Readme:
"Household Inventory, Temporary Sales, and Price Indices"

Kozo Ueda, Kota Watanabe, and Tsutomu Watanabe

June 19, 2023

In this document, when elucidating the data structure of the files employed for input and output, each line described within the file is denoted as $\{v_1, v_2, \cdots, v_n\}$, while columns/data elements that are irrelevant to our analysis are designated as X.

1 Data we use

We use two sets of Japanese scanner data. The first set consists of retailer-side scanner data collected by Nikkei Inc. This dataset contains three kinds of data, referred to as "Nikkei scanner data," "Nikkei category master," and "Nikkei product master," which will be described in Sections 1.1 and 1.2.

The second set consists of household-side data, namely, Shoku-map scanner data collected by Lifescape Marketing Co. This dataset consists of two kinds of data called "ZAIKO" and "TABLE" data, which will be explained in Section 1.4

1.1 Nikkei scanner data

The Nikkei scanner data are a set of retailer-side daily scanner data, running from March 1, 1988, to December 31, 2019. This dataset contains information on the number of units sold and sales amount for each product and retailer on a daily basis. Each product is uniquely identified by the JAN and GEN codes.

The JAN code is a product identifier assigned by GS1 Japan. However, it is not always sufficient for uniquely identifying a product because the JAN code can be reused for different products (e.g., between old and new products). The GEN code, provided by Nikkei, serves to differentiate products that share the same JAN code.

Each product is classified into one of the 6- (small) and 3-digit (middle) product categories defined by Nikkei. While we will calculate price indices, the degree of stockpiling, and other variables for each product category, many files of the Nikkei scanner data lack product category information for each product/record. Therefore, we need to refer to the "Nikkei product master" to identify the product category for each product, as will be explained in Section 1.2.

1.1.1 Data availability statements

Nikkei scanner data is proprietary; please contact Nikkei POS Joho (nkpos.nikkei.co.jp) if you would like to use the data.

1.1.2 Record/Structure

- The files for scanner data are in CSV format. The records in each file correspond to one of the following:
  - {X, Store, Date, JAN code, GEN code, Sales in yen, Quantity}.

    Store : Retailer ID.

Date : YYYY/MM/DD.

JAN code : Product code.

GEN code : Code to distinguish products with the same JAN code.

Sales in yen : Daily ammount of sales for a particular product sold by a particular retailer on a particular date.

Quantity : Daily ammount of quantity sold for a particular product sold by a particular retailer on a particular date.

– {X, Store, Date, JAN code, GEN code, 6-digit product category, Sales in yen, Quantity}.

6-digit product category : Nikkei product category, as will be explained in the next section.

1.1.3 Notes

- We do not use the records in POS data whose product category code cannnot be identified from the Nikkei product master.

1.2 Nikkei category and product masters

In our scanner data, each product belongs to one of the 6- and 3-digit product categories defined by Nikkei [1]; these category codes are listed in files called "Nikkei category master." In addition, the data that record the codes for each product paired with its corresponding product category are listed in files called "Nikkei product master."

There are 13 files for both "Nikkei product master" and "Nikkei category master," and by merging them, we can utilize more data. However, since the product category assigned to each product differs depending on the product master in some cases, in our analysis, we will only use the latest 3-digit product category assigned to each product.

1.2.1 Data availability statements

These data are proprietary; please contact Nikkei POS Joho (nkpos.nikkei.co.jp) if you would like to use the data.

1.2.2 Record/Structure

- "cat.csv"; the number of files is 13.
  These are the "Nikkei category master" files, and the records in each file are as follows:
  - {6-digit product category, Name of product category in Japanese}.
- "itm.csv"; the number of files is 13.
  These are the "Nikkei product master" files, and the records in each file are as follows:
  - {JAN code, GEN code, 6-digit product category, Name of product in Japanese}.

[1] The 3-digit product category of XYZ is represented by XYZ000. The 6-digit product categories included in XYZ are expressed as XYZ001, XYZ002, or so on.

1.3 Reconstruction of original Nikkei data

1.3.1 NikkeiPOS_mergeItemFile.c

We merge the master files for "Nikkei category master" and "Nikkei product master," respectively.

Input files

1. "cat.csv" listed in "list_cat.txt"
2. "itm.csv" listed in "list_itm.txt"

Output files

1. "list_3digit_2008_2019.txt"
   It is the merged category master file, and its records are as follows:
   - {3-digit product catetory}.
2. "list_item_2008_2019.txt"
   It is the merged product master file, and its records are as follows:
   - {JAN code, GEN code, 6-digit product catetory, Name of product in Japanese}.

1.3.2 NikkeiPOS_3digitCategoryLevel.c

We construct the set of scanner data for each 3-digit product category.

Input files

1. "list_3digit_2008_2019.txt"
2. "list_item_2008_2019.txt"
3. The Nikkei scanner data ("sum_XXX.csv") listed in "list_sum.txt"
   The first column in "list_sum.txt" shows the name of the file for each scanner data. The second column represents the type of data field (0 or 1) for each scanner data.
   - If the second column in "list_sum.txt" shows 0, then the records in the corresponding file are as follows:
     - {X, Store, Date, JAN code, GEN code, Sales in yen, Quantity}.
   - If the second column in "list_sum.txt" shows 1, then the records in the corresponding file are as follows:
     - {X, Store, Date, JAN code, GEN code, 6-digit product category, Sales in yen, Quantity}.

Output files

1. Scanner data for each 3-digit product category (e.g., scanner data for the product category 1 are recorded in "1.dat"). [2]
   - {X, Store, NewJAN, JAN code, GEN code, 6-digit product category, Year, Month, Day, Sales in yen, Quantity}.

   NewJAN : New product ID.
2. "JANcode_reconstructed.dat"
   This is the list of NewJAN for each product, and the records in this file are as follows:
   - {JANcode_GENcode, NewJAN}.

   * If "JANcode_GENcode" is -1, there is no "JANcode_GENcode" corresponding to "NewJAN."

[2] Data related to products that could not be identified with their respective product categories from the product master are recorded in "unknown.dat."

1.3.3 NikkeiPOS_3digitCategoryLevel_sortedbyNewJAN.c

We construct the set of scanner data for each 3-digit product category. The records in each file are sorted by using the key "NewJAN."

Input files

1. "list_3digit_2008_2019.txt"
2. Scanner data for each 3-digit product category. The record in each file was explained in the previous subsection.

Output files

1. Scanner data for each 3-digit product category. (e.g., scanner data for the product category 1 are recorded in "1.dat")
   - {X, Store, NewJAN, JAN code, GEN code, 6-digit product category, Year, Month, Day, Sales in yen, Quantity}.
   * Dummy data, {0,0,-1,-1,0,0,0,0,0,0,0}, are inserted at the end of each file.

1.4 Shoku-map data

The Shoku-map data we use are household-side data that provide information on the pattern of purchases and consumption, or inventory for each product and household. The data cover approximately 400 households in each month (approximately 4,000 households in total) and the observation period is from September 1998 to February 2019. Products recorded in this dataset are food only and there is no information on prices. Each product is distinguished by the JAN code and product name, similar to the Nikkei scanner data; however, there is no information on GEN codes in the Shoku-map data. In our analysis, we use the Nikkei product master to identify which Nikkei categories each product recorded in Shoku-map belongs to by assuming that the GEN code of each product is uniformly 1. [3]

This dataset has two types of data files called "ZAIKO" and "TABLE." The "ZAIKO" data (which translates as "inventory") record the date of purchase, and the first and last dates of consumption (i.e., when a product is used up or has gone off, etc.) for each product and household on a daily basis. The data also provide the total number of times each product was used (e.g., if the same product is consumed on three separate days, the total number of times recorded would be 3).

The "TABLE" data provide more detailed information about household characteristics (e.g., number of household members, age of respondent, and income-group) and their consumption behaviors (e.g., when and who consumed the product). The main difference from the "ZAIKO" data in observing consumption patterns is that when a product is used on three separate times/days, the "ZAIKO" data record only the total number of times used, while the "TABLE" data record all the events when the good was used.

1.4.1 Data availability statements

The Shoku-map data is proprietary; please contact Lifescape Marketing Co.(lifescape-m.co.jp) if you would like to use the data.

1.4.2 Record/Structure

- The file for the "ZAIKO" data is in CSV format, and its records are as follows:

[3] JAN code can be reused between products that fall into completely different categories, thus introducing a potential bias in the inventory periods calculated under this assumption. In fact, JAN codes observed in the Shoku-map data are classified as daily necessities in some cases, even though the recorded data are only food. However, we have not taken any specific measures to mitigate such circumstances.

- {Monitor ID, X, Date of purchase, X, X, X, X, JAN code, X, Date of first use, Date of last use, Number of times used, Inventory flag, Final state}.

  Monitor ID : Household ID.

  Date of purchase : YYYY-MM-DD.

  Date of first use : YYYY-MM-DD. First date the product was used.

  Date of last use : YYYY-MM-DD. Last date the product was used.

  Number of times used : Total number of times the product has been used.

