

**SUPPORTING PAPER No 7 (18)\***  
**24 October 2001**

**ENGLISH ONLY**

**STATISTICAL COMMISSION and  
ECONOMIC COMMISSION FOR EUROPE**

**INTERNATIONAL LABOUR  
ORGANISATION (ILO)**

**CONFERENCE OF EUROPEAN  
STATISTICIANS**

**Joint ECE/ILO Meeting  
on Consumer Price Indices  
(Geneva, 1-2 November 2001)**

**APPLICATION OF MIDPOINT-PERIOD METHOD TO COMPILATION OF HEDONIC  
INDICES FROM SCANNER DATA**

Contributed Paper submitted by Statistics Bureau of Japan\*\*

---

\* Paper posted on Internet as submitted by the country.

\*\* Prepared by Masato Okamoto, Consumer Statistics Division, Statistics Bureau, Tokyo, Japan. The views expressed in this paper are solely those of the author and do not necessarily reflect the policies of the Statistics Bureau of Japan.

## Summary

A similar method to midpoint-year basket index, which is used for compilation of a supplementary index in the 2000-base CPI (Okamoto [2001]), is tried as a procedure for compilation of hedonic indices from scanner data, aiming to clarify the following points:

- i). Hill [1999]'s assumption – i.e. smooth change in consumption pattern, or rather smooth change of hedonic function in this case – also stands up all right in elementary index level.
- ii). A month-to-month chaining of price relatives to the previous month estimated from scanner data using hedonic method, which is adopted as the procedure for compilation of price index for PCs in the 2000-base CPI, does not drift away from the proper index.

The study is also carried out for the following practical reason in expectation of possibility of cost-savings such as reduction of the burden to staffs in charge:

- iii). Unbiased hedonic indices can be obtained without every-month update of hedonic regression parameter estimates.

The results indicate the midpoint-period method yields a good approximation to monthly-chained hedonic index if updating hedonic parameters at proper intervals in general, and support the assumption i), ii) and iii). Although the method may be unsuitable for practical use because of its complicatedness, it may be usable as a tool of bias check on hedonic indices estimated by other procedures.

1. Okamoto and Sato [2001] studied application of hedonic approach to compilation of elementary indices for PCs, color TVs and digital (still) cameras from scanner data. Good results obtained from the study led to a decision to adopt hedonic approach for compilation of the official price index for PCs from scanner data in the 2000-base CPI. While carefully considering appropriateness to widen application of hedonic approach and use of scanner data, the possibility of methodological development is also studied in order to make it possible to apply hedonic approach to many categories under limited resources. This study is tied up with purely intellectual interest in applicability of the midpoint-period method to compilation of elementary indices.

2. In this paper, three types of procedures for compilation of hedonic indices including a kind of midpoint-period method are discussed using hedonic regression parameter estimates for PCs and color TVs derived from monthly scanner data.<sup>1</sup>

## Methodology

### - Hedonic regression model used

---

<sup>1</sup> Variables incorporated into the regression models are listed in Okamoto and Sato [2001].

3. Hedonic regression parameters of a semi-log model (1) are estimated by the weighted least square method. Weights are proportionate to numbers of units by product sold. A semi-log model is selected mainly because the coefficient of determination tends to be higher than that of the log-log model although hedonic indices obtained from both models are close to each other as shown in Chart A. Variable for (square of) display size seems to affect the choice of regression model.<sup>2</sup>

$$\ln p_m = \alpha_m + \sum_k \beta_{m,k} x_{m,k} + \varepsilon \quad (1)$$

where

$p_m$  : average unit price of each product in month  $m$

$x_{m,k}$  : variable for characteristic  $k$  in month  $m$

$\alpha_m, \beta_{m,k}$  : partial regression coefficients in month  $m$

4. Parameter estimates obtained from the regression model (1) satisfies the following equation.

$$\ln \bar{p}_m = \alpha_m + \sum_k \beta_{m,k} \bar{x}_{m,k} \quad (2)$$

where

$\bar{p}_m$  : exp[(weighted) average of  $\ln p_m$ ] - i.e. (weighted) geometric mean of  $p_m$

$\bar{x}_{m,k}$  : (weighted) average of characteristic  $k$  in month  $m$

#### - Index compilation based on the base month's parameter estimates (Method B)

5. First, a price relative of the observation month to the base month is calculated using regression parameter estimates for the base month with average price and average specifications in the observation month as follows.

