Appendix

This appendix describes the methods and algorithms that were used for the creation of the various statistics and figures in the paper. Each subsection refers to a specific set of statistics (for example, "Table 1 Statistics", "Statistics Regarding the Markov Chain" etc.) and figures. At the bottom of each subsection (under the title "Programs to Run") we describe which programs should be used in order to replicate the statistics.

Creating Reference Prices

We describe here the method that was used for the creation of the reference prices.

1. For every category find all UPC/Stores pairs.
2. For every UPC/Store pair record the weekly prices, costs, and quantities sold.
3. Divide the 156 weeks of data into 12 quarters of 13 weeks each (12*13=156).
4. Find for each UPC/Store/Quarter the number of weeks in which the UPC was sold.
    - If there are not at least 12 weeks with data do not consider the UPC/Store pair and it is outside of the sample.
    - We thus consider only UPC/Store pairs that appear for all 12 quarters and at least for 12 weeks in each quarter.
5. For each UPC/Store/Quarter calculate:
    - Modal price (i.e. the "reference price")
    - Second modal price (i.e. the second most frequent price)
    - Modal cost (i.e. the "reference cost")
    - Sum of quantities sold at the first modal price.
    - Sum of quantities sold at the second modal price.
    - Sum of quantities sold during the entire quarter.

Programs to Run:

1. *"Create_Reference_Prices".*

Table 1 Statistics

We describe here the method that was used for calculating the first 9 statistics in Table 1. The construction of Statistics 10-13 in Table 1 are explained later.

1. For each of the UPC/Store/Quarter triplets that has passed step 4 in the subsection "Creating Reference Prices" we have recorded:

   - Weekly series of prices within each quarter.
   - 1^{st} and 2^{nd} modal prices within each quarter.
   - Weekly series of costs within each quarter.
   - 1^{st} and 2^{nd} modal costs within each quarter.

2. For each UPC/Store/quarter we find (these statistics are reported in the first six lines of table 1).

   - Fraction of weeks in which the price (cost) is at the modal price (cost).
   - Fraction of weeks in which there is only one price.
   - Fraction of quantities that were sold at the first modal price.
   - Fraction of revenues that were sold at the first modal price.
   - Fraction of quantities that were sold at the second modal price.
   - Fraction of revenues that were sold at the second modal price.
   - For each category calculate the median of each of the statistic and we report the median across the categories.

3. Statistics Regarding Prices that are Above the Reference Price (these statistics are reported in the seventh line of table 1).

   (a) For each UPC/Store/ and quarters 2:11 calculate:

      i. Find the weeks in which the weekly price is above the current modal price – this is done for quarter $t = (2 : 11)$
      ii. For each of the weeks that were identified count the number of weekly prices in those weeks is different from **both** the modal price in quarter $(t-1)$ and quarter $(t+1)$.
      iii. Divide the number you get from the second step by the number you get from the first step.
      iv. For each category calculate the median of each of the statistic and we report the median across the categories.

4. Statistics Regarding the Price Changes (these statistics are reported in the eighth and ninth rows of table 1).

   (a) For each of the UPC/Store/Quarter triplets we identify all price changes within a quarter.

   (b) We then calculate how many of these price changes (within a quarter) are from non-reference to reference, reference to non-reference, and non-reference to another (different) non-reference.

(c) For each category we count the number of transitions that fall into each of the price changes possibilities.

(d) For each category, we count the number of price change from non-reference to reference and divide by the number of all price changes.

(e) The same process is done for cost.

(f) We then report the average across all categories and we report the average across the categories.[1]

Programs to Run:

1. "Table_1_Statistics_price" and "Table_1_Statistics_cost": These programs will generate the statistics and record them.

2. "Table_1_Statistics_price_collect" and "Table_1_Statistics_cost_collect": These programs create the median (or mean when appropriate) value of the statistic per each category.

Statistics Regarding the Markov Chain

We describe here the method that was used for calculating the Markov chain statistics (these statistics are reported in page xxx of the paper).

1. Each weekly price is assigned to "state 1" if that weekly price is the reference price and "state 2" if they differ.

2. Within each quarter we look at the "$Number\ of\ weeks - 1$" weeks and calculate the number of transitions between the two states. We overall have four transition possibilities.

