Japan & The World Economy 65 (2023) 101176

Contents lists available at ScienceDirect

Japan & The World Economy

journal homepage: www.elsevier.com/locate/jwe

Consumer price measurement under the first wave of the COVID-19 spread in Japan: Scanner data evidence for retailers in Tokyo☆

Masahiro Higo [a], Shigenori Shiratsuka [b,*]

[a] Graduate School of Economics, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[b] Faculyt of Economics, Keio University, 2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan

ARTICLE INFO

JEL classification:
C43
E31

Keywords:
Consumer price index
Scanner data
Quality adjustment
Retail service quality
Temporary sales
COVID-19

ABSTRACT

In this paper, we examine the CPI (consumer price index) measurement errors under the first wave of the COVID-19 spread in Japan. To address this question, we construct high-frequency quality-adjusted price indices by employing daily scanner data from retail stores in Tokyo. We demonstrate the importance of using price data with the wide-ranging coverage of products and outlets by making explicit adjustments for temporary sales effects and retail service quality in examining the retail price dynamics under the COVID-19 pandemic as the voluntary lockdown constrained household purchasing behavior. Note that the sources of the CPI measurement errors under the COVID-19 pandemic differ significantly from those in the US, observed as wide-ranging and long-lasting stockouts. We show that downward bias, not upward bias generally advocated, was observed during the first wave of the COVID-19 spread in Japan. The magnitude of the downward bias is estimated at from −0.6 to −0.3 points on the CPI for food less perishables and eating out on the basis of cumulative changes from January 2020 to June. The contribution of the estimates to the overall CPI is −0.3% to −0.15% points on an annualized basis, considering that the estimation covers half-year and the weights are about a quarter of the overall CPI. The magnitude of measurement errors is deemed limited, and the overall trend of the CPI remains unchanged even after incorporating the estimated downward bias. It should be noted that this downward bias arises mainly from the "one-specification-for-one-item" policy by weakening the price representativeness in the Japanese CPI.

1. Introduction

In this paper, we examine the CPI (consumer price index) measurement errors under the first wave of the COVID-19 spread in Japan. To address the question, we construct high-frequency quality-adjusted price indices by employing daily scanner data at retail stores in Tokyo.

During the first wave of the COVID-19 spread in Japan in the first half of 2020, the number of infections began to increase from mid-February and spiked through March. In response, the Government of Japan declared a state of emergency on April 7 in seven prefectures, including Tokyo, and on April 16, the state of emergency was extended to the entire country. Under these circumstances, people refrained from going out. The number of new infections peaked in mid-April and began to decline rapidly. During the first wave of the COVID-19 spread, Japan experienced weaker social and economic constraints under the state of emergency. Unlike the strict lockdown in many countries, Japan implemented a voluntary lockdown, experiencing a significant decline in economic activities.

Many previous studies have used various types of alternative data in Japan to discuss the social and economic issues during the COVID-19 pandemic.[1] For example, Hosono (2021) explored the low infectious cases and the large decline in consumption by extending the SIR-macro-model to incorporate Japan's two key factors of voluntary stay-at-home

☆ We thank Erwin Diewert, Shin'ichi Fukuda, Taisuke Nakata, Tatsuyoshi Okimoto, Mototsugu Shintani, Nao Sudo, Jiro Yoshida, and participants at the workshop on "Government Statistics and Economic Performance" and seminar at the Bank of Japan as well as an anonymous referee for their constructive comments and discussions. INTAGE Inc. provided the daily scanner data used in this paper under the collaborative research project at the Center for Research and Education in Program Evaluation (CREPE), the University of Tokyo. Shiratsuka acknowledges the financial support from Keio University Academic Development Fund. An earlier version of this paper circulated under the title "Was Inflation Observed under the First Wave of the COVID-19 Spread in Japan? Scanner Data Evidence for Retailers in Tokyo".

* Corresponding author.
E-mail addresses: masahiro.higo@e.u-tokyo.ac.jp (M. Higo), shigenori.shiratsuka@keio.jp (S. Shiratsuka).

[1] Japanese Economic Association has set up a webpage for information on academic papers analyzing the impact of COVID-19 (https://covid19.jeaweb.org/scientific.html).

https://doi.org/10.1016/j.japwor.2023.101176
Received 10 May 2022; Received in revised form 31 December 2022; Accepted 20 January 2023
Available online 24 January 2023
0922-1425/

and a request-based lockdown. Using smartphone location data, Watanabe and Yabu (2021) revealed that both the intervention effect of government policy and the information effect of health risks from COVID-19 infection constrain people's behavior in response to the stay-at-home measures. Shibamoto et al. (2022) estimated a dynamic model of the interaction between infection-mobility tradeoff and mobility demand. Fujii and Nakata (2021) pointed out that the tradeoff between output and infection in the short run does not necessarily exist in the long run based on their simulation results with the SIR-macro-model with time-varying parameters.

Some studies have employed various alternative data to analyze the structural changes in household consumption expenditure and retail sales trends. These studies did not only use scanner data for retail stores but also credit card expenditure history and data from household accounting applications. For example, Watanabe and Omori (2020) and Konishi et al. (2021) revealed that household consumption expenditure patterns changed significantly, especially from eating out to home cooking and delivery services. As a result, the retail sales trends also changed drastically, observing that supermarkets enjoyed positive impacts, whereas department stores and restaurant affiliates were negatively affected. However, only a very limited number of studies, including the mask price analysis by Abe et al. (2020), focused on retail price development under the COVID-19 pandemic.[2]

Focusing more on the measurement issues of the consumer price index (CPI) under the COVID-19 pandemic, Diewert and Fox (2020) pointed out the possible downward bias, not upward bias, in the CPI due to the sudden unavailability of goods and services during the pandemic. Cavallo and Kryvtsov (2021) examined the impacts of stockouts on inflation using web-scrapping data. They showed that unexpected product shortages produced significant inflation effects in the US and some advanced economies, but Japan was an outlier in maintaining product variety at the outlets. Cavallo (2020) attempted to quantify the downward bias in the CPI due to changes in consumer expenditure patterns by employing credit and debit transaction data.

In Japan, however, the sources of CPI measurement bias under the COVID-19 pandemic are likely to differ from the drastic changes in consumption baskets. Fig. 1 depicts the CPI less perishables for fixed weight and chain weight bases over time, indicating no significant deviations in the two indices even after 2020 under the COVID-19 pandemic.[3] In addition, the CPI inflation started declining from the beginning of 2020 and continued to decline toward the end of the year. These inflation developments traced by the official CPI generally seem consistent with the household perception of no inflation accelerations during the first half of 2020, including the period for the first wave of the COVID-19 spread, as shown in Fig. 2.

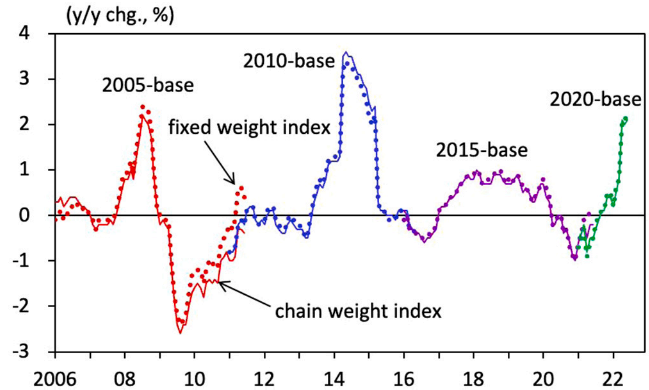

Fig. 1. Consumer price indices: Fixed weight and chain weight basis. Notes: The fixed weight basis corresponds to the headline less perishable based on the fixed weight Laspeyres price index formula. The chain weight basis is computed by the Laspeyres price index formula with annually chained weights. Deviations between the 2015-base and 2020-base indices after April 2021 are attributable to the change in the coverage of cell phone communication charges, which included low-cost cell phone rate structures. *Source:* Statistics Bureau of Japan, Consumer Price Index.

The COVID-19 pandemic, however, may lead to mismeasurement of individual prices due to inadequate adjustments for the temporary sales effects and retail service quality. The same product is sold at different prices at different outlets as observed prices reflect the differences in the temporary sales effects and the retail service quality across outlets. For example, prices at convenience stores are generally higher than those at other outlet channels, reflecting their convenience in terms of operating hours, location, and shopping time. The sudden and unexpected changes in consumer expenditure patterns induced by the COVID-19 spread are likely to produce the measurement bias in the CPI caused by individual prices without making an explicit adjustment for the temporary sales effects and the differences in retail service quality. The Japanese CPI fails to maintain price representativeness under the rapid and significant structural changes in the retail markets due to its price survey method based on the "one-specification-for-one-item" policy. The policy specifies a few popular specifications for each item and continuously surveys their prices at specific outlets.[4]

Fig. 3 depicts the recent development of retail prices using high-frequency scanner price indices for food and beverage products and daily necessities at supermarkets.[5] The T-index aggregates product prices for each outlet at a JAN code level by applying the Törnqvist index formula. The T-7D-MA is a seven-day backward moving average of the T-index. The T-mode-index traces regular price fluctuations by employing mode prices for the daily windows from 28 days before to 28 days after.

The figure depicts that the T-index indicates a moderate upward trend from shortly before the first declaration of a state of emergency in April 2020. The trend then reversed after reaching a peak under the state of emergency, revealing a downward trend toward mid-2021. During this period, two observations are noted. First, the volatility of the T-index declined significantly, suggesting possible declines in the frequency of temporary sales and the size of price reduction. Second,

[2] Retail scanner data research in Japan dates back to the mid-1990s when the Seiyu price index, compiled by one of the major supermarket chains in Japan, highlighted the "price busting" phenomena in Japanese retail markets. In the mid-1990s, Japan experienced a relatively high domestic price level. With continued low inflation under the "lost decades", Japan's price problem turned into a relatively low domestic price level. Scanner data research focuses on stealth inflation with product turnover and downsizing (Imai and Watanabe, 2014; Ueda et al., 2019; Ueda et al., 2019), which tries to raise unit prices by reducing the product volume while keeping tag price unchanged. Moreover, applications of scanner data have continued to expand by linking more broad aspects of macroeconomic issues, such as the frequency of price changes to explore the flattened Phillips curve (Abe and Tonogi, 2010) and the interaction of the frequency of temporary sales with macroeconomic conditions and effectiveness of monetary policy (Sudo et al., 2018).

[3] The Japanese CPI is revised every five years at years ending in zero or five. The revisions are released in July of the following year, considering the availability of the source data for computing weights, the Family Income and Expenditure Survey (FIES) of the base year. As a result, the previous base CPI continues to be released until June, the previous month for the first release of the new base CPI. Thus, the weights for the 2020-base CPI are retroactively revised, reflecting the impacts of the COVID-19 spread. Note that the weights for the 2020-base CPI are computed by using the average of 2019 and 2020, taking into account the impact of the COVID-19.

