Appendix 2 Calculation of price indices of items using POS information

1 Calculation of price indices of items using the hedonic approach

The hedonic approach is one of the methods to adjust quality. This approach is used to identify the qualitative aspect causing price differences between products quantitatively by analyzing the relationship between characteristics (e.g. memory capacity, built-in memory capacity, and display size in case of PCs) and the price of a product using a statistical technique called multiple linear regression analysis on the assumption that the quality of a product can be broken down to a multiple characteristics (performance) that collectively compose the product, and the performance determines the price.

A method to directly evaluate the quality-adjusted price movement through the hedonic approach is used in the CPI for four items having a very short product cycle due to rapid upgrading of quality. These are "TV sets", "Personal computers (desktop)", "Personal computers (notes)", and "Cameras". A significantly large amount of data on the price, quantity, and characteristics of a large number of products is required to make the multiple linear regression analysis highly objective and reliable. Scanner data on sales prices, sales volume, and characteristics of all models from the POS information[51] are used in the hedonic approach when applied to analyze these items.

<Method of index calculation>

(1) A semi-log regression model is formulated with the average sales price of each model as the explained variable, and memory capacity, display size, element size, and other characteristics of each model, as well as its sales period, as the explanatory variables[52].

Multiple linear regression formula using POS information for two consecutive months (t - 1, t)

$$\ln(p_{T,i}) = \alpha + \beta \cdot \delta_{T,t} + \sum_{k} (\gamma_k \cdot x_{k,i}) + \varepsilon_{T,i}$$

T: Month $= t-1$、t　　k: Characteristics used as explanatory variables　　i: Model

$p_{T,i}$: Sales Price　　α、β、γ_k: Partial regression coefficient

$x_{k,i}$: Variables representing characteristic

$\delta_{T,t} = \begin{cases} 0 & (T = t-1) \\ 1 & (T = t\ ) \end{cases}$: Dummy variables for sales period　　$\varepsilon_{T,i}$: Residual error

(2) All models sold[53] in the current month (t) and the previous month (t-1) are subject to regression calculation using the regression model in (1) with the total sales volume of each sold model as the weight to determine the estimated value of the average price (logarithm) for each month

---

[51] Monthly average sales price for each model. The averages of sales prices for each model for one week in the middle of a month are used for the preliminary figures for the Ku-area of Tokyo.

[52] The explanatory variables are revised at an appropriate frequency of about one year. The details of the regression model such as the explanatory variable are explained in the Annual Report on the CPI.

[53] The models sold until the previous month, and the models sold from current month are included.

from the estimated value of the partial regression coefficient.

$$\text{(Previous month)}\quad \ln(P_{t-1}) = \hat{\alpha} + \hat{\beta}\cdot 0 + \sum_k \left(\hat{\gamma}_k \cdot \bar{x}_{k,t-1}\right)$$
$$\text{(Current month)}\quad \ln(P_t) = \hat{\alpha} + \hat{\beta}\cdot 1 + \sum_k \left(\hat{\gamma}_k \cdot \bar{x}_{k,t}\right)$$

(Hat "^" means an estimated value, and bar "‾" means an average value.)

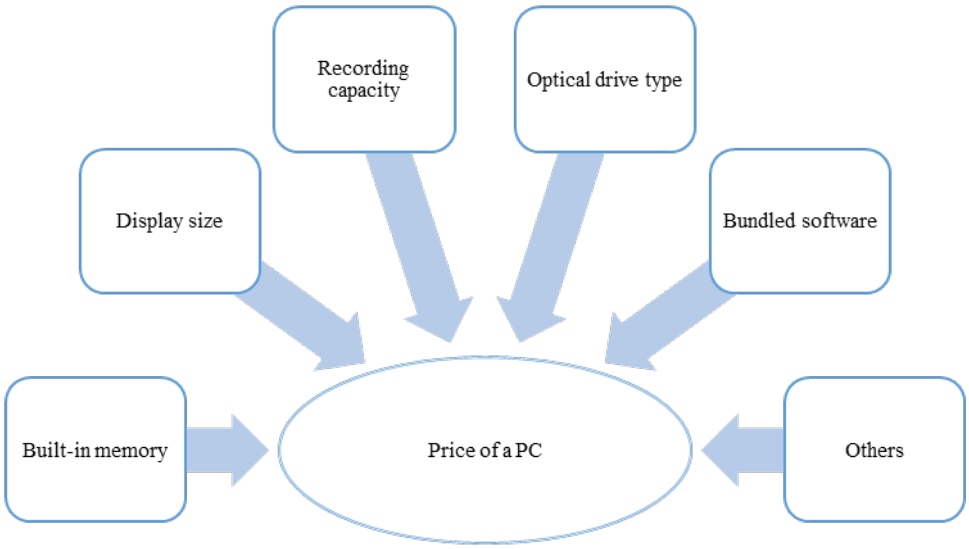
(3) The ratio of prices obtained by excluding the quality difference between the previous month (t-1) and the current month (t) from the estimated value of the average price obtained in (2) above is calculated as the link index.

$$\frac{P_t}{P_{t-1}} = \frac{\exp\left[\hat{\alpha} + \hat{\beta} + \sum_k \left(\hat{\gamma}_k \cdot \bar{x}_{k,t}\right)\right]}{\exp\left[\hat{\alpha} + \sum_k \left(\hat{\gamma}_k \cdot \bar{x}_{k,t-1}\right)\right]}$$
$$= \exp\left[\hat{\beta}\right] \times \exp\left[\sum_k \hat{\gamma}_k \cdot \left(\bar{x}_{k,t} - \bar{x}_{k,t-1}\right)\right]$$
$$I_t^{(L)} = \exp\left[\hat{\beta}\right]$$

(4) The link index calculated in (3) above is multiplied by the index of the previous month (t-1) (2020 = 100) to find the chain index of current month (t).

$$I_t^{(C)} = I_{t-1}^{(C)} \times I_t^{(L)}$$

<Example of quality adjustment using the hedonic approach (personal computers)>



1) Correlations between characteristics and price are analyzed using a large amount of PC sales data.
2) For example, a relation of "1 TB increase of memory capacity makes a 5.0% rise of the price of PC" can be estimated.

