

University of Groningen

A decision theoretic framework for profit maximization in direct marketing

Muus, L.; Scheer, H. van der; Wansbeek, T.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

1996

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Muus, L., Scheer, H. V. D., & Wansbeek, T. (1996). *A decision theoretic framework for profit maximization in direct marketing*. s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

A decision theoretic framework for profit maximization in direct marketing

Lars Muus Hiek van der Scheer Tom Wansbeek*

September 1996

Abstract

One of the most important issues facing a firm involved in direct marketing is the selection of addresses from a mailing list. 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 have to be estimated. These estimates are used to rank the potential addressees and to select the best targets.

Several methods for this selection process have been proposed in the recent literature. All of these methods consider the estimation and selection step separately. Since estimation uncertainty is neglected, these methods lead to a suboptimal decision rule and hence not to optimal profits. We derive an optimal Bayes decision rule that follows from the firm's profit function and which explicitly takes estimation uncertainty into account. We show that the integral resulting from the Bayes decision rule can be either approximated through a normal posterior, or numerically evaluated by a Laplace approximation or by Markov chain Monte Carlo integration. An empirical example shows that higher profits result indeed.

* Muus: Department of Economics, Universitetsparken, 8000 Aarhus C, University of Aarhus, Denmark; Van der Scheer and Wansbeek: Department of Economics, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands.

1. Introduction

Consider a firm engaged in direct marketing, which has to decide which households within a large population to send a mailing. In order to decide which households to solicit it is of crucial importance for the firm to assess how the household's response probability depends on its characteristics (demographic variables, attitudes, etc.) known to the firm. If the effect of the characteristics on the response probability are known, potential addressees can be ranked and the most promising *targets* can be selected.

Of course, these effects are unknown and have to be estimated. Typically, a firm specifies and estimates a *response model* based on a test mailing to obtain some knowledge on the effects of the characteristics on the response probability. For this purpose a number of techniques have been directed, including the traditional probit analysis and various nonparametric methods, e.g. Magidson (1988), Banslaben (1992), Bult (1993), and Bult and Wansbeek (1995). The estimates obtained are then used to formulate a *decision rule* to select households from a mailing list.

This separation of parameter estimation and formulation of decision rules does not, in general, lead to optimal profits since a suboptimal decision rule is specified (Klein et al. 1978). The reason for this is that estimation usually takes place by considering (asymptotic) squared-error loss, which puts equal weight at over- and under-estimating the parameters. However, while a squared-error loss function may be useful when summarizing properties of the response function, it completely ignores the economic objectives of the marketing firm. Rather, the inferential process should be embedded in the firm's decision making framework, taking explicitly into account the firm's objective of maximizing expected profit. Put differently, the decision maker should take the *estimation risk* into account when formulating a decision rule regarding which households to solicit. The loss resulting structure is, in general, asymmetric in contrast to the traditional squared-error loss structure. Consequently, the traditional methods thus yield suboptimal decision rules.

The purpose of this paper is to formulate a strict decision theoretic framework for a marketing firm engaged in direct marketing. In particular, we derive an optimal Bayes rule deciding when to send a mailing to a household with a given set of characteristics. This formal approach has a number of advantages. First of all, a rigorous decision theoretic framework clarifies the essential ingredients entering the marketing firm's decision problem. By deriving the optimal Bayes rule based on an expected profit loss function, the present framework yields admissible decision rules with respect to the marketing firm's economic objective. Furthermore, the estimation uncertainty resulting from the firm's assessment of the characteristics of

the population of potential targets is explicitly taken into account as an integral part of the optimal decision procedure. Thus, the decision theoretic procedure provides a more firm theoretical foundation for optimal decision making on part of the firm. Equally important, the present framework provides decision rules yielding higher profits to the firm.

Integration of the estimation and decision step has been studied thoroughly in statistics (e.g., Berger 1985, DeGroot 1970). This formal decision theoretic framework has been applied in a number of economic decision making situations, including portfolio selection (cf. Bawa, Brown and Klein 1979), real estate assessment (Varian 1975), and agricultural economics (e.g., Lence and Hayes 1994). For further economic applications see Cyert and DeGroot (1987). To the best of our knowledge, only one paper on optimal decision making under uncertainty has been applied to marketing questions (Blattberg and George 1992). These authors consider a firm whose goal it is to maximize profits by determining the optimal price. They conclude that the firm is better off by charging a higher price than the price resulting from traditional methods, which are based on the estimated price sensitivity parameter. However, in contrast with our approach, they consider a loss function that results from a rather ad-hoc specified model, with only one unknown parameter.

The paper is organized as follows. In the next Section we formulate the decision theoretic framework and derive the optimal Bayes decision rule. We show that the decision rule crucially depends on the estimation uncertainty facing the firm. The estimation uncertainty can be incorporated through a posterior density. In Section 3 we derive a closed form expression for the integral resulting from the optimal decision rule by approximating the posterior by the asymptotically normal density of the maximum likelihood (probit) estimator. In Section 4 we discuss the Laplace approximation and Markov chain Monte Carlo integration, which can be used to calculate the integral of interest. In Section 5 we discuss an empirical example, using data provided by a charity firm. Applying the formal decision framework appears to generate the higher profits indeed. We conclude in Section 6.

