Quantitative approaches for profit maximization in direct marketing
van der Scheer, H.R.

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:
1998

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Chapter 5

On the value of information

5.1 Introduction

A database with customer information is an essential tool in direct marketing to maintain a structural relationship. It enables the organization to target the individuals individually. The basis for a database is a customers history file, the so-called internal list. This list includes names and addresses of individuals who have bought products or services from the organization, or who have made an inquiry, or to whom a direct marketing message has been directed. Naturally, the organization must possess sufficient information for efficient fulfilment, billing, credit screening, promotion reporting, customer inquiry and subsequent selection of promotion (Brooks 1989, Katzenstein and Sachs 1992, pp. 156-161). Furthermore, the organization would also like to have a database that contains individual-specific information on product and brand preferences, needs, wants and demands with respect to product attributes and purchase intentions. If all such data are available, an organization is able to deliver its products effectively and efficiently. In the ideal situation sketched above, perfect information is available at the individual level. However, most organizations are faced with less than perfect information.

There are several reasons why an internal list contains imperfect information. First, the internal list traditionally has no roots in the marketing department. For example, financial institutions possess financial transaction records of their clients, for reasons of business administration. These data are used as a starting point for the creation of a customer history file. It goes without saying that the records have to be adjusted in order to be useful for direct marketing. This adjustment process usually takes a lot of time and effort (Shepard 1995). Second, an internal list is a perishable commodity. Individuals
on the list show a fluctuating degree of activity (e.g. they move, marry, die, or change attitudes). According to Baier (1985), in twelve months an average customer list has as many as 25 percent address revisions. Third, most of the direct marketing organizations are involved in highly competitive markets. In order to maintain or gain market share, the organization has to expand its activities. An organization could find (Ansoff 1965) new markets for its current products (market development strategy), it could develop new products of potential interest to its current markets (product development strategy), or it could develop new products for new markets (diversification strategy). For all three strategies, the direct marketing organization needs information which is not available on the internal list. Therefore, it could, for example, rent information from external lists.

An external list contains collections of characteristics of specific individuals that can be accessed either by name and address or by some other specific identifier (Bickert 1992, Lix et al. 1995). External lists comprise lists primarily compiled for direct marketing and lists compiled for other reasons (such as telephone directories or magazine subscriber lists). External lists can be classified based according to whether the information exists at some aggregated level (usually postal code areas) or at the individual level. External lists containing aggregated information are by definition imperfect since the data are gathered at an aggregated level. In the case of postal code information systems, the value of a characteristic is not known for a specific individual but an average is known for clusters of individuals. For a specific individual this means that the value of a characteristic is known with a certain amount of ‘noise’; we call this proxy information. This proxy information on background variables can be used for selecting individuals in the same way as perfect information is used at the individual level. However, since the selection has to take place per cluster and not per individual, the maximally attainable profits will be lower than in the case where exact individual information is available.

Whether the price to be paid for employing imperfect (external) information is outweighed by the possibilities of a better focused mailing campaign depends on the ‘noise-to-signal’ ratio of the proxies and the sensitivity of the profit function to the information content in the external list. The aim of this chapter is to build a conceptual framework for this trade-off between costs and returns.

This chapter, which is based on Bult, Van der Scheer and Wansbeek (1995), is structured as follows. In section 5.2 we introduce the basic model and show the relation between the value of information, defined by $R^2$ and profits. The relation between postal code information and profits is discussed in section 5.3.
An other type of proxy information is obtained when not the variable of interest is measured but some proxy for it, the so-called measurement error. We examine the relation between this kind of proxy information and profits in section 5.4. In section 5.5 we discuss and prove some theoretical results, which are useful throughout the chapter. Section 5.6 contains a number of concluding remarks.

5.2 Model

5.2.1 Basic model

Following on the traditional approaches describing direct mail response, we assume that individual’s \( i \), \( i = 1, \ldots, N \), inclination to respond, \( \eta_i \), is linearly related to a \( k \)-vector of (observable) regressors, \( \xi_i \),

\[
\eta_i = \alpha + \xi_i ^\prime \beta + \epsilon_i, \tag{5.1}
\]

with

\[
y_i = \begin{cases} 
1 & \text{if } \eta_i > 0 \\
0 & \text{otherwise}, \end{cases} \tag{5.2}
\]

where \( \beta \) is a \( k \)-vector of unknown parameters, \( \alpha \) is the unknown intercept, and \( \epsilon_i \) is an (unobservable) disturbance, assumed to be distributed as \( \epsilon_i \sim N(0, \sigma_\epsilon^2) \); \( \sigma_\epsilon^2 \) is also an unknown parameter. We assume that the regressors are distributed as \( \xi_i \sim N_k(0, \Sigma_\xi) \). Moreover, we assume that \( \epsilon_i \) and \( \xi_i \) are both i.i.d. and mutually independent. We do not observe \( \eta_i \) but only whether or not an individual responded, which we call \( y_i \). We set, without loss of generality, \( y_i = 1 \) if \( \eta_i \) is positive and \( y_i = 0 \) otherwise. In other words, \( i \) will respond if \( i \)'s inclination is positive. Note that the choice of the value zero in (5.2) is not restrictive. We can set this threshold at an arbitrary constant value and subtract this value from the intercept. Furthermore, without loss of generality, we might take the mean of \( \xi_i \) equal to zero, since the mean can be absorbed in the intercept.

The probability that the random variable \( y_i \) equals one is

\[
P(y_i = 1 \mid \xi_i) = \Phi \left( \frac{\alpha + \xi_i ^\prime \beta}{\sigma_\epsilon} \right), \tag{5.3}
\]

where \( \Phi(\cdot) \) is the standard normal distribution function. It is clear that the model is identified up to scale and thus we may impose a normalization on one of the parameters. A normalized variance of the disturbance term is the standard formulation of the probit model.
5.2.2 Optimal selection rule and profits

The first step in the selection process consists of the estimation of the parameters of model (5.1) by ML. Next, we use the resulting estimates to select targets; for the sake of simplicity we disregard the distinction between parameters and their consistent estimators. Given that \( i \) receives a mailing, the expected profits of \( i \) are equal to the response probability times the revenues of positive response (\( \equiv a \)) minus the cost of a mailing (\( \equiv c \)), i.e.

\[
a \Phi \left( \frac{\alpha + \xi_i \beta}{\sigma_x} \right) - c.
\]

The (population) mean expected profit is the integral of (5.4) over the distribution of \( \xi \). The sample mean expected profit is simply the mean of (5.4) over all \( i \). Expected profits are maximized when those individuals are selected for which (5.4) is positive. We define \( n_i \equiv \xi_i / \beta \) as individual’s \( i \) index. Then, the optimal selection rule is to select \( i \) if

\[
a \Phi \left( \frac{\alpha + n_i}{\sigma_x} \right) - c \geq 0.
\]

Hence, we order the individuals by increasing values of \( n_i \), and select those with the highest values of \( n_i \). The value of \( n \) for which (5.5) holds with equality defines the optimal cutoff index, denoted by \( n_c \) and defined by

\[
n_c \equiv \tau \sigma_x - \alpha,
\]

where

\[
\tau \equiv \Phi^{-1} \left( \frac{c}{a} \right).
\]

Thus, all individuals with \( n \geq n_c \) should receive a mailing. Note that \( \tau \) is defined for \( 0 < c/a < 1 \). This makes sense since \( c/a = 0 \) means that \( c = 0 \), which of course never holds in practice. Furthermore, \( c/a = 1 \) means that \( a = c \), which indicates that the expected profits are maximally zero. This maximum is only obtained when the response probability equals one, which does not hold in practice. Clearly, \( c/a \) is never negative and never larger than one (\( a \) is always larger than \( c \)), which implies that \( \Phi^{-1}(\cdot) \) is defined for practical situations.

