PUEY UNGPHAKORN INSTITUTE
FOR ECONOMIC RESEARCH

The Movement and Change in Online Price Within and Across Selected Major Retail Stores in Thailand

by

Attakrit Leckcivilize

February 2019

Discussion Paper

No. 105

The movement and change in online price within and across selected major retail stores in Thailand[1]

Attakrit Leckcivilize

Abstract

E-commerce has gained larger market shares in Thailand over the last decade. Yet there is a paucity of studies on online price behaviour and movement. This project is one of the first attempts to explore this topic in the Thai context. Using web scraping technique to acquire the data on price and product information from major retailers that have both physical and online outlets, this paper summarizes its findings into six stylized facts. In short, online price changes more frequent than its offline counterpart, yet the magnitudes of changes are generally much larger. Further, price heterogeneity exists across stores and product categories. However, pricing strategies of the same store seems to differ between its online and offline outlets.

Keywords: *online price, price movement, Thailand, pricing strategy*

[1] The author would like to thank Dr Piti Disyatat, Dr Pym Manopimoke and the research team at PIER for their guidance, help and support. Special thanks to the programmers, Dr Suchit Pongnumkul and Dr Suporn Pongnumkul, as well as the research assistant teams for their dedicated works. The author gratefully acknowledges a support by the Puey Ungphakorn Institute for Economic Research (PIER) Funding and Grants. All of the remaining errors in this report are my own.

1 Introduction

E-commerce and online platform have become another important channel of product distribution. With wider internet access through smartphones, Thai people across different age groups have some experiences buying or selling products online. According to Statista (2018), revenue in the E-commerce market in Thailand steadily grows to the level of US$3,648m in 2018. Given the rising market share of e-commerce, it would be of the interest of policy makers and the public to explore the nature and movement of online price. Hence, this project was set out with three objectives. First, it aims to study the online price movement among selected major retail stores in Thailand (these stores have many chains of actual stores across the country) including frequency, direction and size of the changes. Secondly, it attempts to study behaviours of online sale and promotion. Thirdly, this project provides some evidence on discrepancies in product types, promotion and prices between off- and on-line stores of the same retailer.

In this paper, we discuss the method used to collect the data and their characteristics in Section 2, while challenges during the data cleansing, processing and analysing are presented in Appendix A. In Section 3, we describe our methodology to explore frequency and size of price changes based on statistics commonly applied in the literature. At first, we plan to compare these statistics with some product categories in the CPI. However, without extra programming involving some machine learning techniques, it is too tedious to match all products online to those CPI categories by hand.

Due to an exploratory nature of this paper, we decide to summarize our findings into "six stylized facts". These outlines accompanying by Figures and Tables are presented in Section 4. They demonstrate some differences in price movement as well as magnitudes of changes across stores and product categories. Yet some Figures, especially in Appendix B, show signs of possible outliers. This potential measurement errors seem to be persistent in our data as a result of issues affected data quality and remaining flaws even after the cleansing process.

In a sub-section of Section 4, we compare our scraped data with the actual stores. Despite perfect matches in product information of most products on- and off-line, there are some discrepancies and concerns surrounding product availability as well as the use of quantity related promotion and giveaway. Section 5 suggests possible moves for the next step of this project with one specific example and future works in this field. We conclude in Section 6.

2 Data collection and availability

We employ a technique called web scraping to acquire the data on price and product information. Web scraping is a set of software techniques to extract information from target websites automatically. The main goal is to extract data embedded in webpages and save them in structured databases (such as spreadsheets) for further uses. Web scraping is useful for data collection tasks that target either large quantity data or frequent data collection, which cannot be easily or reliably done manually by human.

During the period from late August 2016 to late September 2017, information on four online shopping platforms of major stores in Thailand was scrapped daily by automated programme. The shop names were anonymized as Shop A, B, C and D. However, due to changes that the retailers made with their websites (please see more discussion on data collection and challenges in Appendix A), there were days when we failed to collect data from the stores' websites, particularly a failure in data collection of all stores from March to mid-April 2017 on top of failures of some websites in some days. Hence, price data from each of these four shops can be classified into three periods, i.e. all daily data successfully scraped, part one of daily data (with small gaps of missing data) and part two. We will use the data from part 2 covering mid-April to early August / mid-September 2017 for our main analysis. Detail information on data availability and product categories sold on these platforms is as follows:

- 241 days of data were collected from Shop A[2] with 1 day that numbers of products were successfully scraped were too few[3]. There were two periods with long discontinuities (more than 5-10 days) of the data collection from Shop A: From January to mid-April 2017 and from mid-August to late September 2017. The distribution of days the products appearing online in our database by the stored own defined categories[4] are shown in Table 1 -3:

(Days)	Beverages	Food	Health & Beauty	Home appliances & Electronic products	Household	Mom, baby & kids	Pet, Outdoor, etc.	Snacks & Sweets	Stationaries	No info	Total
1-30	670	2,123	1,416	145	463	139	127	421	2	1	5,507
31-60	311	729	339	189	324	94	329	308	20	24	2,667
61-90	158	103	242	270	360	87	359	161	43	0	1,783
91-120	96	57	159	146	170	37	113	81	35	0	894
121-150	84	62	219	96	115	36	97	58	15	0	782
151-180	144	82	172	310	282	131	274	62	218	0	1,675
181-210	322	260	542	101	611	100	506	254	47	0	2,743
211-240	1,067	1,345	1,526	370	1,384	572	563	913	158	0	7,898
NA	30	68	0	0	0	0	0	9	0	0	107
Total	2,882	4,829	4,615	1,627	3,709	1,196	2,368	2,267	538	25	24,056

Table 1: Numbers of Product appearing on Shop A website by categories: All (days)

(Days)	Beverages	Food	Health & Beauty	Home appliances & Electronic products	Household	Mom, baby & kids	Pet, Outdoor, etc.	Snacks & Sweets	Stationaries	No info	Total
1-20	7	18	53	12	18	9	36	10	4	0	167
21-40	75	66	132	82	47	14	44	36	7	0	503
41-60	98	13	25	27	55	3	60	17	1	0	299
61-80	106	84	181	141	304	75	230	125	44	0	1,290
81-100	42	19	46	96	73	11	105	3	2	0	397
Almost Every Day	1,534	1,732	2,285	790	2,262	816	1,323	1,257	440	0	12,439
Total	1,862	1,932	2,722	1,148	2,759	928	1,798	1,448	498	0	15,095

Table 2: Numbers of Product appearing on Shop A website by categories (days): Part 1 (29/08/2016 – 29/01/2017 with gaps)

[2] Shop A is a nation-wide hypermarket chain with an online platform and large offline stores across the country.
[3] It is defined by having numbers of products less than one-third of the number of products collected from that website on a median day.
[4] The reasons why some products cannot be categorized and named as "No info" are either: (1) missing information from the website (2) the classification on the website does not make sense or is not meaningful, i.e. containing promotion information rather than categories (3) It might also result from a change in the program used to scrap the data after January 2017.

(Days)	Beverages	Food	Health & Beauty	Home appliances & Electronic products	Household	Mom, baby & kids	Pet, Outdoor, etc.	Snacks & Sweets	Stationaries	No info	Total
1-20	892	2,738	994	200	678	175	104	692	14	25	6,512
21-40	25	32	88	237	79	8	93	22	11	0	595
41-60	120	45	140	256	350	153	548	78	231	0	1,921
61-80	728	772	1,095	200	1,129	181	761	565	63	0	5,494
81-95	831	954	1,252	431	1,063	479	457	686	170	0	6,323
Almost Every Day	49	84	51	5	37	62	23	48	0	0	359
Total	2,645	4,625	3,620	1,329	3,336	1,058	1,986	2,091	489	25	21,204

Table 3: Numbers of Product appearing on Shop A website by categories (days): Part 2 (18/04/2017 – 09/08/2017)

- Shop B[5] was not in our list for web scraping initially. Therefore, the data collection started from late November 2016 to late September 2017 and it covered 204 days in total. However, their website did not provide meaningful or well-defined product categories. Therefore, all Shop B results will be presented without product categories as shown in Table 4 – 6:

(Days)	Count	Percent
1-30	2,876	25.70%
31-60	870	7.80%
61-90	1,000	8.90%
91-120	881	7.90%
121-150	1,280	11.40%
151-180	1,247	11.10%
181-204	3,047	27.20%
Total	11,201	100.00%

Table 4: Numbers of Product appearing on Shop B website: All (days)

(Days)	Count	Percent
1-20	1,050	14.63%
21-40	1,123	15.65%
41-55	1,140	15.88%
Every Day	3,864	53.84%
Total	7,177	100.00%

Table 5: Numbers of Product appearing on Shop B website (days): Part 1 (27/12/2016 – 08/03/2017)

(Days)	Count	Percent
1-20	1,937	20.84%
21-40	587	6.31%
41-60	878	9.44%
61-80	762	8.20%
81-99	2,237	24.06%
Every Day	2,895	31.14%
Total	9,296	100.00%

Table 6: Numbers of Product appearing on Shop B website (days): Part 2 (16/04/2017 – 08/08/2017)

[5] Shop B is a nation-wide chain specializing only in some product categories such as home appliances and electrical products. It has both an online platform and large offline stores across the country albeit fewer stores than other three shops.

- Shop C[6] had 276 days of successfully scrapped data. Shop C changed the URL of its website in late September 2016, so our data collection started from there to mid-September 2017. Although it has the largest number of products sold online, their online products are very different from their offline stores. Particularly, the online platform focuses more beauty, fashion, home & garden products. These products account for more than half of their products sold online by Shop C as presented in Table 7 – 9:

(Days)	Beauty	Electronic & Entertainment	Fashion	Health	Home appliances	Home & Garden	IT, Camera & Gadget	Mobile, Tablet & Accessories	Mom, baby & kids	Supermarket	No info	Total
1-30	1,818	191	2,216	309	422	731	299	1,156	462	224	6	7,834
31-60	1,102	200	1,398	164	334	722	436	559	519	254	17	5,705
61-90	768	124	1,156	302	307	949	256	383	629	452	1	5,327
91-120	456	86	748	72	292	573	392	156	208	162	1	3,146
121-150	563	176	747	105	262	422	167	109	142	158	2	2,853
151-180	928	46	1,063	181	303	687	588	205	436	150	4	4,591
181-210	636	82	826	113	184	1,101	437	190	508	175	0	4,252
211-240	1,166	159	501	196	374	499	557	277	432	154	0	4,315
241-270	995	111	1,160	280	443	451	414	228	232	61	0	4,375
Almost Every Day	1,681	171	739	683	732	1,769	752	425	479	685	0	8,116
Total	10,113	1,346	10,554	2,405	3,653	7,904	4,298	3,688	4,047	2,475	31	50,514

Table 7: Numbers of Product appearing on Shop C website by categories: All (days)

(Days)	Beauty	Electronic & Entertainment	Fashion	Health	Home appliances	Home & Garden	IT, Camera & Gadget	Mobile, Tablet & Accessories	Mom, baby & kids	Supermarket	No info	Total
1-20	1,519	170	1,232	176	125	528	219	1,023	357	59	0	5,408
21-40	1,204	234	1,020	181	395	625	657	590	640	268	0	5,814
41-60	781	214	777	332	492	426	397	398	590	80	0	4,487
61-80	690	232	1,411	486	632	670	560	297	508	161	0	5,647
Every Day	3,095	87	1,429	576	740	2,164	868	760	837	1,093	0	11,649
Total	7,289	937	5,869	1,751	2,384	4,413	2,701	3,068	2,932	1,661	0	33,005

Table 8: Numbers of Product appearing on Shop C website by categories (days): Part 1 (20/09/2016 – 27/12/2016)

(Days)	Beauty	Electronic & Entertainment	Fashion	Health	Home appliances	Home & Garden	IT, Camera & Gadget	Mobile, Tablet & Accessories	Mom, baby & kids	Supermarket	No info	Total
1-20	333	24	926	138	247	414	149	91	126	96	16	2,560
21-40	580	25	584	70	311	278	152	96	65	57	1	2,219
41-60	631	42	691	106	143	296	190	163	392	135	1	2,790
61-80	525	65	747	176	126	405	49	114	94	242	0	2,543
81-100	195	20	447	71	57	288	181	118	192	30	0	1,599
101-120	558	90	372	86	237	337	693	124	201	55	1	2,754
121-140	414	164	470	79	85	270	127	103	59	144	2	1,917
141-155	4,466	45	3919	98	177	4286	287	211	112	41	4	13,646
Every Day	0	473	0	1203	1752	147	1962	859	1644	1036	0	9,076
Total	7,702	948	8,156	2,027	3,135	6,721	3,790	1,879	2,885	1,836	25	39,104

Table 9: Numbers of Product appearing on Shop C website by categories (days): Part 2 (16/04/2017 – 18/09/2017 without interruption)

[6] Shop C is a nation-wide chain of convenient stores. Although it has both an online platform and small offline stores across the country, the products sold are quite different in terms of product categories and size/volume of the similar products.