3. For each category we count the number of transitions that fall into each of the 4 transition possibilities.

4. For each category, in order to get the "category" transition probability from state 1 to state 2 we divide the number of transition from state 1 to state 2 by the sum of (1) the number of transitions from state 1 to state 2, and (2) the number of transitions from state 1 to state 1. We proceed in a similar way for all the other transitions.

5. The same process is done for cost.

6. Once we have calculated the transitions for all categories we report the average across all categories.[2]

[1] Note that we report the average and not the median to assue that the sum of probabilities of all price changes is equal to 1.

[2] Note that we report the average and not the median to assue that the transition probabilities sum to 1.

Programs to Run:

1. "Calculate_Markov": This program will generate the statistics and record it.

2. "Calculate_Markov_collect": This program creates mean value of the statistic per each category.

Table 2-4 Statistics

We describe here the method that was used for calculating the statistics that appear in Tables 2-4. The same method was used to calculate the statistics that are reported in pages xxx of the paper.

1. For each UPC/Store pair we have data over the 156 weeks. We find all the consecutive weeks for which we have data without an interruption (for example, if there is data from weeks 1:30, 32:105, 108:156 we record these three "blocks" of data).

2. We find the block that has the largest number of weeks (in the example above, it will be the second block - i.e. weeks 32:105).

3. For each UPC/Store we then have a sequence of 12 reference prices, costs and the largest block of consecutive weeks (for which we have prices, costs, and quantities sold).

4. We use the data we recorded from the third step and calculate for each UPC/Store the number of price changes (once for weekly data and once for reference prices). We then divide by the total number of "$weekly\ data - 1$" and by 11 for the reference prices (since we are only looking at the UPC/Store pairs that have 12 quarters) to calculate the probability of price changes. We do the same for the cost.

5. We use the data we recorded from step 3 and calculate for each UPC/Store various correlations reported in Table 4.

6. With the data from step 3 recorded for each UPC store we can now calculate all the statistics reported in Table 2-4. Specifically, for each category we calculate the median of the statistic across the UPC/stores that belong to the category. We then calculate the median of the statistic across all categories.

Programs to Run:

1. "Table_2_3_4_Part1"

2. "Table_2_3_4_Part2"

3. Install the function "function_probchange", "function_probchange_dif","function_std"

4. Run "Table_2_3_4_Part3".

The Relation between Prices and Cost Changes

We describe here the method that was used for calculating the statistics regarding the relation between prices and cost changes. These are reported in the bottom of Table 1.

1. For the weekly relation we use the blocks step 2 in the Section "Table 2-4 Statistics".
2. Within a category the stat is calculated as mean.
3. We then calculate the mean of the statistic across all categories.

Programs to Run:

1. Ref_Price_Cost_Relation
2. Ref_Price_Cost_Relation_Collect
3. Week_Price_Cost_Relation
4. Week_Price_Cost_Relation_Collect

Figures 1-4

1. We use the price statistics that were generate in "Table 1 Statistics" and plot the value for each category.

Figures 5-6

1. We identify over the three years the 12 reference prices.
2. We find the most common reference price out of these 12.
3. We find the longest consecutive occurrence of the price from step 2. For example, assume that the 12 reference prices are given by

$$1, 1, 1.5, 1.5, 1, 1, 1, 1, 2, 2, 1, 1.5$$

we will then record as the most common reference price the number "1" and note that it appears 4 consecutive times starting in the fifth quarter. Denote this by $p_{\text{mod}} = 1$ and record the first and last quarter to have this continuos reference price be $dur_{beg} = 5$ and $dur_{end} = 8$.

4. Let the set of weeks which had the price p_{mod} between dur_{beg} and dur_{end} be denoted by j where j is a set of numbers of size N.

5. We record for all the weeks within the set j the cost. Let the vector of costs be $C_j = \left\{c_{j(1)}, c_{j(2)}, ...c_{j(N)}\right\}$

6. We calculate the markup for all the weeks that had the mode price over the three weeks,
$$\mu_j = \left\{\frac{p_{\text{mod}}}{c_{j(1)}}, \frac{p_{\text{mod}}}{c_{j(2)}}, ...\frac{p_{\text{mod}}}{c_{j(N)}}\right\}$$

7. We calculate the average of the markup over all these weeks (note that the price is always the same in these weeks).
$$\overline{\mu} = E(\mu_j)$$