2. The decision theoretic framework

Consider a direct marketing firm that has the option of mailing or not mailing to potential targets. In case a mail is sent to a given household the profit to the firm, π , is given by

$$\pi = rR - c,$$

where r is the revenue from a positive reply, c is the mailing cost, and R is a random variable given by

$$R = \begin{cases} 1 & \text{if the household responds} \\ 0 & \text{if the household does not respond.} \end{cases}$$

Clearly, $c < r$ if the firm has to obtain positive profits at all. We assume that the response is driven by a probit model. Hence, the response probability of a household is

$$P(R = 1 | x, \beta) = \Phi(x'\beta),$$

where $\Phi(\cdot)$ is the standard normal integral, x is a $k \times 1$ vector of regressors and β is a $k \times 1$ vector of regression coefficients ($\beta \in \mathcal{B} \subseteq \mathbf{R}^k$). In case a mail is sent, the expected profit given x and β is

$$E(\pi | x, \beta) = rE(R | x, \beta) - c = r\Phi(x'\beta) - c. \quad (1)$$

With an unknown β the firm has to make a decision whether to send a mail ($d = 1$) or not ($d = 0$) to a given household. The loss function considered in the following is given by

$$\mathcal{L}(d, \beta | x) = \begin{cases} r\Phi(x'\beta) - c & \text{if } d = 1 \\ 0 & \text{if } d = 0. \end{cases} \quad (2)$$

Notice, that the above loss function is naturally induced by the firm's economic profit maximization objective. In this sense, the present decision theoretic framework naturally encompasses the phenomena of estimation uncertainty, without introducing rather ad hoc statistical criteria.

Inference on the parameter vector β is obtained through a test mailing, resulting in the sample

$$\mathcal{S}_n \equiv \{(x_1, R_1), \dots, (x_n, R_n)\}.$$

The posterior density, using Bayes' rule, is given by

$$f(\beta | \mathcal{S}_n, \theta) = \frac{L(\beta | \mathcal{S}_n)f(\beta | \theta)}{f(\mathcal{S}_n | \theta)}, \quad (3)$$

where $L(\beta | \mathcal{S}_n)$ is the likelihood function corresponding to the sample,

$$L(\beta | \mathcal{S}_n) = \prod_{i=1}^n \Phi(x'_i \beta)^{R_i} (1 - \Phi(x'_i \beta))^{1-R_i},$$

and $f(\beta | \theta)$ denotes the prior density, $\theta \in \Theta \subseteq \mathbf{R}^p$ is a $p \times 1$ vector of hyperparameters. Finally, $f(\mathcal{S}_n | \theta)$ denotes the predictive density given by,

$$f(\mathcal{S}_n | \theta) = \int L(\beta | \mathcal{S}_n) f(\beta | \theta) d\beta. \quad (4)$$

The *posterior risk* corresponding to the loss function (2) is then given by

$$\begin{aligned} \mathcal{R}(d | x) &\equiv \mathbb{E}(\mathcal{L}(d, \beta | x) | \mathcal{S}_n) \\ &= \begin{cases} r \int \Phi(x' \beta) f(\beta | \mathcal{S}_n, \theta) d\beta - c & \text{if } d = 1 \\ 0 & \text{if } d = 0. \end{cases} \end{aligned} \quad (5)$$

The *Bayes decision rule* corresponding to the posterior risk (5) is the decision variable d maximizing $\mathcal{R}(d | x)$. It is easily seen that this decision rule is given by

$$d = 1 \quad \text{if and only if} \quad \int \Phi(x' \beta) f(\beta | \mathcal{S}_n, \theta) d\beta \geq \frac{c}{r}. \quad (6)$$

Notice that this decision rule explicitly takes into account the estimation uncertainty inherent when the firm does not know the parameter vector β . The *Bayes optimal mailing region*, denoting the households to whom a mail should be sent, is hence given by

$$\mathcal{M}_B \equiv \left\{ x \in \mathbf{R}^k \mid \int \Phi(x' \beta) f(\beta | \mathcal{S}_n, \theta) d\beta \geq \frac{c}{r} \right\}.$$

The structure of the mailing region may, in general, be quite complicated.

It is often recommended to base the firm's mailing decision on the point estimates obtained from the test mailing. These point estimates are typically derived by implicitly assuming a squared-error loss function, resulting from the use of standard estimation procedures. As this squared-error loss does not reflect the actual loss suffered by the firm, using the point estimate motivated by squared-error loss will be inappropriate. If the firm neglects the estimation uncertainty it would specify a decision rule based on a point estimate of β , say $\hat{\beta}$, e.g. the probit estimator based on \mathcal{S}_n . The point estimate then is used as if it is the true parameter value (e.g., Bult and Wansbeek 1995). The resulting decision rule, which we call the *naive* decision rule, is thus given by

$$d = 1 \quad \text{if and only if} \quad \Phi(x'\hat{\beta}) \geq \frac{c}{r}. \quad (7)$$

This rule evidently ignores the estimation uncertainty surrounding $\hat{\beta}$. Indeed, by a second order Taylor series expansion of $\Phi(x'\beta)$, we obtain

$$\Phi(x'\beta) \approx \Phi(x'\hat{\beta}) + (\beta - \hat{\beta})'x\phi(x'\hat{\beta}) - \frac{1}{2}x'\hat{\beta}\phi(x'\hat{\beta})x'(\beta - \hat{\beta})(\beta - \hat{\beta})'x,$$

using the fact that the derivative of $\phi(t)$ is $-t\phi(t)$, where $\phi(\cdot)$ is the standard normal density. Hence, an approximate Bayes decision rule is given by,

$$\Phi(x'\hat{\beta}) - \frac{1}{2}x'\hat{\beta}\phi(x'\hat{\beta})x'Mx \geq \frac{c}{r} \quad (8)$$

where $M \equiv E(\hat{\beta} - \beta)(\hat{\beta} - \beta)'$ denotes the mean square error matrix of the estimator $\hat{\beta}$. The major difference between the (approximate) Bayes rule (8) and the naive rule (7) is that estimation uncertainty is explicitly taken into account in the former. Evidently, if the estimation uncertainty is small, i.e. M is small, the approximate Bayes rule (8) is adequately approximated by the naive decision rule (7). Notice that the mailing region for the naive rule is the half space given by

$$\mathcal{M}_N \equiv \left\{ x \in \mathbf{R}^k \mid \Phi(x'\hat{\beta}) \geq \frac{c}{r} \right\}$$

The result of applying the naive decision rule is thus to approximate the mailing region \mathcal{M}_N by the halfspace \mathcal{M}_N . As will be demonstrated below this approximation may be rather crude, resulting in a suboptimal level of profits.