- Shop D[7] had 302 days of successfully scrapped data with 57 days that numbers of products successfully scrapped were too few. Data from Shop D suffer from two failures in data collection and unusual long discontinuities between the two successful scrapings (more than 5-10 days), which were from February to mid-April 2017 and in Early August 2017.

(Days)	Baby & Kids	Beverages, Snacks & Desserts	Dry Grocery	Food, Dairy & Bakery	Health & Beauty	Home & Electrical	Household Products	Miscellaneous	Pets	No info	Total
1-40	103	389	67	267	302	61	45	227	14	570	2,045
41-80	69	104	5	76	251	31	123	72	3	751	1,485
81-120	100	367	104	172	409	42	166	166	25	11	1,562
121-160	91	341	164	136	272	25	185	95	10	0	1,319
161-200	299	233	59	152	305	72	98	118	12	0	1,348
201-240	638	721	158	126	2,442	82	661	340	571	0	5,739
241-280	513	2,638	1,479	615	1,745	205	568	308	265	0	8,336
280-294	0	295	675	553	66	0	0	0	0	0	1,589
Total	1,813	5,088	2,711	2,097	5,792	518	1,846	1,326	900	1,332	23,423

Table 10: Numbers of Product appearing on Shop D website by categories: All (days)

(Days)	Baby & Kids	Beverages, Snacks & Desserts	Dry Grocery	Food, Dairy & Bakery	Health & Beauty	Home & Electrical	Household Products	Miscellaneous	Pets	No info	Total
1-20	36	107	58	55	47	1	29	145	0	179	657
21-40	51	219	6	79	15	7	4	0	0	391	772
41-60	332	77	13	96	106	39	19	12	13	467	1,174
61-80	8	51	14	60	89	78	31	1	0	284	616
81-100	17	166	27	51	54	25	21	5	7	11	384
101-120	1,018	120	72	82	2,522	300	1,545	883	854	0	7,396
121-140	252	3,832	1,715	631	2,240	0	0	0	0	0	8,670
Almost Every Day	0	1	675	682	1	0	0	0	0	0	1,359
Total	1,714	4,573	2,580	1,736	5,074	450	1,649	1,046	874	1,332	21,028

Table 11: Numbers of Product appearing on Shop D website by categories (days): Part 1 (29/08/2016–29/01/2017)

(Days)	Baby & Kids	Beverages, Snacks & Desserts	Dry Grocery	Food, Dairy & Bakery	Health & Beauty	Home & Electrical	Household Products	Miscellaneous	Pets	No info	Total
1-20	9	138	34	184	182	27	17	19	1	0	611
21-40	51	136	30	45	175	9	157	144	11	0	758
41-60	39	97	11	26	248	9	64	74	1	0	569
61-80	6	113	30	129	69	5	76	11	5	0	444
81-95	13	276	1,325	1,459	118	11	5	0	0	0	3,207
Almost Every Day	1,481	3,771	1,098	0	4,481	334	1,304	778	853	0	14,100
Total	1,599	4,531	2,528	1,843	5,273	395	1,623	1,026	871	0	19,689

Table 12: Numbers of Product appearing on Shop D website by categories (days): Part 2 (18/04/2017–09/08/2017)

[7] Shop D is another nation-wide hypermarket chain with an online platform as well as large offline stores across the country.

In sum, the largest store by number of products is Shop C, then Shop A, Shop D and Shop B respectively. As for number of days the products appeared online (according to data from Part 2), Shop C, Shop D and Shop B seem to have more than half of their products being sold online almost every day, implying that most of their products are "regular" products, whereas the percentage of regular products seems to be slightly less for Shop A. It might also mean that Shop A introduced more new products online during this period.

3 Methodology

We compute several indicators for price behaviours such as frequency of price change, duration, direction of change and relative size of the changes at the store-level and its own classification of product categories. Assuming that price changes occur at discrete time intervals, we use the frequency approach to describe price stickiness in our data (Aucremanne & Dhyne, 2004; Lunnemann & Wintr, 2006). We define frequency of price changes at the product level of j ($Freq_j$) as the ratio between the number of times a price change was observed and the sum of the number of times that prices changed plus the number of times prices remained unchanged (i.e. the times that the product appears with its price)[8]. We further assume that frequency of price changes and duration of price spells follow an exponential distribution.

Then we can calculate the implied average duration of price spells as:

$$Avg. Duration\ of\ product\ j = -\frac{1}{\ln(1 - Freq_j)}$$

while the implied median duration is:

$$Med. Duration\ of\ product\ j = \frac{\ln(0.5)}{\ln(1 - Freq_j)}$$

Regarding the characteristics of price changes, we count the number of times, which price changed (either increase or decrease) for each product, then report such distribution through tables and figures. In terms of statistics for the size of price changes, we calculate it as percentage of price decrease (or increase) for each product j as follows:

$$Size\ of\ price\ decrease\ for\ product\ j = \left| \frac{\sum_{t=1}^{T_j} Percentage\ decrease\ in\ price_{jt}}{\mathrm{T}_j} \right|$$

Where T_j is the number of times that price reduced, similar formula applies for price increase.

However, due to difficulties to classify each product in our data into the CPI product categories, our results cannot be compared to the existing studies on offline pricing in Thailand (Apaitan, Disyatat, & Manopimoke, 2018). Yet we decide to summarize our preliminary findings as "six stylized facts about online pricing in Thailand". The fifth and sixth facts are a bit explorative. Specifically, in the fifth fact, we discuss some descriptive statistics of sale promotion, while, in the sixth, we compare the online prices with 'on the shelf' price data of selected products

[8] There were days when the price data could not be collected due to challenges mentioned earlier. If the gap in the data is smaller than 3-4 days, those two prices were treated as a two-consecutive day. However, when the gap is too long, they were treated separately or classified as two major discontinuities mentioned in the data section.

collected from the actual stores across the country. These are some examples on how this dataset might be used to analyse pricing strategies and competition in different product categories (on and offline) as well as the role of seasonal effects on pricing and promotion.

4 The six stylized facts

How large is the difference in frequency of price changes across retailers? Does this difference reflect differences in each store's e-commerce strategies? Or does it reflect only different in types of products sold by each retailer? The following facts investigate into the frequency and duration statistics across product categories.

Fact 1: Heterogeneity across product categories or sectors is relevant but seems to be less important than cross-store heterogeneity.

The average frequency and duration for Part 2 in Table 16-18 show that in spite of differences in frequency of price changes among product categories within the same store. Such ranking cannot be applied to other stores. For instance, price changes were very frequent for home appliances & electronic products on Shop A's website with an average duration of only 15.7 days, while it took more than 58.92 and 68.37 days for the similar products of Shop D and Shop C respectively. Likewise, in Shop B where most of its products sold online are in this category, it took 46.21 days on average for the prices to change.

For Shop D, prices of baby & kids were changed the most frequent with around 27.8 days on average followed by household products, pets and fresh food, dairy & bakery respectively. However, for Shop C, none of its product category has the average duration fewer than 53 days with beauty and home & garden as the quickest groups to change.

Despite some expectation that perishable products or fashionable/easily outdated products should experience more price changes, products such as food, fashion or health & beauty seem to rank in the middle for each store. Furthermore, the frequency of price changes does not seem to be related to the number of products sold under each category. Hence, our findings do not demonstrate any clear pattern of underlying factors driving frequency of price changes other than different in each retailer's strategies.

However, owing to lack of actual sale data (or even how active each product was viewed by customers), we cannot rule out each retailer might change the price more or less frequent based on demand and interest of customers. Moreover, there are small differences in the ranking of product categories by average duration of price changes between all data (Table 13 - 15) and part 2 (Table 16 - 18). Thus, it is worth exploring the role of seasonality in the future research as well.

Category	Frequency	Average Duration (days)	Median Duration (days)
Beverages (11%)	0.0275	35.66	24.72
Food (20%)	0.0198	42.69	29.59
Health &Beauty (19%)	0.038	32.24	22.34
Home appliances (6%) & Electronic products	0.0748	23.17	16.06
Household (15%)	0.0321	44.42	30.79
Mom, baby & kids (4%)	0.047	41.13	28.51
Pet, Outdoor, etc. (9%)	0.036	35.96	24.93
Snacks & Sweets (9%)	0.0249	43.38	30.07
Stationaries (2%)	0.0208	45.32	31.41
Total (100%)	0.0333	37.62	26.07

Table 13: Average value of statistics for price changes in Shop A by categories: All[9]

Category	Frequency	Average Duration (days)	Median Duration (days)
Beauty (20%)	0.0102	73.28	50.79
Electronic & Entertainment (2%)	0.0075	86.04	59.64
Fashion (20%)	0.0081	74.97	51.97
Health (4%)	0.0086	100.09	69.38
Home appliances (7%)	0.0089	91.78	63.62
Home & Garden (15%)	0.0126	73.1	50.67
IT, Camera & Gadget (8%)	0.0052	112.1	77.7
Mobile, Tablet & Accessories (7%)	0.0078	82.64	57.28
Mom, baby & kids (8%)	0.008	77.38	53.64
Supermarket (4%)	0.0114	76.9	53.3
Total (100%)	0.0092	80.03	55.47

Table 14: Average value of statistics for price changes in Shop C by categories: All

[9] The numbers of products appearing online in each store are lower than those numbers in the previous section because to calculate the frequency, we include only those products that have at least one two-consecutive price data. Hence, it excludes quite a small number of products with no consecutive price data, for instance appearing only once.

