Taewon Hwang(2020711771)
Industrial Organization 1
Homework 1

Part I

1. Data Processing

The summary statistics for the weekly prices of the 10 best-selling UPCs (in terms of dollar sales over the years 1995-1996) are given as follows. The summary statistics does not evaluate the case in which prices are zero, that is, the item is not sold in the corresponding week.

Table 1
Summary statistics of prices of the top 10 best-selling UPCs, ordered by rank

rank (UPC)	mean	sd	min	p25	p50	p75	max
1 (8000000245)	0.79	0.12	0.56	0.68	0.83	0.85	0.96
2 (4800000024)	0.80	0.08	0.58	0.74	0.83	0.85	0.89
3 (7303010170)	4.47	0.51	2.44	4.44	4.60	4.70	5.05
4 (8660000033)	0.81	0.10	0.41	0.79	0.83	0.85	0.98
5 (7303010314)	2.16	0.42	1.05	2.15	2.32	2.39	2.57
6 (7114000200)	1.94	0.23	0.11	1.94	1.98	2.02	2.08
7 (8660000001)	3.21	0.08	3.04	3.16	3.21	3.26	3.37
8 (3828105005)	0.71	0.08	0.48	0.69	0.71	0.71	0.86
9 (7114010110)	1.37	0.07	1.15	1.35	1.39	1.41	1.49
10 (8660000002)	1.64	0.07	1.43	1.59	1.65	1.68	1.81

Below plots the time series of the weekly prices over 1995-1996 of each UPC. Each line refers to a single UPC(item), and the labels correspond to rank, generated by volume of dollar sales over 1995-1996. Note that week 277 corresponds to the first week of 1995, and week 381 corresponds to the last week of 1996.

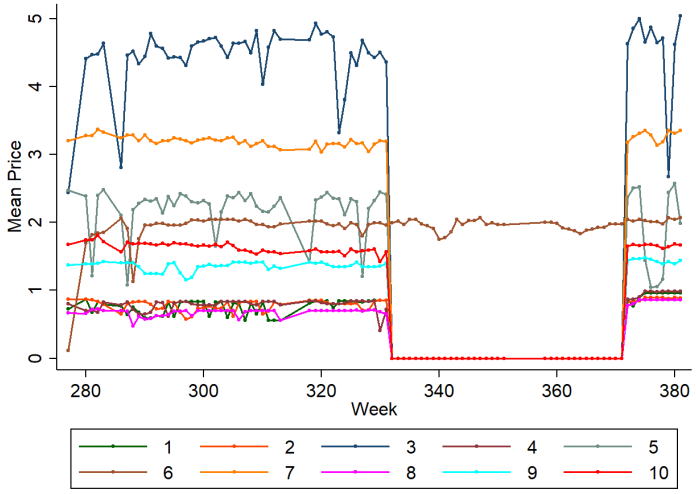

Figure 1
Time series of prices of the top 10 best-selling UPCs, 1995-1996

2. Summary Statistics

Summary statistics of the sales data is given as follows.

Table 2
Summary statistics (Sales)

	mean	sd	min	p25	p50	p75	max
sales	10.27	33.69	0.00	0.00	2.96	11.16	4410.25

Summary statistics of the demographic variables of each store may also provide insight into our data. Here, summary statistics of the income of the log of median income of customers for each store is given. Summary statistics for average household size, ethnicity, proportion of college graduates can also serve as important demographic features, nevertheless, are not reported here due to limit of space.

Table 3
Summary statistics (Profits, by store)

