No. 907

Estimating the maintenance and repair cost in Life Cycle Cost Calculation: A case of automobile ownership

by

Vitoon Puripunyavanich, Satoshi Myojo,
Yuichiro Kanazawa

March 2001

Estimating the maintenance and repair cost in Life Cycle Cost Calculation: A case of automobile ownership

Vitoon Puripunyavanich*   Satoshi Myojo†

Yuichiro Kanazawa‡

February 27, 2001

Abstract

The maintenance and repair cost of durable goods has traditionally been hidden from consumers and yet has been non–negligible part of Life Cycle Cost (LCC) computation. Predicting the maintenance and repair cost is difficult because many of these durable goods do not have constant failure rates. For some durable products, it is often the case that we have at least a rough idea as to their reliability. In this study we propose and illustrate a method to convert that knowledge to monetary cost using automobile ownership as an example. In our

*Doctoral Program in Policy and Planning Sciences

†Doctoral Program in Policy and Planning Sciences

‡Associate Professor of Statistics, Institute of Policy and Planning Sciences, The University of Tsukuba. Correspondence: 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan. E-mail address: kanazawa@sk.tsukuba.ac.jp

method, we first estimate a statistical model from the widely available reliability data. Then we predict the reliability from the model. Finally we convert the predicted reliability to cost figure using the data from the U.S. Consumer Expenditure Survey. The proposed statistical model takes care of the possible bias introduced by partially missing reliability data likely to be caused by lower sales volume of some automobiles and higher response rates from owners of unreliable vehicles. Conversion to cost figure is done on the thirteen popular automobiles whose cost of repairs in major mechanical systems are well documented. We then calculate the LCC of these model cars and discuss the importance of maintenance and repair cost in the LCC.

1 Introduction

The "cost" of a product for consumers has traditionally been equated with its price at the time of purchase. For some products, however, this definition of cost can be quite misleading. For example, in the case of many durable goods, significant cost will be incurred in the use and maintenance of the product over a period of years. Life Cycle Cost (LCC), which includes all the costs associated with acquisition, use, maintenance, and disposal, is more reasonable alternative to evaluate such products.

Consumers themselves have become increasingly aware of not only the cost of acquiring but the cost associated with use, maintenance and disposal of the durable products. This awareness has been partially translated into the US regulations. For example, the US government, through the Energy Policy and Conservation Act (EPCA) of 1975, has been asking home appliance manufacturers to disclose energy consumption on their products (McNeill et al. 1979, Hutton et al. 1980).

The EPCA also included a "New Auto Fuel Economy Program," in which Department of Transportation (DOT) was directed to set "corporate average fuel economy" standard for new car starting in model year 1978, and for new light trucks starting in 1979. Each automaker was required to meet the standard, subject to large fines for non-compliance. The program put Environmental Protection Agency (EPA) in charge of measuring fuel economy for each model, of setting up National Vehicle and Fuel Emissions Laboratory to determine car manufacturers' compliance with federal emissions and fuel economy standards. The program asked Department of Energy (DOE) to publish the Fuel Economy Guide as an aid to consumers considering the purchase of a new car. The Guide lists estimates of miles per gallon (mpg) for each vehicle available for the new model year. These estimates have been provided by the EPA.

Of all durable products consumers purchase, automobile is without doubt the most expensive. For example, according to the U.S. Consumer Expenditure Survey in 1998, expenditure on vehicle purchases, gasoline and motor oil and other vehicle expenses amounted to $6,358 or 17.06% of average household expenditure of $37,260.

Consumers tend to have a pretty good idea on acquisition cost of automobile before the time of purchase from the sticker price and price quotation services from such organizations as Consumer Union. By looking at the window-stickers and using the Fuel Economy Guide, consumers can and are expected to roughly estimate the average yearly fuel cost for any vehicle.

It is difficult, however, to predict repair cost for a specific automobile because it varies from one model to another and the average maintenance and repair cost, for example, in the U.S. Consumer Expenditure Survey, does not apply to the particular automobile of consumer's choice. Until now there has been no regulation requiring Federal and State government office to estimate the repair cost of automobile.

There are state government regulations such as the Lemon Laws stipulating the manufacturers to take some responsibility to the defect of the product they manufactured. For example, California Lemon Law – CA Civil Code Section 1793.22 (Tanner Consumer Protection Act). Although the Lemon Laws like this one are protecting consumers in their first year of car ownership throughout the country, typical consumers still have at least seven more years to think about the repair cost, because "The median age of cars on the road in 1999 is more than 8 years, compared with $6\frac{1}{2}$ years in 1990" (April 1999 Consumer Reports page 97). Public can consult publications such as Consumer Reports or can access to their website for the frequency of repair of the specific make and model. In this paper we will propose and illustrate a method to convert that knowledge of the frequency of repair to "monetary repair cost."

In current practice, repair cost of a general product is estimated through the usage of databases and professional opinions (Taylor 1981). The reason for this is the fact that the repair cost depends on maintainability and reliability parameters. While most electronic components are considered to have constant failure rates (exponential distributions) - which coincidentally simplifies the mathematics for calculating the often used mean time before failure (MTBF) and mean time to repair (MTTR), reliability/maintainability of non-electronic components have non-constant failure rates and can lead to the unwary to intractable mathematics (Fricker 1979, Neumann 1983). Automobiles, as computerized as they may be, have significant mechanical components. This makes it very difficult to obtain theoretical model of the reliability and maintainability of automobiles, leaving us an only choice of statistical method for tracing them.

Our method is as follows: We first estimate a statistical model regressing the reliability summaries in five point ordinal scale published in the Consumer Reports on several design characteristics and price of automobile, and origin of the car manufacturers. We choose these explanatory variables because they are easily available to buyers thinking of purchasing a new car. Then we predict the reliability score from the model. Finally we convert our predicted reliability score to cost figure using the data from the U.S. Consumer Expenditure Survey.

There are two kinds of reliability scores published for automobiles in Consumer Reports, namely "predicted reliability score" and "reliability summaries." Past experiences show that predicted reliability scores themselves may not be as reliable as the name implied. See section 3 for detail. In this study we opt to use reliability summaries for our statistical model.

Reliability summaries are based on Annual Questionnaire, which Consumer Reports mail to and collect from their subscribers every year. There are two potential problems in this kind of survey: The response rate of owners dissatisfied with the reliability of their own cars could be higher than that of satisfied owners; the number of responses from owners of cars with low sales volumes might be too limited to reliably evaluate these cars, making "insufficient data" entries to appear in their reliability summaries.

The potentially "non-ignorable" non-response problem of the former—in sample survey terminology, a variable $\mathbf{Y}$ with unit nonresponse is categorized as "non-ignorably missing" if some of the $\mathbf{Y}$ are missing because of the underlying values it takes—could make unreliable cars and best selling cars overrepresented in the sample and seriously distort the analysis. In the proposed method we take care of these potential problems.

This paper is organized as follows. The methods used in estimating the statistical model and converting the model into cost figure are described in section 2. In section 3, the detailed explanation of the data is presented. In section 4, the result is presented, and in section 5 we discuss the result.

2 Methods

2.1 Estimating the Statistical Model

In this section, we summarize the method for estimating parameters in multinomial logistic regression models when the response variable $\mathbf{Y}$ is partially missing and the missing data mechanism is potentially nonignorable, and the explanatory variables are fully observed. This framework was presented by Ibrahim and Lipsitz (1996) for binomial logistic regression model. Here we extend the model to multinomial logistic regression.