In order to implement the optimal decision rule (6), we need to evaluate the expectation of $\Phi(x'\beta)$ over the posterior density of β . If the posterior admits a closed form solution and is of a rather simple analytical form, this expectation can be solved analytically. Otherwise, numerical methods need to be implemented in order to assess the decision rule (6). In Section 4 we explore various numerical strategies for evaluating the decision rule. However, it is instructive to consider the case where the posterior density is normal in which case we can fully characterize the mailing region.

3. The case of a normal posterior

If the posterior density is normal with mean μ and covariance matrix Ω , we can obtain a closed form expression for (6), namely¹

$$\begin{aligned}
& \int \Phi(x'\beta) f(\beta \mid \mathcal{S}_n, \theta) d\beta \\
&= E_\beta (\Phi(x'\beta)) \quad \text{where } \beta \sim N(\mu, \Omega) \\
&= E_b (\Phi(x'\Omega^{1/2}b + x'\mu)) \quad \text{where } b = \Omega^{-1/2}(\beta - \mu) \sim N(0, I_k) \\
&= E_b E_z I_{\{-\infty, x'\Omega^{1/2}b + x'\mu\}}(z) \quad \text{with } z \sim N(0, 1), \text{ independent of } b \\
&= E_b E_z I_{\{-\infty, x'\mu\}}(z - x'\Omega^{1/2}b) \\
&= P(z - x'\Omega^{1/2}b < x'\mu) \\
&= \Phi\left(\frac{x'\mu}{(1 + x'\Omega x)^{1/2}}\right). \tag{9}
\end{aligned}$$

Hence, the mailing region is given by

$$\begin{aligned}
\mathcal{M}_B &= \left\{ x \in \mathbf{R}^k \mid \Phi\left(\frac{x'\mu}{(1 + x'\Omega x)^{1/2}}\right) \geq \frac{c}{r} \right\} \\
&= \left\{ x \in \mathbf{R}^k \mid x'\mu \geq \gamma(1 + x'\Omega x)^{1/2} \right\}. \tag{10}
\end{aligned}$$

where

$$\gamma \equiv \Phi^{-1}\left(\frac{c}{r}\right).$$

Since in any practical situation $c \ll r$, we assume $\gamma < 0$ whenever the sign of γ is relevant. Notice that, when $\Omega_1 > \Omega_2$, $1 + x'\Omega_1 x > 1 + x'\Omega_2 x$. Thus, since $\gamma < 0$, greater uncertainty as to β implies that the mailing region expands.

Expression (9) enables us to show explicitly that the Bayes decision rule generates higher expected profits than the naive decision rule. The expected profit (cf. (1)), in case mail is sent, is

$$\begin{aligned}
q(x) &\equiv E_\beta(E(\pi \mid x, \beta)) \\
&= r\Phi\left(\frac{x'\mu}{(1 + x'\Omega x)^{1/2}}\right) - c.
\end{aligned}$$

¹ We are indebted to Ton Steerneman for bringing the result to our attention, and for providing this derivation.

For all x in \mathcal{M}_B there holds, by definition, that $q(x) > 0$. Since $\mathcal{M} \subseteq \mathcal{M}_B$ it follows that the expected profit is lower for the naive decision rule.

We consider these mailing region in somewhat more detail. The boundary of the mailing region \mathcal{M}_B is given by

$$\{x \in \mathbf{R}^k \mid x'\mu = \gamma(1 + x'\Omega x)^{1/2}\} \quad (11)$$

We assume that $\Omega > 0$. By squaring and rewriting the argument of (11) we obtain

$$x'(\mu\mu' - \gamma^2\Omega)x = \gamma^2, \quad (12)$$

which can be written as

$$x'\Omega^{1/2}(\Omega^{-1/2}\mu\mu'\Omega^{-1/2} - \gamma^2 I_k)\Omega^{1/2}x = \gamma^2. \quad (13)$$

Let

$$\begin{aligned} A_1 &\equiv \frac{\Omega^{-1/2}\mu\mu'\Omega^{-1/2}}{\mu'\Omega^{-1}\mu} \\ A_2 &\equiv I_k - A_1 \\ \lambda &\equiv \mu'\Omega^{-1}\mu - \gamma^2; \end{aligned}$$

A_1 and A_2 are idempotent matrices of rank 1 and $k - 1$, respectively, $A_1A_2 = 0$, and $A_1 + A_2 = I_k$. Hence, we can write (13) as

$$x'\Omega^{1/2}(\lambda A_1 - \gamma^2 A_2)\Omega^{1/2}x = \gamma^2.$$

Let $A_1 = z_1z_1'$ and $A_2 = Z_2Z_2'$, so (z_1, Z_2) is orthonormal. Then

$$\begin{aligned} G &\equiv \lambda A_1 - \gamma^2 A_2 \\ &= \lambda z_1z_1' - \gamma^2 Z_2Z_2' \\ &= (z_1, Z_2) \begin{pmatrix} \lambda & 0 \\ 0 & -\gamma^2 I_{k-1} \end{pmatrix} \begin{pmatrix} z_1' \\ Z_2' \end{pmatrix} \end{aligned}$$

Hence, the eigenvalues of G are $-\gamma^2$ with multiplicity $k - 1$, and λ with multiplicity one. The sign of λ depends on Ω . Informally speaking, for small values of Ω , $\lambda > 0$, and for large values, $\lambda < 0$. In the first case G has one positive and $k - 1$ negative eigenvalues. Due to ‘Sylvester’s law of inertia’ (e.g. Lancaster and Tismenetsky 1985, p. 188), the same holds for $\mu\mu' - \gamma^2\Omega$. Hence, the matrix is indefinite and the boundary is a hyperboloid in the x -space. When the uncertainty as to β is so large that

