorded	•	*$ 2*	*$ 1*	•	*$ 3*	*$ 2*	•
Sale Indicator		*R*	*S*		*R*	*R*	
Complete Including Sales	•	*$ 2*	*$ 1*	*$ 1*	*$ 3*	*$ 2*	•
Complete Excluding Sales	•	*$ 2*	*$ 2*	*$ 2*	*$ 3*	*$ 2*	•

Notes: (•) represents a missing value. *R* is a regular price, *S* is a sale price.

Missing values are common within price series because products are either out of stock or not correctly recorded in the database on a particular day. Depending on the country, the percentage of these missing values is between 10.75% to 22.62%, as shown in Table 2.[34] Given the high frequency of the data, and the fact that missing gaps do not typically extend for more than a few days, I complete missing values in the price series by imputing the last recorded price until a new price is available.

There are also missing values at the beginning or at the end of a price series. Some of these are caused by new product introductions or goods that are discontinued. Others arise due to left and right censoring in the sample. Since no information of prior or posterior prices is known for these goods, these missing values are not replaced or altered in any way.

Sale events are identified explicitly by the supermarkets with a sales tag. They represent a considerable share of all price observations: 5.78% in Argentina, 2.94% in Brazil, and 6.26% in Colombia (there is no sales indicator in Chile). To exclude sales in some of my results, I replace sale prices with the last recorded *non-sale* price, as shown in Table A1. This means that if the price of a good drops with a sale but later returns to exactly the same value, the non-sale price series will have no variation at all.

Finally, a few price changes in each country seem implausibly large and may be the result of a coding mistake. Table 2 shows they are a negligible part of all observations, but they

[34]Klenow and Kryvtsov (2008) report 12% in monthly US CPI data.

can affect statistics related to the magnitude of price change. Consequently, all daily price changes that exceed 500% are excluded.

A.2 Sales

This section presents general sale statistics in each country. Table 4 shows that price decreases are a significant share of total price changes in each country. This is particularly surprising in Argentina, where price *decreases* are 32% of all price changes and inflation is three times higher than in other countries. Nakamura and Steinsson (2008) found a similar share in the US, with CPI data collected during a period of much lower inflation. Roughly half of the price decreases in Argentinca are caused by sale events.

Indeed, sales can have significant impact on sticky-price facts, as Nakamura and Steinsson (2008) showed. Threfore, in Table A2 I present detailed sales statistics in all countries where sale data are available. In all cases, sale events tend to last only a few days, with a median length between 6 to 13 days.[35] Sales represent between 8.8% and 14% of all price changes, and between 22% and 45% of all price decreases.

Table A2: Sale Events by Country

	Argentina	Brazil	Colombia
Life of sales (in days, median)	6	7	17
Sales as % of price changes	23%	29%	12%
"Sales" with no price change*	25%	18%	10%
"Sales" with price increases*	2.4%	0.4%	3.6%
Sales that end with old price (v-shaped)	86%	60%	53%
Sales that end with higher price	6%	12%	17%
Mean size of sale, given price decrease	-11%	-17%	-21%

Notes: *Price changes occurring on the day a sale indicator appears next to the product.

Surprisingly, many sales are not associated with a reduction in prices at all. If we consider a three-day window around the announcement of a sale, between 10% and 25% of sale events are not related to any price changes. [36] Even more surprisingly, in Argentina and Colombia about 3% of sales are linked to price *increases*. This suggests that an important degree of asymmetric information exists between retailers and consumers, with sale announcements sending misleading signals to consumers who are unable to monitor prices on a daily basis. This behavior is stronger in high-inflation Argentina, where sales also tend to be shorter, more v-shaped, and smaller in size (given a price decrease). Some of the facts suggest that sales play a role in pricing strategies that standard sticky-price models cannot fully explain. For example, sales could potentially play a role in "fair pricing" models such as Rotemberg (2008), as a way to reduce the negative impact that price increases are having on customer anger.

[35]This is an important cause of measurement error when using monthly prices because many sales are either not recorded at all, or assumed to last over a month. See the Appendix for more details.

[36]Price changes occurring within +/- one day from the date a sale indicator appears next to the product. If multiple changes were present during that period, only price decreases were counted.

A.3 Frequency and Inflation

To measure the degree of price stickiness in each country, I follow the frequency approach now standard in the empirical sticky-price literature.[37] This method provides a single parameter to reflect the unconditional probability that a firm will change its price over a given period of time (a day in this case).

To obtain country-level frequencies, I first obtain the daily frequency per individual good by computing the number of daily price changes over the number of total valid change observations for a particular product. Next, I calculate the *mean* frequency per good category, and finally, the *median* frequency across all categories. Given that CPI expenditure weights are different across countries, I use unweighed medians to facilitate cross country comparisons. I also calculate implied durations using $1/-ln(1-frequency)$, which implicitly makes the simplifying assumption of constant hazards, where the probability of a price change is independent from the amount of time elapsed since the previous adjustment. Implied durations provide another intuitive way to compare the degree of price stickiness across countries.