The model consists of the joint distribution of the multinomial ordinal response variable $\mathbf{Y}$ and the binomial observed data indicator $\mathbf{R}$, ith of which takes 0 when the ith of the $\mathbf{Y}$ is not oberved. Since the explanatory variables $\mathbf{X}$ are fully observed, they are treated as fixed throughout. We express the joint distribution $\mathbf{R}$ and $\mathbf{Y}$ by specifying the conditional distributions $\mathbf{Y}|(\boldsymbol{\theta}, \boldsymbol{\beta})$ and $\mathbf{R}|\mathbf{Y}, \boldsymbol{\alpha}$, where $(\boldsymbol{\theta}, \boldsymbol{\beta})$ and $\boldsymbol{\alpha}$ are assumed to be distinct sets of indexing parameters for their respective distributions.

Suppose y_i, $i = 1, \ldots, n$, are independent multinomial observations with the cumulative probability ψ_{ij} up to and including jth category. Further, let $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})$ denote the $1 \times p$ observed vector of explanatory variables for the ith observation, $\mathbf{X}$ is an $n \times p$ matrix of explanatory variables, and let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ denote the corresponding $p \times 1$ column vector of regression coefficients. We use a parallel logistic regression model for the ψ_i's

$$\log\{\psi_{ij}/(1-\psi_{ij})\} = \theta_j - \boldsymbol{\beta}^T \boldsymbol{x}_i, \quad j = 1, \ldots, k-1 \tag{1}$$

with the likelihood for $y_i|\boldsymbol{x}_i$ is given by

$$\begin{aligned} & L_{y_i}(\boldsymbol{\theta}, \boldsymbol{\beta}) \\ = \; & \Pi_{j=1}^{k} (\psi_{ij} - \psi_{ij-1})^{y_{ij}} \\ = \; & \Pi_{j=1}^{k} \left\{ \frac{\exp(\theta_j - \boldsymbol{\beta}^T \boldsymbol{x}_i)}{1+\exp(\theta_j - \boldsymbol{\beta}^T \boldsymbol{x}_i)} - \frac{\exp(\theta_{j-1} - \boldsymbol{\beta}^T \boldsymbol{x}_i)}{1+\exp(\theta_{j-1} - \boldsymbol{\beta}^T \boldsymbol{x}_i)} \right\}^{y_{ij}}, \end{aligned} \tag{2}$$

where $y_{ij} = 1$ if $y_i = j$, otherwise y_{ij}=0.
The negative sign in (1) is a convention ensuring that large values of $\boldsymbol{\beta}^T \boldsymbol{x}$ lead to an increase of probability in the higher-numbered categories. Since θ_j estimates logistic transformation of the cumulative probability up to and including category j, $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_{k-1}$ must be satisfied.

The observed data indicator for the ith response y_i can be written as

$$r_i = \begin{cases} 1 & \text{if } y_i \text{ is observed,} \\ 0 & \text{if } y_i \text{ is missing,} \end{cases}$$

for $i = 1, \ldots, n$. The vector $\boldsymbol{r} = (r_1, \ldots, r_n)^T$ is $n \times 1$ column vector of observed data indicators. We specify a logistic regression model for the r_i's. Let $\boldsymbol{z}_i = (\boldsymbol{x}_i, y_i)$ and let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{p+1})^T$ be a $(p+1) \times 1$ column vector of indexing parameters for r_i. The likelihood for r_i is

$$\begin{aligned} L_{r_i|y_i}(\boldsymbol{\alpha}) &= \left(\frac{p_i}{1-p_i}\right)^{r_i} (1-p_i) \\ &= \exp\left[r_i \boldsymbol{z}_i \boldsymbol{\alpha} - \log\{1+\exp(\boldsymbol{z}_i \boldsymbol{\alpha})\}\right], \end{aligned} \tag{3}$$

where $p_i = \Pr\{r_i = 1|\boldsymbol{z}_i, \boldsymbol{\alpha}\}$ and $\log\{p_i/(1-p_i)\} = \boldsymbol{z}_i \boldsymbol{\alpha}$. If $\alpha_{p+1} \neq 0$, then the missing data mechanism depends on y_i and thus nonignorable. If $\alpha_{p+1} = 0$, then $f(r_i|\boldsymbol{z}_i, \boldsymbol{\alpha})$ does not depend on y_i, but may depend on $\boldsymbol{x}_i$.

When this happens the missing data mechanism is referred as ignorable. If $\alpha_2 = \cdots = \alpha_{p+1} = 0$, then the observed sample is effectively random subsample of the sample.

Under the assumption that $\boldsymbol{\alpha}$ and $(\boldsymbol{\theta}, \boldsymbol{\beta})$ are distinct sets of indexing parameters, the log-likelihood for all of the observations can be decomposed from (2) and (3) as

$$\begin{aligned} l(\boldsymbol{\tau}) &= \sum_{i=1}^{n} l(\boldsymbol{\tau}; \boldsymbol{x}_i, y_i, r_i) = \sum_{i=1}^{n} \{l_{y_i}(\boldsymbol{\theta}, \boldsymbol{\beta}) + l_{r_i|y_i}(\boldsymbol{\alpha})\} \\ &= \sum_{i=1}^{n} \left[y_{ij} \log \left\{ \frac{\exp(\theta_j - \boldsymbol{\beta}^T \boldsymbol{x}_i)}{1 + \exp(\theta_j - \boldsymbol{\beta}^T \boldsymbol{x}_i)} - \frac{\exp(\theta_{j-1} - \boldsymbol{\beta}^T \boldsymbol{x}_i)}{1 + \exp(\theta_{j-1} - \boldsymbol{\beta}^T \boldsymbol{x}_i)} \right\} \right. \\ &\qquad \left. + r_i \boldsymbol{z}_i \boldsymbol{\alpha} - \log\{1 + \exp(\boldsymbol{z}_i \boldsymbol{\alpha})\} \right], \end{aligned} \tag{4}$$

where $\boldsymbol{\tau} = (\theta_1, \ldots, \theta_{k-1}, \beta_1, \ldots, \beta_p, \alpha_1, \ldots, \alpha_{p+1})^T$ is a $(k + 2p) \times 1$ column vector of logistic regression parameters and $l(\boldsymbol{\tau}; \boldsymbol{x}_i, y_i, r_i)$ is the contribution to the log-likelihood from the ith observation. The model in (4) essentially treats the y_i's as missing covariates in the model for $(r_i | \boldsymbol{z}_i, \boldsymbol{\alpha})$. Thus following Ibrahim (1990) and Abe et.al (1998), the maximum likelihood estimates of $\boldsymbol{\tau}$ can be obtained via the EM algorithm by maximizing the expected log-likelihood whose ith individual contribution is

$$E[l(\boldsymbol{\tau}; \boldsymbol{x}_i, y_i, r_i)] = \begin{cases} \sum_{y_i=1}^{k} l(\boldsymbol{\tau}, \boldsymbol{x}_i, y_i, r_i) f(y_i | r_i, \boldsymbol{x}_i, \boldsymbol{\tau}) & \text{if } y_i \text{ is missing,} \\ l(\boldsymbol{\tau}; \boldsymbol{x}_i, y_i, r_i) & \text{if } y_i \text{ is observed.} \end{cases} \tag{5}$$

The E-step in (5) takes the form of a weighted log-likelihood with the conditional probabilities $f(y_i | r_i, \boldsymbol{x}_i, \boldsymbol{\tau})$ of the missing data given the observed data playing the role of weights. The M-step maximizing the function in (4), which is equivalent to calculating complete data maximum likelihood with each incomplete observation replaced by a set of weighted "filled-in" observations with weight $f(y_i | r_i, \boldsymbol{x}_i, \boldsymbol{\tau})$.