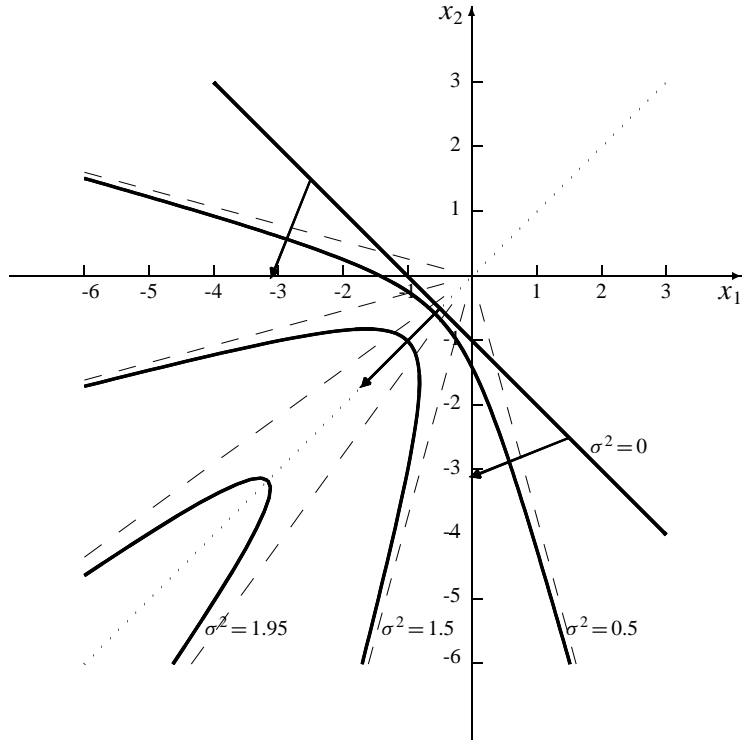


Figure 3.1: The naive and Bayes optimal mailing region compared. The area to the north-east of the straight line is \mathcal{M}_N , and the ellipsoids bound \mathcal{M}_B for various values of σ^2 .

$\lambda < 0$, all eigenvalues of G are negative and (12) does not have a solution. Hence, all households should be included in the mailing campaign.

We illustrate the mailing region it for $k = 2$, $\mu' = (1, 1)$, $\gamma = -1$, and

$$\Omega = \begin{pmatrix} \sigma^2 & \sigma_{12} \\ \sigma_{12} & \sigma^2 \end{pmatrix}.$$

Then, from (10), the mailing region is

$$\mathcal{M}_B = \{x_1, x_2 \mid x_1 + x_2 \geq -\sqrt{1 + \sigma^2(x_1 + x_2) + 2\sigma_{12}x_1x_2}\},$$

which reduces to the halfspace $x_1 + x_2 \geq -1$ if $\sigma^2 = \sigma_{12} = 0$. The matrix in (12) becomes

$$\mu\mu' - \gamma^2\Omega$$

$$\begin{aligned}
&= \begin{pmatrix} 1 - \sigma^2 & 1 - \sigma_{12} \\ 1 - \sigma_{12} & 1 - \sigma^2 \end{pmatrix} \\
&= \frac{1}{2} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2 - \sigma^2 - \sigma_{12} & 0 \\ 0 & -(\sigma^2 - \sigma_{12}) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}. \quad (14)
\end{aligned}$$

Hence, the matrix $\mu\mu' - \gamma^2\Omega$ has one negative eigenvalue, $-(\sigma^2 - \sigma_{12})$, and one eigenvalue that is positive if $\sigma^2 + \sigma_{12} < 2$. Using (14), (12) can be rewritten as

$$(2 - \sigma^2 - \sigma_{12})(x_1 + x_2)^2 - (\sigma^2 - \sigma_{12})(x_1 - x_2)^2 = 2,$$

which is a hyperbola in \mathbf{R}^2 . Its asymptotes are found by putting the left-hand side equal to zero. On letting

$$\varphi \equiv \sqrt{\frac{2 - \sigma^2 - \sigma_{12}}{\sigma^2 - \sigma_{12}}},$$

these asymptotes are found to be

$$\varphi(x_1 + x_2) = \pm(x_1 - x_2),$$

or

$$\frac{x_2}{x_1} = \frac{1 - \varphi}{1 + \varphi} \quad \text{and} \quad \frac{x_2}{x_1} = \frac{1 + \varphi}{1 - \varphi}.$$

Figure 3.1 illustrates the boundary for $\sigma_{12} = 0$, and $\sigma^2 = 0, .5, 1.5$, and 1.95 , respectively. If $\sigma^2 = 0$ we have a straight line. This bounds the mailing region of the naive method or. The mailing region increases as σ^2 increases; the arrows indicate the direction of the increase. When $\sigma^2 \geq 2$, the mailing region is simply \mathbf{R}^2 . The distance between the straight line corresponding with $\sigma^2 = 0$ and the hyperbola is larger when the x -value is larger. This reflects the fact that the uncertainty as to $x'\beta$ increases by the (absolute) value of x .

4. Numerical evaluation of the optimal Bayes rule

Numerical implementation of the optimal Bayes decision rule (6) requires the evaluation, for each value of x , of the integral

$$Q(x) \equiv \int \Phi(x'\beta) f(\beta | \mathcal{S}_n, \theta) d\beta \quad (15)$$

$$= \frac{\int \Phi(x'\beta)L(\beta | \mathcal{S}_n)f(\beta | \theta) d\beta}{\int L(\beta | \mathcal{S}_n)f(\beta | \theta) d\beta}, \quad (16)$$

using (3) and (4) in the last step. We will now explore various methods for evaluating this integral. Henceforth, we denote the probit estimate of β , based on \mathcal{S}_n , by $\hat{\beta}$, and covariance matrix by $\hat{\Omega}$ (e.g. the inverse of the Fisher information matrix evaluated in $\hat{\beta}$).

