The risk function approach to profit maximizing estimation in direct mailing

Lars Muus Hiek van der Scheer Tom Wansbeek *

September 16, 1999

Abstract

When the parameters of the model describing consumers' reaction to a mailing are known, addresses for a future mailing can be selected in a profit-maximizing way. Usually, these parameters are unknown and are to be estimated. Standard estimation are based on a quadratic loss function. In the present context an alternative loss function is suggested by the mailing company's profit function. This leads to different estimators and higher expected profit.

risk function, empirical Bayes estimator, bootstrap, marketing

1 Introduction

We consider the following situation. A direct mailing company (the 'firm', for brevity) has a data base with addresses to which it considers sending a mailing. For each of these 'list members' the value of a number of background variables is known. From a test mailing the influence of these variables on response behavior can be analyzed. To estimate the parameters of this process a number of econometric techniques are available like probit analysis or nonparametric methods, cf. Bult and Wansbeek (1995). When

*Muus: Department of Economics, University of Aarhus, Denmark; Van der Scheer and Wansbeek: Department of Economics, University of Groningen, P.O.Box 800, 9700 AV Groningen, The Netherlands.

the response parameters are known the list members can be ranked and the most promising addresses can be selected.

In this process two steps, estimation and selection, are considered separately. The purpose of this note is to indicate how the expected profit of the firm can be increased by integrating the two steps. Following Blattberg and George (1992), the idea behind this is that estimation takes usually place by considering (asymptotic) squared-error loss, which puts the same price at over- and under-estimates of the parameters. In the case that we consider, the firm's expected profit can be expressed as a function of the parameters, and this function is not symmetric. The expected profit can hence be increased by pursuing an alternative estimation strategy.

In section 2 we outline the model and the risk function. Section 3 derives an empirical Bayes estimator for the crucial parameter in the model. It appears that generically it is optimal to select more addresses than would follow from taking the squared-error based estimator at face value. This approach is based on a strong distributional assumption on the parameter. A way to circumvent this problem is using a bootstrap approach. This is pursued in section 4. An empirical illustration is provided in section 5, and section 6 concludes.

2 The model and the risk function

Let the sample used for the test mailing be indexed by i, $i = 1, \ldots, n$. The characteristics of the list members as known to the firm are contained in the vector x_i. We assume that the inclination to respond positively to a mailing is y_i, which obeys the model

$$y_i = x_i'\beta - \alpha + u_i,$$

where α and β are parameters, and u_i is a disturbance term, distributed independently of x_i. We do not observe y_i itself but only its sign: when $y_i > 0$ an addressee has responded, and when $y_i \leq 0$ he hasn't. We moreover assume that u is normal, and we set its variance freely at 1. The parameters α and β are appropriately estimated by probit analysis.

When these parameters are known, they can be applied to the complete list to select the set of addresses to be mailed. For each member of the list, the value of the 'index' $n \equiv x'\beta$ is computed. High values of n indicate a good prospect, and low values a bad prospect. It remains to find the 'cut-off' point n_c separating the two. As is easily seen, it is determined by the equality of cost to (expected) returns. We arbitrarily normalize the cost of a mailing to 1 and denote by w the returns to a positive reply. The response probability at index value n is $\Phi(n-\alpha)$, so n_c has to satisfy

$$\Phi(n_c-\alpha)=\frac{1}{w},$$

so

$$n_c(\alpha)=\Phi^{-1}\left(\frac{1}{w}\right)+\alpha.$$

So when α increases, hence the average inclination to respond decreases, the cut-off value goes up and less addresses should receive a mailing. The same holds when w decreases, i.e. the mailing is less lucrative on average.

Now assume that the distribution of n is continuous,

$$n\sim f(n),\quad f(n)>0\quad \forall n.$$

Then we can express the expected profit of the firm as

$$\pi(\alpha)=\int_{n_c(\alpha)}^{\infty}\{w\Phi(n-\alpha)-1\}f(n)\mathrm{d}n.$$

The notation suggests the important role played by α. It enters the profit function not only explicitly but also through $n_c(\alpha)$.

When the parameters are not known but an estimate $\hat{\alpha}$ is substituted, the 'feasible' cut-off point is

$$n_c(\hat{\alpha})=\Phi^{-1}\left(\frac{1}{w}\right)+\hat{\alpha}, \tag{1}$$

leading to the profit function

$$\pi(\hat{\alpha})=\int_{n_c(\hat{\alpha})}^{\infty}\{w\Phi(n-\alpha)-1\}f(n)\mathrm{d}n.$$

This suggests $\pi(\alpha)-\pi(\hat{\alpha})$ as the loss function to be employed in profit-maximizing estimation of α. An obvious risk function then is

$$R(\hat{\alpha})=\mathrm{E}_{\alpha}\int_{n_c(\alpha)}^{n_c(\hat{\alpha})}\{w\Phi(n-\alpha)-1\}f(n)\mathrm{d}n, \tag{2}$$

where the expectation is taken with respect to the distribution of α. Minimizing this risk function gives a profit-maximizing estimator of α.

In order to operationalize this procedure we have to become explicit as to the distribution of α. A practical approach is to use the sampling distribution of the estimator of α employing standard optimization based on squared loss. We call the ensuing estimator the empirical Bayes estimator. Another approach is to use a bootstrap technique. The following two sections deal with these approaches.