Category	Frequency	Average Duration (days)	Median Duration (days)
Baby & Kids (8%)	0.0347	45.27	31.38
Beverages,Snacks & Desserts (23%)	0.0229	74.19	51.42
Dry Grocery (12%)	0.0173	79	54.76
Fresh Food, Dairy & Bakery (9%)	0.0229	74.31	51.51
Health & Beauty (26%)	0.0181	78.38	54.33
Home & Electrical (2%)	0.011	99.01	68.63
Household Products (8%)	0.0252	57.45	39.82
Miscellaneous (6%)	0.0269	90.08	62.44
Pets (4%)	0.021	57.74	40.03
Total (100%)	0.022	72.5	50.25

Table 15: Average value of statistics for price changes in Shop D by categories: All

Category	Frequency	Average Duration (days)	Median Duration (days)
Beverages (11%)	0.0272	24.17	16.75
Food (20%)	0.0169	29.02	20.12
Health &Beauty (19%)	0.0474	19.91	13.8
Home appliances (6%) & Electronic products	0.0712	15.7	10.88
Household (15%)	0.0371	22.21	15.39
Mom, baby & kids (4%)	0.0552	21.7	15.04
Pet, Outdoor, etc. (9%)	0.0402	21.34	14.79
Snacks & Sweets (9%)	0.0218	29.33	20.33
Stationaries (2%)	0.0384	19.81	13.73
Total (100%)	0.0349	22.73	15.75

Table 16: Average value of statistics for price changes in Shop A by categories: Part 2

Category	Frequency	Average Duration (days)	Median Duration (days)
Beauty (20%)	0.0127	53.66	37.19
Electronic & Entertainment (2%)	0.0082	74.22	51.44
Fashion (20%)	0.0096	60.84	42.17
Health (4%)	0.009	59.03	40.91
Home appliances (7%)	0.011	68.37	47.39
Home & Garden (15%)	0.0129	53.91	37.37
IT, Camera & Gadget (8%)	0.0044	86.07	59.66
Mobile, Tablet & Accessories (7%)	0.0105	61.52	42.65
Mom, baby & kids (8%)	0.0077	64.07	44.41
Supermarket (4%)	0.0129	62.31	43.19
Total (100%)	0.0104	60.57	41.99

Table 17: Average value of statistics for price changes in Shop C by categories: Part 2

Category	Frequency	Average Duration (days)	Median Duration (days)
Baby & Kids (8%)	0.0375	27.83	19.29
Beverages,Snacks & Desserts (23%)	0.0186	44.34	30.73
Dry Grocery (12%)	0.0192	43.46	30.12
Fresh Food, Dairy & Bakery (9%)	0.0257	36.64	25.4
Health & Beauty (26%)	0.018	41.04	28.45
Home & Electrical (2%)	0.015	58.92	40.84
Household Products (8%)	0.0294	31.84	22.07
Miscellaneous (6%)	0.0212	48.29	33.47
Pets (4%)	0.0222	35.61	24.68
Total (100%)	0.0218	40.02	27.74

Table 18: Average value of statistics for price changes in Shop D by categories: Part 2

<u>*Fact 2*</u>*: Online daily prices change relatively frequent but in many stores slightly more than half of products sold did not experience changes in prices during the period of study.*

According to the data from Part 2 in Table 19, more than half of the products sold by Shop A, Shop B and Shop C did not experience any changes in their price during that period. Nevertheless, average frequency of price changes can be ranked from Shop A (the highest) to Shop D, Shop B and Shop C (the lowest) respectively. Despite gaps and discontinuities, the statistics from both all periods and Part 2 confirm a similar story that Shop A and Shop D changed their price more often than Shop B and Shop C.

The average frequency of online price changes by stores ranges from 1 to 3.5 percent (Table 19). The average duration of a price spell, based on all data by store-product category in Table 13 - 15, ranges from just half a month to two and a half months. This is much smaller than 4 to 7 months according to the monthly Thai CPI data (Apaitan, Disyatat & Manopimoke, 2018). Compared to the CPI data of the US and the euro area (Bils & Klenow, 2004; Aucremanne & Dhyne, 2004), the online price adjustment occurs much more frequently. However, our results are quite comparable to another internet price study like Lünnemann & Wintr (2006). They show that the average frequencies of internet price change in France, Germany and the US are approximately 3.1, 2.7 and 2.5 percent, respectively.

Statistics	Shop A		Shop B		Shop C		Shop D	
	All	Part 2	All	Part 2	All	Part 2	All	Part 2
Min	0	0	0	0	0	0	0	0
Pr25	0	0	0	0	0	0	0	0
Median	0.0111	0.0000	0.0056	0.0000	0.0000	0.0000	0.0117	0.0108
Mean	0.0333	0.0349	0.0174	0.0153	0.0092	0.0105	0.0220	0.0218
Pr75	0.0524	0.0577	0.0226	0.0202	0.0114	0.0131	0.0302	0.0326
Max	1	1	1	1	1	0.6667	0.6667	1
S.D.	0.0480	0.0549	0.0427	0.0432	0.0205	0.0216	0.0319	0.0320
No. Products	24,000	20,879	10,246	8,282	49,369	38,874	23,392	19,654

Table 19: Statistics for frequency of price changes by stores

<u>*Fact 3*</u>*: Online price increases and decreases quite often. Among those products experiencing price reduction(s), the number of times that prices reduce tends to be more often than the number of times that prices rise.*

Figure 1 – 8 show the patterns of price change by the number of times that price rose or fell based on data from all periods and Part 2. Clearly, the same product could have both episodes of price increases and decreases, so the red line, which is the 45 degrees line, marks equal number of times between episodes of price increase(s) and price reduction(s). As more blue dots appear on the lower side of the 45 degrees line in all four stores, it infers that there are more products with more frequent episodes of price decreases than price increases.

To focus on only products appearing often enough online, i.e. appearing online more than two months (60 days) out of all available days of data, Table 20 – 27 present the distribution of product by number of times/days (in ranges of five days) that its price increased or price decreased. The results confirm that all four stores seem to have more products with many more episodes of price falls than the episodes of price rises. For example, in Table 21, among the products from Shop A with 11-15 days of price reduction, we observed 29.5% of these products where price increased for only 6-10 days (fewer days), whereas only 6.3% of these products had 16-20 days (more days) of price increase.

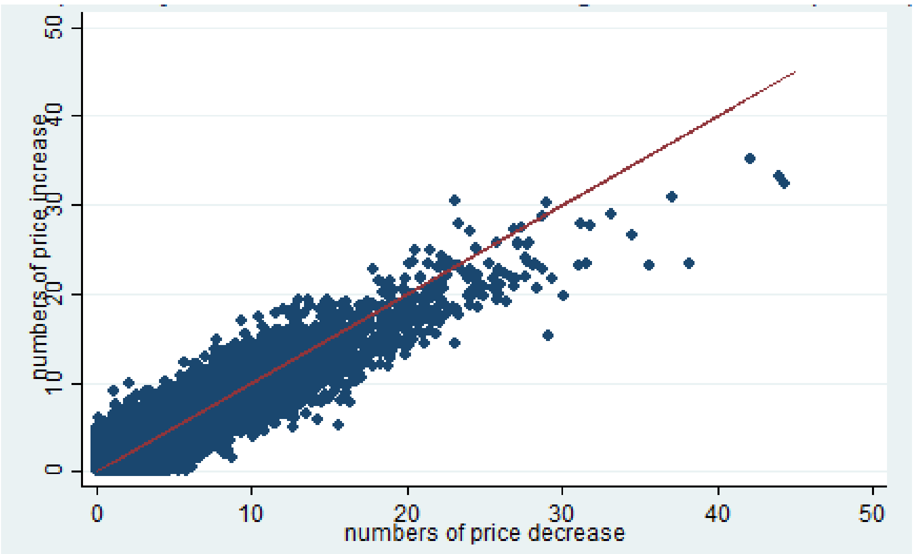

Figure 1: Pattern of price change for Shop A (All 239 days of data collection)

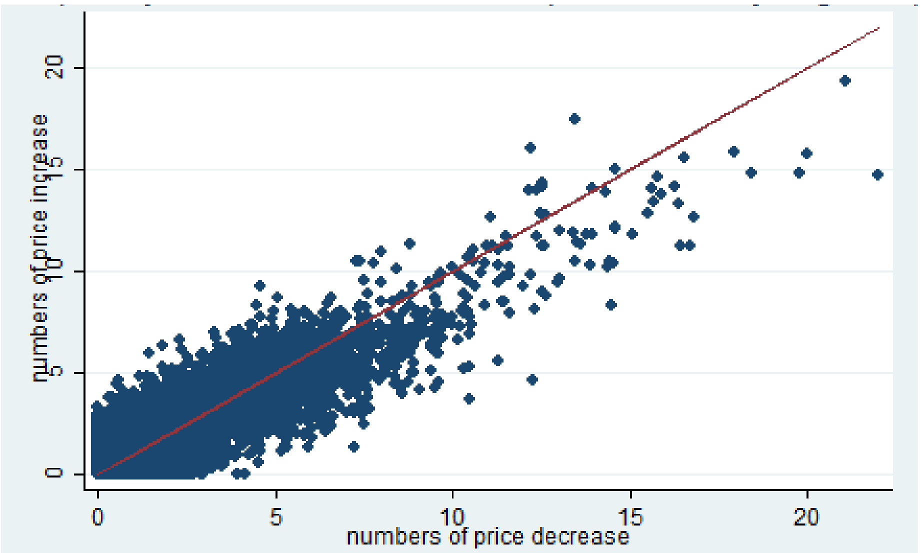

Figure 2: Pattern of price change for Shop A (Part 2, 98 days of data collection)

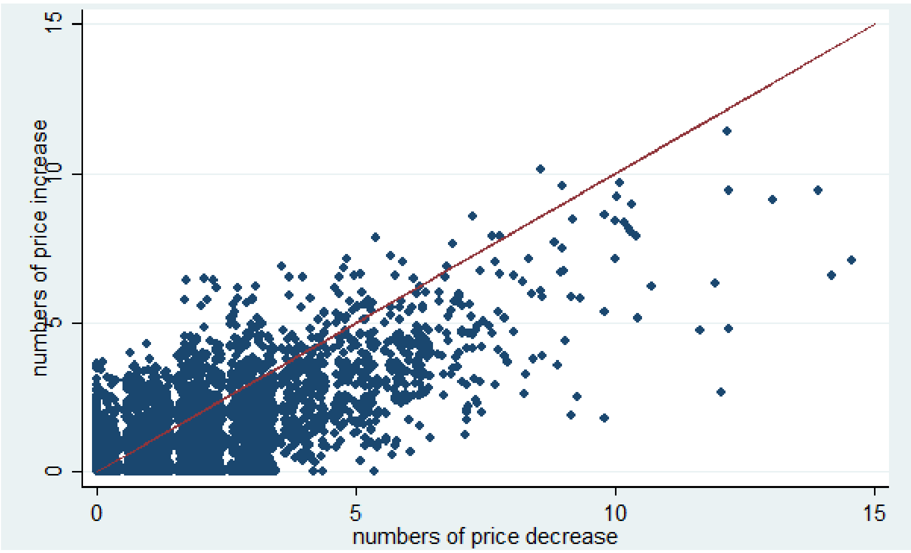

Figure 3: Pattern of price change for Shop B (All 204 days of data collection)

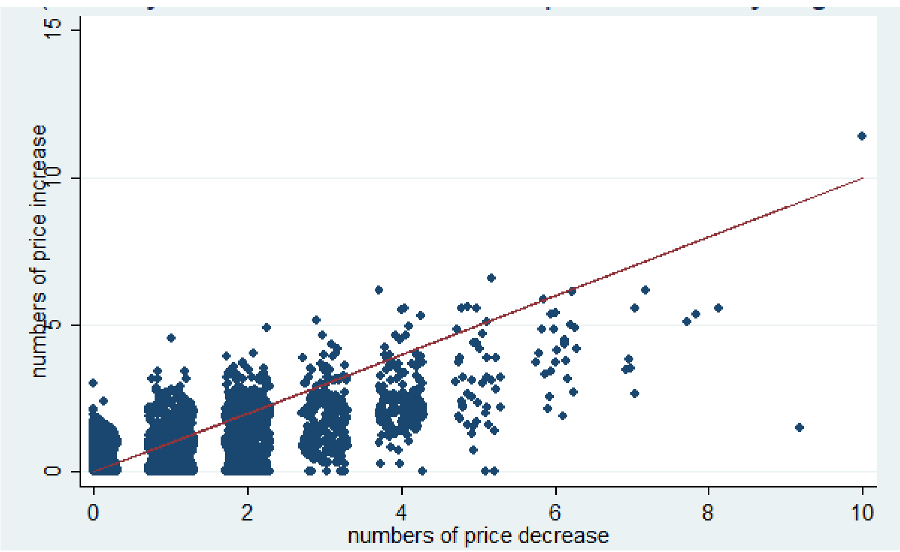

Figure 4: Pattern of price change for Shop B (Part 2, 100 days of data collection)