2 Calculation of price indices of items using the fixed specification method

The fixed specification method uses the characteristic information of each model from the POS information data (scanner data), and only products which have certain fixed specifications - for example, tablet computers which have a storage capacity of 64 GB and a display size of 10.2 inch - are selected from the POS information data. By this method, quality of the products become constant.

The fixed specification method is similar to "detailed specifications" in Retail Price Survey, but the fixed specification method leads the price movement to become more stable than detailed specifications because more models are likely to be selected than with detailed specifications. When changing a trademark due to the introduction of a new product such as a successor model, if it matches the selected specifications, it can be automatically extracted from the month of release. Therefore, quality fixation by the fixed specification method is easier to calculate and simpler than quality adjustment by the hedonic approach and index calculation by a complicated formula.

In the CPI, the fixed specification method is used to make indices for three items of "Video recorders", "Tablet computers", and "PC printers", which have less quality difference between old and new products, and whose prices can be explained with few characteristics.

<Method of index calculation>

(1) The representative characteristics and specifications that determine the price and quality of each model are selected[54], and for all models sold in the current month (t) and the previous month (t-1), only the models that meet all the selected specifications are extracted.

$X_k = \{i | i \in Z, \text{Selection specification conditions (Pk) for characteristic k} = P(x_{k,i})\}$

$$S = \bigcap_{k=1}^{K} X_k = X_1 \cap X_2 \cap \cdots \cap X_K$$

k: Characteristic i: Model $x_{k,i}$: Specification

P_k: Selection specification condition for characteristic k

Z: Set of models included in POS information

(2) For the models extracted in (1) above, the weighted geometric mean price is calculated for the current month (t) and the previous month (t-1) respectively with the sales volume of each model as the weight.

$$P_T = \left(\prod_{i \in S(T)}^{N} {p_{T,i}}^{q_{T,i}} \right)^{\frac{1}{\sum_i q_{T,i}}} = \exp\left[\frac{1}{\sum_i q_{T,i}} \sum_{i \in S(T)}^{N} \left(q_{T,i} \cdot \ln(p_{T,i}) \right) \right]$$

T: Month$= t-1$、t i: Model

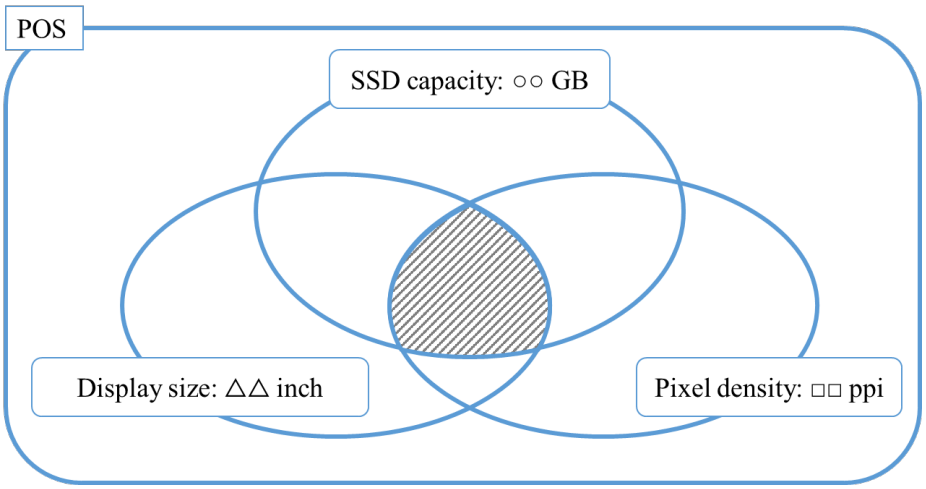
$p_{T,i}$: Sales Price $q_{T,i}$: Sales Volume $S(T)$: Set of models selected

[54] The characteristics and their specifications are reviewed roughly once a year by taking into account changes in the hot-selling line of products. The selected characteristics and specifications are published in the Annual Report on the CPI of the year that has a change in those.

(3) The average ratio of prices between the current month (t) and the previous month (t-1) is used as the link index, which is then multiplied by the index of the previous month (t-1) (2020 = 100) to find the chain index of the current month (t).

$$I_t^{(L)} = \frac{P_t}{P_{t-1}}$$
$$I_t^{(C)} = I_{t-1}^{(C)} \times I_t^{(L)}$$

<Example of model extraction by fixed specification method (tablet computers)>



1) Characteristics that correlate with the price of tablet computers, such as SSD capacity and display size, are selected.
2) The main specifications for each characteristic are selected.
3) The selected specifications are fixed, and models that match the specifications from the monthly POS information data are extracted.