Stylized Facts From Prices at Multi-Channel Retailers in Mexico[1]

Diego Solórzano[2]

Banco de México

CEMLA's XXVI Meeting of the Central Bank Researchers Network
November 11, 2021

---

[2]Directorate General of Economic Research. Email: jsolorzano@banxico.org.mx

Motivation

- Retail price setting is central to a range of economic policy questions.
  Monetary policy, business cycles, income inequality, antitrust analysis, etc.

- Over the last 15 years, price surveys taking place in brick and mortar stores have been the gold-standard for economic research.
  E.g. Microdata from CPI and PPI surveys

- Online prices are increasingly used for measurement and research applications.
  They are cheaper to collect, provide information from all goods on display, numerous products' features, uncensored price spells, collected on a high-frequency basis and without any delays.

Motivation

- There is a growing interest on online price setting and whether they are similar to those observed in (offline) physical stores.

- Differences might arise if
    - Physical stores incur in "menu-costs" when resetting their prices; such costs might be negligible for online websites.
    - Online prices might react promptly to aggregate shocks; offline pricing might react promptly to local shocks.
    - Methodological differences: (offline) sample vs (online) census.

- Understanding these differences (if any) is important as evidence on price rigidity is used to inform macroeconomic modeling.

Research Questions

- Are online price dynamics similar to those observed in brick-and-mortar stores?

- In particular, this paper studies price statistics of product categories in retailers with both online and offline presence and tackles the following questions:
    - Are online prices as sticky as offline prices? In other words, is the frequency of price adjustment the same between online and offline retailing?
    - Given a price change, are the magnitudes of price changes similar online and offline?
    - Do the distributions of price changes differ across sales channels?

- It then analyzes methodological particularities in the online and offline data sources that could potentially explain the differences between them.

- Online prices come from retailers' websites monitored by Banco de México.
  Offline prices are compiled by INEGI as part of the CPI survey.

Where Do We Stand?

- Burgeoning empirical literature on web scraped prices. Along with others:
    - **NSO**: Kony et al. (2019); Flower and Karachalias (2019); Van Loon and Roels (2018); Posse and Reyes (2019). **Central Banks**: Macias et al. (2019); Hull et al. (2017); Peña and Padres (2020). **Academia**: Cavallo (2017, 2018); Cavallo and Rigobon (2016); Aparicio and Bertolotto (2020).

- Cavallo (2017) offers the first large scale comparison of online and offline prices.
    - Compares prices, **at product level**, finding little differences on price levels, they are not fully synchronized but have similar frequencies and sizes of adjustment.
    - An item displayed in the retailer's physical store is generally showed on its website.
    - Studies using few sources of data should be cautious in generalizing. There exists high degree of heterogeneity for different types of retailers, industries and countries.
    - **This paper:** First to compare stylized facts of price setting from **clusters of goods** online and offline.

- Bils and Klenow (2005) and Peña and Padres (2020)
    - Heterogenous price-setting across categories using US CPI and Chilean web scraped data, respectively.

Results

For the retailers and product categories in the sample, price statistics suggest that

- Prices change more frequently in brick-and-mortar stores than in their online channel.

- Given a price change, the magnitude of the change is greater online than in their offline channel.

- For most retailers, product categories changing more frequently online prices are also those adjusting more often offline.

- Online price changes are centered around focal points (multiples of 5% in the $\pm 20\%$ range), while this feature is less prevalent in offline price changes

- Standardized price changes show that online prices adjustments report a smaller fraction of small price variations relative to offline price changes.

Results (cont'd)

- These patterns hold for both non-food and food product categories.

- Also, the above results do not seem to have changed qualitatively in 2020, when the Covid19 pandemic affected offline and online shopping habits.

Potential sources explaining online and offline differences:
- Closer look at the composition of products within and between datasets
    - Price distribution (moments of)
    - Role of flagship products, luxury goods, low-end, special editions, etc.
    - Share of product churn
- Imputations or averages prices in the survey data are discarded as drivers of differences.

As more retailers and product categories are added to the online sample, a more general picture will emerge regarding online price setting in Mexico and how it is related to its offline counterpart.

Outline

Table of Contents

Data at Hand

- The project focuses on 8 retailers with online and offline presence.
  They are known as "multi-channel" by Cavallo (2017).