store	mean	sd	min	p25	p50	p75	max
2	20.69	16.66	-36.85	0.00	27.92	33.50	69.71
5	16.69	16.49	-19.07	0.00	20.58	31.52	70.24
8	18.08	15.79	-73.83	0.00	21.33	31.36	71.48
9	20.05	16.36	-54.88	0.00	26.52	33.14	70.24
12	21.39	17.17	-47.69	0.00	28.87	34.78	69.71
14	19.74	16.93	-41.70	0.00	26.66	33.39	69.71
18	20.18	16.20	-49.58	0.00	26.40	33.03	70.24
21	15.35	15.84	-80.00	0.00	14.31	29.74	71.48
28	19.08	16.49	-49.58	0.00	25.00	32.75	70.24
32	21.22	16.52	-41.70	0.00	28.04	33.61	69.71
33	20.87	17.16	-46.49	0.00	28.36	34.20	73.64
40	15.09	15.83	-80.00	0.00	13.04	29.51	69.50
44	19.98	16.29	-49.58	0.00	26.48	33.10	85.93
45	1.10	5.81	-12.10	0.00	0.00	0.00	62.62
47	17.82	16.66	-54.88	0.00	22.63	32.40	70.24
48	13.85	16.31	-47.69	0.00	0.00	30.06	70.24
49	1.21	6.05	0.00	0.00	0.00	0.00	62.70
50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
51	18.66	16.88	-46.69	0.00	25.21	33.09	70.24
52	20.90	16.63	-36.85	0.00	28.02	33.56	69.71
53	18.72	17.68	-50.94	0.00	25.20	33.92	69.71
54	19.07	16.59	-54.88	0.00	25.03	32.78	71.48
56	18.80	16.45	-54.88	0.00	24.51	32.69	71.48
59	16.10	15.28	-94.73	0.00	17.77	29.49	67.48
62	20.96	16.64	-36.85	0.00	28.02	33.50	85.93
64	0.56	4.31	-42.96	0.00	0.00	0.00	56.14

67	19.60	16.43	-49.58	0.00	25.78	32.94	71.48
68	20.10	16.88	-41.70	0.00	27.15	33.65	69.71
70	15.80	15.24	-88.07	0.00	17.16	28.99	69.46
71	18.45	15.75	-73.83	0.00	22.44	31.27	69.50
72	19.83	16.98	-41.70	0.00	26.96	33.39	69.71
73	19.33	16.46	-49.58	0.00	25.17	32.69	70.24
74	20.35	16.30	-62.92	0.00	26.53	33.23	70.24
75	21.50	17.18	-66.32	0.00	28.87	34.88	69.71
76	17.78	16.71	-79.69	0.00	22.87	32.60	70.24
77	16.38	15.32	-88.07	0.00	18.34	29.74	85.93
78	17.87	15.80	-94.22	0.00	21.16	31.20	69.50
80	18.55	15.72	-73.83	0.00	22.48	31.64	69.50
81	20.14	16.32	-49.58	0.00	26.40	33.19	70.24
83	17.11	15.84	-54.88	0.00	19.84	30.83	69.50
84	17.05	15.44	-73.83	0.00	19.66	30.44	67.75
86	21.96	15.94	-36.85	0.00	28.04	33.50	70.24
88	17.50	16.52	-54.88	0.00	22.30	32.00	70.24
89	16.71	16.68	-82.42	0.00	20.42	31.92	68.66
90	16.53	17.22	-45.20	0.00	14.47	32.69	71.48
91	16.96	16.63	-54.88	0.00	20.93	31.91	70.24
92	14.33	16.14	-20.68	0.00	0.00	29.90	70.24
93	21.03	16.69	-72.22	0.00	28.04	33.56	69.71
94	18.99	16.39	-49.58	0.00	24.77	32.69	71.48
95	19.80	17.03	-36.85	0.00	26.83	33.49	69.71
97	0.00	0.00	0.00	0.00	0.00	0.00	0.00
98	19.21	16.53	-92.93	0.00	25.17	32.75	70.24
100	21.08	16.82	-98.33	0.00	28.04	33.96	69.71
101	19.49	16.42	-61.73	0.00	25.31	32.91	70.24
102	20.08	16.36	-42.93	0.00	26.40	33.23	70.24
103	16.56	16.66	-49.58	0.00	19.74	31.88	70.27
104	19.84	16.47	-89.91	0.00	26.40	33.09	70.24
105	15.97	16.58	-64.79	0.00	14.02	31.52	70.24
106	15.41	16.56	-49.58	0.00	0.00	30.98	71.48
107	19.04	15.59	-73.83	0.00	23.79	31.68	69.50
109	21.09	17.29	-93.27	0.00	28.77	34.36	69.71
110	18.30	16.57	-53.23	0.00	23.81	32.59	70.24
111	17.44	17.34	-38.29	0.00	21.76	33.00	69.71
112	19.19	16.40	-49.58	0.00	25.30	32.83	70.24
113	20.46	16.25	-79.77	0.00	26.84	33.17	70.24
114	18.29	16.75	-74.14	0.00	23.75	32.75	70.24
115	19.69	16.36	-49.58	0.00	26.11	32.94	71.48
116	19.64	16.43	-54.88	0.00	25.93	33.10	70.24
117	13.52	15.90	-15.75	0.00	0.00	29.05	70.24
118	1.37	6.57	0.00	0.00	0.00	0.00	67.54
119	19.16	16.51	-54.88	0.00	25.08	32.78	70.24