The proportion of the variance of the dependent variable that is explained by the regressors, i.e. \( R^2 \), is an obvious choice to examine the value of information. The fact is that information is more valuable if more variation in the dependent
variable is explained. An increase in $R^2$ can either be the result of using more explanatory variables or of substituting an explanatory variable for a better one. In a binary choice model the use of $R^2$ is not straightforward because we estimate the probability of a certain outcome but this cannot be compared with the ‘true’ probability because this is unknown (see Windmeijer (1995) for a discussion on $R^2$-measures for a probit model). However, we can express the expected profits as a function of the $R^2$ of the underlying latent relation (5.1).

To do so, we first decompose the variance of $\eta$, 

$$\sigma^2_{\eta} \equiv \text{var}(\eta_i) = \sigma^2_n + \sigma^2_\epsilon,$$

where we use the independence of $\xi_i$ and $\epsilon_i$. Hence, 

$$R^2 = \frac{\sigma^2_n}{\sigma^2_{\eta}} = \frac{\sigma^2_n}{\sigma^2_n + \sigma^2_\epsilon}.$$ 

An increase in $R^2$ means that $\sigma^2_n$ increases and that $\sigma^2_\epsilon$ decreases by the same value since the inclination to respond and thus $\sigma^2_\epsilon$ does not change. We use this to impose an alternative normalization on (5.1). We choose $\sigma^2_\eta = 1$, so that $R^2$ is simply $\sigma^2_n$, and $\sigma^2_\epsilon = 1 - R^2$. Hence, (5.3) can be written as 

$$P(y_i = 1 | \xi_i) = \Phi \left( \frac{\alpha + n_i}{\sqrt{1 - R^2}} \right),$$

and the optimal cutoff index (5.6) becomes 

$$n_c \equiv \tau \sqrt{1 - R^2} - \alpha. \quad (5.7)$$

To assess the value of information we have to express the expected profits as a function of $R^2$. For the individuals that receive the mailing, the expected profits are 

$$E \left( a \Phi \left( \frac{\alpha + n}{\sqrt{1 - R^2}} \right) - c | n \geq n_c \right) = a \int_{n_c}^{\infty} \Phi \left( \frac{\alpha + n}{\sqrt{1 - R^2}} \right) \phi \left( \frac{n}{R} \right) \frac{du}{1 - \Phi \left( \frac{n_c}{R} \right)} - c, \quad (5.8)$$

where $\phi \left( \frac{n}{R} \right)$ is the density of $n$. The denominator is the fraction of selected individuals, 

$$P(n \geq n_c) = 1 - \Phi \left( \frac{n_c}{R} \right)$$

$$= 1 - \Phi \left( \tau \sqrt{\frac{1 - R^2}{R^2}} - \frac{\alpha}{R} \right). \quad (5.9)$$
where we use (5.7). To determine the expected profits per individual on the list we have to multiply (5.8) by (5.9), which gives
\[ a \int_{n_c}^{\infty} \Phi \left( \frac{\alpha + n}{\sqrt{1 - R^2}} \right) \phi \left( \frac{R}{R} \right) \frac{1}{R} \, dn - c \left( 1 - \Phi \left( \frac{n_c}{R} \right) \right). \] (5.10)

The expected total returns are obtained by multiplying this by \( N \), the number of individuals on the list.

As noted before, profits are positively related with \( R^2 \). This results from the economic principle that better decisions will be made with better, or more informative, information. Note, however, that this relation should hold in theory but that in practice we face the problem of the optimal number of regressors. That is, statistical procedures may show a decreasing actual performance when the number of variables is increased beyond a certain bound (e.g. Breiman and Freedman 1983).

To prove algebraically that (5.8) is increasing in \( R^2 \) we should take the derivative of (5.8) with respect to \( R^2 \). Unfortunately this becomes too complicated because \( R^2 \) does not only appear in argument of the integral but also in \( n_c \). It is, however, numerically quite easy to compute (5.8) for different values of \( R^2 \). Figure 5.1 shows the relation for various values of \( \alpha \), with \( a = 10 \) and \( c = 1 \). It shows indeed that profits increase in \( R^2 \). The difference in expected profits for the three values of \( \alpha \) vanishes when \( R^2 \) is close to one. This makes sense intuitively. The fact is that we are perfectly able to discriminate between targets and non-targets, which means that \( n \) will respond if \( n_i \) is above a certain threshold. Hence, all three curves go to \( a - c = 9 \). In contrast, if the importance of the explanatory variables vanishes, i.e. \( R^2 \approx 0 \), it is no longer useful to select individuals on the basis of \( n \). Consequently, either all or none of the individuals should receive a mailing. The former holds if \( a \Phi(\alpha) - c \geq 0 \), which are then the expected profits for each individual. This is depicted in figure 5.2. It shows that if \( R^2 = 0 \), either all or none of the individuals are selected. The organization is indifferent between these two when \( \alpha = \tau \). When \( R^2 = 1 \), the fraction equals \( \Phi(\alpha) \), which follows directly from (5.9).

Expression (5.7) suggests that \( n_c \) increases in \( R^2 \), because \( \tau \) is generally negative (\( \frac{\alpha}{n} < 0.5 \)). This implies that the fraction of selected individuals, i.e. \( P(n \geq n_c) \), decreases in \( R^2 \). However, according to figure 5.2, this does not always hold: it depends on the value of \( \alpha \). This paradox can be explained as follows. A change in \( R^2 \) not only affects \( n_c \) but also the shape of the distribution since \( R^2 \) is the variance of \( n \). As illustrated in figure 5.3, an increase of \( R^2 \) generates a larger \( n_c \) and a flatter distribution, resulting in an increase of
Figure 5.1: Relation between expected profits per selected individual and $R^2$.

Figure 5.2: Fraction of selected individuals, i.e. $P(n \geq n_c)$. 
the fraction of selected individuals. We can also show this algebraically. The derivative of (5.9) with respect to $R$ is

$$
\frac{dP(n \geq n_c)}{dR} = -\phi \left( \frac{\alpha}{R} \left( \sqrt{\frac{1 - R^2}{R^2}} - \frac{\alpha}{R} \right) \left( -\tau \sqrt{\frac{1}{1 - R^2}} + \alpha \right) \right) \frac{1}{R^2}.
$$

It is easy to see that the derivative is decreasing in $R$ if

$$
\alpha > \tau \sqrt{\frac{1}{1 - R^2}}.
$$

since $\phi(\cdot) > 0$ and $\frac{1}{R^2} > 0$. The maximum of the right-hand side is in $R^2 = 0$; hence, the inequality holds for all values of $R^2$ when $\alpha > \tau$, which implies that the fraction of selected individuals decreases in $R^2$. If $\alpha < \tau$, there is an interval of $R^2$ in which the derivative is positive. For example, when $\alpha = -2$, this interval for $R^2$ is $(0, 0.59)$, which becomes immediately apparent from figure 5.2.

By combining figure 5.1 and figure 5.2, i.e. expression (5.10), we obtain the expected profits for all individuals on the mailing list. This is depicted in figure 5.4. Here, we also see that the expected profits increase in $R^2$. Profits
increase, however, with a slower rate than in figure 5.1 since we have to multiply (5.8) with a value smaller than one to obtain (5.10). Note that expected profits of (5.10) equal that of (5.8) if $R^2 \approx 0$, since all or none of the individuals are selected. The curves in figure 5.4 are rather flat but it must be realized that a direct mail campaign may involve millions of individuals and this increase in profits could turn out to be a very large number for the total database. Moreover, the relation holds for a particular $x$-vector. When an additional explanatory variable is added to the equation, the value of $\alpha$ changes and so the curve changes.