- The types of goods offered by retailer vary: some are food-oriented; some offer apparel and electronics but no food; some are mixed.

- "Retailer" should be taken as a retail chain.
  They have physical stores across different locations in Mexico

- The comparison of stylized facts is carried out by product category and retailer.
  E.g. The frequency of online price changes for "Refrigerators" is greater than the frequency of offline price changes for Retailer X.

- Aggregates by retailer are then produced in order to have an overall picture.

- Benchmark results come from Mexico's core CPI components.
  Also, because of the Covid19 pandemic, there are two sets of results: pre-2020 and 2020.

Online Price Data

- The online price dataset is compiled at Banco de México

- A robot parses retailers' websites and gathers data from all items displayed at the time of collection. Importantly, it revises the product's identifier and price.

- Goods are then classified into clusters of fairly homogeneous goods known as "product categories" or *genéricos* by INEGI.
  This classification allows the direct comparison with the offline (CPI) survey.
  E.g. Milk, Eggs, Women Trousers, Men Trousers or Televisions.

- The start dates, as well as the frequency of price collection vary by retailer.

- Prices do not include any delivery fee and the dataset does not report any geographical dimension.

Offline Price Data

- Gathered by INEGI, the offline dataset comes from the CPI survey.

- The survey takes place on a regular basis and through direct visits to numerous outlets across Mexico.[3] Hence, no uneven pricing gaps as in the online price survey.

- Data from retailers appearing in the online dataset only.

- The survey considers a sample of goods per product category and retailer.

- Reported prices are not averaged nor imputed prices but actual prices observed 14 or 7 days apart.

- In order to maximize the sample sizes per price category in each retailer, all prices are considered regardless the geographical location of the store.

---

[3]Until March 2020, online price collection was considered by INEGI for few price categories but not for the ones analyzed in this study. Due to the Covid19 pandemic, INEGI followed the retailers in this study through their websites but pricing only a sample (fixed basket) of goods as usual.

Descriptive Statistics

Table: Data by Retailer
: Non-Food Categories Only. Data from 2016 to 2019.

	Online								Offline					
	Start date	End date	Days	Fortnights	Observations (Thousands)	Products (Thousands)	Frequency of Observation (Days)	Sales	Start date	End date	Observations (Thousands)	Products (Thousands)	Outlets Locations	CPI Weight (%)
Retailer 1	01jan2016	01jan2020	1,087	76	1,984.1	4.9	1.3		01jan2016	01jan2020	12.2	0.3	11	0.4
Retailer 2	31may2016	01jan2020	1,280	88	2,069.9	4.8	1.0		01jan2016	01jan2020	31.2	0.7	30	1.9
Retailer 3	21nov2017	01jan2020	627	48	1,228.1	5.2	1.2	✓	01jan2016	01jan2020	168.6	4.8	76	4.6
Retailer 4	21nov2017	01nov2019	440	40	169.2	1.2	1.6	✓	01jan2016	01jan2020	3.4	0.1	8	0.1
Retailer 5	21nov2017	27dec2019	511	47	378.5	2.2	1.5	✓	01jan2016	01jan2020	31.5	1.2	40	0.4
Retailer 6	21nov2017	06aug2019	91	39	918.6	50.6	6.9	✓	01jan2016	01jan2020	133.5	4.7	53	0.6
Retailer 7	11aug2016	29dec2019	320	78	556.8	20.9	3.9	✓	01jan2016	01jan2020	91.2	3.5	42	0.5
Retailer 8	12aug2016	01jan2020	561	83	778.0	62.3	2.2	✓	01jan2016	01jan2020	14.3	0.6	8	0.1

Note: A fortnight is counted if there is at least one day in the fortnight. Fortnights are defined from the 1st until the 15th, and from the 16th until the last day of the month. Observations are the number of prices in the dataset. Products represent the number of unique identifiers in the retailer. Frequency of observation is the mean number of days between price observations. Outlets locations stand for the number of stores in the retail chain priced in the CPI survey. The column CPI weight represents the total weight from the products priced at the retailer (includes weights from food-categories).

Table of Contents

Do Online and Offline Prices Change at the Same Frequency?

- The frequency of price adjustment is defined as the fraction of products reporting a price change.
- It is calculated using one day of the week. The weekday is selected by retailer as the one that maximizes the Spearman correlation coefficient across its categories.