121	20.84	16.10	-49.71	0.00	27.49	33.23	70.24
122	18.17	15.82	-73.83	0.00	22.00	31.53	69.50
123	20.40	16.82	-36.85	0.00	27.61	33.55	69.71
124	20.17	16.86	-36.85	0.00	27.31	33.55	69.71
126	20.47	16.22	-49.58	0.00	27.31	33.22	70.24
128	21.62	16.56	-46.66	0.00	28.31	34.13	69.71
129	21.07	15.99	-49.58	0.00	27.48	33.23	70.24
130	18.80	17.20	-41.70	0.00	25.72	33.32	69.71
131	20.05	16.30	-49.58	0.00	26.40	33.10	70.24
132	18.54	15.76	-73.83	0.00	22.66	31.49	69.50
133	20.68	16.19	-79.60	0.00	27.29	33.23	70.24
134	18.65	16.48	-49.58	0.00	24.33	32.69	70.24
135	20.69	15.00	-94.73	0.00	24.87	32.32	60.60
136	18.59	16.55	-82.44	0.00	24.05	32.69	85.93
137	20.61	16.70	-36.85	0.00	27.94	33.46	70.30
139	20.65	16.18	-74.15	0.00	27.31	33.23	70.24
140	21.29	14.60	-88.07	11.69	25.07	31.76	64.10
141	21.59	14.56	-88.07	12.43	25.30	32.11	62.41
142	22.88	15.16	-73.83	8.72	28.18	33.33	65.56
143	26.49	14.21	-54.88	22.94	30.64	35.15	60.60
144	26.18	14.49	-54.88	21.70	30.22	35.15	67.75
146	23.40	15.27	-49.58	10.94	28.87	34.60	68.65
Total	18.12	16.62	-98.33	0.00	23.43	32.69	85.93

3. HHI

Summary statistics for the HHI of each product sold in Chicago are given as follows. Items in which the generated HHI value is 0 are excluded from the sample, as items with no aggregate market share during this period are of little interest. Hence, a total of 22 items were dropped from the sample.

Table 4
Summary statistics (HHI, by UPC)

	mean	sd	min	p25	p50	p75	max
HHI	140.59	833.22	2.00	23.00	34.50	57.00	10000.00

Results indicate that the degree of competition between stores for each UPC is relatively low. Note that a very high HHI value may imply that the item is being supplied to minimal, or selective stores.

Part II

In this section, price variation across various strata are explored. It should be noted that to adequately assess variations in price, a specific unit price among items must be constructed. In this section, prices of items are standardized in proportion to a certain volume and package size of the item; the prices correspond to a 12 pack of 16 oz. tuna units. While there may be heterogeneity amongst individual items observed in the data set, for the first two questions, I assume that standardizing the price is suffice to aggregate the items into an singular item.

1. Price Variation across Cities

Here, the price variation between cities with the highest 10% median income and those with the lowest 10% median income is investigated. Median income value data of the customers of each store are provided in the store-specific demographics dataset. These are aggregated to construct the median incomes of each city.