[4] As Shiratsuka (2021) pointed out, the Japanese CPI has problems with the broadly-defined lower-level substitution bias, which corresponds to the weakened price representativeness stemming from the "one-specification-for-one-item" policy of the price survey method.

[5] NIKKEI CPINow is the real-time daily price index computed from scanner data of supermarkets. It covers food and beverage products and other daily necessities with a fairly long time series from 1989.

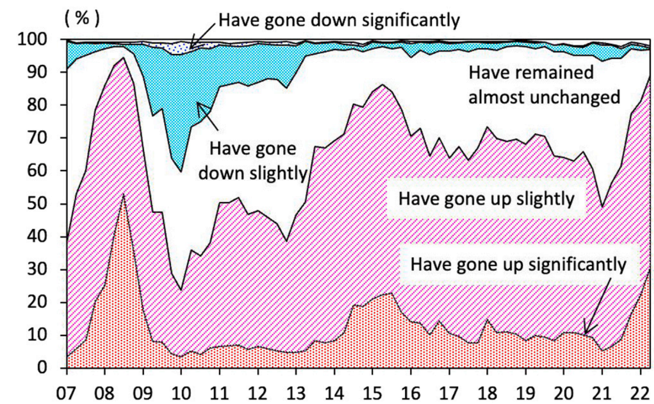

Fig. 2. Inflation perception by household.
Source: Bank of Japan, Opinion Survey on the General Public's Views and Behavior.

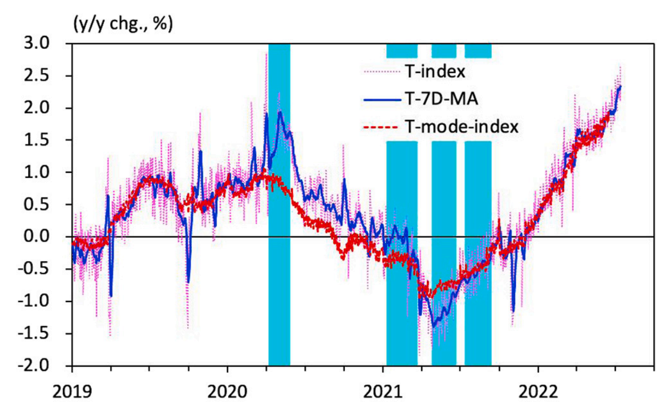

Fig. 3. Daily retail price developments. Notes: The T-index aggregates product prices for each outlet at a JAN code level by applying the Törnqvist index formula. The T-7D-MA is a seven-day backward moving average of the T-index. The T-mode-index employs mode prices for the daily windows from 28 days before to 28 days after. The light blue areas indicate the periods under the declaration of states of emergency in Tokyo—(i) from April 7 to May 25, 2020, (ii) from January 8 to March 21, 2021, (iii) from April 25 to June 20, 2021, and (iv) from July 12 to September 30, 2021. Year-on-year changes are calculated as percentage changes from 364 days earlier to adjust the week-day effects. The plotted data for CPINow are nationwide, as it does not publish the data for Tokyo metropolitan area, though it publishes other prefectural data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Source: Nowcast, NIKKEI CPINow.

the T-mode-index deviated from the upward trend of the T-index significantly, although the T-index and the T-mode-index experienced similar movements outside this period. This observation indicates that increases in the T-index are attributed to decreases in the frequency and the size of price reductions of temporary sales. Moreover, the voluntary lockdown under the state of emergency induced people to minimize shopping frequency, travel distance, and shopping time, influencing household purchasing behavior, such as the choice of retail outlets and price sensitivity of consumption expenditure decisions.

To examine the measurement bias of the CPI under the COVID-19 pandemic, we construct high-frequency quality-adjusted price indices by making explicit adjustments to daily scanner data on product characteristics, temporary sales effects, and retail service quality. As the voluntary lockdown constrained household purchasing behavior, we emphasize the importance of using price data with the wide-ranging coverage of products and outlets by making explicit adjustments for the effects of temporary sales and retail service quality in examining the retail price dynamics under the COVID-19 pandemic. Households responded to the voluntary lockdown by switching shopping sites to reduce shopping frequency, travel distance, and shopping time.

Our approach is closely related to de Haan and Hendriks (2013), who applied the time-product dummy (TPD) method to daily web-scraping data of Dutch online stores.[6] They compared the TPD index with other conventional price index formulas, indicating downward biases with volatile price trends. We extend the approach of de Haan and Hendriks (2013) in three respects. First, we extend the TPD model by taking explicit account of the difference in the retail service quality of outlet channels (the extended TPD model). We employ the combinations of the two fixed effects, i.e. products and outlet channels, in the cross-sectional direction as our dataset includes sales records of various store types in the same product. Second, we construct a high-frequency quality-adjusted price index on a weekly basis using daily scanner data. Due to the strong weekly effects observed in the retail sales trends, we construct weekly quality-adjusted price indices based on the estimates for weekly dummies applied to both daily and weekly converted data. Third, we apply the extended TPD model to all food and beverage products in a unified manner, thereby aggregating them into the overall food and beverage product price indices.

Our conclusions are summarized as follows. The magnitude of the downward bias is estimated at from −0.6 to −0.3 points on the CPI for food less perishables and eating out on the basis of cumulative changes from January 2020 to June. The contribution of the estimates to the overall CPI is −0.3% to −0.15% points on an annualized basis, considering that the estimation covers half-year and the weights are about a quarter of the overall CPI. The magnitude of measurement errors is deemed limited, and the overall trend of the CPI remains unchanged even after incorporating the estimated downward bias. However, it should be noted that this downward bias arises mainly from the "one-specification-for-one-item" policy by weakening the price representativeness, called the broadly-defined lower-level substitution bias in Shiratsuka (2021).

The rest of this paper is as follows. Section 2 describes the extension of the TPD model to incorporate the explicit quality adjustment for retail service quality. Section 3 explains the INTAGE SRI (Nationwide Retail Store Panel Survey) scanner data and summarizes the retail sales trends around the first wave of the COVID-19 spread. Section 4 explains our quality adjustment strategy and constructs high-frequency quality-adjusted price indices based on the baseline estimation results for the extended TPD model. Section 5 describes comprehensive robustness checks on the baseline estimation results. Section 6 discusses the CPI measurement errors under the first wave of the COVID-19 spread in Tokyo. Finally, Section 7 concludes the paper by discussing the implications of our estimation exercises for the official CPI compilation methods.

2. Empirical framework

We extend the TPD model in the studies by de Haan and Hendriks (2013) and de Haan and Krsinich (2014) by explicitly considering the difference in the retail service quality of outlet channels (extended TPD

[6] de Haan and Hendriks (2013), de Haan and Krsinich (2014), and de Haan (2015) proposed two estimation procedures to incorporate scanner and web-scraping data into price indices—the time dummy hedonic (TDH) index (when information on item characteristics is readily available) and the TPD index (when unavailable). Krsinich (2016) examined the application of the TPD index combined with rolling estimations, called the fixed effects window splicing (FEWS) index. She argued that the FEWS index is robust to retroactive revisions from data accumulation over time, but it is less applicable to products with rapid changes in product characteristics and consumer preferences. de Haan et al. (2021) pointed out that the TPD index is distorted toward zero due to overfitting. They argued that the TDH method makes robust and explicit quality adjustments based on quality characteristics, whereas the TPD method has problems with implicit quality adjustments.

model). Assuming that N different products are sold during the sample period from 0 to T, the period 0 is the base period, and the periods from 1 to T are the comparison periods. $p_{i,t}$ denotes the price of a product i at period t ($t = 0, 1, \ldots, T$). By using log-transformed prices as a dependent variable, the log-linear time-dummy hedonic (TDH) model is stated as follows:

$$lnp_{i,t} = \alpha + \sum_{t=1}^{T} \lambda_t TD_{i,t} + \sum_{k=1}^{K} \beta_k z_{k,i} + \epsilon_{i,t}, \tag{1}$$

where $TD_{i,t}$ is the time dummy variable, which takes the value of 1 if the observation is the period t and 0 otherwise. $z_{k,i}$ is the indicators of product characteristics k for a product i, and β_k is the corresponding estimated parameter. $\epsilon_{i,t}$ is the error term. The price index from period 0 to t is computed by taking the exponential of the estimated parameter for the time dummy λ_t.

Similarly, the log-linear TPD model is defined as follows:

$$lnp_{i,t} = \alpha + \sum_{t=1}^{T} \lambda_t TD_{i,t} + \sum_{i=1}^{N-1} \gamma_i PD_i + \epsilon_{i,t}, \tag{2}$$

where PD_i is the product dummy variable, which takes the value of 1 if the observation relates to a product i and 0 otherwise, and the dummy for an arbitrary product N is excluded in the estimation specification to identify the model. The estimated parameters λ_t and γ_i are fixed effects for time and product, respectively. As with the TPD model, the price index from period 0 to t is computed by taking the exponential of the estimated parameter for the time dummy λ_t.

We extend the log-linear TPD model by making an explicit quality adjustment for retail services by introducing an additional cross-sectional control variable for stores as follows:

$$lnp_{i,t} = \alpha + \sum_{t=1}^{T} \lambda_t TD_{i,t} + \sum_{i=1}^{N-1} \gamma_i PD_i + \sum_{s=1}^{S-1} \delta_s ST_{s,i} + \epsilon_{i,t}, \tag{3}$$

where $ST_{s,i}$ is the store dummy variable, which takes the value of 1 if a product i is sold at a store s and 0 otherwise, and the dummy for an arbitrary store S is excluded in the estimation specification to identify the model.

As discussed by Krsinich (2016), in theory, the price indices based on the TDH model and the TPD model are the same because the fixed effect for a product i corresponds to the net effect of the characteristics for a product i. However, in practice, the TPD model is less efficient than the TDH model if sufficient information on the characteristics is available. The number of relevant characteristics can be narrowed down considerably by applying the TDH model.[7] However, the TDH model is often difficult to apply to analyses with scanner data because, in many cases, it is hard to link scanner data to a dataset of the characteristics of a wide range of products.

Thus, choosing either the TDH or TPD model is a practical issue based on data availability. We consider the TPD model a realistic option in this paper because it needs to handle scanner data covering a wide variety of product categories in a unified manner.

3. Data

In this section, we explain the INTAGE SRI scanner data and then summarize the retail sales trends under the first wave of the COVID-19 spread.