Table: Spearman Correlation of the Frequency of Online and Offline Prices Changes

	Sunday		Monday		Tuesday		Wednesday		Thursday		Friday		Saturday		Avg 7-Days		Max Corr
Retailer 1	0.78	***	0.76	***	0.77	***	0.76	***	0.73	***	0.75	***	0.76	***	0.76	***	Sunday
Retailer 2	0.63	***	0.64	***	0.64	***	0.64	***	0.68	***	0.64	***	0.65	***	0.65	***	Thursday
Retailer 3	-0.14		-0.15		-0.14		-0.15		-0.13		-0.13		-0.13		-0.14		Saturday
Retailer 4	0.18		0.20		0.21		0.17		0.10		0.29		0.25		0.20		Friday
Retailer 5	0.55	***	0.55	***	0.60	***	0.67	***	0.62	***	0.55	***	0.55	***	0.58	***	Wednesday
Retailer 6							-0.54	***			0.66	***			0.06	***	Friday
Retailer 7	0.09										0.21		0.35	**	0.22		Saturday
Retailer 8											0.27		0.36	**	0.31	*	Saturday

Note: Each Spearman correlation coefficient is calculated as follows. First, price changes are calculated as the 14-day log price difference using the weekday in question only. Then, the frequencies of online price adjustments by category and retailer are calculated. Finally, the frequency of online price adjustments is compared with the frequency of offline price changes via a Spearman correlation coefficient. The column under the title "Avg 7-Days" considers the 14-day apart price changes using all days and not one day at the time. Figures in boxes are the maximum by retailer. The lack of online price collection in certain days for few retailers prevents calculating the Spearman correlation, generating few empty cells in the table. Data from 2016 to 2019 and non-food categories only. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
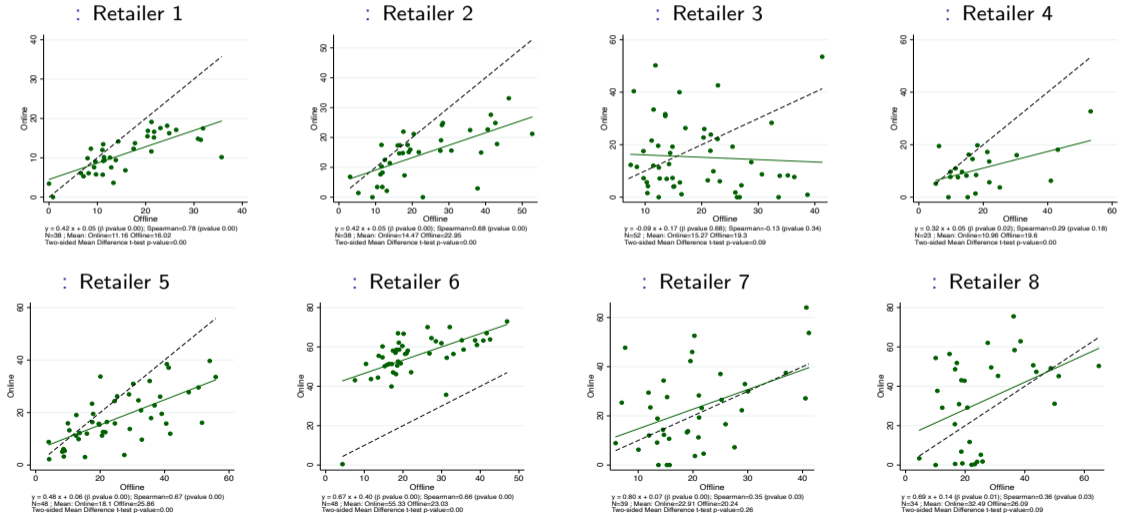
Frequency of Online and Offline Price Changes