In the sections below we consider (5.1) with one regressor. It is therefore useful to give the relevant expression for the optimal cutoff index and profit function here. Henceforth we also use the traditional normalization of $\sigma_x^2 = 1$. The optimal cutoff index reduces to an optimal cutoff point, which is defined by

$$\xi_c \equiv \frac{\tau - \alpha}{\beta}, \quad (5.11)$$
and (5.8) becomes

\[
E \left( a \Phi(\alpha + \xi \beta) - c \mid \xi \geq \xi_L \right) = a \frac{\int_{\xi_L}^{\infty} \Phi(\alpha + \xi \beta) \phi \left( \frac{\xi}{\sigma_\xi} \right) \frac{1}{\sigma_\xi} \, d\xi}{1 - \Phi \left( \frac{\xi_L}{\sigma_\xi} \right)} - c, \quad (5.12)
\]

where \( \phi \left( \frac{\xi}{\sigma_\xi} \right) \frac{1}{\sigma_\xi} \) is the density of \( \xi \). Analogous to (5.8), the denominator gives the fraction of selected individuals.

5.3 Postal code information

5.3.1 What is postal code information?

The Netherlands consist of almost 400,000 postal code areas, which divide the total of approximately 6.4 million street name/house number combinations and is a standard part of the address of virtually every individual and organization. Each postal code has a unique combination of four digits and two letters. Postal code information systems (Geo-system and Mosaïc) collect information on a variety of characteristics at the postal code level. In the Netherlands, information at the postal code level comprises on average 16 households or firms. The information systems can, for example, be added to an internal list, such that the organization is able to create a customer profile, which can be helpful for further direct marketing activities (e.g. Lix et al. 1995).

The information supplied by Geo-Marktprofiel is collected by specially developed, individualized single-source research on all postal code areas. This information is linked with information contained in the postal address list owned by the Dutch Postal Services. The last information is updated every four months, while every year information of about 50,000 individuals, gathered through questionnaires, is updated. The system contains over 80 neighborhood characteristics such as average income, family composition, urbanization, access to shops, school and sport facilities and house moving frequency. Based on this information, Geo-Marktprofiel classified the postal codes into 336 classes according to the dimensions 1) average income (six levels), 2) family composition and life cycle (seven levels) and 3) urbanization (eight levels).

The Mosaïc system describes the postal codes by characteristics such as: types of homes, composition of the family, occupation, education, average income, social class, agricultural employment, the percentage of working woman, religion and various other characteristics. The Mosaic system is made
up of three separate information systems. First, the internal list of Wehkamp (the largest catalog retailer in the Netherlands). Second, the Wehkamp list is linked with the postal address list owned by the Dutch Postal Services. Third, individually specific information on 800,000 households is used, obtained by a written questionnaire. The three sources are updated every year.

Although individuals in the same postal code area do not possess the same characteristics, it is assumed that those individuals have a certain level of homogeneity. According to Baier (1985), people with comparable interests tend to cluster geographically. Furthermore, their purchase decisions are frequently influenced by their desire to emulate friends and neighbors. Therefore, postal codes provide the means to identify clusters of individuals that have a certain degree of similarity in purchase behavior. Postal code information systems rely on the principles of reference group theory as well as on the concept of environmental influences on buying behavior (Baier 1985).

5.3.2 Profits and information

A special case of the model (5.1) is when the regressor is not observed at the individual level. For example, information about income may not be available at the individual level, so $\xi_i$ is not observable, but a proxy $z_i$ is available at some aggregate level, i.e., for the postal code area to which household $i$ belongs. The income of an individual in a specific postal code area is then the average income of that postal code area plus a deviation because obviously not all the individuals in that area have exactly the same income. We call this grouping on the basis of external information.

Consider the model with one regressor, i.e.,

$$\eta_i = \alpha + \xi_i \beta + \epsilon_i,$$

(5.13)

with $\xi$, a scalar, which we do not observe. Please recall that we normalize the variance of $\epsilon$, hence, $\epsilon_i \sim N(0, 1)$. We observe $z_i$, given by

$$\xi_i = z_i + \omega_i,$$

(5.14)

where we assume that $z_i \sim N(0, \sigma_z^2)$ and $\omega_i \sim N(0, \sigma_\omega^2); z_i$ and $\omega_i$ are i.i.d. and mutually independent, and independent from $\epsilon_i$. Note that, although suppressed by the notation, $z_i$ is the average of a group of individuals. Substituting (5.14) into (5.13) yields

$$\eta_i = \alpha + z_i \beta + \epsilon_i,$$

(5.15)
where \( e_i = \varepsilon_i + \omega_i \beta \). It is clear that \( E(\varepsilon_i e_i) = 0 \). The variance of \( e_i \) is:

\[
\sigma_e^2 = E(e_i + \omega_i \beta)^2 = 1 + \sigma_\omega^2 \beta^2. \tag{5.16}
\]

Dividing (5.15) by \( \sigma_e \), we obtain

\[
\frac{\eta_i}{\sigma_e} = \frac{\alpha}{\sigma_e} + z_i \frac{\beta}{\sigma_e} + \frac{\varepsilon_i}{\sigma_e} = \tilde{\alpha} + z_i \tilde{\beta} + \tilde{\varepsilon}_i, \tag{5.17}
\]

where \( \tilde{\beta}, \tilde{\alpha}, \) and \( \tilde{\varepsilon}_i \) are implicitly defined. We now have a model that is fully analogous to (5.13): on the left-hand side we have an unobservable variable of which we only observe the sign (\( \eta_i/\sigma_e \) has the same sign as \( \eta_i \)); on the right-hand side we have a regressor (now \( z_i \)) and a disturbance that, by construction, has unit variance and is uncorrelated with the regressor.

Assume that a direct marketing organization neglects the noise (5.14) and just uses \( z_i \) as the regressor in the analysis, obtaining consistent estimates for \( \tilde{\alpha} \) and \( \tilde{\beta} \) (cf. Yatchew and Griliches 1985). The consequence is that the optimal profits are reduced. This is not only due to the use of \( \tilde{\alpha} \) and \( \tilde{\beta} \) instead of \( \alpha \) and \( \beta \), but also due to the fact that the ordering of the population by \( z_i \) is different from the one by \( \xi_i \). Hence, two effects of imperfect information can be distinguished: the selected fraction is not optimal and the population is incorrectly ordered. These effects are shown above in figure 5.5. The upper axis gives the ordering of four individuals based on \( \xi_i \) and the optimal cutoff point \( \xi_i \). Because \( \xi_3 \) and \( \xi_4 \) are larger than \( \xi_i \), individuals 3 and 4 should be selected. The middle and the lower axes give the ordering of the four individuals using the proxy \( z_i \). If the same cutoff point is used only individual 2 would be selected, instead of 3 and 4. However, if the cutoff point \( z_i \) is adopted, individuals 1, 2 and 4 would be selected.