3.1. INTAGE scanner data

We employ daily scanner data of food and beverage products, SRI (Nationwide Retail Store Panel Survey) provided by INTAGE Inc. The dataset contains the daily sales, size, and quantity information of nationwide retail stores in Japan at a barcode level from January 2019 to June 2020. The sales records cover sales of food and beverage products with JAN codes, recorded as a stock-keeping unit (SKU) in the dataset. Food and beverage products cover about 20% of the consumption expenditure basket for the CPI. The outlets are classified into seven types-(i) general merchandise stores (GMS), having a sales floor of 3000 or more square meters and more than 50 employees, (ii) supermarkets (SM), having 500 square meters or above, (iii) mini supermarkets (M-SM), (iv) convenience stores (CVS), (v) home centers and discount stores (HC/DS), (vi) large drug stores (Drug/L), and (vii) discount liquor stores (Liquor/DS).[8]

We focus on the data for retail stores in Tokyo from January to June 2020 to analyze the effects of the first wave of the COVID-19 spread in Tokyo. Table 1 summarizes information on the dataset from the product side on the upper panel and the store side on the lower panel. The upper panel reveals that the total number of SKUs is just over 76,000, categorized into 151 items, such as rice and bread. The items are further classified into eight categories, including staple foods. The lower panel presents the daily average data on retail trends. GMS has the largest number of price observations (products) and attracts many customers, registering large sales amounts in a relatively small number of stores. SM has the largest total sales amounts and a relatively large number of price observations and customers. In contrast, CVS and Drug/L have a limited number of price observations and relatively low total sales amounts in a relatively large number of stores.

3.2. Unit price and effective quantity indices

We employ unit prices as the dependent variables in estimating high-frequency quality-adjusted price indices at an elementary aggregation level. Prior to our analysis, we examine how the relationship between unit prices and quantity changed across retail channels during the first wave of the COVID-19 spread.

The sales amounts are calculated by multiplying the price (P) and the quantity (Q) of an individual product and summing them up over all products. The price is decomposed into the product of the unit price (UP) and unit size (US). The unit size multiplied by the quantity is defined as the effective quantity (EQ). Thus, the sales amounts (SA) can be redefined as the summation of the product of the unit price and effective quantity of all products, as follows:

$$SA_{k,t} = \sum_i P_{i,k,t} \times Q_{i,k,t} = \sum_i UP_{i,k,t} \times US_{i,k,t} \times Q_{i,k,t} = \sum_i UP_{i,k,t} \times EQ_{i,k,t}, \tag{4}$$

where subscript i, k, and t indicate a product, store, and time, respectively.

Based on the decomposition of Eq. (4), we construct the unit price index (UPI) and effective quantity index (EQI) using the real-time weight of the daily sales amounts as shown by Eqs. (5) and (6) below:

$$UPI_t = \sum_i \sum_k sw_{i,k,t} UPI_{i,k,t}, \tag{5}$$

$$EQI_t = \sum_i \sum_k sw_{i,k,t} EQI_{i,k,t}, \tag{6}$$

where sw denotes weight for a product i at store k at time t. The indices are constructed using the first week of January 2020 as the base period.

[7] For example, see the study by Shiratsuka (1995) for selecting explanatory variables in the THD model. He applied the TDH model to automobiles in Japan by selecting three key performance characteristics—max power, wheelbase, and interior space—out of 11 characteristics.

[8] INTAGE data divide SM into large and small sizes by floor size, but we merge them as similar retail sales trends are observed in the two outlet types. We also drop outlets classified as small drug stores (Drug/S) and small liquor shops (Liquor/S) because of the small outlet samples and fewer SKUs. INTAGE scanner data expand coverage of stores from January to March 2019, especially in HC/DS.

Table 1
Summary statistics.

(1) Items and SKUs					
Category	Number of items	Number of SKUs	SKUs per item		
			Mean	Maximum	Minimum
Staple foods	17	10,580	622	3204	30
Seasonings	36	11,890	330	1696	11
Processed foods	45	20,220	449	2117	22
Sweets/Snacks	19	16,270	856	2695	10
Milk beverages	4	789	197	441	55
Coffee/Tea	8	2049	256	706	61
Soft drinks	16	5074	317	724	73
Alcoholic beverages	6	9058	1510	2744	491
Total	151	75,930	503	3204	22
(2) Retail sales trend (Daily average)					
Outlet type	Sales amount (million yen)	Number of stores	Per outlet		Per customer
			Number of price observations	Number of customers	Sales amount (yen)
GMS	71.8	15.1	4641.0	6505.7	743.8
SM	126.0	60.9	2927.8	2639.8	783.9
M-SM	13.6	17.6	1469.5	1297.8	586.3
CVS	7.3	51.4	473.7	728.1	204.1
HC/DS	12.9	14.6	909.7	2510.3	335.9
Drug/L	12.9	85.6	380.3	704.0	193.6
Liquor/DS	4.7	9.7	406.5	335.2	1413.0
Total	249.2	254.9	1366.4	1645.5	450.7

Notes: Figures in the table include SKUs sold from January 6, 2020, to June 28, 2020.

Fig. 4 depicts the computed results for the UPI and EQI across store types. Three points should be noted. First, the sharp increase in the UPIs for GMS, SM, M-SM, and HC/DS shortly before the first declaration of the state of emergency corresponds to the spike in the EQIs. The UPI for Drug/L increased slightly in early March. The UPIs for other store types did not experience such significant increases, especially those for CVS experienced a very flat trend over time, even under the first declaration of the state of emergency. Second, the EQIs for most store types, except for CVS and Liquor/DS, experienced two spikes in the early and end of March, suggesting stock-piling behavior at many households in preparation for the stay-at-home request under the state of emergency. Third, the UPIs for GMS, SM, and M-SM, which increased shortly before the first declaration of the state of emergency, started declining after lifting the state of emergency.

Based on the first and second points, the increases in the UPIs are associated with the increases in the EQIs, suggesting that all the increases in the UPIs should not be considered as price hikes on a quality-adjusted basis. During the first wave of the COVID-19 spread, due to the increased risk of COVID-19 infection, households decided to purchase at higher prices than usual to reduce shopping frequency, travel distance to stores, and shopping time.[9]

3.3. Number of SKUs

We next ascertain the impact of stockouts by examining the number of SKUs at outlets over time. Since canner data collect sales records of products at outlets, it is impossible to distinguish the dates of temporary stockouts from those of no sales records for a product. Instead, we count the number of SKUs sold at each outlet every day and aggregate them by the store types on a weekly frequency basis.

Fig. 5 plots the indices for the number of SKUs sold at a store by store type using the first week of January 2020 as the base period.

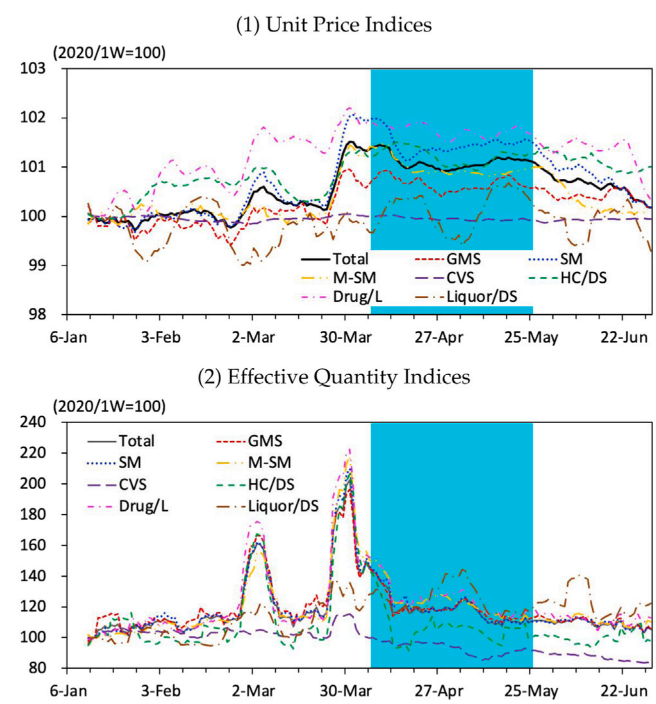

Fig. 4. Unit price indices and effective quantity indices by outlet types. Notes: The plotted figures are backward seven-day moving averages, starting from the date on the horizontal axis. The light blue areas indicate the period of the first declaration of a state of emergency in Tokyo. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The figure shows that Liquor/DS increased significantly, while CVS followed a mild declining trend. Other store types remained almost unchanged, with small jumps in the weeks when stock-piling purchasing behavior was observed in the early and the end of March. These observations are consistent with the sales trend of EQIs, just shown in the previous subsection, confirming the findings in Cavallo and Kryvtsov (2021) that the stockouts had very limited impacts on household consumption behavior and inflation trend in Japan.

[9] As seen in Fig. 4, the effective sales value remained flat at a high level during the first declaration of the state of emergency. Although not depicted in the figure, the number of customers visiting stores decreased during this period, whereas the sales amounts per customer increased. This observation also suggests that households tried to minimize shopping frequency by making bulk purchases in a single shopping trip to control the risk of COVID-19 infection.

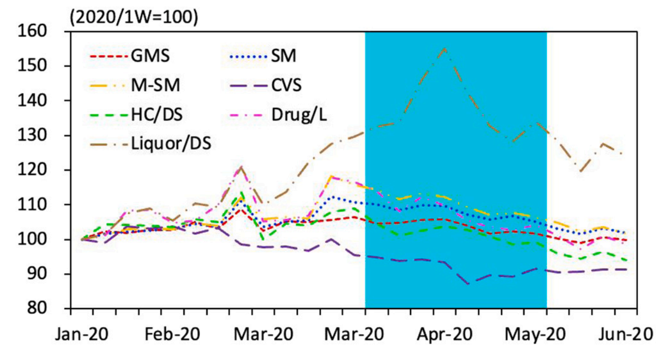

Fig. 5. Number of SKUs. Notes: The light blue areas indicate the period of the first declaration of a state of emergency in Tokyo. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.4. Temporary sales

We now examine the effects of temporary sales. We define temporary sales as a day that the observed price of a product at a store is at least two yen lower than the mode price for the past seven days, including the current day, following the study of Abe and Tonogi (2010). We also define a price reduction as the deviation of the observed price at a temporary sale from the mode price. We aggregate the frequency and size of price reductions over items by computing simple mean and weighted mean with sales amounts.

Fig. 6 depicts the frequency of temporary sales and the size of price reductions at temporary sales over time.[10] These figures reveal that temporary sales declined rapidly from early March toward the first declaration of the state of emergency, as indicated by the significant divergence between the T-index and the T-mode-index of CPINow in Fig. 3. Before March, the differences between the simple and weighted means remained almost constant, but they started narrowing in early March. These observations confirm that both the frequency of temporary sales and the size of price reductions declined during this period. That suggests the importance of considering the temporary sales effects on the retail sales trends when computing the high-frequency quality-adjusted price indices.

To confirm the structural changes in temporary sales under the first wave of the COVID-19 spread, we perform a statistical test for structural changes with an unknown break point. Fig. 7 plots the sup-Wald statistics for structural changes in the differences between simple and weighted means for the share of temporary sales and the size of price reductions, respectively. The sup-Wald statistics are computed from January 6 to June 28 in 2020, with 15% trimming for detecting an unknown break point. The sup-Wald statistics start increasing from mid-February and reach their peaks around the end of March. The detected breakpoints for the share of temporary sales and the size of price reductions are March 28 and 24, 2020, respectively.