Table: Frequency of Price Adjustments by Retailer

| | Frequency of Price Changes | | | | Equality Test | | | Spearman | | Linear Fit | | |
| | Average | | Ho:Equality | Categories | Categories | | | ρ | Ho: $\rho=0$ | β | Ho: $\beta=0$ | Ho: $\beta=1$ |
	Online	Offline	p-value		On > Off	Equal	Off > On		p-value		p-value	p-value
Retailer 1	11.16	16.02	0.00	38	1	16	21	0.78	0.00	0.42	0.00	0.00
Retailer 2	14.47	22.95	0.00	38	2	9	27	0.68	0.00	0.42	0.00	0.00
Retailer 3	15.27	19.30	0.09	52	15	9	28	-0.13	0.34	-0.09	0.68	0.00
Retailer 4	10.96	19.60	0.00	23	1	10	12	0.29	0.18	0.32	0.02	0.00
Retailer 5	18.10	25.86	0.00	48	5	19	24	0.67	0.00	0.48	0.00	0.00
Retailer 6	55.33	23.03	0.00	48	47	0	1	0.66	0.00	0.67	0.00	0.02
Retailer 7	22.91	20.24	0.26	39	14	10	15	0.35	0.03	0.80	0.00	0.45
Retailer 8	32.49	26.09	0.09	34	18	4	12	0.36	0.03	0.69	0.01	0.26

Note: Using data from 2016 to 2019 and non-food categories only. The average frequency of online price changes comes from the day maximizing the Spearman between online and offline prices. Averages are calculated across unweighted product categories. The third column reports the p-value from a two-sided t-test of mean differences. Spearman and linear fit are computed using the categories within each retailer. Beta coefficients come from regressions using online prices as dependent variable, offline prices as the independent variable, plus a constant. The Equality Test columns report the number of categories where, using a two-sided z-test of proportion differences, the difference is statistically insignificant at 5% (Equal), greater fraction of online price changes than offline changes at 5% significance level (On > Off), and greater proportion of offline price changes than online adjustments at 5% significance level (Off > On).

Frequency of Online and Offline Price Changes

Figure: Frequency of Price Adjustments by Retailer
: Non-Food Categories. Data from 2016 to 2019.

: Retailer 1

y = 0.42 x + 0.05 (β pvalue 0.00); Spearman=0.78 (pvalue 0.00)
N=38 ; Mean: Online=11.16 Offline=16.02
Two-sided Mean Difference t-test p-value=0.00

: Retailer 2

y = 0.42 x + 0.05 (β pvalue 0.00); Spearman=0.68 (pvalue 0.00)
N=38 ; Mean: Online=14.47 Offline=22.95
Two-sided Mean Difference t-test p-value=0.00

: Retailer 3

y = -0.09 x + 0.17 (β pvalue 0.68); Spearman=-0.13 (pvalue 0.34)
N=52 ; Mean: Online=15.27 Offline=19.3
Two-sided Mean Difference t-test p-value=0.09

: Retailer 4

y = 0.32 x + 0.05 (β pvalue 0.02); Spearman=0.29 (pvalue 0.18)
N=23 ; Mean: Online=10.96 Offline=19.6
Two-sided Mean Difference t-test p-value=0.00

: Retailer 5

y = 0.48 x + 0.06 (β pvalue 0.00); Spearman=0.67 (pvalue 0.00)
N=48 ; Mean: Online=18.1 Offline=25.86
Two-sided Mean Difference t-test p-value=0.00

: Retailer 6

y = 0.67 x + 0.40 (β pvalue 0.00); Spearman=0.66 (pvalue 0.00)
N=48 ; Mean: Online=55.33 Offline=23.03
Two-sided Mean Difference t-test p-value=0.00

: Retailer 7

y = 0.80 x + 0.07 (β pvalue 0.00); Spearman=0.35 (pvalue 0.03)
N=39 ; Mean: Online=22.91 Offline=20.24
Two-sided Mean Difference t-test p-value=0.26

: Retailer 8

y = 0.69 x + 0.14 (β pvalue 0.01); Spearman=0.36 (pvalue 0.03)
N=34 ; Mean: Online=32.49 Offline=26.09
Two-sided Mean Difference t-test p-value=0.09

Is the Average Size of Price Adjustments Similar Across Channels?