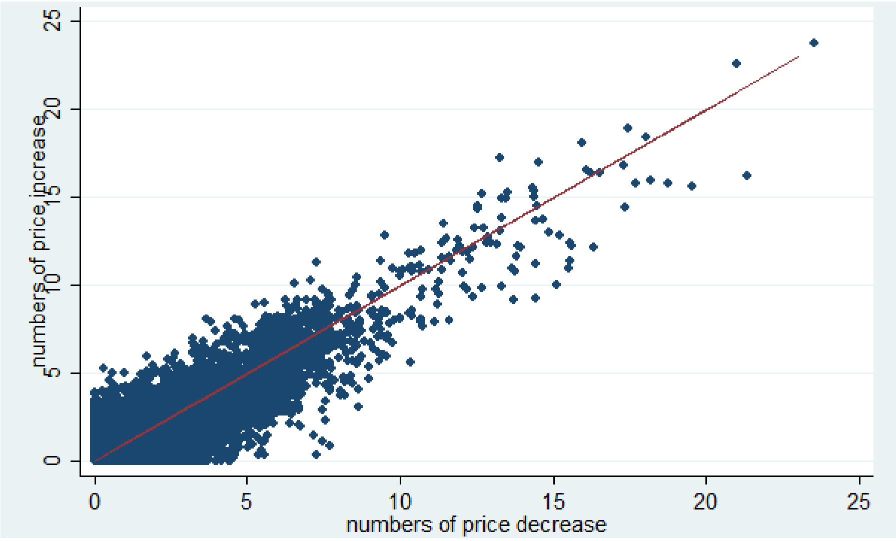

Figure 5: Pattern of price change for Shop C (All 276 days of data collection)

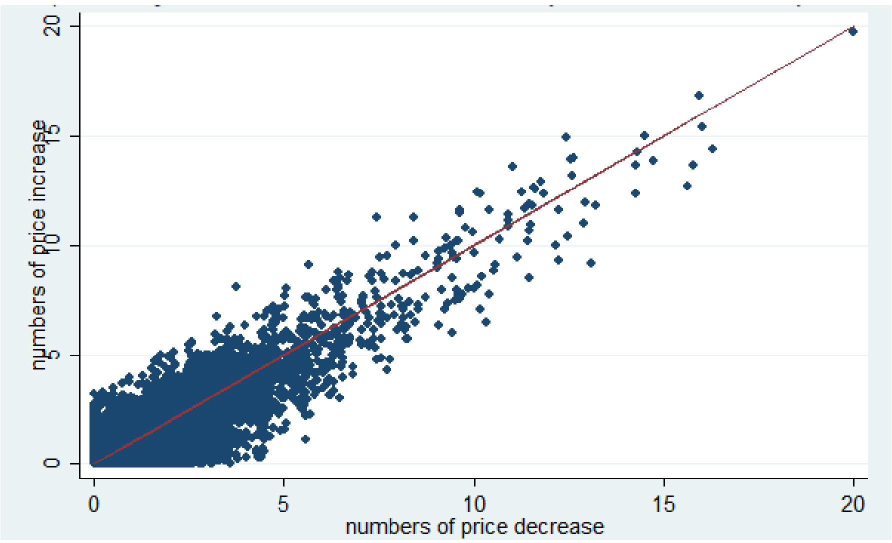

Figure 6: Pattern of price change for Shop C (Part 2, 156 days of data collection)

Figure 7: Pattern of price change for Shop D (All 302 days of data collection)

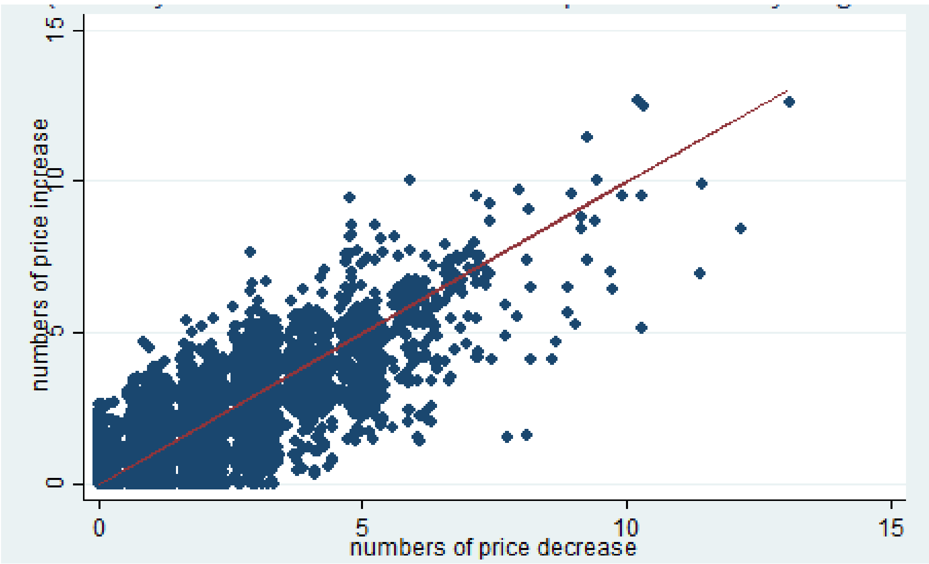

Figure 8: Pattern of price change for Shop D (Part 2, 100 days of data collection)

Price Increase (days)	Price Decrease (days)								
	0	1-5	6-10	11-15	16-20	21-25	26-30	31+	Total
0	4,663	499	0	0	0	0	0	0	5,162
1-5	654	5,913	467	0	0	0	0	0	7,034
6-10	0	392	1,944	233	3	0	0	0	2,572
11-15	0	0	171	508	125	0	0	0	804
16-20	0	0	0	50	115	49	4	0	218
21-25	0	0	0	0	9	42	16	4	71
26-30	0	0	0	0	0	4	6	6	16
31+	0	0	0	0	0	0	1	3	4
Total	5,317	6,804	2,582	791	252	95	27	13	15,881

Table 20: Number of products in Shop A appearing online at least 2 months with number of days / times[10] its price decrease and increase: All

[10] We sometimes use the word “days” and “times” interchangeably but it counts the days that price did change either rise or fall, not the length of each spell of reduction or rise in price.

Price Increase (days)	Price Decrease (days)								
	0	1-5	6-10	11-15	16-20	21-25	26-30	31+	Total
0	87.7%	7.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	32.5%
1-5	12.3%	86.9%	18.1%	0.0%	0.0%	0.0%	0.0%	0.0%	44.3%
6-10	0.0%	5.8%	75.3%	29.5%	1.2%	0.0%	0.0%	0.0%	16.2%
11-15	0.0%	0.0%	6.6%	64.2%	49.6%	0.0%	0.0%	0.0%	5.1%
16-20	0.0%	0.0%	0.0%	6.3%	45.6%	51.6%	14.8%	0.0%	1.4%
21-25	0.0%	0.0%	0.0%	0.0%	3.6%	44.2%	59.3%	30.8%	0.4%
26-30	0.0%	0.0%	0.0%	0.0%	0.0%	4.2%	22.2%	46.2%	0.1%
31+	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	3.7%	23.1%	0.0%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 21: Proportion of products (by column) in Shop A appearing online at least 2 months with number of days / times its price decrease and increase: All

Price Increase (days)	Price Decrease (days)				
	0	1-5	6-10	11-15	Total
0	2,707	1,292	0	0	3,999
1-5	489	2,693	188	3	3,373
6-10	0	28	47	7	82
11-15	0	0	0	1	1
Total	3,196	4,013	235	11	7,455

Table 22: Number of products in Shop B appearing online at least 2 months with number of days / times its price decrease and increase: All

Price Increase (days)	Price Decrease (days)				
	0	1-5	6-10	11-15	Total
0	84.70%	32.20%	0.00%	0.00%	53.60%
1-5	15.30%	67.10%	80.00%	27.30%	45.20%
6-10	0.00%	0.70%	20.00%	63.60%	1.10%
11-15	0.00%	0.00%	0.00%	9.10%	0.00%
Total	100.00%	100.00%	100.00%	100.00%	100.00%

Table 23: Proportion of products (by column) in Shop B appearing online at least 2 months with number of days / times its price decrease and increase: All

Price Increase (days)	Price Decrease (days)						
	0	1-5	6-10	11-15	16-20	21-25	Total
0	20,204	2,819	0	0	0	0	23,023
1-5	1,737	11,033	276	0	0	0	13,046
6-10	0	107	707	16	0	0	830
11-15	0	0	7	50	4	0	61
16-20	0	0	0	2	10	1	13
21-25	0	0	0	0	0	2	2
Total	21,941	13,959	990	68	14	3	36,975

Table 24: Number of products in Shop C appearing online at least 2 months with number of days / times its price decrease and increase: All

Price Increase (days)	Price Decrease (days)						
	0	1-5	6-10	11-15	16-20	21-25	Total
0	92.1%	20.2%	0.0%	0.0%	0.0%	0.0%	62.3%
1-5	7.9%	79.0%	27.9%	0.0%	0.0%	0.0%	35.3%
6-10	0.0%	0.8%	71.4%	23.5%	0.0%	0.0%	2.2%
11-15	0.0%	0.0%	0.7%	73.5%	28.6%	0.0%	0.2%
16-20	0.0%	0.0%	0.0%	2.9%	71.4%	33.3%	0.0%
21-25	0.0%	0.0%	0.0%	0.0%	0.0%	66.7%	0.0%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 25: Proportion of products (by column) in Shop C appearing online at least 2 months with number of days / times its price decrease and increase: All

Price Increase (days)	Price Decrease (days)								
	0	1-5	6-10	11-15	16-20	21-25	26-30	31+	Total
0	5,440	2,066	0	0	0	0	0	0	7,506
1-5	1,123	8,971	464	4	1	0	0	0	10,563
6-10	0	423	1,457	108	7	0	0	0	1,995
11-15	0	1	146	243	33	1	0	0	424
16-20	0	0	0	20	50	11	0	0	81
21-25	0	0	0	1	1	10	1	0	13
26-30	0	0	0	0	0	3	4	1	8
31+	0	0	0	0	0	0	0	1	1
Total	6,563	11,461	2,067	376	92	25	5	2	20,591

Table 26: Number of products in Shop D appearing online at least 2 months with number of days / times its price decrease and increase: All

Price Increase (days)	Price Decrease (days)								
	0	1-5	6-10	11-15	16-20	21-25	26-30	31+	Total
0	82.9%	18.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	36.5%
1-5	17.1%	78.3%	22.4%	1.1%	1.1%	0.0%	0.0%	0.0%	51.3%
6-10	0.0%	3.7%	70.5%	28.7%	7.6%	0.0%	0.0%	0.0%	9.7%
11-15	0.0%	0.0%	7.1%	64.6%	35.9%	4.0%	0.0%	0.0%	2.1%
16-20	0.0%	0.0%	0.0%	5.3%	54.3%	44.0%	0.0%	0.0%	0.4%
21-25	0.0%	0.0%	0.0%	0.3%	1.1%	40.0%	20.0%	0.0%	0.1%
26-30	0.0%	0.0%	0.0%	0.0%	0.0%	12.0%	80.0%	50.0%	0.0%
31+	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	50.0%	0.0%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 27: Proportion of products (by column) in Shop D appearing online at least 2 months with number of days / times its price decrease and increase: All

Fact 4: The magnitude of price changes, either increases or decreases, are sizeable compared to the inflation rate of comparable product categories. Heterogeneity in the size of changes by product categories seems to be more important than heterogeneity across stores. Yet the magnitude (in percentage) of price reductions are on average smaller than price increases.