  Inventory flag : 1. Regular inventory, 2. Initial inventory that was already in stock when the household became the monitor/respondent.

  Final state : Product usage at the end (e.g., in use, used up (code 105), out of sample period, discarded before/after starting to use, gave it before use, and unknown).

- The file for the "TABLE" data is in CSV format, and its records are as follows:

  - {Monitor ID, X, X, X, X, X, Age of respondent, X, X, X, Date of use, X, X, X, X, X, X, X, X, X, X, X, X, X, X, X, JAN code, X, Date of purchase, X, X}.

    Monitor ID : Household ID.

    Age of respondent : Age of homemaker in the household.

    Date of use : YYYY-MM-DD. Date when the product was used.

    JAN code : Japanese Article Number for the purchased product.

    Date of purchase : YYYY-MM-DD.

1.5 Consumer Price Index data

We use the group-/item-level price indices for the 2015-Base Consumer Price Index provided by the Statistics Bureau of Japan.

1.5.1 Data availability statements

You can download the data from the following portal.

`https://www.e-stat.go.jp/en/stat-search/files?page=1&layout=datalist&toukei=00200573&tstat=000001084976&cycle=0&tclass1=000001085995&tclass2=000001085936&tclass3=000001085996&tclass4=000001085976&tclass5val=0`
We have downloaded the data on October 18, 2022.

1.5.2 Record/Structure

- The CPI data ("zmi2015a.csv") include the following informations.

  類・品目 (Group/Item) : Group/item name described in both Japanese and English.

  類・品目符号 (Group/Item code) : Group/item code expressed as a 1-, 2-, or 4-digit number.[4]

  ウェイト (Weight) : Expenditure weight for each group/item.

  ウェイト 1 万分比 (Weight per 10000) : Expenditure weight for each group/item when the total weight is normalized as 10000.

  Date : Date of month represented by YYYYMM.

  Price : The price index for each group/item.

[4] For example, code 1000 (Non-glutinous rice) is composed of codes 1001 (Non-glutinous rice-A) and 1002 (Non-glutinous rice-B), and is one of the components of code 4 (Rice). Further, code 4 is one of the components of code 3 (Cereals), which is one of the components of code 2 (Food).

1.5.3 Price indices used

We compare the price indices calculated from two distinct datasets: our scanner data and the Consumer Price Index (CPI) data. These datasets are characterized by different components of goods. The CPI data encompass a wide range of goods and services purchased by households, while our scanner data are limited to processed foods and daily necessities.

Then, we manually list the comparable categories/items present in both our scanner (218 categories at the 3-digit level) and CPI data (585 items in the most detailed classification). In the file named "ListofItemUsed_zmi2015.txt," we list 58 CPI groups/items matched against the product categories in our scanner data.

We calculate the following "CPI groceries," which is an index of only those CPI groups/items that can be matched to our scanner data:

$$P_{t'}^{groceries} = \frac{\sum_{j|p_{jt'} \text{ is available}} w_j p_{jt'}}{\sum_{j|p_{jt'} \text{ is available}} w_j},$$

where w_j is the expenditure weight for group/item j, and $p_{jt'}$ represents the monthly price index for group/item j on month t'. The effect of consumption tax increases are excluded by dividing price indices by 1.03, 1.05, or 1.08 in the month after the tax rate increased to 3%, 5%, or 8%, respectively.

1.5.4 CPIGroceries.c

Input file

- "zmi2015a.csv"
- "ListofItemUsed_zmi2015.txt"
  - {Group/Item code}

Output file

- "cpi_groceries.txt"
  - {MM/1/YYYY, $P_{t'}^{groceries}/P_{t'-12}^{groceries} - 1$, share}

  share : Percentage of expenditure on goods we used relative to the total expenditure.

2 Estimate the degree of stockpiling, consumption prices, quantities consumed, and other variables associated with sales

In this document, to simplify the notation for each variable, we refer to the combination of product and retailer as "product."

2.1 Overview

We take the following steps to estimate the elasticity of subsitution, and infer the timeseries of consumption prices and quantities for each 3-digit product category. Our estimation strategy is comprised of two principal phases. First, the elasticity of substitution is estimated from the time series of prices and quantities purchased, as well as the estimated normal prices. Second, the time series of consumption prices and quantities are estimated by using the estimated elasticity.

Steps in the first phase $1 \to 3 \to 4 \to 5$.

Steps in the second phase $1 \to 2 \to 3 \to 4 \to 6$.

1. Collect the timeseries of daily prices and quantities purchased for product i, which are represented by $\{p_{it}\}_t$ and $\{q_{it}\}_t$, respectively.
2. Remove outlier prices.
3. Estimate the timeseries of daily regular prices $\{\overline{p}_{it}\}_t$ from $\{p_{it}\}_t$.
4. Impute non-recorded prices.
5. Estimate the elasticity of substitution for the product category j by using the timeseries $\{p_{it}\}_t$, $\{q_{it}\}_t$, and $\{\overline{p}_{it}\}_t$ for all products in the corresponding category j.
6. Estimate the degree of stockpiling m for each sale period of product i, and infer the timeseries of consumption prices $\{r_{it}\}_t$ and quantities $\{c_{it}\}_t$.

2.2 Collect the timeseries of daily prices and quantities purchased

We denote daily sales, the quantity purchased, and price for product i and date t as s_{it}, q_{it}, and $p_{it} := \lfloor s_{it}/q_{it} \rfloor$, respectively. We temporarily impute prices as zero if the quantity purchased is not recorded (i.e., zero).

2.3 Remove outlier prices

We remove outlier prices as follows.

1. If the number of days with $q_{it} > 0$ is less than seven days for a certain product i, we exclude this product from our study.
2. We calculate the median price for product i, which is denoted by $p_i^{med} = median(\{p_{it} | \text{corresponding } q_{it} > 0\}_t)$.
3. We reset the price and quantity at $p_{it} = 0$ and $q_{it} = 0$, respectively, if $|\log(p_{it}/p_i^{med})| > \theta_{dp}$, where we set θ_{dp} at $\ln(7)$.

2.4 Estimate the timeseries of daily regular prices

We estimate the timeseries of daily regular prices for each product by applying the sales filter, which is based on the V-shape filter by Nakamura and Steinsson (2008). The algorithm consists of the following six steps (steps 1 to 6), with step 1 having the highest priority and step 6 the lowest. Exceptionally, in the first time period of $\{p_{it}\}_t$, steps 1 to 3 are not applicable, and thus, step 4 is put the highest priority.

1. If $|p_{it} - \overline{p}_{it-1}| \leq 1$, then $\overline{p}_{it} = \overline{p}_{it-1}$
2. If $p_{it} > \overline{p}_{it-1} + 1$, then $\overline{p}_{it} = p_{it}$
3. If $\overline{p}_{it-1} \in \{p_{it+1}, ..., p_{it+J}\}$ and the price never rises above $\overline{p}_{it-1}$ before returning to $\overline{p}_{it-1}$, then $\overline{p}_{it} = \overline{p}_{it-1}$.
   In our study, if there is a first time period $n \in [1, J]$ that satisfies $|\overline{p}_{it-1} - p_{it+n}| \leq 1$ and $(p_{it+n'} - \overline{p}_{it-1}) < -1$ for all $n' \in [1, n-1]$, then $\overline{p}_{it} = \overline{p}_{it-1}$.
4. If the set $\{p_{it}, p_{it+1}, ..., p_{it+L}\}$ has K or more different elements, then $\overline{p}_{it} = p_{it}$.
5. Define $p_{max} = \max(\{p_{it}, p_{it+1}, ..., p_{it+L}\})$ and t_{max} =first-time $\max(\{p_{it}, p_{it+1}, ..., p_{it+L}\})$. If $p_{max} \in \{p_{it_{max}+1}, ..., p_{it_{max}+L}\}$, then $\overline{p}_{it} = p_{max}$.
   In our study, if there is a time period $n \in [1, L]$ exists that satisfies $|p_{max} - p_{it_{max}+n}| \leq 1$, then $\overline{p}_{it} = p_{max}$.
6. $\overline{p}_{it} = p_{it}$.

We set $L = K = J = 42$ days.

2.5 Impute non-recorded prices

1. If $\overline{p}_{it}$ is zero, impute it as $\overline{p}_{it} = \overline{p}_{it-1}$.
2. If p_{it} is zero, impute it as $p_{it} = \overline{p}_{it}$.

2.6 Find first and last periods for each product

We find the following time periods for each product.

- t'_{first}: First period that satisfies $q_{it} > 0$.
- t'_{last}: Last period that satisfies $q_{it} > 0$.
- t_{first}: First period to use in our analysis.
  - To estimate the elasticity of substitution and consumption variables, we need to use X_H values greater than zero. If X_H is equal to zero at the sale event of our focus, a non-zero value of X_H observed in the most recent past is utilized. However, if X_H is zero at the first sales event of a product, we cannot utilize the most recent non-zero X_H. Therefore, we search for the earliest sale event that satisfies $X_H > 0$ and move back in time as far as possible, while maintaining $p_{it} = \overline{p}_{it}$, to determine the first time period $t_{first} (\geq t'_{first})$ that we can use in our analysis.
- t_{last}: Last period to use in our analysis.
  - We need to calculate X^1_L and X^2_L to estimate the elasticity of substitution and consumption variables; however, these variables cannot be calculated if a bargain sale continues in the last period(t'_{last}). Then, we find the last period $t_{last} (\leq t'_{last})$ that satisfies $|\overline{p}_{it_{last}} - p_{it_{last}}| \leq 2$.

