# Binary Choice Models with Discrete Regressors: Identification and Misspecification

Tatiana Komarova [*]

London School of Economics

May 24, 2012

**Abstract**

In semiparametric binary response models, support conditions on the regressors are required to guarantee point identification of the parameter of interest. For example, one regressor is usually assumed to have continuous support conditional on the other regressors. In some instances, such conditions have precluded the use of these models; in others, practitioners have failed to consider whether the conditions are satisfied in their data. This paper explores the inferential question in these semiparametric models when the continuous support condition is not satisfied and all regressors have discrete support. I suggest a recursive procedure that finds sharp bounds on the components of the parameter of interest and outline several applications, focusing mainly on the models under the conditional median restriction, as in Manski (1985). After deriving closed-form bounds on the components of the parameter, I show how these formulas can help analyze cases where one regressor's support becomes increasingly dense. Furthermore, I investigate asymptotic properties of estimators of the identification set. I describe a relation between the maximum score estimation and support vector machines and also propose several approaches to address the problem of empty identification sets when a model is misspecified. Finally, I present a Monte Carlo experiment and an empirical illustration to compare several estimation techniques.

JEL Classification: C2, C10, C14, C25

Keywords: Binary response models; Discrete regressors; Partial identification; Misspecification; Support vector machines

[*] First version: *November 2007.* London School of Economics and Political Science, Houghton Street, London WC2A 2AE, England. Tel.: +44-(0)20-7852-3707; fax: + 44-(0)20-7955-6592. E-mail: t.komarova@lse.ac.uk. I am deeply indebted to Elie Tamer for many helpful discussions, guidance and encouragement throughout this project. I am very grateful to Joel Horowitz, Charles Manski and Roza Matzkin for their valuable advice and suggestions. I thank the participants of CEMMAP Conference on Quantile Regressions and seminar participants at Oxford University, Toulouse School of Economics and Universidad Carlos III de Madrid.

# 1 Introduction

The econometrics literature on inference in semiparametric binary response models have used support conditions on observable regressors to guarantee point identification of the vector parameter of interest. These support conditions always require continuity of one (or more) regressors. In practice though, it is not uncommon to have data sets where all regressors have discrete support, such as age, years of education, number of children and gender. In these cases, the parameter of interest is not point identified, that is, a large set of parameters will be consistent with the model. Therefore, it is important to develop methods of drawing accurate inferences without continuous support conditions on the data.

This paper examines the question of identification in semiparametric binary response models in the absence of continuity. Consider

$$Y = 1(X\beta + U \geq 0), \tag{$BR$}$$

where $Y$ is an observable binary outcome, $U$ is an unobservable, real-valued, scalar random variable, $\beta$ is a $k$-dimensional parameter, and $X$ is an observable random variable with discrete support. I impose a weak median condition on the error term, as in Manski (1985):

$$M(U|X = x) = 0 \quad \text{for any } x \text{ in support of } X. \tag{$MED$}$$

In this framework, I provide a methodologically new approach to the analysis of binary response models. The paper makes the following contributions.

I note that the parameter's identification region is described by a system of a finite number of linear inequalities and therefore represents a convex polyhedron. To construct this system, it is enough to know whether conditional probabilities $P(Y = 1|X = x)$ are greater or less than 0.5. As was shown by Manski and Thompson (1986), under the median condition the sign of index $x\beta$ is the same as the sign of $P(Y = 1|X = x) - 0.5$. Moreover, Manski (1988) used this fact to establish a general non-identification result for the case of discrete regressors. The first contribution of the paper is to provide a recursive procedure that allows us to easily find sharp bounds on the components of the parameter of interest $\beta$. Although this procedure was outlined, for example, in Kuhn (1956) and Solodovnikov (1977), it has not been used in the context of identification.

I derive formulas for bounds on the components of the parameter, which prove useful in analyzing cases when the support of one regressor can become increasingly dense, that is, when we approach point identification conditions in Manski (1988). Furthermore, I show that the recursive procedure can be used not only to find sharp bounds but also to determine other characteristics of the identification region. Moreover, it can be employed in the extrapolation problem when we want to learn about $P(Y = 1|X = x_0)$ for a point $x_0$ that is off the support. In addition, because identification regions in ordered response and in single-index models with a monotone link function are described by systems of linear inequalities, the recursive procedure can be applied to them too.

Another contribution of the paper is to link binary response models to support vector machines (SVMs) in statistical learning theory. When the support of $X$ is discrete and the median condition $(MED)$ holds, binary response models classify points in the support into two groups and every parameter value from the identification set defines a hyperplane that separates these groups. SVMs, in their turn, is a learning technique that focuses on finding a special hyperplane that efficiently separates two classes of given training data. The major difference is that binary response models aim to find all separating hyperplanes, whereas SVMs seek only one hyperplane.

Because models might carry some degree of specification error, the recursive procedure may cease working in some situations. Therefore, it is important to develop techniques that address the consequences of model misspecification. The third contribution of this paper is to offer several methods for dealing with the issue, all of which are based on the optimization of certain objective functions. One approach is the maximum score estimation method presented in Manski (1975, 1985). Another allows us to measure the degree of misspecification by finding the minimal number of classification errors. Each method features a crucial property: The set of solutions coincides with the identification set when the model is well specified. The third approach is a modification of a soft margin hyperplane approach in SVMs and it lets us determine the extent of misspecification by determining the minimal size of a general classification error. For a well specified model, this approach gives the closure of the identification set.

Another contribution of this paper is to explore the estimation of the identification region and the sharp interval bounds. Although this paper focuses on identification, it is of interest to analyze cases where conditional probabilities $P(Y=1|X=x)$ are not known, but their estimates $\widehat{P}(Y=1|X=x)$ are available. In this situation, we can find estimates of identification sets and sharp interval bounds from a system of linear inequalities that uses $\widehat{P}(Y=1|X=x)$ instead of $P(Y=1|X=x)$. I show that when the model is well specified, such set estimators of identification sets (sharp interval bounds) converge to the true identification set (true sharp interval bounds) arbitrarily fast in terms of Hausdorff distances. I find that the sets of maximum score estimates possess the same property. I also construct confidence regions for the identification set and the sharp interval bounds and show that because of the discrete nature of the problem, they are usually conservative.

The paper presents the results of a Monte Carlo experiment with a well-specified model. The error term satisfies the median condition but is not independent of the regressors. I show that the estimates of the sharp interval bounds obtained from the system of inequalities that uses estimated conditional probabilities coincide with the identification intervals for the components of the parameter. The same is true for the sets of maximum score estimates for individual components of the parameter. For coefficients corresponding to non-constant regressors, I find the set of maximum rank correlation estimates, which turn out to lie inside the identification intervals but form much smaller sets. I also present normalized probit and logit estimates. Though these estimates are located inside the identification intervals, they are far from the value of the parameter, which was used to generate the model.

The last contribution of this paper is an empirical application which is based on data regarding the labor force participation of married women. The decision of women to participate in the labor force is treated as a dependent binary variable and regressed on education, age, labor market experience and number of children. I use different estimation techniques and compare their results. Given that misspecification or sampling error leaves the system of inequalities constructed from the estimates of conditional probabilities without solutions, I use methods suggested for dealing with the misspecification problem. I also find normalized probit and logit estimates, ordinary least squares and least absolute deviation estimates, and compare them to other estimates.

This paper is related to two strands of the literature. The first one embodies a considerable amount of work on partially identified models in econometrics. Studies on partial identification were largely initiated and advanced by Manski (see, for example, Manski (1990, 1995, 2003)), Manski and Tamer (2002) and carried further by other researchers.

The second strand analyzes models with discrete regressors. This topic is relatively underdeveloped in econometric theory, in spite of its importance for empirical work. An example of a paper that touches upon this subject is Honore and Tamer (2006). The authors describe how to characterize the identification set for dynamic random effects discrete choice models when points in the support have discrete distributions. For single-index models $E(Y|X = x) = \phi_\theta(x\theta)$ with discrete explanatory variables and no assumption on the link function $\phi_\theta$ except for measurability, Bierens and Hartog (1988) show that there is an infinite number of observationally equivalent parameters. In particular, the identification set of the $k$-dimensional parameter $\theta = (\theta_1, \ldots, \theta_k)$ normalized as $\theta_1 = 1$ will be whole space $\Re^{k-1}$, with the exception of a finite number of hyperplanes (or a countable number of hyperplanes if the regressors have a discrete distribution with an infinite number of values). In binary response models with discrete regressors, Manski (1988) provides a general non-identification result and Horowitz (1998) demonstrates that the parameter can be identified only in very special cases. Magnac and Maurin (2005) also address identification issues in binary response models with discrete regressors. Their framework, however, is different from the framework in this paper. They consider a case where there is a special covariate among the regressors and assume that the model satisfies two conditions related to this covariate - partial independence and large support conditions.

The rest of the paper is organized as follows. Section 2 explains the problem and defines the identification set. Section 3 contains the mathematical apparatus and describes the recursive procedure. It also outlines applications of the recursive procedure, in particular to single-index and ordered-response models. Section 4 analyzes the case in which the discrete support of regressors grows increasingly dense. Section 5 draws an analogy between identification in bi to SVMs Section 5.2 considers misspecification issues and suggests techniques for dealing with them.

Section 6 considers the estimation of the identification set from a sample and statistical inference. Section 7 contains the results of estimations in a Monte Carlo experiment and the empirical application based on MROZ data. Section 8 concludes and outlines ideas for future research. The proofs of theorems and propositions are collected in the Appendix.

# 2 Partial identification

I begin by reviewing the main point identification results in the literature as well as the support conditions that guarantee point identification.

Manski (1985) proved that, when coupled with the median condition ($MED$), the following conditions on the support of $X$ guarantee that $\beta$ in ($BR$) is identified up to scale:

1. The support of the distribution of $X$ is not contained in any proper linear subspace of $\Re^k$.

2. There is at least one component $X_m$, $m \in \{1, \dots, k\}$ with $\beta_m \neq 0$ such that for almost every $\widetilde{x} = (x_1, \dots, x_{m-1}, x_{m+1}, \dots, x_k)$, the distribution of $X_m$ conditional on $\widetilde{X} = \widetilde{x}$ has a density positive almost everywhere with respect to the Lebesgue measure.