	Statistics	Shop A	Shop B	Shop C	Shop D
Price Increase	Average size	23.04%	20.32%	22.33%	18.12%
	Median size	17.86%	11.17%	11.22%	14.70%
Price Decrease	Average size	16.13%	14.52%	13.89%	14.55%
	Median size	14.72%	11.12%	10.52%	13.38%

Table 28: Size of average and median price changes for products appearing online at least 2 months in each store: Part 2

According to Table 28, despite small differences in the average and median of the absolute size of price changes across retailers, there are a few similarities. First, all eight distributions of prices movement (increasing & decreasing for four stores) are right-skewed with higher mean than its median, i.e. long tails with a few cases of large price adjustment. Despite more days (or times) when the product prices moved downward (*fact 3*), the size of price increases seem to dominate the size of reduction in general. As shown in Table 28, both average and median size of price increases are higher than those of price decreases in all four stores. Also, the spreads between average and median price increases are larger than those spreads for price decreases (see figures B1 – B4 for the scatter plots between average size of price increases and price decreases in each store). One key difference between stores is that most of price increases in Shop B and Shop C are slightly more than 11% and those cases with sizeable increases are quite large. Thus, their average size of price increases can overtake Shop D and become quite closed to Shop A.

Category	Price Increase		Price Decrease	
	Average size	Median size	Average size	Median size
Beverages (11%)	12.63%	9.26%	10.66%	8.38%
Food (20%)	14.28%	11.34%	11.50%	10.00%
Health &Beauty (19%)	27.25%	25.14%	20.31%	20.06%
Home appliances (6%) & Electronic products	39.15%	15.25%	13.91%	11.90%
Household (15%)	28.51%	20.41%	19.34%	16.81%
Mom, baby & kids (4%)	17.14%	13.97%	13.86%	11.79%
Pet, Outdoor, etc. (9%)	30.74%	20.41%	19.90%	16.95%
Snacks & Sweets (9%)	13.47%	10.74%	11.02%	10.00%
Stationaries (2%)	21.83%	19.28%	16.91%	15.87%
Total (100%)	23.04%	17.86%	16.13%	14.72%

Table 29: Size of average and median price changes for products appearing online at least 2 months in Shop A by product categories: Part 2

Category	Price Increase		Price Decrease	
	Average size	Median size	Average size	Median size
Beauty (20%)	20.53%	12.83%	16.44%	13.19%
Electronic & Entertainment (2%)	12.69%	8.71%	11.64%	9.58%
Fashion (20%)	14.08%	10.00%	12.46%	10.00%
Health (4%)	39.39%	10.42%	12.07%	9.16%
Home appliances (7%)	16.16%	9.17%	14.67%	9.34%
Home & Garden (15%)	17.93%	13.73%	13.91%	12.52%
IT, Camera & Gadget (8%)	83.07%	9.26%	13.25%	7.00%
Mobile, Tablet & Accessories (7%)	16.69%	11.62%	14.09%	12.06%
Mom, baby & kids (8%)	25.04%	12.79%	13.04%	10.55%
Supermarket (4%)	19.96%	9.03%	10.74%	8.21%
Total (100%)	22.33%	11.22%	13.89%	10.52%

Table 30: Size of average and median price changes for products appearing online at least 2 months in Shop C by product categories: Part 2

Category	Price Increase		Price Decrease	
	Average size	Median size	Average size	Median size
Baby & Kids (8%)	16.06%	17.97%	13.29%	15.19%
Beverages,Snacks & Desserts (23%)	14.61%	11.24%	11.91%	10.11%
Dry Grocery (12%)	13.36%	11.11%	11.64%	10.26%
Fresh Food, Dairy & Bakery (9%)	19.21%	13.33%	15.68%	13.64%
Health & Beauty (26%)	24.82%	22.52%	18.48%	17.91%
Home & Electrical (2%)	30.06%	28.44%	21.09%	16.39%
Household Products (8%)	18.47%	16.50%	15.55%	15.15%
Miscellaneous (6%)	26.46%	22.12%	21.43%	21.15%
Pets (4%)	9.58%	8.01%	8.50%	6.90%
Total (100%)	18.12%	14.70%	14.55%	13.38%

Table 31: Size of average and median price changes for products appearing online at least 2 months in Shop D by product categories: Part 2

There are stark differences in the size of increases and decreases in prices across product categories within the same store. These differences across product categories clearly outweigh any heterogeneity across retailers as shown in Table 29 – 31. Yet both average and median size of price increases dominate those of price falls in all product categories. Almost all product categories in all stores illustrate the right-skewed distribution of price movement with larger means than medians except for baby & kids in Shop D. However, average size of price movement has to be interpreted with caution because these averages can be influenced by extreme outliers as shown in Figure B1 – B4 and probably in the case of IT, camera & gadgets for Shop C. Such outliers could be the website's typos, some mistakes embedded in our scraping program or some mistakes during the data management and analysis.

	Statistics	Shop A	Shop B	Shop C Home Appliance	Shop C Electronic & Entertainment	Shop D
Price Increase	Average size	39.15%	20.32%	16.16%	12.69%	30.06%
	Median size	15.25%	11.17%	9.17%	8.71%	28.44%
Price Decrease	Average size	13.91%	14.52%	14.67%	11.64%	21.09%
	Median size	11.90%	11.12%	9.34%	9.58%	16.39%

Table 32: Size of average and median price changes for home appliances and electrical products appearing online at least 2 months in each store: Part 2

	Statistics	Shop A	Shop C Health	Shop C Beauty	Shop D
Price Increase	Average size	27.25%	39.39%	20.53%	24.82%
	Median size	25.14%	10.42%	12.83%	22.52%
Price Decrease	Average size	20.31%	12.07%	16.44%	18.48%
	Median size	20.06%	9.16%	13.19%	17.91%

Table 33: Size of average and median price changes for health and beauty products appearing online at least 2 months in each store: Part 2

	Statistics	Shop A	Shop C	Shop D
Price Increase	Average size	17.14%	25.04%	16.06%
	Median size	13.97%	12.79%	17.97%
Price Decrease	Average size	13.86%	13.04%	13.29%
	Median size	11.79%	10.55%	15.19%

Table 34: Size of average and median price changes for mom and kid products appearing online at least 2 months in each store: Part 2

	Statistics	Shop A Beverages (11%)	Shop A Snacks & Sweets (9%)	Shop D Beverages,Snacks & Desserts (23%)
Price Increase	Average size	12.63%	13.47%	14.61%
	Median size	9.26%	10.74%	11.24%
Price Decrease	Average size	10.66%	11.02%	11.91%
	Median size	8.38%	10.00%	10.11%

Table 35: Size of average and median price changes for beverage, snacks and sweets products appearing online at least 2 months in each store: Part 2

We further explore the heterogeneity of the magnitude of price changes across retailers among those product categories which are roughly comparable. Starting from home appliances and electrical products, there are sizeable differences between stores, particularly the average and median of price increases. Then the differences across stores seem to narrow down among health and beauty as well as mom and kid products. Lastly, the differences between Shop A and Shop D in a category of beverage, snacks and sweets are quite minimal. Although it is not clear from the data, it seems that retailers' online pricing strategies (in terms of size of price changes) are interrelated to the nature of products as define by these broad classifications.

<u>Fact 5</u>: Frequency of price promotion online is different across retailers. Out of four retailers, one store has more than 65% of its products on promotion every day. On the contrary, other two stores have only around 2% of their products on promotion every day.

Based on our scraped data, we define "Promotional date" as the days when the webpage claims that the product is on Promotion or displays original and/or special price of that product. In particular, Shop A, Shop B and Shop C showed the "original price" of their products alongside the actual selling price. On the other hand, Shop D also displayed the date that the "Promotion" would expire, i.e. Promotion till xxx as well as the "original price". Some retailers even calculated and showed how much the customers could save from this promotion based on the "original price". For Shop D, where customers had information on both the sale promotion and the length of the promotion, it allowed customers to know not only the percentage of potential discount they "receive" from the stores but also the time when they have to/ought to buy the products. Conceptually, this strategy needs time limit, so Shop D might not have as high percentage of discounted products comparing to our stores.

Cross tabulations between the proportion of days (times) that product appeared online with "promotion"[11] and numbers of days those products were sold on the platform with the price data are shown in Table 36-43. Shop C was a store where 60-75% of its products were on promotion every day regardless of how long the product appeared on its platform. Only slightly less than 20% of its products did not have promotion at all. Shop B seems to have a mixture of new products on promotion immediately and those old products promoted after some time. As for the stores selling wide variety of products, like Shop A and Shop D, they have roughly 30-40% of their products which were on sale 11-75% of the time (days), whereas very few of their products were on promotion every day, i.e. 1.79% and 2.05% for Shop A and Shop D respectively.

In sum, there seems to be no relationship between how long the product stayed on the platform and how likely it is to be on promotion. In other words, these four retailers do not seem to treat either new or regular products differently in terms of promotion. Possibly, the promotion used could be the store's reaction to customers' demand and seasonality or actions against its competitors. However, such strategic reactions and other types of promotion are beyond the scope of this paper.

[11] These are the products that the shop did deliberately claim that they gave discounts. In this paper, we focus on promotion based on price only (e.g. promotion based on quantity, gifts or product samples are excluded).

Appearing Online (Days)	Proportion of times the product appearing Online with Special Prices/Promotions								Total
	0%	1-10%	11-25%	26-50%	50-75%	76-90%	91-99%	Every Day	
1-30	3,577	1	211	79	43	115	154	196	4,376
31-60	1,398	23	208	164	169	207	119	167	2,455
61-90	290	47	70	116	53	22	12	2	612
91-120	189	38	95	113	81	26	14	1	557
121-150	404	36	58	120	85	37	31	10	781
151-180	926	37	134	286	161	61	67	3	1,675
181-210	1,511	290	308	414	134	56	30	0	2,743
211-240	2,762	489	1,205	1,795	1,011	369	267	0	7,898
NA	48	3	20	14	3	9	10	0	107
Total	11,105	964	2,309	3,101	1,740	902	704	379	21,204

Table 36: Proportion of days (times) that product appears online with promotion by numbers of days appearing with the price data: Shop A Part 2 (count)

Appearing Online (Days)	Proportion of times the product appearing Online with Special Prices/Promotions								Total
	0%	1-10%	11-25%	26-50%	50-75%	76-90%	91-99%	Every Day	
1-30	81.74%	0.02%	4.82%	1.81%	0.98%	2.63%	3.52%	4.48%	100%
31-60	56.95%	0.94%	8.47%	6.68%	6.88%	8.43%	4.85%	6.80%	100%
61-90	47.39%	7.68%	11.44%	18.95%	8.66%	3.59%	1.96%	0.33%	100%
91-120	33.93%	6.82%	17.06%	20.29%	14.54%	4.67%	2.51%	0.18%	100%
121-150	51.73%	4.61%	7.43%	15.36%	10.88%	4.74%	3.97%	1.28%	100%
151-180	55.28%	2.21%	8.00%	17.07%	9.61%	3.64%	4.00%	0.18%	100%
181-210	55.09%	10.57%	11.23%	15.09%	4.89%	2.04%	1.09%	0.00%	100%
211-240	34.97%	6.19%	15.26%	22.73%	12.80%	4.67%	3.38%	0.00%	100%
NA	44.86%	2.80%	18.69%	13.08%	2.80%	8.41%	9.35%	0.00%	100%
Total	52.37%	4.55%	10.89%	14.62%	8.21%	4.25%	3.32%	1.79%	100%