Table 3 presents each country's frequency and duration estimates. A puzzling result is that, counter to standard predictions in sticky-price models, countries with higher inflation rates can also have stickier prices. In particular, Argentina is a puzzling case: the annual inflation rate is three times the level of any other country in the sample, but price changes are relatively sticky with an implied median duration of 66 days, 25% longer than in Colombia and 70% longer than in Brazil. The median frequencies of price increases and decreases, when considered separately, also fail to significantly correlate with inflation levels.

The only frequency statistic that is strongly correlated with inflation across countries is the relative frequency of increases over frequency of decreases. This is also the case in the cross-section of products within each country, as seen in Table A3. This suggests that the overall degree of stickiness is less important than the relative flexibility of increases over decreases in order to predict the short-term effects of monetary policy on inflation and output.

Table A3: Correlation of Frequency and Inflation - Individual Products

	Argentina	Brazil	Chile	Colombia
	Inflation	Inflation	Inflation	Inflation
Frequency	0.05	0.06	0.03	0.05
Frequency +	0.23	0.15	0.13	0.19
Frequency -	-0.14	-0.05	-0.08	-0.08
Frequency +/Frequency-	0.54	0.33	0.49	0.37

Argentina's low frequency is not explained by a difference in the type of goods being sampled in each country. As can be seen in Table A4, Argentina is stickier than Brazil in 54 out of 60 common categories (or 90%). An alternative explanation for Argentina's relative stickiness is the existence of price controls, but Table A4 that Argentina is stickier than Brazil even in categories which are never under price controls. In a previous version of this paper, I showed that price controls are correlated with *higher* frequencies at the good

[37] See Bils and Klenow (2004), Nakamura and Steinsson (2008) and Gopinath and Rigobon (2008)

level. If these controls are playing a role in increasing stickiness, they must be affecting pricing behaviors in unrelated goods categories. It is likely that firms are afraid of being "selected" by the government for future controls, or concerned about boycotts and other negative consequences associated with frequent price changes.

A.4 Non-Parametric Tests of Modality

The departure from unimodality in the distributions shown in Figures 1 and 2 can be testes formally with two non-parametric tests: Hartigan's Dip and Siverman's Bandswidth tests. Non-parametric methods are important to avoid making ex-ante assumptions on the number of modes.

Hartigan's Dip is a simple test of unimodality. It relies on the fact that the cumulative distribution function of a density function f with a single mode at m_f is convex on the interval $(-\infty, m_f)$ and concave on the interval (m_f, ∞).[38] In other words, at the left side of the mode, the density is non decreasing, while the opposite occurs at the left of the mode. With this insight, one can find the unimodal distibution that minimizes the difference with the observed empirical distribution. This difference is measured by the dip statistic, which can be used as a sort of "score" to measure the departure from unimodality. Positive dip values provide evidence to reject the null hypothesis of unimodality. To determine the statistical significance of a positive dip, Hartigan and Hartigan (1985) sets the null hypothesis equal to the uniform distribution, for which, asymptotically, the dip value is stochastically largest among all unimodal distributions. Hartigan and Hartigan (1985) also show that this is not always the case with small samples. To address this concern, I use a calibration of the dip test proposed by Cheng and Hall (1998).

Silverman's Bandwidth method can be used to test for multiple modes. It uses the non-parametric smoothed kernel density to evaluate the number of modes in an empirical distribution. The basic insight in Silverman (1981) is that the larger the smoothing applied, the fewer the number of modes in the estimated density. So for the null hypothesis of unimodality, he proposed using as a test statistic the minimum smoothing required for the density to have a single mode. Large values of this statistic (the "critical bandwidth") are evidence against the null hypothesis of unimodality, because they mean that larger degrees of smoothing are needed to eliminate additional modes in the density estimate. The statistical significance of the score can be evaluated using a smoothed bootstrap method.[39]

[38] See Hartigan and Hartigan (1985)

[39] See Henderson et al. (2008) for more details for both statistical tests.