I find that the cities with the highest 10% median income are `BANNOCKBURN, GLENVIEW, LAKE ZURICH, NORTHBROOK, NORTHFIELD`, and those with the lowest 10% median income are `CALUMET CITY, CHICAGE, NORTH RIVERSIDE, OAK PARK, RIVER GROVE`. Dummy variables `highinc` and `lowinc` are constructed, taking the value 1 if the observation is recorded in respective groups of cities.

I calculate the standard deviation of standardized prices within each city to serve as a proxy for price variation between the high income and low income cities. The summary statistics of the standard deviation of standardized prices of the cities with higher incomes and those with lower incomes are given as follows.

Table 5
Summary statistics (Price variation, low income cities)

	mean	sd	min	max
sd	278.65	20.64	246.16	299.84

Table 6
Summary statistics (Price variation, high income cities)

	mean	sd	min	max
sd	286.16	21.12	267.33	317.91

Note that cities with higher median income tended to have larger levels of price variation, which may be generated from (1) the plausibly larger heterogeneity in non-item properties, such as customer service or (2) higher search costs. That is, one may suggest that those with higher incomes have larger search costs, leading them to accept prices above the minimum charged in the market. If firms choose to exploit this demographic by implementing a mixed-strategy equilibrium, the variation between prices may increase.

2. Price Variation across Zones

The variation of standardized prices across the 16 zones are given as follows.

Table 7
Summary statistics (Price variation, by zone)

	mean	sd	min	max
sd	268.58	29.25	183.46	311.74

Note that the level of price variation across zones is lower than that of both levels of price variation in question 1, which corresponds to the uniform zone pricing strategies employed by DFF.

3. Price Dispersion between Online/Offline Markets

To test the claim that price dispersion is less substantial in online markets, I construct a dummy variable which takes the value 1 if a zone assignment is not made (hence, an offline store). Redefining price definition as price variation within each UPC, the summary statistics regarding price variation within offline markets and online markets are given as follows.

Table 8
Summary statistics (Price variation, by offline/online market)

sd	mean	sd	min	max
Offline	49.96	161.01	0.00	1228.30
Online	51.71	160.92	0.00	1202.87

Note that the level of variation in UPC prices are lower in offline markets, opposing the claim that price dispersion should be more substantial. This may be caused by stickier prices in an offline environment, that is, prices may not be as flexibly adjusted as in the case of online markets.

4. Variation in Markups across Zones

The variation in markups across zones is conveniently investigated via the summary statistics of `profit` by zone, as `profit` indicates the gross margin in percent that the store makes on the sale of the item. The results are as follows.

Table 9
Summary statistics (Markup, by zone)

	mean	sd	min	max
markup	18.21	2.97	8.97	21.35

The mean of the average markups for each zone is 18.21%, and the standard deviation of average markups across zones is 2.97%, which serves as a measure for variation in markup. While the magnitude of variation in markup cannot be stated explicitly, barring one outlier with a average markup of 8.97%, the average markup of remaining zones exhibit a bell-shaped distribution concentrated between 16˜22%, as plotted in Figure 2.

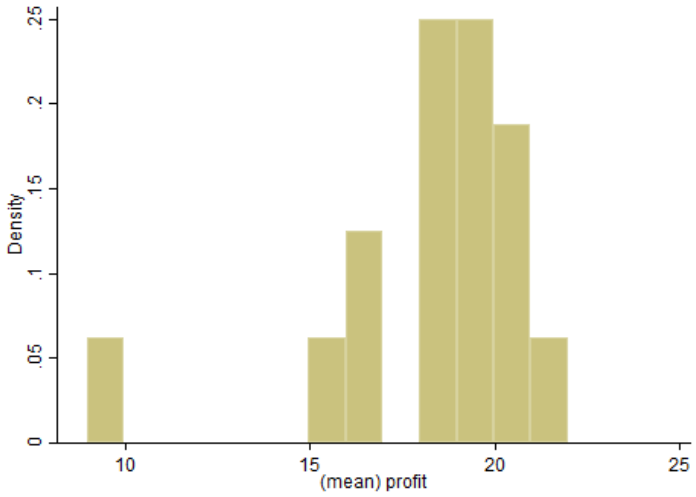

Figure 2
Distribution of markup, by zone

Part III

Table 2, Chevalier, Kashyap, and Rossi(2003)

Chevalier, Kashyap, and Rossi(2003)(hereby referred to as "CKR") attempt to substantiate the empirical findings of falling prices during demand peaks, which is largely anomalous to the theoretical implications of a standard competitive model of price determination. Regarding data processing, CKR(2003) provides instructions on how to construct variables that represent the periods of seasonal demand peaks in detail, and the procedure is listed as follows.