Table 37: Proportion of days (times) that product appears online with promotion by numbers of days appearing with the price data: Shop A Part 2 (percent)

Appearing Online (Days)	Proportion of times the product appearing Online with Special Prices/Promotions								Total
	0%	1-10%	11-25%	26-50%	50-75%	76-90%	91-99%	Every Day	
1-30	1,072	3	8	38	21	10	6	360	1,518
31-60	247	8	30	38	61	35	17	114	550
61-90	391	19	51	66	19	27	40	160	773
91-120	424	37	62	78	50	34	51	145	881
121-150	664	52	105	120	62	68	56	153	1,280
151-180	580	54	130	146	56	48	68	165	1,247
181-204	1,699	82	262	338	175	78	113	300	3,047
Total	5,077	255	648	824	444	300	351	1,397	9,296

Table 38: Proportion of days (times) that product appears online with promotion by numbers of days appearing with the price data: Shop B Part 2 (count)

Appearing Online (Days)	Proportion of times the product appearing Online with Special Prices/Promotions								Total
	0%	1-10%	11-25%	26-50%	50-75%	76-90%	91-99%	Every Day	
1-30	70.62%	0.20%	0.53%	2.50%	1.38%	0.66%	0.40%	23.72%	100%
31-60	44.91%	1.45%	5.45%	6.91%	11.09%	6.36%	3.09%	20.73%	100%
61-90	50.58%	2.46%	6.60%	8.54%	2.46%	3.49%	5.17%	20.70%	100%
91-120	48.13%	4.20%	7.04%	8.85%	5.68%	3.86%	5.79%	16.46%	100%
121-150	51.88%	4.06%	8.20%	9.38%	4.84%	5.31%	4.38%	11.95%	100%
151-180	46.51%	4.33%	10.43%	11.71%	4.49%	3.85%	5.45%	13.23%	100%
181-204	55.76%	2.69%	8.60%	11.09%	5.74%	2.56%	3.71%	9.85%	100%
Total	54.61%	2.74%	6.97%	8.86%	4.78%	3.23%	3.78%	15.03%	100%

Table 39: Proportion of days (times) that product appears online with promotion by numbers of days appearing with the price data: Shop B Part 2 (percent)

Appearing	Proportion of times the product appearing Online with Special Prices/Promotions								
Online (Days)	0%	1-10%	11-25%	26-50%	50-75%	76-90%	91-99%	Every Day	Total
1-30	522	28	11	36	21	0	7	2,072	2,697
31-60	477	11	20	176	36	138	31	1,848	2,737
61-90	560	59	104	148	70	21	25	1,833	2,820
91-120	384	18	56	118	33	25	49	1,665	2,348
121-150	507	26	89	116	73	18	51	1,973	2,853
151-180	1046	142	155	277	61	40	70	2,800	4,591
181-210	899	63	235	127	94	59	25	2,750	4,252
211-240	716	118	115	241	61	131	51	2,882	4,315
241-270	727	95	64	78	56	106	39	3,210	4,375
Almost Every Day	1,442	229	271	152	123	89	70	5,740	8,116
Total	7,280	789	1,120	1,469	628	627	418	26,773	39,104

Table 40: Proportion of days (times) that product appears online with promotion by numbers of days appearing with the price data: Shop C Part 2 (count)

Appearing	Proportion of times the product appearing Online with Special Prices/Promotions								
Online (Days)	0%	1-10%	11-25%	26-50%	50-75%	76-90%	91-99%	Every Day	Total
1-30	19.35%	1.04%	0.41%	1.33%	0.78%	0.00%	0.26%	76.83%	100%
31-60	17.43%	0.40%	0.73%	6.43%	1.32%	5.04%	1.13%	67.52%	100%
61-90	19.86%	2.09%	3.69%	5.25%	2.48%	0.74%	0.89%	65.00%	100%
91-120	16.35%	0.77%	2.39%	5.03%	1.41%	1.06%	2.09%	70.91%	100%
121-150	17.77%	0.91%	3.12%	4.07%	2.56%	0.63%	1.79%	69.16%	100%
151-180	22.78%	3.09%	3.38%	6.03%	1.33%	0.87%	1.52%	60.99%	100%
181-210	21.14%	1.48%	5.53%	2.99%	2.21%	1.39%	0.59%	64.68%	100%
211-240	16.59%	2.73%	2.67%	5.59%	1.41%	3.04%	1.18%	66.79%	100%
241-270	16.62%	2.17%	1.46%	1.78%	1.28%	2.42%	0.89%	73.37%	100%
Almost Every Day	17.77%	2.82%	3.34%	1.87%	1.52%	1.10%	0.86%	70.72%	100%
Total	18.62%	2.02%	2.86%	3.76%	1.61%	1.60%	1.07%	68.47%	100%

Table 41: Proportion of days (times) that product appears online with promotion by numbers of days appearing with the price data: Shop C Part 2 (percent)

Appearing	Proportion of times the product appearing Online with Special Prices/Promotions								
Online (Days)	0%	1-10%	11-25%	26-50%	50-75%	76-90%	91-99%	Every Day	Total
1-40	370	2	5	13	11	21	1	29	452
41-80	361	16	59	73	27	23	8	19	586
81-120	422	43	206	150	62	22	25	18	948
121-160	401	57	79	76	35	19	12	36	715
161-200	659	167	173	151	87	31	27	29	1,324
201-240	2,555	410	952	967	405	179	149	122	5,739
241-280	3,617	540	1,233	1,629	714	291	162	150	8,336
280-294	773	140	282	276	90	19	9	0	1,589
Total	9,158	1,375	2,989	3,335	1,431	605	393	403	19,689

Table 42: Proportion of days (times) that product appears online with promotion by numbers of days appearing with the price data: Shop D Part 2 (count)

Appearing Online (Days)	Proportion of times the product appearing Online with Special Prices/Promotions								Total
	0%	1-10%	11-25%	26-50%	50-75%	76-90%	91-99%	Every Day	
1-40	81.86%	0.44%	1.11%	2.88%	2.43%	4.65%	0.22%	6.42%	100%
41-80	61.60%	2.73%	10.07%	12.46%	4.61%	3.92%	1.37%	3.24%	100%
81-120	44.51%	4.54%	21.73%	15.82%	6.54%	2.32%	2.64%	1.90%	100%
121-160	56.08%	7.97%	11.05%	10.63%	4.90%	2.66%	1.68%	5.03%	100%
161-200	49.77%	12.61%	13.07%	11.40%	6.57%	2.34%	2.04%	2.19%	100%
201-240	44.52%	7.14%	16.59%	16.85%	7.06%	3.12%	2.60%	2.13%	100%
241-280	43.39%	6.48%	14.79%	19.54%	8.57%	3.49%	1.94%	1.80%	100%
280-294	48.65%	8.81%	17.75%	17.37%	5.66%	1.20%	0.57%	0.00%	100%
Total	46.51%	6.98%	15.18%	16.94%	7.27%	3.07%	2.00%	2.05%	100%

Table 43: Proportion of days (times) that product appears online with promotion by numbers of days appearing with the price data: Shop D Part 2 (percent)

<u>Online versus Offline prices</u>

<u>*Fact 6*</u>*: Depending on the type of retailers, products online match well with their offline in hypermarket stores but not with those in convenient stores. Online products can be different, in terms of sizes and models/versions, from those sold in the actual stores. Slightly more than half of the product-day matched had exactly the same price between online platform and offline stores. Although differences in the percentage of matched between the two hypermarket chains are small, percentages of product-days matched vary markedly by product categories.*

In order to verify the consistency between the online and on-the-shelf prices, our research assistants surveyed prices of some random products (based on our scraped data) from the "actual" stores (namely Shop A and Shop D) four times in Bangkok between late March and mid-July 2017 and once in the following provinces: one province in the North, two from the Central, two from the Northeast and two provinces from the South. Our team collected information such as product name, barcode, price and promotion from each store as shown in Figure 9 and 10 below:

Figure 9: An example of a product that exists in both stores with its price

Figure 10: Examples of products with their promotional prices on the price tags

Regarding Shop C, it has almost mutually exclusive products between their online and on-shelf outlets, possibly due to the relatively small size of most of its convenient stores. Particularly, its website focuses on product categories such as beauty, fashion or electronic appliances, while those categories that exist in both platforms, for example beverages or personal care, tend to have different package size, i.e. much larger size on its online outlet. In contrast, Shop B imposes a policy of no photo taking in its stores. Thus, we decide to exclude it from our offline study.

We randomly selected our offline survey sample from the scraped data and found most of our 20-30 randomly selected products in those two stores. There are around 20-30% of products with their barcodes online so that we can match these products between the two stores as we show one example in Figure 9. Our surveys also gathered information on existing sales and promotion as shown in Figure 10.

There are some differences between online and on-the-shelf products in these stores. First, although we could match the price of the same product many times, other times we could not. It seems that we could not match some price or quantity promotions, especially when the offers are buy one get one free or buy 2 units of A and receive 1 unit of B for free. Moreover, products available in stores can be of different types and sizes from the online outlets. For instance, the products sold online might contain older versions or older models of the same product, particularly in the gadgets and electronic category. This might result from limited "physical" space on the shelf compared to less limited space online. Sometimes, among products with the same barcode and/or Stock Keeping Unit (SKU)[12], the product information on the website is incorrect or outdated, for example, wrong package size or slightly different types and colours.

Out of all product-days that we can match with our scraped data, slightly more than half of them were sold at the same price off- and on-line. In Shop A, around 47% of the product-days in the sample was sold at different prices, while the mismatch in Shop D was around 35%. Table 44 and 45 present the number of match and mismatch between online and offline price by product categories. The percentages of mismatched prices are very high in both Shop A and

[12] Other than these three stores, we also surveyed the products in a multinational furniture store, IKEA, in Bangkok. Yet the price and product information in this store perfectly matches the information provided on its website.

D for health & beauty products. Only half of the product-days shows exact price matched for home appliance and electronic in both Shop A and D, whereas the percentage of mismatched reduces to only 20- 25% for household products. Other product categories such as food, dry grocery and snacks seem to have very few mismatches. Interestingly, beverages is another product category with a large proportion of price mismatches (and this might pull up the number of mismatches among Beverages, Snacks and Desserts of Shop D). Although there seems to be some pattern in the discrepancies between online VS. offline price along the line of product categories, we are not willing to draw any strong conclusion due to our small offline sample size. Further, the actual stores did have some more bunching, bundling or giveaway promotion. Thus, comparing prices alone is ambiguous to conclude whether the online or offline products are cheaper.

Category	matched	mismatched	total number product-days	% mismatch
Beverages	2	11	13	84.62%
Clothes&Shoes	2	0	2	0.00%
Food	13	2	15	13.33%
Health-Beauty	6	19	25	76.00%
Home appliance/Electronic	2	2	4	50.00%
Household	3	1	4	25.00%
Mom-Baby-Kid	4	0	4	0.00%
Pet-Outdoor & etc.	2	2	4	50.00%
Snacks-Sweets	10	3	13	23.08%
Stationary	2	0	2	0.00%

Table 44: Number of product-days with the price matched between the same online and offline products based on our own survey by category: Shop A

Category	matched	mismatched	total number product-days	% mismatch
Baby & Kids	3	3	6	50.00%
Beverages, Snacks & Desserts	10	7	17	41.18%
Dry Grocery	17	5	22	22.73%
Fresh Food, Dairy & Bakery	4	0	4	0.00%
Health & Beauty	8	17	25	68.00%
Home & Electrical	3	3	6	50.00%
Household Products	8	2	10	20.00%
Miscellaneous	7	0	7	0.00%
Pets	5	0	5	0.00%

Table 45: Number of product-days with the price matched between the same online and offline products based on our own survey by category: Shop D

As for regional differences, we found that in all three stores price differences in our selected sample across provinces are very minimal. These large chains regularly release their own hard copy catalogues, which are the same catalogues across stores and exactly match the actual price and promotion. Local stores seem to have limited capacity to set their own price except for some product categories such as electronic appliances, where they can set their own store-level promotion.