<sup>2</sup> It is known that (display size)<sup>2</sup> is a better variable than display size. As (display size)<sup>2</sup> is actually identical to display size in the log-log model - i.e.  $\ln(\text{display size})^2 = 2\ln(\text{display size})$ ,  $[\ln(\text{display size})]^2$  was tried instead of  $\ln(\text{display size})$  but the results were not better than the semi-log model. A non-log linear model as shown in formula (18), which corresponds to a non-log & non-linear model with time dummy variables discussed by Diewert [2001] (formula (22) in his paper), did not yield higher coefficients of determination either. Price level of new products relative to the existing equivalent products and its changes seem to cause not outstanding performance of the linear model. (See Chart A.)

$$\bar{p}_0 = \exp \left( \alpha_0 + \sum_k \beta_{0,k} \bar{x}_{t,k} \right) \quad (3)$$

$$BI_{t/0} = \frac{\bar{p}_t}{\bar{p}_0} \quad (4)$$

where

$\alpha_0, \beta_{0,k}$  : partial regression coefficients in the base month 0

$\bar{x}_{t,k}$  : (weighted) average of charachtersitc  $k$  in month  $t$

$\bar{p}_t = \exp[(\text{weighted}) \text{ average of logarithm of } p_t]$

6. The above calculation is performed with periodical change of the base month, and price relatives are chained as follows.

$$t = NL + m \quad (5)$$

$$B_t = BI_{(NL+m)/NL} \prod_{n=1}^N BI_{nL/(n-1)L} \quad (6)$$

where

$L$  : interval for the base month's update

$m < L$

**- Index compilation based on both the base month's & the observation month's parameter estimates (Method D)**

7. Two price relatives of the observation month to the base month are calculated using regression parameter estimates for the base month with average of price and specifications in the observation month and vice versa. Then, a geometric mean of the two price relatives is taken.

$$\bar{p}_0 = \exp\left(\alpha_0 + \sum_k \beta_{0,k} \bar{x}_{t,k}\right) \quad (7)$$

$$\hat{p}_t = \exp\left(\alpha_t + \sum_k \beta_{t,k} \bar{x}_{0,k}\right) \quad (8)$$

$$DI_{t/0} = \sqrt{\frac{\bar{p}_t}{\bar{p}_0} \cdot \frac{\hat{p}_t}{\hat{p}_0}} \quad (9)$$

where

$\alpha_t, \beta_{t,k}$  : partial regression coefficients in month  $t$

$\alpha_0, \beta_{0,k}$  : partial regression coefficients in the base month 0

$\bar{x}_{t,k}$  : (weighted) average of charactersitc  $k$  in month  $t$

$\bar{x}_{0,k}$  : (weighted) average of charactersitc  $k$  in the base month 0

$\bar{p}_t = \exp[(\text{weighted}) \text{ average of logarithm of } p_t]$

$\hat{p}_0 = \exp[(\text{weighted}) \text{ average of logarithm of } p_0]$

8. The above calculation is performed with periodical change of the base month, and price relatives are chained as follows.

$$t = NL + m \quad (10)$$

$$D_t = DI_{(NL+m)/NL} \prod_{n=1}^N DI_{nL/(n-1)L} \quad (11)$$

where

$L$  : interval for the base month's update

$m < L$

9. This method is identical to the 'single month method – geometric-mean' in Okamoto and Sato [2001] in the case that the base month is changed every month. 'Method D - every month' – i.e. the 'single month method – geometric-mean' - is used as the benchmark in this paper.

#### - Index compilation based on the midpoint-month's parameter estimates (Method M)

10. A price relative of the midpoint month of an earlier month and a later month to the earlier month is calculated using regression parameter estimates for the midpoint month with average of price and specifications in the earlier month. Another price relative of the later month to the midpoint month is calculated using regression parameter estimates for the midpoint month with average of price and specifications in the later month. Then, the product of the two price relatives is taken.

$$\hat{p}_L = \exp\left(\alpha_L + \sum_k \beta_{L,k} \bar{x}_{L-m,k}\right) \quad (12)$$

$$\bar{p}_L = \exp\left(\alpha_L + \sum_k \beta_{L,k} \bar{x}_{L+m,k}\right) \quad (13)$$

$$MI_{(L+m)/(L-m)} = \sqrt{\frac{\hat{p}_L \cdot \bar{p}_{L+m}}{\hat{p}_{L-m} \cdot \bar{p}_L}} \quad (14)$$

where

$\alpha_L, \beta_{L,k}$  : partial coefficients in the midpoint month  $L$

$\bar{x}_{L+m,k}$  : (weighted) average of charactersitc  $k$  in a later month  $L + m$

