Bank of Japan Working Paper Series

New Hedonic Quality Adjustment Method using Sparse Estimation

Sahoko Furuta*
sahoko.furuta@boj.or.jp

Yudai Hatayama**
yuudai.hatayama@boj.or.jp

Atsushi Kawakami***
atsushi.kawakami@boj.or.jp

Yusuke Oh**
yuusuke.ou@boj.or.jp

No.21-E-8
July 2021

Bank of Japan
2-1-1 Nihonbashi-Hongokucho, Chuo-ku, Tokyo 103-0021, Japan

* Research and Statistics Department

** Research and Statistics Department (currently, Financial System and Bank Examination Department)

*** Research and Statistics Department (currently, International Department)

New Hedonic Quality Adjustment Method using Sparse Estimation[*]

Sahoko Furuta,[†] Yudai Hatayama,[‡] Atsushi Kawakami,[§] Yusuke Oh[**]

July 2021

Abstract

In the application of the hedonic quality adjustment method to the price index, multicollinearity and the omitted variable bias arise as practical issues. This study proposes the new hedonic quality adjustment method using 'sparse estimation' in order to overcome these problems. The new method deals with these problems by ensuring two properties: the 'grouped effect' that gives robustness for multicollinearity and the 'oracle property' that provides the appropriate variable selection and asymptotically unbiased estimators. We conduct an empirical analysis applying the new method to the producer price index of passenger cars in Japan. In comparison with the conventional standard estimation method, the new method brings the following benefits: 1) a significant increase in the number of variables in the regression model; 2) an improvement in the fit of the regression model to actual prices; and 3) reduced overestimation of the product quality improvements due to the omitted variable bias. These results suggest the possible improvement in the accuracy of the price index while enhancing the usefulness of the hedonic quality adjustment method.

JEL Classification: C43, E31, C52

Keywords: Price Index, Quality Adjustment, Hedonic Regression Model, Multicollinearity, Omitted Variable Bias, Sparse Estimation, Adaptive Elastic Net

[*] The authors thank Naohito Abe, Fumio Funaoka, Yukinobu Kitamura, Chihiro Shimizu, Shigenori Shiratsuka, and the staff of the Bank of Japan for their valuable comments. We also thank Yuto Ishikuro, Marina Eguchi, Taiki Kubo, and Kotaro Shinma for their cooperation in data calculations. All remaining errors are our own. The views expressed in this study are those of the authors and do not necessarily reflect the official views of the Bank of Japan.

[†] Research and Statistics Department, Bank of Japan (E-mail: sahoko.furuta@boj.or.jp)

[‡] Research and Statistics Department, (currently, Financial Systems and Bank Examination Department), Bank of Japan (E-mail: yuudai.hatayama@boj.or.jp)

[§] Research and Statistics Department, (currently, International Department), Bank of Japan (E-mail: atsushi.kawakami@boj.or.jp)

[**] Research and Statistics Department, (currently, Financial Systems and Bank Examination Department), Bank of Japan (E-mail: yuusuke.ou@boj.or.jp)

1. Introduction

The hedonic quality adjustment method is one of the quality adjustment methods for the price index.[1] As the price index indicates 'pure' price changes of products over time, it is essential to adjust for differences in quality between old and new products in response to the renewal of products in the market. In the hedonic approach, based on the assumption that the quality of a product can be represented by the accumulation of its individual characteristics, we decompose the difference in the observed prices between old and new products into a quality change and a pure price change using the regression model which estimates the relationship between characteristics and prices. The hedonic quality adjustment method has two main advantages: 1) it can objectively evaluate the quality changes of products using data and statistical methods rather than the subjective judgement by the authorities; and 2) even if there are various changes in characteristics of products, it can comprehensively evaluate the effects of these changes on product prices. Therefore, the hedonic approach has been applied to the compilation of the consumer price index and the producer price index in many countries.

However, there are some issues in applying the hedonic quality adjustment method in practice.[2] First, in the regression model, if the characteristics of the products are highly correlated, the problem of multicollinearity on the explanatory variables is likely to occur, and the estimated parameters for the variables may become unstable. In addition, the parameters of the variables included in the regression model can be biased due to the omitted variables when it is difficult to obtain all the characteristic data of the products. Furthermore, considering that the relationship between characteristics and prices is not always linear, in the hedonic approach we often estimate non-linear models. However, it is known that the problems of multicollinearity and omitted variable bias can be more serious as the functional form for the model becomes more complex.[3]

Although these issues of the hedonic quality adjustment method long been known,

---

1 For the representative study of the hedonic approach, see Shiratsuka (1998).

2 For the practical issues of the hedonic approach, see Triplett (2006).

3 See Cropper et al. (1988) for details.

practically sufficient solutions have not been available until now. Therefore, in this study, we attempt to overcome these problems by improving the estimation method. Specifically, we propose a method to deal with the problems of multicollinearity and omitted variable bias using ‘sparse estimation’ as an estimation method for the hedonic regression model. Sparse estimation has gone through a process of improvement in statistics over a long time, and is often used in many academic fields, such as machine learning, in recent years. True to its meaning, ‘sparse’ estimation selects only meaningful explanatory variables from a large number of candidates, and estimates the parameters of the other variables to be exactly zero. Because of this property, in comparison with the conventional estimation method used in the hedonic model—for example, the ordinary least squares (OLS)—the new method with sparse estimation has the advantage that it can automatically select variables in the model. In particular, among sparse estimation methods, the adaptive elastic net (AEN), which is used in the new estimation method proposed in this study, is superior in that it has two desirable properties (Zou and Zhang (2009)): the ‘group effect’ that enhances robustness for multicollinearity, and the ‘oracle property’ that ensures appropriate variable selection and asymptotically unbiased estimators. In this paper, we show that these properties of the AEN can help to solve the above-mentioned problems of the hedonic approach. To our knowledge, there have been empirical analyses using AEN in various fields in recent years, however, there is no previous study applying AEN to the hedonic regression model.

The results of the analysis in this paper are as follows. We perform an empirical analysis applying the new method to passenger car prices in the Corporate Goods Price Index (CGPI) in Japan, which mostly corresponds to the producer price index, compiled by the Research and Statistics Department of the Bank of Japan. As a result, compared with the conventional estimation method used in the hedonic approach for the CGPI, the new method using AEN brings the following multiple and varied benefits. First, the number of variables incorporated into the regression model increase significantly, and this leads to an expansion of the characteristics that can be taken into account in the quality adjustment. Second, the fit of the regression model improves not only for the sample prices during the estimation period, but also after the estimation. In addition, we confirmed that estimated parameters are more stable than the conventional method with the change in estimation period. Third, when we examine the effect of change in the estimation method on the actual price index, it is confirmed that the rate of decline of the price index estimated by the new method becomes more gradual than that of the conventional method. This fact suggests that the conventional method may overestimate the quality improvement rate due to the omitted variable bias, while new estimation method could solve this problem. These results suggest that the new method can contribute to improvement in the accuracy of the price index while enhancing the usefulness of the hedonic quality adjustment method.

The rest of this paper is organized as follows. Section 2 describes the overview of conventional hedonic regression model and its problems. Section 3 explains the new hedonic quality adjustment method using sparse estimation and its properties. Section 4 shows the results of empirical analysis applying the new method to the producer price of passenger cars in Japan. Section 5 summarizes the paper.

2. Conventional method and issues

2-1. Conventional method

In this section, we provide an overview of the conventional method used in the hedonic quality adjustment, taking the CGPI compiled by the Research and Statistics Department of the Bank of Japan as an example. In the hedonic approach, a regression analysis is performed using the prices of the products as the dependent variables and the data representing the characteristics of products as the explanatory variables. Then, the estimated parameters are applied for the quality adjustment between new and old products. In the regression procedure, although we have to assume some specific functional form for the hedonic model, from the perspective of economic theory, it is known that there are no a priori restrictions on this form.[4] Since there are innumerable functional form

[4] The hedonic function is theoretically described as an envelope with respect to a bid function for a characteristic through consumer's utility maximization and an offer function derived from producer's profit maximization, in a perfectly competitive market where all characteristics can be selected continuously. Therefore, there are no a priori restrictions on the functional form. See Rosen (1974) for details.

candidates for the estimation, in practice, it is necessary to choose a proper functional form in terms of goodness of fit and consistency of the estimated parameters, e.g., significance, sign, and so on.[5] However, it is necessary to consider non-linearity because the relationship between product prices and characteristics is not always linear. From this point of view, in order to take into account the non-linearity, previous research has proposed using the regression model with the Box-Cox transformation of variables as follows.[6]

Box-Cox transformation

$$x^{(\lambda)} = \begin{cases} \dfrac{x^{\lambda} - 1}{\lambda} & (\lambda \neq 0) \\ \log x & (\lambda = 0) \end{cases}$$

λ in the above indicates the Box-Cox parameter and is a coefficient that determines the degree of nonlinearity of the function. Conventional hedonic regression model with the Box-Cox transformed term is as follows.

Conventional hedonic regression model

$$y_i{}^{(\lambda_0)} = \beta_0 + \sum_{j=1}^{p_c} \beta_{cj} x_{cj,i}{}^{(\lambda_j)} + \sum_{k=1}^{p_d} \beta_{dk} x_{dk,i} \tag{2}$$

y_i: theoretical price, $x_{cj,i}$: continuous variable, $x_{dk,i}$: dummy variable,
β_0: constant term, β_{cj}: coefficient on a continuous variable,
β_{dk}: coefficient on a dummy variable,
λ_0: Box-Cox parameter for theoretical price,
λ_j: Box-Cox parameter for a continuous variable,
p_c: number of continuous variables, p_d: number of dummy variables

According to the values of λ_0 and λ_j, the above formula is classified as follows;

(a) Log-Linear model when both the dependent and explanatory variables are log-linear

---

[5] Shiratsuka (1997) states that the criteria for function selection in hedonic methods should include goodness of fit and coherence of the parameters as well as value interpretability and estimation burden.

[6] For more details on the Box-Cox transformation, see Box and Cox (1964). In addition, Halvorsen and Pollakowski (1981) advocate utilizing the Box-Cox transformation as a general functional form for the hedonic model and performing the likelihood ratio test to select a specific functional form.

$(\lambda_0 = \lambda_j = 0)$

(b) Semi Log-Linear model when only the dependent variable is log-linear
$(\lambda_0 = 0, \lambda_j = 1)$

(c) Linear model when both the dependent and explanatory variables are linear
$(\lambda_0 = \lambda_j = 1)$

(d) Semi Box-Cox model when only the dependent variable is applied the Box-Cox transformation $(\lambda_j = 1)$

(e) Double Box-Cox model when both the dependent and explanatory variables are applied the Box-Cox transformation

These five regression models should be tested when selecting the functional form. It is known that, in the hedonic regression, the Double Box-Cox model is selected in many cases as a result of such a test.[7]

2-2. Issues (i): Multicollinearity

One of the issues that the conventional hedonic estimation method is likely to face is multicollinearity. Multicollinearity refers to a state in which there is a high intercorrelation among explanatory variables in a regression model. Multicollinearity makes it difficult to identify the effects of variables and estimate the parameters accurately. As a result, the parameters of the variables that are supposed to have an important effect on the dependent variable become insignificant.

It is known that the hedonic regression model is prone to the problem of multicollinearity. Using the example of the passenger car, total length and weight of the car body are highly correlated, leading to the problem of multicollinearity (Chart 1). In the dataset of items to which the hedonic quality adjustment method is applied, there are correlations among many variables, which are not limited to those that inevitably arise from technologically-based relationships such as the example of total length and weight. This is why companies have multiple product lines in different price ranges as a marketing

---

[7] Triplett (2006) mentions that statistical tests are more likely to reject linear model and log-linear model rather than Box-Cox. Actually, most of the hedonic regression model used for quality adjustment on CGPI in Japan is Double Box-Cox model.

strategy. Such as high-end products are equipped with many various functions, while those functions are reduced to the minimum necessary in low-end products. As a result, correlations are likely to occur between variables that do not necessarily have a strong technologically-based relationship, for example, between maximum power output and whether there is a power-controlled seat.[8]

There are two major approaches to deal with multicollinearity. The first is to perform principal component analysis beforehand and use some of the obtained principal components as explanatory variables on the regression of hedonic function. In Shiratsuka (1995), the hedonic regression for passenger cars is performed with the principal component added as an explanatory variable. It notes that improvements in the coefficient of determination of the regression equation are marginal and that it is difficult to interpret the estimated parameters because the effect of each characteristic to the principal components varies over time. On that basis, it concludes that while principal component analysis is useful to guess important characteristics, it is not always an appropriate method for dealing with multicollinearity using the principal components as explanatory variables. Therefore, a second, simpler approach is widely used in practice. This method excludes one of the correlated variables from the equation (stepwise method). In other words, if an effect of multicollinearity is suspected from the estimation results, it can be avoided to a certain extent by reestimating without the variables that may be the cause. However, as in the passenger car example above, it is not always easy to select the variables properly under a strong correlation between characteristics. Therefore, there is the inevitable burden of repeating the estimation until a plausible result is obtained.

2-3. Issues (ii): Omitted variable bias

The second issue that conventional hedonic estimation method is likely to face is bias in the parameters caused by omitted variables. An omitted variable is a variable that is not included in the regression model, although it is highly relevant to the explained variable.

---

[8] Triplett (2006) distinguishes between "multicollinearity in the universe" due to technologically-based correlation like length and weight and "multicollinearity in the sample" due to the correlation of functions depending on the grades of products.

In the hedonic method in price statistics, this is the case when the regression model does not include characteristics and performance that would have affected the price of the product.