In terms of the offline price changes, according to our randomly selected products, it seems that each store changes its products' prices only when it releases the new catalogue. During the period of a given catalogue, prices seem to be fixed across (actual) stores until the new catalogue is released. We find this observation on the ground very interesting and in the future work we might be able to carefully verify if the prices in the online outlet are changed around the day of catalogue released dates or not.

5 Potential for future works

As this project is one of the very first studies exploring the internet price movement in Thailand, we attempts to document some stylized facts drawn from descriptive statistics of the data. Yet web scraping has a real potential that can open more possibilities of researches on price movement and price setting in Thailand. Questions, such as are there any differences in price behaviours or sale strategies between online platforms with and without physical stores? or how strong is a pass-through pricing across ASEAN countries?, can be answered by the data gathered from online platform.

As for future works based on this dataset, we believe that further exploration into issues, for example, differences between "normal" price reduction and "promotional" price, price movement of the same product across stores matched by barcodes, or other types of sale promotion online, has good potentials of providing new insight into nature and characteristics of Thai retail market. So far, we use only summary statistics of the scraped online price; however, the time-series aspect of this dataset is still underexplored. In this section, we demonstrate one possibility of using the online price to supplement and potentially solve some problems stemmed in the price collection offline for a construction of CPI.

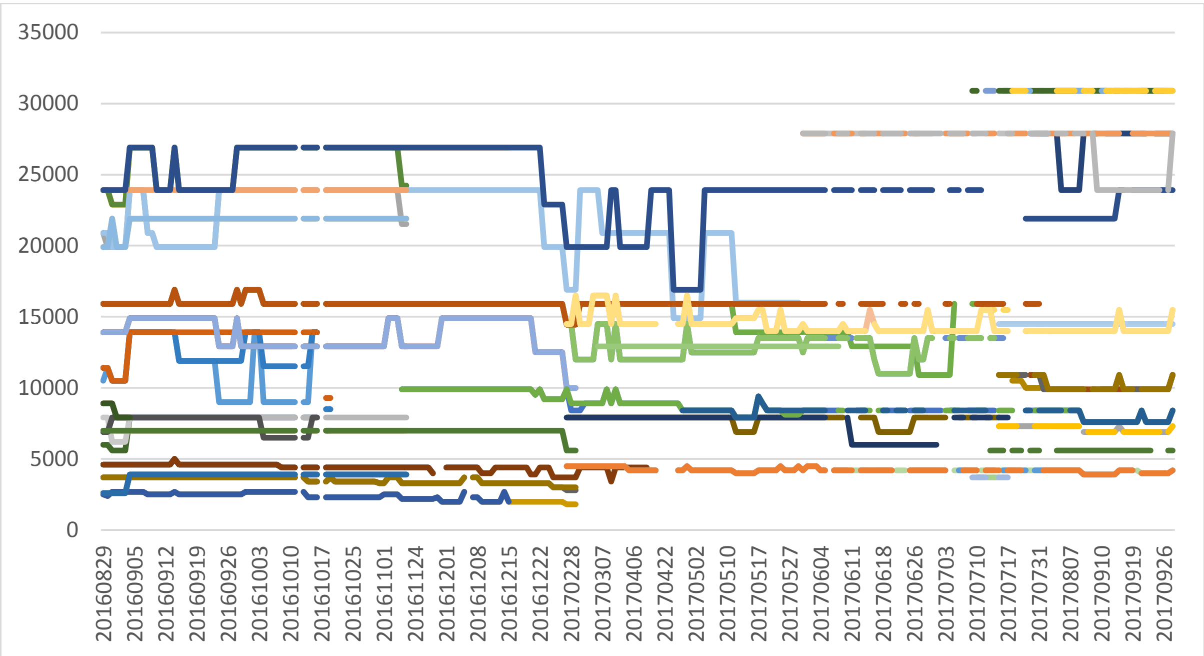

Figure 11: Price movement of all 70 Samsung smartphones sold on one retailer's online outlet

One of an issues facing the statistics offices responsible for producing indices such as CPI is how to find comparable products to replace discontinued products in the (e.g. CPI) basket. Hedonic Quality Adjustment is one of the techniques used to adjust for the differences in quality between old and new products (Wells & Restieaux, 2014). Yet Cavallo and Rigobon (2016) show that a simple index constructed from online prices of different models and brands of television in the US can approximate an official price index for TV that use complex hedonic quality-adjustment methods. This method is called overlapping qualities proposed by Armknecht and Weyback (1989)[13].

Figure 11 illustrates how our scraped data look like using Samsung smartphones as an example. We can document prices of both discontinued products and new releases in our data. Thus, it provides an opportunity to construct a price index of this broad product category and use it as a supplement to existing smartphone CPI data. Of course, we can extend this diagram to cover all brands of smartphones (or other products) sold in all online four stores. However, this is not straightforward because some of the stores' own defined categories and product names might not be distinct enough to be used to classify some products out of the rest (see for example, Challenges in data management and analysis in Appendix B) .

6 Conclusion

This paper presents six stylized facts of online (and offline) price in Thailand based on some descriptive statistics computed from web-scraped data. We show that price changes in both directions are not uncommon but there exist heterogeneities in such movement across stores and product categories. We also provide some evidence that each retailer seems to engage in different pricing strategies between its online outlet and physical stores.

Notwithstanding its great potential, our dataset is far from perfect. There are several issues that complicate our data collection and data management as well as jeopardize the data quality and consistency. For example, there are days of miss data owing to failure and changes made within our program or the retailers' websites. Some figures and statistics clearly indicate outlier problems within our data. These outliers and/or discontinued series might result from changes in products' URL and/or the spelling of product name over time. Despite our best effort in data cleansing, we urge readers to interpret our results with caution.

Nevertheless, we believe that this exercise is an important step toward understanding online price movement in Thailand. Given the potential to tap into this vast resource of product, price and sale promotion data, we would like to encourage other researchers to continue finding the best practice to work with such big data in the context of Thai E-commerce. It would be very interesting if machine learning techniques can be used to enable us to match our product to the official CPI categories efficiently. Or when data from more online stores can be scraped on a regular basis. Then we will be able to properly address more questions about price behaviours, retailers' strategies or the interaction between online outlets and physical stores in Thailand.

[13] To be precise, if two goods coexist for some time, their overlapping prices can be used to obtain an estimate of quality change (Cavallo & Rigobon, 2016).

References

Alvarez, L. J., Dhyne, E., Hoeberichts, M., Kwapil, C., Bihan, H., Lünnemann, P., ... & Vilmunen, J. (2006). STICKY PRICES IN THE EURO AREA: A SUMMARY OF NEW MICRO-EVIDENCE. *Journal of the European Economic association*, *4*(2-3), 575-584.

Apaitan, T., Disyatat, P., & Manopimoke, P. (2018). *Price Setting Behavior in Thailand: Evidence from Micro CPI Data* (No. 85). Puey Ungphakorn Institute for Economic Research.

Aucremanne, L., & Dhyne, E. (2004). *How frequently do prices change? Evidence based on the micro data underlying the Belgian CPI* (No. 44). Working Paper Research.

Cavallo, A. (2013). Online and official price indexes: Measuring Argentina's inflation. *Journal of Monetary Economics*, *60*(2), 152-165.

Cavallo, A., & Rigobon, R. (2016). The billion prices project: Using online prices for measurement and research. *Journal of Economic Perspectives*, *30*(2), 151-78.

Klenow, P. J., & Kryvtsov, O. (2008). State-dependent or time-dependent pricing: Does it matter for recent US inflation?. *The Quarterly Journal of Economics*, *123*(3), 863-904.

Lünnemann, P., & Wintr, L. (2006). *Are internet prices sticky?* (No. 645). ECB Working Paper.

Statista. 2018. *eCommerce in Thailand*. [ONLINE] Available at: https://www.statista.com/outlook/243/126/ecommerce/thailand. [Accessed 21 August 2018].

Wells, J., & Restieaux, A. (2014). *Review of Hedonic Quality Adjustment in UK Consumer Price Statistics and Internationally*. Office for National Statistics UK.

Appendix A

Data collection, management, and analysis: Issues and lessons learned

Challenges in data collection

In the early-stage of this project, good-level price and promotional data were scraped from multiple retailers' websites both in Thailand and in CLMV countries. Yet due to limited availability of retailed stores in those countries, which had proper online platforms to sell their products, the author decided to modify the project title and objective to focus mainly in Thailand. Although the automated process allowed us to obtain useful data, there are some challenges we faced in the process.

The first challenge is how the data is stored in some websites which prevents data from being scraped. For example, a website of a retailer in CLMV has information about prices and promotions of their products on their website but the price and promotion information is embedded in the image of the product. While customers who visit their website can visibly see the information, a web scraper cannot obtain the data. Therefore, such websites are unable to be included in this project.

Figure A1: A screenshot from a website showing products and their prices and promotional information

Figure A2**:** An example of a product image from a website. The price and promotional information is embedded in the image, which allows users to view it. Yet a web scraper obtains text data from websites; therefore, is unable to obtain the price data from this website.

The second challenge is the changes that the retailers made with their websites. The competition of retailers is fierce and the web technology advances quickly. Therefore, there are often changes made to the retailer's website. The changes normally disrupt web scraping process and the product data cannot be obtained. The following are examples of the changes that were made.

1. **Changes of website's URL.** Shop C changed its URL around September 2016.
2. **Changes of website's user interfaces.** Shop A's website was redesigned to create new images for its brand. As web scraping use the website's structure to locate relevant data, such changes made the web scraping program fail to obtain any data.
3. **Changes of website's categories.** Shop D removed alcohol from the list of categories of its website. As the web scraping program then relied on a specific list of category, parts of the process failed and the number of items retrieved went down from around 18,600 to 4,000 for several days before we spotted the issue.
4. **Changes of website's security.** On September 2, 2016, a website from another ASEAN country started using cloudflare, a technology for DDoS attack protection. DDoS attack is an attempt to flood the bandwidth or resources of a targeted system. Such technology also prevent automated web scraping. Therefore, the program that worked on the previous day was unable to access the website and unable to obtain any data.

For the second challenge, changes were made to the web scraping program during this project to address the issues. However, we can expect that retailers will continue to make changes in the future. Therefore, it is important to add monitoring process of the progress and failure of web scraping and ensure resources to promptly modify the web scraping program for continuity of data.

The last challenge is the growing size of data. The retailer that carried the most number of products in their website for this project carried approximately 26,000 products in 2017. This results in 3GB of data in csv format per year. To continue scraping data, an archival of data should be considered, because a large storage on a cloud server can be costly. The current rented cloud server has 20GB of disk space and the researchers manually move the data to local storage.

Challenges in data management and analysis

We use STATA 14 to cleanse, compile and analyse the data. In terms of data cleansing, although the programme can handle Thai characters, for example, in some of the product names, there are several issues leading to minor and major concerns on the data set quality.

First, a common data format such as comma-separated values (*.csv) is used to store scrapped data. However, some stores use commas in their websites to indicate, for instance, prices higher 1,000. So it requires extra programing to manage those problematic cases. Therefore, we suggest that the tab-delimited text file could be a better option to store the data in future research. Furthermore, STATA sometimes cannot handle string information properly, especially when extra quotation marks appear in the data cells unexpectedly such as product names with their size in inches (4"*6"). This issue leads to some common symptoms that we could spot and fix them systematically. Nevertheless, given the size of our data, we cannot be 100% certain that our cleansed data are free from other uncommon anomaly.