8. We calculate per each week that had the mode price the percentage differences of the markup in that week relative to the average markup from step 4.
$$\Delta\mu_j = \left\{\log\left(\frac{p_{\text{mod}}}{c_{j(1)}}/\overline{\mu}\right), \log\left(\frac{p_{\text{mod}}}{c_{j(2)}}/\overline{\mu}\right), ... \log\left(\frac{p_{\text{mod}}}{c_{j(N)}}/\overline{\mu}\right)\right\}$$

9. We create a vector of 201 bins that goes from -1 to 1 in increments of 0.01. Each element of $\Delta\mu_j$ is put into the appropriate bin that records the value of the elements. For example if the second element $\log\left(\frac{p_{\text{mod}}}{c_{j(2)}}/\overline{\mu}\right)$ is 0.04 then it is inserted into the 105th bin. This way we record per each good to which bins the elements of $\Delta\mu_j$ correspond.

10. We repeat steps 1-7 for all goods and for each good we also record the duration of the longest consecutive occurrence, i.e. $(dur_{end} - dur_{beg} + 1)$. Note that this duration can be any number between 1 and 12.

11. For each of the possible 12 durations we then identify all UPC/Stores that had that duration (again, as captured by $(dur_{end} - dur_{beg} + 1)$) and we sum their vectors. We thus get one vector of 201 elements that has in each bin an integer number that represents the occurrences of $\Delta\mu_j$ that belong to that bin within all the goods of the same duration. We then construct the distribution for each of the possible 12 durations.

12. Figure 6 plots those 12 durations from step 11 while Figure 5 plots it for all the durations together.

Programs to run:

1. Dist_Markup_over_qtrly_refprice_part1

2. Dist_Markup_over_qtrly_refprice_part2

Figure 7

1. We use the statistics that were generate in "Table 2-4 Statistics" and plot the scatter plot of all the categories.

Figure 8

1. We identify over the three years the 12 reference prices and reference costs.
2. We calculate the markups for each of the 12 quarters.
3. We calculate the mean of the markup from step 2.
4. For each of 12 quarters we calculate the % deviations of the quarterly markup relative to the mean from step 3.
5. We create a vector of 201 bins that goes from -1 to 1 in increments of 0.01. Each element of from step 4 is put into the appropriate bin that records the value of the elements as it was done in the discussion of figures 5-6
6. For each bin we then have the number of times we were there and for how many of those the price was changed.
7. Similarly, we record for each bin and for all the cases of the prices that were changed, the bin to which the resulting markup went to (in terms of % deviations from the average markup).
8. We repeat this process for all the UPC/stores. We then sum across all these pairs (within a category) the information from the bins (from steps 6-7) and we thus have a "probability of a change and "where do prices go after a change" per each category. We then calculate for each bin the average across the categories.

Programs to run:

1. V_shapre_part1
2. V_shapre_part2
3. V_shapre_part3

Figure 9-10

1. The first seven steps are identical to those used in Figure 8.

2. We then record for each UPC/Store pair its (i) duration of reference price, and (ii) cost volatility.

3. For each of the possible cases of reference price duration and cost volatility we repeat step 8 from the discussion of Figure 8. That is, assuming there are X upc/stores pair that fall into the "4 quarters duration" group then we sum across all the bins of the X products in this group and calculate the "probability of a change and "where do prices go after a change" for the "4 quarters duration" group. This process is repeated for all 12 duration and for the 6 groups of cost volatility we consider.

Figure 11

1. For each upc/store identify the quarters in which the reference price does not change - assume for example these are periods $\{t, t+1, t+2, t+3\}$

2. Calculate the markup over those periods.

3. Calculate the % deviations of the markups from step 2 relative to the realized markups once the price changes (i.e. in the example period $t+4$).

4. Take the median within a category from step 3 over all the goods at the last period before the change (in the example that would be $t+3$). Then proceed and calculate the median from step 3 over all the goods at two periods before the change (in the example it would be $t+2$). Keep on repeating this process until you arrive to a maximum of 11 (as there are 12 observations we could see 11 constant prices and then in the last 12th quarter a price change). Within each iteration we calculate the median only for those goods that indeed had a price constant sufficiently in the past. That is, in the example above, that good will not enter into the calculation starting at 5 periods before the change since such an observation does not exist for the good.

5. We then take medians across categories for each of the possible 11 steps. We truncate the figure at 10 since there are almost no observations at "11 periods before the price change.

Programs to run:

1. Markup_before_change_ref_1

2. Markup_before_change_ref_1_collect