It should be noted that declines in the frequency of temporary sales and the size of price reduction influence how to measure the central tendency of the unit price distributions over time. Fig. 8 plots three statistical measures for the central tendency of the unit price distributions over time—mean, median, and mode of weekly windows for a product at a store—aggregated by the time-varying sales amount weights. The three measures are indexed at 100 with the mode values for the first week of January 2020.

The order of the three statistics remains unchanged over time: mode, median, and mean from top to bottom, indicating that the unit price distributions are skewed toward the left. However, the mean moves closer to the mode, reflecting a decrease in the temporary sales effects from the end of February, and the leftward-skewness of the unit price distributions also declined. That implies that inflation trends are estimated differently, depending on the measures for the central tendency of the unit price distributions used. The mean values for unit prices are most influenced by the declines in temporary sales under the first wave of the COVID-19 spread, and the mode values are the least influenced.

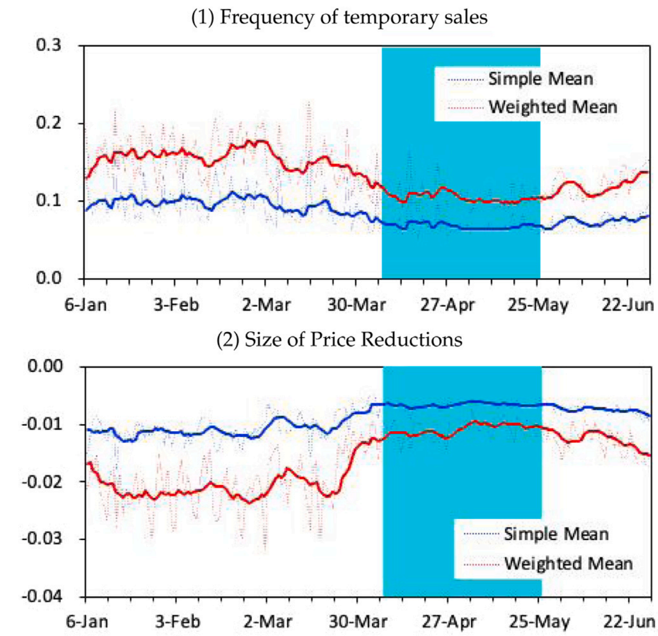

Fig. 6. Frequency and price discount of temporary sales. Notes: The bold lines are backward seven-day moving averages starting from the date on the horizontal axis. The light blue areas indicate the period of the first declaration of a state of emergency in Tokyo. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

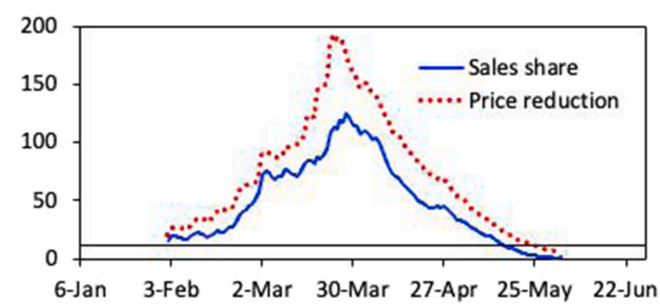

Fig. 7. Test for structural changes with an unknown breakpoint. Notes: The plotted figures are the sup-Wald statistics for detecting an unknown break point for the differences between the simple and weighted means for sales share and price reductions. The bold horizontal line indicates the critical value for 1% with 15% trimming for detecting an unknown break point. The breakpoints for sales share and price reductions are March 28 and 24, 2020, respectively.

[10] Sudo et al. (2018) examine the effects of temporary sales on the macroeconomic implication by focusing on the households' allocation of time for work, leisure, and bargain hunting. They compute various measures for the frequency of temporary sales based on two types of sales filters: the mode price during a certain window length (Eichenbaum et al., 2011, and Kehoe and Midrigan, 2015) and filtered V-shaped price fluctuations as a temporary sale (Nakamura and Steinsson, 2008). Their estimates for the closest definition to ours, mode price for a seven-day widow, are around 0.2 in the 2010s and slightly higher than our estimates in early 2020. In this respect, our estimates for the frequency of temporary sales become slightly higher, around 0.2, when limiting data only for GMS and SM, which seems consistent with the estimates in Sudo et al. (2018).

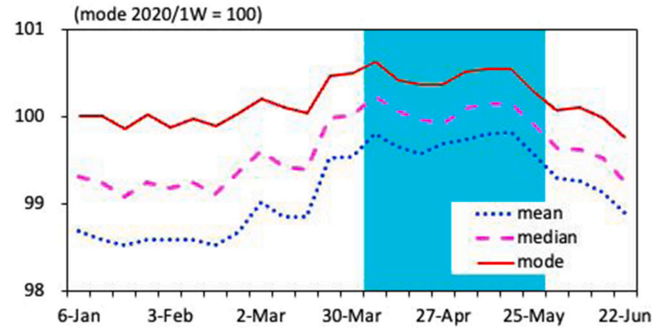

Fig. 8. Summary statistics for central tendency of unit price distributions. Notes: The light blue area indicates the period for the first declaration of a state of emergency in Tokyo. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

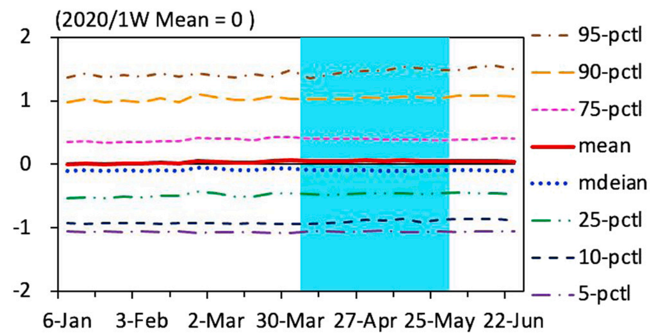

Fig. 9. Unit size distributions. Notes: The light blue area indicates the period of the first declaration of a state of emergency in Tokyo. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

However, the unit size of products purchased by households did not change significantly during the first wave of the COVID-19 spread. Fig. 9 plots quantiles of the unit size distributions for products purchased by households on a weekly basis. The plotted figures are normalized using the mean and standard deviation for the first week of January 2020. The distributions remain almost unchanged, slightly shifting toward a larger direction. In response to the declined temporal sales, households continued to purchase products with approximately the same unit size, suggesting that, due to the increased risk for COVID-19 infection, households chose to purchase at higher prices to reduce travel distance to stores and shopping time.

4. High-frequency quality-adjusted price indices

In this section, we explain our quality adjustment strategy and construct high-frequency quality-adjusted price indices based on the baseline estimation results for the extended TPD model.

4.1. Quality adjustment strategy

Given the observations about the significant structural changes in the retail sales trends under the first wave of COVID-19 in the previous section, we point out the importance of the quality adjustments in two respects, in addition to the standard quality adjustments for product characteristics.

First, we need to pay careful attention to accounting for the structural changes in the frequency and size of price reductions of temporary sales. Due to the significant structural changes in the relative relationship between the regular and discount prices of temporary sales, the treatment of discount prices significantly influences the estimated price index. It should be noted that both supply- and demand-side factors induce such structural changes in the retail sales trends. On the one hand, retailers try to raise prices in real terms by reducing the frequency and size of price reductions of temporary sales to cover the increased costs of infection prevention measures. On the other hand, consumers are more likely to purchase at higher-than-normal prices to minimize the risk of COVID-19 infection under voluntary lockdown.

Second, we also need to consider outlet substitution bias. This bias is induced by the structural change in retail markets, as observed by the shift in shopping sites across the outlet types. The shifts in nearby outlets indicate possible changes in consumers' preferences due to the stay-at-home request under the voluntary lockdown and the risk of COVID-19 infection. Under such circumstances, households are likely to alter their shopping behavior to minimize shopping frequency, travel distance, and shopping time. Such changes in consumers' behavior indicate that, on a quality-adjusted basis, they perceive that the prices for the same products sold at supermarkets nearby are less expensive than before.[11]

To deal with the temporary sales effects and the retail service quality differences, we employ four baseline specifications in Table 2—D-ROL_SK+ST, D-ROL_SK+TY, W-MO_SK+ST, and W-MO_SK+TY.

The four specifications are a combination of strategies to deal with the two factors—the temporary sales effects and the retail service quality differences. The first part of the abbreviations, D-ROL or W-MO, corresponds to the strategy to deal with the effects of structural changes in temporary sales. D-ROL employs two-week staggered rolling estimations for daily frequency data with weekly time dummies. Rolling estimations are expected to absorb item-wide structural changes, including the frequency and price discount of temporary sales, by changes in constant terms over time.[12] W-MO converts daily frequency data to weekly frequency data using the weekly mode of unit prices. The weekly mode conversion is a direct adjustment for temporary sales effects on unit prices, as discussed in the previous section.

The second part of the abbreviations, SK+ST or SK+TY, represents the fixed effects for a cross-sectional direction to control differences in the product quality and the retail service quality across the outlet channels. SK+ST employs two fixed effects for SKU code and store identity (ID) separately, thereby controlling the retail service quality differences at an individual store level. SK+TY employs SKU code and store type dummies for GMS, SM, M-SM, CVS, HC/DS, Drug/L, and Liquor/DS separately, thus, controlling the retail service quality differences at the outlet channel level. Comparing the estimation results for the two types of fixed effects enables us to confirm whether differences in retail service quality matter at a store or an outlet channel level. In addition, the second type of fixed effects also has the advantage that the estimates of fixed effects provide convenience premiums or discounts across outlet channels.

4.2. Estimation procedure

We construct the high-frequency quality-adjusted price indices in two steps. The first step is to estimate the extended TPD model for all

[11] Shiratsuka (1999) pointed out that the CPI failed to reflect consumers' shift from retail shops and department stores to discount stores in the mid-1990s due to the limited coverage of discount stores in price surveys. Such structural changes in retail markets are called "price busting". It was induced by the massive expansion of large-scale shopping sites, such as general merchandise stores and large discount stores. After the sharp appreciation of the Japanese yen after the Plaza accord in 1985, it is often pointed out that the inefficiency of retail markets in Japan is the cause of price level differences between Japan and other advanced economies.

[12] Another way to adjust the temporary sales effects seems to include the temporary sales dummy as an additional explanatory variable in the staggered rolling regression. However, such estimation eliminates the temporary price fluctuation at the individual price observation level, resulting in the over-adjustments of the temporary sales effects by eliminating all the temporary price reductions.

Table 2
Specification for baseline estimations.

