ON ESTIMATING THE RATE OF ACTUAL PURCHASERS BY THE PURCHASE INTENTION SURVEY

— In the Case of Washing Machines and Color TV Sets —

YOSHIAKI SATŌ
Assistant
Faculty of Economics
and
Business Administration
Hokkaido University

1. INTRODUCTION

The purchase intention survey is a conventional marketing research technique especially for a new product development. Though it does not provide the exact rate of actual purchasers,(1) [2] proposes a model, in which a beta prior distribution plays an important part in estimation of a purchase probability of those who have a stated purchase intention.

This paper presents a model to estimate the rate of actual purchasers out of those who have intentions to buy a commodity on the assumption of a normal prior distribution in their purchasing outcome. Our model is applied to data made available by Economic Planning Agency, instead of marketing data which are hardly available in actualities, as the agency's data have some similarity to such data as are obtained through purchase intention surveys.

2. INTENTION-TO-BUY RATES AND PURCHASE RATES FROM A PURCHASE INTENTION SURVEY

Respondents to a purchase intention survey are looked on as having an intention to buy if they mark either of the first two in the following fivepoint format, which exemplifies such a survey asked of respondents who have been allowed to use a new product:

I definitely would buy it.
I probably would buy it.
I might or might not buy it.
I probably would not buy it.
I definitely would not buy it.

(1) [5] proposes to measure not only buying intentions but needs in order to get more correct estimates.

Concerning the intention-to-buy, respondents are divided into four groups: Those who have shown an intention and buy; those who have not shown an intention but buy; those who have shown an intention and do not buy; those who have not shown an intention and do not buy. Rates of respondents grouped above against the total number are denoted by t, u, s, and v respectively. Hense, s plus t equals the rate, denoted by q, of respondents who have shown the intention-to-buy, whereas t plus u equals the rate, denoted by p, of respondents who buy. The relations are shown below.

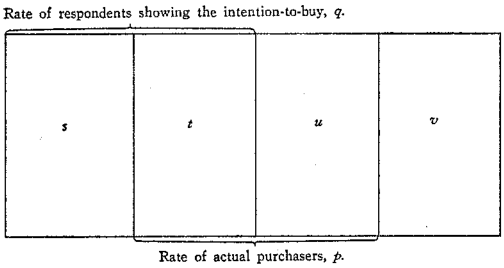

Fig. 1.

A research example in [6] tells us $u=0$, therefore $q>p$, which seems to be a general case in the marketing context. The data, to which we apply our model at section 4, are not the ones collected at some stage of development of a new product. Instead, they are the data obtained by Economic Planning Agency's survey from a viewpoint of national economy, not referring to a specific brand of a product but to a category representing the product. A questionnaire by this survey seeks such information as follows concerning the intention-to-buy:

> In the following list of durable goods, which did you buy in the last three months, and which do you plan to buy during the three months to come?

In this case, in which any of them is, of course, not made available for respondents to use, $u>s$, therefore $p>q$ almost in the periods concerned. (See Table 1.)

3. A THEORETICAL MODEL

The fundamental equation is given by

$$p=q+w, \tag{1}$$

where $w=u-s$, being called adjustment term. We are called on to estimate p after observing the estimated value of q. Suppose that we have histories of q and w,(2) or their estimated values, then we can specify prior distributions for those which are denoted by $h_0(q)$ and $g_0(w)$. They are given by

$$h_0(q) = N(\bar{q}_0, \sigma_{q0}^2)\,, \tag{2}$$

$$g_0(w) = N(\bar{w}_0, \sigma_{w0}^2)\,, \tag{3}$$

where $N(\bar{q}_0, \sigma_{q0}^2)$ represents a normal distribution of q with a mean, $\bar{q}_0$, and a variance, σ_{q0}^2. That is,

$$N(\bar{q}_0, \sigma_{q0}^2) \equiv \frac{1}{\sqrt{2\pi}\ \sigma_{q0}} \exp\left[-\frac{(q-\bar{q}_0)^2}{2\sigma_{q0}^2}\right]. \tag{4}$$

We read $g_0(w)$ and the corresponding notations likewise. They must have their parameters so that their values of density might be nearly zero except in the interval $[0-1]$ of variables because rates vary only from 0 to 1.