–APPENDIX–

Table A4: Implied Median Duration by Categories

CATEGORY	ARGENTINA	BRAZIL	CHILE	COLOMBIA
FLOUR AND PREPARED FLOUR MIXES	91*	40		66
CEREALS	89	52	109	66
PASTA	92*	40	88	66
RICE	64	40	105	
BREAD	83*	45	88	87
FRESH BISCUITS, ROLLS	119	21	176	
CAKES AND CUPCAKES		61	58	83
COOKIES		34	56	83
CRACKERS AND BREAD & CRACKER PRODUCTS	76*	71	176	
SWEETROLLS, COFFEE CAKE & DOUGHNUTS	89	61	206	
FROZEN BAKERY PRODUCTS				51
PIES, TARTS, TURNOVERS				95
UNCOOKED GROUND BEEF	65*	17	88	
UNCOOKED BEEF STEAKS	92	14	64	46
OTHER UNCOOKED BEEF AND VEAL	370	47	35	
BACON, BREAKFAST SAUSAGE, AND RELATED PRODUCTS	82*	52	88	
HAM	74*	36	58	
PORK CHOPS				55
OTHER PORK INCLUDING ROASTS AND PICNICS			85	
FRANKFURTERS	105	38		49
LAMB, ORGAN MEATS, AND GAME	41		176	
CHICKEN	127	52	58	47
OTHER POULTRY INCLUDING TURKEY			58	
FRESH FISH AND SEAFOOD	78*	94	81	
PROCESSED FISH AND SEAFOOD	51*	61	157	41
EGGS	67	36	117	41
MILK	51*	29	25	78
CHEESE AND RELATED PRODUCTS	92*	45	89	117
ICE CREAM AND RELATED PRODUCTS	185*	23	52	66
OTHER DAIRY AND RELATED PRODUCTS		25	58	55
OTHER FRESH FRUITS	19	28	48	55
LETTUCE	25			
TOMATOES		11		
OTHER FRESH VEGETABLES INCLUDING FRESH HERBS	74*	61	116	83
CANNED FRUITS AND VEGETABLES	32	69	176	111
FROZEN FRUITS AND VEGETABLES	74	40		47
OTHER PROCESSED FRUITS AND VEGETABLES INCLUDING DRIED	126	74	171	111
CARBONATED DRINKS	59*	36	88	55
FROZEN NONCARBONATED JUICES AND DRINKS				141
NON-FROZEN NON-CARBONATED JUICES AND DRINKS	61*	40	172	83
COFFEE		52	176	
TEA	89*	79	353	
SUGAR AND ARTIFICIAL SWEETENERS	96*	66	353	111
CANDY AND CHEWING GUM		30	353	
OTHER SWEETS	82*	75	25	84
MARMALADE & JAMS		73		166
CHOCOLATE	56*	52	168	111
BUTTER AND MARGARINE		25	88	55
OTHER FATS AND OILS	91*	40	119	66
OTHER MISCELLANEOUS FOODS	117*	26		
SOUPS	74*	40	176	33
FROZEN AND FREEZE DRIED PREPARED FOODS	92*	28	88	55
SNACKS		52	52	
SALT AND OTHER SEASONINGS AND SPICES	128	73	179	47
OLIVES, PICKLES, RELISHES		72	176	
SAUCES AND GRAVIES	85	47	176	83
OTHER CONDIMENTS	123	73	353	55
BABY FOOD		21		
PREPARED SALADS	61	9		166
BEER, ALE AND OTHER MALT BEVERAGES AT HOME	106*	45	117	47
DISTILLED SPIRITS AT HOME	68*	52	58	58
WINE AT HOME	61*	52	64	58
CIGARETTES				47
DENTAL & NONELECTRIC SHAVING PRODUCTS	123*	40		93
DEODORANT/SUNTAN PREPARATIONS	80*	95	263	66
COSMETICS NAIL PREPARATIONS & IMPLEMENTS	92*	45	353	66
SHAMPOO, BATH PRODUCTS	87*	28	176	66
BABY CARE PRODUCTS	117*	73	174	83
SANITARY/FOOTCARE PRODUCTS	123	51	353	54
PERFUME	136		167	66
LAMPS & LIGHTING FIXTURES	123	122		
PAINT, WALLPAPER TOOLS & SUPPLIES	123		353	
TOOLS	185			
LAWN & GARDEN SUPPLIES & INSECTICIDES	108*	91	242	
CLEANING PRODUCTS	92*	73	176	97
LAUNDRY PRODUCTS	46*	47	167	41
HOUSEHOLD PAPER PRODUCTS	89*	66	78	165
MISCELLANEOUS HOUSEHOLD PRODUCTS	123	100	353	111
TIRES	49			
VEHICLE PARTS & EQUIPMENT OTHER THAN TIRES	92	73		
TOYS, GAMES, HOBBIES, & PLAYGROUND EQUIPMENT	293	183	41	
PET FOOD	85	52	88	111
PURCHASE OF PETS, PET SUPPLIES, ACCESSORIES	173	91	192	111
OVER-THE-COUNTER DRUGS	359		353	
TOPICALS AND DRESSINGS				66
APPLIANCES	234			
BOOKS				
KITCHEN & DINING ROOM FURNITURE	33			
OUTDOOR FURNITURE	42			
BEDROOM AND BATHROOM LINENS	49			
SUPERMARKET BRAND			117	
INTERNATIONAL PRODUCTS			343	
LOW CALORIES			176	
ORGANIC FOOD	123	73		99
OTHER GOODS		72	176	

*Categories with price controls.