$\bar{x}_{L-m,k}$  : (weighted) average of charactersitc  $k$  in an earlier month  $L - m$

$\hat{p}_{L-m} = \exp[(\text{weighted}) \text{ average of logarithm of } p_{L-m}]$

$\bar{p}_{L+m} = \exp[(\text{weighted}) \text{ average of logarithm of } p_{L+m}]$

11. The above calculation is performed with periodical change of the midpoint month, and price relatives are chained as follows.

$$t = NL + m \quad (15)$$

$$M_t = MI_{(NL+m)/(NL-m)} M_{NL-m} \quad (16)$$

where

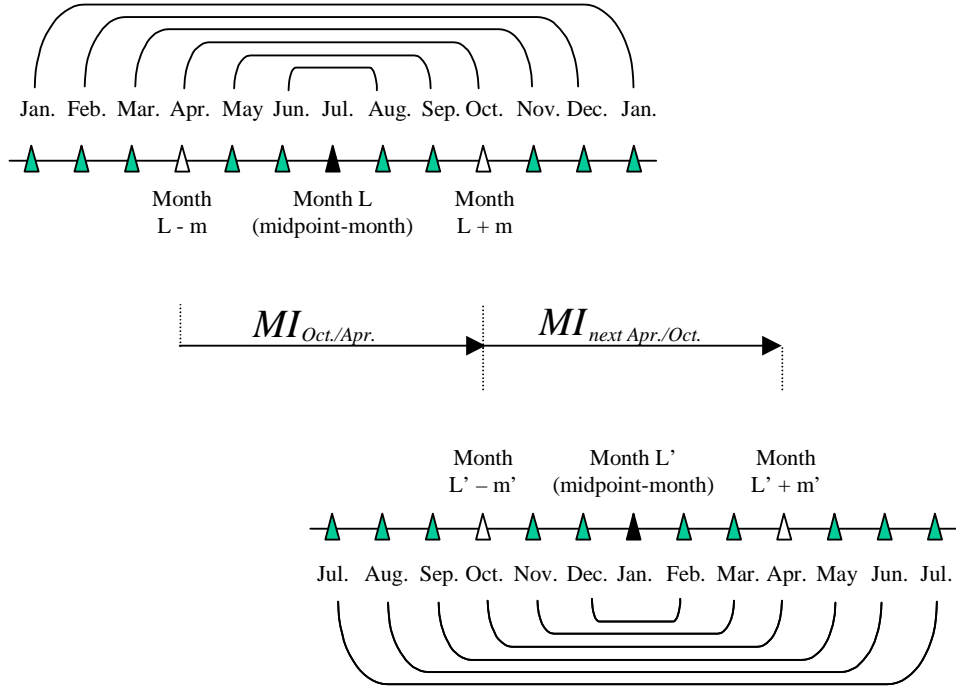
$L$  : interval for the midpoint - month's update

$m < L$

12. Note that indices for the first  $L+1$  months from month 0 to month  $L$  must be given beforehand as initial values. Two sets of initial values are used for the test calculation. One set is a series of indices derived from Method D ('Method M - d'). Another set is a series of indices derived from Method B ('Method M - b'). The following chart illustrates the procedure in the case that the 'midpoint month' is renewed every six months.

