Webscraped Prices Comparison Method adopted for the PPI in Japan

— including interim outcomes of further analysis with the use of machine learning techniques —

15th Meeting of the Ottawa Group on Price Indices
held in Eltville am Rhein, Germany

May 10-12, 2017

Kimiaki Shinozaki

Bank of Japan

I. Overview

- The Bank of Japan conducted the rebasing of the PPI on February 2017. Taking that opportunity, the Bank introduced a new quality adjustment method, so-called **webscraped prices comparison method**, to electric appliances. It is a foolproof way to regard **50% of the price difference between old and new products as the quality improvement part** if its magnitude is not known.
- In this presentation, I demonstrate the appropriateness of the method empirically by targeting at individual products of durable consumer goods sold in Japan. (For details, see BOJ Working Paper #16-E-5).
- Also, I would like to show some interim outcomes of further analysis with the use of supervised machine learning techniques. The Bank compiles **experimental price indices** automatically from the webscraped data set by applying the Support Vector Machines (SVMs), programmed in Python.

II. Empirical Analysis

Data Sets

- Develop the unbalanced panel data sets by integrating the following:
  1. Product specifications: registered at the ***Kakaku.com*** between December 2012 and December 2015.
  2. Weekly average prices: registered at the paid ***Kakaku.com Trend Search*** between December 2013 and December 2015.
- Coverage:

  Home electrical appliances: 8 commodities

  Digital consumer electronics: 12 commodities
- Data Volume:

  Number of products: 4,500

  Size of panel data: 150,000

  Total data volume: 5.6 million

Obtained from the ***Kakaku.com*** charged website by applying the **webscraping technology**.

II. Empirical Analysis

Estimation of Hedonic Functions

- We estimate the following **semi-logarithmic hedonic functions** with a dummy variable to control the elapse of time from the launch of products to capture the price transition through the product life-cycle.

$$\ln(p_{i,t}) = \alpha + \sum_{k} \beta_k X_{i,k} + \sum_{\tau} \gamma_\tau D_t(\tau_i + \tau) + \sum_{\tau} \delta_\tau D_t(\tau) + \varepsilon_{i,t}$$

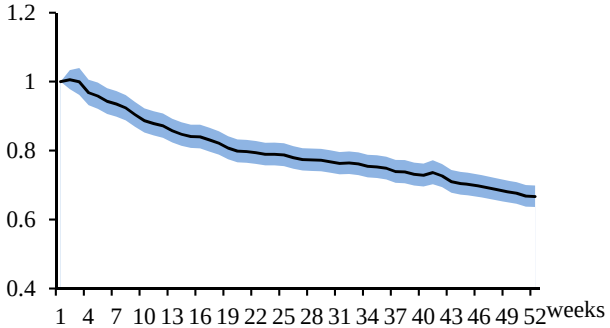
$$D_t(T) = \begin{cases} 1 & (\text{if } \ t = T) \\ 0 & (\text{if } \ t \neq T) \end{cases}$$

$p_{i,t}$: price of product i at time t, $X_{i,k}$: kth specification of product i
$D_t(\tau_i + \tau)$: dummy variable to control the elapsed weeks from the launch of product
$D_t(\tau)$: dummy variable to control macroeconomic shocks in each quarter

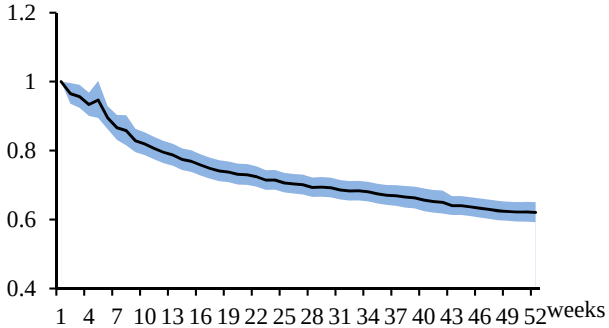
- Then we match pairs of products based on the following criteria: (a) The launch date of a new product is later than that of the old product, (b) old and new products are made by the same manufacturer, etc.