In particular, if we normalize $\beta_1 = 1$, then $\beta$ is identified.

The smoothed maximum score method described in Horowitz (1992), the Klein and Spady's (1993) approach and the maximum rank correlation method presented in Han (1987) require these conditions. It is worth mentioning that Manski (1988) presents other identification results. For instance, under certain conditions, even when $\beta$ is not identified, the signs of $\beta_1$,…,$\beta_k$ can be identified. Horowitz (1998) contains a thorough review of identification results for binary response models.

Now I turn to the case of discrete support. Let $X$ be a random variable with the discrete finite support

$$S(X) = \{x^1, \dots, x^d\}. \tag{2.1}$$

Following Manski and Thompson (1986, 1989), notice that the median condition allows us to rewrite the binary response model in a form that contains only conditional probabilities $P(Y = 1|X = x)$ and linear inequalities:

$$Pr(Y = 1|X = x) \geq 0.5 \quad \Leftrightarrow \quad x\beta \geq 0. \tag{$BRM$}$$

Thus, model ($BR$) together with ($MED$) is equivalent to model ($BRM$), and the identification problem comes down to solving a system of a finite number of inequalities. If for some $l = 1, \ldots, d$, we have $Pr(Y = 1|X = x^l) \geq 0.5$, then the inequality corresponding to $x^l$ is

$$z_{l1} + z_{l2}\beta_2 + \ldots + z_{lk}\beta_k \geq 0,$$

where $z_l = x^l$. If $Pr(Y = 1|X = x^l) < 0.5$, then the inequality corresponding to $x^l$ is

$$z_{l1} + z_{l2}\beta_2 + \ldots + z_{lk}\beta_k > 0,$$

where $z_l = -x^l$. Though this system contains strict and non-strict inequalities, for the sake of notational convenience, I will write it as the system of non-strict inequalities

$$z_{l1} + z_{l2}\beta_2 + \ldots + z_{lk}\beta_k \geq 0, \quad l = 1, \ldots, d.$$

It is important to keep in mind, however, that some inequalities are strict; this property is what allows us to separate the points with $P(Y = 1|X = x) \geq 0.5$ from the points with $P(Y = 1|X = x) < 0.5$.

Throughout this paper, I use normalization $\beta_1 = 1$, along with the notations $x^l$, which denote points in the support, and $d$, which stands for the number of these points, as in (2.1). I assume that all points in the support are different. Furthermore, $q^l$ is used to denote the probability of $x^l$ in the population and $P^l$ is used to indicate conditional probabilities $P(Y = 1|X = x^l)$, assuming that $0 < q^l < 1$ for any $l$. The parameter's identification set is denoted as $B$.

$N$ stands for the number of observations in a sample, $\hat{q}^l$ denotes the sample frequency estimator of $q^l$ and $\hat{P}^l$ signifies an estimator of $P^l$. An estimator of $B$ is denoted as $\widehat{B}$. Throughout the paper, $z_l = sgn(P^l - 0.5)\, x^l$, where function $sgn(\cdot)$ is defined as

$$sgn(t) = \begin{cases} 1, & t \geq 0 \\ -1, & t < 0. \end{cases}$$

In several instances $z_l$ will mean $z_l = sgn(\hat{P}^l - 0.5)\, x^l$. These cases will be clear from the context.

**Theorem 2.1.** *The closure of the identification set of parameter $\beta$ in (BRM) is a $k_0$-dimensional convex polyhedron, where $k_0 \leq k - 1$.*

**Corollary 2.2.** *The identification set of $\beta_m$, $m \neq 1$, in (BRM) is a connected interval.*

Theorem 2.1 does not need any proof because, by definition, a convex polyhedron is the set of solutions to some finite system of non-strict linear inequalities (see, for instance, Rockafellar (1972)). Corollary 2.2 holds because the identification interval for $\beta_m$, $m \neq 1$, is the projection of the identification set on the axis $x_l$. Due to the convexity of $B$, it is a connected interval.

Every bounded convex polyhedron can be equivalently described pointwise as the convex hull of a finite number of points.[1] (The minimal set of such points is the so-called set of the vertices of the polyhedron.) Any unbounded convex polyhedron can be represented as a Minkowski sum of a bounded convex polyhedron and a convex cone (see, for example, Padberg (1999)). Even though some methods for finding all the vertices from a system of linear inequalities were suggested in the literature,[2] finding the vertices is not easy and such methods have not proved useful in theoretically analyzing the properties of convex polyhedra.

An easier, effective approach in the theoretical analysis of the identification set $B$ in $(BRM)$ is finding the smallest rectangular superset of $B$. This rectangle is the Cartesian product of the identification intervals for $\beta_m$, $m \neq 1$. Its dimension can be smaller than $k-1$ if some $\beta_m$, $m \neq 1$, are point identified.

Figure 1 shows an identification set on the left and its smallest rectangular superset on the right.

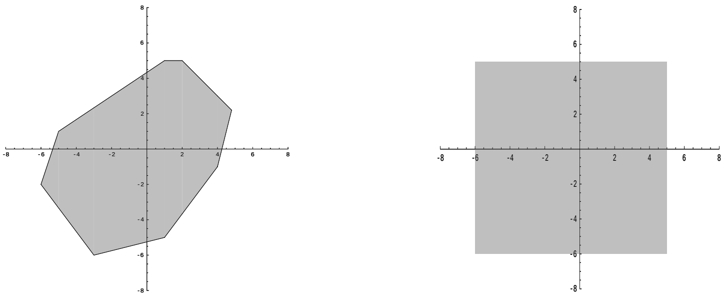

Figure 1. Convex polyhedron and its smallest rectangular superset

# 3 Mathematical tools

## 3.1 Recursive method

In this section, I describe a recursive procedure that finds identification intervals for $\beta_m$, $m \neq 1$. Applied to a system of linear inequalities, this method excludes from the system one unknown variable at each step until only a single variable is left. From there, identifying the sharp bounds for the remaining variable is straightforward. Although this approach is outlined, for instance, in Kuhn (1956) and Solodovnikov (1977), I supplement it by discussing cases in which the system has an unbounded or empty set of solutions and by deriving formulas for parametric bounds.

[1] See Rockafellar (1972).

[2] For instance, Motzkin et. al. (1953), Balinski (1961), Chernikova (1965), Manas and Nedoma (1968), Matheiss (1973) and Matheiss and Rubin (1980).

Consider an arbitrary system of linear inequalities with $k-1$ unknown variables:

$$\begin{aligned} z_{11}+z_{12}b_2+\ldots+z_{1k}b_k &\geq 0 \\ z_{21}+z_{22}b_2+\ldots+z_{2k}b_k &\geq 0 \\ &\ldots \\ z_{d1}+z_{d2}b_2+\ldots+z_{dk}b_k &\geq 0. \end{aligned} \tag{$S_1$}$$

Suppose we want to find sharp bounds for variable $b_k$. Consider $i$'s inequality in the system:

$$z_{i1}+z_{i2}b_2+\ldots+z_{ik}b_k \geq 0.$$

If $z_{i2}>0$, then this inequality is equivalent to

$$-\frac{z_{i1}}{z_{i2}}-\frac{z_{i3}}{z_{i2}}b_3\ldots-\frac{z_{ik}}{z_{i2}}b_k \leq b_2.$$

If $z_{j2}<0$, then it is equivalent to

$$-\frac{z_{j1}}{z_{j2}}-\frac{z_{j3}}{z_{j2}}b_3-\ldots-\frac{z_{jk}}{z_{j2}}b_k \geq b_2.$$

Suppose system $(S_1)$ has $I$ inequalities with $z_{\cdot 2}>0$, $J$ inequalities with $z_{\cdot 2}<0$ and $M$ inequalities with $z_{\cdot 2}=0$. In this case, $(S_1)$ can be equivalently written as the system

$$\begin{aligned} b_2 \geq D_i, &\quad i=1,\ldots,I, \\ N_j \geq b_2, &\quad j=1,\ldots,J, \\ Z_m \geq 0, &\quad m=1,\ldots,M, \end{aligned}$$

where $D_i$, $N_j$, $Z_m$ do not contain $b_2$ and are linear in $b_3,\ldots,b_k$. This system implies that

$$\begin{aligned} N_j \geq D_i, \quad i=1,\ldots,I, \quad j=1,\ldots,J, \\ Z_m \geq 0, \quad m=1,\ldots,M. \end{aligned} \tag{$S_2$}$$

$(S_2)$ is a system of linear inequalities with $k-2$ unknown variables. In other words, system $(S_2)$ consists of $M$ inequalities that do not contain $b_2$, and $I\times J$ linear combinations of inequalities $i\in I$ and inequalities $j\in J$ with positive coefficients $-z_{j2}$ and $z_{i2}$, respectively. This first step is illustrated in example 9.1 in the Appendix.

The process above is repeated and a variable (for example, $b_3$) is excluded from $(S_2)$ to obtain a system $(S_3)$ of linear inequalities with unknown variables $(b_4,\ldots,b_k)$. Then this procedure is being repeated again, removing one variable each time, until a system with only one unknown variable $b_k$ is obtained:

$$\begin{aligned} a_s + c_s b_k \geq 0, &\quad s = 1, \dots, S, \\ h_q \geq 0, &\quad q = 1, \dots, Q, \end{aligned} \tag{$S_{k-1}$}$$

where $c_s \neq 0$, $s = 1, \dots, S$.

The next proposition establishes a relationship between solutions to $(S_1)$ and $(S_{k-1})$.