Now suppose that n respondents are randomly selected from among consumers in point and that x of them give the intention-to-buy answers. If n is large enough, a distribution of $\hat{q}=x/n$ denoted by $m(\hat{q}|q)$ becomes approximately a normal distribution, as given by

$$m(\hat{q}|q) = N\big(q, q(1-q)/n\big)\,, \tag{5}$$

from which we know q is a mean of $\hat{q}$ and a variance of $\hat{q}$ is $q(1-q)/n$. Given $\hat{q}$, a distribution, $h(q|\hat{q})$, of q, that is, a posterior distribution, $h_1(q)$, of q is obtained from Bayes' theorem.(3)

$$h_1(q) = h(q|\hat{q}) = \frac{h_0(q)\, m(\hat{q}|q)}{\int_{-\infty}^{\infty} h_0(q)\, m(\hat{q}|q)\, dq} = N(q_1, \sigma_{q1}^2)\,; \tag{6}$$

where

$$q_1 = (\hat{q}\sigma_{q0}^2 + \bar{q}_0\sigma_{\hat{q}}^2)/(\sigma_{q0}^2 + \sigma_{\hat{q}}^2)\,, \tag{7}$$

$$\sigma_{q1}^2 = \sigma_{q0}^2\sigma_{\hat{q}}^2/(\sigma_{q0}^2 + \sigma_{\hat{q}}^2)\,. \tag{8}$$

We assume that q is statistically independent of w and that a distribution of w is not changed by observing $\hat{q}$. Then a distribution, $f_1(p)$, of p after observing $\hat{q}$ is given by

$$f_1(p) = N(\bar{p}_1, \sigma_{p1}^2)\,, \tag{9}$$

where $\bar{p}_1 = \bar{q}_1 + \bar{w}_0$, $\sigma_{p1}^2 = \sigma_{q1}^2 + \sigma_{w0}^2$.
For the standard normal distribution $\varphi(z)$, the value of z, of which probability on the right is $\frac{\alpha}{2}$, is designated as $K_{\frac{\alpha}{2}}$, that is

(2) In case we have histories of p and w, see [4].

(3) A standard textbook of statistics, for example, section 2.8 in [1] gives us the ideas of Bayesian estimations.

$$0.5-\int_0^{K_{\frac{\alpha}{2}}}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z}{2}\right)^2 dz=\frac{\alpha}{2}. \tag{10}$$

Then, a $100\times(1-\alpha)$ percent Bayesion interval estimate of p is given by

$$p_1-K_{\frac{\alpha}{2}}\sigma_{p1}\leq p\leq p_1+K_{\frac{\alpha}{2}}\sigma_{p1}. \tag{11}$$

Next, we get a sample size n, given a width of the estimation interval, E, and the degree of a confidence of $(1-\alpha)\times100\%$. From (9), we get

$$E=2K_{\frac{\alpha}{2}}\sigma_{p1}=2K_{\frac{\alpha}{2}}\sqrt{\sigma_{q1}^2+\sigma_{w0}^2}. \tag{12}$$

Hence,

$$\sigma_{q1}^2=\sigma_{q0}^2\sigma_{\hat{q}}^2/(\sigma_{q0}^2+\sigma_{\hat{q}}^2)=\left\{E/(2K_{\frac{\alpha}{2}})\right\}^2-\sigma_{w0}^2. \tag{13}$$

Replacing $\sigma_{\hat{q}}^2$ with $q(1-q)/n$, we solve (13) for n:

$$n=q(1-q)\Big/\left\{\left(\frac{E}{2K_{\frac{\alpha}{2}}}\right)^2-\sigma_{w0}^2\right\}^{-1}-\frac{q(1-q)}{\sigma_{q0}^2}. \tag{14}$$