There are two types of situations in which omitted variables occur in the hedonic method: (a) the case that occurs at the stage of data set construction; and (b) the case that occurs as a result of variable selection. In the case of (a), the problem arises from the fact that the characteristics have an impact on price but cannot be observed. For example, it is inherently difficult to include characteristics into the regression model, which we could not quantify well such as product design, style, and brand value. We can only deal with this problem partially by using dummy variables that identify the manufacturer as a proxy variable. Besides, if a new function emerges due to technological innovation, it is necessary to wait to incorporate variables into the regression model until a product with that function has penetrated the market to a certain extent. In the case of (b), the problem arises from inadvertent inclusion of variables with a slight impact on prices and exclusion of variables that truly have an impact on prices under the circumstances where we could not help selecting the limited number of variables due to multicollinearity.

Due to the presence of omitted variables, the parameters of the variables selected in the estimated regression model are distorted. If the distortion causes bias in the price index, the difference in the relative rate of quality improvement of the omitted and employed variables determines whether the distortion causes upward or downward bias. For example, if the omitted variable has a significant improvement over the employed variable, the quality improvement is underestimated, resulting in an upward bias in the price index. Conversely, if the employed variable with a distorted parameter improves significantly in quality while the omitted variable improves only slightly, there is a downward bias in price index as a result of overestimating quality improvement. Triplett (2006) applies the hedonic method to PC prices and finds that in the presence of omitted variables, a downward bias of about -0.2% to -1.0% arises in the price index over a five-month period. He provides the contextual background that employed variables such as processing speed and memory size may have improved better than the omitted variables. Sawyer and So (2018) also estimate how much the rate of price decline of microprocessors derived from hedonic regression differs among possible subsets of the regressors. He shows that the rate of price decline (on average over four years) when only one variable is employed is up to -45.11%, lower than -8.77% when all characteristics are employed, due to the omitted variable bias. Such previous research suggests that the presence of an omitted variable in the hedonic regression model may lead to a downward bias in the price index (an overestimation of the rate of quality improvement).

It is known that the omitted variable bias becomes more severe on complex functional forms. Cropper et al. (1988) states that it is appropriate to select the simpler functions through the hedonic estimation for a real estate price when there may be omitted variables. Particularly for functional forms with Box-Cox transformed terms, there is a risk of extreme values of Box-Cox parameters depending on the subset of explanatory variables.[9] Adopting a distorted functional form causes a problem—that the fit to the dataset used for the estimation is good, but the fit to the new product that comes after the estimation is poor ('overfitting'). Therefore, it is necessary to repeat estimation with changing the subset of the variables each time so that the Box-Cox parameters do not become excessively high order. We sometimes observe that the hedonic estimation result change greatly after re-estimation. For example, the Box-Cox parameter of passenger cars (Minivans) of CGPI changed from 3.4 to almost zero, logarithmic form (Chart 2). These changes may suggest parameter instability due to the presence of omitted variables.

2-4. Issues (iii): Interactions between characteristics

An additional issue faced in applying the hedonic model is the issue of 'interactions' between characteristics. The hedonic regression model is often performed under the assumption that the parameters for characteristics are the same among the products, but in practice, the assumption is not always valid. For example, there may be an interaction where a quality improvement in one characteristic increases the impact of quality improvement of another characteristic on price, or we estimate under the assumption that

[9] Graves et al. (1988) also estimates the hedonic regression model for the real estate value with various subset of variables and various functional forms. They note that within the functional form including the Box-Cox transformed terms, the choice of specification greatly affected the estimation results.

they are the same product but, in fact, they are classified by more detailed categories.

To deal with these interactions, it is useful to introduce a cross term for the variables in the regression. This allows us to capture the situation in which the impact of each characteristic on price depends on the state of another characteristic. However, it is difficult in practice to employ all of the cross terms when estimating because there is a huge number of potential combinations of cross term variables. The more cross terms employed, the higher the correlation between the explanatory variables, potentially leading to multicollinearity, and possibly making the parameters more unstable. As a result, the conventional hedonic regression model has been limited in the employment of cross terms. However, given the omitted variable bias stated above, there is a risk that the parameters of the variables and cross terms may be biased when estimating without the important cross terms. We consider that the presence of interactions has not been resolved in the hedonic regression, as it faces both multicollinearity and omitted variables bias, as described above.

3. New hedonic quality adjustment method using sparse estimation

3-1. Sparse estimation

This section will explain the hedonic regression model using sparse estimation. Sparse estimation selects only the meaningful variables from many candidates of explanatory variables and gives zero coefficients to the rest of the variables (called ‘sparsity’). Sparse estimation performs variable selection and coefficient estimation at the same time under sparsity. This method has an advantage over the conventional one using OLS estimation in which it can automatically derive a stable and well fitted model. Sparse estimation has been used in various fields of empirical analysis, not only in economics. In this section, we explain how this type of method is useful in dealing with the issues of hedonic regression for price statistics (multicollinearity and omitted variable bias).[10]

[10] Sparse estimation is also useful when analyzing observational data, for example, it was used in the world’s first black hole imaging by the international project (The Event Horizon Telescope Collaboration (2019)). In addition, also in the field of geographic information science, sparse estimation is used in quantifying inter-regional heterogeneity. For example, Jin and Lee (2020) estimate housing prices with a

Many methods of sparse estimation have been proposed to date, starting with the "Lasso" (least absolute shrinkage and selection operator) proposed by Tibshirani (1996). The new estimation method proposed in this study employs an adaptive elastic net (AEN), which enjoys two desirable properties: the 'group effect' that gives robustness for multicollinearity and the 'oracle property' that ensures the adequacy of variable selection and estimated coefficients.[11] To our knowledge, there have been no studies applying AEN to the hedonic regression model.[12] In the following, we will provide an overview of sparse estimation and the above two properties in turn.

First, we will outline how sparsity is satisfied using Lasso, a typical sparsity estimation. Lasso is a method for estimating $\boldsymbol{\beta}$, which minimizes the function of the sum of the squared error in the equation including the L_1 norm (sum of absolute values) of $\boldsymbol{\beta}$ as a regularization term.[13]

<u>Lasso</u>

$$\widehat{\boldsymbol{\beta}}(Lasso) = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left(|\boldsymbol{Y} - \boldsymbol{X\beta}|^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right) \tag{3}$$

$\lambda > 0$: regularization parameter

(A relatively smaller number of variables are selected if λ is large)

A traditional method that could deal with multicollinearity is ridge regression.[14] It is same as Lasso, in that it minimizes the function of the sum of the squared error including the regularization term, but ridge regression uses the L_2 norm (the sum of the squares) of $\boldsymbol{\beta}$ as the regularization term. It leads to a key difference that Lasso satisfies sparsity, while ridge regression does not.

---

spatial vector auto regression model using sparse estimation and Wheeler (2009) suggests adopting sparse estimation on a geographically weighted regression model.

[11] For details of estimation method and each of the properties, see Zou and Zhang (2009).

[12] There are some studies using Lasso among sparse estimation for hedonic regression model. For example, Zafar and Himpens (2019) apply Lasso to analyze webscraped laptop prices and characteristics and compare the result with other estimation methods which consider nonlinearity.

[13] We centralize the dependent variables and standardize the explanatory variables. That is, for the number of observations n, we set $\frac{1}{n}\sum_{i=1}^{n} y_i = 0$, $\frac{1}{n}\sum_{i=1}^{n} x_{j,i} = 0$, and $\frac{1}{n}\sum_{i=1}^{n} {x_{j,i}}^2 = 1$.

[14] For details, see Hoerl and Kennard (1970).

Ridge regression

$$\widehat{\boldsymbol{\beta}}(Ridge) = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left(|\boldsymbol{Y} - \boldsymbol{X\beta}|^2 + \lambda \sum_{j=1}^{p} {\beta_j}^2 \right) \tag{4}$$

$\lambda > 0$: regularization parameter
(Coefficient is estimated to be smaller if λ is large)

We show how differences in regularization terms satisfy or does not satisfy the sparsity intuitively in Chart 3 in line with the discussion described in Tibshirani (1996). If there are two variables, from the Karush-Kuhn-Tucker condition, $\widehat{\boldsymbol{\beta}}(Lasso)$ and $\widehat{\boldsymbol{\beta}}(Ridge)$ can be transformed into the formulas described in Chart 3. Considering a plane consisting of β_1-axis and β_2-axis, the sum of the squared error illustrates an ellipse centered on $\widehat{\boldsymbol{\beta}}(OLS)$. The constraint corresponding to each regularization term illustrates a rhombus in Lasso and a circle in ridge regression. Under these conditions, $\boldsymbol{\beta}$ is derived from the tangent point of the sum of the squared error (ellipse) and the constraint (rhombus or circle). Here, in Lasso, the constraint represents rhombus, and the two conditions are likely to intersect at the corners. In other words, the corner solution is likely to be selected in the constraint of $\boldsymbol{\beta}$. In this case, one of the parameters is estimated to be exactly zero and the variable is selected automatically. On the other hand, ridge regression is not prone to automatic variable selection. This is because the constraint region is a circle, and the sum of the squared error and the constraint are not likely to intersect at any particular point, making it unlikely that one parameter will be estimated at exactly zero.

3-2. Group effect

For Lasso, the results of variable selection are known to be unstable in data with strong multicollinearity. For example, suppose that the true values of the parameters for two variables are ${\beta_1}^*$ and ${\beta_2}^*$. As an extreme example, if the values of these two variables are exactly the same, then the solution to the optimization by Lasso is not uniquely determined because there are innumerable solutions as follows.

$$\widehat{\boldsymbol{\beta}}(Lasso) = \begin{pmatrix} s({\beta_1}^* + {\beta_2}^*) \\ (1-s)({\beta_1}^* + {\beta_2}^*) \end{pmatrix} \text{ for any } s \in [0,1] \tag{5}$$

Similarly, when there are two highly correlated variables, Lasso's variable selection is greatly affected by slight changes in data and the variable into the regression is not stable.

Since the variables for hedonic regression model are often highly correlated, it is necessary to adopt sparse estimation, which is robust under the multicollinearity conditions described above. One of the typical sparse estimation with this property is the elastic net (EN). EN is a method for estimating $\boldsymbol{\beta}$, which minimizes the function of the sum of the squared error in the equation plus both the L_2 norm and the L_1 norm of $\boldsymbol{\beta}$ as a regularization term.[15] This enables EN to have the advantages of both Lasso and ridge regression: variable selection and robustness for multicollinearity.

<u>Elastic Net (EN)</u>

$$\widehat{\boldsymbol{\beta}}(EN) = \left(1 + \frac{\lambda_2}{n}\right)\left\{\underset{\boldsymbol{\beta}}{\operatorname{argmin}}\left(|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}|^2 + \lambda_2 \sum_{j=1}^{p} {\beta_j}^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j|\right)\right\} \tag{6}$$

$\lambda_2 > 0$: L_2 norm regularization parameters
$\lambda_1 > 0$: L_1 norm regularization parameters
n: number of observations

The robustness of EN for multicollinearity is called the ‘group effect’. Group effect is a property that gives smaller differences between the coefficients on those variables when the correlation between the variables is high.[16] As an extreme case, if the values of two variables are exactly the same, the EN estimates the parameters on those two variables as exactly equal, as follows. This allows for stable variable selection and parameter estimation, even in situations where it is difficult to discern which variables surely have impact on price from the data under multicollinearity.

$$\widehat{\boldsymbol{\beta}}(EN) = \begin{pmatrix} \frac{1}{2}({\beta_1}^* + {\beta_2}^*) \\ \frac{1}{2}({\beta_1}^* + {\beta_2}^*) \end{pmatrix} \tag{7}$$

---

[15] For details, see Zou and Hastie (2005).

[16] To be more specific, the maximum absolute value of difference between the parameters is directly proportional to $\sqrt{1-\rho}$, when the sample correlation ρ is greater than zero.

3-3. Oracle property

Another property that must be satisfied by the estimator derived from sparse estimation is the 'oracle property'. Specifically, with the true coefficient $\boldsymbol{\beta}^*$, it is defined that an estimator $\widehat{\boldsymbol{\beta}}$ has the oracle property when it satisfies the following two conditions.

<u>Oracle property</u>

(1) Variable selection consistency

$$\lim_{n\to\infty} P(\hat{\beta}_j = 0) = 1 \qquad\qquad with\ \ \beta_j{}^* = 0$$

(2) Asymptotic normality of the non-zero coefficients

$$\lim_{n\to\infty} \frac{(\hat{\beta}_j - \beta_j{}^*)}{\sigma(\hat{\beta}_j)} \sim \mathrm{N}(0,1) \qquad with\ \ \beta_j{}^* \neq 0$$

$\sigma^2(\hat{\beta}_j)$: asymptotic variance of estimator

Of the two conditions above, (1) 'variable selection consistency' means that the estimator of the coefficient satisfies consistency for a variable whose true coefficient is zero. The 'asymptotic normality of the non-zero coefficients' in (2) means that for variables whose true coefficients are non-zero, the estimation error on those coefficients follows an asymptotic normal distribution.

The oracle property is an important property that asymptotically guarantees the appropriateness of both the 'variable selection' and the 'coefficient estimation' that sparse estimation simultaneously performs. However, Lasso and EN are known not to satisfy oracle property depending on the data, no matter how properly the regularization parameters are chosen. Therefore, we adopt the following adaptive elastic net (AEN) as a new estimation method for the hedonic regression model, which satisfies the oracle property in sparse estimation.

<u>Adaptive elastic net (AEN)</u>

$$\widehat{\boldsymbol{\beta}}(AEN) = \left(1 + \frac{\lambda_2}{n}\right)\left\{\underset{\boldsymbol{\beta}}{\operatorname{argmin}}\left(|\boldsymbol{Y} - \boldsymbol{X\beta}|^2 + \lambda_2 \sum_{j=1}^{p} \beta_j{}^2 + \lambda_1{}^* \sum_{j=1}^{p} \widehat{w}_j |\beta_j|\right)\right\} \qquad (8)$$

$$\widehat{w}_j = \left(|\hat{\beta}_j(EN)|\right)^{-\gamma}$$

$\lambda_1{}^* > 0$: L_1 norm regularization parameters (2nd stage)

$\widehat{w}_j > 0$: adaptive weight

$\gamma > 0$: adaptive parameter

(Larger γ imposes larger penalties corresponding to the absolute value of the coefficient)

The AEN estimation is performed in two stages. At the first stage, we estimate the coefficients with EN. Then, EN is performed again after the regularization term of the L_1 norm is adjusted for each variable to impose greater penalties for variables with small absolute values of the coefficients.[17] This two-step estimation allows us to enjoy oracle property with almost no dependence on the properties of the dataset.