**Proposition 3.1.** *If $(b_2^*, b_3^*, \dots, b_k^*)$ is a solution to $(S_1)$, then $b_k^*$ is a solution to $(S_{k-1})$. If $b_k^*$ is a solution to $(S_{k-1})$ , then there exists $(b_2, \dots, b_{k-1})$ such that $(b_2, b_3, \dots, b_{k-1}, b_k^*)$ is a solution to $(S_1)$.*

When system $(S_{k-1})$ is obtained, I find

$$\underline{b}_k = \max \left\{ -\frac{a_s}{c_s} : c_s > 0 \right\},$$

$$\overline{b}_k = \min \left\{ -\frac{a_s}{c_s} : c_s < 0 \right\}.$$

The set of solutions to $(S_{k-1})$ is $\left[\underline{b}_k, \overline{b}_k\right]$. Proposition 3.1 implies that $\underline{b}_k$ and $\overline{b}_k$ are sharp bounds for $b_k$. This last step is illustrated in example 9.2 in the Appendix. If the set of solutions to $(S_1)$ is nonempty and bounded, then the described procedure always goes through till the very last step of getting system $(S_{k-1})$. In order to obtain a system with only variable $b_k$ in the last step, variables $b_2, \dots, b_{k-1}$ can be excluded in an order different from the one described here. For instance, variables could be eliminated in the order $b_{k-1}, b_{k-2}, \dots, b_2$. Proposition 3.1 implies that any order of elimination of the variables gives the same interval $\left[\underline{b}_k, \overline{b}_k\right]$ in the last step.

Though it has been implicitly assumed until now that $(S_1)$ has solutions, it is possible that the system of inequalities derived from econometric models does not have solutions because of model misspecification, or because of a sampling error if the system is constructed based on some estimates $\hat{P}^l$ of conditional probabilities.

When $(S_1)$ does not have solutions, it is intuitive that the recursive procedure breaks at some point. In this situation, at some step we may get an obvious contradiction $C \geq 0$ where $C$ is a negative constant. We may also be able to reach the last step of the procedure, only to discover that $(S_{k-1})$ contains clear contradictions, such as $h_q \geq 0$ where $h_q$ is a negative constant or $\overline{b}_k < \underline{b}_k$. These situations are illustrated in examples 9.3 and 9.4 in the Appendix.

If $(S_1)$ has an unbounded set of solutions, this can be easy to spot if at one step in the process we notice that all the coefficients corresponding to some variable $b_m$ have the same strict sign. In this case we can conclude right away that the solution set is unbounded. Nevertheless, we may still be able to continue the procedure and learn more about the solution set if there is a variable with coefficients of both signs. These cases are considered in example 9.4 in the Appendix. If at some step in the procedure we notice that, first, no variables in the system have both negative and positive signs, and second, each variable has zero coefficients, then the system needs further investigation because its solution set may be either unbounded or empty.

Clearly, the recursive procedure can be used to find sharp bounds on any parameter. For example, to find the sharp bounds on parameter $b_3$, we can exclude $b_2$, then $b_4$,. . . , $b_k$ in the manner described.

**Example 3.1.** Consider a model

$$Y = 1(X_1 + \beta_2 X_2 + \beta_3 X_3 + U \geq 0),$$

where $X_1$ takes values from $\{-5, -4, \ldots, 4, 5\}$, $X_2$ is the constant term ($X_2 = 1$), and $X_3$ takes values from $\{0, 1, \ldots, 7\}$. Thus, the support of $X$ contains 88 points. Whether conditional probabilities $P^l$, $l = 1, \ldots, 88$, are above or below 0.5 is determined by the rule

$$\begin{aligned} x_1^l + 1.25 - 0.5x_3^l \geq 0 \quad &\Rightarrow \quad P^l \geq 0.5 \\ x_1^l + 1.25 - 0.5x_3^l < 0 \quad &\Rightarrow \quad P^l < 0.5. \end{aligned}$$

Now, knowing which $P^l$ are above and which are below 0.5, I can construct the system of linear inequalities that defines the identification set $B$:

$$P^l \geq 0.5 \quad \Leftrightarrow \quad x_1^l + x_2^l b_2 + x_3^l b_3 \geq 0, \quad l = 1, \ldots, 88.$$

The recursive procedure finds the bounds on $\beta_2$ and $\beta_3$: $\beta_2 \in (1, 1.6)$, $\beta_3 \in (-0.6, -0.4286)$. The values of 1.25 and $-0.5$ are, of course, within those bounds. Figure 2 shows both the identification set and bounds. □

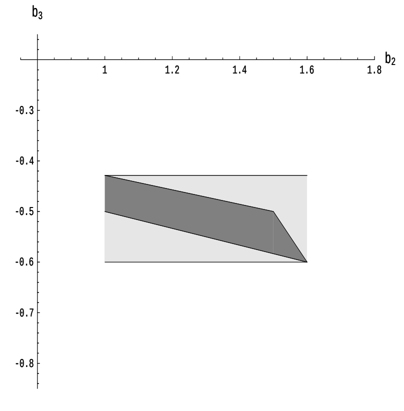

Figure 2. Identification set and its smallest rectangular superset in Example 3.1

## 3.2 Formulas for bounds on $\beta_m$, $m \neq 1$

In this section, the recursive procedure is used to derive formulas for the bounds on $\beta_m$, $m \neq 1$, in model $(BRM)$. These bounds are expressed in terms of $z_l = sgn(P^l - 0.5)x^l$, $l = 1, \dots, d$. For clarity, I first show formulas for the case $k = 3$ and then present them in general.

**Proposition 3.2.** *Let $k = 3$. Suppose the solution set to $(S_1)$ is non-empty and bounded. Then*

$$b_{3L} \leq \beta_3 \leq b_{3U}, \qquad \textit{where}$$

$$b_{3U} = \min_{i,j} \left\{ -\frac{\begin{vmatrix} z_{j1} & z_{j2} \\ z_{i1} & z_{i2} \end{vmatrix}}{\begin{vmatrix} z_{j3} & z_{j2} \\ z_{i3} & z_{i2} \end{vmatrix}} : z_{j2} < 0, z_{i2} > 0, \begin{vmatrix} z_{j3} & z_{j2} \\ z_{i3} & z_{i2} \end{vmatrix} < 0 \right\} = \min_{i,j} \left\{ -\frac{\begin{vmatrix} z_{j1} & z_{j2} \\ z_{i1} & z_{i2} \end{vmatrix}}{\begin{vmatrix} z_{j3} & z_{j2} < 0 \\ z_{i3} & z_{i2} > 0 \end{vmatrix} < 0} \right\},$$

$$b_{3L} = \max_{i,j} \left\{ -\frac{\begin{vmatrix} z_{j1} & z_{j2} \\ z_{i1} & z_{i2} \end{vmatrix}}{\begin{vmatrix} z_{j3} & z_{j2} \\ z_{i3} & z_{i2} \end{vmatrix}} : z_{j2} < 0, z_{i2} > 0, \begin{vmatrix} z_{j3} & z_{j2} \\ z_{i3} & z_{i2} \end{vmatrix} > 0 \right\} = \max_{i,j} \left\{ -\frac{\begin{vmatrix} z_{j1} & z_{j2} \\ z_{i1} & z_{i2} \end{vmatrix}}{\begin{vmatrix} z_{j3} & z_{j2} < 0 \\ z_{i3} & z_{i2} > 0 \end{vmatrix} > 0} \right\}.$$

*Assume there are sets of indices $i$, $j$ such that the conditions in the definition of $b_{3U}$ and $b_{3L}$ are satisfied.*

The formulas for $b_{3U}$ and $b_{3U}$ are in a certain symmetric. In order to find the upper (lower) bound $b_{3U}$ ($b_{3L}$), we choose inequalities for which the determinant in the denominator is negative (positive). This symmetry gives results in section 4, which considers situations when the support of $X$ becomes increasingly dense.

Note that bounds $b_{3U}$ and $b_{3L}$ are not necessarily sharp. According to the recursive procedure, if in some inequalities the coefficients corresponding to $b_2$ are 0, then those inequalities are carried over to the next step without any changes. The formulas for $b_{3U}$ and $b_{3L}$, however, ignore situations in which some coefficients corresponding to $b_2$ can be 0. Therefore, they do not necessarily describe the sharp bounds. However, because an inequality with $z_{i2} = 0$ and $z_{i3} < 0$ gives $b \leq -\frac{z_{i1}}{z_{i3}}$, and an inequality with $z_{i2} = 0$ and $z_{i3} > 0$ gives $b \geq -\frac{z_{i1}}{z_{i3}}$, then the sharp upper and lower bounds, which account for cases $z_{i2} = 0$, can be written, respectively, as follows:

$$\min\left\{ b_{3U}, \min_i \left\{ -\frac{z_{i1}}{z_{i3}} \; : \; z_{i2} = 0, \; z_{i3} < 0 \right\} \right\}, \quad \max\left\{ b_{3L}, \min_i \left\{ -\frac{z_{i1}}{z_{i3}}, : \; z_{i2} = 0, \; z_{i3} > 0 \right\} \right\}.$$

If there is a constant term among regressors, these formulas are simpler. Without a loss of generality, $x_2^i = 1$, $i = 1, \dots, d$. Then $z_{j2} < 0$ is possible if and only if $z_{j2} = -1$, and $z_{i2} > 0$ is possible if and only if $z_{i2} = 1$. The formulas can be written as follows:

$$b_{3U} = \min_{i,j} \left\{ -\frac{x_1^i - x_1^j}{x_3^i - x_3^j} : x_3^i - x_3^j < 0 \right\}, \quad b_{3L} = \max_{i,j} \left\{ -\frac{x_1^i - x_1^j}{x_3^i - x_3^j} : x_3^i - x_3^j > 0 \right\}.$$

Now I obtain formulas for bounds for any $k \geq 3$. Define

$$A_1(m, i, j) = \begin{vmatrix} z_{jm} & z_{j2} \\ z_{im} & z_{i2} \end{vmatrix},$$

where $z_{j2}$, $z_{i2}$ are such that $z_{j2} < 0$, $z_{i2} > 0$. Let me write this formula as

$$A_1(m, i, j) = \begin{vmatrix} z_{jm} & z_{j2} < 0 \\ z_{im} & z_{i2} > 0 \end{vmatrix}.$$

Then, according to Proposition 3.2,

$$b_{3U} = \min_{i,j} \left\{ -\frac{A_1(1, i, j)}{A_1(3, i, j) < 0} \right\}, \quad b_{3L} = \max_{i,j} \left\{ -\frac{A_1(1, i, j)}{A_1(3, i, j) > 0} \right\}.$$

Note that $A_1(m, i, j)$ can be written in the form

$$A_1(m, i_1, j_1, i_2, j_2) = \begin{vmatrix} A_0(m, j) & A_0(2, j) < 0 \\ A_0(m, i) & A_0(2, i) > 0 \end{vmatrix},$$

where $A_0(m, i) = z_{im}$. The proposition below presents a general result.