**Chained hedonic index derived from Method M  
(in the case that the 'midpoint month' is changed every six months)**

Page 7



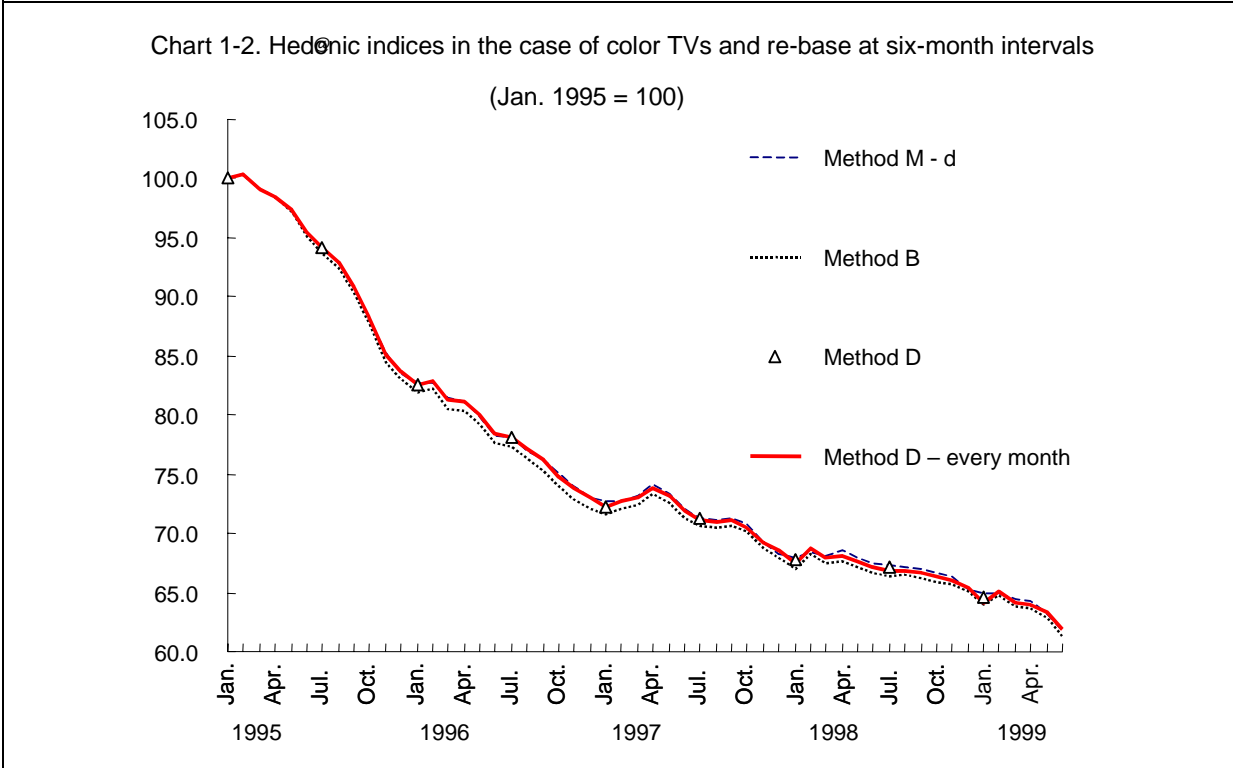
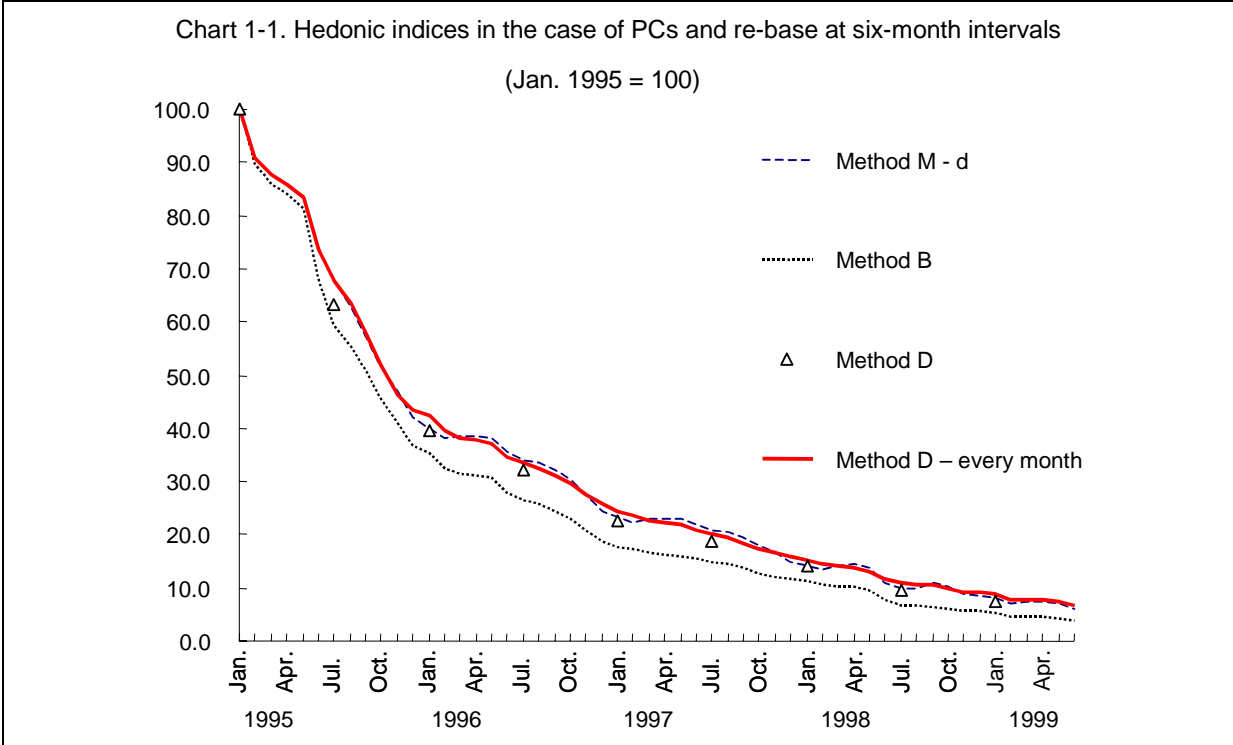
**Results**