Chart 4 provides an intuitive explanation of the reason why AEN satisfies oracle property, referring to the discussion in Zou (2006). Here, we artificially generate the matrix $\boldsymbol{X}$ of the explanatory variables and the vector $\boldsymbol{\varepsilon}$ of the disturbance terms, then calculate the vector $\boldsymbol{Y}$ of the dependent variable based on the true model ($\boldsymbol{Y} = \boldsymbol{X\beta}^* + \boldsymbol{\varepsilon}$). Then we check how OLS, Lasso, and AEN estimate $\boldsymbol{\beta}$ when $\boldsymbol{Y}$ and $\boldsymbol{X}$ are the observed values. The true coefficient β^* is on horizontal axis, and the coefficients of $\hat{\beta}(OLS)$, $\hat{\beta}(Lasso)$ and $\hat{\beta}(AEN)$ are plotted on the vertical axis. First, at Lasso, we see that $\hat{\beta} = 0$ when $|\beta^*| < \lambda$ and it is clear that it satisfies sparsity. On the other hand, when $|\beta^*| \geq \lambda$, $\hat{\beta}$ is estimated to be smaller by λ in absolute value than the true value β^*. In other words, we can see that the regularization parameter λ and the condition of the oracle property are in a trade-off; as λ increases, it becomes easier to estimate the zero coefficient and satisfy the consistency of variable selection, while the estimated value becomes smaller by λ in absolute value, making it difficult to satisfy the asymptotic normality for the non-zero coefficient.

In contrast, at AEN, when $|\beta^*|$ is small, a great penalty is imposed based on the small value of the coefficients in the first-stage estimation, and $\hat{\beta} = 0$ is derived. On the other hand, when $|\beta^*|$ is large, we can see that $\hat{\beta}$ approaches β^* asymptotically due to lower penalties. Thus, by adjusting the penalties corresponding to estimates at the first-stage, $\hat{\beta}$ is likely to be estimated zero when the coefficient is small, while the shrinkage

[17] In the second-stage of the EN estimation, we drop the variables whose parameter is estimated zero at the first-stage.

of the estimates in absolute value is minimized when the coefficient is large. This makes it easier to satisfy the two conditions for oracle property.

3-4. Selection of functional form

In the new estimation method proposed in this study, the hedonic function is formulated as a quadratic polynomial. AEN determines which terms should be included in the regression model and performs both variable selection and functional form selection at the same time. Since all cross terms are considered, it is possible to include the interaction effects into the regression model, unlike in the conventional method.[18] The reason we limit the degree of equation to second is to prevent overfitting caused by higher-degree terms, which sometimes occur in Box-Cox method.[19]

According to the above, the hedonic regression model with the new method is estimated as follows.

Hedonic regression model using AEN

$$Y_i \equiv \log y_i$$

$$Y_i = \hat{\beta}_{00} + \sum_{j=1}^{p} \hat{\beta}_{0j} x_{j,i} + \sum_{j=1}^{p} \hat{\beta}_{jj} {x_{j,i}}^2 + \sum_{k>j\geq 1} \hat{\beta}_{jk} x_{j,i} x_{k,i} \tag{9}$$

$where$

$$\widehat{\boldsymbol{\beta}} = \left(1 + \frac{\lambda_2}{n}\right)\left\{\underset{\boldsymbol{\beta}}{\operatorname{argmin}}\left(|\boldsymbol{Y} - \boldsymbol{X\beta}|^2 + \lambda_2 \sum_{k\geq j\geq 0} {\beta_{jk}}^2 + {\lambda_1}^* \sum_{k\geq j\geq 0} \widehat{w}_{jk}\left|\beta_{jk}\right|\right)\right\}$$

---

[18] If we calculate the cross terms for all subset of variables, a perfect multicollinearity occurs in most cases. Therefore, in this study, we drop some variables which are linearly dependent on other variables before estimation.

[19] Another argument on the estimation using AEN is the setting of hyperparameter ($\lambda_1,\ {\lambda_1}^*,\ \lambda_2,\ \gamma$). From several methods of the setting the parameters mentioned in Zou and Zhang (2009), in this study, we select K-fold cross validation which is often used in fields of machine learning. We split the dataset into K groups, then take one group as test data and take the remaining *K-1* groups as a training data. We fit a model on the training data and evaluate it on the test data. We can evaluate the model by retaining this procedure K times with resampling group. When we select the appropriate degree for the K, we have to pay attention to a trade-off between the bias on the coefficients which affects estimation accuracy and the variances due to differences in training data. We choose K=10 in our analysis, which is commonly used.

$$\widehat{w}_{jk} = \left(\left|\hat{\beta}_{jk}{}^{1st}\right|\right)^{-\gamma}$$

$$\widehat{\boldsymbol{\beta}}^{1st} = \left(1 + \frac{\lambda_2}{n}\right)\left\{\underset{\boldsymbol{\beta}}{\operatorname{argmin}}\left(|\boldsymbol{Y} - \boldsymbol{X\beta}|^2 + \lambda_2 \sum_{k \geq j \geq 0} \beta_{jk}{}^2 + \lambda_1 \sum_{k \geq j \geq 0} |\beta_{jk}|\right)\right\}$$

y_i: theoretical price, $x_{j,i}$: explanatory variable, $\hat{\beta}_{jk}$: coefficient on $x_{j,i}x_{k,i}$,
p: number of candidate explanatory variables, n: number of observations,
$\lambda_1 > 0$: L_1 norm regularization parameter (1st stage),
$\lambda_1{}^* > 0$: L_1 norm regularization parameter (2nd stage),
$\lambda_2 > 0$: L_2 norm regularization parameter,
$\gamma > 0$: adaptive parameter, $\widehat{w}_{jk} > 0$: adaptive weight

4. Empirical analysis using new estimation method

4-1. Dataset for estimation

In this section, we apply the new hedonic regression model using AEN to passenger cars in Japan and discuss its properties.

We use the same data for estimation as used in the hedonic regression for CGPI in Japan compiled by the Research and Statistics Department of the Bank of Japan. Specifically, retail price data are taken from the Goo-net by the PROTO CORPORATION and average discounts are taken from the Monthly Car Magazine JIKAYOSHA by the Naigai Publishing Corp. Price data on passenger cars are compiled by the retail prices and average discounts. The period examined is from the 3rd quarter of 2016 to the 2nd quarter of 2018 and number of observations is 940.

The product specification data are basically taken from the Goo-net as well, but other important specifications unlisted in the database are taken from the specification sheet of each passenger car. The characteristics and performance used are shown in Chart 5. The data contains about 20 continuous variables measuring quantitative characteristics and about 100 dummy variables measuring qualitative characteristics.[20] The large amount of

[20] This includes vehicle configuration dummy, brand dummy and time dummy besides characteristics.

variables are due to the complicated characteristics of passenger cars, and how to select appropriate variables is particularly challenging in these complicated products. As stated in the previous section, the method using sparse estimation is superior in that it selects variables automatically, and this advantage is expected to be especially great when adjusting the quality of products with a lot of quality characteristics, such as passenger cars.

4-2. Comparison of new and old estimation results

Here, we show the results of applying the conventional estimation method and the new method using AEN. First, the result of the conventional estimation method is shown in Chart 6. As explained in Section 2, the conventional hedonic regression model is performed with the Box-Cox transformed term and the double Box-Cox model is selected based on the results of likelihood ratio test. Among the explanatory variables, only room space, fuel efficiency × equivalent inertia weight, and maximum output were selected for continuous variables. Note that here we employ dummy variables for each vehicle configuration to account for the difference of the impact of characteristics on the price. For example, the room space was not significant for sedans and wagons but was significant only for minivans. Dummy variables were significant for powertrain (e.g., 4WD, RWD), interior and exterior equipment (e.g., leather seats, LED headlamps, etc.), and brand (dummies for each automakers), respectively.

Next, the results from the new hedonic method using AEN are shown in Chart 7. As mentioned earlier, sparse estimation, such as AEN, can estimate with a large number of explanatory variables and perform both the ‘variable selection’ and the ‘coefficient estimation’ simultaneously. In this study, we limit the order of non-linearity in the equation to second and employ many cross terms to account for the presence of interactions between variables. As a result, compared to the conventional method, a large number of variables are employed and many cross terms are captured in the regression model.[21,22]

However, the regression model in the conventional and new methods have very different functional forms, and the parameters derived from the estimation cannot be simply compared. Therefore, we calculate the contribution of each variable to the theoretical price as follows and compare the results.

$$\pi_l{}^{func} = \frac{y^{func}(\bar{x}_l + \Delta x_l, \bar{\boldsymbol{x}}_{-l}) - y^{func}(\bar{x}_l, \bar{\boldsymbol{x}}_{-l})}{y^{func}(\bar{x}_l, \bar{\boldsymbol{x}}_{-l})} \times 100 \tag{10}$$

$$\log y^{AEN}(\bar{x}_l, \bar{\boldsymbol{x}}_{-l}) = \hat{\beta}_{00} + \sum_{j=1}^{p} \hat{\beta}_{0j}\bar{x}_j + \sum_{j=1}^{p} \hat{\beta}_{jj}\bar{x}_j{}^2 + \sum_{k>j\geq 1} \hat{\beta}_{jk}\bar{x}_j\bar{x}_k \tag{11}$$

$$y^{Box-Cox}(\bar{x}_l, \bar{\boldsymbol{x}}_{-l})^{(\lambda_0)} = \hat{\beta}_0 + \sum_{j=1}^{p_c} \hat{\beta}_{cj}\bar{x}_{cj}{}^{(\lambda_j)} + \sum_{k=1}^{p_d} \hat{\beta}_{dk}\bar{x}_{dk} \tag{12}$$

$\pi_l{}^{func}$: contribution rate of x_l for $func = AEN$ or $Box - Cox$

$\bar{x}_l$: average of explanatory variable l

Δx_l: standard deviation in continuous variable or 1 in dummy variable l

Chart 8 shows the estimation results of the contribution to passenger car prices of the continuous and dummy variables employed in Chart 6 and 7 (including cross terms). Specifically, we show the rate of change in theoretical price $\pi_l{}^{func}$ due to one standard deviation increase in continuous variables or one unit increase in dummy variables where a hypothetical sample with all variables are set at the mean value over the sample period.

First, with the continuous variables, more variables are employed in the new method than in the conventional one. The number of adopted variables increased from just two in the conventional method to nine in the new method. This can be interpreted as improved applicability of the hedonic quality adjustment using the new method. Next, with the

---

[21] Some papers consider residual bootstrapping method to test the significance of estimators in AEN, for example Chatterjee and Lahiri (2013), but there is still no consensus.

[22] For AEN estimation, the same number of variables as degrees of freedom can be employed in the model at most. However, if the number of samples is large enough and the degree of freedom is high, we can avoid extremely complex model and calculation burden by restricting the number of employed variables. In this study, we confine the maximum number of variables to 140 for 939 degrees of freedom, although we confirm that the improvement in fit is limited at even larger number of variables.

dummy variables, the number of the variables employed in the new method also increased significantly compared to the conventional method. Notably, the new method captures more dummy variables measuring characteristics than the conventional method, while the contribution on price of brand dummies reduces. This means that quality, which was previously captured as a manufacturer-specific factor, can now be captured as specific product characteristics by each of variables. Actually, the quality adjustment for passenger cars is often applied to the cases of model change occurred in the same brand. If the theoretical price is estimated mainly on brand dummies in the model, there will be no room to apply the quality adjustment as long as the manufacturer does not change, even though the quality improvement seems to occur through the model change in fact. Therefore, not relying on brand dummies may provide significant benefits in the quality adjustment.

These are the comparisons of the estimation results about the parameters. In order to compare the performance of the new method with that of the conventional method, we need to compare the fit of the regression model. To make this visible, we show in Chart 9 the mean squared error of the two regression models calculated with products released in each quarter, using the recent dataset. Applying the new method, the error is reduced over the whole quarter compared to the conventional method, confirming that the estimation accuracy is improved. In particular, the new method reduces the error not only during the estimation period, but also for the sample after the estimation period. Since we usually apply the quality adjustment to products appeared in the market after the estimated period, the improvement of the fit to out-of-sample is important.

In addition, when applying the hedonic method in practice, it is necessary to periodically re-estimate the regression model. The new method also confirmed the modest change in the estimation results when changing the sample period of the dataset (see the Appendix for details). Such an enhancement in time-stability of the estimation results may also improve the applicability of the hedonic quality adjustment.

4-3. Impact of new estimation on the price index

Here we see how the introduction of the new hedonic regression model using AEN affects the price index. Specifically, we estimate how the price index would have changed if the old and new hedonic quality adjustment had been applied to all the sample price replacements — the replacement of the surveyed product due to the EOL of the old product or change in representativeness of products in the market — occurred after 2017 for PPI “Standard Passenger Cars (Gasoline Cars)”.

We then compare the results of the new and old method with the officially released index of PPI to examine whether the results are plausible. In practice in the compilation of the CGPI, even if the hedonic quality adjustment method is available for a product, the Research and Statistics Department of the Bank of Japan would choose the most appropriate method, based mainly on a plausibility check of an estimated quality improvement with a surveyed company and a comparison using estimates of other quality adjustment methods such as the production cost method. In other words, if the applying the conventional hedonic quality adjustment is not judged to be appropriate from a practical view point, then we decide not to apply it. Therefore, comparing the price index when the new method is applied to all the sample price replacements with the released index of PPI, which is compiled based on practical judgement, we can generally assess whether the new method accurately estimates the rate of quality improvement.