1) Dummy variables (hereby referred to as "the variable") that take the value 1 for two shopping weeks before each holiday are generated.
2) For holidays that land on a Thursday, the variable is set to 1 for the two weeks prior to the holiday, by 0 for the week including the holiday.
3) For holidays that land on any other day, the variable is set to 1 for the week before the holiday and the week including the holiday.
4) The variable for `Christmas` remains equal to 1 for the week following the holiday, relaxing the need for a separate variable for New Years.
5) The variable for `Post-Thanksgiving` takes the value of 1 for the week following Thanksgiving.
6) The variable for `Lent` takes the value of 1 for the four weeks preceding the two-week Easter shopping period.

The references below provide data on the date and weekday of the holidays in 1995 and 1996, and are therefore used to construct the holiday dummies as given in the literature.

`https://print-a-calendar.com/1995-holidays`
`https://print-a-calendar.com/1996-holidays`

Note that in 1995, Thanksgiving Day lands on a Thursday, and in 1996, Independence Day, Halloween, and Thanksgiving Day are on Thursdays. This is accounted for in the construction of the respective dummy variables.

Next, I construct an item aggregate focusing on a select group of items. Since the items "Chicken of the Sea 6 oz. Chunk Lite Tuna in Oil" and "Chicken of the Sea 6 oz. Chunk Lite Tuna in Water" are not explicitly found in the database, I arbitrarily select the closest matches to these items. A total of 5 items are selected for analysis:

```
(1780000012) COS CHUNK LIGHT OIL 6 OZ
(1780000024) C O S CHUNK LIGHT TU 6 OZ
(4800000012) COS CHUNK LIGHT OIL 6 OZ
(4800000024) C O S CHUNK LIGHT TU 6 OZ
(4800000120) COS CHUNK LIGHT OIL 6 OZ
```

Two regressions are implemented to investigate the price effects at peak demand periods. The results of the conventional pooled OLS model are given in (1), while the second regression controls for the categories of UPCs included in the aggregate item bundle. The results of the latter regression are provided in column (2).

Table 10
Estimated Price Effects at Peak Demand Periods for Individual Item Aggregates

	(1) OLS		(2) UPC fixed effects	
	Std. Price		Std. Price	
Independence Day	-1.782***	(-29.57)	-1.710***	(-27.11)
Christmas	0.420***	(4.27)	0.390***	(3.96)
Easter	1.384***	(8.61)	1.361***	(8.45)
Halloween	1.715***	(11.20)	1.669***	(11.01)
Labor Day	-1.016***	(-6.83)	-0.943***	(-6.41)
Memorial Day	-6.309***	(-65.97)	-6.236***	(-63.05)
Presidents Day	-0.696***	(-5.62)	-0.569***	(-4.48)
Thanksgiving	1.008***	(9.48)	0.965***	(9.12)
Post-Thanksgiving	1.009***	(6.93)	0.946***	(6.52)
Lent	-2.111***	(-11.49)	-2.049***	(-11.17)
Constant	26.78***	(642.35)	26.78***	(648.09)
N	9987		9987	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As exhibited in CKR(2003), during `Lent`, a period in which demand for canned tuna is anticipated to increase, statistically significant negative price effects can be estimated. Meanwhile, the downward price effect is estimated to be even stronger during the shopping period adjacent to `Memorial Day`, which leaves an inquisitive question for further investigation.

References

Chevalier, J. A., A. K. Kashyap, and P. E. Rossi (2003): "Why Don't Prices Rise during Periods of Peak Demand? Evidence from Scanner Data," American Economic Review, 93(1), 15-37.