**Proposition 3.3.** *Let $k \geq 3$. Suppose the set of solutions to $(S_1)$ is non-empty and bounded. Then*

$$b_{kL} \leq \beta_k \leq b_{kU},$$

*where*

$$b_{kU} = \min_{i_1, \dots, j_{2^{k-3}}} \left\{ -\frac{A_{k-2}(1, i_1, \dots, j_{2^{k-3}})}{A_{k-2}(k, i_1, \dots, j_{2^{k-3}}) < 0} \right\},$$

$$b_{kL} = \max_{i_1, \dots, j_{2^{k-3}}} \left\{ -\frac{A_{k-2}(1, i_1, \dots, j_{2^{k-3}})}{A_{k-2}(k, i_1, \dots, j_{2^{k-3}}) > 0} \right\},$$

*where $A_{k-2}(m, i_1, \dots, j_{2^{k-3}})$ is defined recursively as*

$$A_{k-2}(m, i_1, \dots, j_{2^{k-3}}) = \begin{vmatrix} A_{k-3}(m, i_1, \dots, j_{2^{k-4}}) & A_{k-3}(k-1, i_1, \dots, j_{2^{k-4}}) < 0 \\ A_{k-3}(m, i_{2^{k-4}+1}, \dots, j_{2^{k-3}}) & A_{k-3}(k-1, i_{2^{k-4}+1}, \dots, j_{2^{k-3}}) > 0 \end{vmatrix},$$

$$A_0(m, i) = z_{im}.$$

*Assume there are sets of indices $i_1, \dots, i_{2^{k-3}}$, $j_1, \dots, j_{2^{k-3}}$ such that the conditions in the definition of $b_{kU}$ and $b_{kL}$ are satisfied.*

Bounds $b_{kU}$ and $b_{kL}$ are not necessarily sharp. The reason for that is similar to the explanation described earlier for the case of $k = 3$: The formulas ignore instances in which some inequalities are carried over to the next step without any changes. When $k > 3$, it is notationally difficult to keep track of these inequalities at each step of the procedure; that is why formulas for the sharp bounds for $k > 3$ are not presented here.

## 3.3 Applications of the recursive method

This section presents several applications of the recursive method.

### 3.3.1 Extrapolation

Consider a point $x^* = (x_1^*, \dots, x_k^*) \notin S(X)$ for which we want to learn whether the projected value of $P(Y = 1|X = x^*)$ is above or below 0.5.

We proceed by finding the sharp bounds on the values of linear function

$$x_1^* + x_2^* b_2 + \ldots + x_k^* b_k, \quad (b_2, \dots, b_k) \in B.$$

If the lower bound is non-negative, then $P(Y = 1|X = x^*) - 0.5 \geq 0$. If the upper bound is negative, then $P(Y = 1|X = x^*) - 0.5 < 0$. Finally, if the interval between the lower and upper bound contains zero, then no conclusions can be drawn about the value of $sgn(P(Y = 1|X = x^*) - 0.5)$.

In order to find these bounds, introduce a new variable $b'$ which is the value of the following linear function:

$$b' = x_1^* + x_2^* b_2 + \ldots + x_k^* b_k.$$

Assume that at least one of $x_1^*$, $x_2^*$, …, $x_k^*$, is different from 0. Without a loss of generality, $x_k^* \neq 0$. Then $b_k$ can be expressed through $b_2$, …, $b_k$, $b'$ as follows:

$$b_k = \frac{1}{x_k^*} b' - \frac{x_1^*}{x_k^*} - \frac{x_2^*}{x_k^*} b_2 - \ldots - \frac{x_{k-1}^*}{x_k^*} b_{k-1}.$$

Next, substitute this expression into the system that defines the identification set:

$$z_{l1} + z_{l2} b_2 + \ldots + z_{lk} b_k \geq 0, \quad l = 1, \dots, d,$$

and obtain the system of linear inequalities with unknown variables $b_2$, …, $b_{k-1}$, $b'$:

$$z_{l1} - z_{lk} \frac{x_1^*}{x_k^*} + \left( z_{l2} - z_{lk} \frac{x_2^*}{x_k^*} \right) b_2 + \ldots + \left( z_{l,k-1} - z_{lk} \frac{x_{k-1}^*}{x_k^*} \right) b_{k-1} + \frac{z_{lk}}{x_k^*} b' \geq 0, \quad l = 1, \dots, d.$$

The recursive procedure can be used to find the sharp bounds for $b'$ – that is, the sharp bounds on the values of the linear function.

As an example, consider a situation in which $Y = 1$ if an individual has a high deductible insurance, and $Y = 0$ if an individual has a low deductible insurance. Using extrapolation we can predict which insurance the people who currently do not have coverage will choose with a higher probability when they are forced to buy an insurance.

### 3.3.2 Single-index models with a monotone link function

Consider a single-index model

$$E(Y|X = x) = G(x\beta) \tag{3.1}$$

with an *unknown increasing* function $G(\cdot)$. In semiparametric binary choice models this is the case when, for instance, the error term is independent of the regressors and its distribution is not specified. Normalize $\beta_1 = 1$. The identification set is the set of $(b_2, \dots, b_k)$ such that $b = (1, b_2, \dots, b_k)$ solves the following system of strict linear inequalities:

$$\forall\,(x^l, x^m \in S(X)) \qquad E(Y|X = x^l) > E(Y|X = x^m) \;\Rightarrow\; (x^l - x^m)b > 0. \tag{3.2}$$

Indeed, the inequalities in this system are necessary conditions from (3.1). Also, for any $b$ that satisfies (3.2), there exists an increasing function $G(\cdot)$ that takes the value of $E(Y|X = x^l)$ at $x^l b$, $l = 1, \dots, d$, and thus, can generate (3.1). This means that (3.2) completely characterized the identification set. To find the sharp bounds on the components of $\beta$, we need to apply the recursive procedure to the system of inequalities defined in (3.2).

If, for example, $E(Y|X = x^1) > E(Y|X = x^2) > \ldots > E(Y|X = x^d)$, then the identification set is described by the following system of $d - 1$ inequalities:

$$x^1 b > x^2 b > \ldots > x^d b,$$

or, equivalently,

$$(x^1 - x^2)b > 0, \quad (x^2 - x^3)b > 0, \quad \ldots, \quad (x^{d-1} - x^d)b > 0.$$

In general, there may be many more than $d-1$ inequalities defining the identification set. For example, if $d$ is even and $E(Y|X = x^l) = 1$ for $l = 1, \dots, d/2$, and $E(Y|X = x^l) = 0$ for $l = d/2 + 1, \dots, d$, then the system describing the identification set comprises $d^2/4$ inequalities:

$$(x^l - x^m)b > 0, \quad l = 1, \dots, d/2, \quad m = d/2 + 1, \dots, d.$$

If it is known that function $G$ is *strictly increasing*, then the values in the identification set also satisfy conditions

$$\forall (x^l, x^m \in S(X)) \qquad E(Y|X = x^l) = E(Y|X = x^m) \;\Leftrightarrow\; x^l b = x^m b. \tag{3.3}$$

Each equality $x^l b = x^m b$ can be written as the system of two non-strict inequalities $(x^l - x^m)b \geq 0$ and $(x^m - x^l)b \geq 0$. Then the identification set is completely characterized by the system of strict inequalities in (3.2) and non-strict inequalities in (3.3).

Note that if there is a constant term among the components of the covariates, then under the single-index restriction no bounds can be obtained on the coefficient corresponding to the constant term since it gets canceled out in the difference $x^l - x^m$.

**Example 3.2.** *(Example 3.1 continued.)* Consider a model

$$Y = 1(X_1 + \beta_2 X_2 + \beta_3 X_3 + U \geq 0),$$

where $X_1$ takes values from $\{-5, -4, \ldots, 4, 5\}$, $X_2$ is the constant term ($X_2 = 1$), and $X_3$ takes values from $\{0, 1, \ldots, 7\}$. Suppose $U$ is independent of $X$. Then

$$P(Y = 1|X = x) = G(x_1 + \beta_2 + \beta_3 x_3),$$

where $G(\cdot)$ is the distribution function of $U$. Construct the system of linear inequalities, defining the identification set, in the following way. For any $x^l, x^m \in S(X)$, if

$$x_1^l + 1.25 - 0.5x_3^l > x_1^m + 1.25 - 0.5x_3^m,$$

then the inequality

$$x_1^l - x_1^m - b_3(x_3^l - x_3^m) > 0$$

belongs to the system.

This system gives the following bounds on $\beta_3$: $\beta_3 \in (-0.5714, -0.4286)$. As for $\beta_2$, only the trivial conclusion $\beta_2 \in (-\infty, \infty)$ can be made. □

If both the single-index restriction (3.1) and the median condition ($MED$) hold, then the system for the identification set includes inequalities from both (3.2) and ($BRM$). It is worth noting, however, that the sharp bounds for the coefficients corresponding to non-constant covariates remain the same as those obtained from (3.2) only. This is intuitive because we can always think about (3.1) as the model obtained from the independence of $U$ and $X$ in ($BR$). The independence property implies that the conditional median $Med(U|X = x)$ does not depend on $x$. Therefore, restriction ($MED$), which sets this conditional median to be equal to 0, only provides a "location normalization" in the sense that it allows us to obtain non-trivial bounds on the coefficient corresponding to the constant term.