Chart 10 shows the results of the calculation. The dashed line in the chart is an estimated price index when the conventional hedonic quality adjustment is applied to all the sample price replacements. We can see a somewhat larger decline in the price index of conventional hedonic regression model. On the other hand, regarding the solid line where the new hedonic method using AEN is applied, the price index shows gradual decline compared to the conventional method. Thus, the difference in the estimation results between the old and new methods indicates a quantitatively non-negligible impact on the price index of passenger cars.

Chart 10 also shows the officially released index of PPI as a dotted line, and the trend is more similar to the new method with the AEN than to the conventional method. The results show that if the old and new hedonic quality adjustment are applied to all the sample price replacements, using the old method would risk overestimating the rate of quality improvement, resulting excessive decline in the price index, whereas the new method may be able to estimate the rate of quality improvement more accurately in general.

These results are consistent with the results of previous studies on the bias of missing variables described in Section 2. In the conventional method, a limited number of explanatory variables due to multicollinearity are more likely to cause omitted variable bias, which leads to distortions in the parameters of variables in the hedonic model—which seems to overestimate the rate of quality improvement. On the other hand, in the AEN estimation, the increase in the number of explanatory variables is likely to reduce omitted variable bias, and the small distortion of the parameters results in an accurate calculation of the quality improvement rate. This is reflected in the differences in the price index.

5. Final Remarks

In this study, we survey the issues of the hedonic regression model and then explain the details of the new estimation method using sparse estimation and its results. The new estimation method proposed in this study employs an adaptive elastic net (AEN), which enjoys two desirable properties: the 'grouped effect' that gives robustness for multicollinearity and the 'oracle property' that ensures the adequacy of variable selection and asymptotic unbiasedness of coefficients. It has a possibility to overcome the practical issues of the hedonic regression model. In fact, the empirical analysis of passenger car prices in Japan in this study shows that the new method using the AEN improved in terms of: 1) a significant increase in the number of adopted variables; 2) improvement in fit; and 3) elimination of omitted variable bias. In particular, applying the new estimation method instead of the conventional one, the price index of passenger cars shows more moderate decline, and this method reduces the risk of overestimation of the quality improvement rate due to the omitted variable bias present in the conventional method. It is expected this change will make the hedonic quality adjustment more accurate and improve its applicability when the sample price replacement occurs. As mentioned in Section 1, the hedonic method has strengths in evaluating quality objectively based on data and statistical methods, and it is compatible even for a large number of changes in characteristics between new and old products. The increased usability of the hedonic regression model with these strengths is expected to make the price index more accurate.

In this study, we used passenger cars as an example, but the method proposed is based on the versatile approaches using ‘sparse estimation’ and ‘polynomial regression’, which are also applicable to another products. In applying the hedonic regression approach, we have to gather the data and construct the model for regression, considering the characteristics of each product sufficiently. The issues pointed out in this study are generally common to all products, and the new method which intends to overcome such issues could improve the performance of hedonic methods in a variety of products. Also, because of the versatile approach, we can flexibly customize the method corresponding to advances in statistical methods research and practical requirements. For example, whether the estimation accuracy and parameter stability can be improved by applying more advanced sparse estimation, or whether the generalization performance can be further enhanced by using more advanced cross-validation methods in hyperparameter setting, are some of the remaining issues. In addition, if the estimation accuracy required in practice is not always high, an alternative approach that emphasizes interpretability for the hedonic model can be fully envisioned while maintaining the framework of the new estimation method. For example, we can select a simpler functional form or variable composition by setting a lower upper limit of the number of the variables employed in the model, as well as we can limit the number of variables for cross terms from the outset.

This study focuses on dealing with the issues about multicollinearity and omitted variable bias by applying sparse estimation to the hedonic regression model, however, there are a number of other issues surrounding the hedonic approach. For example, the method of gathering the dataset is an important issue that is also related to omitted variable bias. As the adage ‘garbage in garbage out’ suggests, it is important to maintain the quality of the dataset for estimation by accurately grasping the technological innovation of the products and adopting variables related to new characteristics as necessary. In the field of hedonic approach, the subject of how to utilize recent advanced information processing technology, such as big data analysis, to gathering the dataset is also under study.[23] The use of large dataset is expected to become easier in the future. Under these circumstances, the estimation method proposed in this study is a highly efficient method as it can automatically construct a good performance model by extracting all necessary information even with the large dataset. We expect further utilization of the new method proposed in this study for empirical research and statistical practice in the future.

[23] For the research on hedonic regression model with the web scraping data, see Zafar and Himpens (2019) or Efthymiou and Antoniou (2013).

References

Bonaldi, P., Hortaçsu, A., and Kastl, J., "An Empirical Analysis of Funding Costs Spillovers in the Euro-Zone with Application to Systemic Risk," NBER Working Paper, No. 21462, National Bureau of Economic Research, 2015.

Box, G. E. P. and Cox, D. R., "An Analysis of Transformations," *Journal of the Royal Statistics Society Series B*, Vol. 26, pp. 211-252, 1964.

Chatterjee, A. and Lahiri, S. N., "Rates of Convergence of the Adaptive LASSO Estimators to the Oracle Distribution and Higher Order Refinements by the Bootstrap," *The Annals of Statistics*, Vol. 41(3), pp. 1232-1259, 2013.

Cropper, M., Deck, L. B., and McConnell, K. E., "On the Choice of Functional Form for Hedonic Price Functions," *The Review of Economics and Statistics*, Vol. 70(4), pp. 668-675, 1988.

Efthymiou, D. and Antoniou, C., "How Do Transport Infrastructure and Policies Affect House Prices and Rents? Evidence from Athens, Greece," *Transportation Research Part A*, Vol. 52, pp. 1-22, 2013.

The Event Horizon Telescope Collaboration, "First M87 Event Horizon Telescope Results. IV. Imaging the Central Supermassive Black Hole," *The Astrophysical Journal Letters*, Vol. 875(1), 2019.

Graves, P., Murdoch, J. C., Thayer, M. A., and Waldman, D., "The Robustness of Hedonic Price Estimation: Urban Air Quality," *Land Economics*, Vol. 64(3), pp. 220-233, 1988.

Halvorsen, R. and Pollakowski, H. O., "Choice of Functional Form for Hedonic Price Equations," *Journal of Urban Economics*, Vol. 10(1), pp. 37-49, 1981.

Hirakata, N., "The Time Variation of the Hedonic Regression Model and Its Effect on the Price Index: A case of Personal Computers in Japan" *Bank of Japan Working Paper Series*, No.05-J-1, 2005(in Japanese).

Hoerl, A. E. and Kennard, R. W., "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics*, Vol. 12, pp. 55-67, 1970.

Jin, C., and Lee, G., "Exploring spatiotemporal dynamics in a housing market using the spatial vector autoregressive Lasso: A case study of Seoul, Korea," *Transactions in GIS*, Vol. 24(1), pp. 27-43, 2020.

Pakes, A., "A Reconsideration of Hedonic Price Indexes with an Application to PC's," *American Economic Review*, Vol. 93(5), pp. 1578-1596, 2003.

Rosen, S., "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, Vol. 82(1), pp. 34-55, 1974.

Sawyer, S. D. and So, A., "A New Approach for Quality-Adjusting PPI Microprocessors," Monthly Labor Review, Bureau of Labor Statistics, 2018.

Shiratsuka, S., "Automobile Prices and Quality Changes: A Hedonic Price Analysis of Japanese Automobile Market," *Monetary and Economic Studies*, Vol. 13(2), pp. 1-44, 1995.

Shiratsuka, S., "Measuring Quality Changes using Hedonic Approach: Theoretical framework and its application to empirical research," *IMES discussion paper series*, No. 97-J-6, Bank of Japan, 1997(in Japanese).

Shiratsuka, S., *An Economic Analysis of Pricing*, University of Tokyo Press, 1998(in Japanese).

Tibshirani, R., "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistics Society Series B*, Vol. 58, pp. 267-288, 1996.

Triplett, J. E., *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products*, OECD Publishing, 2006.

Wheeler, D. C., "Simultaneous coefficient penalization and model selection in geographically weighted regression: the geographically weighted LASSO," *Environment and Planning A*, Vol. 41, pp. 722–742, 2009.

Zafar, J. D. and Himpens, S., "Webscraping Laptop Prices to Estimate Hedonic Models and Extensions to Other Predictive Methods," presented at the 16th meeting of the Ottawa Group on Price Indices, Rio de Janeiro, 2019.

Zou, H., "The Adaptive Lasso and Its Oracle Properties," *Journal of the American*

Statistical Association, Vol. 101, pp. 1418-1429, 2006.

Zou, H. and Hastie, T., "Regularization and Variable Selection via the Elastic Net," *Journal of the Royal Statistics Society Series B*, 67, pp. 301-320, 2005.

Zou, H. and Zhang, H. H., "On the Adaptive Elastic-Net with a Diverging Number of Parameters," *The Annals of Statistics*, Vol. 37(4), pp. 1733-1751, 2009.

Appendix; Time-stability of the hedonic regression model

It is widely known that the hedonic regression model is unstable. This is because the relationship between characteristics and prices may change over time, influenced by advancement in technology, changes in consumer preferences, and other factors. For example, Pakes (2003), in estimating the regression model for personal computers, points out that when the price of a microprocessor falls significantly due to technological innovation, the equation for personal computers equipped with this microprocessor may also change, and then it shows that the estimated parameters actually may change significantly. In order to measure these changes properly, it is necessary to periodically re-estimate the regression model and flexibly adopt to changes in the functional form and the subset of variables.

Here, as in the main text, we analyze the stability of the parameters by running the regression on different sample periods. This analysis is conducted for passenger cars and the samples are one year older than the one used in the main text (the period examined is from the 3rd quarter of 2015 to the 2nd quarter of 2017 and number of observations is 1,188). The result of the conventional estimation method is shown in Appendix Chart 1 and that of the new method using AEN is shown in Appendix Chart 2. In the following, we estimate how much changes in the sample period affecting the estimation results for the old and new methods.

First, we summarize how the parameters change when the samples are one year older. We calculate how much the contribution of each variable (calculated by the same procedure as in Chart 8) changes due to the replacement to the older sample. The results for both the old and new methods are shown respectively in Appendix Chart 3.[24]

The difference of contribution rate in variable l = ${\pi_l}^{func,OLD} - {\pi_l}^{func,NEW}$ (A1)

$$ {\pi_l}^{func,smpl} = \frac{y^{func,smpl}(\bar{x}_l + \Delta x_l, \bar{\boldsymbol{x}}_{-l}) - y^{func,smpl}(\bar{x}_l, \bar{\boldsymbol{x}}_{-l})}{y^{func,smpl}(\bar{x}_l, \bar{\boldsymbol{x}}_{-l})} \times 100 \qquad \text{(A2)} $$

${\pi_l}^{func,smpl}$: contribution rate of x_l (*func* = AEN or Box-Cox, *smpl* = NEW or OLD)

$y^{func,smpl}$: theoretical price (*func* = AEN or Box-Cox, *smpl* = NEW or OLD)

[24] For the variables that are adopted in either one model, the contribution rate of the other one is taken zero.

As shown in Appendix Chart 3, the differences in parameters between datasets occur within ±5 percentage points for the new method using AEN, however, within ±20 percentage points for the old method. The result suggests that each parameter is more stable in the new method than the old one.

However, the stability of the parameters for each variable does not necessarily lead to the stability of the results of quality adjustment immediately. Even if the changes in individual parameters are small, the rate of quality change (the rate of change in the theoretical price) can be great when the sign of parameters is same. We can say that for the opposite situation as well because parameter increase of some variables may well be offset by parameter decrease of another variable, under the situation where there is correlation between variables. Therefore, we have to pay attention to how the rate of quality change indicated from the theoretical prices differ when we evaluate the stability of the results of quality adjustment. Based on these points, Hirakata (2005) estimates the regression model for desktop computers in several functional forms, and then compares the results of quality adjustment using regression models with different sample periods to analyze how the price index can change. In the following, we follow Hirakata's (2005) method and compare the stability of the rate of quality change between the old and new methods, taking a sample price replacement in a passenger car as an example.

First, for the data set used in Chart 9, we extract samples classified as sedans or wagons and group them by released date quarterly.[25] Then, we build a hypothetical sample product where all variables are set at the mean value over the group for each quarter. Finally, we calculate the quality improvement rate due to the sample price replacement between these hypothetical products for the new and conventional method.[26,27] By comparing the quality improvement rates obtained in this way between the original sample and the old sample, it is possible to comprehensively consider the impact of

---

[25] The period examined is from the 3rd quarter of 2016 to the 2nd quarter of 2019. The 2nd quarter of 2017, the 1st quarter of 2018 and the 1st quarter of 2019 are not subject to the estimation because there are not sedans and wagons.

[26] When calculating the quality change rate, brand dummy and time dummy is fixed to a single value, not the average.