Normal posterior approximation

It is well known that the posterior density converges under suitable regularity conditions to a normal distribution, with mean $\hat{\beta}$ and covariance matrix $\hat{\Omega}$, when the sample size is sufficiently large (Jeffreys 1967, p. 193, Heyde and Johnstone 1979). Obviously, the approximation may be rather crude, since it is solely based on the asymptotic equivalence of the Bayes and maximum likelihood estimator. Thus, this approximation completely ignores the prior distribution $f(\beta | \theta)$. However, as we showed in Section 3, this property appears to be very valuable since it enables us to obtain a closed form expression for (15), which is given in (9) by substitution of $\hat{\beta}$ for μ and $\hat{\Omega}$ for Ω . Moreover, Zellner and Rossi (1984) showed that, for moderate sample sizes ($n = 100$), the normal posterior approximation works well for the logit model.

Laplace approximation

A more refined asymptotic approximation is the Laplace approximation proposed by Tierney and Kadane (1986) (see also Kass et al. 1988, and Tierney et al. 1989). The Laplace approximation of (16) is given by

$$\hat{Q}(x) = \frac{\Psi_1(\hat{\beta}_1)|H_1(\hat{\beta}_1)|^{-1/2}}{\Psi_0(\hat{\beta}_0)|H_0(\hat{\beta}_0)|^{-1/2}}$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the maximizers of $\Psi_0(\cdot)$ and $\Psi_1(\cdot)$, respectively, and

$$\begin{aligned} \Psi_0(\beta) &\equiv L(\beta | \mathcal{S}_n)f(\beta | \theta) \\ \Psi_1(\beta) &\equiv \Phi(x'\beta)L(\beta | \mathcal{S}_n)f(\beta | \theta), \end{aligned}$$

and

$$H_0(\beta) \equiv -\frac{\partial^2 \ln \Psi_0(\beta)}{\partial \beta \partial \beta'}$$

$$H_1(\beta) \equiv -\frac{\partial^2 \ln \Psi_1(\beta)}{\partial \beta \partial \beta'}.$$

By means of the Laplace approximation, the integral $Q(x)$ is thus evaluated without any need for numerical integration. Instead the Laplace approximation requires maximization, in order to determine $\hat{\beta}_0$ and $\hat{\beta}_1$, and differentiation, in order to find $H_0(\cdot)$ and $H_1(\cdot)$. For $\hat{\beta}_0$ and $\hat{\beta}_1$ we use the values obtained by a single Newton-Raphson step from $\hat{\beta}$ when maximizing $\ln \Psi_0(\beta)$ and $\ln \Psi_1(\beta)$, which does not affect the rate at which the approximation error vanishes. As demonstrated by Tierney and Kadane (1986), Kass et al. (1988), and Tierney et al. (1989), the general error of the approximation vanishes at rate n^{-2} . As these authors demonstrate, this approximation is often very accurate.

We apply this approximation for an informative prior and an uninformative prior. As to the former we choose for $f(\beta | \theta)$ the normal density with mean $\hat{\beta}$ and covariance matrix $\hat{\Omega}$. Since, $\partial \ln f(\beta | \theta) / \partial \beta = -\hat{\Omega}^{-1}(\beta - \hat{\beta})$, we have $\hat{\beta}_0 = \hat{\beta}$, and in the Appendix we show that

$$\hat{\beta}_1 = \hat{\beta} + \xi(\hat{\beta}) H_0(\hat{\beta})^{-1} x, \quad (17)$$

where $\xi(\cdot)$ is a scalar function defined in (18).

For the uninformative prior we use Jeffreys' prior (e.g. Berger 1985, pp. 82-89, and Zellner 1971, pp. 41-53), given by

$$f(\beta | \theta) = \left| -E \left(\frac{\partial^2 \ln L(\beta | \mathcal{S}_n)}{\partial \beta \partial \beta'} \right) \right|^{1/2}.$$

Notice that no hyperparameters are involved here. Within the context of binary response models this prior has been examined by, among others, Ibrahim and Laud (1991), and Poirier (1994). These authors support the use of Jeffreys' prior as an uninformative prior but notice that it can be quite cumbersome to work with analytically as well as numerically.

Monte Carlo integration

The recent development of *Markov chain Monte Carlo* (MCMC) procedures has revolutionized the practice of Bayesian inference. See, for example, Tierney (1994), and Gilks et al. (1995) for expositions of basic Markov chain Monte Carlo procedures. These algorithms are easy to implement and have the advantage that they do not require evaluation of the normalizing constant of the posterior density, given by

(4). As a *candidate density* it is natural to select the asymptotic approximation, $q(\beta) \sim N(\hat{\beta}, \hat{\Omega})$. The density of interest, the so-called *target density*, is given by

$$h(\beta) \equiv L(\beta | \mathcal{S}_n) f(\beta | \theta).$$

The independence sampler (Tierney 1994), a special case of the Hastings-Metropolis algorithm, is used to generate random variates β_j , $j = 1, \dots, J$, from the (unnormalized) density $h(\beta)$ through the following algorithm, where β_0 is arbitrarily selected:

1. draw a *candidate point*, β_j^* , from $q(\cdot)$
2. draw u_j from the uniform density on $(0, 1)$
3. if $u_j \leq \alpha(\beta_{j-1}, \beta_j^*)$, then $\beta_j = \beta_j^*$, else $\beta_j = \beta_{j-1}$.

Here

$$\alpha(\beta_{j-1}, \beta_j^*) \equiv \begin{cases} \min\left(\frac{h(\beta_j^*)q(\beta_{j-1})}{h(\beta_{j-1})q(\beta_j^*)}, 1\right) & \text{if } h(\beta_{j-1})q(\beta_j^*) > 0 \\ 1 & \text{else.} \end{cases}$$

The generated β_j 's, $j = 1, \dots, J$ are used to evaluate the integral by

$$\hat{Q}(x) = \frac{1}{J} \sum_{j=1}^J \Phi(x' \beta_j).$$

We use this algorithm instead of more advanced MCMC procedures, like the Gibbs sampler (e.g. Albert and Chib 1993), since we have a candidate density that is a good approximation of the target distribution (Roberts 1995). Again, we apply this algorithm for the (informative) normal prior and for the (uninformative) Jeffreys' prior.