In the following sections, we use only periods from t_{first} to t_{last} for each product.

2.7 Estimate the elasticity of substitution for each 3-digit product category j

We calculate two kinds of elasticity of substitution for each 3-digit product cateogory j.

2.7.1 Variables used

- The timeseries of price and quantity purchased for each product: $\{p_{it}\}_t$ and $\{q_{it}\}_t$ for all $i \in I_j$, where I_j represents the set of products in the 3-digit product category j.
- The timeseries of estimated normal price for each product: $\{\overline{p}_{it}\}_t$ for all $i \in I_j$.

2.7.2 Procedure

We take the following steps for each sale event and product to estimate the elasticity of substitution Γ.

1. Find the time period that satisfy $|\overline{p}_{it} - p_{it}| \leq 2$ and $|\overline{p}_{it+1} - p_{it+1}| > 2$.
2. Find the smallest integer T that is greater than or equal to one, and satisfies $|\overline{p}_{it+T+1} - p_{it+T+1}| \leq 2$.
3. Set $P_H = p_{it}$ and $X_H = q_{it}$ if $q_{it} > 0$. Otherwise, $P_H = P_H^{last}$ and $X_H = X_H^{last}$. Then, we set $P_H^{last} = P_H$ an $X_H^{last} = X_H$.
4. Calculate P_L and X_L only when a sale lasts more than one day ($T > 1$).
   - Calculate the average prices and quantities purchased in the first and second half of a sale.

$$\begin{aligned} P_L^1 &= \frac{\sum_{s=1}^{\lfloor T/2 \rfloor} p_{it+s}}{\lfloor T/2 \rfloor}, & (1)\\ P_L^2 &= \frac{\sum_{s=1}^{T-\lfloor T/2 \rfloor} p_{it+\lfloor T/2 \rfloor+s}}{T-\lfloor T/2 \rfloor}, & (2)\\ X_L^1 &= \frac{\sum_{s=1}^{\lfloor T/2 \rfloor} q_{it+s}}{\lfloor T/2 \rfloor}, & (3)\\ X_L^2 &= \frac{\sum_{s=1}^{T-\lfloor T/2 \rfloor} q_{it+\lfloor T/2 \rfloor+s}}{T-\lfloor T/2 \rfloor}. & (4) \end{aligned}$$

   If $X_L^1 \geq X_L^2$, we set $P_L = P_L^2$ and $X_L = X_L^2$. Otherwise, we set $P_L = P_L^1$ and $X_L = X_L^1$.
   - Calculate Γ if $P_L \neq P_H$:

$$\Gamma = -\frac{\log(X_L/X_H)}{\log(P_L/P_H)}. \tag{5}$$

   - Go back to step 1.

Further, we take the following steps to estimate the elasticity of substition σ^{simple}.

1. Find the time periods that satisfy $|p_{it+1} - p_{it}| > 2$.
2. Calculate σ^{simple} if $p_{it} \neq p_{it+1}$, $q_{it} > 0$, and $q_{it+1} > 0$.

$$\sigma^{simple} = -\frac{\log(q_{it+1}/q_{it})}{\log(p_{it+1}/p_{it})}. \tag{6}$$

2.7.3 Output variables

Finally, we calculate the unweighted average of Γ and σ^{simple} across sale events and products for each 3-digit product category, respectively.

2.7.4 Associated files

Program file

- EstimateConsumptionVariablesandPriceIndices.c

Output files

- "Elasticity.dat"
  - $\{\Gamma, \sigma^{simple}, j\}$
- "C.dat" in the folder named "PH"
  - $\{p_{it}, p_{it+T+1}, q_{it}, q_{it+T+1}$,X,X,X,X,X,X,X,X,X,$y$,X$\}$

  C : 3-digit product category code.

  * Records only if a sale begins at period $t+1$.
- "C.dat" in the folder named "PL"
  - $\{$X,X,X,X,$P_L^1, P_L^2, X_L^1, X_L^2$,X,X,X,X,X,X,$y$,X$\}$

  * Records only if a sale begins at period $t+1$ and lasts more than one day ($T > 1$).

In the following sections, we represent Γ as σ for simplicity and use only 3-digit product categories with $\Gamma > 0$.

2.8 Estimate degree of stockpiling (m^{cont}), consumption price (r), quantity consumed (c), and variables associated with sales (δ) for each 3-digit product category j

2.8.1 Variables used

- The timeseries of price and quantity purchased for each product: $\{p_{it}\}_t$ and $\{q_{it}\}_t$ for all $i \in I_j$.
- The timeseries of estimated normal price for each product: $\{\overline{p}_{it}\}_t$ for all $i \in I_j$.
- The elasticity of substitution for the 3-digit product category j: σ.

2.8.2 Procedure

We take the following steps for each product i within the product category j.

1. If $t > t_{last}$, then finish the process.
2. If $|\overline{p}_{it} - p_{it}| \leq 2$, then $r_{it} = p_{it}$ and $c_{it} = q_{it}$.
   (a) Reset t to $t+1$ and go back to step 1.
3. If $|\overline{p}_{it} - p_{it}| > 2$, then find the smallest $T \geq 1$ that satisfies $|\overline{p}_{it+T} - p_{it+T}| \leq 2$.
   (a) If $p_{it-1} = r_{it-1}$, then $P_H = p_{it-1}$ and $X_H = q_{it-1}$. If $q_{it-1} = 0$, then $X_H = X_H^{last}(P_H/P_H^{last})^{-\sigma}$. Then, we set $P_H^{last} = P_H$ and $X_H^{last} = X_H$.
   (b) If $p_{it-1} \neq r_{it-1}$, then $P_H = P_H^{last}$ and $X_H = X_H^{last}$.

(c) If $T = 1$

- $r_{it} = p_{it}$, $c_{it} = c_L^*$ =max$(0.01, X_H(p_{it}/P_H)^{-\sigma})$.
- $I_L = \max(0, q_{it} - c_L^*)$.

(d) If $T > 1$

- If $X_L^1 \geq X_L^2$, then $P_L = P_L^2$ and $X_L = X_L^2$. Otherwise, $P_L = P_L^1$ and $X_L = X_L^1$.
- $r_{it+t'} = P_L$ and $c_{it+t'} = c_L^* = \max(0.01, X_H(P_L/P_H)^{-\sigma})$ for $0 \leq t' < T$.
- $I_L = \max\left(0, \sum_{t'=0}^{T-1} q_{it+t'} - Tc_L^*\right)$.

(e) If $I_L = 0$

- $r_{it+T} = P_H$ and $c_{it+T} = X_H$.
- Reset t to $t + T + 1$ and go back to step 1.

(f) If $I_L > 0$

- $m^{cont} = \frac{P_H - P_L}{P_L} \frac{\sigma - 1}{1 - (P_H/P_L)^{-\sigma+1}} \frac{I_L}{c_L^*}$, $m = \lceil m^{cont} \rceil$.
- $\Delta I_m = \sum_{n=1}^{m} c_{it+T-1+n} - I_L$, where $c_{it+T-1+n} = (r_{it+T-1+n}/P_L)^{-\sigma} c_L^*$ and $r_{it+T-1+n} = \frac{n}{m}(P_H - P_L) + P_L$.
- If $\Delta I_m > 0$
  Find the smallest $m^*(\geq m)$ that satsifies $\Delta I_{m^*} \geq 0$ and $\Delta I_{m^*+1} < 0$.
- If $\Delta I_m < 0$
  Find the largest $0 \leq m^* < m$ that satsifies $\Delta I_{m^*} \geq 0$ and $\Delta I_{m^*+1} < 0$.
- Find the smallest $l \geq 1$ that satisfies $|p_{it+T+l} - \overline{p_{it+T+l}}| > 2$.
- $r_{it+T-1+n} = \frac{n}{m^*}(P_H - P_L) + P_L$ and $c_{it+T-1+n} = (r_{it+T-1+n}/P_L)^{-\sigma} c_L^*$ for $1 \leq n \leq$min(m^*, l). If $m^* = 0$, then we set $r_{it+T} = P_H$ and $c_{it+T} = Q_H$.
- Reset t to $t + T+$min(m^*, l) and go back to step 1.