2.2 Converting the predicted reliability summaries into cost figure

We compute the maintenance and repair expenditure of these automobiles over the median age of the cars on the road, which is rounded to be eight years, according to the following steps: In step 1, we predict the expected number of troubles each of the automobiles will encounter from 1997 to through 2004 based on the statistical model; In step 2, we compute the average expenditure of maintenance and repair per trouble using the data from the U.S. Consumer Expenditure Survey; In step 3, we calculate the weight showing the vehicle–specific maintenance and repair expenditure relative to the average computed in step 2. This weight will reflect automobile-to-automobile maintenance and repair expenditure differential; In step 4, the maintenance and repair expenditure of each of the automobiles is calculated by multiplying the yearly expected number of troubles in step 1 with the average expenditure of maintenance and repair per trouble in step 2 and with the weight showing the vehicle–specific maintenance and repair expenditure in step 3 from 1997 through 2004. Detailed calculation algorithm is in Appendix A.

3 Data

Each year, Consumer Reports ask their readers to report "serious problem" on sixteen trouble spots which include engine, cooling system, fuel system, ignition, transmission, clutch, electrical system, air conditioning, suspension, brakes, exhaust, body rust, paint/trim, body integrity and body hardware on cars going back for eight years (April 1995 Consumer Reports page 285-286). A "serious problem"is one that requires a costly repair, puts the car out of commission for a time, or causes vehicle failure or a safety problem (ibid., page 287). Then the data are compiled and analyzed by Consumer Reports' statisticians.

3.1 Reliability Summaries and Predicted Reliability Scores

The "reliability summaries" are the weighted sum of the trouble rates of all trouble spots year by year, car by car (April 1996 Consumer Reports page 50, April 1997 Consumer Reports page 70), although the weights they used have not been published. "The reliability summaries show how each model compares with the overall average for that model year" and "the scores in reliability summaries are on relative scale, compared with the average for all models of the same year, from much worse than average to much better than average" in five-point scale (April 1998 Consumer Reports, page 74-75). The "predicted reliability scores", on the other hand, are the judgment based on the "reliability indexes," which Consumer Reports constructs from the three most recent years of reliability summaries (ibid., page 75).

Out of 103, 163 and 150 of 1996 models surveyed in 1997, 1998, and 1999, 46.6 %, 45.4 %, and 44.0 % of them registered reliability summaries that were different from the reliability scores predicted in 1996. We choose to use "reliability summaries" as the measurement of reliability since it reflects the actual response, not prediction, from the readers in Annual Questionnaire.

There are three reliability summaries—1997, 98 and 99—for 1996 models. If all the model deteriorated at the same rate, not only the shapes but also the locations of the three reliability summary distributions would be the same. Figure 1 shows histograms of the Reliability Summaries of 1996 models in 1997, 1998 and 1999 respectively. These histograms are drawn from the

102 models whose reliability summaries were recorded for the three years in Consumer Reports.

These histograms suggest that the three reliability summaries vary from year to year. We therefore assume that the underlying distributions represented in these histograms are likely to be different. We analyze three reliability summaries separately.

3.2 Data Handling

Data are obtained from two primary sources: all the issues of Consumer Reports from January 1992 to April 2000 and Motor Vehicle – Facts & Figures 1994 to 1997. Although there were 188 makes and models of the 1996 cars listed in the former, there were only 88 of the 188 whose sales data were reported in the U.S. Retail Sales of Passenger Cars and Top Selling Vehicles published in the latter. The sales information is vital to correct possible biases associated with the underreporting problem we discussed on page 5. Therefore, we analyze these 88 make and model cars.

3.2.1 Response variable

Reliability summary is the response variable for the multinomial logistic regression model in (1). We assign scores 5 to 1 to entries from much better than average to much worse than average. There were 29 (32.95%), 5 (5.68%) and 11 (12.50%) missing values out of 88 models in 1997, 98 and 99 respectively. We code them as -9999 and include them in the model.

The observed indicator is the response variable for the binomial logistic regression model in (3). The models (2) and (3) are simultaneously estimated.

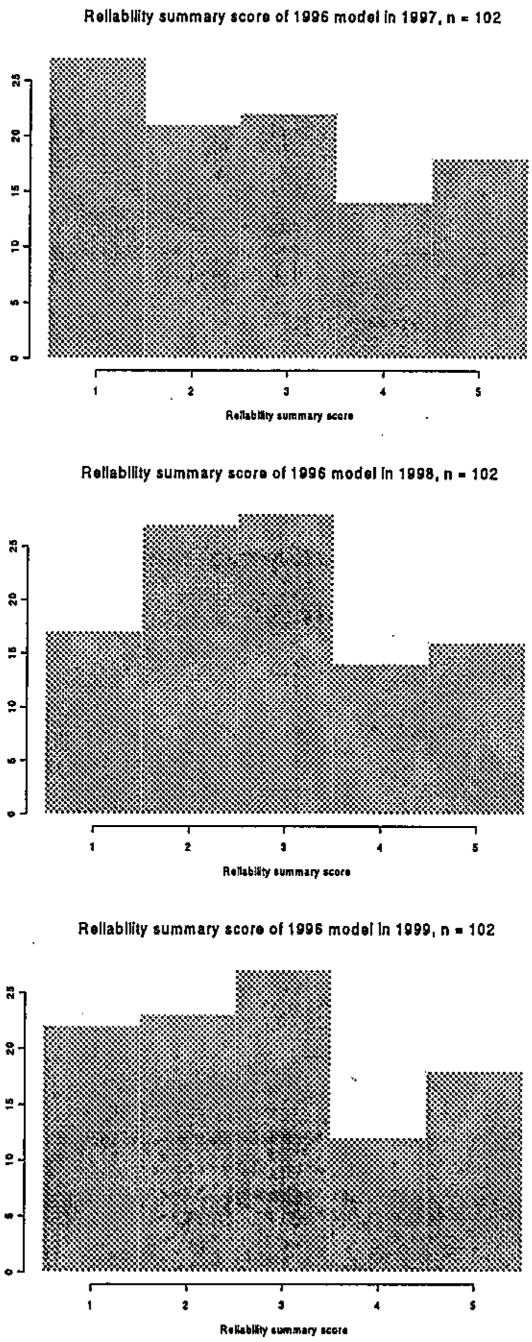

Figure 1: Histograms of the Reliability Summaries of 1996 models, $n = 102$

3.2.2 Explanatory variables

Due to the quality of the parts, the nature of the design, or the craftsmanship in producing the car, some models suffer problem at a rate far lower or higher than what one might expect from sheer aging. Explanatory variables we believe potentially related to automobile reliability in model (2) are: sticker selling price; cars' design characteristics—maximum horsepower, displacement in liters, weight in pounds and length in inches, power displacement ratio (PDR), power weight ratio (PWR); three dummy variables, which we call "origin", indicating country origin—U.S., Japan and Germany—of car manufacturers; nine "segment" category dummy variables indicating whether a vehicle is large, luxury, sport/sporty, small, coupe, pickup, SUV, minivan and medium. We use manufacturer indicator variables—"Chrysler", "Ford", "GM", "Honda", "Nissan", and "Toyota"—in place of "origin" in our preliminary analysis and found that there are no significant differences between coefficients of "Chrysler", "Ford" and "GM", just as are no differences between "Honda", "Nissan" and "Toyota." In Appendix B we explain how we proceed when these design characteristics are not available.