5. Application

We illustrate our approach with an application based on data from a charitable foundation in the Netherlands. This foundation heavily rests on direct mailing. Every year it sends mailings to almost 1.2 million households in the Netherlands. The dependent variable is the response/nonresponse in 1991. The explanatory variables are the amount of money (in NLG) donated in 1990 (A90) and 1989 (A89), the interaction between these two (INT), the date of entry on the mailing list (ENTRY), the family size (FS), own opinion on charitable behavior in general (CHAR; four categories: donates never, donates sometimes, donates regularly, and donates always).

The data set consists of 40 000 observation. All the households on the list donated at least once to the foundation since entry on the mailing list.

In order to have a sufficiently large validation sample we used 1 000 observations for estimation. The response rate in the estimation sample is 31.8%. This rather high response rate is not surprising since charitable foundations have in general high response rates (Statistical Fact Book 1994-1995), and the mailing list consists of households that responded to this particular foundation before. The average amount of donation in the estimation sample is NLG 14.56, the cost of a mailing is NLG 3.50. We use the average amount of donation for household selection and to determine the profit implications. Table 5.1 gives the probit estimates and the average of the coefficients based on the independence sampler with the normal and Jeffreys' prior, respectively. The donation in 1990 and 1989 are, as expected, positively related with the response probability. The negative sign of the interaction term can be interpreted as a correction for overestimation of the response probability if a households responded in 1990 and 1989. The other three coefficients do not significantly differ from zero. As expected, the average value of the coefficients for the independence sampler are similar to the probit estimates. The standard deviations, however, of the normal prior are much smaller.

The basic difficulty in MCMC procedures is the decision when the generated sequence of parameters has converged to a sample of the target distribution. Many diagnostic tools to address this convergence problem have been suggested in the recent literature (see Cowles and Carlin 1996 for an extensive overview). Following the recommendations of these authors, we generated six parallel sequences of parameters with starting points systematically chosen from a large number of drawings from a distribution that is overdispersed with respect to the target distribution. We inspected the sequences of each parameter by displaying them in a common graph and in separate graphs. We used the Gelman-Rubin statistics (Gelman and Rubin 1992) to quantitatively analyze the sequences. The results of these diagnostics are satisfying, indicating an almost immediate convergence of the sample.

Table 5.2 shows the profit implications for the various approaches to determine the posterior risk function and the naive approach for the validation sample. As a benchmark we also give the situation in which the foundation sends all the households a mailing. Of these 39 000 households, 13 274 responded, generating a net profit of NLG 56 784. If the foundation would have used the naive selection approach they would have selected 87.03% (33 946) of the households, with a net profit of NLG 59 345. Using the Bayes decision rule, the foundation would have selected more households, as expected. This ranges from 34 018 of the Laplace approximation with the normal prior to 34 271 of the independence sampler with Jeffreys' prior.

Table 5.1: Probit estimates and results of the independence sampler

	Probit Estimates ¹	Independence Sampler ²	
		Normal prior	Jeffreys' prior
Constant	-0.3938 (0.4511)	-0.3964 (0.3120)	-0.3948 (0.4539)
A90	0.0052 (0.0014)	0.0053 (0.0010)	0.0051 (0.0014)
A89	0.0074 (0.0030)	0.0074 (0.0021)	0.0072 (0.0030)
INT	-0.0056 (0.0029)	-0.0057 (0.0019)	-0.0053 (0.0027)
ENTRY	-0.0063 (0.0048)	-0.0063 (0.0033)	-0.0063 (0.0048)
FS	-0.1526 (0.1408)	-0.1503 (0.1003)	-0.1513 (0.1397)
CHAR	0.0683 (0.0537)	0.0680 (0.0371)	0.0685 (0.0530)

¹ Asymptotic standard errors in parentheses

² Standard deviation, based on $J = 10\,000$, in parentheses

Except for the Laplace approximation with the normal prior, the additional selected households generate sufficient response to increase the net profits, reinforcing the importance of the Bayes decision rule. Net profits increase with 4.5% if the naive selection is used instead of selecting all the households. This percentage increases to 5.3% if we apply the normal posterior approximation, and to 5.4% when using the independence sampler with Jeffreys' prior. Given that the foundation's database contains 1.2 million targets, these increases turn out to be quite substantial. Notice that the figures of the Laplace approximation and independence sampler with the normal prior are much closer to those of the naive approach than those with Jeffreys' prior. This makes intuitive sense since informative priors put more weight to values of β near $\hat{\beta}$. In the case of the posterior density degenerating at $\hat{\beta}$, i.e. perfect prior information on β , the decision rule is equivalent to the naive rule.

Table 5.2: Target selection and profit implications

	Number selected	Response	Actual profit (NLG)
No Selection	39 000	13 274	56 784
Naive approach	33 946	12 236	59 345
Normal posterior	34 240	12 337	59 787
Laplace approximation:			
Normal prior	34 018	12 250	59 297
Jeffreys' prior	34 256	12 341	59 789
Independence sampler:			
Normal prior	34 153	12 310	59 698
Jeffreys' prior	34 271	12 347	59 824

6. Discussion and conclusion

In order to select addresses from a list for a direct mailing campaign, a firm can build a response model and use the (consistently) estimated parameters for selection. The decision rule for selection is often defined on the basis of the estimated parameters taken as the true parameters. This paper shows that this leads to suboptimal results. The reason for this is that the estimation uncertainty resulting from the firm's assessment of the characteristics of the potential targets is not taken into account. Put differently, both steps of a target selection process, estimation and selection, should be considered simultaneously. We formulated a rigorous theoretic framework, based on the firm's profit maximizing behavior, to derive an optimal Bayes decision rule. We demonstrated theoretically as well as empirically that this approach generates higher profits.