2.8.3 Output variables

We calculate the unweighted average of $\log m^{cont}$ and $\log \delta = \log(P_L/P_H)$ across sale events and products within the 3-digit product cateogory j:

$$\log m_{jt}^{cont} = \frac{\sum_{i \in I_j} \log m_{it}^{cont} \cdot 1_{m_{it}^{cont} > 0}}{\sum_{i \in I_j} 1_{m_{it}^{cont} > 0}}, \tag{7}$$

$$=: \frac{m_{jt}^{cont,1}}{m_{jt}^{cont,2}}, \tag{8}$$

$$\log \delta_{jt} = \frac{\sum_{i \in I_j} \log \delta_{it} \cdot 1_{m_{it}^{cont} > 0}}{\sum_{i \in I_j} 1_{m_{it}^{cont} > 0}}, \tag{9}$$

$$=: \frac{\delta_{jt}^1}{\delta_{jt}^2}, \tag{10}$$

where the indicator function $1_{m_{it} > 0}$ equals one if m^{cont} is calculated on date t, i.e., the last day of a sale event. Otherwise, if a sale does not end on date t or the household does not hold a positive inventory, $1_{m_{it} > 0}$ equals zero.

2.8.4 Associated files

Program file

- EstimateConsumptionVariablesandPriceIndices.c

Output files

- "m_forhistogram_catC.dat"
  - $\{m^*\}$

  C : 3-digit product category code.

  * Output only for C=1 (tofu products) and C=137 (instant cup noodle products).

- "Freq.dat"
  - $\{\overline{q}^1_{jt}, \overline{q}^2_{jt}, \underline{q}^1_{jt}, \underline{q}^2_{jt}, m^{cont,1}_{jt}, m^{cont,2}_{jt}, \delta^1_{jt}$, X, δ^2_{jt}, X, X, $\pi^{T,2}_{jt}, \pi^{T,1}_{jt}, S^{T,2}_{jt}, S^{T,1}_{jt}, t, \sigma, j\}$

  * See also Section 2.9.

2.8.5 Notes

- We do not update X_H if the consumption price deviates from the purchase price just before a sale begins. In this case, if we update X_H to a recent quantity purchased q, there is no longer a guarantee that the order r index will return to its original level.
- Inventory levels are always reset to zero when a sale begins.
- The degree of stockpiling m^{cont} is derived from the continuous-time equation, while the unit of time in our analysis is discrete. We first set $m = \lceil m^{cont} \rceil$ and search for the maximum natural number $m = m^*$ (≥ 0) satisfying $\sum_{n=1}^{m^*} c_{it+T-1+n} - I_L \geq 0$ and $\sum_{n=1}^{m^*+1} c_{it+T-1+n} - I_L < 0$.
- If $P_H = P_L$, then we use $m = \left\lfloor \frac{I_L}{c^*_L} \right\rfloor$.
- If $m = 0$, consumption price and quantity just after a sale ends are set to P_H and X_H, respectively.

2.9 Estimate variables associated with sale ($\overline{q}$, $\underline{q}$, and fr) for each 3-digit product category j

2.9.1 Variables used

- The timeseries of price purchased for each product: $\{p_{it}\}_t$ for all $i \in I_j$.
- The timeseries of estimated normal price for each product: $\{\overline{p}_{it}\}_t$ for all $i \in I_j$.

2.9.2 Output variables

We calculate $\overline{q}$, $\underline{q}$, and fr for each period t and 3-digit product category j:

$$\overline{q}_{jt} = \frac{\sum_{i \in I_{jt-1} \cap I_{jt}} 1_{\Delta p_{it} > 2 | \Delta p_{it-1} \leq 2}}{\sum_{i \in I_{jt-1} \cap I_{jt}} 1_{\Delta p_{it-1} \leq 2}},, \tag{11}$$

$$=: \frac{\overline{q}^1_{jt}}{\overline{q}^2_{jt}}, \tag{12}$$

$$\underline{q}_{jt} = \frac{\sum_{i \in I_{jt-1} \cap I_{jt}} 1_{\Delta p_{it} > 2 | \Delta p_{it-1} > 2}}{\sum_{i \in I_{jt-1} \cap I_{jt}} 1_{\Delta p_{it-1} > 2}}, \tag{13}$$

$$=: \frac{\underline{q}^1_{jt}}{\underline{q}^2_{jt}}, \tag{14}$$

$$fr_{jt} = \frac{\sum_{i \in I_{jt-1} \cap I_{jt}} 1_{\Delta p_{it} > 2}}{\sum_{i \in I_{jt-1} \cap I_{jt}} 1}, \tag{15}$$

$$=: \frac{fr^1_{jt}}{fr^2_{jt}}, \tag{16}$$

where I_{jt} represents the set of products on period t for the product category j and $\Delta p_{it} = |\overline{p}_{it} - p_{it}|$. 1_S is the indicator function, such that $1_S = 1$ if statement S is true, and $1_S = 0$ otherwise.

2.9.3 Associated files

Program file

- EstimateConsumptionVariablesandPriceIndices.c

Output files

- "Freq_P.dat"
  - $\{fr^1_{jt}, fr^2_{jt}, t, j\}$
- "Freq.dat"
  - $\{\overline{q}^1_{jt}, \overline{q}^2_{jt}, \underline{q}^1_{jt}, \underline{q}^2_{jt}, m^{cont,1}_{jt}, m^{cont,2}_{jt}, \delta^1_{jt}$, X, δ^2_{jt}, X, X, $\pi^{T,2}_{jt}, \pi^{T,1}_{jt}, S^{T,2}_{jt}, S^{T,1}_{jt}, t, \sigma, j\}$

3 Calculate the changes in the price indices for each 3-digit product category j

3.1 Calculate the changes in the bilateral price indices

3.1.1 Variables used

- The timeseries of price and quantity purchased for each product: $\{p_{it}\}_t$ and $\{q_{it}\}_t$ for all $i \in I_j$.
- The timeseries of price and quantity consumed for each product: $\{r_{it}\}_t$ and $\{c_{it}\}_t$ for all $i \in I_j$.
- The elasticity of substitution for the 3-digit product category j: σ.

3.1.2 Output variables

We calculate the changes in the price indices from $t - dt$ to t for each 3-digit product category j and index formula $X(= L, P, T, CT)$, π^X_{jt}, as follows:

$$\pi^L_{j,t-dt,t} = \frac{\sum_{i \in I_{jt-dt} \cap I_{jt}} p_{it-dt} q_{it-dt} \log\left(\frac{p_{it}}{p_{it-dt}}\right)}{\sum_{i \in I_{jt-dt} \cap I_{jt}} p_{it-dt} q_{it-dt}}, \tag{17}$$

$$\pi^P_{j,t-dt,t} = \frac{\sum_{i \in I_{jt-dt} \cap I_{jt}} p_{it} q_{it} \log\left(\frac{p_{it}}{p_{it-dt}}\right)}{\sum_{i \in I_{jt-dt} \cap I_{jt}} p_{it} q_{it}}, \tag{18}$$

$$\pi^T_{j,t-dt,t} = \frac{1}{2}\left\{\frac{\sum_{i \in I_{jt-dt} \cap I_{jt}} p_{it-dt} q_{it-dt} \log\left(\frac{p_{it}}{p_{it-dt}}\right)}{\sum_{i \in I_{jt-dt} \cap I_{jt}} p_{it-dt} q_{it-dt}} + \frac{\sum_{i \in I_{jt-dt} \cap I_{jt}} p_{it} q_{it} \log\left(\frac{p_{it}}{p_{it-dt}}\right)}{\sum_{i \in I_{jt-dt} \cap I_{jt}} p_{it} q_{it}}\right\}, \tag{19}$$

$$\pi^{CT}_{j,t-dt,t} = \frac{1}{2}\left\{\frac{\sum_{i \in I_{jt-dt} \cap I_{jt}} r_{it-dt} c_{it-dt} \log\left(\frac{r_{it}}{r_{it-dt}}\right)}{\sum_{i \in I_{jt-dt} \cap I_{jt}} r_{it-dt} c_{it-dt}} + \frac{\sum_{i \in I_{jt-dt} \cap I_{jt}} r_{it} c_{it} \log\left(\frac{r_{it}}{r_{it-dt}}\right)}{\sum_{i \in I_{jt-dt} \cap I_{jt}} r_{it} c_{it}}\right\}, \tag{20}$$

We also use the following order r superlartive index $\pi^S_{j,t-dt,t}$ to calculate the cost-of-living index:

$$\pi^S_{j,t-dt,t} = \frac{1}{1-\sigma}\left\{\log\left(\frac{\sum_{i\in I_{jt-dt}\cap I_{jt}} r_{it-dt}c_{it-dt}\left(\frac{r_{it}}{r_{it-dt}}\right)^{\frac{1-\sigma}{2}}}{\sum_{i\in I_{jt-dt}\cap I_{jt}} r_{it-dt}c_{it-dt}}\right) - \log\left(\frac{\sum_{i\in I_{jt-dt}\cap I_{jt}} r_{it}c_{it}\left(\frac{r_{it-dt}}{r_{it}}\right)^{\frac{1-\sigma}{2}}}{\sum_{i\in I_{jt-dt}\cap I_{jt}} r_{it}c_{it}}\right)\right\}. \tag{21}$$