Since we assume sales volume and reliability summary of each model could affect missingness of its reliability summary, we use the sales volume figures and reliability score as explanatory variables in model (3). Sales volume data come from Fact & Figures published by the American Automobile Manufacturers Association(AAMA).

4 Result

4.1 Statistical estimation of 1997, 98, and 99 reliability summaries

The selected models are listed below. The process according to which the model selection is done is detailed in Appendix C. Asymptotic t-values of the coefficients appear in parentheses above the respective coefficients, AIC scores are also listed.

1997 AIC:180.553

$$\begin{aligned}\mathrm{logit}(\hat{\gamma}_j) = {} & \hat{\theta}_j + \overset{(3.13)}{0.1533} *\mathrm{PDR} + \overset{(3.60)}{0.0042} *\mathrm{weight} \\ & - \overset{(-1.96)}{9.8179} * 10^{-5} *\mathrm{price} - \overset{(-3.94)}{5.4786} *\mathrm{Japan} \\ & + \overset{(1.66)}{3.3411} *\mathrm{sport/sporty} + \overset{(3.60)}{4.1494} *\mathrm{small} \\ & + \overset{(2.90)}{4.0119} *\mathrm{coupe} + \overset{(1.94)}{2.9033} *\mathrm{minivan}\end{aligned} \tag{6}$$

$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
$\overset{(-3.99)}{-24.6233}$	$\overset{(-3.74)}{-22.5170}$	$\overset{(-3.41)}{-19.8861}$	$\overset{(-2.90)}{-16.94}$

$$\mathrm{logit}(\hat{p}) = \overset{(-1.05)}{-0.8164} + \overset{(3.65)}{2.9474} * 10^{-5} *\mathrm{sales\ volume} - \overset{(-1.00)}{0.2449}\, Y \tag{7}$$

1998 AIC:217.543

$$\begin{aligned}\mathrm{logit}(\hat{\gamma}_j) = {} & \hat{\theta}_j + \overset{(2.17)}{0.0719} *\mathrm{PDR} \\ & + \overset{(3.52)}{0.002366} *\mathrm{weight} - \overset{(-3.90)}{19.2419} * 10^{-5} *\mathrm{price} \\ & - \overset{(-4.06)}{3.0533} *\mathrm{Japan} + \overset{(1.87)}{3.1185} *\mathrm{luxury} \\ & + \overset{(3.24)}{2.7932} *\mathrm{sport/sporty} + \overset{(4.00)}{4.5260} *\mathrm{coupe} \\ & + \overset{(1.81)}{2.2520} *\mathrm{SUV}\end{aligned} \tag{8}$$

$\hat\theta_1$	$\hat\theta_2$	$\hat\theta_3$	$\hat\theta_4$
$\overset{(-3.40)}{-10.6874}$	$\overset{(-2.80)}{-8.6128}$	$\overset{(-2.06)}{-6.2020}$	$\overset{(-1.34)}{-4.0180}$

$$\text{logit}(\hat p) = \overset{(0.58)}{+1.6869} \overset{(2.00)}{+6.5860 * 10^{-5}} * \text{sales volume} \overset{(-0.64)}{-0.5010}\, Y \tag{9}$$

1999 AIC: 235.03

$$\begin{aligned}\text{logit}(\hat\gamma_j) = {} & \hat\theta_j \overset{(2.86)}{+0.1039} * \text{PDR} \\ & \overset{(2.52)}{+0.001766} * \text{weight} \overset{(-4.04)}{-24.0305 * 10^{-5}} * \text{price} \\ & \overset{(-4.20)}{-3.4679} * \text{Japan} \overset{(2.26)}{+1.5856} * \text{large} \\ & \overset{(2.99)}{+5.6990} * \text{luxury} \overset{(2.32)}{+2.087} * \text{sport/sporty} \\ & \overset{(1.87)}{+1.7414} * \text{coupe} \overset{(2.92)}{+4.1336} * \text{SUV} \\ & \overset{(1.49)}{+2.0883} * \text{minivan}\end{aligned} \tag{10}$$

$\hat\theta_1$	$\hat\theta_2$	$\hat\theta_3$	$\hat\theta_4$
$\overset{(-2.63)}{-8.6408}$	$\overset{(-2.25)}{-7.3076}$	$\overset{(-1.59)}{-5.0701}$	$\overset{(-0.98)}{-3.1299}$

$$\text{logit}(\hat p) = \overset{(1.46)}{+2.0286} \overset{(2.47)}{+2.7013 * 10^{-5}} * \text{sales volume} \overset{(-1.51)}{-0.5969}\, Y \tag{11}$$

Throughout we find the following variables to be significant: "PDR" and "weight" as design characteristics, "Japan" as manufacturer origin, "sport/sporty" and "coupe" as car type, and "price." Decreases in PDR or weight as well as increases in price will increase the probability that the reliability summary falls into upper category. Switching manufacturer origin from "the U.S." or "Germany" to "Japan" will also increase the reliability. "Sport/sporty" cars and "coupes" are less reliable.

In our analysis, price has always been negative and significant. Furthermore, the model fitting will be noticeably worsened if the price variable is not included. So we consider that price of an automobile must have reflected its reliability or "perception"of reliability. We surmise of two explanations. First, price is simply the representative of the quality of parts used in a car: The higher the price, the better the quality of parts. Second, it is possible that perception of quality is affected by the price car owners paid for the automobile: If a subscriber of Consumer Reports had purchased an expensive car, s/he might have psychologically been predisposed to think that higher priced cars would be more reliable and accordingly might have minimized troubles when they responded to the Annual Questionnaire. The point to notice, however, is that the price is significantly and positively correlated with reliability when other variables; "PDR", "weight", "Japan", and car types are controlled. That is, among the cars with same values of PDR, weight, manufacturer origin, and the car types, the higher priced cars tend to be more reliable than the lower priced cars.

In the 1997 model, "small" cars are significantly less reliable, though they are not in 1998 and 1999. This is probably due to the same reason why "car type"—small is correlated with decreased reliability in initial analysis detailed in Appendix C.

In 1998 and 1999, "luxury" cars appear to be significantly correlated with low reliability summaries. As noted above, higher prices of automobiles are significantly correlated with elevated reliability. That "luxury" appears to be correlated with lower reliability in the presence of "price" probably points to the fact that purchasing luxury cars is not necessarily justified in terms of reliability, though it may be so in terms of comfort.

Normally, PDR may be considered as an engine variable and expresses its performance. However, it is difficult to explain the effect of PDR on reliability of automobile in terms of engine because most of cars in 1996 have "better than average" or "much better than average" engine reliabilities in April issues of Consumer Reports from 1997 to 2000. Therefore it will be better to regard PDR as a surrogate variable for the way that automobiles are designed and engineered. Higher PDR implies greater mechanical stress on moving components such as clutch, transmission, brakes, suspension, tires, and body.

The missingness in reliability summaries seems to be highly correlated with sales volume but much less so with reliability summaries, although in 1997 and 1999 their asymptotic t-values exceed unity. Thus we know that a vehicle with higher sales volume is less likely to have its reliability summary missing. We can not totally exclude the possibility that an automobile whose reliability summary is worse or much worse than average registers higher response rate to Annual Questionnaire than those with better or much better than average.

4.2 Maintenance and repair expenditure computation

The August 2000 issue of Consumer Reports gives us an idea on how to estimate the repair costs for eighteen different 1992 model cars, minivans, and SUVs. Their idea is that we can estimate the expected cost of repairs from the cost information of eight trouble spots and their trouble rates calculated from the reliability. The trouble spots are: air conditioner, electrical system, engine, fuel system, ignition system, suspension, and transmission. Labor and parts estimated compiled from Mitchel Estimating Guide, an industry manual, except for transmissions, where they used estimates from Portland Transmission in Connecticut, and for radiators, from Modine, a manufacturer. The total cost for each of the trouble spots is calculated by summing up the costs of its subsystems.