[27] In total, 144 (= 36($_9C_2$)×2 methods (AEN or Box-Cox)×2 datasets (NEW or OLD)) different quality improvement rates are calculated.

changes in the regression model on the price index, taking into account the correlation between variables.

$$\Pi^{func,smpl} = \frac{y^{func,smpl}(\overline{\boldsymbol{x}}') - y^{func,smpl}(\overline{\boldsymbol{x}})}{y^{func,smpl}(\overline{\boldsymbol{x}})} \times 100 \tag{A3}$$

$\Pi^{func,smpl}$: the rate of quality change（*func* = AEN or Box-Cox, *smpl* = NEW or OLD）

$y^{func,smpl}$: theoretical price（*func* = AEN or Box-Cox, *smpl* = NEW or OLD）

$\overline{\boldsymbol{x}}$: specification of hypothetical sample（before model change）

$\overline{\boldsymbol{x}}'$: specification of hypothetical sample（after model change）

In Appendix Chart 4, we compare the rate of quality change for a hypothetical model change between the old and new methods, where the horizontal axis shows the estimation result $\Pi^{func,NEW}$ and the vertical axis shows $\Pi^{func,OLD}$. The scatter plots of the new method are roughly distributed around the diagonal, while that of the conventional method are far from the diagonal. This suggests that the deviation in the rate of quality change caused by change in the estimation period is smaller when the new method is applied. In order to evaluate this point quantitatively, the deviation (absolute value) of the rate of quality change with the change in the estimation period, calculated as follows, is shown for each the new and conventional methods in Appendix Chart 5.

$$\text{The deviation in the rate of quality change} = \left|\Pi^{func,OLD} - \Pi^{func,NEW}\right| \tag{A4}$$

$$\Pi^{func,smpl} = \frac{y^{func,smpl}(\overline{\boldsymbol{x}}') - y^{func,smpl}(\overline{\boldsymbol{x}})}{y^{func,smpl}(\overline{\boldsymbol{x}})} \times 100 \tag{A5}$$

$\Pi^{func,smpl}$: the rate of quality change（*func* = AEN or Box-Cox, *smpl* = NEW or OLD）

$y^{func,smpl}$: theoretical price（*func* = AEN or Box-Cox, *smpl* = NEW or OLD）

$\overline{\boldsymbol{x}}$: specification of hypothetical sample（before model change）

$\overline{\boldsymbol{x}}'$: specification of hypothetical sample（after model change）

The deviation of the quality improvement rates in new method is roughly half of that of the conventional method on average for the entire period. This indicates that the application of AEN has increased the stability of the estimation results. This is consistent with that the fit to out-of-sample is good as well as in-sample in the new method, as confirmed in the main text. This suggests that, in the new method, the estimation error in the quality improvement rate tends to remain relatively small, even when the relationship between price and characteristics changes over time.

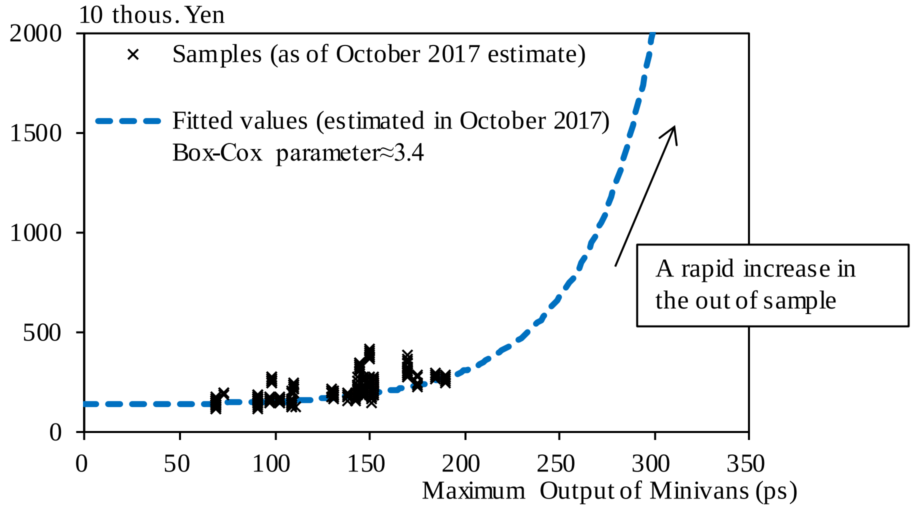
In addition, in applying the quality adjustment, we put the first priority on evaluating whether the product quality improves or deteriorates. In this regard, Appendix Chart 4 shows that in some samples of model change, as circled in red, the sign of quality change rate is clearly different due to the change of the estimation period in the conventional method. This suggests that due to the regression model obsoleting over time, we may wrongly evaluate that the quality is deteriorate (improve) from the hedonic model, even though the quality in fact improves (deteriorate) when applying the quality adjustment. On the other hand, in the new method, there are few cases where the sign of quality change rates reverse between the estimation models. This improvement with the introduction of AEN could also lead to an increasing the applicability of the hedonic quality adjustment.

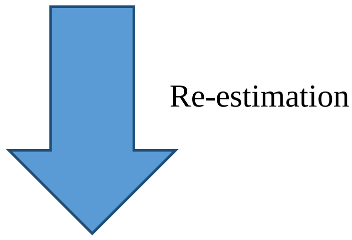
Correlation coefficients of variables for passenger cars

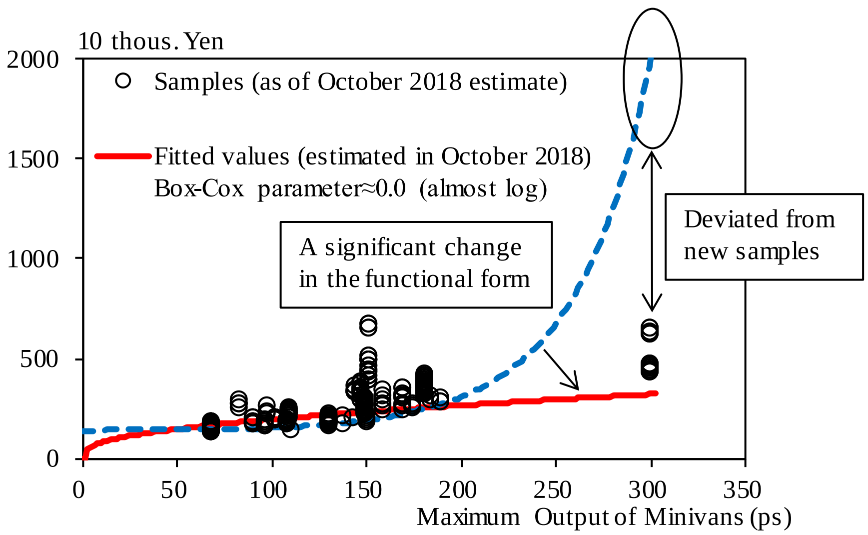
	SC	L	W	H	WT	FE	MO	MT	RS	NG
SC	1.000									
L	0.420	1.000								
W	0.223	0.849	1.000							
H	0.787	0.317	0.251	1.000						
WT	0.521	0.885	0.844	0.553	1.000					
FE	-0.252	-0.501	-0.618	-0.268	-0.563	1.000				
MO	0.017	0.628	0.696	0.018	0.665	-0.661	1.000			
MT	0.008	0.599	0.711	-0.006	0.618	-0.535	0.812	1.000		
RS	-0.051	0.625	0.785	0.004	0.584	-0.489	0.645	0.711	1.000	
NG	-0.058	0.198	0.152	-0.081	0.213	-0.080	0.409	0.319	0.159	1.000

SC: Seating Capacity, L: Length, W: Width, H: Height, WT: Weight, FE: Fuel Efficiency,
MO: Maximum Output, MT: Maximum Torque, RS: Rim Size, NG: Number of Gears

Change in the functional form
(Maximum output of Minivans)







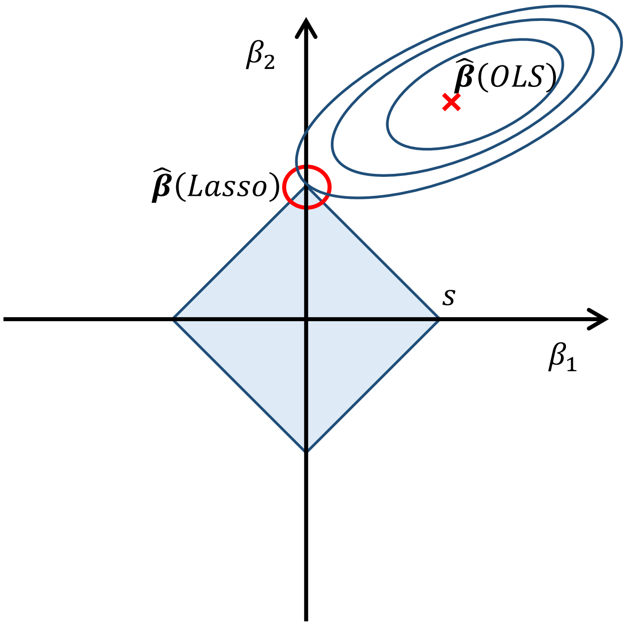
Schema of sparse estimation

Lasso

$$\underset{\beta_1,\beta_2}{\operatorname{argmin}} \sum_{i=1}^{n} \left(Y_i - \beta_1 X_{1,i} - \beta_2 X_{2,i}\right)^2$$

s.t. $|\beta_1| + |\beta_2| \leq s$

s $>$ 0: 1 to 1 corresponding to λ

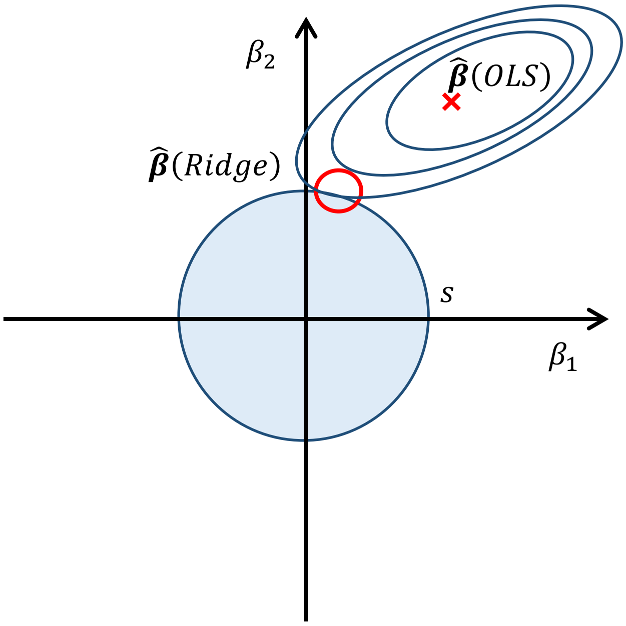


Ridge Regression

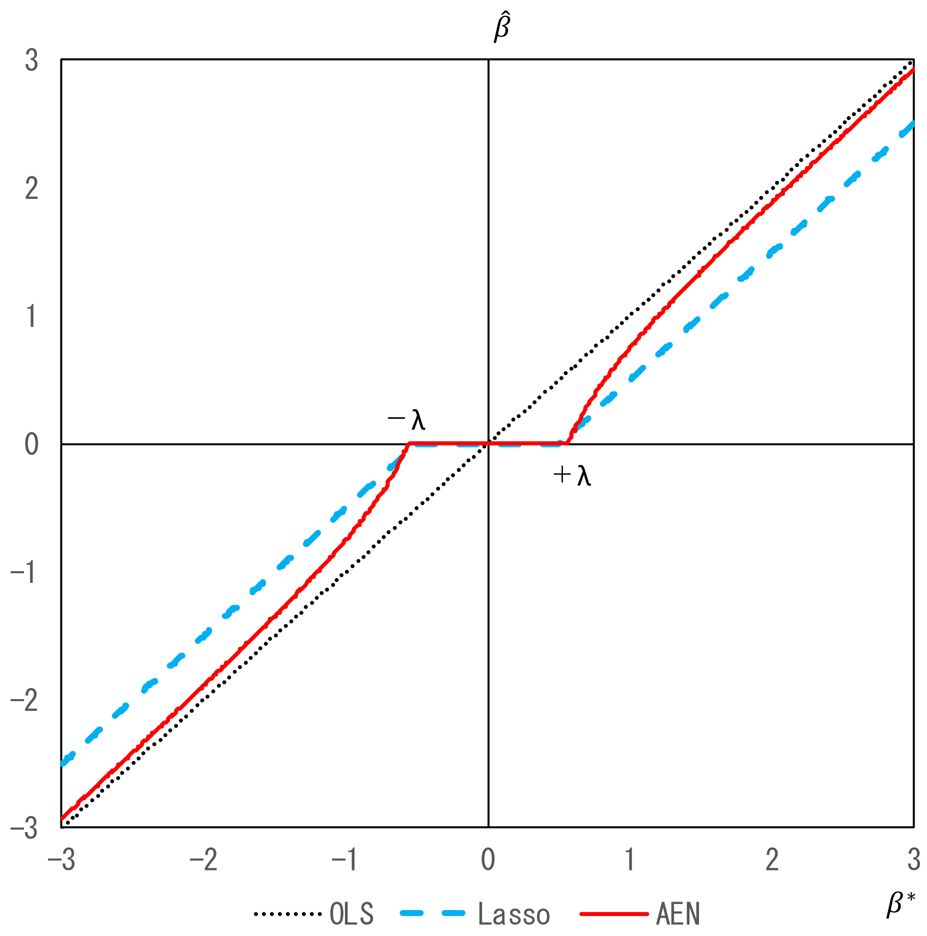
$$\underset{\beta_1,\beta_2}{\operatorname{argmin}} \sum_{i=1}^{n} \left(Y_i - \beta_1 X_{1,i} - \beta_2 X_{2,i}\right)^2$$

s.t. ${\beta_1}^2 + {\beta_2}^2 \leq s^2$

s $>$ 0: 1 to 1 corresponding to λ



Statistical properties of AEN



Notes: 1. The estimated $\hat{\beta}$ are plotted with each estimation method using the artificial data, where $\boldsymbol{X}$ = 120,200 × 601 design matrix, $\beta_i{}^* = -3 + 0.01\mathrm{i}$ $(\mathrm{i} = 0 \sim 600)$, $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{\varepsilon}$ $(\boldsymbol{\varepsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{I}))$. Note that the mean is set to 0 and the standard deviation to 1 for all columns of $\boldsymbol{X}$.