13. The test calculation is performed using monthly scanner data from January 1995 to June 1999. The base month and the midpoint-month are renewed every month, every three months or every six months. Results are shown in Chart 1-1 and 1-2 below, Chart 2-1 – 3-2 and Table 1-1 – 3-2 annexed on page 11 - 18. The Chart 1-1 and 1-2 shows results in the case of periodical re-base at six-month intervals.

14. (Chained) hedonic indices derived from Method B, which measures a price change of an average specification in a later month based on a relation between prices and specifications in an earlier month, tend to be lower than that of 'Method D - every month', which measures a price change of an average specification in an earlier month based on a relation between prices and specifications in a later month and vice versa, and takes a geometric mean of the two. However, Method B seems fairly good in the case of color TVs and re-base at three-month or six-month intervals as shown in Chart 1-2 and 2-2 while monthly re-base causes noticeable downward bias as shown in Chart 3-2.

15. (Chained) hedonic indices derived from Method M, which measures a price change of an average specification in an earlier month between an earlier month and the midpoint-month based on a relation between prices and specifications in the midpoint-month, and a price change of an average specification in a later month between the midpoint-month and a later month based on a relation between prices and specifications in the midpoint-month, and takes the product of the two, is close to that of 'Method D - every month' in the case of color TVs. Method M also yields a good approximation to hedonic indices derived from 'Method D - every month' in

the case of PCs and re-base at every-month or six-month intervals as shown in Chart 1-1 and 3-1. However, it comes apart from the benchmark in the case of PCs and re-base at three-month intervals as shown in Chart 2-1. Those results indicate Method M yields a good approximation to ‘Method D - every month’ unless quality changes as fast as PCs.



16. As Method M requires initial values, two sets of initial values are tried. ‘Method M – d’ uses indices derived from ‘Method D - every month’, and ‘Method M – b’ uses indices derived from Method B with re-base at the same intervals. However, both yield almost the same indices on annual averages as shown in Table 1-1 – Table3-2. Thus, Method M seems insensitive to initial values.

17. Method M yields good approximations to ‘Method D - every month’ even if a different regression model is used as shown in Chart A. However, since (chained) hedonic indices derived from a non-log linear hedonic model (17) substantially differ from those of a semi-log model (1) and a log-log model in the case of color TVs, good performance of Method M does not mean appropriateness of the regression model used.