Pricing Patterns over Product Life-Cycle: Home Electrical Appliances

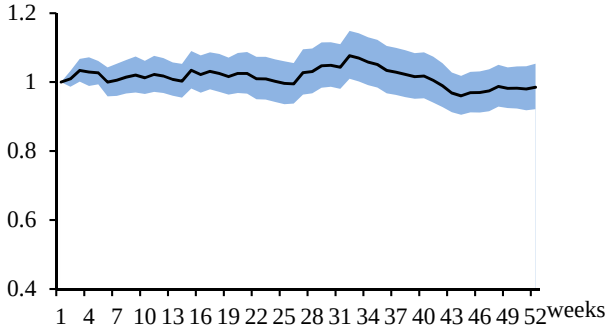
(1) Air conditioners



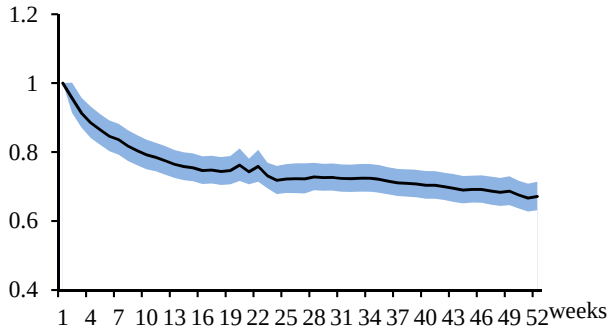
(2) Refrigerators and freezers



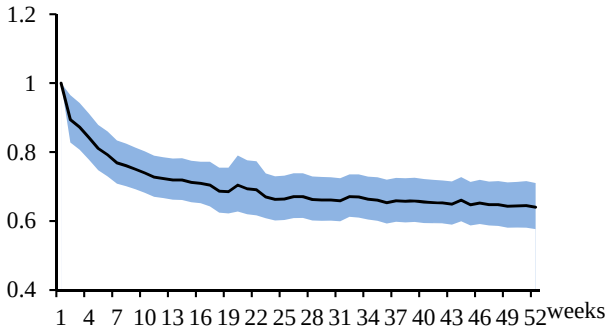
(3) Washers and dryers



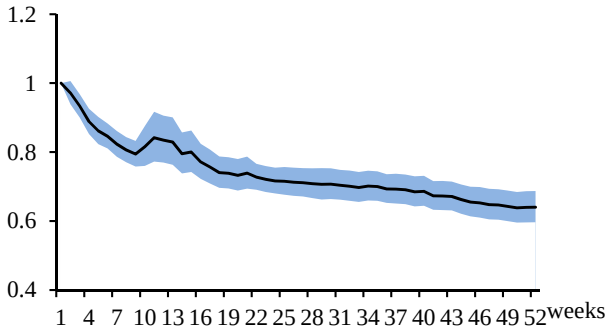
(4) Rice cookers



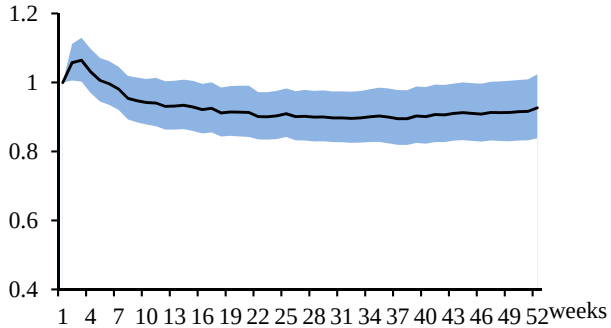
(5) Vacuum cleaners



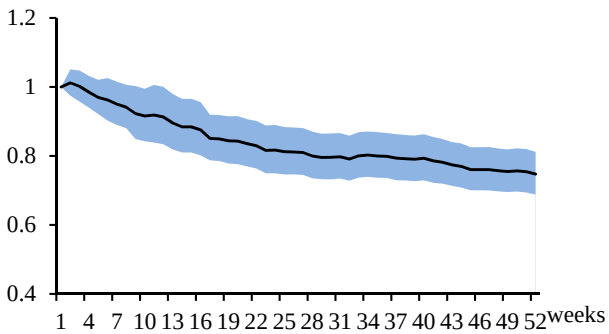
(6) Microwaves



(7) Hair dryers and curling irons



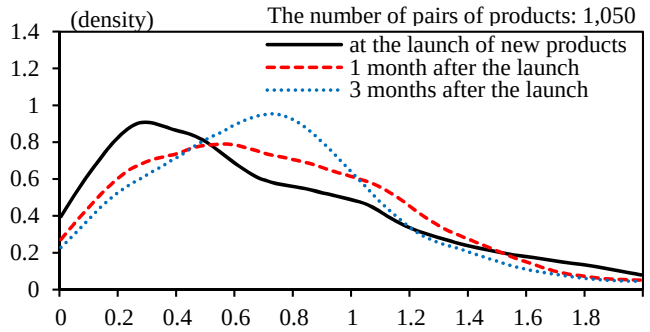
(8) Air purifiers



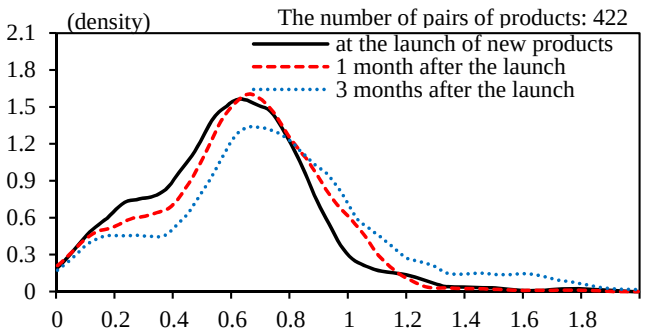
Note: The vertical axis indicates the **coefficient estimate of week dummy variables** with exponential transformation ($\exp(\hat{\gamma})$) as time proceeds.