Abbreviation	Frequency conversion	Sample period	Fixed effects		
			SKU	Store	Store type
D-ROL_SK+ST	None	Two-week staggered rolling sample period	✓	✓	
D-ROL_SK+TY			✓		✓
W-MO_SK+ST	Weekly mode	Whole sample period	✓	✓	
W-MO_SK+TY			✓		✓

151 items, thereby computing elementary price indices for all items. The second step aggregates elementary indices to an overall index using sales amounts as weights.

As the first step for computing high-frequency quality-adjusted price indices, we estimate the extended TPD model for all 151 items based on the four baseline specifications, as shown in Table 2. The specifications for the daily frequency estimation of D-ROL_SK+ST and D-ROL_SK+TY are given by Eqs. (7) and (8), respectively.

$$lnUP_{i,k,t} = \lambda_w D\_WK^w_{i,k,t} + \gamma_i + \delta_k + \epsilon_{i,k,t}, \tag{7}$$

$$lnUP_{i,k,t} = \lambda_w D\_WK^w_{i,k,t} + \gamma_i + \mu_j + \epsilon_{i,k,t}, \tag{8}$$

$$D\_WK^w_{i,k,t} = \begin{cases} 1 & t \text{ belongs to the week } w \\ 0 & \text{otherwise,} \end{cases}$$

where $UP_{i,k,t}$ and $D\_WK^w_{i,k,t}$ represent unit price and week dummy of second-week observations in the rolling subsample period for SKU code i at store ID k and time t, respectively.[13] γ_i, δ_k, and μ_j are fixed effects for a product with SKU code i, a store with store ID k, and store type j, respectively. The fixed effect for product γ_i controls all the product characteristics, including the unit size, tracing a shadow price for the quality of a product i. The fixed effects for store δ_k and store type μ_j control all the quality differences related to retail services at store and store type levels, respectively. Here store ID k is included in store type j, i.e. GMS, SM, M-SM, CVS, HC/DS, Drug/L, and Liquor/DS. The estimated coefficient for week dummies (λ_w) corresponds to the quality-adjusted price index using week one as the base period.

The specifications for the weekly frequency estimation of W-MO_SK+ST and W-MO_SK+TY are given by Eqs. (9) and (10) below:

$$lnUP^{MO}_{i,k,w} = \sum_{v=2}^{W} \lambda_v D\_WK^v_{i,k,w} + \gamma_i + \delta_k + \epsilon_{i,k,w}, \tag{9}$$

$$lnUP^{MO}_{i,k,w} = \sum_{w=2}^{W} \lambda_v D\_WK^v_{i,k,w} + \gamma_i + \mu_j + \epsilon_{i,k,w}, \tag{10}$$

$$D\_WK^v_{i,k,t} = \begin{cases} 1 & t \text{ belongs to the week } v \\ 0 & \text{otherwise,} \end{cases}$$

where $UP^{MO}_{i,k,w}$ and $D\_WK^v_{i,k,t}$ indicate the weekly mode of unit price and week dummies for SKU code i, at store ID k, and week v, respectively. γ_i, δ_k, and μ_j indicate the fixed effects for a product with SKU code i, a store with store ID k, and store type j, which are the same as the daily frequency estimation specifications.

Table 3 summarizes the estimation results for W-MO_SK+ST and W-MO_SK+TY on the three selected items—sweetened buns with the largest number of SKUs, honey with the median number of SKUs, and skimmed milk with the smallest number of SKUs. We find two important points about the estimation results in the table. First, the adjusted R-squared for all the estimations is generally high, indicating that the two types of fixed effects for a cross-sectional direction—SKU and store as well as SKU and store type—effectively control the time-invariant factors for product characteristics and retail service quality. Second, the estimated quality-adjusted price indices differ from product to product, even with generally high precision in the estimation results. The price indices for honey reveal statistically significant increases from the first week of January 2020 during most of the sample period, whereas those for skimmed milk remain almost unchanged, as seen in the statistically insignificant estimates for most of the weeks, and those for sweetened buns exhibit statistically significant decreases during the second half of the sample period.

We repeat the full sample estimation of the extended TPD model of Eqs. (9) and (10) for all 151 items for the weekly frequency estimations. For the daily frequency estimation, we first conduct the two-week staggered rolling regressions on the extended TPD model of Eqs. (7) and (8) 24 times over the sample period of 25 weeks and then repeat that process for all 151 items. That is the first step of computing the elementary price indices of the items.

In the second step, we aggregate elementary price indices to the overall price index by using sales amounts as weights. In the daily frequency estimation for the two-week subsample periods, the estimated coefficients for week dummies correspond to quality-adjusted price changes from the previous week. The cumulative quality-adjusted price changes are computed by linking the estimated week-on-week price changes from the base week. In the weekly frequency estimation for the full sample period, the estimated coefficients for week dummies correspond to the cumulative quality-adjusted price changes from the base week. Thus, the high-frequency quality-adjusted price indices for the week w in the daily and weekly frequency estimations—$HFQAPI^{D-ROL}_w$ and $HFQAPI^{W-MO}_w$—are computed as follows:

$$HFQAPI^{D-ROL}_w = \sum_{v=2}^{w} \sum_{i=1}^{I} sw_i \lambda_{v,i}, \tag{11}$$

$$HFQAPI^{W-MO}_w = \sum_{i=1}^{I} sw_i \lambda_{w,i}, \tag{12}$$

where sw_i is the sales amounts weight for item i, and I is the number of SKUs of the item. We use the fixed weight in this paper because of two reasons. First, using the fixed or time-varying weights in the daily frequency estimations produces similar results over time. Second, the weekly frequency estimations with full sample data do not fit well with the time-varying weights aggregation.

4.3. Baseline estimation results

Fig. 10 summarizes the estimated HFQAPIs based on the baseline estimation results for D-ROL_SK+ST, D-ROL_SK+TY, W-MO_SK+ST, and W-MO_SK+TY. The plotted figures are the cumulative price changes on a quality-adjusted basis, using the week starting on January 6, 2020, as the base period. The figure also includes the UPI for reference. The light blue shaded area in the figures indicates the period of the first declaration of a state of emergency in Tokyo.

The figure depicts three points regarding the inflation developments for food and beverage products during the first half of 2020. First, all the indicators, including UPI, reveal very similar trends until the end of February. Second, UPI and other quality-adjusted price indices reveal different trends since then because of the heightened concern over the wide and rapid spread of COVID-19 in Tokyo. Third, again, all the indicators reveal a similar downward trend from the end of the first declaration of a state of emergency.

[13] Regarding the subsample period of two weeks, estimation results remain almost unchanged when extending it to three weeks.

Table 3
Representative estimation results for weekly frequency estimations.

	Sweetened buns (Largest number of SKUs)				Honey (Median number of SKUs)				Skimmed milk (Smallest number of SKUs)			
	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE
2wk_id	−0.003***	(0.001)	−0.003***	(0.001)	0.005*	(0.003)	0.004	(0.003)	0.000	(0.004)	−0.001	(0.006)
3wk_id	0.001	(0.001)	0.001	(0.001)	−0.010***	(0.003)	−0.010***	(0.003)	−0.009*	(0.005)	−0.009	(0.007)
4wk_id	0.002*	(0.001)	0.002*	(0.001)	0.001	(0.003)	0.002	(0.003)	−0.003	(0.004)	−0.002	(0.006)
5wk_id	0.000	(0.001)	0.000	(0.001)	0.010***	(0.003)	0.010***	(0.003)	−0.003	(0.005)	−0.004	(0.007)
6wk_id	−0.001	(0.001)	−0.001	(0.001)	−0.013***	(0.003)	−0.013***	(0.003)	−0.002	(0.005)	−0.004	(0.007)
7wk_id	0.001	(0.001)	0.001	(0.001)	0.009***	(0.003)	0.010***	(0.003)	0.000	(0.004)	0.000	(0.006)
8wk_id	0.002**	(0.001)	0.002*	(0.001)	0.007***	(0.003)	0.008***	(0.003)	0.001	(0.004)	0.001	(0.006)
9wk_id	−0.002**	(0.001)	−0.003**	(0.001)	0.011***	(0.003)	0.011***	(0.003)	−0.001	(0.004)	−0.001	(0.007)
10wk_id	−0.001	(0.001)	−0.002	(0.001)	0.015***	(0.003)	0.015***	(0.003)	−0.006	(0.006)	−0.002	(0.008)
11wk_id	−0.003**	(0.001)	−0.003***	(0.001)	0.001	(0.003)	0.001	(0.003)	−0.008	(0.006)	−0.007	(0.007)
12wk_id	0.005***	(0.001)	0.004***	(0.001)	0.014***	(0.003)	0.013***	(0.003)	−0.004	(0.005)	−0.002	(0.007)
13wk_id	−0.001	(0.001)	−0.001	(0.001)	0.013***	(0.003)	0.012***	(0.003)	−0.007	(0.005)	−0.004	(0.007)
14wk_id	0.001	(0.001)	0.000	(0.001)	0.018***	(0.003)	0.017***	(0.003)	−0.001	(0.004)	0.002	(0.007)
15wk_id	−0.001	(0.001)	−0.002	(0.001)	0.017***	(0.003)	0.016***	(0.003)	−0.003	(0.005)	0.002	(0.007)
16wk_id	−0.002	(0.001)	−0.002**	(0.001)	0.011***	(0.003)	0.012***	(0.003)	−0.002	(0.005)	0.000	(0.007)
17wk_id	−0.005***	(0.001)	−0.005***	(0.001)	0.016***	(0.003)	0.017***	(0.003)	0.000	(0.004)	0.003	(0.006)
18wk_id	−0.009***	(0.001)	−0.009***	(0.001)	0.018***	(0.002)	0.018***	(0.003)	0.001	(0.004)	0.006	(0.006)
19wk_id	−0.004***	(0.001)	−0.005***	(0.001)	0.016***	(0.003)	0.017***	(0.003)	0.001	(0.003)	0.007	(0.006)
20wk_id	−0.002	(0.001)	−0.002	(0.001)	0.014***	(0.003)	0.014***	(0.003)	0.000	(0.004)	0.003	(0.006)
21wk_id	−0.004***	(0.001)	−0.005***	(0.001)	0.015***	(0.003)	0.015***	(0.003)	−0.001	(0.004)	−0.003	(0.006)
22wk_id	−0.010***	(0.001)	−0.010***	(0.001)	0.011***	(0.003)	0.011***	(0.003)	−0.003	(0.004)	−0.007	(0.008)
23wk_id	−0.007***	(0.001)	−0.007***	(0.001)	0.012***	(0.003)	0.013***	(0.003)	−0.002	(0.004)	−0.002	(0.006)
24wk_id	−0.006***	(0.001)	−0.006***	(0.001)	0.014***	(0.003)	0.015***	(0.003)	−0.001	(0.004)	0.001	(0.006)
25wk_id	−0.006***	(0.001)	−0.006***	(0.001)	0.013***	(0.003)	0.013***	(0.003)	−0.002	(0.004)	−0.002	(0.006)
Constant	4.744***	(0.001)	4.744***	(0.001)	0.737***	(0.002)	0.737***	(0.002)	0.728***	(0.003)	0.727***	(0.004)
Observations	449,013		449,013		36,414		36,414		4111		4111	
Adj. R-squared	0.882		0.870		0.988		0.986		0.965		0.931	
Fixed effects	SKU & Store		SKU & Store type		SKU & Store		SKU & Store type		SKU & Store		SKU & Store type	

Notes: Robust standard errors in parentheses.
*$p < 0.1$.
**$p < 0.05$.
***$p < 0.01$.