However, we do not know q or $\hat{q}$ before a survey. If we substitute 1/2 into q, we get the largest value of $q(1-q)$, and (14) is changed into the following inequation:

$$n\leq\frac{1}{4}\left[\left\{\left(\frac{E}{2K_{\frac{\alpha}{2}}}\right)^2-\sigma_{w0}^2\right\}^{-1}-\sigma_{q0}^{-2}\right]. \tag{15}$$

The value given by the right-hand side of (15) guarantees the first conditions at least.

4. EMPIRICAL EXAMPLES

We use only data of the washing machine and the color TV set among about forty consumer durable goods surveyed in [3]. That, in fact, does not tell us true p and q but estimated values of them. It is impossible to know true rates because all the consumers cannot be surveyed in this case. We do not know true rates concerning new products developed until they are sold. What to do for us then is to compare $\overset{*}{p}$ with $\hat{p}$, the former being the theoretical value of p on the basis of our model and the latter the surveyed estimated value of p, almost regarded as true p.

First we make four prior distributions corresponding to four kinds of research periods, that is, Jun.-Aug., Sept.-Nov., Dec.-Feb. and Mar.-May periods, respectively denoted by A, B, C, and D. Prior means in those periods of q and w are denoted by $\bar{q}_{A0}$, $\bar{q}_{B0}$, $\bar{q}_{C0}$, $\bar{q}_{D0}$ and $\bar{w}_{A0}$, $\bar{w}_{B0}$, $\bar{w}_{C0}$, $\bar{w}_{D0}$, respectively, variances are denoted by σ_{qA0}^2, σ_{qB0}^2, σ_{qC0}^2, σ_{qD0}^2 and σ_{wA0}^2, σ_{wB0}^2, σ_{wC0}^2, σ_{wD0}^2 respectively. Those parameters are calculated from data for each

TABLE 1.

	Washing Machine			Color TV Set		
	Planned ($\hat{q}$)	Actual ($\hat{p}$)	Adjustment ($\hat{w}$)	Planned ($\hat{q}$)	Actual ($\hat{p}$)	Adjustment ($\hat{w}$)
Jun. 1971–Aug.	0.038	0.043	0.005	0.035	0.048	0.013
Sept.–Nov.	0.026	0.036	0.010	0.036	0.046	0.010
Dec.–Feb. 1972	0.019	0.023	0.004	0.038	0.062	0.024
Mar.–May	0.022	0.042	0.020	0.021	0.042	0.021
Jun.–Aug.	0.027	0.036	0.009	0.022	0.048	0.026
Sept.–Nov.	0.026	0.028	0.002	0.026	0.044	0.018
Dec.–Feb. 1973	0.023	0.034	0.011	0.034	0.060	0.026
Mar.–May	0.021	0.033	0.012	0.019	0.045	0.026
Jun.–Aug.	0.026	0.037	0.011	0.015	0.032	0.017
Sept.–Nov.	0.021	0.033	0.012	0.020	0.038	0.018
Dec.–Feb. 1974	0.019	0.017	−0.002	0.018	0.052	0.034
Mar.–May	0,015	0.032	0.017	0.007	0.031	0.024
Jun.–Aug.	0.023	0.028	0.005	0.011	0.033	0.022
Sept.–Nov.	0.031	0.026	−0.05	0.014	0.028	0.014
Dec.–Feb. 1975	0.012	0.025	0.013	0.020	0.039	0.019
Mar.–May	0.013	0.026	0.013	0.007	0.045	0.038

(This table is derived from the survey result of agricultural households in [3]. And $\hat{w}$ represents surveyed estimated value of w.)

of the four periods starting in the Jun.-Aug. period in 1972, as shown in Table 1, since which period the survey of planned purchases and that of actual purchases have been compiled in the same period.