An important aspect of our approach is the evaluation of the integral resulting from the Bayes decision rule. We used a normal posterior, Laplace approximation, and Monte Carlo integration to evaluate the Bayes rule numerically. Although the normal posterior approach may be rather crude it has the advantage that a closed form expression can be derived, and, moreover, it performs quite well in the empirical illustration. As a consequence of the former, we do not need the computational intensive methods. Moreover, we obtain a transparent expression for the expected profit, which explicitly shows the effect of estimation risk. It has to be realized, however, that the empirical results indicate that the decision rule is affected by the chosen prior density. Since the normal posterior approximation ignores the prior density, it has to be used with caution when prior information is available.

This paper has some limitations. First, we considered only the question of selecting households for one direct mailing campaign. That is, we did not consider the long-term impact of the selection process. Second, we solely considered the binary response choice to the mailing and not the amount of money donated. Third, we made the implicit assumption that the parameters are constant across households. This assumption may be unrealistic in practice. It runs, for example, counter to the idea of trying to customize promotions through direct marketing. A company could deal with this kind of heterogeneity by using, for example, latent class analysis (DeSarbo and Ramaswamy 1994, Wedel et al. 1993). We want to stress, however, that these assumptions are commonly made in direct marketing research. Furthermore, our method results from a general decision theoretic framework that can be extended, in principle, to situations that do suffer from these limitations, in a straightforward manner.

References

- Albert, J.H. and S. Chib (1993), “Bayesian analysis of binary and polychotomous response data”, *Journal of the American Statistical Association*, **88**, 669–679.
- Balestra, P. (1976), *La Dérivation Matricielle*, Sirey, Paris.
- Banslaben, J. (1992), “Predictive modeling”, in E.L. Nash, editor, *The Direct Marketing Handbook*, McGraw-Hill, New York, 620–636.
- Bawa, V.S., S.J. Brown, and R.W. Klein, editors (1979), *Estimation Risk and Optimal Portfolio Choice*, North-Holland, Amsterdam.
- Berger, J.O. (1985), *Statistical Decision Theory and Bayesian Analysis*, second edition, Springer-Verlag, New York.
- Blattberg, R.C. and E.I. George (1992), “Estimation under profit-driven loss function”, *Journal of Business & Economic Statistics*, **10**, 437–444.
- Bult, J.R. (1993), “Semiparametric versus parametric classification models: An application to direct marketing”, *Journal of Marketing Research*, **30**, 380–390.
- Bult, J.R. and T.J. Wansbeek (1995), “Optimal selection for direct mail”, *Marketing Science*, **14**, 378–394.
- Cowles, M.K. and B.P. Carlin (1996), “Markov chain Monte Carlo convergence diagnostics: A comparative review”, *Journal of the American Statistical Association*, **91**, 883–904.
- Cyert, R.M. and M.H. DeGroot (1987), *Bayesian Analysis and Uncertainty in Economic Theory*, Chapman and Hall, London.
- DeGroot, M.H. (1970), *Optimal Statistical Decisions*, McGraw-Hill, New York.
- DeSarbo, W.S. and V. Ramaswamy (1994), “Crisp: Customer response based iterative segmentation procedures for response modeling in direct marketing”, *Journal*

- of *Direct Marketing*, **8** (3), 7–20.
- Gelman, A. and D.B. Rubin (1992), “Inference from iterative simulation using multiple sequences”, *Statistical Science*, **7**, 457–511.
- Gilks, W.R., S. Richardson, and D.J. Spiegelhalter, editors (1995), *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London.
- Heyde, C.C. and I.M. Johnstone (1979), “On asymptotic posterior normality for stochastic processes”, *Journal of the Royal Statistical Society B*, **41**, 184–189.
- Ibrahim, J.G. and P.W. Laud (1991), “On Bayesian analysis of generalized linear models using Jeffreys’s prior”, *Journal of the American Statistical Association*, **86**, 981–986.
- Jeffreys, H. (1967), *Theory of Probability*, third edition, Oxford University Press, London.
- Kass, R.E., L. Tierney, and J.B. Kadane (1988), “Asymptotics in Bayesian computation”, in J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith, editors, *Bayesian Statistics 3*, Oxford University Press, Oxford.
- Klein, R.W., L.C. Raftery, D.S. Sibley, and R.D. Willig (1978), “Decisions with estimation uncertainty”, *Econometrica*, **46**, 1363–1387.
- Lancaster, P. and M. Tismenetsky (1985), *The Theory of Matrices*, second edition, Academic Press, Orlando.
- Lence, S.H. and D.J. Hayes (1994), “The empirical minimum-variance hedge”, *American Journal of Agricultural Economics*, **76**, 94–104.
- Magidson, J. (1988), “Improved statistical techniques for response modeling”, *Journal of Direct Marketing*, **2** (4), 6–18.
- Poirier, D. (1994), “Jeffreys’ prior for logit models”, *Journal of Econometrics*, **63**, 327–339.
- Roberts, G.O. (1995), “Markov chain concepts related to sampling algorithms”, in W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London, 45–57.
- Statistical Fact Book (1994-1995)*, Direct Marketing Association, Inc., New York, NY.
- Tierney, L. (1994), “Markov chains for exploring posterior distributions (with discussion)”, *Annals of Statistics*, **22**, 1701–1762.
- Tierney, L. and J.B. Kadane (1986), “Accurate approximations for posterior moments and marginal densities”, *Journal of the American Statistical Association*, **81**, 82–86.
- Tierney, L., R.E. Kass, and J.B. Kadane (1989), “Fully exponential Laplace approximations to expectations and variances of nonpositive functions”, *Journal of the American Statistical Association*, **84**, 710–716.
- Varian, H.R. (1975), “A Bayesian approach to real estate assessment”, in S.E. Fienberg

and A. Zellner, editors, *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage*, North-Holland, Amsterdam.