3.1.3 Associated files

Program file

- EstimateConsumptionVariablesandPriceIndices.c

We represent equations (17) to (21) for $dt = 1$ as follows:

$$\begin{aligned}
\pi^L_{j,t-1,t} &= \frac{x^L_{jt}}{S^L_{jt}}, \\
\pi^P_{j,t-1,t} &= \frac{x^P_{jt}}{S^P_{jt}}, \\
\pi^T_{j,t-1,t} &= \frac{1}{2}\left(\frac{x^{T,1}_{jt}}{S^{T,1}_{jt}} + \frac{x^{T,2}_{jt}}{S^{T,2}_{jt}}\right), \\
\pi^{CT}_{j,t-1,t} &= \frac{1}{2}\left(\frac{x^{CT,1}_{jt}}{S^{CT,1}_{jt}} + \frac{x^{CT,2}_{jt}}{S^{CT,2}_{jt}}\right), \\
\pi^S_{j,t-1,t} &= \frac{1}{1-\sigma}\left\{\log\left(\frac{x^{S,1}_{jt}}{S^{S,1}_{jt}}\right) - \log\left(\frac{x^{S,2}_{jt}}{S^{S,2}_{jt}}\right)\right\}.
\end{aligned}$$

Output files

- "PriceIndices_PurchaseBase.dat"
  - {X, X, X, X, X, X, X, X, $x^{T,2}_{jt}$, $x^{T,1}_{jt}$, $S^{T,2}_{jt}$, $S^{T,1}_{jt}$, x^L_{jt}, S^L_{jt}, x^P_{jt}, S^P_{jt}, $s_{jt'}$, π^{RWGEKS}_{jt}, t,j}
    - The variable π^{RWGEKS}_{jt} is explained in Section 3.2.
    - The variable $s_{jt'}$ represents the sales of product category j for month t' that includes the day t.
- "PriceIndices_ConsumptionBase.dat"
  - {$x^{S,1}_{jt}$, $S^{S,1}_{jt}$, $x^{S,2}_{jt}$, $S^{S,2}_{jt}$, $x^{CT,2}_{jt}$, $x^{CT,1}_{jt}$, $S^{CT,2}_{jt}$, $S^{CT,1}_{jt}$, $s_{jt'}$, σ, t, j}

3.1.4 Notes

- If the total sales for the matched sample in $t - dt$ and t equal zero, we cannot calculate the price inflation on period t; that is, π^X_{jt}. Then, we exclude such periods from our calculations.

3.2 Calculate the RWGEKS-Törnqvist price index for each 3-digit product category j

3.2.1 Variables used

- The timeseries of price and quantity purchased for each product: $\{p_{it}\}_t$ and $\{q_{it}\}_t$ for all $i \in I_j$.

3.2.2 Output variables

The changes in the Rolling Window GEKS(RWGEKS)-Törnqvist price index between $t-1$ and t is calculated by the following:

$$\pi_{jt}^{RWGEKS} = \frac{1}{d} \sum_{\tau=0}^{d-1} \left(\pi_{j,t-1,t-\tau}^{T} - \pi_{j,t-\tau,t}^{T} \right), \tag{22}$$

where d represents the length of the rolling window in days, and π_{j,t_1,t_2}^{T} is the Törnqvist price inflation from t_1 to t_2 for each 3-digit product category j, which is defined by equation (19).

3.2.3 Associated files

Program file

- EstimateConsumptionVariablesandPriceIndices.c

Output files

- "PriceIndices_PurchaseBase.dat"
  - {X, X, X, X, X, X, X, X, $x_{jt}^{T,2}$, $x_{jt}^{T,1}$, $S_{jt}^{T,2}$, $S_{jt}^{T,1}$, x_{jt}^{L}, S_{jt}^{L}, x_{jt}^{P}, S_{jt}^{P}, $s_{jt'}$, π_{jt}^{RWGEKS}, t,j}

3.2.4 Notes

- If the total sales for the matched sample between two periods equals zero, we cannot calculate the price inflation on that period; that is π_{jt}^{X}. Then, we exclude such periods from our calculations.

3.3 Chain drift test for each price index formula X and the 3-digit product category j

3.3.1 Variables used

- The timeseries of price and quantity purchased for each product: $\{p_{it}\}_t$ and $\{q_{it}\}_t$ for all $i \in I_j$.
- The timeseries of price and quantity consumed for each product: $\{r_{it}\}_t$ and $\{c_{it}\}_t$ for all $i \in I_j$.
- The elasticity of substitution for the 3-digit product category j: σ.

3.3.2 Output variables

We represent the annual chain drift $d_{0,\tau,dt}^{X}$ for each year y, product category j, time-inverval dt, and the price index X as follows:

$$d_{jy,dt}^{X} = \sum_{s=1}^{(\tau-1)/dt} \pi_{j,(s-1)dt,sdt}^{X} - \pi_{j,0,\tau-1}^{X}, \tag{23}$$

where τ represents the interval over which chain drift is measured, which we set to 365 days, and π_{j,t_1,t_2}^{X} represents the change in the price index from t_1 to t_2 calculated by the respective index formula $X(=T, CT, S)$ given in equations (19) to (21). The first term on the right-hand side of equation (23) represents the change in the chained price index when it is incrementally chained forward by interval dt from 0 (base period) to $\tau - 1$. We set the base period on January 1st for each year y. The second term represents the change in the price index when we calculate the price change from 0 to $\tau - 1$ all at once.

We also calculate the annual chain drift for the RWGEKS as follows:

$$d_{jy,dt}^{RWGEKS} = \pi_{j,0,d-dt}^{GEKS} + \sum_{s=d/dt}^{(\tau-1)/dt} \pi_{j,(s-1)dt,sdt}^{RWGEKS} - \pi_{j,0,\tau-1}^{RWGEKS}, \tag{24}$$

where π_{j,t_1,t_2}^{RWGEKS} is given by equation (22). The first term on the right-hand side of this equation represents the GEKS index, not the RWGEKS index, defined as follows:

$$\pi_{j,0,d-dt}^{GEKS} = \frac{dt}{d} \sum_{s=0}^{(d-dt)/dt} (\pi_{j,0,s\cdot dt}^{T} - \pi_{j,d-dt,s\cdot dt}^{T}). \tag{25}$$

3.3.3 Procedure

We take the following steps for each product $i \in I_j$ to calculate the chain drift for each index formula X and year y .

1. Set d to 7 or 28 days, and choose the time interval $dt \in \{1, 7, 14, 28, 52, 91, 182, 364\}$.
2. Set the time period on January 1st for each year y as zero (base period).
3. Use only a domain of products that are available for the period from 0 to 364.
4. For $X = T, CT, S$
   - Calculate $\pi_{j,(s-1)dt,sdt}^{X}$ for $1 \leq s \leq (\tau-1)/dt$.
   - Calculate $\pi_{j,0,\tau-1}^{X}$.
5. If $d > dt$
   - Calculate $\pi_{j,0,d-dt}^{GEKS}$.
   - Calculate $\pi_{j,(s-1)dt,sdt}^{RWGEKS}$ for $d/dt \leq s \leq (\tau-1)/dt$.
   - Calculate $\pi_{0,\tau-1}^{RWGEKS}$.
6. Calculate the chain drift for each price index formula X $(= T, CT, S, RWGEKS)$ and year y by applying equations (23) and (24)

3.3.4 Associated files

Program file

- EstimateConsumptionVariablesandPriceIndices.c

Output files

- "ChainDriftTest.dat"
  - $\{d_{jy,dt}^{T}$, X, X, $d_{jy,dt}^{CT}$, $d_{jy,dt}^{S}$, X, $d_{jy,dt}^{RWGEKS}$, s_{jy}, Indicator, y, $j\}$

  Indicator : It is a natural number from 0 to 11, corresponding to one of $dt \in \{1, 2, 4, 7, 13, 14, 26, 28, 52, 91, 182, 364\}$.

3.3.5 Notes

- If the total sales for the matched sample between two periods equal zero, we cannot calculate the price inflation on that period; that is π_{j,t_1,t_2}^{X}. Then, we exclude such periods from our calculations.

4 Aggregation of variables calculated from the scanner data

4.1 Price inflation for each index formula X

4.1.1 Calculate the average Törnqvist inflation for each month and product category

We calculate the averaged daily Törnqvist inflation rate in month t' for each category j as follows:

$$\pi^T_{jt'} = \frac{\sum_{t \in M_{t'}} \pi^T_{jt}}{\sum_{t \in M_{t'}} 1_{\pi^T_{jt} \text{ is calculated}}}, \tag{26}$$

where $M_{t'}$ represents the set of dates in month t'.

4.1.2 Calculate the average Törnqvist inflation for each year and product category

We calculate the averaged daily Törnqvist inflation rate in year y for each product category j as follows:

$$\pi^T_{jy} = \frac{\sum_{t \in Y_y} \pi^T_{jt}}{\sum_{t \in Y_y} 1_{\pi^T_{jt} \text{ is calculated}}}, \tag{27}$$

where Y_y represents the set of dates in year y.

4.1.3 Calculate the aggregate chained price index

When we construct the changes in the aggregate price index for each X $(= L, P, T, CT, S, RWGEKS)$, we aggregate the price inflations across product categories as follows.

$$\pi^X_{t-dt,t} = \sum_j \frac{s_{jt'}}{\sum_j s_{jt'}} \pi^X_{j,t-dt,t}, \tag{28}$$

where $s_{jt'}$ represents the sales for month t' that includes the day for which the aggregation is done; that is $s_{jt'} = \sum_{i \in I_j} \sum_{t \in M_{t'}} p_{it} q_{it}$.