For instance, we calculate $\bar{q}_{A0}$ and σ^2_{qA0} of the washing machine like below from the above Table:

$$\bar{q}_{A0} = (0.038+0.027+0.026+0.023)/4 = 0.029\,,$$

$$\sigma^2_{qA0} = \left\{(0.038-0.029)^2+(0.027-0.029)^2+(0.026-0.029)^2+(0.023-0.029)^2\right\} \div 4$$
$$= 0.0057^2\,.$$

We calculate other values likewise. Hence, we get

$$g_0(q) = N(0.029,\ 0.0057^2)\,, \tag{16}$$

$$h_0(w) = N(0.0075,\ 0.0026^2)\,. \tag{17}$$

Other parameters are shown in Table 2.

Now we estimate p from $\hat{q}$ in the Jun.-Aug. period in 1976. Table 3 tells us $\hat{q}=0.018$. From (5), we get $\sigma^2_{\hat{q}}=0.018\times0.982\div720=0.0049^2$. From (7) and (8), the posterior mean and the variance of q in that period, which

TABLE 2.

Washing Machine		Color TV Set	
$\bar{q}_{A0}=0.029$	$\sigma_{qA0}^2=0.0057^2$	$\bar{q}_{A0}=0.021$	$\sigma_{qA0}^2=0.0091^2$
$\bar{q}_{B0}=0.026$	$\sigma_{qB0}^2=0.0035^2$	$\bar{q}_{B0}=0.024$	$\sigma_{qB0}^2=0.0080^2$
$\bar{q}_{C0}=0.018$	$\sigma_{qC0}^2=0.0040^2$	$\bar{q}_{C0}=0.028$	$\sigma_{qC0}^2=0.0086^2$
$\bar{q}_{D0}=0.018$	$\sigma_{qD0}^2=0.0038^2$	$\bar{q}_{D0}=0.014$	$\sigma_{qD0}^2=0.0065^2$
$\bar{w}_{A0}=0.0075$	$\sigma_{wA0}^2=0.0026^2$	$\bar{w}_{A0}=0.020$	$\sigma_{wA0}^2=0.0049^2$
$\bar{w}_{B0}=0.0048$	$\sigma_{wB0}^2=0.0068^2$	$\bar{w}_{B0}=0.014$	$\sigma_{wB0}^2=0.0030^2$
$\bar{w}_{C0}=0.0065$	$\sigma_{wC0}^2=0.0059^2$	$\bar{w}_{C0}=0.026$	$\sigma_{wC0}^2=0.0054^2$
$\bar{w}_{D0}=0.016$	$\sigma_{wD0}^2=0.0032^2$	$\bar{w}_{D0}=0.027$	$\sigma_{wD0}^2=0.0065^2$

TABLE 3.

	Washing Machine			Color TV Set			Sample Size
	$\hat{q}$	$\overset{*}{p}$	$\hat{p}$	$\hat{q}$	$\overset{*}{p}$	$\hat{p}$	(n)
Jun.-Aug. 1975	0.018	0.031	0.030	0.011	0.033	0.030	720
Sept.-Nov.	0.011	0.024	0.024	0.021	0.036	0.023	715
Dec.-Feb. 1976	0.013	0.022	0.023	0.017	0.046	0.037	713
Mar.-May	0.018	0.034	0.036	0.015	0.042	0.036	703

(The values of $\hat{q}$ and $\hat{p}$ are from [3]. Those of $\overset{*}{p}$ are derived from the theoretical model.)

are denoted by $\bar{q}_{A1}$ and σ_{Aq1}^2, respectively, are given by

$$\begin{aligned}\bar{q}_{A1} &= (0.029\times 0.0049^2+0.018\times 0.0057^2)\div(0.0057^2+0.0049^2)\\ &= 0.023\end{aligned}$$

$$\sigma_{qA1}^2 = 0.0057^2\times 0.0049^2\div(0.0057^2+0.0049^2)=0.0036^2 .$$

Hence, (9) tells us

$$\begin{aligned}f_1(p) &= N(0.023+0.0075,\ 0.0036^2+0.0026^2)\\ &= N(0.031,\ 0.0044^2) .\end{aligned}$$