If we simply use the figures in their table, however, we are likely to overestimate the actual expenses on repairs. This is because their cost figures assume a complete breakdown for each trouble spot, whereas in reality repairs are likely to be done only on the broken subsystem that needs repair. For example their yearly repair costs of Ford Taurus for third, fourth, and fifth year would be $277, $565, and $684 respectively in our estimation; while according to the U.S. Consumer Expenditure Survey the actual annual expenditure is $681 in 1997 for average household with 1.98 vehicles, which also include the expenditure for maintenance such as oil change, battery and tires. Therefore we base our expenditure calculation on the U.S. Consumer Expenditure Survey's information.

We estimate the lifetime maintenance and repair expenditure of thirteen model cars, minivan, and SUVs out of the eighteen models in the August 2000 issue of Consumer Reports. These thirteen models are selected because they are included for estimation of the statistical models in section 4.1. These are: Ford Escort, Honda Civic, and Toyota Corolla, as representative small cars; Ford Taurus, Honda Accord, and Toyota Camry as medium cars; Buick Le Sabre, Ford Crown Victoria, Cadillac DeVille, and Lincoln Town Car as large cars; Ford Explorer, and Chevrolet Blazer as SUVs; Dodge Caravan as minivan. The selected models were, respectively the top, the second and the fifth highest selling models for small cars, the three highest selling models for medium cars, the four highest selling models for large cars, the top and the third highest selling models for SUVs; the top selling model for minivan respectively in 1996.

Although Cadillac DeVille, and Lincoln Town Car are categorized as luxury cars in April 2000 issue of Consumer Reports, they are classified as large cars in April 1996 issue. We use the 1996 classification. On the calculation we assume the repair cost required in each trouble spot does not change between 1992 and 1996 car models. The properties of selected models—country origins, car types, PDRs in horsepower/liter (hp/l), weights in lb. and prices in U.S. dollars—are shown in Table 1.

Automobile Model	Country	Car type	PDR(hp/l)	Weight(lb.)	Price($)
Ford Escort	U.S.	small	46.32	2565	14030
Honda Civic	Japan	small	66.25	2440	16740
Toyota Corolla	Japan	small	58.33	2540	18973
Ford Taurus	U.S.	medium	66.67	3516	23065
Honda Accord	Japan	medium	62.96	3255	25480
Toyota Camry	Japan	medium	44.33	3230	23818
Buick Le Sabre	U.S.	large	53.95	3450	23888
Ford Crown Victoria	U.S.	large	45.65	4010	25283
Cadillac DeVille	U.S.	large	59.78	4020	39184
Lincoln Town Car	U.S.	large	45.65	4055	39435
Ford Explorer	U.S.	SUV	40.00	4440	36028
Chev. Blazer	U.S.	SUV	44.19	4225	29054
Dodge Caravan	U.S.	Minivan	47.88	3985	25940

Table 1: The properties of selected cars

We convert the predicted reliability summaries into cost figure according to the method described in section 2.2.

In step 1, we compute the expected number of troubles of each thirteen models from two sources. One is the estimated probabilities that the automobile's reliability belongs to each of the five categories—much better than average to much worse than the average. The estimated probabilities shown in Table 2, from 1997 to 1999 are calculated using the statistical model in section 4.1, the figures from 2000 to 2004 are the average of the previous three years probabilities. The other is the trouble rates one would expect from her/his automobile if it belongs to one of the aforementioned five categories. The trouble rates are published in Consumer Reports every year. Since the trouble rate figures vary from year to year, we use the average of 1997 through 2000 and these numbers are shown in Table 3. The expected number of troubles of each thirteen models is shown in Table 4.

	Year	5	4	3	2	1		Year	5	4	3	2	1
Ford	1997	0.0240	0.2948	0.5478	0.1150	0.0184	Honda	1997	0.3796	0.5413	0.0729	0.0054	0.0008
Escort	1998	0.0641	0.3141	0.4932	0.1104	0.0182	Civic	1998	0.4393	0.4351	0.1112	0.0112	0.0016
	1999	0.0552	0.2338	0.5030	0.1432	0.0648		1999	0.3608	0.4363	0.1764	0.0193	0.0071
	2000-4	0.0475	0.2809	0.5147	0.1229	0.0338		2000-4	0.3932	0.4709	0.1208	0.0119	0.0032
Toyota	1997	0.6275	0.3423	0.0280	0.0020	0.0003	Ford	1997	0.0031	0.0523	0.3934	0.4212	0.1300
Corolla	1998	0.6267	0.3104	0.0569	0.0052	0.0008	Taurus	1998	0.0094	0.0684	0.4069	0.3975	0.1178
	1999	0.6481	0.2795	0.0641	0.0061	0.0022		1999	0.0114	0.0629	0.3548	0.3112	0.2597
	2000-4	0.6341	0.3107	0.0497	0.0044	0.0011		2000-4	0.0080	0.0611	0.3850	0.3766	0.1692
Honda	1997	0.8316	0.1579	0.0098	0.0007	0.0001	Toyota	1997	0.9878	0.0115	0.0006	0.0000	0.0000
Accord	1998	0.4368	0.4364	0.1139	0.0112	0.0016	Camry	1998	0.6952	0.2578	0.0426	0.0039	0.0006
	1999	0.6060	0.3086	0.0756	0.0072	0.0026		1999	0.8819	0.0992	0.0168	0.0015	0.0005
	2000-4	0.6248	0.3086	0.0664	0.0064	0.0014		2000-4	0.8550	0.1228	0.0200	0.0018	0.0004
Buick	1997	0.0310	0.3478	0.5156	0.0915	0.0142	Ford	1997	0.0119	0.1747	0.5745	0.2021	0.0368
Le Sabre	1998	0.0336	0.2022	0.5389	0.1901	0.0353	Crown	1998	0.0201	0.1340	0.5158	0.2718	0.0583
	1999	0.0130	0.0710	0.3781	0.3031	0.2348	Victoria	1999	0.0147	0.0794	0.3991	0.2937	0.2132
	2000-4	0.0259	0.2070	0.4775	0.1949	0.0947		2000-4	0.0156	0.1293	0.4965	0.2559	0.1028
Cadillac	1997	0.0052	0.0847	0.4884	0.3403	0.0815	Lincoln	1997	0.0385	0.3940	0.4812	0.0750	0.0113
DeVille	1998	0.0952	0.3878	0.4293	0.0758	0.0119	Town Car	1998	0.2192	0.4945	0.2515	0.0302	0.0045
	1999	0.0870	0.3118	0.4626	0.0979	0.0407		1999	0.2924	0.4496	0.2222	0.0261	0.0097
	2000-4	0.0625	0.2615	0.4601	0.1713	0.0447		2000-4	0.1834	0.4460	0.3183	0.0438	0.0085
Ford	1997	0.0133	0.1914	0.5767	0.1856	0.0329	Chev.	1997	0.0088	0.1355	0.5564	0.2499	0.0494
Explorer	1998	0.0092	0.0668	0.4021	0.4013	0.1206	Blazer	1998	0.0030	0.0228	0.2018	0.4735	0.2989
	1999	0.0128	0.0701	0.3758	0.3040	0.2373		1999	0.0023	0.0135	0.1147	0.2322	0.6372
	2000-4	0.0118	0.1094	0.4515	0.2970	0.1303		2000-4	0.0047	0.0573	0.2910	0.3186	0.3285
Dodge	1997	0.0006	0.0099	0.1173	0.4185	0.4538							
Caravan	1998	0.0206	0.1368	0.5181	0.2676	0.0569							
	1999	0.0087	0.0488	0.3062	0.3207	0.3156							
	2000-4	0.0099	0.0652	0.3139	0.3356	0.2754							