**Example 3.3.** *(Examples 3.1 and 3.2 continued.)* Consider the setting in examples 3.1 and 3.2 and suppose that both the single-index restriction and the median condition ($MED$) hold. Combining the system of linear inequalities in example 3.1 with that in example 3.2, I obtain the following sharp bounds on $\beta_2$ and $\beta_3$: $\beta_2 \in (1, 1.5714)$, $\beta_3 \in (-0.5714, -0.4286)$. □

If no monotonicity assumptions or any other assumptions except for measurability are imposed on $G(\cdot)$, then, as shown by Bierens and Hartog (1988), the set of observationally equivalent parameters is almost the whole space (for more details, see the Introduction section).

### 3.3.3 Ordered response models

In ordered response models, agents choose among $M$ alternatives according to the rule

$$Y = \sum_{m=1}^{M+1} m1(\alpha_{m-1} < Y^* \leq \alpha_m),$$

where $\alpha_0 < \alpha_1 < \ldots < \alpha_M < \alpha_{M+1}$, $\alpha_0 = -\infty$, $\alpha_{M+1} = +\infty$, $Y^* = X\beta + U$. In general, threshold levels $\alpha_m$ are not known. If we assume the median condition $M(U|X = x^l) = 0$, $l = 1, \ldots, d$, then

$$P(Y \leq m|X = x^l) = P(x^l\beta + U \leq \alpha_m) = F_{U|x^l}(\alpha_m - x^l\beta), \quad l = 1, \ldots, d, \quad m = 1, \ldots, M.$$

The identification set is described by a system of linear inequalities with a maximum of $2d + M - 1$ inequalities. Each $x^l$ contributes one or two inequalities to the system, and each $x^l$ has three possibilities: $P(Y \leq 1|X = x^l) \geq 0.5$, or $P(Y \leq M|x = x^l) < 0.5$, or $P(Y \leq 1|X = x^l) < 0.5$ and $P(Y \leq M|x = x^l) \geq 0.5$. In the first case, $x^l$ contributes

$$\alpha_1 - x^l b \geq 0.$$

In the second case, it provides

$$\alpha_M - x^l b < 0.$$

In the third case, find $m(x^l) \in \{2, \ldots, M\}$ such that

$$P(Y \leq m(x^l) - 1|X = x^l) < 0.5, \quad P(Y \leq m(x^l)|x = x^l) \geq 0.5.$$

Then $x^l$ contributes two inequalities

$$\alpha_{m(x^l)-1} - x^l b < 0, \quad \alpha_{m(x^l)} - x^l b \geq 0. \tag{3.4}$$

Thus, points from $S(X)$ form a system of at most $2d$ linear inequalities. However, we also have to add $M-1$ inequalities

$$\alpha_m - \alpha_{m-1} > 0, \quad m = 2, \ldots, M.$$

Some of these additional inequalities will be excessive because, for example, (3.4) implies that $\alpha_{m(x^l)-1} < \alpha_{m(x^l)}$.

After $b$ is normalized by setting $b_1 = 1$, we have a system of at most $2d + M - 1$ linear inequalities with $k - 1 + M$ unknown variables $b_2, \ldots, b_k, \alpha_1, \ldots, \alpha_M$. This system contains both strict and non-strict inequalities. The recursive method can be applied to find the sharp bounds on the values of these variables.

# 4 Dense support

One of the sufficient conditions for point identification in $(BR)$ given by Manski (1988) is that for almost every $x_2, \ldots, x_k$ the distribution of $X_1$ conditional on $x_2, \ldots, x_k$, has a density positive almost everywhere. It is intuitive that when covariates are discrete but the values of $x_1$ corresponding to a fixed vector $(x_2, \ldots, x_k)$ form a rather dense set for many values of $(x_2, \ldots, x_k)$, then the identification set should be small. Proposition 4.1 formalizes this suggestion.

**Proposition 4.1.** *Consider system $(S_1)$ with $k = 3$. Suppose that its solution set is non-empty and bounded. Also, suppose that the system contains four inequalities*

$$\begin{aligned}
z_{i_1,1} + z_{i_1,2}b_2 + z_{i_1,3}b_3 &\geq 0 \\
z_{i_2,1} + z_{i_2,2}b_2 + z_{i_2,3}b_3 &\geq 0 \\
z_{j_1,1} + z_{j_1,2}b_2 + z_{j_1,3}b_3 &\geq 0 \\
z_{j_2,1} + z_{j_2,2}b_2 + z_{j_2,3}b_3 &\geq 0
\end{aligned}$$

*such that*

$$z_{i_1,2} > 0, \; z_{i_2,2} > 0$$

$$z_{i_1,2}z_{i_2,3} - z_{i_1,3}z_{i_2,2} > 0$$

$$(z_{j_1,2}, z_{j_1,3}) = -(z_{i_1,2}, z_{i_1,3}), \;\; z_{i_1,1} + z_{j_1,1} < \Delta$$

$$(z_{j_2,2}, z_{j_2,3}) = -(z_{i_2,2}, z_{i_2,3}), \;\; z_{i_2,1} + z_{j_2,1} < \Delta$$

*for a fixed* $\Delta > 0$*. Then*

$$b_{3U} - b_{3L} \leq \frac{\begin{vmatrix} \Delta & -z_{i_2,2} \\ \Delta & z_{i_1,2} \end{vmatrix}}{\begin{vmatrix} z_{i_2,3} & z_{i_2,2} \\ z_{i_1,3} & z_{i_1,2} \end{vmatrix}} = \Delta \frac{\begin{vmatrix} 1 & -z_{i_2,2} \\ 1 & z_{i_1,2} \end{vmatrix}}{\begin{vmatrix} z_{i_2,3} & z_{i_2,2} \\ z_{i_1,3} & z_{i_1,2} \end{vmatrix}},$$

*where* $b_{3U}$*,* $b_{3L}$ *are defined as in Proposition 3.2.*

This result is obtained using the symmetry of the formulas for $b_{3U}$ and $b_{3L}$. If we take four other inequalities that satisfy the conditions of the proposition, then we obtain a different bound for $b_{3U} - b_{3L}$, and we can choose the lower of two.

The role of Proposition 3.2 may be better appreciated if we formulate an analogous result in terms of the properties of the support.

**Corollary 4.2.** *Let* $B$ *be non-empty and bounded. Suppose that there exist* $(x_2, x_3)$ *and* $(x_2^*, x_3^*)$ *such that*

$$x_2 x_3^* - x_2^* x_3 \neq 0,\; x_2 \neq 0,\; x_2^* \neq 0. \tag{4.1}$$

*Also suppose that*

$$\exists (x_1, \tilde{x}_1 : (x_1, x_2, x_3), (\tilde{x}_1, x_2, x_3) \in S(X)) \forall (b \in B)$$

$$x_1 + x_2 b_2 + x_3 b_3 \geq 0, \quad \tilde{x}_1 + x_2 b_2 + x_3 b_3 < 0, \quad x_1 - \tilde{x}_1 < \Delta$$

*and*

$$\exists (x_1^*, \tilde{x}_1^* : (x_1^*, x_2^*, x_3^*), (\tilde{x}_1^*, x_2^*, x_3^*) \in S(X)) \forall (b \in B)$$

$$x_1^* + x_2^* b_2 + x_3^* b_3 \geq 0, \quad \tilde{x}_1^* + x_2^* b_2 + x_3^* b_3 < 0, \quad x_1^* - \tilde{x}_1^* < \Delta$$

*for a given* $\Delta > 0$*. Then*

$$b_{3U} - b_{3L} \leq \Delta \frac{|x_2| + |x_2^*|}{|x_2 x_3^* - x_2^* x_3|}. \tag{4.2}$$

Proposition 4.1 and Corollary 4.2 tell us that the length of the identification interval for $\beta_3$ can be bounded from above by a value that depends, first, on the "denseness" $\Delta$ of the support of $X_1$ conditional on fixed values $(X_2, X_3) = (x_2, x_3)$ and $(X_2, X_3) = (x_2^*, x_3^*)$, and second, on the variation in the values of $(X_2, X_3)$ that satisfy conditions (4.1) and give the "denseness" of level $\Delta$. For instance, if $X_2$ is the constant term, then the expression on the right-hand side in (4.2) becomes $\frac{2\Delta}{|x_3^* - x_3|}$. This value is inversely proportional to the absolute variation in the values of $X_3$ that give "denseness" $\Delta$.

Proposition 4.3 is an analog of Proposition 4.1 for the case of $k = 4$.

**Proposition 4.3.** *Consider system* $(S_1)$ *with* $k = 4$*. Let its solution set be non-empty and* *bounded. Suppose that $(S_1)$ contains eight inequalities*

$$\begin{aligned} z_{i_1,1} + z_{i_1,2}b_2 + z_{i_1,3}b_3 + z_{i_1,4}b_4 \geq 0 \\ \dots \\ z_{i_4,1} + z_{i_4,2}b_2 + z_{i_4,3}b_3 + z_{i_4,4}b_4 \geq 0 \\ z_{j_1,1} + z_{j_1,2}b_2 + z_{j_1,3}b_3 + z_{j_1,4}b_4 \geq 0 \\ \dots \\ z_{j_4,1} + z_{j_4,2}b_2 + z_{j_4,3}b_3 + z_{i_4,4}b_4 \geq 0 \end{aligned}$$

*such that*

$$(z_{j_m,2}, z_{j_m,3}, z_{j_m,4}) = -(z_{i_m,2}, z_{i_m,3}, z_{i_m,4}), \quad z_{i_m,1} + z_{j_m,1} < \Delta, \quad m = 1, 2, 3, 4,$$

*for some $\Delta > 0$, and also*

$$z_{i_1,2} > 0, \quad z_{i_2,2} > 0, \quad z_{i_3,2} > 0, \quad z_{i_4,2} > 0,$$

$$z_{i_2,3}z_{i_1,2} - z_{i_2,2}z_{i_1,3} > 0, \quad z_{i_4,3}z_{i_3,2} - z_{i_4,2}z_{i_3,3} > 0,$$

$$D > 0,$$

*where*

$$D = \begin{vmatrix} \begin{vmatrix} z_{i_4,4} & z_{i_4,2} \\ z_{i_3,4} & z_{i_3,2} \end{vmatrix} & \begin{vmatrix} z_{i_4,3} & z_{i_4,2} \\ z_{i_3,3} & z_{i_3,2} \end{vmatrix} \\ \\ \begin{vmatrix} z_{i_2,4} & z_{i_2,2} \\ z_{i_1,4} & z_{i_1,2} \end{vmatrix} & \begin{vmatrix} z_{i_2,3} & z_{i_2,2} \\ z_{i_1,3} & z_{i_1,2} \end{vmatrix} \end{vmatrix}.$$

*Then*

$$b_{4U} - b_{4L} \leq \frac{\Delta}{D} \begin{vmatrix} \begin{vmatrix} 1 & -z_{i_4,2} \\ 1 & z_{i_3,2} \end{vmatrix} & - \begin{vmatrix} z_{i_4,3} & z_{i_4,2} \\ z_{i_3,3} & z_{i_3,2} \end{vmatrix} \\ \\ \begin{vmatrix} 1 & -z_{i_2,2} \\ 1 & z_{i_1,2} \end{vmatrix} & \begin{vmatrix} z_{i_2,3} & z_{i_2,2} \\ z_{i_1,3} & z_{i_1,2} \end{vmatrix} \end{vmatrix},$$

*where $b_{4U}$, $b_{4L}$ are defined as in Proposition 3.3.*

As in the case $k = 3$ we can obtain bounds on $b_{4U} - b_{4L}$ in terms of the properties of the support $S(X)$. Results for $k = 3$ and $k = 4$ can be generalized for the case of any $k$.