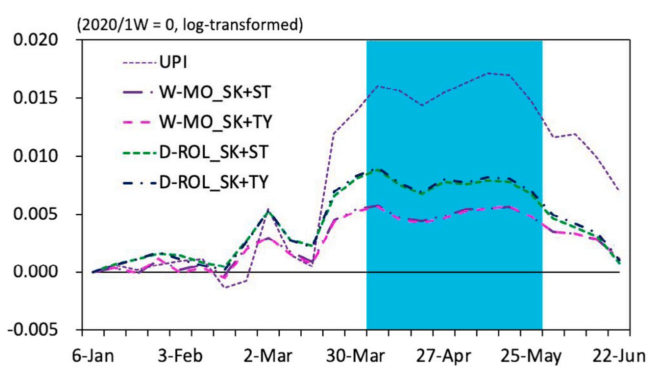

Fig. 10. Baseline estimation results. Notes: The light blue area indicates the period for the first declaration of a state of emergency in Tokyo. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

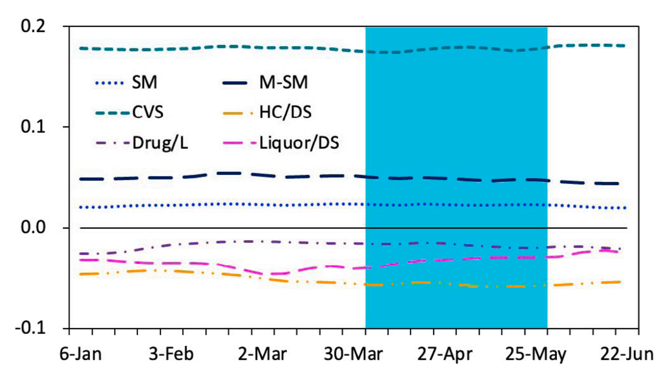

Fig. 11. Estimates for store type fixed effect. Notes: GMS is used as the base of the fixed effect on the store type, and the plotted figures indicate the convenience premiums or discounts of the store type against GMS. The light blue area indicates the period for the first declaration of a state of emergency in Tokyo. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The cumulative inflation from the beginning of January to the peak is estimated at 0.9% in the daily frequency estimations and 0.6% in the weekly frequency estimations. The estimates are slightly different but stay between UPI and zero inflation. In addition, the estimates for the daily and weekly specifications are not influenced by the two types of fixed effects in a cross-sectional direction, SKU and store as well as SKU and store type. That suggests that the retail service quality differences are well controlled at the store type level.

Fig. 11 plots the aggregated fixed effects of store types, δ, over all items using sales amount share as weights. The base store type is GMS, and the plotted figures indicate the convenience premiums or discounts of the store type against GMS. The estimates reveal very stable movements over time for all store types. These observations suggest that convenience premium or discount for store types remained unchanged under the first wave of the COVID-19 spread. CVS has the largest premium, reflecting its high retail service quality in terms of operating hours, location, and shopping time. SM and M-SM have smaller but positive premiums. HC/DS, Drug/L, and Liquor/DS give discounts, which is consistent with the sales strategy of lowering prices by reducing retail services. Discounts for Liquor/DS started shrinking slightly from March, reflecting the growing demand for home drinking under a voluntary lockdown.

Table 4
Robustness checks.

	Abbreviations	Details on robustness checks
(1)	D-ROL_SKxST W-MO_SKxST	Alternative fixed effect for product-store combination (SKxST)
(2)	D-WHL_SK+ST D-WHL_SK+TY	Whole sample estimation using daily frequency data (WHL)
(3)	W-MN_SK+TY W-MD_SK+TY	Alternative summary statistics for weekly conversion: mean (MN) and median (MD)

The results of the baseline estimations seem consistent with Fig. 3, which indicates that the inflation of UPI (CPINow T-index) accelerated, whereas there was no acceleration in the mode UPI (CPINow T-mode-index). The T-index fully reflects the decrease in temporary sales as a price increase, whereas the T-mode-index does not account for changes in temporary sales. The quality-adjusted price indices, in terms of both product characteristics and retail service quality, are somewhere between the two CPINow indicators because households try to minimize the risk of COVID-19 infection by reducing travel distance to outlets and shopping time. In the next section, we examine the appropriateness of the estimated quality-adjusted price index with various robustness checks.

5. Robustness checks

We now carry out the robustness checks from the following three perspectives: (1) alternative fixed effect for product-store combination, (2) whole sample estimation using daily frequency data, and (3) alternative summary statistics for weekly conversion (mean and median). Table 4 summarizes the three robustness checks.[14]

5.1. Robustness check 1: Fixed effect for the product —store combination

The first robustness check examines the impacts of alternative fixed effects for the product-store combination as the cross term of SKU code and store ID. In this case, the same products sold at different stores are considered different goods. The estimation specifications of D-ROL_SKxST and W-MO_SKxST are given by Eqs. (13) and (14) for the daily and weekly frequency estimations, respectively:

$$lnUP_{i,k,t} = \lambda_w D\_WK^w_{i,k,t} + \eta_{i,k} + \epsilon_{i,k,t}, \tag{13}$$

$$lnUP^{MO}_{i,k,w} = \sum_{v=2}^{W} \lambda_v D\_WK^v_{i,k,w} + \eta_{i,k} + \epsilon_{i,k,w}, \tag{14}$$

where $\eta_{i,k}$ is a fixed effect for the product—store combination for SKU code i and store ID k.

Fig. 12 plots D-ROL_SKxST and W-MO_SKxST along with the baseline specifications of D-ROL_SK+ST and W-MO_SK+ST as well as the UPI. Conducting the estimations with the cross term of SKU and store as a fixed effect for a cross-sectional direction produces lower estimates for the quality-adjusted price indices in both the daily and weekly frequency estimations. The downward revision is larger in the daily frequency estimation. This estimation results suggest the possibility of the over-adjustments of SKU-store specific price increases when using the cross term of SKU and store as a fixed effect. This tendency is further amplified by the two-week staggered rolling estimations for daily frequency data through the time-varying constant terms.

[14] In the third robustness check exercise, we show the estimation results for using only the combinations of SKU and store type as fixed effects for cross-sectional direction. We confirmed that almost identical results are obtained using the combinations of SKU and store as well as SKU and store types, as in the baseline estimation shown in Fig. 10.

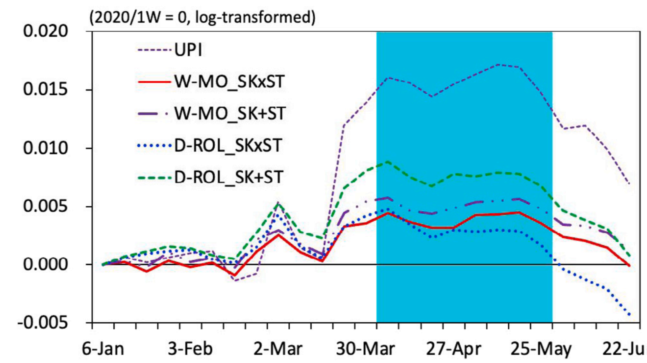

Fig. 12. Robustness check 1: Fixed effect for the product-store combination. Notes: The light blue area indicates the period of the first declaration of a state of emergency in Tokyo. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5.2. Robustness check 2: Daily estimation for the full sample period

The second robustness check focuses on the effects of structural changes in temporary sales in the daily frequency estimations by comparing the results of the daily estimation for the full sample period data with those of the baseline two-week staggered rolling estimations. The staggered rolling estimations are expected to absorb item-wide structural changes, including the frequency and price discount of temporary sales, by changes in constant terms over time. The estimation specifications of D-WHL_SK+ST and D-WHL_SK+ST are given by Eqs. (15) and (16) for the two types of fixed effects, respectively:

$$lnUP_{i,k,t} = \sum_{w=2}^{W} \lambda_w D\_WK^w_{i,k,t} + \gamma_i + \delta_k + \epsilon_{i,k,t}, \tag{15}$$

$$lnUP_{i,k,t} = \sum_{w=2}^{W} \lambda_w D\_WK^w_{i,k,t} + \gamma_i + \mu_j + \epsilon_{i,k,t}, \tag{16}$$

Fig. 13 depicts the estimation results for D-WHL_SK+ST and D-WHL_SK+TY, along with the baseline estimation results for D-ROL_SK+ST and D-ROL_SK+TY as well as the UPI. The cumulative inflation for D-WHL_SK+ST and D-WHL_SK+TY are both estimated at 1.2%, which is higher than the baseline estimation results. However, the estimates are not influenced by the two types of fixed effects in a cross-sectional direction for SKU and store as well as SKU and store type.

The differences between D-ROL and D-WHL are estimated at 0.3%. The differences come from the time-varying constant terms in the two-week staggered rolling estimations. The time-varying constant terms are deemed effective in absorbing item-wide structural changes, including the frequency and price discount of temporary sales. Thus, we regard the differences as the effects of reduced frequency and price reduction of temporary sales in the daily frequency estimations. In addition, the differences between UPI and D-WHL of around 0.5% are regarded as outlet substitution effects stemming from the households' behavior to purchase from a store nearby at slightly higher prices because of the risk of COVID-19 infection.