Table 2: Probability distribution of reliability in each category

In step 2, the average annual expenditure of maintenance and repair per trouble is calculated to be \$352.17 in 1997. We arrive at the number using maintenance and repairs expenses per vehicle in 1997 from the U.S. Consumer

year(s) old i	Much Better than average (j=5)	Better than average (j=4)	Average (j=3)	Worse than average (j=2)	Much Worse than average (j=1)
1	0.126	0.192	0.258	0.311	0.364
2	0.224	0.345	0.467	0.571	0.675
3	0.310	0.494	0.678	0.812	0.947
4	0.313	0.582	0.851	1.052	1.254
5	0.368	0.666	0.965	1.228	1.491
6	0.511	0.838	1.165	1.476	1.787
7	0.719	1.067	1.416	1.801	2.186
8	0.953	1.306	1.660	1.928	2.197

Table 3: The number the trouble rate of i year old model whose reliability falls into j category

Expenditure Survey and the number of vehicles 1989 to 1996 year model on the road from the AAMA motor vehicle facts and figures 1997, and the total number of troubles expected from the trouble rate figure.

In step 3, the weight showing the vehicle–specific maintenance and repair expenditure relative to the average is calculated using the information of frequency-of-repair charts of 1989 to 1996 year model, the information of suggested repair costs of 1992 car models, and sales volume information. They are respectively taken from April 1997 issue of Consumer Reports, August 2000 issue of Consumer Reports, and the AAMA Facts & Figures 1997. First, we calculate suggested repair cost of each of the thirteen models from 1989 to 1996. Second, we divide these suggested repair cost by the number of troubles for each of the models. This obtains suggested repair cost per trouble for each of the models. Third, we obtain average suggested repair cost per trouble. This cost is calculated by the total suggested repair cost and the total number of troubles for the thirteen models sold between 1989 to 1996. Here we assume the thirteen models are adequate representatives[1] of all the automobiles in the sense that the joint distribution of suggested repair

[1]The thirteen models are very popular models and cover about 32% of all the automobiles on the road in 1997.

	expected number of troubles							
Automobile Model	1997	1998	1999	2000	2001	2002	2003	2004
Ford Escort	0.24	0.43	0.65	0.79	0.91	1.10	1.36	1.60
Honda Civic	0.17	0.31	0.47	0.52	0.60	0.76	0.99	1.23
Toyota Corolla	0.15	0.28	0.39	0.43	0.50	0.65	0.87	1.11
Ford Taurus	0.29	0.52	0.77	0.98	1.13	1.37	1.67	1.84
Honda Accord	0.14	0.31	0.40	0.44	0.51	0.66	0.88	1.12
Toyota Camry	0.13	0.27	0.34	0.36	0.42	0.57	0.78	1.01
Buick Le Sabre	0.24	0.48	0.80	0.90	1.03	1.25	1.54	1.73
Ford Crown Victria	0.26	0.49	0.75	0.90	1.03	1.25	1.54	1.73
Cadillac DeVille	0.28	0.41	0.61	0.80	0.92	1.13	1.39	1.61
Lincoln Town Car	0.23	0.36	0.50	0.65	0.74	0.93	1.16	1.41
Ford Explorer	0.26	0.52	0.77	0.93	1.08	1.30	1.59	1.78
Chev. Blazer	0.27	0.57	0.88	1.03	1.21	1.45	1.77	1.91
Dodge Caravan	0.33	0.48	0.79	1.01	1.18	1.42	1.73	1.88

Table 4: Expected number of troubles of each the thirteen models

costs and sales volume for the remaining automobiles are the same as that of thirteen models. Finally, the weight of the vehicle–specific maintenance and repair expenditure is the ratio of suggested repair cost per trouble of each of the thirteen models to the average suggested repair cost per trouble. These weights are shown in Table 5. We repeat this process (steps 1 to 4) for eight years for the thirteen automobiles. See Table 6 for final result.

We summarize the result. In Table 5, we notice the differences in weights reflect those of repair costs between the thirteen models when they encounter one trouble. The weights of Honda Civic and Toyota Collora are higher than those of Ford Escort and so are the weights of Honda Accord and Toyota Collora relative to Ford Taurus. This is because the cost of parts and labor needed in repairs for these Japanese models are more expensive.

Furthermore, the weights of the Honda tend to be higher than the Toyota. This is because parts and labor needed to repair Honda are in general more expensive than those of Toyota. Consequently, the maintenance and repair expenditures of Honda Civic and Accord are higher than Toyota Collora and Camry respectively in Table 6 although the trouble rates of Honda are very close to those of Toyota as seen in Table 4. The weights of six, seven, and eight year old Honda Civic are noticeably higher than these of previous years. This is due to Civic's sagging reliability for engine from 5 to 4 or 3 in these years, which is very expensive to repair or replace. The weights of Lincoln Town Car are so high at first, but relatively slow increase in their repair cost to the average models results in lower weights in later years. The reliability summaries of Cadillac DeVille fluctuate and so do their weights. In Table 6, we observe that older models tend to cost more to maintain and repair. As expected these small cars are least expensive to maintain and repair and the medium cars follow the suit. Toyota Collora are especially economical to run. Ford Taurus is not as expensive as expected from their trouble rates in Table 4 because cost of repair is lower than those of the Honda Accord or Toyota Camry. Honda Accord is found to be less expensive than Honda Civic in some years due to their lower trouble rate in Table 4. Although the large cars and SUVs are expensive to maintain and repair than small or medium cars, by far the most expensive in terms of maintenance and repair cost in the thirteen models are Dodge Caravan.

4.3 Simplified LCC calculation

In theory automobile LCC must include acquisition cost, fuel cost, maintenance and repair cost, insurance cost and disposal cost. Insurance premium

Automobile Model	Weight							
	1997	1998	1999	2000	2001	2002	2003	2004
Ford Escort	0.84	0.78	0.89	0.78	0.74	0.79	0.79	0.85
Honda Civic	1.09	1.09	1.08	1.05	1.05	1.21	1.21	1.21
Toyota Collora	1.03	1.03	0.88	0.79	1.00	0.87	0.88	1.10
Ford Taurus	0.92	0.78	0.76	0.76	0.76	0.84	0.85	0.94
Honda Accord	1.53	1.32	1.33	1.16	1.05	1.44	1.28	1.31
Toyota Camry	1.21	1.17	1.17	0.99	0.96	1.15	1.17	1.43
Buick Le Sabre	1.00	0.98	1.11	0.92	0.86	0.77	0.79	0.86
Ford Crown Victoria	0.88	0.78	0.89	0.84	0.82	0.86	0.94	0.89
Cadillac DeVille	1.62	1.14	1.29	1.49	1.32	1.44	1.15	1.20
Lincoln Town Car	1.55	1.40	1.45	1.25	0.93	0.82	0.82	0.95
Ford Explorer	0.88	0.76	0.77	0.86	1.01	0.98	0.95	0.95
Chev. Blazer	1.27	1.10	0.88	0.82	0.75	0.83	0.91	0.89
Dodge Caravan	1.10	0.93	1.23	1.28	1.15	1.35	1.54	1.42

Table 5: The weight showing the vehicle–specific maintenance and repair expenditure relative to the average maintenance and repair expenditure

depends on the driver's profile and can be obtained from the insurance company where the car owner wants to insure his/her car. Disposal cost may be substituted by trade in values and can be obtained from used car dealers. In this study, we calculate the automobile LCC without insurance and disposal cost.