When $dt = 1$, we can construct the chained daily price indices using the cumulative sum of the past price inflations for each X:

$$P^X_t = \exp\left(\sum_{s=1}^{t} \pi^X_{s-1,s} \right). \tag{29}$$

4.2 Variables associated with temporary sales

4.2.1 Calculate the monthly variables associated with sales: m^{cont}, δ, $\overline{q}$, $\underline{q}$, and fr for each product category

Using equations (8), (10), (12), (14), and (16), we calculate the following variables for each month t':

$$\log m^{cont}_{jt'} = \frac{\sum_{t \in M_{t'}} m^{cont,1}_{jt}}{\sum_{t \in M_{t'}} m^{cont,2}_{jt}}, \tag{30}$$

$$\log \delta_{jt'} = \frac{\sum_{t \in M_{t'}} \delta^1_{jt}}{\sum_{t \in M_{t'}} \delta^2_{jt}}, \tag{31}$$

$$\overline{q}_{jt'} = \frac{\sum_{t \in M_{t'}} \overline{q}^1_{jt}}{\sum_{t \in M_{t'}} \overline{q}^2_{jt}}, \tag{32}$$

$$\overline{q}_{jt'} = \frac{\sum_{t \in M_{t'}} \overline{q}^1_{jt}}{\sum_{t \in M_{t'}} \overline{q}^2_{jt}}, \tag{33}$$

$$fr_{jt'} = \frac{\sum_{t \in M_{t'}} fr^1_{jt}}{\sum_{t \in M_{t'}} fr^2_{jt}}. \tag{34}$$

4.2.2 Calculate the annual variables associated with sales: m^{cont}, δ, fr, $\overline{q}$ and $\underline{q}$ for each product category

Using equations (8), (10), (12), (14), and (16), we calculate the following variables for each year y:

$$\log m_{jy}^{cont} = \frac{\sum_{t\in Y_y} m_{jt}^{cont,1}}{\sum_{t\in Y_y} m_{jt}^{cont,2}}, \tag{35}$$

$$\log \delta_{jy} = \frac{\sum_{t\in Y_y} \delta_{jt}^{1}}{\sum_{t\in Y_y} \delta_{jt}^{2}}, \tag{36}$$

$$\overline{q}_{jy} = \frac{\sum_{t\in Y_y} \overline{q}_{jt}^{1}}{\sum_{t\in Y_y} \overline{q}_{jt}^{2}}, \tag{37}$$

$$\overline{q}_{jy} = \frac{\sum_{t\in Y_y} \overline{q}_{jt}^{1}}{\sum_{t\in Y_y} \overline{q}_{jt}^{2}}, \tag{38}$$

$$fr_{jy} = \frac{\sum_{t\in Y_y} fr_{jt}^{1}}{\sum_{t\in Y_y} fr_{jt}^{2}}. \tag{39}$$

4.3 Annual chain drift for each price index formula X

We aggregate the annual chain drifts at the 3-digit product category level, as given in equations (23) and (24), using the annual sales of each product category as weights:

$$d_{y,dt}^{X} = \frac{\sum_j s_{jy} d_{jy,dt}^{X}}{\sum_j s_{jy}}, \tag{40}$$

where $s_{jy} = \sum_{i\in I_j} \sum_{t\in Y_y} p_{it} q_{it}$.

4.4 Two kinds of annual asymmetry in quantity purchased for each product category

We define the following variables for each product i and sale event n: $P_{H,in}^{1}$ and $X_{H,in}^{1}$ denote the price and quantity purchased just before a sale, respectively; $P_{L,in}^{1}$ and $X_{L,in}^{1}$ denote the average price and quantity purchased during the first half of the sale, respectively; $P_{L,in}^{2}$ and $X_{L,in}^{2}$ denote the average price and quantity purchased during the second half of the sale, respectively; and $P_{H,in}^{2}$ and $P_{H,in}^{2}$ denote the price and quantity purchased just after the sale, respectively. While these variables are essentially the same as those defined in Section 2.7, no imputation is made to P_H and X_H even if X_H is zero.

Next, we introduce two kinds of asymmetry in the quantity purchased for each product i and sale event n as follows:

$$Q_{H,in} = \log(X_{H,in}^{2} + \sqrt{1 + (X_{H,in}^{2})^2}) - \log(X_{H,in}^{1} + \sqrt{1 + (X_{H,in}^{1})^2}), \tag{41}$$

$$Q_{L,in} = \log(X_{L,in}^{2} + \sqrt{1 + (X_{L,in}^{2})^2}) - \log(X_{L,in}^{1} + \sqrt{1 + (X_{L,in}^{1})^2}). \tag{42}$$

We then calculate the weighted average of $Q_{H,in}$ and $Q_{L,in}$ over i and n for each 3-digit product category j and year y as follows:

$$Q_{H,jy} = \frac{\sum_{i\in I_j} \sum_{n\in N_{iy}} (P_{H,in}^{1} X_{H,in}^{1} + P_{H,in}^{2} X_{H,in}^{2}) Q_{H,in}}{\sum_{i\in I_j} \sum_{n\in N_{iy}} (P_{H,in}^{1} X_{H,in}^{1} + P_{H,in}^{2} X_{H,in}^{2})}, \tag{43}$$

$$Q_{L,jy} = \frac{\sum_{i\in I_j} \sum_{n\in N_{iy}} (P_{L,in}^{1} X_{L,in}^{1} + P_{L,in}^{2} X_{L,in}^{2}) Q_{L,in}}{\sum_{i\in I_j} \sum_{n\in N_{iy}} (P_{L,in}^{1} X_{L,in}^{1} + P_{L,in}^{2} X_{L,in}^{2})}, \tag{44}$$

where N_{iy} represents the set of sale events of product i started within year y.

4.5 The weighted quartiles of the distribution of annual variables at the product category level

We compute the weighted quartiles of the respective distributions of $v_{j,y} = d_{jy,dt=1}$, $Q_{H,jy}$, and $Q_{L,jy}$ at the 3-digit category level using the annual sales of each category $s_{j,y}$ $(= s_{jy})$ as weights.

We represent the set of pairs of variables $\{(v_{j,y}, s_{j,y})\}_j$ sorted by $v_{j,y}$ in ascending order as $\{(v_{k,y}, s_{k,y})\}_k$ and search for the smallest l that satisfies:

$$p_{y,2} = \frac{\sum_{k=1}^{l} s_{j,k} - s_{l-1,y}/2}{\sum_k s_{j,k}} \geq p, \tag{45}$$

where p equals 0.25, 0.5, or 0.75, representing the first, second, and third quartiles, respectively. The weighted quartile of $v_{k,y}$ for each p is defined as follows:

$$x_{y,p} = v_{l-1,k} + \frac{p - p_{y,1}}{p_{y,2} - p_{y,1}}(v_{l,y} - v_{l-1,y}), \tag{46}$$

where $p_{y,1} = (\sum_{k=1}^{l-1} s_{k,y} - s_{l-1,y}/2)/\sum_k s_{k,y}$.

5 Variables calculated from the Shoku-map data

5.1 Calculate the average duration of inventories and average quantities purchased

We calculate the average duration of inventories and average quantities purchased for each 3-digit product category and month using the Shoku-map "ZAIKO" data.

This dataset is recorded in pairs of three dates: the date of purchase (t_p), and when a household started and finished consuming (t_f and t_l) the product. If multiple units of the same product are purchased by a particular household on the same date, each unit will be recorded as a separate data entry.

Here, we denote the dates when consumption begins and ends for the n-th unit of product i, out of q_{iht} units purchased by the household h on date t, as $t_{f,iht,n}$ and $t_{l,iht,n}$, respectively.

5.1.1 Variables used

We define the duration of inventories for the product i purchased on date $t = t_p$ as follows:

$$l_{iht} = \max_{1 \leq n \leq q_{iht}} (t_{l,iht,n}) - t + 1. \tag{47}$$

5.1.2 Output variables

We calculate the average duration of inventories and average quantities purchased for each product category j on month t' as follows:

$$\begin{aligned}
\log l_{jt'} &= \frac{\sum_{i \in I_j} \sum_h \sum_{t \in M_{t'}} \log l_{iht}}{\sum_{i \in I_j} \sum_h \sum_{t \in M_{t'}} 1_{l_{iht} \text{ is observable}}}, &(48)\\
&=: \frac{l^1_{jt'}}{l^2_{jt'}}, &(49)\\
\log q_{jt'} &= \frac{\sum_{i \in I_j} \sum_h \sum_{t \in M_{t'}} \log q_{iht}}{\sum_{i \in I_j} \sum_h \sum_{t \in M_{t'}} 1_{l_{iht} \text{ is observable}}}, &(50)\\
&=: \frac{q^1_{jt'}}{q^2_{jt'}}. &(51)
\end{aligned}$$

5.1.3 Associated files

Program file

- Shokumap.c

Output files

- "MeanInventory_CatLevel_Monthly.dat"
  - $\{l^1_{jt}, q^1_{jt}, l^2_{jt}, t, j\}$
- "BeerConsumption_ParticularHousehold.dat"
  - $\left\{ \begin{array}{l} \text{X, X, } t,\ t_{f,ih^*t,n},\ N \\ \text{X, X, } t_{f,ih^*t,n},\ t_{l,ih^*t,n},\ N \end{array} \right\}$

  N : Cumulative number of beer products purchased by a particular household h^*.