Distribution of Quality Improvement Ratios: Home Electrical Appliances

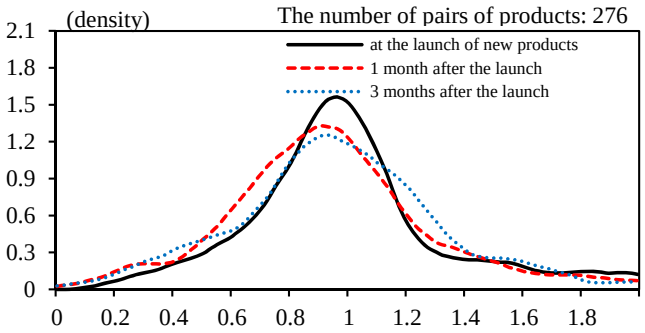
(1) Air conditioners



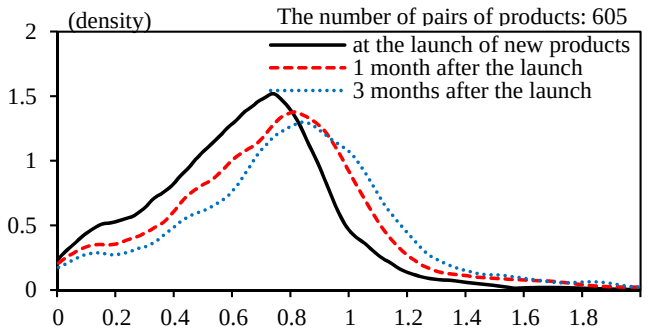
(2) Refrigerators and freezers



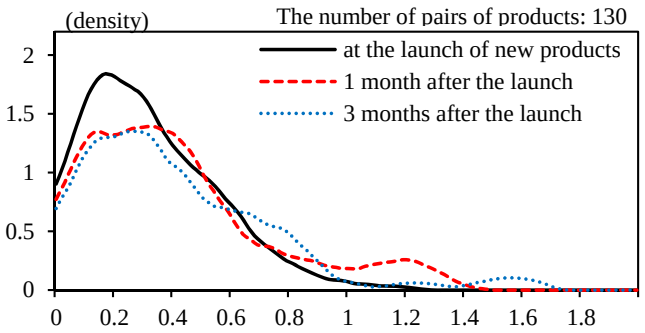
(3) Washers and dryers



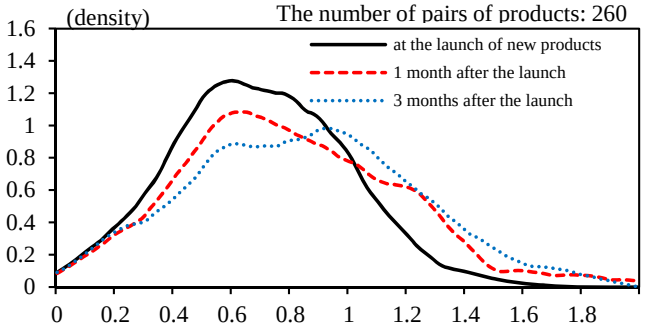
(4) Rice cookers



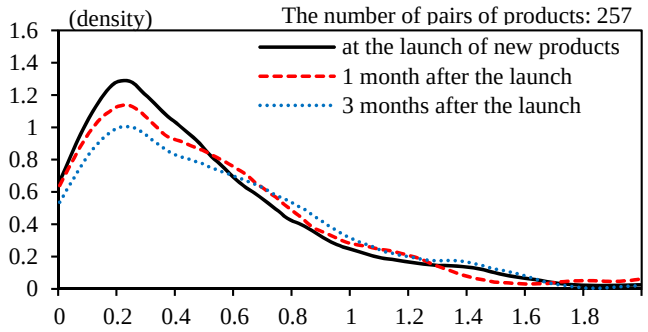