3 Empirical Bayes estimation

The first approach, the empirical Bayes approach, takes $\alpha \sim N(\mu_\alpha, \sigma_\alpha^2)$, where μ_α is the point estimator of α and σ_α^2 its sampling variance. The estimator of $\hat{\alpha}$ follows from minimizing the risk (2). The first-order condition is

$$\begin{aligned}
\frac{\mathrm{d}R(\hat{\alpha})}{\mathrm{d}\hat{\alpha}} &= \mathrm{E}_\alpha \left\{ w\Phi\left(n_c(\hat{\alpha}) - \alpha\right) - 1 \right\} f(n_c(\hat{\alpha})) \\
&= \mathrm{E}_\alpha \left\{ w\Phi\left(\Phi^{-1}\left(\frac{1}{w}\right) + \hat{\alpha} - \alpha\right) - 1 \right\} f(n_c(\hat{\alpha})) \\
&= w\sigma_\alpha \mathrm{E}_{\alpha^*} \left\{ \Phi\left(\Phi^{-1}\left(\frac{1}{w}\right) + \hat{\alpha} - \sigma_\alpha\alpha^* - \mu_\alpha\right) - \frac{1}{w} \right\} f(n_c(\hat{\alpha})) \\
&= w\sigma_\alpha \left\{ \Phi\left(\frac{\Phi^{-1}\left(\frac{1}{w}\right) + \hat{\alpha} - \mu_\alpha}{\sqrt{1+\sigma_\alpha^2}}\right) - \frac{1}{w} \right\} f(n_c(\hat{\alpha})) = 0, \qquad (3)
\end{aligned}$$

where the third step is based on the transformation

$$\alpha^* \equiv \frac{\alpha - \mu_\alpha}{\sigma_\alpha}$$

and the last step is based on the fact that, for $x \sim N(0, 1)$, there holds[1]

$$\mathrm{E}_x \Phi(ax + b) = \mathrm{E}_x \mathrm{E}_z I_{(-\infty, ax+b)}(z),$$

with $z \sim N(0, 1)$, independent of x, and

$$\begin{aligned}
\mathrm{E}_x \mathrm{E}_z I_{(-\infty, ax+b)}(z) &= \mathrm{E}_x \mathrm{E}_z I_{(-\infty, b)}(z - ax) \\
&= P\{z - ax \le b\} \\
&= \Phi\left(\frac{b}{\sqrt{1+a^2}}\right).
\end{aligned}$$

From the last line of (3) the solution for $\hat{\alpha}$ appears to be

$$\hat{\alpha} = \mu_\alpha + \Phi^{-1}\left(\frac{1}{w}\right)\left\{\sqrt{1+\sigma_\alpha^2} - 1\right\}, \qquad (4)$$

so the optimal cut-off point is

$$n_c(\hat{\alpha}) = \mu_\alpha + \Phi^{-1}\left(\frac{1}{w}\right)\sqrt{1+\sigma_\alpha^2}$$

[1]We are indebted to Ton Steerneman for the result and its derivation.

instead of (1).

Summarizing, from (4) we see that the firm increases its expected profit by adjusting the estimator μ_α from a first, squared-error based estimation step by a term whose sign depends on $\Phi^{-1}\left(\frac{1}{w}\right)$. This term is negative for $w > 2$, i.e. the returns to a positive reply exceed twice the cost of a mailing. We may consider this to be the typical case. Then, μ_α has to be updated in a downward direction, so more addresses have to be mailed than the outcome of classical estimation implies. The size of the term increases with σ_α^2.

Note that the distribution $f(n)$ of n plays no role in the estimator.

4 Bootstrap estimation

The empirical Bayes method as we implemented it in the previous section is attractive in the sense that it yields a closed-form expression for the estimator. The price to be paid is the introduction of a possibly unrealistic distributional assumption on α. This can be avoided by the bootstrap approach.

This approach starts out by estimating α a number of times, say $M = 200$. This yields estimators $\alpha_1, \ldots, \alpha_M$. The risk function then is

$$R_B(\hat{\alpha}) = \frac{1}{M}\sum_{m=1}^{M}\int_{n_c(\alpha_m)}^{n_c(\hat{\alpha})}\{w\Phi(n-\alpha_m)-1\}\, f(n)\mathrm{d}n,$$

leading to the first-order condition

$$\frac{\mathrm{d}R_B(\hat{\alpha})}{\mathrm{d}\hat{\alpha}} = \frac{1}{M}\sum_{m=1}^{M}\left\{w\Phi\left(n_c(\hat{\alpha})-\alpha_m\right)-1\right\} f(n_c(\hat{\alpha})).$$

So $\hat{\alpha}$ is the solution of

$$\frac{1}{M}\sum_{m=1}^{M}\Phi\left(\Phi^{-1}\left(\frac{1}{w}\right)+\hat{\alpha}-\alpha_m\right) = w.$$

This equation is easily solved by numerical methods.

5 An empirical application

6 Conclusion

We have shown how in direct mailing profit maximization leads to a natural form of the loss function in estimation. (bla bla bla)

There are a number of imperfections, though. One is that we have concentrated on the intercept parameter α, neglecting β while doing so. The role played by β is much less explicit, and the only way β enters into the analysis is through the index n. For our analysis to be valid knowledge of β is not required, but we have implicitly assumed that the ordering according to values $n = x'\beta$ is correct even though we have to substitute an estimator for β when computing the indices n.

Another extension would be to introduce second-order considerations. (bla bla) This can be used to determine the optimal size of a test mailing, atopic which has been neglected in the literature so far.

Bounded support of n, or discrete n

Graphs!!!!

References

Blattberg, R.C. and E.I. George (1992), "estimation under profit-driven loss functions", *Journal of Business & Economic Statistics*, **10**, 437–444.

Bult, J.R. and T.J. Wansbeek (1995), "Optimal selection for direct mail", *Marketing Science*, forthcoming.