For acquisition cost, we use the prices Consumer Reports paid for their tested car and calculate it to 1996 price as mentioned in Appendix B. For fuel cost, we assume that each automobile travels 12,000 miles per year and use automobile fuel–consumption–rate priced at 1.23 dollars per gallon, which was the national average price in 1997. The resulting LCC without insurance and disposal cost is listed in Table 7.

Automobile Model	1997	1998	1999	2000	2001	2002	2003	2004	Total
Ford Escort	77.12	128.18	219.07	232.89	253.80	330.55	410.16	514.50	2166.28
Honda Civic	71.03	127.54	192.98	206.64	237.92	349.15	454.44	566.27	2205.98
Toyota Corolla	59.31	108.55	129.75	128.35	187.54	214.40	289.02	462.86	1579.78
Ford Taurus	100.64	154.04	222.15	280.25	325.49	434.03	536.82	653.71	2707.13
Honda Accord	79.83	154.87	201.94	191.43	200.61	362.10	427.97	555.70	2174.46
Toyota Camry	58.29	118.70	149.05	134.27	152.89	246.58	345.10	550.65	1755.53
Buick Le Sabre	89.27	172.04	321.86	299.05	324.85	349.34	444.82	547.75	2548.99
Ford Crown Victoria	86.29	144.88	255.70	288.52	322.35	411.58	550.72	591.34	2651.39
Cadillac DeVille	169.86	176.60	299.90	452.00	459.51	611.45	606.13	729.97	3505.41
Lincoln Town Car	135.53	190.21	272.35	306.43	261.17	287.12	358.87	504.47	2316.17
Ford Explorer	84.72	150.49	222.83	303.13	409.66	479.84	573.50	642.55	2866.72
Chev. Blazer	126.97	240.17	294.25	319.44	343.34	456.70	612.55	649.77	3043.19
Dodge Caravan	137.00	171.04	371.84	487.78	513.01	727.19	1010.14	1009.70	4427.71

Table 6: An automobile specific expected maintenance and repair expenditure

Three medium cars—Ford Taurus LX, Honda Accord LX V6, Toyota Camry LE V6—and Buick Le Sabre Custom as large car were tested in January 1996 issue of Consumer Reports because they were comparable in acquisition cost. Based on acquisition cost alone, Ford Taurus LX seems to be reasonably priced among the four automobiles. However, after LCC is taken into consideration, it becomes about $700 more expensive than Honda Accord LX V6 due to the latter's lower maintenance and repair cost. Also it is almost equally expensive relative to Toyota Camry LE V6 because the Camry consumes less fuel and is much more reliable.

We find in Table 7 that the amount of maintenance and repair costs appear to be smaller than fuel costs. However our maintenance and repair cost calculation does not include the opportunity cost that is incurred when these automobiles need maintenance and/or repair and thus out of service.

Automobile Model	reported date	Acquision Cost	Fuel Cost	M&R Cost	Total Cost
Ford Escort LX	93,9	14030	4373	2166	20569
Honda Civic LX	96,3	16740	3809	2206	22755
Toyota Corolla LE	93,8	18973	3936	1580	24489
Ford Taurus LX	96,1	23065	5623	2707	31395
Honda Accord LX V6	96,1	22880	5623	2174	30677
Toyota Camry LE V6	96,1	23818	5134	1756	30708
Buick Le Sabre Custom	96,1	23888	5904	2549	30756
Ford Crown Victoria LX	94,3	25283	6215	2651	34149
Cadillac DeVille Base	98,6	39184	5904	3505	48593
Lincoln Town Car	NA	39435	6215	2316	47966
Ford Explorer Limited 4WD	95,8	36028	6946	2867	45841
Chev. Blazer LT 4WD	95,8	29054	7872	3043	39969
Dodge Caravan LE V6	96,7	25940	6215	4428	36583

Table 7: An automobile specific LCC without insurance cost and disposal cost

5 Discussion

The result in section 4 tells us how much one should realistically expect to pay to maintain and repair the popular thirteen vehicles purchased in 1996 in their median life time of eight years. We find that on average 1) one is likely to pay $1984 for the three small cars in those years; 2) one is expected to pay $2212 or $228 more for the three medium cars than for those small cars; 3) one is to pay $2755 or $771 more for the four the large cars than for those small cars; 4) one pays $2955 or $971 more for the SUVs than for those small cars; 5) for the only minivan—Dodge Caravan, one pays $4428 or $2444 more. This calculation shows that some models could be regarded as unacceptably expensive to maintain and repair, even though they offer, for example, large interiors or four–wheel drive capability for traction and stability.

It has been said that vehicles manufactured by the Japanese Car Companies are more reliable. However we find that this fact does not necessarily translate into their lower maintenance and repair expenditure because higher parts and labor costs of those vehicles could partially offset its reliability differential with vehicles made by the Big Three. For example, Ford Escort have been less reliable than Honda Civic, but it turns out that the former are less expensive to maintain and repair than the latter, albeit by small margin. In any event two models from Toyota are likely to be very economical to run according to their extremely lower trouble rates and their reasonable expenditures for repairs.

In this paper we propose the statistical model which is new to management sciences literature in the topic of life time repair cost estimation. Nonetheless, we believe our model is a simple and powerful approach that enables researchers to replicate and generalize the results across durable products.

References

[1] Abe, Y., Y. Kanazawa, I. Kabashima. 1998. Modelling Party Support of Japanese Voters in the Early 1996. *Behaviormetrika,* Vol.25, No.1, 25-44.

[2] Consumer Reports. January 1992 to August 2000.

[3] Fricker, D.V. 1979. Life Cycle Cost: The Key Roles of Reliability and Maintainability. *Engineerig,* Vol.219, No.2. 164-166.

[4] Hutton, R. and L. Wilkie. 1980. Life-Cycle Cost: Consumer Responses to New Information. *Journal of Consumer Research,* Vol.6, March. 349-360.

[5] Ibrahim, J.G. 1990. Incomplete Data in Generalized Linear Models. *Journal of the American Statistical Association.* No.85. 765-769.

[6] Ibrahim, J.G., S.R. 1996. Lipsitz. Parameter Estimation from Incomplete Data in Binomial Regression When the Missing Data Mechanism is Nonignorable. *Biometrics,* No.52. 1071-1078.

[7] McNeill D.L., W.L. Wilkie. 1979. Public Policy and Consumer Information: Impact of the New Energy Labels. *Journal of Consumer Research,* Vol.6, June. 1-11.

[8] Motor vehicle Facts & Figures 1994 to 1997. American Automobile Manufacturers Association.

[9] Neumann B. J.K., Skwizynski. 1983. *Electronic Systems Effectiveness and Life Cycle Costing, NATO ASI Series, Series F: Computer and Systems Sciences No.3.* Springer-Verlag.

[10] Taylor W.B. 1981. The use of life cycle costing in acquiring physical assets. *Long Range Planning.* 14(6) 32-43.