### 5.3.3 Effects of using proxy information

It is straightforward to find the relevant expression for the fraction of individuals when the proxy is used. We define the optimal cutoff point, analogous to (5.11), as

\[
z_c = \frac{\tau - \tilde{\alpha}}{\tilde{\beta}}, \tag{5.18}
\]

indicating that all individuals with \( z_i \geq z_c \) are selected. Note that (5.18) reduces to (5.11) if there is no noise. Then \( \sigma_\omega^2 = 0, \sigma_\varepsilon^2 = 1, \) and thus \( (\tilde{\alpha}, \tilde{\beta}) = (\alpha, \beta) \).
The expected profits deriving from this strategy should of course still be based on the $\xi$; this is given by

$$E \left( a \Phi(\alpha + \xi \beta) - c \mid z \geq z_c \right).$$

In section 5.5 we prove that

$$E \left( a \Phi(\alpha + \xi \beta) - c \mid z \geq z_c \right) = E \left( a \Phi(\tilde{\alpha} + z \tilde{\beta}) - c \mid z \geq z_c \right). \quad (5.19)$$

This result indicates that there is no bias in the expected profits, if they are determined on the proxy information and the estimates for $\tilde{\alpha}$ and $\tilde{\beta}$ instead of $\alpha$ and $\beta$. This is an interesting result: even though the $\xi_i$s are unobservable for the direct marketing organization, which therefore does not know $\alpha$ and $\beta$, it can still make a correct appraisal of the expected profits using the imperfect information. Thus, the organization can make an unbiased decision whether the external information is worth its costs. From (5.19) an expression for the expected profits follows directly:

$$E \left( a \Phi(\alpha + \xi \beta) - c \mid z \geq z_c \right) = a \int_{z_c}^{\infty} \Phi(\tilde{\alpha} + z \tilde{\beta}) \phi \left( \frac{z}{\sigma} \right) \frac{1}{\sigma} \, dz - c. \quad (5.20)$$

This clearly reduces to (5.12) if $z = \xi$. We illustrate the relation between the optimally obtainable profits on the one hand and the quality of the information...
of the individuals on the other hand graphically. In figure 5.6 we depict (5.20), interpreted as a function of $\sigma_z^2$. For simplicity of presentation we use the *intra-class correlation*, $\rho$, defined by

$$\rho \equiv \frac{\sigma_z^2}{\sigma_x^2},$$

rather than $\sigma_z^2$ itself, since $\rho$ is restricted to the $[0, 1]$-interval. The figure, based on the choice $\beta = 2, \sigma_x^2 = 1, a = 10$ and $c = 1$, presents the expected profits *per selected individual*, for four values of $\alpha$. It shows that the expected profits are an increasing function of the intra-class correlation, as they should be.

Figure 5.7 shows how the fraction of selected individuals, i.e. $P(z \geq z_c)$, depends on $\rho$. On combining figure 5.6 and figure 5.7 we obtain the graphs giving expected profits for all individuals on the list (see figure 5.8). This is (5.20) multiplied by $P(z \geq z_c)$. Also here, of course, we see that expected profits increase with the reliability of the proxy.

The interesting feature of figure 5.8, is the behavior at $\rho = 0$ (i.e. $\sigma_z^2 = 0$ or $z_i = 0 \forall i$). Then the importance of the proxy variable vanishes, i.e. it has no predictive power anymore. This means that it is no longer helpful to use
Figure 5.7: Fraction of selected individuals, i.e. $P(z \geq z_c)$

Figure 5.8: Expected profits per individual on the mailing list
for the selection of individuals. Hence the direct marketing organization is confronted with the choice between mailing to all or none of the individuals; this decision depends, of course, on $\alpha$. There are two ways to look at the behavior at $\rho = 0$. First, by using (5.17). As $\sigma_u^2 = 0$ it follows that $\sigma_u^2 = \sigma_e^2$, so $\sigma_w^2 = 1 + \sigma_e^2\beta^2$. Hence, $\tilde{\alpha} = \alpha/\sqrt{1 + \sigma_e^2\beta^2}$, and the organization should send all individuals a mailing when $a\Phi(\tilde{\alpha}) \geq c$. Second, by notifying that we actually have a situation without a regressor. Hence, the organization uses the mean expected response probability, which is obtained by integrating over $\xi$ (cf. result (4.9) in chapter 4),

$$
P(\eta > 0) = \int_{-\infty}^{\infty} \Phi(\alpha + \xi\beta)\Phi\left(\frac{\xi}{\sigma_\xi}\right) \frac{1}{\sigma_\xi} \, d\xi
= \Phi\left(\frac{\alpha}{\sqrt{1 + \sigma_e^2\beta^2}}\right)
= \Phi(\tilde{\alpha}).
$$

For one particular value of $\alpha$ the organization is indifferent between mailing all or none of the individuals. This occurs if $a\Phi(\tilde{\alpha}) - c = 0$, i.e. $\alpha = \tau\sqrt{1 + \sigma_e^2\beta^2} \approx -2.87$, for the parameter setting underlying the graphs.

A glance at these figures makes clear that they are similar to those in section 5.2. This is not surprising since we can express the $R^2$ of (5.15) as a function of $\rho$ times a constant, i.e.

$$
R^2 = \frac{\sigma_e^2\beta^2}{\sigma_\eta^2} = \rho \frac{\sigma_e^2\beta^2}{\sigma_\eta^2}.
$$

Note that the constant is the $R^2$ of model (5.13), which is 0.8 for the given parameter setting. Hence, figures 5.6 to 5.8 also give the relation between $R^2$ and expected profit, when the $x$-axes are rescaled from 0 to 0.8; the additional $x$-axes show this.

We are now in a position to discuss the use of these graphs. Assume that $z$ has already been observed. A test mailing (the cost of which we neglect for the sake of simplicity) is held to estimate consistently the values of $z_c$, $\tilde{\alpha}$, $\tilde{\beta}$ and $\sigma_e^2$. Now information is offered with intra-class correlation $\rho_1$. In order to assess the value of information the organization needs to have an idea of the reliability of the information it already has; let this be denoted by $\rho_0$. Combining this with
the values of $\tilde{\alpha}$, $\tilde{\beta}$ and $\sigma^2_\varepsilon$ as computed from the test mailing gives us values of $\alpha$, $\beta$ and $\sigma^2_w$. It holds that

$$\tilde{\alpha} = \frac{\alpha}{\sigma_e} \quad (5.22)$$

$$\tilde{\beta} = \frac{\beta}{\sigma_e} \quad (5.23)$$

$$\sigma^2_\varepsilon = \sigma^2_\varepsilon - \sigma^2_w \quad (5.24)$$

$$\rho_0 = \frac{\sigma^2}{\sigma^2_\varepsilon} \quad (5.25)$$

where (5.22) and (5.23) follow from (5.17), $\sigma_e$ is given by (5.16), and (5.24) implicitly follows from (5.14). The system (5.22)–(5.24) can be interpreted as a set of three equations in four unknown parameters ($\alpha$, $\beta$, $\sigma^2_\varepsilon$, $\sigma^2_w$). With an idea $\rho_0$ on $\rho$ added as in (5.25), however, the system can be solved to yield

$$\sigma^2_\varepsilon = \frac{\sigma^2}{\rho_0}$$

$$\sigma^2_w = \sigma^2_\varepsilon \left( \frac{1 - \rho_0}{\rho_0} \right)$$

$$\alpha = \frac{\tilde{\alpha}}{\sqrt{1 - \sigma^2_w \tilde{\beta}^2}}$$

$$\beta = \frac{\tilde{\beta}}{\sqrt{1 - \sigma^2_w \tilde{\beta}^2}}.$$  

Given this solution the relevant graph in the sense of figure 5.8 can be drawn. Information with intra-class correlation $\rho_i$ is worthwhile if the expected profit at $\rho_i$ minus the expected profit at $\rho_0$ exceeds the purchase cost of information per individual on the mailing list.