  * Data blocks.
- "InventoryPeriods_Shokumap.dat"
  - $\{t_p,\ \min_{1\le n\le q_{iht_p}}(t_{f,iht_p,n}),\ \max_{1\le n\le q_{iht_p}}(t_{l,iht_p,n})\}$

5.1.4 Notes

We calculate the variables $\log l_{jt'}$ and $\log q_{jt'}$ using only records that satisfy the following conditions.

- "Iventory flag=1 (Regular inventory)."
  If "Iventory flag=2 (Initial inventory)," we cannot know when a product was purchased and cannot calculate the inventory period.
- "Final state=105 (used up).".
  In other final states, products are discarded or given to others; therefore, we cannot properly calculate the inventory period.
- We only use records that allow us to identify the Nikkei 3-digit product category to which the product belongs.

5.2 Other output variables

5.2.1 Associated files

Program file

- Shokumap.c

Output files

- "ConsumprionPattern_SaltProduct.dat"
  - $\{t^*, 1, t_{t^*,h^*,i^*}\}$

  t_{t^*,h^*,i^*} : The date on which the household h^* consumed product i^* purhcased on date t^*. These observations are recorded in the Shoku-map "TABLE" data.
- "BasicStats_ShokuMap.dat"

  N of households per month : Number of households that made purchases in each month.

  N of products purchased per month : Number of purchase records in each month.

  N of products purchased per month and per household : Number of purchase records divided by number of purchasing houhoselds in each month.

N of months for which a household answered : Number of months for which a household purchases products.

Age of the wife in the household : Age of the respondent. If the age of the respondent has increased during the response period, the highest age among them is recorded.

* Calculate the average of monthly variables for the period from October 1998 to January 2019.

6 Flow

6.1 Overview

We take the following steps to construct our figures and tables.

1. Run "EstimateConsumptionVariablesandPriceIndices.c."
2. Run "Shokumap.c."
3. Run "Aggregation_Index.c."
4. Run "Asymmetry_QHorQL_Annual_3digitCategoryLevel.c."
5. Run "PanelandQuartiles_QH_QL_drift.c."
6. Run "Panel_TimeAverage_monMar2014_Monthly_Jan1989toDec2018.c."
7. Run "TimeAverage_m_fr_DecompTerms_TornPI_Drift_Annual_1989to2018.c."
8. Run "PoSvsShokumap.c."
9. Run "fig_cumulativeprob_m.c."
10. Run in R: Regress fr,$\overline{q}$,$\underline{q}$,$\log(P_L/{}_PH)$ and $\log m$ on d^T:

    Input "TimeAverage_m_fr_q_d_TornPI_Drift_1989to2018.dat"

    R function "lm()"

    Results "R_table2.txt"

11. Run in R: Regress Q_H and Q_L on d^T:

    Input "Average_drift_QH_QL_1989to2018.dat"

    R function "lm()"

    Results "R_table2.txt"

12. Run in R: Calculate the Spearman's correlations between $< \log l_{jt't} >_{t'}$ and $< \log m_{jt'} >_{t'}$, $< \log q_{jt't} >_{t'}$ and $< \log m_{jt'} >_{t'}$, $< \log l_{jt'} >_{t'}$ and $\Delta \log m_{jt'=Mar.2014}$, $\log m_{jt'=Mar.2014}$ and $\log m_{jt'=Mar.2014}$, respectively:

    Input "Averageofmfrom1989to2018_vs_AverageInventory.dat"

    Input "dmonMar2014_vs_AverageofInventoryDuration.dat"

    Input "monMar2014_vs_qonMar2014.dat"

    R function "cor.test(,method="spearman")"

    Results "R_fig7and9.txt"

13. Run " 'load 'fig.plt' " in gnuplot.

6.2 EstimateConsumptionVariablesandPriceIndices.c

6.2.1 Overview

- Estimate the elasticity of substitution for each product category j (see Section 2.7).
- Estimate variables associated with sale ($\overline{q}, \underline{q}, \delta$) and the degree of stockpiling (m^{cont}) for each product category j and period t (see Sections 2.8 and 2.9).
- Calculate the frequency of temporally sale (fr) for each product category j and period t (see Section 2.9).
- Calculate the purchase-weighted price indices for each index formula X, product category j, and period t (see Sections 3.1 and 3.2).
- Calculate the consumption-weighted price indices for each index formula X, product category j, and period t (see Section 3.1).
- Calculate the annual chain drift for each price index formula X, product category j, and year y (see Section 3.3).

6.2.2 Output files

- "Elasticity.dat"
  - $\{\sigma, \sigma^{simple}, j\}$
- "PriceIndices_PurchaseBase.dat"
  - $\{$X, X, X, X, X, X, X, X, $x_{jt}^{T,2}$, $x_{jt}^{T,1}$, $S_{jt}^{T,2}$, $S_{jt}^{T,1}$, x_{jt}^{L}, S_{jt}^{L}, x_{jt}^{P}, S_{jt}^{P}, s_{jm}, π_{jt}^{RWGEKS}, $t, j\}$
- "PriceIndices_ConsumptionBase.dat"
  - $\{x_{jt}^{S,1}, S_{jt}^{S,1}, x_{jt}^{S,2}, S_{jt}^{S,2}, x_{jt}^{CT,2}, x_{jt}^{CT,1}, S_{jt}^{CT,2}, S_{jt}^{CT,1}, s_{jm}, \sigma, t, j\}$
- "ChainDriftTest.dat"
  - $\{d_{jy,dt}^{T}$, X, X, $d_{jy,dt}^{CT}$, $d_{jy,dt}^{S}$, X, $d_{jy,dt}^{RWGEKS}$, s_{jy}, Indicator, y, $j\}$

  Indicator : It is a natural number from 0 to 11 corresponding to one of $dt \in \{1, 2, 4, 7, 13, 14, 26, 28, 52, 91, 182, 364\}$.
- "Freq_P.dat"
  - $\{fr_{jt}^{1}, fr_{jt}^{2}, t, j\}$
- "Freq.dat"
  - $\{\overline{q}_{jt}^{1}, \overline{q}_{jt}^{2}, \underline{q}_{jt}^{1}, \underline{q}_{jt}^{2}, m_{jt}^{cont,1}, m_{jt}^{cont,2}, \delta_{jt}^{1}$, X, δ_{jt}^{2}, X, X, $\pi_{jt}^{T,2}, \pi_{jt}^{T,1}, S_{jt}^{T,2}, S_{jt}^{T,1}, t, \sigma, j\}$
- ⧫ "C.dat" in the folder named "PH"
  - $\{p_{it}, p_{it+T+1}, q_{it}, q_{it+T+1}$,X,X,X,X,X,X,X,X,X,$y$,X$\}$

  C : 3-digit product category code.

  * Records only if a sale begins at period $t+1$.
- ⧫ "C.dat" in the folder named "PL"
  - $\{$X,X,X,X,$P_L^1, P_L^2, X_L^1, X_L^2$,X,X,X,X,X,X,$y$,X$\}$

  * Records only if a sale begins at period $t+1$ and lasts more than one day ($T > 1$).
- ⧫ "Path_PriceandQuantity_ParticularInstantNoodle.dat"

- $\{\overline{p}_{it}, p_{it}, q_{it}, r_{it}, c_{it}, t\}$

- "m_forhistogram_catC.dat"

  - {m}

  C : 3-digit product category code (1 or 137).

  * The degree of stockpiling m is the natural number we actually used to estimate consumption prices and quantities.

6.3 Shokumap.c

6.3.1 Overview

- Calculate average hoarding ammount and its inventory period for each product category j and month t' (see equations (48) to (51) in Section 5).
- Record data on stockpiling behavior of beer products for a particular household.
- Record data on the purchase and consumption pattern of a particular salt product purchased by a particular household on a particular date.
- Calculate the basic statistics for the Shoku-map data.

6.3.2 Output files

- "MeanInventory_CatLevel_Monthly.dat"

  - $\{l^1_{jt'}, q^1_{jt'}, l^2_{jt'}, t', j\}$

- ⧫ "BeerConsumption_ParticularHousehold.dat"

  - $$\left\{ \begin{array}{l} \text{X, X, } t, t_f^{t,h^*,i:n}, N \\ \text{X, X, } t_f^{t,h^*,i:n}, t_l^{t,h^*,i:n}, N \end{array} \right\}$$

  N : Cumulative number of beer products purchased by a particular household.

  * Data blocks.