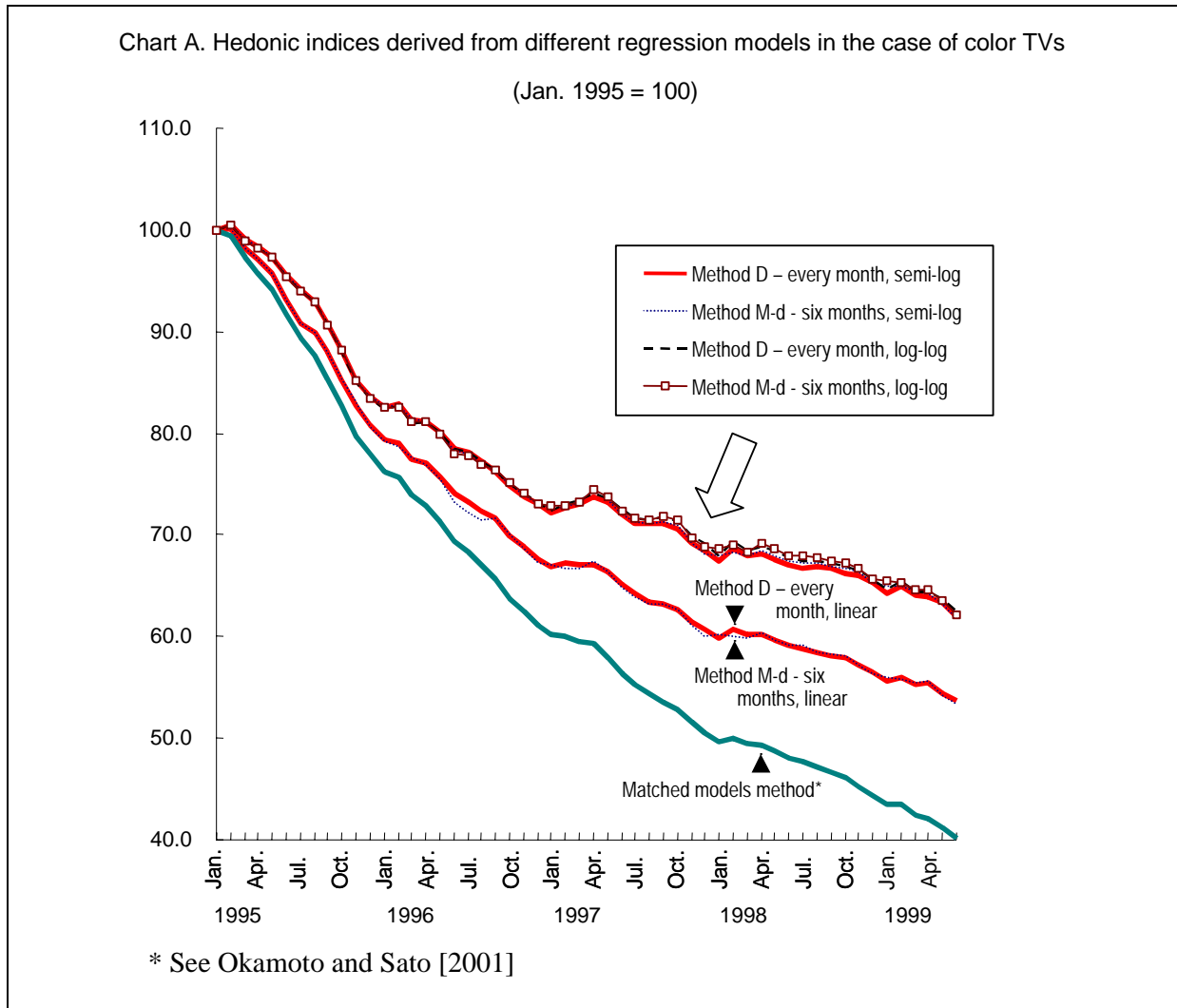
$$p_m = \alpha_m + \sum_k \beta_{m,k} x_{m,k} + \varepsilon \quad (17)$$

where

$p_m$  : average unit price of each product in month  $m$

$x_{m,k}$  : variable for characteristic  $k$  in month  $m$

$\alpha_m, \beta_{m,k}$  : partial regression coefficients in month  $m$



**Conclusion**

18. Good results of application of midpoint-period method to compilation of hedonic indices from scanner data can be regarded as supporting Hill [1999]’s assumption – i.e. smooth change in consumption pattern, or rather smooth change of hedonic function in this case – also stands up all right in elementary index level. If taking a look from another angle, the results indicate a month-to-month chaining of price relatives to the previous month, estimated from scanner data using ‘Method D - every month’, does not have a nature of drift away from the proper index resulting from iteration of linking.

19. If updating hedonic regression parameter estimates at proper intervals, Method M probably yields good approximations to ‘Method D - every month’ in the case of ordinary electric appliances such as TVs of which quality tends to change not so fast as PCs. It means unbiased hedonic indices can be obtained without monthly update of hedonic regression parameter estimates in general. However, this method may be difficult to put into practice because of its complicatedness although a burden of sensitive work for maintaining hedonic function may

be reduced if adopting Method M. In addition, it cannot be deniable so far that peculiar fluctuation observed in the case of PCs as shown in Chart 2-1 may occur in the case of ordinary electric appliances also. One possible practical use is for bias check on hedonic indices estimated by other procedures.<sup>3</sup>

20. If downward bias produced by Method B is within tolerance level, this method may be preferable because of its simplicity. The results indicate Method B may be a choice in the case of ordinary electric appliances if appropriate re-base intervals are known beforehand.

## REFERENCES

Hill, T. P. [1999], *Inflation, the Cost of Living and the Domain of a Consumer Price Index*, Joint ECE/ILO Meeting on consumer price indices, Geneva, November 1999.

Okamoto, Masato [2001], *Midpoint-Year Basket Index as a Practical Approximation to Superlative Indexes*, the Sixth Meeting of the International Working Group on Price Indices, Canberra, Australia, April 2001.