(5) Vacuum cleaners



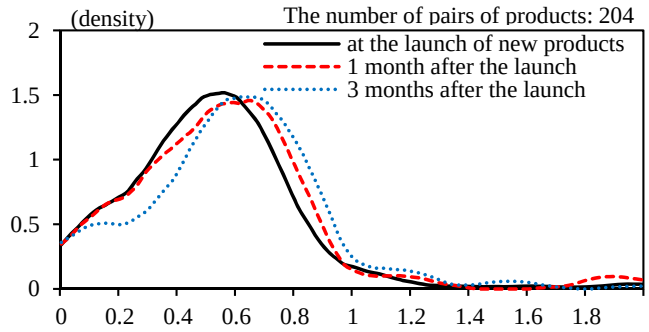
(6) Microwaves



(7) Hair dryers and curling irons



(8) Air purifiers

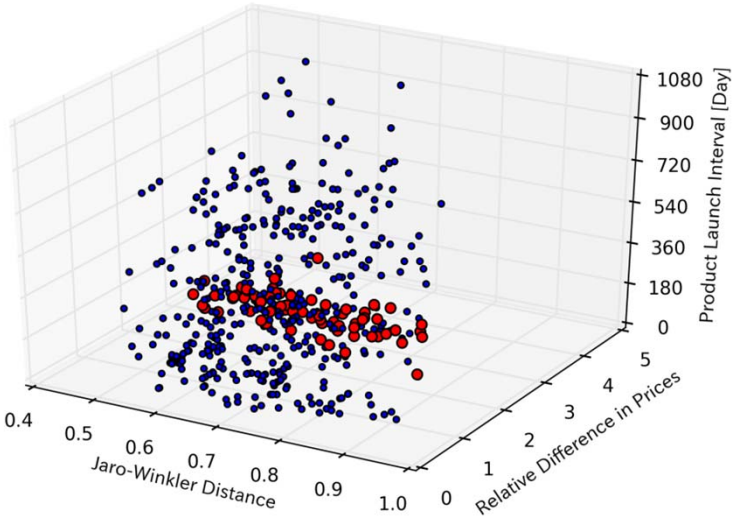


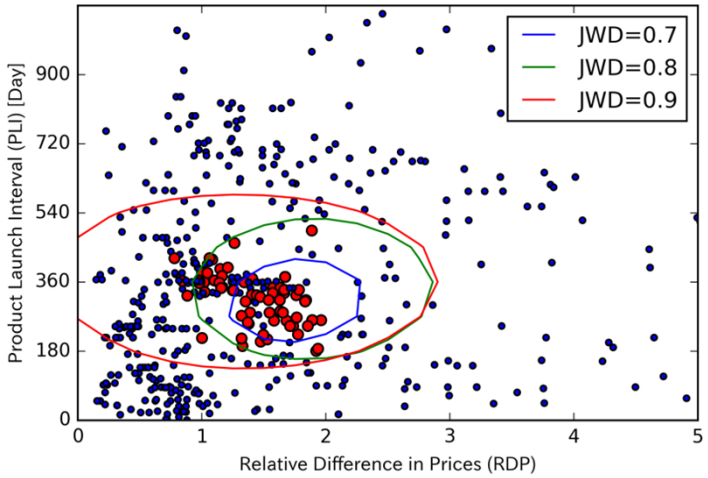
Note: The horizontal axis indicates the **Quality Improvement Ratios** of individual pairs of old and new products defined as follows:

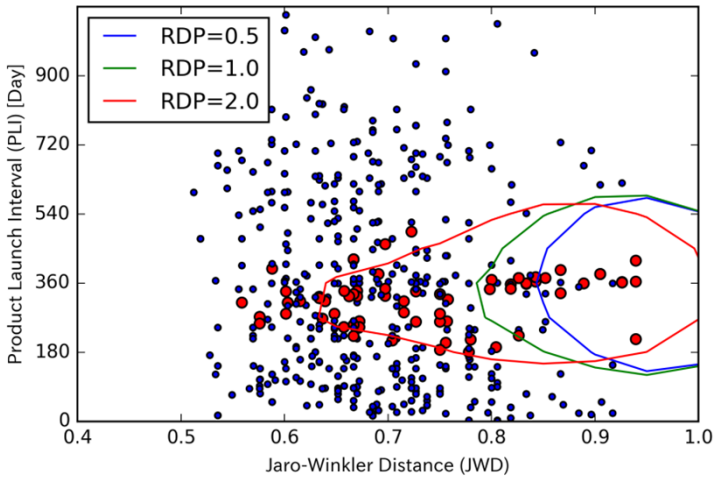
$$\mu_{\tau}^{i,j} \equiv \frac{\sum_k \beta_k \left(X_{j,k} - X_{i,k}\right)}{\ln\left(p_{j,\tau_j+\tau}\right) - \ln\left(p_{i,\tau_j+\tau}\right)}$$

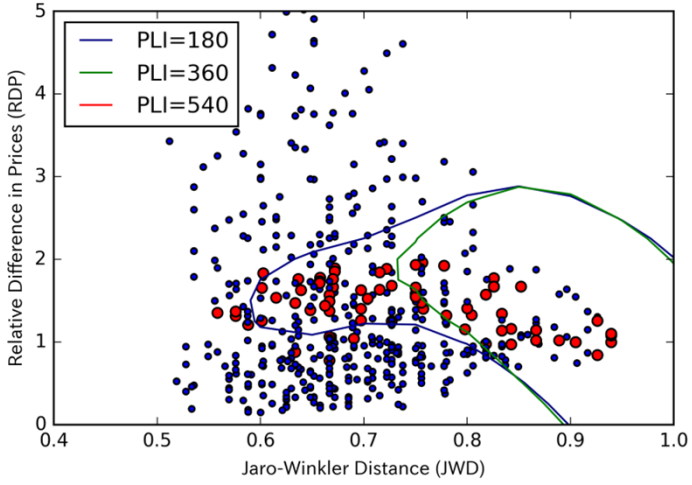
III. Interim Outcomes of Further Analysis

Automated pairing of products for “Refrigerators and Freezers”









To find true pairs of old and new products, we set up the three features for the SVMs: (1) **Jaro-Winkler Distance**, (2) **Relative Difference in Prices**, and (3) **Product Launch Interval** for each pair of products to enable **automated pairing for webscraped data sets**.

Jaro-Winkler Distance is a kind of the edit distance that quantifies similarities of names between old and new products.

In scatter plots, **Red circles indicate the TRUE pairs of products** while **Blue circles are the FALSE ones**.

III. Interim Outcomes of Further Analysis

Experimental price indices for “Refrigerators and Freezers”

- We make up the experimental price indices by a variety of quality adjustment methods. The index compiled with **Webscraped prices comparison method** is similar to that with **Hedonic method**.
- The detailed methodology and complete outcomes of the further analysis will be released in mid-2017. The Bank of Japan and I would like to present it to you in the near future!