2. $\lambda = 0.5$ for Lasso and $\lambda_1 = \lambda_1{}^* = 0.2$, $\lambda_2 = 0.001$, $\gamma = 0.5$ for AEN

Candidate variables for passenger car

List of Candidate Variables					
Continuous Variables	Seating Capacity (person)	Spec Dummy Variables	ETC	Spec Dummy Variables	Rear View Camera
	Length (mm)		Navigation System		Side View Camera
	Width (mm)		DVD Player		Front View Camera
	Height (mm)		Blu-ray Player		Surround View Camera
	Weight (kg)		AM/FM Radio		AFS
	Wheelbase (mm)		USB Input		Hill Start Assist
	Minimum Turning Radius (m)		No Idling		Cold Climate Version
	Fuel Efficiency (JC08 mode, km/l)		Full Auto Air Conditioner		Rain Sensor
	Fuel Tank Capacity (l)		Dual Zone Air Conditioner		Anti-Theft System
	Maximum Output (ps)		Front Dual Zone Air Conditioner	Body-Type Dummy Variables	Sedan
	Maximum Torque (kg·m)		Driver's Seat Heater		Wagon
	Number of Cylinders (#)		Driving Position Memory System		Coupe
	Total Displacement (cc)		Split-Folding Rear Seat		Convertible
	Rim Size (inch)		Front Power Seat		Minivan
	Tire Width (mm)		Passenger's Power Seat		SUV
	Tire Flatness (%)		Rear Power Seat		Hatchback
	Number of Gears (#)		Leather Seat	Manufacturer Dummy Variables	Domestic Car A
	Indoor Space (m^3)		Leather Steering		Domestic Car B
Spec Dummy Variables	Diesel		Telescopic Steering Device		Domestic Car C
	Hybrid		Steering Controller		Domestic Car D
	Plug-In Hybrid		Wood Panel		Domestic Car E
	Unleaded Premium Gasoline		Aluminum Wheel		Domestic Car F
	Turbo		LED Headlamp		Domestic Car G
	Supercharger		LED Fog Lamp		Domestic Car H
	Twin-Turbo		Front Fog Lamp		Domestic Car I
	Flat Engine		Rear Fog Lamp		Imported Car A
	FF		Xenon Headlamp		Imported Car B
	FR		Projector Headlamp		Imported Car C
	Full-Time 4WD		LSD		Imported Car D
	Part-Time 4WD		Cruise Control	Time Dummy Variables	2016Q3
	AT		ACC		2016Q4
	MT		ACC (No speed limitation)		2017Q1
	CVT		Clearance Sonar		2017Q2
	Front Spoiler		LDWS		2017Q3
	Rear Spoiler		LKAS		2017Q4
	Rear Window Wiper		Traction Control		2018Q1
	Sunroof		Unintended Start Prevention		2018Q2
	Glasstop		AEBS		
	Privacy Glass		Brake Assist		
	Side Airbag		Parking Assist		

Estimation result with conventional method

Estimated Model		Double Box-Cox Model
Box-Cox Parameter of Dependent Variable		-0.280
Intercept		3,472.763 ***
Room Space (m^3)	Sedans & Station Wagons	--
	Box-Cox Parameter	--
	Minivans	1.360E-05 ***
	Box-Cox Parameter	3.400
Fuel Efficiency JC08 (km/l) ×Equivalent Inertia Weight (kg)	Sedans & Station Wagons	2.543E-09 ***
	Box-Cox Parameter	1.372
	Minivans	1.606E-09 ***
	Box-Cox Parameter	1.455
	SUVs	6.841E-09 ***
	Box-Cox Parameter	1.330
	Hatchbacks	7.152E-18 ***
	Box-Cox Parameter	3.351
Horsepower (PS)	Sedans & Station Wagons	2.846E-04 ***
	Box-Cox Parameter	0.647
	Minivans	0.007 ***
	Box-Cox Parameter	6.240E-06
	SUVs	5.880E-06 ***
	Box-Cox Parameter	1.337
	Hatchbacks	0.008 ***
	Box-Cox Parameter	3.621E-06
Dummy Variables		
Car Configuration		
Minivans		-1,162.565 ***
SUVs		0.006 ***
Hatchbacks		-2,306.764 ***
Motor		
Hybrid Vehicles		--
Plug-in Hybrid Electric Vehicles		--
Powertrain		
AWD (Full time or Part time)		0.002 ***
FR (Front-engine, rear-wheel-drive)		0.002 ***
Standard Equipment		
Leather Seats		0.001 ***
Side Airbags		4.504E-04 **
Power Seats		0.002 ***
Aluminum Wheel		0.002 ***
LED Headlamp		0.001 ***
Privacy Glass		--
Limited Slip Differential (LSD)		0.002 ***
Advanced Emergency Braking System (AEBS)		--
Adaptive Cruise Control (ACC)		--
Adaptive Cruise Control (ACC) <No speed limitation>		0.001 ***
Lane Departure Warning System (LDWS)		0.001 ***
Adaptive Front-Lighting System (AFS)		0.001 ***
Parking Assist		0.001 ***
Brand		
Brand A		-0.002 ***
Brand B		-0.003 ***
Brand C		--
Brand D		--
Brand E		-0.001 ***
Brand F		0.004 ***
Brand G		0.003 ***
Brand H		--
Brand I		0.006 ***
Brand J		0.008 ***
Brand K		0.006 ***
R-squared		0.957
Adjusted R-squared		0.956
Standard Error of Regression		0.002
Mean of Dependent Variable		3.509
Number of Observations		1,155
(release period)		(from 3Q 2016 to 2Q 2018)
Tests for Double Box-Cox Model		
(H_1: Double Box-Cox)		
H_0: Semi Box-Cox (λ_i=1)		85.560 ***
H_0: Log Linear (λ_0=λ_i=0)		273.705 ***
H_0: Semi Log Linear (λ_0=0,λ_i=1)		130.257 ***
H_0: Linear (λ_0=λ_i=1)		1,905.192 ***

Source: Bank of Japan

Notes: 1. The equivalent inertia weight of a vehicle is measured as its curb weight with an additional 110kg of weight to a vehicle, which is set to chassis dynamometer while measuring its fuel efficiency under JC08 emission test cycle.

2. In addition to the explanatory variables listed above, the model includes release period dummy variables.

Estimation result with new method

Explanatory Variables	
Constant	12.939
Imported Car C	0.244
Imported Car A	0.204
Supercharger	0.119
Navigation System	0.142
Rear Power Seat	0.053
Aluminum Wheel	0.032
LDWS	0.007
Blu-ray Player	0.057
Population Density (person/m^2)	-0.055
Curb Weight(kg)/Volume(m^3)	0.001
Rim Size (inch)	0.001
FF	-0.025
Curb Weight(kg)/Volume(m^3)×: quadratic term	2.550E-05
Length(m)×Width(m)×: quadratic term	0.009
Maximum Output (ps)/Weight(kg)×: quadratic term	4.854
Domestic Car G×2017Q4	-0.003
Domestic Car G×Front Fog Lamp	-0.034
Domestic Car G×Front Spoiler	-0.056
Domestic Car E×2018Q2	0.248
Domestic Car E×Leather Seat	0.018
Domestic Car E×AFS	0.032
Domestic Car E×Maximum Output (ps)/Weight(kg)	0.865
Domestic Car D×2016Q4	0.350
Domestic Car D×2017Q3	-0.012
Domestic Car D×Hatchback	-0.003
Domestic Car D×Minivan	-0.135
Domestic Car D×Height (mm)	-4.036E-05
Imported Car B×Maximum Torque (kg·m)	0.003
Imported Car B×Full Auto Air Conditioner	0.117
Imported Car B×Rim Size (inch)	0.005
Domestic Car F×2017Q3	-0.130
Domestic Car F×Hybrid	-0.061
Domestic Car C×CVT	-0.135
Domestic Car B×CVT	-0.053
Domestic Car B×Xenon Headlamp	0.088
Imported Car A×2017Q3	0.073
2016Q4×Maximum Torque (kg·m)	-0.003

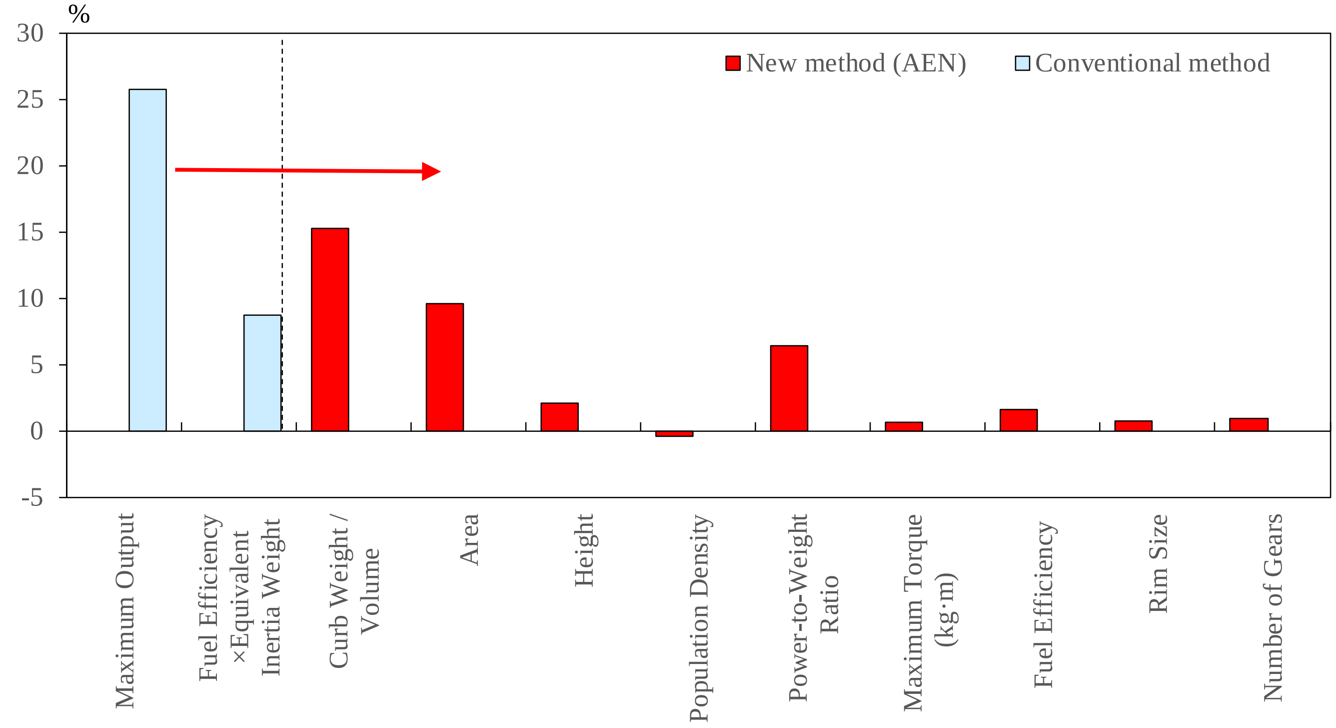
Explanatory Variables	
2016Q4×Leather Seat	-0.031
2017Q3×Height (mm)	-1.943E-05
2017Q4×CVT	-0.038
2018Q1×Front Fog Lamp	0.005
2018Q2×FF	-0.037
Coupe×Maximum Output (ps)/Weight(kg)	0.223
Hatchback×Maximum Output (ps)/Weight(kg)	-0.315
Height (mm)×Hybrid	4.580E-05
Height (mm)×Front Fog Lamp	1.029E-05
Height (mm)×Dual Zone Air Conditioner	3.248E-05
Height (mm)×Sunroof	1.827E-06
Height (mm)×Driver's Seat Heater	4.776E-06
Height (mm)×Driving Position Memory System	2.691E-05
Height (mm)×ACC (No speed limitation)	1.741E-05
Height (mm)×Curb Weight(kg)/Volume(m^3)	2.412E-06
Fuel Efficiency (JC08 mode, km/l)×Maximum Torque (kg·m)	1.906E-05
Fuel Efficiency (JC08 mode, km/l)×Front Spoiler	0.001
Fuel Efficiency (JC08 mode, km/l)×Navigation System	2.649E-04
Fuel Efficiency (JC08 mode, km/l)×Dual Zone Air Conditioner	0.001
Fuel Efficiency (JC08 mode, km/l)×Leather Steering	0.001
Fuel Efficiency (JC08 mode, km/l)×Sunroof	9.915E-05
Fuel Efficiency (JC08 mode, km/l)×Driver's Seat Heater	0.001
Maximum Torque (kg·m)×Rear Spoiler	0.001
Maximum Torque (kg·m)×Maximum Output (ps)/Weight(kg)	2.818E-04
Unleaded Premium Gasoline×Front Spoiler	0.024
Unleaded Premium Gasoline×Rear Spoiler	0.002
MT×LED Headlamp	0.019
Number of Gears (#)×Rim Size (inch)	0.001
LSD×Leather Steering	0.015
LSD×Leather Seat	0.049
Cruise Control×Curb Weight(kg)/Volume(m^3)	3.507E-04
Leather Seat×Length(m)×Width(m)	0.004
LED Headlamp×Curb Weight(kg)/Volume(m^3)	4.970E-04