Okamoto, Masato, and Sato, Tomohiko [2001], *Comparison of Hedonic Method and Matched Models Method using Scanner Data: The Case of PCs, TVs and Digital Cameras*, the Sixth Meeting of the International Working Group on Price Indices, Canberra, Australia, April 2001.

Diewert, Erwin [2001], *Comment on the Measurement of Quality-Adjusted Price Changes by Mick Silver and Saeed Heravi*, the Sixth Meeting of the International Working Group on Price Indices, Canberra, Australia, April 2001.

---

<sup>3</sup> As noted in para 17, the fact that Method M gives good approximates to 'Method D - every month' does not indicate appropriateness of the regression model selected from non-log linear model, semi-log model etc.

Annex

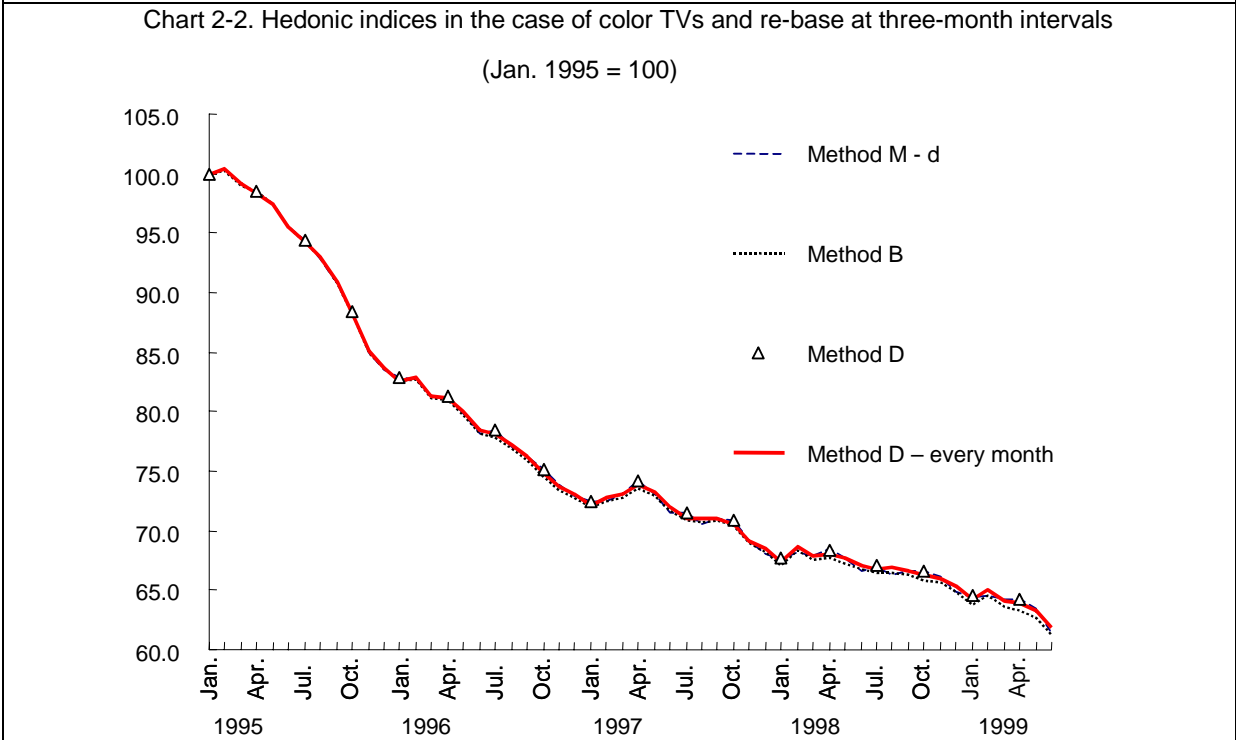
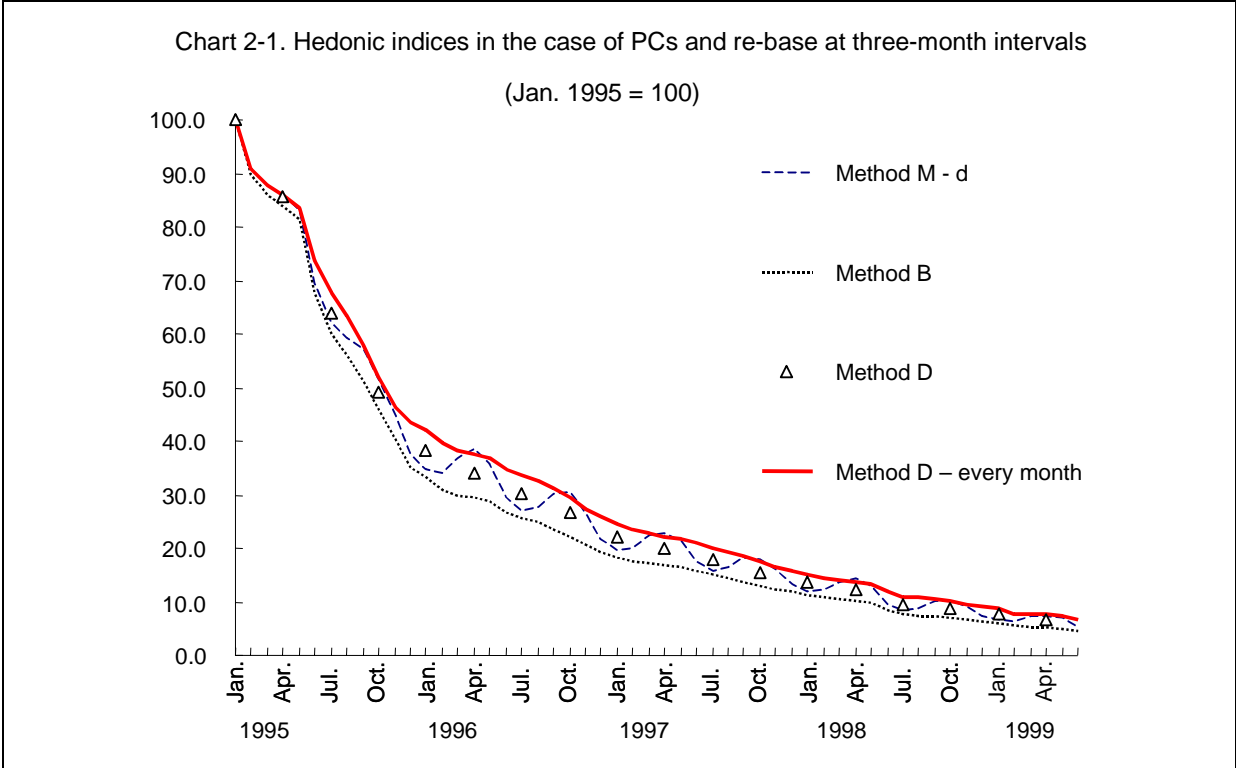