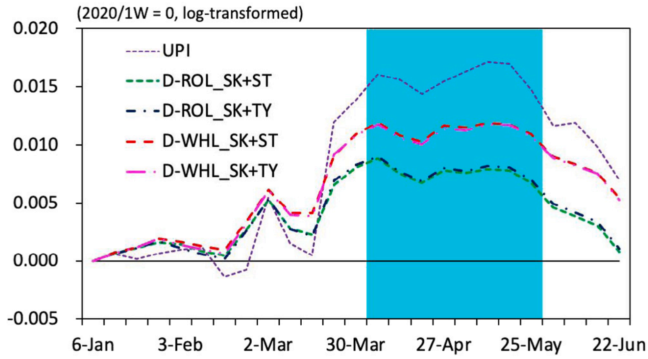

Fig. 13. Robustness check 2: Daily estimation for the full sample period. Notes: The light blue area indicates the period of the first declaration of a state of emergency in Tokyo. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

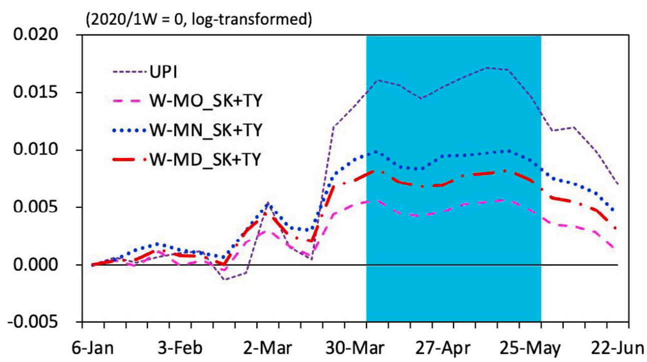

Fig. 14. Robustness check 3: Weekly data conversion measures. Notes: The light blue area indicates the period of the first declaration of a state of emergency in Tokyo. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5.3. Robustness check 3: Weekly data conversion measures

The third robustness check focuses on the effects of structural changes in temporary sales in the weekly frequency estimations by comparing the estimation results using two alternative summary statistics for data conversion on a weekly basis—mean (MN) and median (MD). The estimation specifications of W-MN_SKxTY and W-MD_SKxTY are given by Eqs. (17) and (18) for the fixed effects using the combination of SKU and store type:

$$lnUP^{MN}_{i,k,w} = \sum_{w=2}^{W} \lambda_v D\_WK^{v}_{i,k,w} + \gamma_i + \mu_j + \epsilon_{i,k,w}, \tag{17}$$

$$lnUP^{MD}_{i,k,w} = \sum_{w=2}^{W} \lambda_v D\_WK^{v}_{i,k,w} + \gamma_i + \mu_j + \epsilon_{i,k,w}. \tag{18}$$

The estimation results in Fig. 14 are ordered from highest to lowest as follows: W-MN, W-MD, and W-MO, with almost identical estimates for the two types of fixed effects in a cross-sectional direction. The results suggest that W-MN is mostly influenced by declines in the frequency and size of price reductions of temporary sales. That is consistent with the observation in Fig. 8 that compares the inflation trend with the three statistical measures for the central tendency of unit price distributions, revealing that the mean is the highest, followed by the mean and median. Thus, it is likely that W-MN and W-MD make over-adjustments of the temporary sales effects.

As in the previous robustness check for the daily full sample estimations, the differences between W-MN and W-MO (0.4%) and those between UPI and W-MN (0.7%) are regarded as the effects of temporary sales and outlet substitution, respectively. The differences between W-MN and W-MO solely come from the summary statistics measures in weekly frequency conversion, that is, the mean or mode of unit prices in a week, which corresponds to the effects of temporary sales in terms of reduced frequency and price reduction.

Table 5
Weights for CPI food.

	Tokyo		Nationwide	
	2015-base	2020-base	2015-base	2020-base
Food	24.96	25.29	26.23	26.26
Less perishables and eating out	15.14	16.10	16.88	17.70
Perishables	4.12	3.85	4.14	3.96
Eating out	5.70	5.34	5.21	4.60

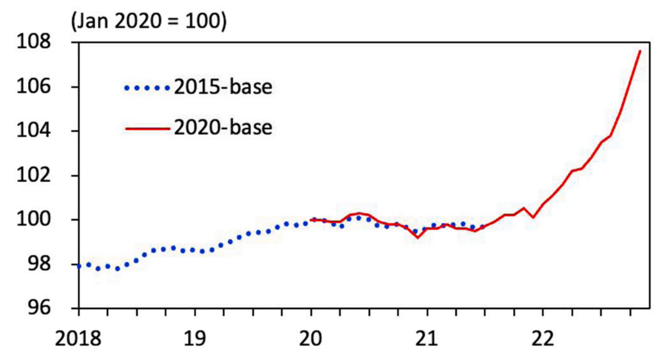

Fig. 15. CPI for food less perishables and eating out (Tokyo). Notes: Plotted data is computed by subtracting the CPI for eating out from the CPI food less perishables. *Source:* Statistics Bureau of Japan, Consumer Price Index.

6. CPI measurement errors under the first wave of the COVID-19 spread

In this section, we assess the CPI measurement errors under the first wave of the COVID-19 spread, based on the estimation results using retail scanner data in Tokyo.

6.1. CPI revision from 2015-base to 2020-base

The Japanese CPI is revised every five years at years ending in zero and five. The revisions are released in July of the following year, considering the availability of the source data for computing weights, the Family Income and Expenditure Survey (FIES) of the base year. As a result, the previous base CPI continues to be released until June, the previous month for the release of the new base CPI. In the case of the 2020-base revision, the CPI weights are retroactively revised and released in July 2021, reflecting the impacts of the COVID-19 spread.[15] In addition, the CPI for the 2015-base and 2020-base are both available from January 2020 to June 2021.

The consumption basket for food and beverage products remained mostly the same even under the COVID-19 spread, as shown in Table 5. As a result, no significant deviations arise in the CPI for food less perishables and eating out in Tokyo, which is comparable to the coverage of scanner data we employ, even after the weight revision from the 2015-base to the 2020-base, as shown in Fig. 15.

[15] The weights for calculating the index are generally computed using the FIES of the base year. However, the weights for the 2020-base are computed by using the average of 2019 and 2020, taking into account the impact of the COVID-19.

Table 6
Decomposition of unit price changes.

	Daily frequency estimation	Weekly frequency estimation
Unit price index	1.6% (= UPI)	
Outlet substitution effects	0.5% (=UPI−D-WHL)	0.7% (=UPI−W-MN)
Temporary sales effects	0.4% (=D-WHL−D-ROL)	0.4% (=W-MN−W-MO)
Quality-adjusted price index (benchmark)	0.7% (=D-ROL)	0.5% (=W-MO)
Measurement errors	−0.6/−0.5% (=CPI−D-ROL)	−0.4-/−0.3% (=CPI−W-MO)
Consumer price index (2015-base/2020-base)	0.1/0.2% (=CPI)	0.1/0.2% (=CPI)

6.2. Quantitative assessment of the CPI measurement bias

Table 6 summarizes the quantitative assessments on the measurement errors for the CPI for food less perishables and eating out under the first wave of the COVID-19 spread. The data are transformed to a monthly basis, using the data for the weeks when the CPI price surveys are conducted: Wednesday, Thursday, or Friday of the week including the 12th day of the month. The data are the cumulative changes from January to June when the price increases peaked.

The UPI increased by 1.6% from January to June, and the quality-adjusted price indices increased by 0.7% in the daily frequency estimation and 0.5% in the weekly frequency estimation. The differences between the UPI and the quality-adjusted price indices are attributable to the two factors: outlet substitution effects and temporary sales effects: 0.5% and 0.4% in the daily frequency estimation and 0.7% and 0.4% in the weekly frequency estimation.

In the meantime, the CPI increased by 0.1% in the 2015-base and 0.2% in the 2020-base. The differences between the quality-adjusted price indices and the CPIs, ranging from −0.6 to −0.3 points, correspond to the measurement errors of the CPI for food less perishables and eating out. Considering the estimation covers half-year and the weights are about a quarter of the overall CPI, the measurement errors in the CPI for food less perishable and eating out produced a downward bias of −0.3 to −0.15% points on an annualized basis.

The magnitude of measurement errors is deemed limited, and the overall trend of the CPI remains unchanged even after incorporating the estimated downward bias. However, it should be stressed that the CPI for food less perishable and eating out was under-evaluated, not over-evaluated, for the true inflation under the COVID-19 pandemic.

6.3. Possible sources of measurement errors

We next examine possible sources of the CPI measurement errors under the COVID-19 pandemic. The most likely source is the weak price representativeness due to the "one-specification-for-one-item" policy, called the broadly-defined lower-level substitution bias in Shiratsuka (2021).

The Japanese CPI conducts the price survey based on the "one-specification-for-one-item" policy. The policy specifies a few popular specifications for each item and continuously surveys their prices at specific outlets with large sales share in the local market. If very popular national brands exist, such as *Suntory Kaku* for whisky and *Kikkoman* or *Yamasa* for soy sauce, the CPI surveys such popular products at all the outlets. However, without such very popular national brands in the items, the CPI continues to survey the most popular product at each store, such as a standard product for bread, nuts, and castella. Such a price survey method makes it difficult to maintain price representativeness under the rapid and significant structural changes in the retail markets.

We carry out two additional empirical exercises regarding the effects of the coverage of store types and prices surveyed at an item level. First, we estimate the baseline specifications for the extended TPD model using data for GMS and SM, where the CPI price surveys for food and beverage products are mostly conducted. Fig. 16 depicts the estimation results for D-ROL_SK+TY_GM+SM and W-MO_SK+TY_GM+SM using data only for GM and SM, along with the baseline estimation results for D-ROL_SK+TY and W-MO_SK+TY.[16] The estimated quality-adjusted price indices exhibit very similar movements over time regardless of whether the data is restricted to GMS and SM, indicating that the differences in the data coverage for store types do not significantly influence the estimation results. Sales share of GMS and SM is fairly stable at around 80% (about 30% for GMS and 50% for SM) even under the first declaration of a state of emergency. In addition, the results suggest that the extended TPD model using store type dummies as cross-sectional fixed effects succeeds in adjusting the retail services quality differences across outlet channels.

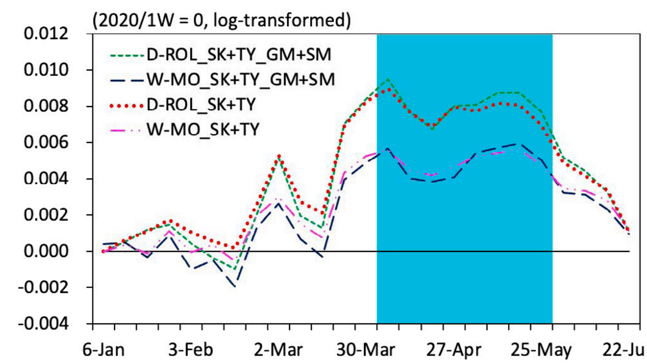

Fig. 16. Estimation results for GMS and SM. Notes: The light blue area indicates the period of the first declaration of a state of emergency in Tokyo. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Second, we construct proxy indicators of the CPI for food less perishables and eating out, using the INTAGE scanner data for GMS and SM: MPP-ITM and MPP-STR. In selecting products surveyed, MPP-ITM uses the most popular SKU in the items (common to all the stores), while MPP-STR uses the most popular SKU in the items for each store. At an elementary aggregation level, we employ geometric mean to produce item indices, considering the heterogeneity of unit prices across stores, especially in the case of MPP-STR. We then aggregate them into the overall indices with fixed weight for the whole sample period.

Fig. 17 plots the computed results for MPP-ITM and MPP-STR, along with the baseline estimation results for D-ROL_SK+TY and W-MO_SK+TY. Comparing the two proxy indicators, MPP-STR exhibits a much faster increase than MPP-ITM, comparable to D-ROL_SK+TY. Price increases become milder when restricting prices surveyed by strictly following the "one-specification-for-one-item" policy as in the case of MPP-ITM, while price increases become relatively higher when widening the specifications for prices surveyed by choosing the most popular SKU for items at each store as in the case of MPP-STR.