Hyperparameters	
λ_1	0.013
λ_1^*	1.970E-05
λ_2	1.000E-05
γ	0.5

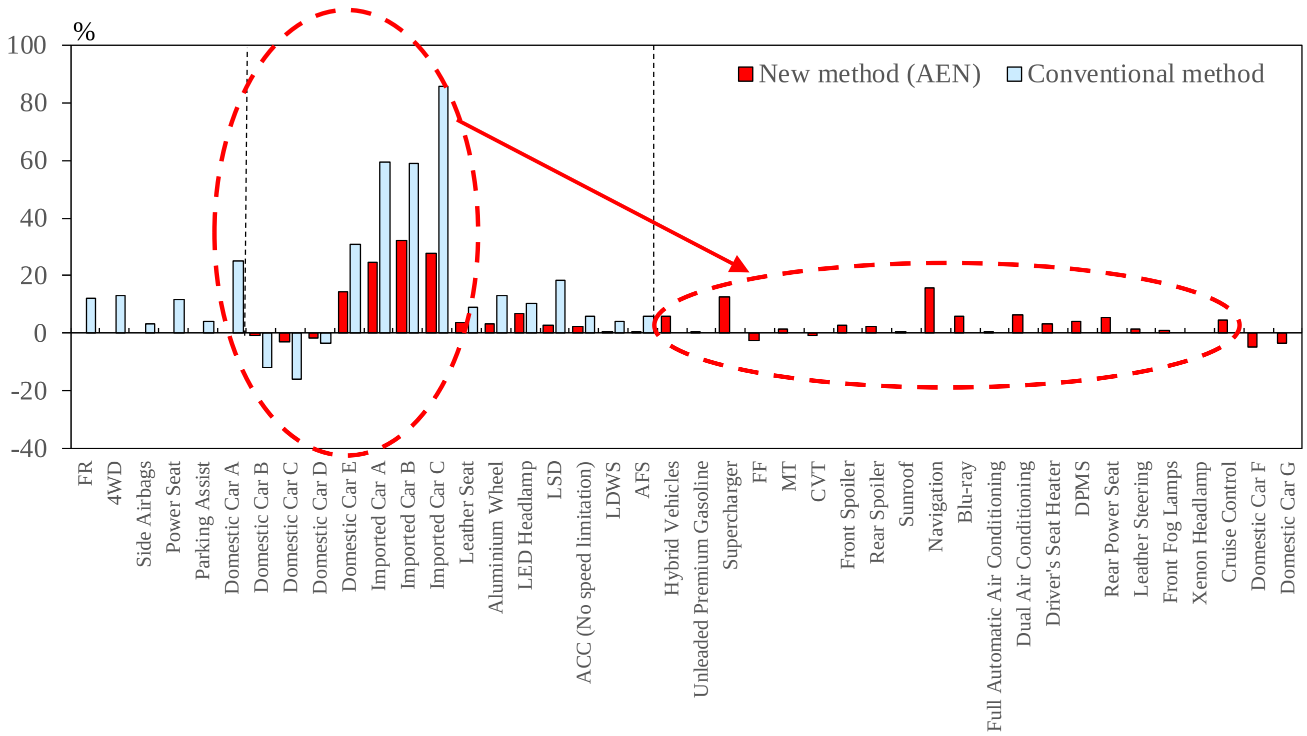
Notes: The sample period is from the 3rd quarter of 2016 to the 2nd quarter of 2018. Volume＝Length×Width×Height.
Population Density＝seating capacity÷(Length×Width).

Contribution of variables to the theoretical price

1. Continuous variables



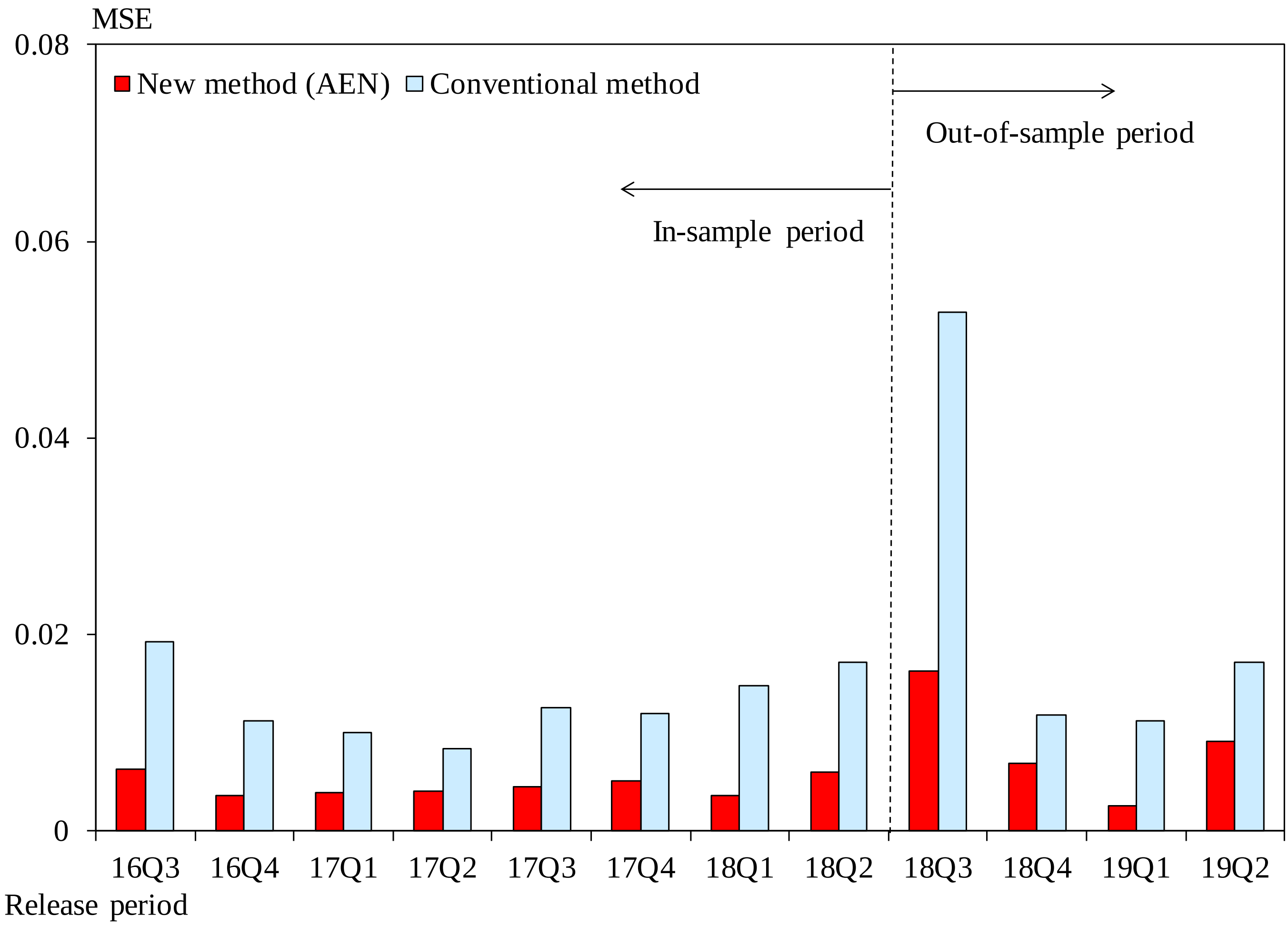
2. Dummy variables



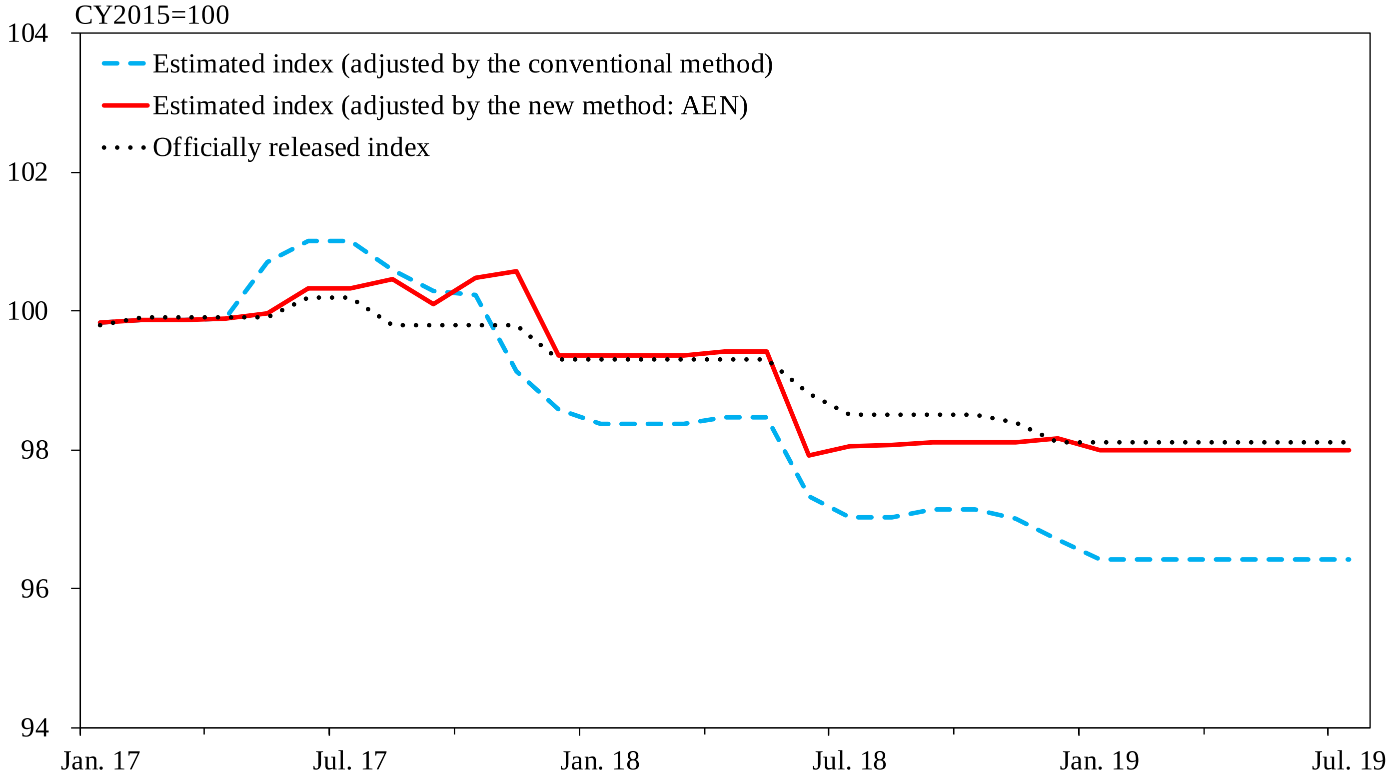
Notes: In addition to the variables listed above, the model includes dummy variables for car configuration and release period.

Chart 9

Comparison of fit between old and new methods



Estimated price index by old and new methods
PPI “Standard Passenger Cars (Gasoline Cars)”



Estimation result with conventional method（old sample）

Estimated Model		Double Box-Cox Model
Box-Cox Parameter of Dependent Variable		0.150
Intercept		1,664.131 ***
Room Space (㎥)	Sedans & Station Wagons	2.433 ***
	Box-Cox Parameter	1.066
	Minivans	0.039 ***
	Box-Cox Parameter	2.770
Fuel Efficiency JC08 (km/l) ×Equivalent Inertia Weight (kg)	Sedans & Station Wagons	9.512E-08 ***
	Box-Cox Parameter	1.637
	Minivans	1.754E-09 ***
	Box-Cox Parameter	2.097
	SUVs	3.147 ***
	Box-Cox Parameter	0.002
	Hatchbacks	1.951E-26 ***
	Box-Cox Parameter	5.773
Horsepower (PS)	Sedans & Station Wagons	5.993 ***
	Box-Cox Parameter	0.003
	Minivans	3.825E-07 ***
	Box-Cox Parameter	3.384
	SUVs	3.243 ***
	Box-Cox Parameter	0.040
	Hatchbacks	4.408 ***
	Box-Cox Parameter	0.018
Dummy Variables		
Car Configuration		
Minivans		2,275.621 ***
SUVs		819.239 ***
Hatchbacks		2,019.161 ***
Motor		
Hybrid Vehicles		0.393 ***
Plug-in Hybrid Electric Vehicles		2.137 ***
Powertrain		
AWD (Full time or Part time)		0.846 ***
FR (Front-engine, rear-wheel-drive)		--
Standard Equipment		
Leather Seats		1.003 ***
Side Airbags		0.559 ***
Power Seats		0.869 ***
Aluminum Wheel		--
LED Headlamp		--
Privacy Glass		0.782 ***
Limited Slip Differential (LSD)		0.630 ***
Advanced Emergency Braking System (AEBS)		0.358 ***
Adaptive Cruise Control (ACC)		0.405 ***
Adaptive Cruise Control (ACC) <No speed limitation>		--
Lane Departure Warning System (LDWS)		0.184 **
Adaptive Front-Lighting System (AFS)		0.624 ***
Parking Assist		--
Brand		
Brand A		-1.557 ***
Brand B		-1.353 ***
Brand C		-0.523 ***
Brand D		-1.803 ***
Brand E		-1.237 ***
Brand F		2.648 ***
Brand G		-0.896 ***
Brand H		-0.611 ***
Brand I		2.987 ***
Brand J		4.723 ***
Brand K		3.825 ***
R-squared		0.962
Adjusted R-squared		0.961
Standard Error of Regression		0.792
Mean of Dependent Variable		52.810
Number of Observations (release period)		994 (from 3Q 2015 to 2Q 2017)
Tests for Double Box-Cox Model (H_1: Double Box-Cox)		
H_0: Semi Box-Cox (λ_i=1)		220.310 ***
H_0: Log Linear (λ_0=λ_i=0)		158.589 ***
H_0: Semi Log Linear (λ_0=0,λ_i=1)		238.038 ***
H_0: Linear (λ_0=λ_i=1)		641.781 ***

Source: Bank of Japan

Notes: 1. The equivalent inertia weight of a vehicle is measured as its curb weight with an additional 110kg of weight to a vehicle, which is set to chassis dynamometer while measuring its fuel efficiency under JC08 emission test cycle.