Wedel, M., W.S. DeSarbo, J.R. Bult, and V. Ramaswamy (1993), “A latent class poisson regression model for heterogeneous count data”, *Journal of Applied Econometrics*, **8**, 397–411.

Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, John Wiley, New York.

Zellner, A. and P.E. Rossi (1984), “Bayesian analysis of dichotomous quantal response models”, *Journal of Econometrics*, **25**, 365–393.

Appendix: On the Laplace approximation

We first prove (17), then we give the derivatives of Jeffreys’ prior. Let

$$\begin{aligned} g_0(\beta) &\equiv \frac{\partial \ln \Psi_0(\beta)}{\partial \beta} = \frac{\partial \ln L(\beta \mid \mathcal{S}_n)}{\partial \beta} - \hat{\Omega}^{-1}(\beta - \hat{\beta}) \\ g_1(\beta) &\equiv \frac{\partial \ln \Psi_1(\beta)}{\partial \beta} = \frac{\phi}{\Phi} x + g_0(\beta) = \zeta x + g_0(\beta), \end{aligned}$$

where $\phi \equiv \phi(x' \beta)$, $\Phi \equiv \Phi(x' \beta)$, and $\zeta \equiv \frac{\phi}{\Phi}$ is the inverse of Mills’ ratio. Notice that $g_0(\hat{\beta}) = 0$. Further,

$$\begin{aligned} H_0(\beta) &\equiv -\frac{\partial^2 \ln L(\beta \mid \mathcal{S}_n)}{\partial \beta \partial \beta'} + \hat{\Omega}^{-1} \\ H_1(\beta) &\equiv \frac{\phi(\phi + x' \beta \Phi)}{\Phi^2} x x' + H_0(\beta) = \zeta(\zeta + x' \beta) x x' + H_0(\beta). \end{aligned}$$

Then $\hat{\beta}_1$ follows from the Newton-Raphson step

$$\begin{aligned} \hat{\beta}_1 &= \hat{\beta} + H_1(\hat{\beta})^{-1} g_1(\hat{\beta}) \\ &= \hat{\beta} + \left(\hat{\zeta}(\hat{\zeta} + x' \hat{\beta}) x x' + H_0(\hat{\beta}) \right)^{-1} g_1(\hat{\beta}) \\ &= \hat{\beta} + \frac{1}{1 + \hat{\zeta}(\hat{\zeta} + x' \hat{\beta}) x' H_0(\hat{\beta})^{-1} x} H_0(\hat{\beta})^{-1} g_1(\hat{\beta}) \\ &= \hat{\beta} + \xi(\hat{\beta}) H_0(\hat{\beta})^{-1} x, \end{aligned}$$

where $\hat{\zeta}$ denotes ζ evaluated in $\hat{\beta}$, and

$$\xi(\beta) \equiv \frac{\zeta}{1 + \zeta(\zeta + x' \beta) x' H_0(\beta)^{-1} x}. \quad (18)$$

We will now derive the first and second derivative of Jeffreys' prior, given by

$$f(\beta | \theta) = \left| -E \left(\frac{\partial^2 \ln L(\beta | \mathcal{S}_n)}{\partial \beta \partial \beta'} \right) \right|^{1/2} = |A|^{1/2}$$

where

$$A \equiv \sum_{i=1}^n \frac{\phi_i^2}{D_i} x_i x_i',$$

with $\phi_i \equiv \phi(x_i' \beta)$ and $D_i \equiv \Phi_i(1 - \Phi_i)$, where $\Phi_i \equiv \Phi(x_i' \beta)$. Using some well known properties of matrix differentiation (e.g. Balestra 1976), we obtain the logarithmic first derivative

$$\frac{\partial \ln |A|^{1/2}}{\partial \beta} = \frac{1}{2} ((\text{vec} A^{-1})' \otimes I_k) \text{vec} \left(\frac{\partial A}{\partial \beta} \right).$$

Let

$$M \equiv \text{vec} I_k \otimes I_k,$$

then we can write, using the product rule for matrices, the second derivative as

$$\begin{aligned} \frac{\partial^2 \ln |A|^{1/2}}{\partial \beta \partial \beta'} &= \\ & \frac{1}{2} \left[((\text{vec} A^{-1})' \otimes I_k) \left(I_k \otimes \frac{\partial^2 A}{\partial \beta \partial \beta'} \right) \right. \\ & \quad \left. - M' \left(I_k \otimes \left((A^{-1} \otimes I_k) \frac{\partial A}{\partial \beta} A^{-1} \frac{\partial A}{\partial \beta'} \right) \right) \right] M. \end{aligned}$$

Finally, to complete the derivatives we need an expression for $\partial A / \partial \beta$ and $\partial^2 A / \partial \beta \partial \beta'$, which are given by

$$\begin{aligned} \frac{\partial A}{\partial \beta} &= - \sum_{i=1}^n \left(\frac{2x_i' \beta \phi_i^2}{D_i} + \frac{\phi_i^3(1 - 2\Phi_i)}{D_i^2} \right) x_i x_i' \otimes x_i \\ \frac{\partial^2 A}{\partial \beta \partial \beta'} &= \sum_{i=1}^n \left(\frac{2\phi_i^2(2(x_i' \beta)^2 - 1)}{D_i} + \frac{5x_i' \beta \phi_i^3(1 - 2\Phi_i) + 2\phi_i^4}{D_i^2} \right. \\ & \quad \left. + \frac{2\phi_i^4(1 - 2\Phi_i)^2}{D_i^3} \right) x_i x_i' \otimes x_i x_i', \end{aligned}$$

which enables us to calculate the derivatives of Jeffreys' prior.