- ⧫ "ConsumprionPattern_SaltProduct.dat"

  - $\{t^*, 1, t_c^{t^*,h^*,i^*}\}$

  $t_c^{t^*,h^*,i^*}$: The date on which household h^* consumed product i^* purhcased on date t^*.

- "BasicStats_ShokuMap.dat"

  N of households per month : Number of households that made purchases in each month.

  N of products purchased per month : Number of purchase records in each month.

  N of products purchased per month and per household : Number of purchase records divided by number of purchasing houhoselds in each month.

  N of months for which a household answered : Number of months for which a household purchases products.

  Age of the wife in the household : Age of the respondent.

  * Calculate the average of monthly variables for the period from October 1998 to January 2019.

  * We create Table C.1 with the results listed in this file.

6.4 Aggregation_Index.c

6.4.1 Overview

- Aggregate the price inflations and COLIs at the 3-digit product category level (see equations (26) and (27) in Section 4.1).
- Calculate the aggregate price index for each index formula X (see equation (28) in Section 4.1).
- Calculate the average of π_t^X for the period from January 1, 1990, to December 31, 2018, as

  $\pi^X = \sum_t \pi_t^X / \sum_t 1_{\pi_t^X \text{ is calculated}}$.
- Calculate the aggtegate chain drift for each index formula $X = \{T, CT, S, RWGEKS\}$ (see. equation (40) in Section (4.3)).
  1. Calculate the average of $d_{y,dt}^X$ over y.
  2. We apply the binomial test for the null hypothesis that the probabilities that the annual chain drift ($d_{y,dt}^X$) is positive or negative are equal to 0.5. If the number of observations for the positive and negative annual chain drifts are denoted by n^+ and n^-, respectively, then we employ the binomial test for the case of n^+ successes in a sample of size $n^+ + n^-$.

6.4.2 Output files

- "DailyPriceInflation_Mean.dat"

  In this file, we record $100 \times \{\exp(365 \times \pi^X) - 1\}$ and π^X for each X.
- "ChainedPriceIndices.dat"
  - {X, X, P_t^T, P_t^L, P_t^P, X, Month/Day/Year}
- "ChainedPriceInflations_Raw_Imputed_U_OrderR.dat"
  - {X, $\log(P_t^T/P_{t-365}^T)$, $\log(P_t^{CT}/P_{t-365}^{CT})$, $\log(P_t^S/P_{t-365}^S)$, $\log(P_t^{RWGEKS}/P_{t-365}^{RWGEKS})$, Month/Day/Year}
- "ChainDriftTest_Mean_forTable.txt"
  - {dt, $100\{\exp(< d_{y,dt}^T >_y) - 1\}$, $100\{\exp(< d_{y,dt}^{CT} >_y) - 1\}$, $100\{\exp(< d_{y,dt}^S >_y) - 1\}$}

  $< x_{y,dt} >_y$: Average of $x_{y,dt}$ from 1989 to 2018.
  - In this file, ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively, when we apply the binomial test for the null hypothesis that the probabilities that the annual chain drift is positive or negative are equal to 0.5.

  * We create Table 1 with the results listed in this file.
- "ChainDriftTest_Mean_forFig.txt"
  - {$< d_{y,dt}^T >_y$, X, $< d_{y,dt}^L >_y$, X, $< d_{y,dt}^P >_y$, X, X, X, X, X, X, X, dt}

  $< x_{y,dt} >_y$: Average of $x_{y,dt}$ from 1989 to 2018.

6.5 Asymmetry_QHorQL_Annual_3digitCategoryLevel.c

6.5.1 Overview

- Calculate two kinds of asymmetry in quantity purchased for each 3-digit product category (see equations (43) and (44) in Section 4.4).

6.5.2 Output file

- "dQ_mean_annual.dat"

  PH : $\{$X, X, $Q_{H,jy}$, X, y, $j\}$
  PL : $\{$X, X, $Q_{L,jy}$, X, y, $j\}$

6.6 PanelandQuartiles_QH_QL_drift.c

6.6.1 Overview

- Calculate the weighted quartiles of the distribution of $Q_{H,jy}$, $Q_{L,jy}$, and $d^T_{jy,dt=1}$ across product categories for each year (see Section 4.5).

6.6.2 Output files

- "Quartiles_Asymmetry_QH_QL_drift.dat"
  - $\{x_{H,y,0.25}$, $x_{H,y,0.5}$, $x_{H,y,0.75}$, X, $x_{L,y,0.25}$, $x_{L,y,0.5}$, $x_{L,y,0.75}$, X, $x_{d,y,0.25}$, $x_{d,y,0.5}$, $x_{d,y,0.75}$, X, $y\}$
  * The variables $x_{H,y,p}$, $x_{L,y,p}$, and $x_{d,y,p}$ represent the weighted quartiles of $Q_{H,jy}$, $Q_{L,jy}$, and $d_{jy,dt=1}$, respectively.
- "Average_drift_QH_QL_1989to2018.dat"
  - $\{< d^T_{jy,dt=1} >_y, < Q_{H,jy} >_y, < Q_{L,jy} >_y\}$

  $< \cdot >_y$: Average of annual variable from $y = 1989$ to $y = 2018$.
  * In this calculation, we only use product categories whose variables are continuously observable from 1989 to 2018.

6.7 Panel_TimeAverage_monMar2014_Monthly_Jan1989toDec2018.c

6.7.1 Overview

- Calculate monthly variables for each product cateogory which are represented by equations (26), and (30) to (33) in Section 4.
- Calculate $\log m^{cont}_{jt'}$ and $\Delta \log m^{cont}_{jt'} = \log m^{cont}_{jt'} - \log m^{cont}_{jt'-12}$ on t' =March 2014.

6.7.2 Output files

- "m_Mar2014.dat"
  - $\{\log m^{cont}_{jt'=Mar.2014}, j\}$
- "m_YoYDif_Mar2014.dat"
  - $\{\Delta \log m^{cont}_{jt'=Mar.2014}, j\}$
- "TimeMean_DecompTerm_3digit.dat"
  - $\{< \log m^{cont}_{jt'} >_{t'}$, X, X, X, X, $j\}$

  $< x_{jt'} >_{t'}$: Average of $x_{jt'}$ for thev period from January 1989 to December 2018.
- "Panel_decomposedm_catlevel_Jan1989toDec2018.dat"
  - $\{j, t'$ $\log m^{cont}_{jt'}$, $\log(1 - \overline{q}_{jt'})$, $\log(1 - \underline{q}_{jt'})$, $\log(\exp(-\log \delta_{jt'}) - 1)$, $\pi^T_{jt'}\}$
  * Records only product categories whose variables can be continuously observed in the entire period.

6.8 TimeAverage_m_fr_DecompTerms_TornPI_Drift_Annual_1989to2018.c

6.8.1 Overview

- Calculate the annual variables for each product cateogory which are represented by equations (27), and (35) to (39) in Section 4.
- Calculate the averages of these variables for the period from 1989 to 2018.
- Calculate the average chain drift from 1989 to 2018 for $d^T_{jy,dt=1}$ given in equation (23).

6.8.2 Output files

- "TimeAverage_m_fr_q_d_TornPI_Drift_1989to2018.dat"
  - $\{< \log m^{cont}_{jy} >_y, < fr_{jy} >_y, < \overline{q}_{jy} >_y, < \underline{q}_{jy} >_y, < \delta_{jy} >_y, < \pi_{jy} >_y, < d^T_{jy,dt=1} >_y, j\}$

  $< x_{jy} >_y$**:** Average of x_{jy} from 1989 to 2018.

6.9 PoSvsShokumap.c

6.9.1 Overview

- Calculate the average of $\log m^{cont}_{jt'}$ for the period from January 1989 to December 2018.
- Calculate the year-on-year difference in $\log m^{cont}_{jt'}$ on t' =Mar. 2014, a month before tax increases.
- Calculate the averages of $\log l_{jt'}$ and $\log q_{jt'}$ over the observed periods in the Shoku-map data.

6.9.2 Output files

- "monMar2014_vs_qonMar2014.dat"
  - $\{\log m^{cont}_{jt'=Mar.2014}, \log q_{jt'=Mar.2014}, j\}$
- "dmonMar2014_vs_AverageofInventoryDuration.dat"
  - $\{\Delta m^{cont}_{jt'=Mar.2014}, < \log l_{jt'} >_{t'}, j\}$

  $< \cdot >_{t'}$: Average of monthly variable across time periods t'.
- "Averageofmfrom1989to2018_vs_AverageInventory.dat"
  - $\{< \log l_{jt'} >_{t'}, < \log q_{jt'} >_{t'}, < \log m^{cont}_{jt'} >_{Jan.1989 \le t' \le Dec.2018}, j\}$

6.10 fig_cumulativeprob_m.c

6.10.1 Overview

- Create the cumulative probability for the degree of stockpiling ($m > m^*$).

6.10.2 Output files

- Tofu category: "cumulativeprob_m_1.dat"
  - $\{m^*$, cumulative probability $(m > m^*)\}$
- Instant cup noodle category: "cumulativeprob_m_137.dat"
  - $\{m^*$, cumulative probability $(m > m^*)\}$