2. In addition to the explanatory variables listed above, the model includes release period dummy variables.

Estimation result with new method（old sample）

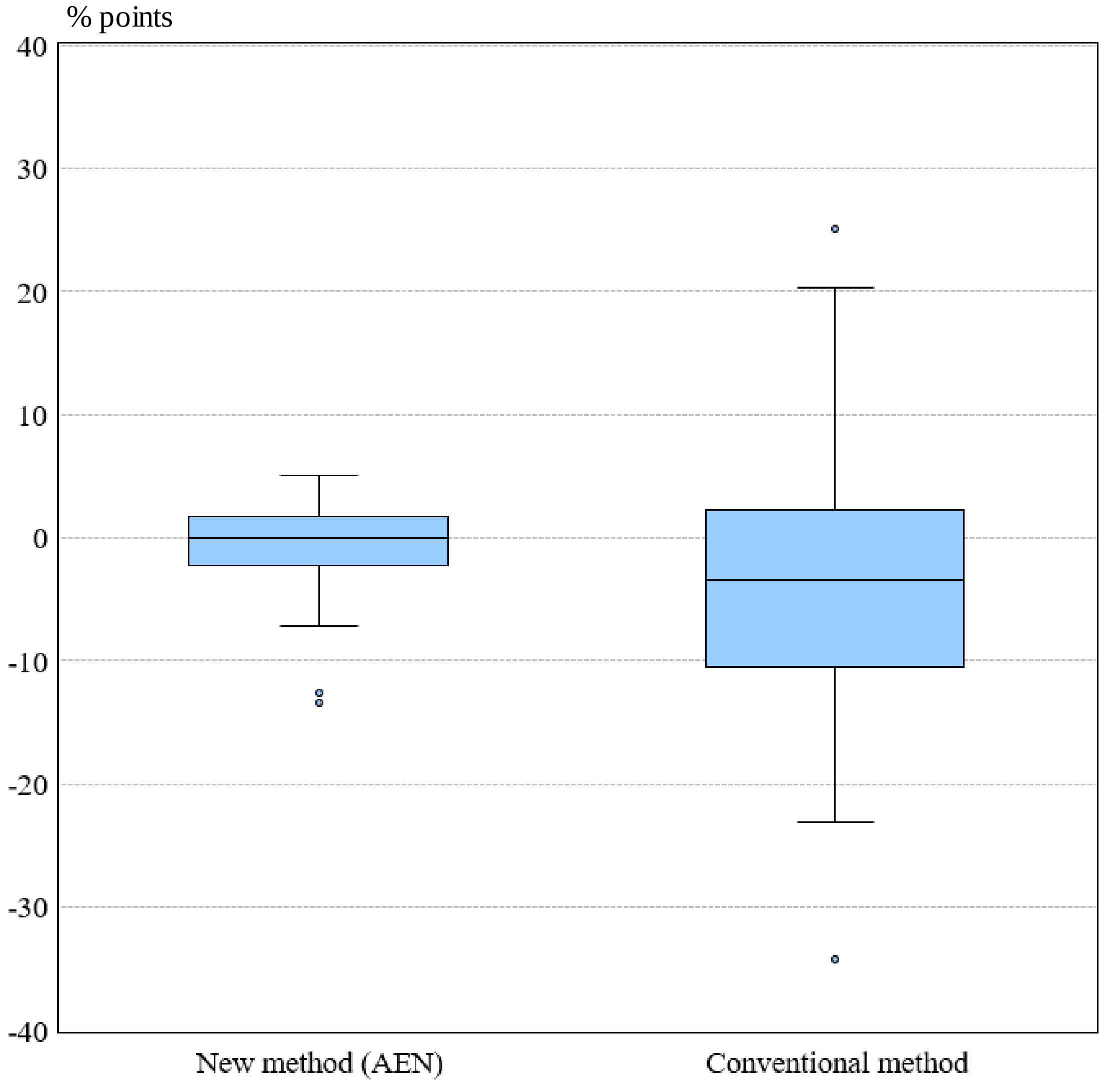
Explanatory Variables	
Constant	12.879
Population Density (person/m^2)	-0.192
Curb Weight(kg)/Volume(m^3)	0.002
Length(m)×Width(m)	0.032
Navigation System	0.039
Leather Seat	1.432E-05
Rear Power Seat	0.091
Length(m)×Width(m)×: quadratic term	0.002
Imported Car C×Height (mm)	1.123E-05
Imported Car C×Maximum Output (ps)/Weight(kg)	2.093
Domestic Car G×CVT	-0.025
Domestic Car G×Aluminum Wheel	-0.067
Domestic Car E×Maximum Output (ps)/Weight(kg)	1.211
Domestic Car E×Driver's Seat Heater	0.014
Domestic Car D×2016Q3	-0.034
Domestic Car D×2016Q4	0.203
Domestic Car D×Minivan	-0.019
Domestic Car D×Population Density (person/m^2)	-0.114
Imported Car B×Maximum Output (ps)/Weight(kg)	1.016
Imported Car B×Maximum Torque (kg·m)	0.002
Imported Car B×Full Auto Air Conditioner	0.137
Imported Car B×Wood Panel	0.018
Imported Car B×Privacy Glass	0.011
Domestic Car F×2016Q1	-0.068
Domestic Car F×Fuel Efficiency (JC08 mode, km/l)	-0.001
Domestic Car C×CVT	-0.146
Domestic Car B×2015Q4	-0.124
Domestic Car B×Hatchback	-0.074
Imported Car A×Leather Steering	0.145
2016Q3×Cruise Control	0.009
2016Q3×Wood Panel	0.012
2016Q3×ETC	0.056
2016Q3×Driver's Seat Heater	0.019
Hatchback×Population Density (person/m^2)	-0.047
Minivan×Hybrid	0.043
Minivan×Driver's Seat Heater	0.029
Population Density (person/m^2)×Front Dual Zone Air Conditioner	0.027
Curb Weight(kg)/Volume(m^3)×Length(m)×Width(m)	2.605E-04

Explanatory Variables	
Curb Weight(kg)/Volume(m^3)×Height (mm)	2.436E-06
Curb Weight(kg)/Volume(m^3)×Rim Size (inch)	9.914E-05
Curb Weight(kg)/Volume(m^3)×Telescopic Steering Device	1.396E-04
Curb Weight(kg)/Volume(m^3)×Anti-Theft System	3.299E-04
Curb Weight(kg)/Volume(m^3)×Blu-ray Player	3.530E-04
Length(m)×Width(m)×Rim Size (inch)	0.002
Height (mm)×Leather Seat	3.468E-05
Height (mm)×LED Headlamp	3.315E-05
Height (mm)×LDWS	4.677E-06
Fuel Efficiency (JC08 mode, km/l)×Hybrid	0.001
Fuel Efficiency (JC08 mode, km/l)×Navigation System	0.002
Fuel Efficiency (JC08 mode, km/l)×Dual Zone Air Conditioner	0.002
Fuel Efficiency (JC08 mode, km/l)×Cruise Control	0.001
Fuel Efficiency (JC08 mode, km/l)×Leather Steering	0.001
Fuel Efficiency (JC08 mode, km/l)×Front Power Seat	0.001
Fuel Efficiency (JC08 mode, km/l)×Driver's Seat Heater	0.001
Hybrid×Rain Sensor	0.013
Maximum Output (ps)/Weight(kg)×Maximum Torque (kg·m)	0.007
Maximum Output (ps)/Weight(kg)×Unleaded Premium Gasoline	0.167
Maximum Output (ps)/Weight(kg)×FF	-0.312
Maximum Output (ps)/Weight(kg)×Navigation System	0.065
Maximum Output (ps)/Weight(kg)×Dual Zone Air Conditioner	0.143
Maximum Output (ps)/Weight(kg)×Wood Panel	0.268
Maximum Output (ps)/Weight(kg)×Aluminum Wheel	0.591
Maximum Output (ps)/Weight(kg)×Side Airbag	0.193
Maximum Output (ps)/Weight(kg)×Anti-Theft System	0.009
Maximum Output (ps)/Weight(kg)×Blu-ray Player	0.010
Front Fog Lamp×AEBS	0.014
Aluminum Wheel×AEBS	0.006

Hyperparameters	
λ_1	0.017
λ_1^*	2.068E-05
λ_2	0.010
γ	0.5

Notes: The sample period is from the 3rd quarter of 2015 to the 2nd quarter of 2017. Volume＝Length×Width×Height.
Population Density＝seating capacity÷(Length×Width).

Change in contribution rate of each variable between sample periods

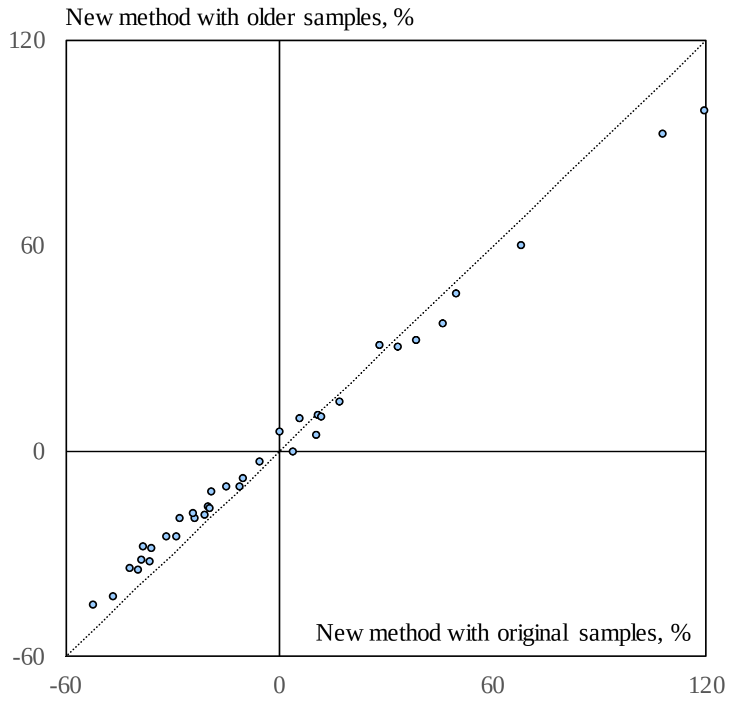


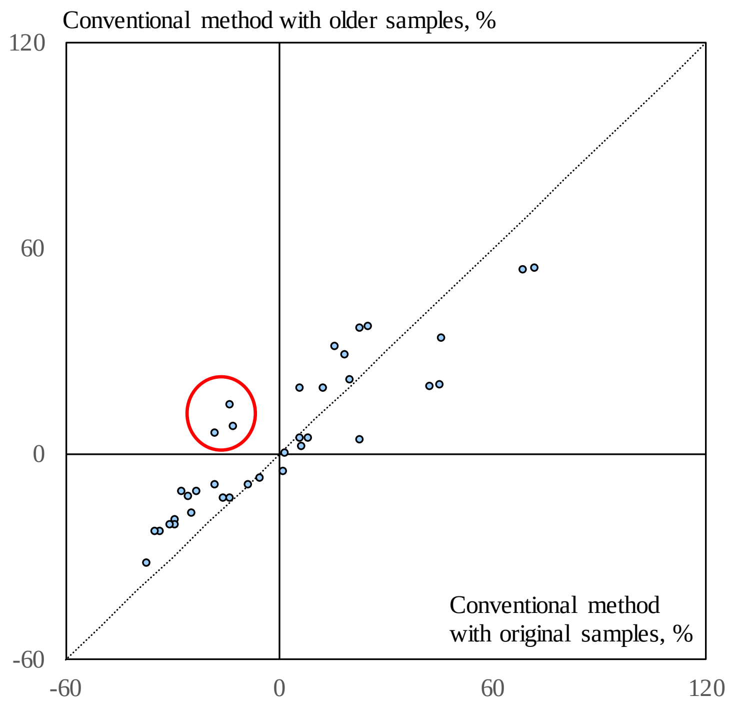
Notes: 1. We build hypothetical sample prices where all variables are set at the mean value for sedans and wagons. For these samples, we calculate the rate of change in theoretical price due to one standard deviation increase in continuous variables or one unit increase in dummy variables with the regression models derived from each original and older dataset. Here, we show the difference between the contribution rates for each variables from original and older dataset. Dummy variables regarding body type and release period are not subject to the calculation.

2. Numbers of variables subject to the calculation are 55 for the new method (AEN) and 31 for the conventional method. For the variables that are adopted in either one model, the contribution rate of the other one is taken zero.

3. The dotted plots, values below “the first quartile－1.5×the quartile range” or above “the third quartile＋1.5×the quartile range” are indicated as outliers.

Change in quality change rate between sample periods





Notes: Assume that a model change occurs from a hypothetical product with average specifications that is launched in one quarter to a hypothetical product built in the same way for another quarter. We calculate the rate of quality change (the rate of change in theoretical price) with the regression models derived from each original and older dataset and plot the combinations.

Deviation in quality change rate between sample periods

Old / New model	New method (AEN)							
	16Q4	17Q1	17Q3	17Q4	18Q2	18Q3	18Q4	19Q2
16Q3	4.6	1.8	3.3	4.3	0.1	3.8	6.6	2.9
16Q4	-	5.3	6.2	7.0	3.9	2.1	9.5	6.1
17Q1	-	-	2.3	3.6	2.6	8.4	6.3	1.6
17Q3	-	-	-	1.7	6.9	15.8	4.9	1.2
17Q4	-	-	-	-	9.8	20.3	3.1	3.3
18Q2	-	-	-	-	-	4.1	7.5	3.3
18Q3	-	-	-	-	-	-	6.5	3.8
18Q4	-	-	-	-	-	-	-	6.2
Average	5.3							

Old / New model	Conventional method							
	16Q4	17Q1	17Q3	17Q4	18Q2	18Q3	18Q4	19Q2
16Q3	4.6	8.3	9.2	7.8	6.8	1.2	6.8	20.4
16Q4	-	11.6	11.9	10.7	2.5	6.2	9.8	23.6
17Q1	-	-	2.4	0.7	18.8	12.0	1.0	12.7
17Q3	-	-	-	1.9	25.6	18.5	4.1	11.2
17Q4	-	-	-	-	22.8	15.2	2.0	13.6
18Q2	-	-	-	-	-	9.9	12.6	27.8
18Q3	-	-	-	-	-	-	4.9	16.1
18Q4	-	-	-	-	-	-	-	15.3
Average	10.8							

Notes: Assume that a model change occurs from a hypothetical product with average specifications that is launched in one quarter to a hypothetical product built in the same way for another quarter. We calculate the rate of quality change (the rate of change in theoretical price) with the regression models derived from each original and older dataset and list the absolute value of the difference.