5.3.4 Application

In order to illustrate the concepts introduced above, we discuss an application based on synthetic data. We consider model (5.13) with $\alpha = -1.5$, $\beta = 1$, $\xi_i \sim N(0, 0.25)$, and $\varepsilon_i \sim N(0, 1)$. Hence, the $R^2$ of the underlying model is 0.2. We choose $\alpha = 25$ and $\varepsilon = 2$. We generate the data in such a way that, apart from household information, we have two levels of grouping. At the
Table 5.1: Estimates of probit model and optimal cutoff point

<table>
<thead>
<tr>
<th>variable</th>
<th>coefficient</th>
<th>estimate</th>
<th>theoretical</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi$</td>
<td>$\alpha$</td>
<td>-1.526</td>
<td>-1.500</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>0.979</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>$\xi_1$</td>
<td>0.123</td>
<td></td>
</tr>
<tr>
<td>$z_1$</td>
<td>$\alpha$</td>
<td>-1.450</td>
<td>-1.420</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>0.917</td>
<td>0.947</td>
</tr>
<tr>
<td></td>
<td>$z_c$</td>
<td>0.048</td>
<td></td>
</tr>
<tr>
<td>$z_2$</td>
<td>$\alpha$</td>
<td>-1.399</td>
<td>-1.367</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>0.900</td>
<td>0.911</td>
</tr>
<tr>
<td></td>
<td>$z_c$</td>
<td>-0.006</td>
<td></td>
</tr>
</tbody>
</table>

$^1$ Asymptotic standard error in parentheses

first level the average is determined on the basis of 16 observations and at the second level of 256 observations. We generate 50 176 observations (to have balanced groups), of which 5 000 are used for estimation.

A certain postal code area, with 16 households, has a specific mean, which we call $\mu_1$. We have 50176/16=3136 different values of $\mu_1$. A group of 16 postal code areas (i.e. 256 households) has mean $\mu_2$ of which we have 3136/16=196 different values. To generate the data we start with drawing 196 values of $\mu_2 \sim N(0, 0.3)$. For each $\mu_2$ we draw 16 values of $\mu_1$, $\mu_1 \sim N(\mu_2, 0.3)$. Finally, we draw 16 values, $\xi_i \sim N(\mu_1, 1)$, for each $\mu_1$; these are the household data. The variances of $\xi$, $\mu_1$ and $\mu_2$ are chosen in such a way that the intra-class correlations have the reasonable values of 0.5 and 0.15, respectively. The data have to be transformed such that $\sigma_2^2 = 0.25$. The values of $\xi_i$ are used to determine the sample means for the two levels of grouping: $z_{i1}$ for grouping over 16 households and $z_{i2}$ for grouping over 256 households. These sample means are used for estimation.

We follow the method described in this section. The estimated intra-class correlation coefficients are 0.536 and 0.182, respectively. First, we estimate the coefficients by ML using $\xi$, $z_1$ and $z_2$; table 5.1 gives the results. The theoretical results are based on the asymptotic results of (5.17). The coefficients decrease as we go from $\xi$ to $z_1$ and $z_2$. Note that the theoretical values of the coefficients are within two times the standard error. The standard errors of $\alpha$ decrease and
Table 5.2: Profits and the probability of being selected

<table>
<thead>
<tr>
<th>variable</th>
<th>method I</th>
<th>method II</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \xi )</td>
<td>profits 39 471</td>
<td>37 538</td>
</tr>
<tr>
<td>( \xi )</td>
<td>( P(\xi \geq \xi_1) ) 0.40</td>
<td>0.39</td>
</tr>
<tr>
<td>( z_1 )</td>
<td>profits 29 662</td>
<td>27 497</td>
</tr>
<tr>
<td>( z_1 )</td>
<td>( P(z \geq z_1) ) 0.45</td>
<td>0.43</td>
</tr>
<tr>
<td>( z_2 )</td>
<td>profits 18 876</td>
<td>17 423</td>
</tr>
<tr>
<td>( z_2 )</td>
<td>( P(z \geq z_2) ) 0.51</td>
<td>0.52</td>
</tr>
<tr>
<td>no info.</td>
<td>profits 8 648</td>
<td></td>
</tr>
<tr>
<td>no info.</td>
<td>( P(\text{selection}) ) 1</td>
<td></td>
</tr>
</tbody>
</table>

1 Based on (5.20)

those of \( \beta \) increase from \( \xi \) to \( z_1 \) and \( z_2 \). This reflects the uncertainty of the effect of the observed variable on \( \eta \).

We determine the profits (for the whole mailing list) for the three types of information in two ways. The easiest way is to use the optimal cutoff point to determine the households that would have been selected, and count the households who actually responded. That is,

\[
\sum_{i=1}^{N} (a_i y_i - c) I(\xi_i \geq \xi_1),
\]

where \( I(\cdot) \) is an indicator function which is one if the argument is true and zero otherwise. We call this method I. Note that this gives the actual profits of the \( N \) households. The other way to determine the profits, method II, is by computing (5.20) and multiplying this by the number of selected households; this gives the expected profits. Table 5.2 gives the results of both methods. The case of using no information at all is presented as a bench mark. The table shows that profits increase and the fraction of selected households decreases with the intra-class correlation (cf. figures 5.6 and 5.7).

In order to examine our approach in somewhat more detail we consider the case that the organization has observed \( z_2 \) and knows the intra-class correlation of this information. Then, a mailing is held to estimate the values of \( (\tilde{\alpha}, \tilde{\beta}, z_c, \sigma^2) \) consistently. These values, combined with the intra-class correlation, give the values of \( (\alpha, \beta, \sigma^2_\alpha, \sigma^2_\beta) \). These values enable us to determine the
Table 5.3: Break-even price (per household) for buying better information. Entry \((i, j)\) denotes the break-even price between information on level \(i\) and information on level \(j\).

<table>
<thead>
<tr>
<th>Entry</th>
<th>(\xi)</th>
<th>(z_1)</th>
<th>(z_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(z_1)</td>
<td>0.195</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(z_2)</td>
<td>0.410</td>
<td>0.215</td>
<td></td>
</tr>
<tr>
<td>no info</td>
<td>0.614</td>
<td>0.419</td>
<td>0.204</td>
</tr>
</tbody>
</table>

expected profits for every value of \(\rho\). The expected profits for the intra-class correlation for \(z_1\) and \(\xi\), i.e. \(\rho = 0.536\) and \(\rho = 1\), are 27 803 and 37 139 respectively. These values are close to the expected profits given in the last column of table 5.2. Hence, if the organization knows the intra-class correlations of the observed and the new data, the proposed method is indeed able to determine the expected profits.

The bottom line is to attach a monetary value to information. In other words, what is an organization willing to pay for better information? We determine the price at which an organization is indifferent between buying and not buying better information; we called this the break-even price. Table 5.3 presents the break-even prices. For example, if the organization has information on level \(z_2\) and it can buy information on level \(z_1\), it is willing to pay 0.215 per household at most.

### 5.4 Measurement error

An other form of proxy information is when the regressor of interest, \(\xi\), is not measured exactly but an indicator is used instead,

\[
x_i = \xi_i + v_i,
\]

where \(x_i\) is the indicator and \(v_i\) is the noise when using \(x_i\) instead of \(\xi_i\), assumed to be independently normally distributed, \(v_i \sim N(0, \sigma^2_i)\); \(v_j\) is assumed to be independently distributed from \(v_j, j \neq i\), from \(\varepsilon\) and from \(\xi\). This is the so-called measurement error model (e.g. Judge et al. 1985, chapter 17). We still consider the case with one explanatory variable. Note that the observed value of the proxy variable \(x_i\) is in principle different for each individual, unlike
the case of grouping where the observed value of the proxy variable \( z_i \) is similar for a group of individuals. Moreover, here the proxy variable \( x_i \) is not independently distributed from the error term \( \epsilon_i \), whereas the proxy variable \( z_i \) is independently distributed of the error term \( \epsilon_i \) in the case of grouping.