A Detail of the calculation of maintenance and repair expenditure

In section 4.2, we compute the maintenance and repair expenditure for the selected thirteen 1996 models. Calculation for automobile k in ith year, where $k = 1, 2, \ldots, 13$, one of the thirteen selected automobile models, i is the age of automobile from 1 year old to 8 years old, starting from 1997 to 2004, is according to the following steps:

In step 1, we predict the expected number of troubles by

$$\text{(expected trouble rate)}_{ik} = \sum_{j=1}^{5} r_{ij} \times \hat{p}_{ijk}, \tag{12}$$

where r_{ij} is the trouble rate of an automobile whose reliability belongs to jth category—$j = 1$ is much worse than average, $j = 2$ is worse than average, $j = 3$ is average, $j = 4$ is better than average, $j = 5$ is much better than average—in i year old, published in Consumer Reports every year. Table 3 shows the average value over 1996 to 2000, we utilize. $\hat{p}_{ijk}$ is the estimated probability in section 4.1, which automobile k's reliability summary belongs to category j in i th year.

In step 2, we compute the average maintenance and repair expenditure per trouble in 1997 by

$$\text{average expenditure per trouble} = \frac{c \times \sum_{m=1}^{8} v_m}{\sum_{m=1}^{8} v_m r_{m3}}, \tag{13}$$

where c represents the average maintenance and repair expenditure per one automobile in 1997 and v_m is the number of automobiles on the road of $m = 1, 2, \ldots, 8$ years old model in 1997. r_{m3} is the trouble rate of m year old average reliability ($j = 3$) model per year. The numerator of the right hand side of (13) equals to the total maintenance and repair expenditure in

1997 and the denominator is total number of troubles in 1997. We substitute the number of automobiles purchased in each of the eight years backwards from 1996 to 1989 for v_m as seen in table 8. This substitution implies that we assume the vehicles purchased between the eight years are still left on the road in 1997.

age m	year purchased	v_m (Units in Thousands)
1	1996	10098
2	1995	14246
3	1994	13395
4	1993	11879
5	1992	12078
6	1991	12184
7	1990	13577
8	1989	13669

Table 8: The number v_m of automobiles on the road in 1997 classified by aging from one year old model up to eight years old model

In step 3, the weight of the automobile–specific maintenance and repair cost relative to the average maintenance and repair cost, w_{mk} is calculated as follows: First, we calculate suggested repair cost by

$$(\text{suggested repair cost})_{mk} = \sum_{l=1}^{8} \text{TR}_{mkl} \times C_{kl}, \tag{14}$$

where TR_{mkl} is the trouble rate of the trouble spot l $(= 1, 2, \ldots, 8$, there are eight primal trouble spots) and C_{kl} is the repair cost needed in spot l for its complete breakdown. We apply labor and part cost of 1992 automobile model published in August 2000 issue of Consumer Reports into C_{kl}.

Second, we calculate the suggested repair cost per trouble by dividing suggested repair cost with the total trouble rate over the eight troubles spots:

$$\text{(suggested repair cost per trouble)}_{mk} = \frac{\text{(suggested repair cost)}_{mk}}{\sum_{l=1}^{8} \text{TR}_{mkl}}. \quad (15)$$

Third, we obtain the average suggested repair cost per trouble over the thirteen automobiles by dividing total suggested repair cost of the thirteen automobiles up to eight years with the total number of troubles of those as

$$\text{average suggested repair cost per trouble} = \frac{\text{TSRC}}{\text{TNT}}, \quad (16)$$

where total suggested repair cost of the thirteen automobiles up to eight years, TSRC, is calculated by

$$\text{TSRC} = \sum_{k=1}^{13} \sum_{m=1}^{8} \text{(suggested repair cost)}_{mk} \times \text{(sales volume)}_{mk}. \quad (17)$$

Furthermore, the total number of troubles of the thirteen automobiles up to eight years, TNT, is computed by

$$\text{TNT} = \sum_{k=1}^{13} \sum_{m=1}^{8} \left\{ \left(\sum_{l=1}^{8} \text{TR}_{mkl} \right) \times \text{(sales volume)}_{mk} \right\}. \quad (18)$$

Finally, we obtain the weight of the automobile–specific maintenance and repair cost relative to the average maintenance and repair cost, w_{mk} by

$$w_{mk} = \frac{\text{(suggested repair cost per trouble)}_{mk}}{\text{average suggested repair cost per trouble}}. \quad (19)$$

We assume the weight of the past automobiles equal to the weight of automobiles in the future, that is $w_{ik} = w_{mk}$.

In step 4, we obtain the maintenance and repair cost by

$$\begin{aligned} &\text{(maintenance and repair cost)}_{ik} \\ &= \text{(expected trouble rate)}_{ik} \\ &\quad \times \text{average expenditure per trouble} \times w_{ik}. \end{aligned} \quad (20)$$

B Explanation when automobiles design characteristics are not available

We try to use design characteristics of the models tested by Consumer Reports. When we can not obtain them because some models have not been tested, we impute them as follows:

- In many cases, Consumer Reports' staff have tested only one model from similarly designed cars like Chrysler Sebring and Dodge Avenger twin, and they assign the same tested values to these models. In this case, we use the values assigned by Consumer Reports.
- Some models such as BMW Z3 were not tested, but their upgraded models were tested later; the other models such as Cadillac Fleetwood were never tested, although the number of these model is few. We choose to use design characteristics of their 1996 model published in "Profile of the 1996 cars" in April 1996 issue of Consumer Report because using these characteristics from years other than 1996 would have confounded the analysis unnecessarily.

In some models, there are a lot of trim lines; and their sticker selling prices vary in accordance with their specifications. We use the prices of tested trim line. Nevertheless, many of models were tested before or after 1996 and we have to predict the sticker selling price in 1996.

To do so, we proceed as follows: The fifth best selling car in 1996 was Ford Escort, but its LX trim line was tested in September 1993 for example. At that time the car was purchased at $12,539. The model's price range was $8,355 – $11,933 according to September 1993 issue of Consumer Reports. In 1996 its reported price range was $10,065 – $ 13,205. So we calculate that 1996 Ford Escort LX was sold at around $14,030; the midrange price differential of $1,491 added to the 1993 selling price.

Notice that we do not have to invoke this formula for the four best selling cars in 1996—Ford Taurus, Honda Accord, Toyota Camry, and Honda Civic—because they were tested in 1996 and their purchasing prices were available. In general we find that the more popular cars tended to have 1996 purchasing prices available because they were frequently tested.

C Covariate selection process

First we investigate the relationship between reliability summaries and "car type" variables—small cars, sports/sporty cars, medium cars, coupes, large cars, luxury cars, minivans, SUVs, pickup trucks. This is to obtain an overall picture as to which car types tend to correlate with better/worse reliability. Using medium car as baseline, we find that small car and coupe categories in 1997 and coupe and sport car categories in 1998 and 1999 tend to be associated with decreased reliability at 95% significant level. Pickup trucks, SUVs, and minivans tend to show somewhat decreased reliability, although their significances are less pronounced with the corresponding asymptotic t-values just above unity.

Note that small cars in 1997 are associated with decreased reliability, which contradicts to general understanding that small cars, sharing many components with their larger siblings, tend to be more reliable than medium cars. In our data Volkswagen Golf and Jetta happen to have registered unusual "much worse than average" in 1997, though they recover in 1998 and 1999 to "average."

To find the optimal statistical model, we run multinomial logistic regression analysis of reliability summaries on the explanatory variables listed in 3.3.2, while binary logistic regression of observed indicator variable begins with sales volume and reliability summaries as covariates.

We eliminate a variable one at a time when its asymptotic t-value is least significant. We repeat this elimination process until AIC value starts to rebound and increase. For analysis of observed indicator, we intentionally leave insignificant covariate—reliability summaries—whose asymptotic t-value is less than unity in absolute terms in year 1998, though that covariate appears more significant in years 1997 and 1999. This is because the fact that this value is not so significant tells us the nature of missingness in our data.