[16] We also confirm that almost the same estimation results are obtained using the store dummies as fixed effects.

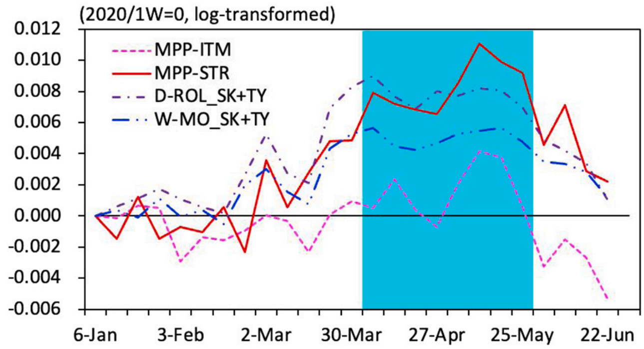

Fig. 17. Proxy indicators for the CPI. Notes: The light blue area indicates the period of the first declaration of a state of emergency in Tokyo. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

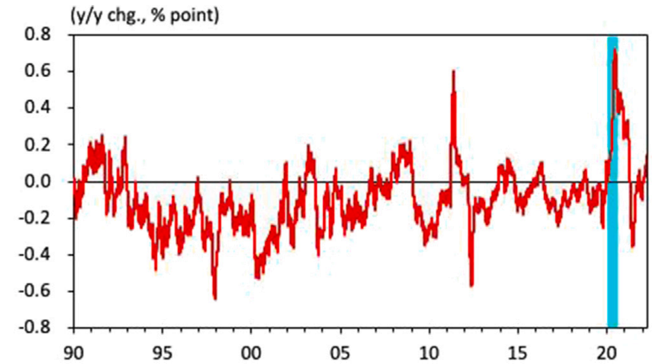

Fig. 18. Spreads between T-index and T-mode-index. Notes: The figure plots the divergence between year-on-year changes in T-index and T-mode-index. The light blue area indicates the first half of 2020, the period under primary analysis in this paper. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Source: Nowcast, CPINow.

This observation indicates the limitations of the current price survey method for the CPI based on the "one-specification-for-one-item" policy due to weakened price representativeness under large-scaled structural changes in the retail market. It is thus deemed effective to widen the coverage of prices surveyed in terms of products and outlets in tracing heterogeneous price responses across products and outlets.[17]

7. Conclusions

We examined the CPI measurement errors under the first wave of the COVID-19 spread in Japan. To address this question, we constructed high-frequency quality-adjusted price indices by employing daily scanner data from retail stores in Tokyo. We demonstrated the importance of using price data with the wide-ranging coverage of products and outlets by making explicit adjustments for the effects of temporary sales and retail service quality in examining the retail price dynamics under the COVID-19 pandemic as the voluntary lockdown constrained household purchasing behavior. Note that the sources of the CPI measurement errors under the COVID-19 turmoil differ significantly from those in the US, observed as wide-ranging and long-lasting stockouts.

We concluded that downward bias, not upward bias generally advocated, was observed during the first wave of the COVID-19 spread in Japan. The magnitude of the downward bias is estimated at from −0.6 to −0.3 points on the CPI for food less perishables and eating out on the basis of cumulative changes from January 2020 to June. The contribution of the estimates to the overall CPI is −0.3% to −0.15% points on an annualized basis, considering that the estimation covers half-year and the weights are about a quarter of the overall CPI. The magnitude of measurement errors is deemed limited, and the overall trend of the CPI remains unchanged even after incorporating the estimated downward bias. However, this downward bias arises mainly from the "one-specification-for-one-item" policy by weakening the price representativeness, called the broadly-defined lower-level substitution bias in Shiratsuka (2021).

Our empirical results revealed that the construction of quality-adjusted price indices becomes very difficult when facing large-scaled structural changes in the retail markets, including the frequency of temporary sales and the size of price reductions. As discussed in Section 1, the widening deviations between the two types of daily price indices—the T-index and the T-mode-index—indicate the possibility of structural changes in the retail markets. Fig. 18 plots the long time series of the deviations from 1990. As depicted in the figure, the period of our focus, January to June 2020, experienced the largest positive deviations between the two indices, suggesting that structural changes in retail markets were unprecedentedly large, although for a short period.[18]

However, it should be noted that the mid-1990s to the early 2000s also experienced significant and continued deviations between the two indices in a negative direction. This period corresponds to observing the "price busting" phenomenon when large-scale structural changes in the retail markets, symbolized by the expansion of large-scale retail stores, were advanced. As Shiratsuka (1999) has pointed out, the outlet substitution bias severely affected the Japanese CPI during that period.

We stress that the CPI needs to consider how to deal with possible large-scaled structural changes in the retail markets as it is hard to anticipate in advance. In that respect, it is deemed important to respond flexibly to such structural changes in the retail markets once it occurs. However, the current price survey method for the CPI based on the "one-specification-for-one-item" policy is unable to deal with that issue, as the method specifies a few most popular specifications for each item and continuously surveys their prices at specific outlets.[19] The most promising alternative price information is scanner data, which enables us to expand the coverage of products and outlets, as examined in this paper. Although scanner data cover about 25% of the CPI basket, these products represent a significant portion of the Retail Price Survey, which is the primary data source for the CPI. It is a future challenge to explore an estimation framework to produce stable results while flexibly incorporating the effects of structural changes with longer-term scanner data.

Data availability

The authors do not have permission to share data.

[17] We also compute MPP-ITM and MPP-STR by further restricting data for high sales in the items, such as outlets for the top 10% and 25% sales share. Such indices exhibit milder price increases, compared with the indices for all outlets, suggesting that price increases at smaller outlets were higher than those at large outlets under the COVID-19 turmoil. This observation suggests the importance of widening the coverage of price surveys in terms of outlets in dealing with large-scaled structural changes in the retail market.

[18] Another major divergence was observed immediately after the Great East Japan Earthquake in March 2011.

[19] Shiratsuka (2021) examined the lower-level substitution bias using the micro-data for the Retail Price Survey, which is the primary source data for the Japanese CPI. He revealed that the lower-level substitution bias stemming from the elementary aggregation formula is very limited but volatile in both positive and negative directions.

References

Abe, N., Inoue, T., Sato, H., 2020. Price Index Numbers Under Large-Scale Demand Shocks: The Japanese Experience of the COVID-19 Pandemic. RCESR Discussion Paper Series DP20-2, Research Center for Economic and Social Risks, Institute of Economic Research, Hitotsubashi University.

Abe, N., Tonogi, A., 2010. Micro and macro price dynamics in daily data. J. Monetary Econ. 57 (6), 716–728.

Cavallo, A., 2020. Inflation with Covid Consumption Baskets. Working Paper 27352, National Bureau of Economic Research.

Cavallo, A., Kryvtsov, O., 2021. What Can Stockouts Tell Us About Inflation? Evidence from Online Micro Data. Working Paper 29209, National Bureau of Economic Research.

Diewert, W.E., Fox, K.J., 2020. Measuring Real Consumption and CPI Bias under Lockdown Conditions. Working Paper 27144, National Bureau of Economic Research.

Eichenbaum, M., Jaimovich, N., Rebelo, S., 2011. Reference prices, costs, and nominal rigidities. Amer. Econ. Rev. 101 (1), 234–262.

Fujii, D., Nakata, T., 2021. COVID-19 and output in Japan. Jpn. Econ. Rev. 72 (4), 609–650.

de Haan, J., 2015. A framework for large scale use of scanner data in the dutch CPI. In: Paper Presented at the Fourteenth Ottawa Group Meeting, Tokyo, Japan.

de Haan, J., Hendriks, R., 2013. Online data, fixed effects and the construction of high-frequency price indexes. In: Paper Presented at the 2013 Economic Measurement Group, Sydney, Australia.

de Haan, J., Hendriks, R., Scholz, M., 2021. Price measurement using scanner data: Time-product dummy versus time dummy hedonic indexes. Rev. Income Wealth 67 (2), 394–417.

de Haan, J., Krsinich, F., 2014. Scanner data and the treatment of quality change in nonrevisable price indexes. J. Bus. Econom. Statist. 32 (3), 341–358.

Hosono, K., 2021. Epidemic and economic consequences of voluntary and request-based lockdowns in japan. J. Jpn. Int. Econ. 61, 101147.

Imai, S., Watanabe, T., 2014. Product downsizing and hidden price increases: Evidence from Japan's deflationary period. Asian Econ. Policy Rev. 9 (1), 69–89.

Kehoe, P., Midrigan, V., 2015. Prices are sticky after all. J. Monetary Econ. 75, 35–53.

Konishi, Y., Saito, T., Ishikawa, T., Kanai, H., Igei, N., 2021. How did Japan cope with COVID-19? Big data and purchasing behavior. Asian Econ. Pap. 20 (1), 146–167.

Krsinich, F., 2016. The FEWS index: Fixed effects with a window splice. J. Off. Stat. 32 (2), 375–404.

Nakamura, E., Steinsson, J., 2008. Five facts about prices: A reevaluation of menu cost models*. Q. J. Econ. 123 (4), 1415–1464.

Shibamoto, M., Hayaki, S., Ogisu, Y., 2022. COVID-19 infection spread and human mobility. J. Jpn. Int. Econ. 64, 101195.

Shiratsuka, S., 1995. Automobile prices and quality changes : A hedonic price analysis of Japanese automobile market. Monet. Econ. Stud. 13 (2), 1–44.

Shiratsuka, S., 1999. Measurement errors in the Japanese consumer price index. Monet. Econ. Stud. 17 (3), 69–102.

Shiratsuka, S., 2021. Lower-Level Substitution Bias in the Japanese Consumer Price Index: Evidence from Government Micro Data. Discussion Paper Series A No.722, Hitotsubashi University, Institute for Economic Research.

Sudo, N., Ueda, K., Watanabe, K., Watanabe, T., 2018. Working less and bargain hunting more: Macroimplications of sales during Japan's lost decades. J. Money Credit Bank. 50 (2–3), 449–478.

Ueda, K., Watanabe, K., Watanabe, T., 2019. Product turnover and the cost-of-living index: Quality versus fashion effects. Am. Econ. J.: Macroecon. 11 (2), 310–347.

Watanabe, T., Omori, Y., 2020. How Much Did People Refrain from Service Consumption due to the Outbreak of COVID-19?. CARF F-Series CARF-F-477, Center for Advanced Research in Finance, Faculty of Economics, The University of Tokyo.

Watanabe, T., Yabu, T., 2021. Japan's voluntary lockdown: further evidence based on age-specific mobile location data. Jpn. Econ. Rev. (Oxford England) 72, 333–370.