From (5.26) we have

\[
x_i \sim N(0, \sigma_z^2) = N(0, \sigma_x^2 + \sigma_c^2).
\]

Then

\[
q \equiv \frac{\sigma_x^2}{\sigma_z^2} = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_c^2},
\]

is defined as the reliability of \( x \) as a proxy for \( \xi \). We now reformulate (5.13) into a form that is based on \( x \) rather than \( \xi \):

\[
\eta_i = \alpha + \xi_i \beta + \epsilon_i
\]

\[
= \alpha + x_i \beta + \{ \epsilon_i - \nu_i \beta \}
\]

\[
= \alpha + x_i \frac{\sigma_x^2}{\sigma_z^2} \beta + \left\{ \epsilon_i - \nu_i \beta + x_i \beta - \frac{\sigma_x^2}{\sigma_z^2} \beta \right\}
\]

\[
= \alpha + x_i \frac{\sigma_x^2}{\sigma_z^2} \beta + \left\{ \epsilon_i - \nu_i \beta + x_i \frac{\sigma_x^2}{\sigma_z^2} \beta \right\}
\]

\[
= \alpha + x_i \frac{\sigma_x^2}{\sigma_z^2} \beta + u_i,
\]

(5.27)

where \( u_i \) is implicitly defined, and has the property

\[
E(x_i u_i) = E \left\{ (\xi_i + \nu_i) (\epsilon_i - \nu_i \beta + (\xi_i + \nu_i) \frac{\sigma_x^2}{\sigma_z^2} \beta) \right\}
\]

\[
= \sigma_x^2 \frac{\sigma_x^2}{\sigma_z^2} \beta - \sigma_c^2 \beta + \frac{\sigma_x^4}{\sigma_z^4} \beta
\]

\[
= 0.
\]

For the variance of \( u_i \) there holds

\[
\sigma_u^2 = E \left\{ \epsilon_i - \nu_i \beta + x_i \frac{\sigma_x^2}{\sigma_z^2} \beta \right\}^2
\]

\[
= 1 + \beta^2 \sigma_x^2 + \beta^2 \frac{\sigma_x^4}{\sigma_z^4} - 2 \beta^2 \frac{\sigma_x^4}{\sigma_z^4}.
\]
\[
\begin{align*}
1 + \beta^2 & \left( \frac{\sigma_x^2}{\sigma_x^4} \right) \\
& = 1 + \beta^2 \frac{\sigma_x^2 \sigma_c^2}{\sigma_x^4}.
\end{align*}
\]

(5.29)

Note that \(\frac{\sigma_x^2}{\sigma_x^4}\) is the probability limit of the OLS estimator of the regression of \(\eta\) on \(x\). Now we divide (5.28) by \(\sigma_u\) to obtain

\[
\frac{\eta_i}{\sigma_u} = \frac{\alpha}{\sigma_u} + x_i \frac{\sigma_x^2 \beta}{\sigma_u} + \frac{u_i}{\sigma_u}
\]

\[
= \alpha^* + x_i \beta^* + u_i^*,
\]

(5.30)

where \(\beta^*, \alpha^*\) and \(u_i^*\) are implicitly defined. Again we have a model that is fully analogous to (5.13): on the left-hand side we have a variable of which we only observe the sign; on the right-hand side we have a regressor (now \(x_i\)) and a disturbance that, by construction, has unit variance and is uncorrelated with the regressor. As before, the direct marketing organization neglects the noise (5.26) and just uses \(x_i\) as the regressor in the analysis. Equation (5.28) shows the implication: because \(u_i\) (hence \(u_i^*\)) and \(x_i\) are uncorrelated, the analysis yields a consistent estimator of \(\alpha^*\) and \(\beta^*\), but not for \(\alpha\) and \(\beta\). Still neglecting the difference between parameters and their consistent estimators, we investigate the implications for expected profits when \(\alpha^*\) and \(\beta^*\) are used instead of \(\alpha\) and \(\beta\). Note that this is not a matter of choice on the organization’s behalf since \(\beta\) is not identified in a linear latent variable model where all variables are normally distributed (cf. Aigner et al. 1984).

The relevant expressions are analogous to (5.11) and (5.12). The cutoff point of \(x, x_c\), is defined as

\[
x_c = \frac{\tau - \alpha^*}{\beta^*},
\]

(5.31)

which reduces to (5.11) if there is no noise, \(\sigma_x^2 = 0\). Hence, the organization should select individuals with \(x_i \geq x_c\). Expected profits deriving from this strategy should of course still be based on the \(\xi\); this is given by

\[
\mathbb{E} \left( a \Phi (\alpha + \xi \beta) - c \mid x \geq x_i \right),
\]

instead of the left hand side of (5.12). In section 5.5 we prove that

\[
\mathbb{E} \left( a \Phi (\alpha + \xi \beta) - c \mid x \geq x_i \right) = \mathbb{E} \left( a \Phi (\alpha^* + x \beta^*) - c \mid x \geq x_i \right).
\]

(5.32)
This result indicates that there is no bias in the expected profits, if they are determined on the proxy information and the inconsistent estimates. Hence, even though \((\xi, \beta, \alpha)\) is unknown to the direct marketing organization, it can still make a correct appraisal of the expected profits. From (5.32) an expression for the expected profits follows directly:

\[
E \left( a \Phi(\alpha + \xi \beta) - c \mid x \geq x_c \right) = a \frac{\int_{x_c}^{\infty} \Phi(\alpha^* + x \beta^*) \phi \left( \frac{x}{\sigma_\xi} \right) \frac{1}{\sigma_\xi} \, dx}{1 - \Phi \left( \frac{x_c}{\sigma_\xi} \right)} - c,
\]

which equals (5.12) when \(x = \xi\).

The relation between the optimally obtainable profits and the quality of the information on the individuals can be depicted in figures similar to figures 5.6, 5.7 and 5.8. Here we summarized the quality of information in terms of intra-class correlation. Now we do so using the reliability (5.27). In section 5.5 we prove that, if the underlying distributions are normal, exactly the same figures are obtained. That is, if \(\rho = q\), the expected profits and the percentage of selected individuals have the same value for both forms of imperfect information. The intuition behind this result is that the explanatory power of the two specifications is equal if \(\rho = q\). This is obvious when we express the \(R^2\) as a function of \(q\):

\[
R^2 = \frac{\sigma_\xi^2 \sigma^4}{\sigma^2 \sigma^2} = q \frac{\sigma_\xi^2 \beta^2}{\sigma^2},
\]

analogous to (5.21).

Hence, the two types of imperfect information, grouping and measurement error, are mathematically largely comparable. It should be realized, however, that the assumption underlying the structure of the data is completely different. That is, in the case of measurement error, the data generating process is fully specified, in contrast with the case of grouping. Loosely speaking, it shows the difference between an econometric specification and a statistical specification.

5.5 Proofs

In this section we prove (5.19) and (5.32).