Table: Size of Price Changes

| | Size of Price Adjustments | | | | Equality Test | | | Spearman | | Linear Fit | | |
| | Average | | Ho:Equality | Categories | Categories | | | ρ | Ho: $\rho = 0$ | β | Ho: $\beta = 0$ | Ho: $\beta = 1$ |
	Online	Offline	p-value		On > Off	Equal	Off > On		p-value		p-value	p-value
Retailer 1	12.69	10.95	0.01	31	15	13	3	0.49	0.01	0.44	0.01	0.00
Retailer 2	24.29	9.72	0.00	33	27	6	0	-0.24	0.18	-0.21	0.71	0.04
Retailer 3	22.65	13.47	0.00	41	34	5	2	0.19	0.23	0.21	0.48	0.01
Retailer 4	22.22	14.27	0.00	13	10	2	1	0.45	0.13	0.57	0.10	0.21
Retailer 5	15.79	9.44	0.00	41	30	9	2	0.28	0.08	0.15	0.75	0.08
Retailer 6	19.16	16.85	0.00	48	26	9	13	0.23	0.11	0.16	0.38	0.00
Retailer 7	21.45	14.13	0.00	33	24	5	4	0.12	0.50	0.48	0.23	0.19
Retailer 8	20.93	18.64	0.10	25	10	10	5	0.32	0.12	0.26	0.12	0.00

Note: Using data from 2016 to 2019 and non-food categories only. Prices changes do not consider the sign of adjustment (absolute value). The average size of online price changes comes from the weekday maximizing the Spearman between online and offline prices. Averages are calculated across unweighted product categories. The third column reports the p-value from a two-sided t-test of mean differences. Spearman and linear fit are computed using the categories within each retailer. Beta coefficients come from regressions using online prices as dependent variable, offline prices as the independent variable, plus a constant. The Equality Test columns report the number of categories where, using a two-sided t-test of mean differences, the difference is statistically insignificant at 5% (Equal), greater absolute value mean size of online price changes than offline changes ($On > Off$) or greater offline price adjustments than online changes ($Off > On$) at 5% significance level.
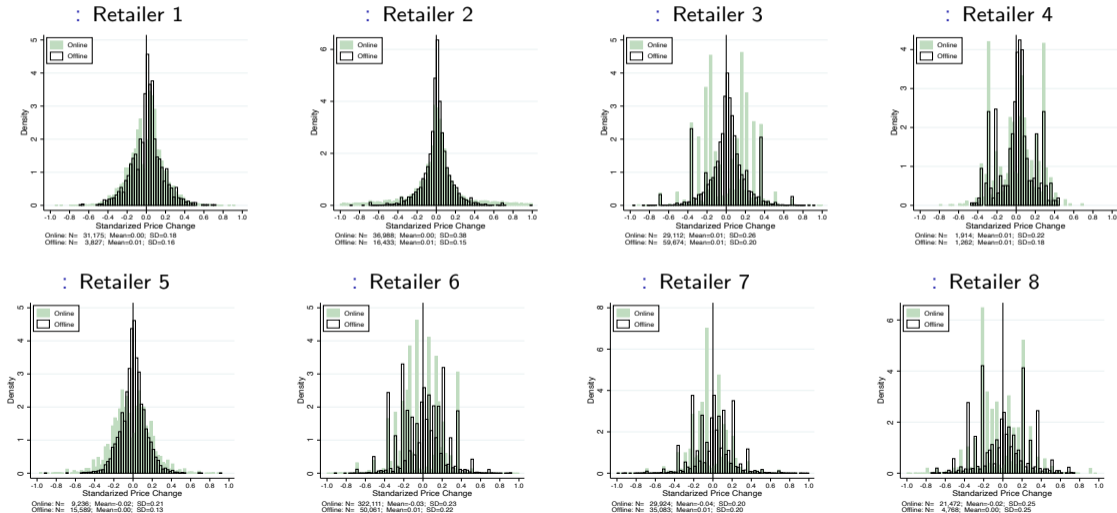
Average Size of Price Adjustments