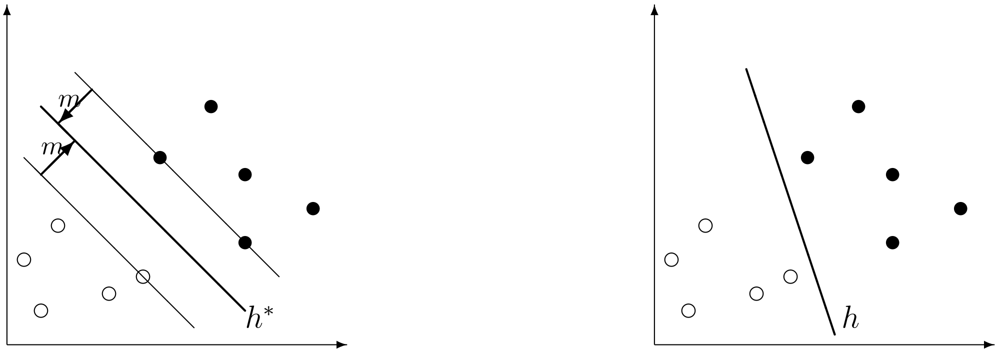

Figure 3. The optimal hyperplane $h^*$ with maximal margin $m$ (left) and a non-optimal separating hyperplane $h$ (right)

# 5 Extensions

## 5.1 Support vector machines

In $(BRM)$ all points from $S(X)$ are grouped into two classes. If $P^l \geq 0.5$, then $x^l$ is assigned to one class (where $x^l\beta \geq 0$); otherwise, it is assigned to the other class (where $x^l\beta < 0$). Manski and Thompson (1986, 1989) interpret the equivalence condition in $(BRM)$ as a “single-crossing” condition for response probabilities. It implies, in particular, that the two classes can be separated by a hyperplane. This observation provides a clear link between the identification problem in binary response models and support vector machines in the statistical learning theory, which is discussed in this section.

To describe this relationship I have to assume that one of the regressors in $X$ is a constant term. Without a restriction of generality, $X_k = 1$. Define a set $\tilde{S}(X) \subset \Re^{k-1}$ that consists of points from $S(X)$ with an omitted last regressor $x_k$. The problem of finding the identification region in $(BRM)$ is equivalent to determining all the hyperplanes that separate two classes of points from $\tilde{S}(X)$: those for which $P(Y = 1|X = x) \geq 0.5$ and those for which $P(Y = 1|X = x) < 0.5$.

Support vector machines (SVMs) are learning techniques used for classification problems. They separate a given set of binary-labeled training data with a hyperplane that maximizes the distance (or, the margin) between itself and the closest training data point. It is called the *optimal* or the *maximal margin hyperplane*.

SVMs are similar to the identification problem for $(BRM)$ in that both are hyperplane-based classifiers. However, in contrast with SVMs, which select only one hyperplane, the identification set in $(BRM)$ contains all the vectors that define separating hyperplanes. Figure 3 shows optimal and a non-optimal separating hyperplanes for the two classes of points.

Let me briefly present the mathematical description of SVMs (for more details, see

Vapnik (2000)). For a better comparison with existing literature, I adopt their notations.

Suppose we have a training set of $n$ independent and identically distributed observations of $(X, Y)$ drawn according to the unknown probability distribution $P(x, y)$:

$$\{(x_i, y_i)\}_{i=1}^n, \ x \in \Re^k, \ y \in \{-1, 1\},$$

where observation $i$ is assigned to one of two classes (-1 or 1) according to its value $y_i$. Also suppose that the two classes of points in the training set can be separated by the linear hyperplane $w_1 x_1 + \ldots + w_k x_k - a = 0$. To find the optimal hyperplane, solve the quadratic programming problem

$$\min_{w_1,\ldots,w_k,a} \sum_{j=1}^k w_j^2 \tag{5.1}$$

subject to

$$y_i[w_1 x_{i,1} + \ldots + w_k x_{i,k} - a] \geq 1, \quad i = 1, \ldots, n. \tag{5.2}$$

The solution of this problem is unique, and it defines the optimal hyperplane.

Let me now return to the identification set $B$ in $(BRM)$. As mentioned earlier, any point in $B$ can serve as a classification rule. Generally, the quality of a classification rule is measured by a loss function. In Lin (2000, 2002), the following expected misclassification rate (or generalization error) is regarded as a loss function:

$$R(\eta) = P(\eta(X) \neq Y) = \int 0.5|y - \eta(x)|dP, \tag{5.3}$$

where $\eta$ is a classification rule. The classification rule with the smallest generalization error is

$$\eta^*(x) = sgn(P(Y = 1|X = x) - 0.5).$$

Thus, the median condition $(MED)$ in the binary response model $(BR)$ guarantees that any point in the identification set $B$ constitutes a linear decision rule that is optimal with regard to the expected misclassification rate criterion (5.3).

It is important to stress that in this section it is assumed that $(BRM)$ is well specified.

## 5.2 Misspecification

In general, this paper maintains well specification of model $(BRM)$. If it is misspecified, then it is possible that $(S_1)$ has no solutions. In this section, I briefly discuss how one might think of approaching the misspecification problem.

Misspecification may be a result of omitting relevant covariates from the model, or the model may not be driven by a linear index of covariates. Any cause of misspecification that leads to $(S_1)$ having no solutions can be considered as a case of the violation of the conditional median restriction ($MED$). In this case, generalizations of ($BRM$) can be considered.

I suggest two different approaches to generalizing ($BRM$). The first one is based on considering criteria functions with certain properties and finding the values of the parameter that maximize them. This approach is in the spirit of Manski (1985) whose maximum score estimator is a solution to several maximization problems. The second approach uses the link between the identification problem in this paper and the statistical learning theory, outlined in section 5.1, to suggest a method in the spirit of the support vector machines.

### 5.2.1 Maximization of criteria functions

This section suggests considering an optimization problem

$$\max_{b:b_1=1} Q(b),$$

where criterion $Q(\cdot)$ is a piecewise continuous function and such that the set of its maximizers coincides with the identification set $B$ (or its closure) if the model is well specified. If the model is misspecified, this set of maximizers gives the set of the values of the parameter that are meaningful in a certain sense. There are potentially many functions $Q(\cdot)$ that satisfy these properties and there is no "best" one among them. Choosing $Q(\cdot)$ for a particular application is not an easy task and preferences for a certain approach usually change depending on the situation.

Below I give several examples of functions $Q(\cdot)$ that have clear interpretations in terms of the model. They are a just a few of the possible options.

**Maximum score**

Manski (1985) shows that the maximum score estimation problem can be described in several equivalent ways. One way to define it (on the population level) is to consider the maximization of the score function:[3]

[3] It can also be described as the minimization of the expected absolute distance between the conditional expectation of $Y$ conditional on $X$ and $1(Xb \geq 0)$ (if the median condition ($MED$) holds, then $M(Y|X) = 1(X\beta \geq 0)$):

$$\min_{b:b_1=1} E|E(Y|X) - 1(Xb \geq 0)|,$$

or as the minimization of the expected squared distance between $E(Y|X)$ and $1(Xb \geq 0)$:

$$\min_{b:b_1=1} E(E(Y|X) - 1(Xb \geq 0))^2,$$

or as optimization problems

$$\min_{b:b_1=1} E|Y - 1(Xb \geq 0)|,$$

$$\min_{b:b_1=1} E(Y - 1(Xb \geq 0))^2.$$

$$\max_{b:b_1=1} E(sgn(2Y-1)sgn(Xb)). \tag{5.4}$$

When all the regressors are discrete, this is tantamount to the maximization of the function

$$S^{ms}(b) = 2\sum_{l=1}^{d}(P^l - 0.5)sgn(x^l b)q^l. \tag{5.5}$$

where, I remind, $P^l = P(Y=1|X=x^l)$, $q^l = P(X=x^l)$.

Let $B^{ms}$ stand for the set of maximizers of $S^{ms}(\cdot)$:

$$B^{ms} = \left\{(b_2,\dots,b_k) : (1,b_2,\dots,b_k) \in \text{Arg}\max_{b:b_1=1} S^{ms}(b)\right\}.$$

The proposition below show that if the model is well specified, then set $B^{ms}$ coincides with $B$ if $P^l \neq 0.5$ for all the points in the support, but is strictly larger if this is not the case.

**Proposition 5.1.** *Suppose that model (BRM) is well specified. Then $B^{ms}$ is a convex polyhedron and*
*(a) if $P^l \neq 0.5$ for all $x^l \in S(X)$, then $B = B^{ms}$.*
*(b) if $P^l = 0.5$ for some $x^l \in S(X)$, then $B \subset B^{ms}$ with a strict inclusion.*

If $(BRM)$ is misspecified, then $B^{ms}$ is not necessarily a convex polyhedron. This may happen, for instance, when many identical values exist among $(P^l - 0.5)q^l$, $l = 1,\dots,d$. However, $B^{ms}$ is always a finite union of disjoint convex polyhedra.