Chart 3-1. Hedonic indices in the case of PCs and every-month re-base  
(Jan. 1995 = 100)

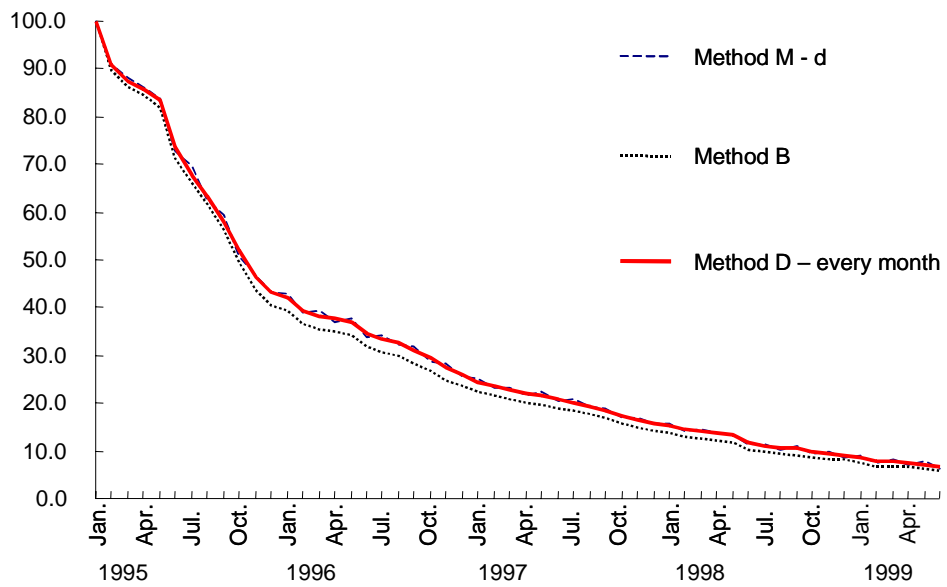


Chart 3-2. Hedonic indices in the case of color TVs and every-month re-base  
(Jan. 1995 = 100)