Figure: Size of Price Changes
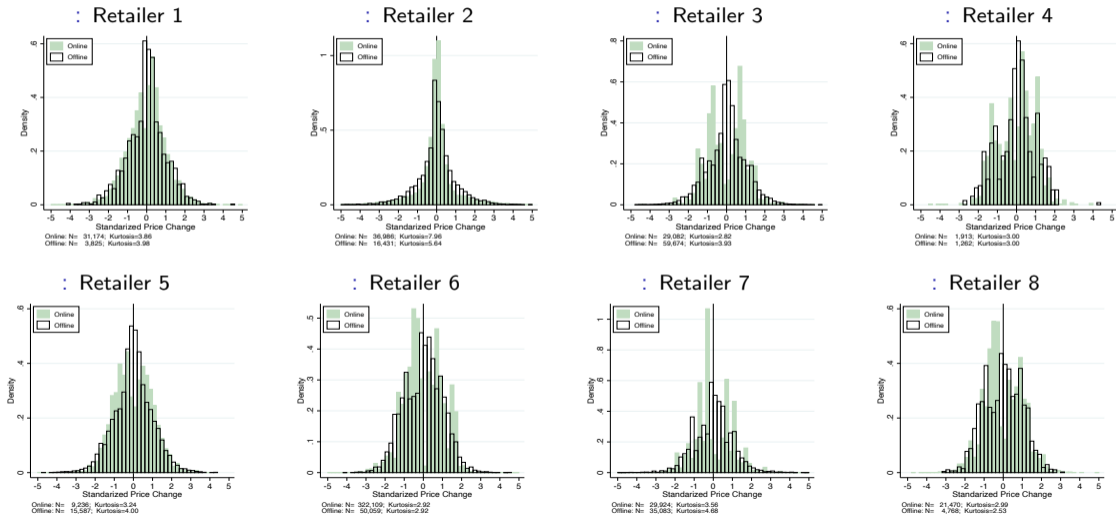: Non-Food Categories. Data from 2016 to 2019.
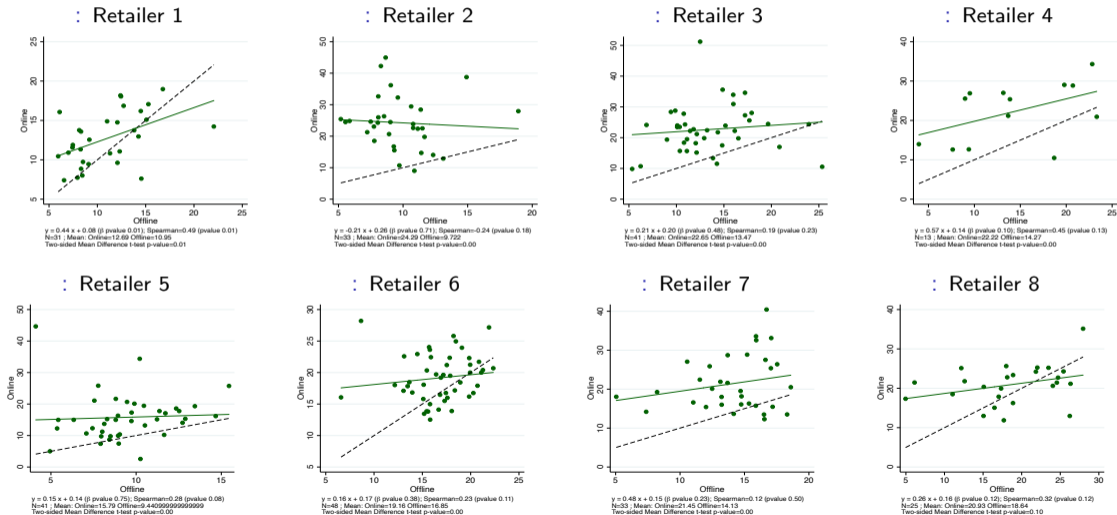
Distribution of Price Changes

Figure: Distribution of Non-Zero Price Changes
: Non-Food Categories. Data from 2016 to 2019.

Distribution of Standardized Price Changes

Figure: Distribution of Standardized Non-Zero Price Changes
: Non-Food Categories. Data from 2016 to 2019.

In the paper...

- Results are similar for food-categories.

- Closer look at the 2020 data:
  - The frequency of price changes increased, on average, by around 5 p.p. in both sales channels for the product categories and retailers in the study.
  - In general, the average size of price adjustments did not change relative to previous years.

- Evidence on the differences in the product composition across sales channel: Universe (online) vs sample (offline) of products.
  - Greater average price in the online samples than in the offline counterparts.
  - More product churn (turnover) online than offline.

Table of Contents

Takeaways

- Online prices are increasingly available for measurement and research applications.
  Cheaper to collect, information from all goods on display and collected on a high-frequency basis.

- This paper shows that the stylized facts about price changes differ across sales channels for eight multi-channel retailers in Mexico.

- Methodological differences should be taken into account when comparing price statistics based on survey data to price moments drawn from big data.

- This is important as evidence on price rigidity is used to inform macroeconomic modeling.