Using (11), with a 95% degree of confidence (therefore $K_{\frac{\alpha}{2}}=1.96$ from (10)), the estimation interval of p is given by

$$0.031-1.96\times 0.0044 \le p \le 0.031+1.96\times 0.0044 .$$

Calculating this down to 3 places of decimals, we get 0.031 ± 0.009 at the above inequation. Table 3 and Fig. 2 show theoretical point estimates of p, which is denoted by $\overset{*}{p}$, for one year starting in the Jun.-Aug. period in 1975.

5. DISCUSSION AND SUMMARY

After estimating $\overset{*}{p}$ from $\hat{q}$, the question is whether or not $\overset{*}{p}$ is in the neighborhood of p. Discussion of the result of the empirical examples reveals the following: Assuming $\hat{p}$ to be true p, Table 3 and Fig. 2 can tell us that both values have a good fitness in the case of the washing machine, but have some discrepancy in the case of the color TV set.

Such discrepancy should be filled with prior information related to the purchase above and beyond the purchase intention. This means that a researcher is called on to incorporate other prior information in $h_0(q)$ and $g_0(w)$. What information is incorporated in the specified distribution and how it is done are depend upon researcher's decisions.

Therefore, when he specifies prior distributions, he becomes a decision maker rather than just a researcher. Our model proposed allows us to have the logical framework in which subjective factors are incorporated in the process of estimating p.

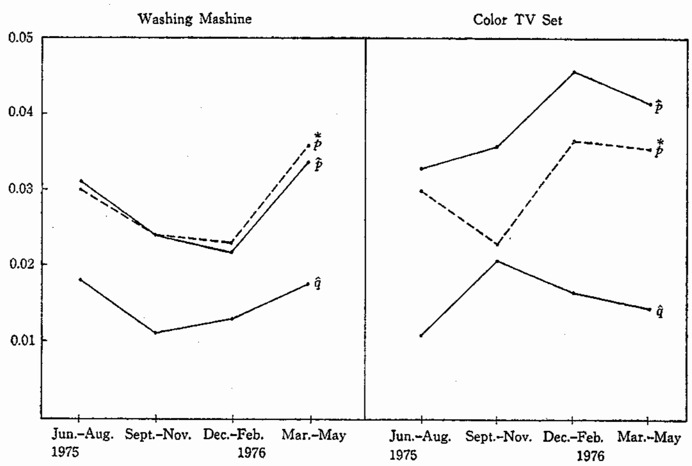

Fig. 2.

REFERENCES

[1] Hoel, P. G., S. Port and C. J. Stone, *Introduction to Statistical Theory*, Moughton Mifflin Company, 1971.

[2] Morrison, D. G., "Purchase Intentions and Purchase Behavior," *Journal of Marketing*, Vol. 43, No. 2, 1979, pp. 65–74.

[3] Research Bureau of Economic Planning Agency, *Movement of Consumption and Savings*, 1972–1977. Keizai Kikakuchō Chōsakyoku (Research Bureau of Economic Planning Agency), Shōhi to Chochiku no Dōkō (Movement of Consumption and Savings), 1972–1977.

[4] Satō, Y. "Estimating the Proportion of Buyers from the Purchase Intention Survey—Its Formations by Personal Probability Approaches," *Economic Studies* (Hokkaido Univ.), Vol. 29, No. 1, 1979, pp. 269–288.

[5] Tauber, E. M., "Reduce New Product Failures: Measure Needs as well as Purchase Interest," *J. M.*, Vol. 37, No. 3, 1973, pp. 61–64.

[6] Taylor, J. W., J. J. Houlahan and A. C. Gabriel, "The Purchase Intention Question in New Product Development: a Field Test," *J. M.*, Vol. 39, No. 1, 1975, pp. 90–92.