**Proof of (5.19)** Consider model (5.13) with \(\xi = z_i + \omega_i\). Write the response probability as

\[
P(y_i = 1 \mid z_i \geq z_c)
\]
On the value of information

\begin{align*}
&= P(\eta_i > 0 \mid z_i \geq z_e) \\
&= P(\alpha + \xi_j \beta + \epsilon_i > 0 \mid z_i \geq z_e) \\
&= P(\alpha + (z_i + \omega_i) \beta + \epsilon_i > 0 \mid z_i \geq z_e) \\
&= P(\epsilon_i + \omega_i \beta > -\alpha - z_i \beta \mid z_i \geq z_e) \\
&= P(\tilde{\epsilon}_i > -\tilde{\alpha} - z_i \tilde{\beta} \mid z_i \geq z_e). \quad (5.33)
\end{align*}

The important aspect is the equality between (5.33) and (5.34). Evaluating these probabilities enables us to prove (5.19). That is,

\begin{align*}
P(\tilde{\epsilon} > -\tilde{\alpha} - z \tilde{\beta} \mid z \geq z_e)P(z \geq z_e) &= P(\tilde{\epsilon} + z \tilde{\beta} > -\tilde{\alpha} & z \geq z_e) \\
&= \int_{z_e}^{\infty} \int_{-\tilde{\alpha} - z \tilde{\beta}}^{\infty} f_{\tilde{\epsilon},z}(s, t) \, ds \, dt \\
&= \int_{z_e}^{\infty} \int_{-\tilde{\alpha} - z \tilde{\beta}}^{\infty} f_{\tilde{\epsilon}}(s) f_z(t) \, ds \, dt \\
&= \int_{z_e}^{\infty} P(\tilde{\epsilon} \geq -\tilde{\alpha} - t \tilde{\beta}) f_z(t) \, dt \\
&= E(P(\tilde{\epsilon} > -\tilde{\alpha} - z \tilde{\beta}) \mid z \geq z_e)P(z \geq z_e) \\
&= E \left( \Phi(\tilde{\alpha} + z \tilde{\beta}) \mid z \geq z_e \right) P(z \geq z_e), \quad (5.34)
\end{align*}

where the third step uses the independence between \( \tilde{\epsilon} \) and \( z \). Moreover,

\begin{align*}
P(\epsilon > -\alpha - \xi \beta \mid z \geq z_e)P(z \geq z_e) &= P(\epsilon + z \beta + \omega \beta > -\alpha \& z \geq z_e) \\
&= \int_{z_e}^{\infty} \int_{-\alpha - z \beta}^{\infty} \int_{-(\alpha + t \beta)}^{\infty} f_{\epsilon,\alpha,z}(s, k, t) \, ds \, dk \, dt \\
&= \int_{z_e}^{\infty} \int_{-\alpha - z \beta}^{\infty} \int_{-(\alpha + t \beta)}^{\infty} f_{\epsilon,\alpha}(s, k) f_z(t) \, ds \, dk \, dt \\
&= \int_{z_e}^{\infty} P(\epsilon > -\alpha - (\omega + t) \beta \& -\infty < \omega < \infty) f_z(t) \, dt \\
&= \int_{z_e}^{\infty} P(\epsilon > -\alpha - \xi \beta) f_z(t) \, dt \\
&= E \left( \Phi(\alpha + \xi \beta) \mid z \geq z_e \right) P(z \geq z_e), \quad (5.38)
\end{align*}
where the fourth step uses the independence between \( e, \omega \) and \( z \). Since (5.35) equals (5.37), and (5.36) equals (5.38), it follows that

\[
E \left( a \Phi (\alpha + \xi \beta) - c \mid z \geq z_c \right) = E \left( a \Phi (\bar{\alpha} + z \bar{\beta}) - c \mid z \geq z_c \right),
\]

which proves (5.19).

**Proof of (5.32)** This proof is analogous to the proof of (5.19). Consider model (5.13) with \( x_i = \xi_i + \nu_i \). Write the response probability, using (5.28) and (5.30), as

\[
P(y_i = 1 \mid x_i \geq x_c) = P(\eta_i > 0 \mid x_i \geq x_c)
= P(\alpha + \xi_i \beta + \epsilon_i > 0 \mid x_i \geq x_c)
= P \left( \frac{u_i}{\sigma_u} > -\frac{\alpha}{\sigma_u} - x_i \frac{\sigma_x^2}{\sigma_u^2} \beta \mid x_i \geq x_c \right)
= P(u^*_i > -\alpha^* - x_i \beta^* \mid x_i \geq x_c)
\]

where the latter step holds since \( u^*_i \) and \( x_i \) are, by construction, mutually independent. Again, we evaluate these probabilities. Using the independence between \( u^* \) and \( x \), it is clear, analogous to the equality between (5.35) and (5.36), that

\[
P(u^* + x \beta^* > -\alpha^* \mid x \geq x_c)P(x \geq x_c) = E \left( \Phi(\alpha^* + x \beta^*) \mid x \geq x_c \right) P(x \geq x_c).
\]

Furthermore,

\[
P(e + \xi \beta > -\alpha \mid x \geq x_c)P(x \geq x_c)
= P(e + x \beta - \nu \beta > -\alpha \mid x \geq x_c)
= \int_{x_c}^{\infty} \int_{-\infty}^{\infty} \int_{-\alpha + (k - 1)\beta}^{\infty} f_{\epsilon, \nu, \beta}(s, k, t) \ ds \ dk \ dt
= \int_{x_c}^{\infty} \int_{-\alpha + (k - 1)\beta}^{\infty} f_{\epsilon, \nu, \beta}(s, k \mid x = t) f_s(t) \ ds \ dk \ dt
= \int_{x_c}^{\infty} \Phi(\epsilon > -\alpha - (x - \nu \beta \mid x = t) f_s(t) \ dt
= \int_{x_c}^{\infty} \Phi(\alpha + \xi \beta) f_s(t) \ dt
= E \left( \Phi(\alpha + \xi \beta) \mid x \geq x_c \right) P(x \geq x_c),
\]
which, using (5.41), and the equality between (5.39) and (5.40), implicitly proves (5.32).

In words the proofs can be sketched as follows. The probabilities of interest ((5.33) and (5.39)) are conditioned on variables that do not appear explicitly in argument \( z \) and \( x \) respectively. However, we can rewrite the argument in terms of either \( z \) or \( x \). Consequently, we have a probability that depends on two independent variables (by construction in the case of \( x \)) conditional on \( z \) or \( x \). Finally, we express the conditional probability as a conditional expectation. Since the parameters are identified up to scale, we can write the probability as a standard normal distribution function.

As indicated in section 5.3 and 5.4, the results imply that even though the direct marketing organization cannot consistently estimate the parameters and does not observe the variable of interest, it can still make a correct appraisal of the expected profits for the whole range of \( p \) and \( q \).

Now we prove that if \( \rho = q \),

\[
E \left( \Phi(\alpha + z\beta) \mid z \geq z_c \right) = E \left( \Phi(\alpha^* + x\beta^*) \mid x \geq x_c \right).
\]

(5.42)

First we prove the following three auxiliary results:

\[
\frac{\sigma^2_x}{\sigma^2_z} f_{\text{obj}} \left( k - \frac{\sigma^2_z}{\sigma^2_x} t \mid \xi = k \right) = f_x(t - k). \quad (5.43)
\]

\[
\frac{\sigma^2_x}{\sigma^2_z} x_c = z_c. \quad (5.44)
\]

\[
P(x \geq x_c) = P(z \geq z_c). \quad (5.45)
\]