Regarding compilation of the data sets, key information of the products published on each website, which are product name, product price, product category, sales and promotions, and barcode (if available), is collected everyday (that our programme works). Therefore, the size of the data we have to handle for each store is massive (e.g. around 20,000 products daily for roughly 200-250 days). Due to limited physical memory in standard personal computers or laptops, the huge data set sometimes needs several hours to generate simple variables or even days in case of reshaping the data. To prevent failures during the compilation process, we run the program with a trial subset of the data so that irrelevant information or duplication can be dropped; hence, reduce the size of the data as early as possible. Yet for future works with many years of data, better hardware with greater memory and computational power is needed.

Another issue on compilation is how to uniquely match the same products across dates. Some stores assign unique category and product ID for each product. In these cases, we can keep only one unique product though they might be listed in more than one product category (for example, baby milk powder could be classified as both beverages and moms & kids). Other stores do not have any information equivalence to a unique ID. So either product name or its URL is used to match the product and create a time series for its prices and promotions. Of course, this method is susceptible to matching failures due to slight changes in the URL or product name over the study period but it seems to be the best strategy available.

As for data analysis, we first plan to construct a comparable index to the Consumer Price Index (CPI) for some product categories. However, product categories on the website of each store are assigned by its own criteria and cannot directly match to other stores or the official CPI. To classify each product in our data into the detailed product categories of the CPI requires thorough investigation into each product name with some levels of judgement. For instance, a product name contains a word "rice" could belong to the CPI category other than rice such as rice vermicelli, which is a type of noodles. Therefore, our analysis focuses on descriptive statistics based on store-level or major category-store level only. For future works, we see a potential of using machine learning techniques to deal with this classification issue.

Appendix B

Extra Figures

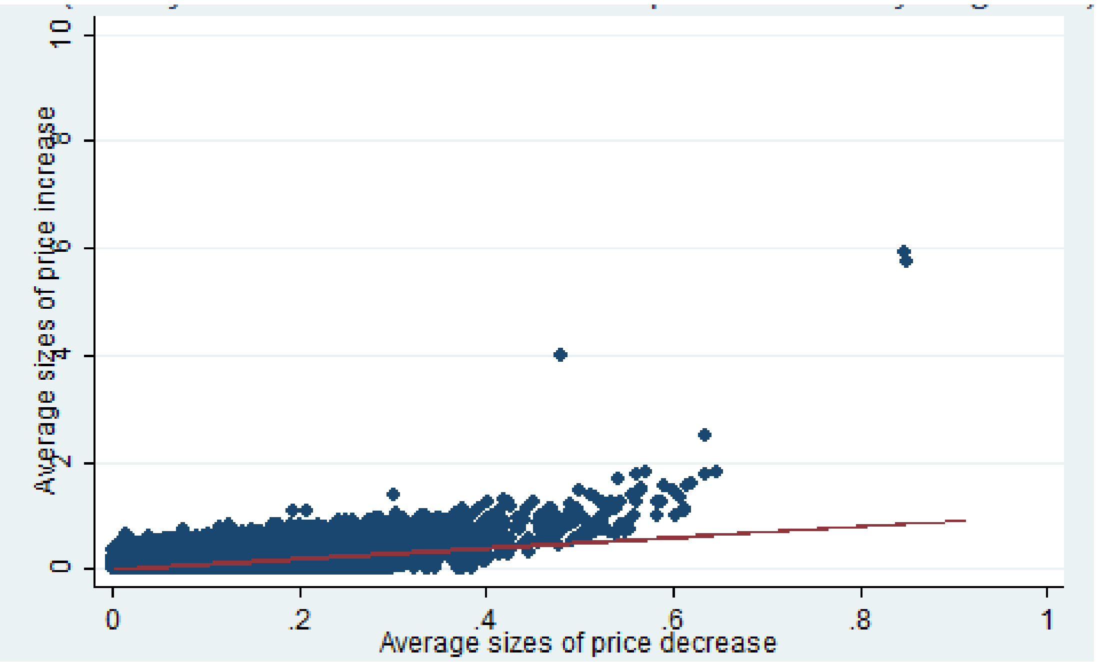

Figure B1: Average sizes of price increases and price decreases for Shop A: Part 2 (98 days of data)

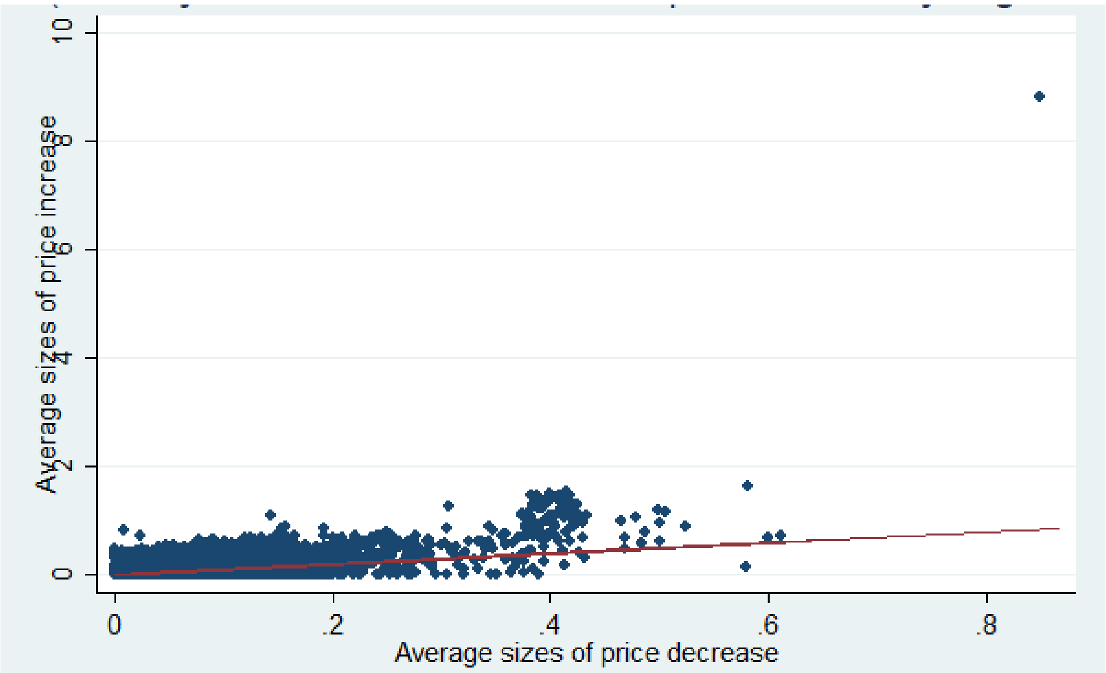

Figure B2: Average sizes of price increases and price decreases for Shop B: Part 2 (100 days of data)

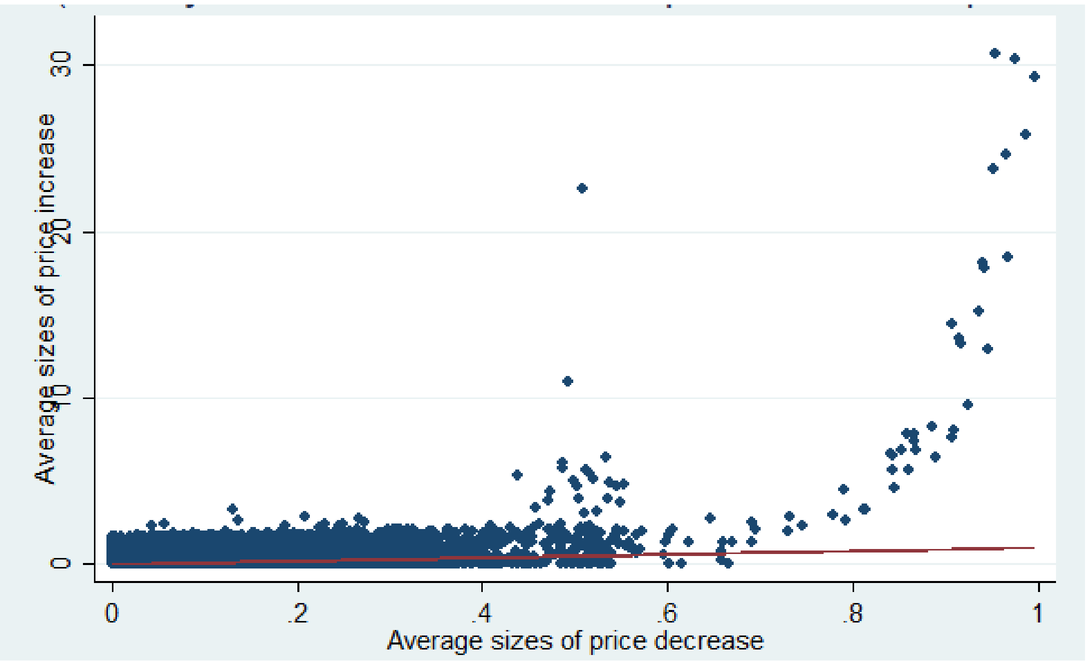

Figure B3: Average sizes of price increases and price decreases for Shop C: Part 2 (156 days of data)

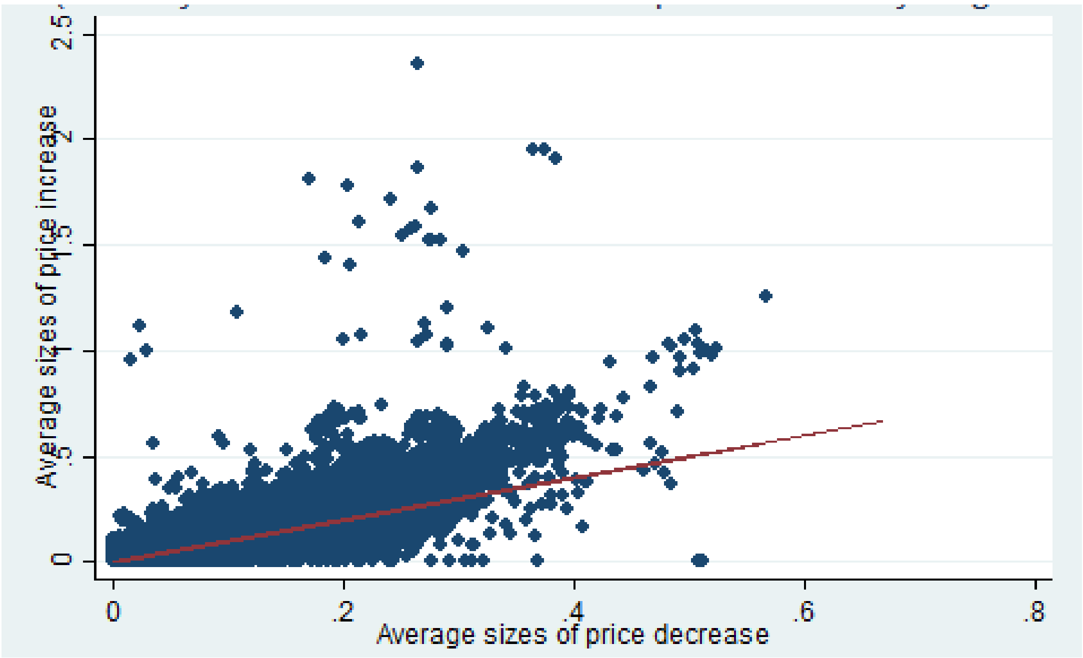

Figure B4: Average sizes of price increases and price decreases for Shop D: Part 2 (100 days of data)