Function $S^{ms}(\cdot)$ is among possible criteria functions $Q(\cdot)$.

**Minimal number of classification errors**

In $(BRM)$ all points from $S(X)$ can be grouped into two classes. If $P^l \geq 0.5$, then $x^l$ is assigned to one class; otherwise, it is assigned to the other class. If the model is misspecified and the system of inequalities defining $B$ has no solutions, this means that some classification errors have been made. One may be interested in finding the values of the parameter that minimize the number of classification errors. These values can be found by maximizing

$$Q(b) = \sum_{l=1}^{d} sgn(P^l - 0.5)sgn(x^l b). \tag{5.6}$$

Because the value of $Q(b)$ is equal to $d$ minus the number of classification errors, the maximization of $Q(b)$ minimizes the number of these errors. If $(BRM)$ is well specified, then $Q(\cdot)$ attains its maximum possible value $d$ and the set of maximizers for (5.6) coincides with $B$.

A criterion which is similar to (5.6) but takes into account the probabilities of $x^l$ in the support, and thus, assigns different degrees of importance to classification errors for different support points, is

$$Q(b) = \sum_{l=1}^{d} sgn(P^l - 0.5) sgn(x^l b) q^l. \tag{5.7}$$

The maximization of this $Q(\cdot)$ is equivalent to the minimization of the expected absolute distance between $M(Y|X)$ and $1(Xb \geq 0)$:

$$\min_{b:b_1=1} E|M(Y|X) - 1(Xb \geq 0)|.$$

The sets of maximizers for (5.6) and (5.7) are finite unions of disjoint convex polyhedra.

### 5.2.2 Minimal general classification error

This section discusses an approach to misspecification based on the soft margin hyperplane method in support vector machines, a technique suggested in Cortes and Vapnik (1995) for handling cases in which two classes of training data are not linearly separable. Roughly speaking, this method deals with errors in the data by allowing some anomalous points to fall on the wrong side of the hyperplane. This approach is discussed in detail later in this section.

The relation of this method to misspecification issues in model $(BRM)$ is the following. As mentioned earlier, any cause of misspecification can be considered as a case of the violation of the median condition $Med(Y|X = x^l) = 0$, $l = 1, \ldots, d$. A more general model specification would allow the conditional median $Med(Y|X = x^l)$ to depend on $x^l$ but at the same time to satisfy conditions

$$Med(Y|X = x^l) \in [-\mu^l, \mu^l], \quad \mu^l \geq 0, \quad l = 1, \ldots, d. \tag{5.8}$$

The method described in this section essentially finds the values of the parameter (or, in the terminology of the SVMs literature, finds soft margin hyperplanes) that minimize the sum $\sum\limits_{l=1}^{d} \mu^l$ – the total deviation from the median condition $(MED)$ – subject to restrictions (5.8).[4] If the model is well specified, then the optimal value of each $\mu^l$ is 0, and therefore, the total deviation from the median condition $(MED)$ is 0 as well.

For SVMs, only one soft margin hyperplane is considered: the optimal one. I am interested in finding a set of these hyperplanes, however, rather than just one. In fact, I would like to determine the set of all soft margin hyperplanes, if possible. An attractive feature of the approach outlined here is that it only requires solving linear programming problems, so it is easy to implement.

As in section 5.1, first assume that there is a constant term among regressors. Without

[4] We can also consider a weighted sum of $\mu^l$'s.

a restriction of generality, $X_k = 1$. Let set $\tilde{S}(X) \subset \Re^{k-1}$ consist of points from $S(X)$ with an omitted last regressor $x_k$. For now, let me abandon normalization $b_1 = 1$. Without this normalization, the identification set is a subset of $\Re^k$. Denote it as $\tilde{B}$. Geometrically, $\tilde{B}$ consists of $k$-dimensional vectors $(b_1, b_2, \ldots, b_{k-1}, b_k)$ such that hyperplanes with the slope coefficients $(b_1, \ldots, b_{k-1})$ and the location coefficient $b_k$ separate two groups of points in $\tilde{S}(X)$: the class for which $P^l \geq 0.5$ and that for which $P^l < 0.5$:

$$\forall(l = 1, \ldots, d) \quad x_1^l b_1 + x_2^l b_2 + \ldots + x_{k-1}^l b_{k-1} + b_k \geq 0 \quad \Leftrightarrow \quad P^l \geq 0.5.$$

For each class, consider the convex hull of its points. Because the closures of the convex hulls do not intersect, there are separating hyperplanes that do not contain any points from $\tilde{S}(X)$. In other words, these hyperplanes are separated from either class by a strictly positive distance:

$$\exists(b \in \tilde{B})\exists(\delta > 0)\forall(l = 1, \ldots, d) \quad sgn(P^l - 0.5)(x_1^l b_1 + x_2^l b_2 + \ldots + x_{k-1}^l b_{k-1} + b_k) \geq \delta. \quad (5.9)$$

Note that in this assertion, I have incorporated the finite support of $X$ and the constant term among regressors. Because for any $b \in \tilde{B}$, vector $\alpha b$, $\alpha > 0$, defines the same hyperplane as $b$, $\delta$ can be any positive number. Without a loss of generality, suppose that $\delta = 1$. Let $h$ stand for the separating hyperplane $x_1^l b_1 + x_2^l b_2 + \ldots + x_{k-1}^l b_{k-1} + b_k = 0$. Let $h_1$ denote the hyperplane $x_1^l b_1 + x_2^l b_2 + \ldots + x_{k-1}^l b_{k-1} + b_k = 1$ and $h_2$ denote the hyperplane $x_1^l b_1 + x_2^l b_2 + \ldots + x_{k-1}^l b_{k-1} + b_k = -1$. Then, according to (5.9), all points from one class lie above or lie on hyperplane $h_1$ and all points from the other class lie below or lie on hyperplane $h_2$.

If model $(BRM)$ is misspecified, then the two classes of points are not necessarily linearly divisible. In this case, I introduce non-negative slack variables to allow for some error in separation and to find a soft margin hyperplane.

Let $v_l \geq 0$, $l = 1, \ldots, d$, denote slack variables. The value of $v_l$ is interpreted as the size of the classification error for point $x^l$. Consider the following linear programming problem:

$$\min_{b, \{v_l\}_{l=1}^d} Q(b, v) = \sum_{l=1}^{d} v_l \quad (5.10)$$

subject to

$$sgn(P^l - 0.5)(x_1^l b_1 + x_2^l b_2 + \ldots + x_{k-1}^l b_{k-1} + b_k) \geq 1 - v_l,$$

$$v_l \geq 0, \quad l = 1, \ldots, d.$$

Denote its set of solutions as $D^* \subset \Re^{k+d}$. Let

$$B^* = \{b \in \Re^k : (b, v) \in D^* \text{ for some } v \in \Re^d\}, \; V^* = \{v \in \Re^d : (b, v) \in D^* \text{ for some } b \in \Re^k\}.$$

Notice that when $(BRM)$ is well specified, the optimal value of the objective function is 0, and $V^* = \{(0, \ldots, 0)\}$.

For any $(b^*, v^*) \in D^*$, the hyperplane $x_1 b_1^* + \ldots + x_{k-1} b_{k-1}^* + b_k^* = 0$ defined by $b^*$ is called a soft margin hyperplane. Because $\sum_{l=1}^{d} v_l$ can be interpreted as a general classification error, soft margin hyperplanes minimize this error.

Take the following example. In Figure 4, points with $P^l \geq 0.5$ are depicted as dark circles, and points with $P^l < 0.5$ are depicted as white circles. As we can see, the two classes of points are not linearly separable. Consider a hyperplane $h$ defined by a vector $b$ (with $b_1 > 0$): $x_1 b_1 + x_2 b_2 + b_3 = 0$. Also picture two hyperplanes parallel to $h$ that are separated from it by an equal distance: hyperplane $h_1$, defined as $x_1 b_1 + x_2 b_2 + b_3 = 1$, and hyperplane $h_2$, defined as $x_1 b_1 + x_2 b_2 + b_3 = -1$. In the case of separability, we could find a $b$ such that all dark points would lie above or on hyperplane $h_1$, and all the white points would lie below or on hyperplane $h_2$. From this point of view, points 7, 8 and 10 are located on the correct side of $h$ and $h_1$. Therefore, their classification errors $v_7$, $v_8$ and $v_{10}$ are 0. Point 9, on the other hand, is located on the correct side of $h$ but the incorrect side of $h_1$; the distance from point 9 to its correct location (that is, to $h_1$) is $v_9$, the point's classification error. For 6, located on the incorrect side of $h$, the distance to $h_1$ is $v_6$. In the second class only point 3, located on the incorrect side of $h$, has a classification error; its distance to $h_2$ is $v_3$. In fact, hyperplane $h$, as shown on Figure 4, is a soft margin hyperplane for the depicted two classes of points.