Proof of (5.43)

\[
\omega \mid \xi = k \sim N \left( \frac{\sigma^2_\omega}{\sigma^2_x} k, \frac{\sigma^4_\omega}{\sigma^2_x} \right) = N \left( \frac{\sigma^2_\omega}{\sigma^2_x} k, \frac{\sigma^2_\omega^2}{\sigma^2_x} \right),
\]

since, using \( \sigma^2_x/\sigma^2_x = \sigma^2_x/\sigma^2_x \), i.e. \( q = \rho \),

\[
\frac{\sigma^2_\omega}{\sigma^2_x} - \frac{\sigma^4_\omega}{\sigma^4_x} = \frac{\sigma^2_\omega}{\sigma^2_x} (\sigma^2_x - \sigma^2_x) = \frac{\sigma^2_\omega^2}{\sigma^2_x} = \frac{\sigma^2_\omega^2}{\sigma^2_x}.
\]

so

\[
\frac{\sigma^2_x}{\sigma^2_z} f_{\text{obj}} \left( k - \frac{\sigma^2_z}{\sigma^2_x} t \mid \xi = k \right)
\]
\[
\begin{align*}
\text{Proofs 107} & \\
\frac{1}{\sqrt{2\pi}} \frac{\sigma_z}{\sigma_z \sigma_w} & \exp\left( -\frac{1}{2} \frac{\sigma_z^2}{\sigma_z^2 \sigma_w} \left( \frac{\sigma_z^2}{\sigma_z^2 \sigma_w} - \frac{\sigma_w^2 k}{\sigma_z^2} \right)^2 \right) \\
= & \frac{1}{\sqrt{2\pi}} \frac{\sigma_z}{\sigma_z \sigma_w} \exp\left( -\frac{1}{2} \frac{\sigma_z^2}{\sigma_z^2 \sigma_w} \left( \frac{\sigma_z^2}{\sigma_z^2 \sigma_w} - \frac{\sigma_z^2 (t - k)}{\sigma_z^2} \right)^2 \right) \\
= & \frac{1}{\sqrt{2\pi}} \frac{\sigma_z}{\sigma_z \sigma_w} \exp\left( -\frac{1}{2} \frac{(t - k)^2}{\sigma_z^2} \right) \\
= & f_x(t - k),
\end{align*}
\]

where we use
\[
1 - \frac{\sigma_w^2}{\sigma_z^2} = \frac{\sigma_z^2}{\sigma_z^2} = \frac{\sigma_z^2}{\sigma_z^2},
\]

and
\[
\frac{\sigma_z^2}{\sigma_z^2 \sigma_w^2} \frac{\sigma_z^4}{\sigma_z^4 \sigma_w^4} = \frac{\sigma_z^2}{\sigma_z^2 \sigma_w^2} = \frac{1}{\sigma_z^2},
\]

in the second and third step respectively. This proves (5.43).

**Proof of (5.44)**

\[
\begin{align*}
\frac{\sigma_z^2}{\sigma_z^2} x_c & = z_c \\
\frac{\sigma_z^2}{\sigma_z^2} (\Phi^{-1}(c/a) - \alpha^*) & = \Phi^{-1}(c/a) - \tilde{\alpha} \iff \\
\frac{\sigma_z^2}{\sigma_z^2} \beta^* & = \frac{\Phi^{-1}(c/a) - \tilde{\alpha}}{\beta}.
\end{align*}
\]

using (5.31) and (5.18). Note that the numerators are equal if \(\alpha^* = \tilde{\alpha}\), i.e. \(\alpha / \sigma_u = \alpha / \sigma_\epsilon\), and this holds if \(\sigma_u = \sigma_\epsilon\). The same argument holds for the denominators:

\[
\sigma_z^2 \beta^* = \sigma_z^2 \tilde{\beta} \Rightarrow \sigma_z^2 \beta / \sigma_u = \sigma_z^2 \tilde{\beta} / \sigma_\epsilon.
\]

Note that \(\sigma_\epsilon = \sigma_u \iff 1 + \beta^2 \sigma_\epsilon^2 / \sigma_z^2 = 1 + \beta^2 \sigma_u^2 / \sigma_z^2\), i.e. \(\sigma_z^2 \sigma_\epsilon^2 / \sigma_z^2 = \sigma_u^2\), where the latter is obtained from (5.46); this proves (5.44).

**Proof of (5.45)**

\[
P(x \geq x_c) = 1 - \Phi \left( \frac{x_c}{\sigma_z} \right) = 1 - \Phi \left( \frac{\sigma_z^2}{\sigma_z^2} x_c \right) = 1 - \Phi \left( \frac{1}{\sigma_z} z_c \right) = P(z \geq z_c),
\]
where we use (5.44) and \( \sigma_z = \frac{\sigma_x^2}{\sigma_z} \).

**Proof of (5.42)**

\[
E \left( \Phi(\alpha + \xi \beta) \mid z \geq z_c \right) \mathbb{P}(z \geq z_c)
\]

\[
= \int_{z_c}^{\infty} \mathbb{P}(\varepsilon \geq -\alpha - \xi \beta) \int_{-\infty}^{\infty} f_{x,\omega}(k, k - s) \, dk \, ds
\]

\[
= \int_{z_c}^{\infty} \mathbb{P}(\varepsilon \geq -\alpha - \xi \beta) \int_{-\infty}^{\infty} f_{\omega|x}(k - s \mid \xi = k) \, f_x(k) \, dk \, ds
\]

\[
= \int_{z_c}^{\infty} \mathbb{P}(\varepsilon \geq -\alpha - \xi \beta) \int_{-\infty}^{\infty} \frac{\sigma_z^2}{\sigma_x^2} f_{\omega|x}(k - s \mid \xi = k) \, f_x(k) \, dk \, dt
\]

\[
= \int_{z_c}^{\infty} \mathbb{P}(\varepsilon \geq -\alpha - \xi \beta) \int_{-\infty}^{\infty} f_x(t - k) \, f_x(k) \, dk \, dt
\]

\[
= \int_{z_c}^{\infty} \Phi(\alpha + \xi \beta) \, f_x(t) \, dt
\]

\[
= \mathbb{E} \left( \Phi(\alpha + \xi \beta) \mid x \geq x_c \right) \mathbb{P}(x \geq x_c),
\]

where we use the substitution \( s = \frac{\sigma_x^2 t}{\sigma_z^2} \) and (5.44) in the fourth step and (5.43) in the fifth step. Using (5.45) we obtain (5.42), which completes the proof.

### 5.6 Discussion and conclusion

We have shown how the monetary value of information in direct marketing can be assessed and used in a decision process where benefits and costs of data acquisition are weighed against each other. Our approach is stylized and theoretical, but we wish to emphasize that this approach would be the core of any more extended model. Our approach provides the statistical aspects involved in assessing the value of information. There are various steps that could be taken on the road to more realism.

In order to decide whether it is profitable to buy new information, the organization should have an idea of the intra-class correlation of the new information. These could be obtained from small studies. Alternatively, a minimum intra-class correlation of new information could be specified. The result of the analysis would be a *lower bound* to the maximally attainable increase in
Discussion and conclusion

profits. A more refined way to assess the value of new information is by specifying a distribution of the intra-class correlation, indicating the organization’s uncertainty about $\rho$. By integrating the expected profits over $\rho$ we obtain the expected value of new information.

An important extension would be to incorporate more regressors. This extension is not trivial, because all regression coefficients are estimated inconsistently even if only one regressor is measured imperfectly (grouping or measurement error). A possible solution of this problem is to neglect the imperfect information, i.e. exclude the variable from the model. However, we then have an omitted variable problem and that also gives us inconsistent estimates (e.g. Yatchew and Griliches 1985), even if the variables are uncorrelated. A natural question that arises is whether it is preferable to include or to omit the proxy variable, which of course depends on the correlation with the other variables. If, for example, the correlation between the proxy variable and an other variable is nearly perfect, the additional information of the proxy variable is superfluous for selection and it will probably be better to omit the proxy variable.