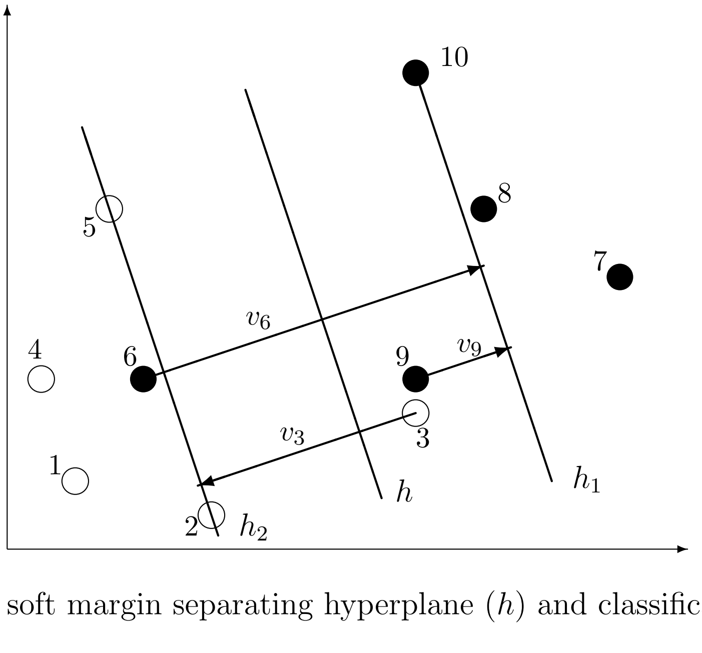

Figure 4. A soft margin separating hyperplane ($h$) and classification errors

Let me now return to the general problem. If instead of $\delta = 1$ we took a different $\delta > 0$ and solved (5.10) subject to

$$sgn(P^l - 0.5)(x_1^l b_1 + x_2^l b_2 + \ldots + x_{k-1}^l b_{k-1} + b_k) \geq \delta - v_l,$$

we would obtain the same soft margin hyperplanes, though the solution set $D^*(\delta)$ would differ from $D^*$. Namely, set $D^*(\delta)$ would comprise the elements of $D^*$ multiplied by $\delta$: $D^*(\delta) = \delta D^*$. Given this scale effect, we are interested in the set

$$B_1^* = \{(b_2, \dots, b_k) : \exists(\gamma > 0) \quad (\gamma, \gamma b_2, \dots, \gamma b_k) \in B^*\} \subset \Re^{k-1}$$

rather than set $B^*$ itself. (At this stage, without a loss of generality I assume that $b_1^* > 0$ for any $b^* \in B^*$.)

Finding the smallest multidimensional rectangle superset for $B_1^*$ would be difficult. To do that, we would have to consider every solution in $B^*$, divide all of its coordinates by the first coordinate to find ratios, and summarize the results for all solutions.

The problem becomes much easier if we are satisfied with finding the smallest multidimensional rectangle superset for a subset of $B_1^*$. Let $(b^*, v^*)$ be any element from $D^*$. Define $B_{1,s}^* \subset \Re^{k-1}$ as the solution set $(b_2, \dots, b_k)$ to the following system:

$$sgn(P^l - 0.5)(x_1^l + x_2^l b_2 + \dots + x_{k-1}^l b_{k-1} + b_k) \geq \frac{1 - v_l^*}{b_1^*}, \quad l = 1, \dots, d. \tag{5.11}$$

Clearly, $B_{1,s}^* \subset B_1^*$, and system (5.11) can be rewritten in the form of $(S_1)$. Therefore, we can find the smallest multidimensional rectangle superset for $B_{1,s}^*$ by using the recursive procedure.

We can also consider some modifications of the soft margin hyperplane method. In the general classification error in (5.10), all the slack variables have the same weight. In SVMs, all the training data points are of equal importance, therefore, SVMs consider a general classification error only in this form. Nevertheless, we can discriminate between points in $S(X)$ and assign different weights to slack variables. In other words, the general classification error can take the form $\sum_{l=1}^{d} \lambda_l v_l$, where $\sum_{l=1}^{d} \lambda_l = 1$, $\lambda_l \geq 0$, $l = 1, \dots, d$. For instance, we may be willing to assign more importance to points with a higher probability of occurring and consider the objective function $\sum_{l=1}^{d} q^l v_l$. Different weights $\lambda_l$ will yield different solution sets $B_1^*$ and, consequently, different sets $B_{1,s}^*$.

In the soft margin approach in SVMs, instead of problem (5.1), the method solves the following quadratic programming problem:

$$\min_{w_1, \dots, w_k, a, \xi} \sum_{j=1}^{k} w_j^2 + C \sum_{i=1}^{n} \xi_i^\sigma$$

subject to

$$y_i[w_1 x_{i,1} + \dots + w_k x_{i,k} - a] \geq 1 - \xi_i,$$

$$\xi_i \geq 0, \quad i = 1, \dots, n,$$

for given $0 < \sigma \leq 1$ and $C > 0$. The objective function presents the trade-off between the maximal margin and the minimal penalty. For $\sigma = 1$, the penalty function is linear in $\xi$. When $\sigma$ is small, the penalty function is close to the number of classification errors.

The idea of using linear programming problems with slack variables in the context of identification is suggested in Honore and Tamer (2006). In the authors' framework, a system of linear equations describes the identification set. To check whether a particular parameter value belongs to the set, they introduce non-negative slack variables into the equation constraints and minimize their sum subject to these modified restrictions. If the optimal function value is 0, they conclude that the parameter value belongs to the identification set.

# 6 Asymptotic properties

## 6.1 Consistency

The system of inequalities that defines the identification set can be constructed once it is known whether conditional probabilities $P^l$ are below 0.5 or not. Usually, only estimates of conditional probabilities are available, so natural questions are: a) how close does an estimated identification set get to the true set as the sample size increases; b) how well do estimates of the sharp interval bounds approximate the true sharp interval bounds for the components of the parameter. This section gives answers to these questions.

In this section, I assume that model $(BRM)$ is well specified, and I distinguish two cases. In one case, all conditional probabilities are different from 0.5. In the other one, some conditional probabilities are 0.5. In the latter instance, the problem of finding the identification set is ill-posed in the sense that small changes in the estimators of conditional probabilities may cause considerable changes in the estimator of the identification set, no matter how large the sample size is. Let $R(B)$ denote the minimal multidimensional rectangle superset of $B$. In other words, $R(B)$ is the Cartesian product of the identification intervals for the components of $\beta$. For a set estimate $\widehat{B}$, the notation for its minimal multidimensional rectangle superset is $R(\widehat{B})$.

**Theorem 6.1.** *Suppose that*

$$\hat{P}^l \xrightarrow{p} P^l \text{ as } N \to \infty, \tag{6.1}$$

*and $P^l \neq 0.5$, $l = 1, \ldots, d$. Let $\widehat{B}$ be a solution set for the system of linear inequalities derived by the rule*

$$\forall (x^l \in S(X)) \quad \hat{P}^l \geq 0.5 \quad \Leftrightarrow \quad x^l b \geq 0. \tag{6.2}$$

*Then*

$$Pr(\widehat{B} \neq B) \to 0 \quad as \quad N \to \infty,$$
$$Pr(R(\widehat{B}) \neq R(B)) \to 0 \quad as \quad N \to \infty.$$

The statement of Theorem 6.1 does not hold if there exist $P^l$ equal to 0.5. Nevertheless, in this case a consistent estimators of $B$ and $R(B)$ can be derived by introducing slack variables $\epsilon_N$.

**Theorem 6.2.** *Suppose that*

$$\tau_N(\hat{P}^l - P^l), \quad l = 1, \dots, d, \tag{6.3}$$

*has a non-degenerate distribution limit as* $N \to \infty$, *where* $0 < \tau_N < \infty$, $\tau_N$ *is increasing and* $\tau_N \to \infty$ *as* $N \to \infty$. *Let* $\epsilon_N$ *be a sequence of numbers such that*

$$\epsilon_N > 0 \textit{ and } \epsilon_N \to 0, \quad \epsilon_N \tau_N \to \infty \quad as \quad N \to \infty. \tag{6.4}$$

*Let* $\widehat{B}$ *be a solution set for the system of linear inequalities derived by the rule*

$$\forall (x^l \in S(X)) \quad \hat{P}^l \geq 0.5 - \epsilon_N \quad \Leftrightarrow \quad x^l b \geq 0.$$

*Then*

$$Pr(\widehat{B} \neq B) \to 0 \quad as \quad N \to \infty,$$
$$Pr(R(\widehat{B}) \neq R(B)) \to 0 \quad as \quad N \to \infty.$$

Theorems 6.1 and 6.2 imply properties of convergence in terms of Hausdorff distances. The next corollary shows that the Hausdorff distance between $B$ and $\widehat{B}$ and the Hausdorff distance between $R(B)$ and $R(\widehat{B})$ converge in probability to 0 with an arbitrary rate.

**Corollary 6.3.** *Under the conditions of Theorem 6.1 or Theorem 6.2,*

$$r_N H(\widehat{B}, B) \xrightarrow{p} 0 \quad as \quad N \to \infty$$
$$r_N H(R(\widehat{B}), R(B)) \xrightarrow{p} 0 \quad as \quad N \to \infty$$

*for any* $0 < r_N < \infty$. *(For instance, one can take* $r_N = N^c$, $c > 0$.*)*

Given that the support of $X$ is finite, it may be convenient to use frequency estimates of conditional probabilities:

$$\hat{P}^l = \frac{\sum y_i 1(x_i = x^l)}{\sum 1(x_i = x^l)}, \tag{6.5}$$

based on a random sample $(y_i, x_i)_{i=1}^N$. These estimates give stronger consistency result.

**Theorem 6.4.** *Let $\hat{P}^l$ be defined as in (6.5) and $\widehat{B}$ be the solution set for the system of linear inequalities derived by the rule (6.2).*

*If $P^l \neq 0.5$ for any $x^l \in S(X)$, then $\exists\,(0 < \rho < 1)$ such that*

$$Pr(\widehat{B} \neq B) = o(\rho^N),$$
$$Pr(R(\widehat{B}) \neq R(B)) = o(\rho^N)$$

*as $N \to \infty$.*

*If $P^l = 0.5$ for some $x^l \in S(X)$, then asymptotically, the estimator $\widehat{B}$ differs from the identification set $B$ with a positive probability:*

$$\exists(p_0 > 0) \quad Pr(\widehat{B} \neq B) \geq p_0 \text{ as } N \to \infty.$$

The second claim of Theorem 6.4 does not necessarily hold for $Pr(R(\widehat{B}) \neq R(B))$ because there may be special cases when $R(\widehat{B}) = R(B)$ but $\widehat{B} \neq B$.

Random sampling errors in estimated response probabilities may cause the system of inequalities based on $\hat{P}^l$ to have no solutions. In this case, $\widehat{B}$ would be empty. One way to address this problem is to consider sample analogs of the objective functions in sections 5.2.1 and **??**. For instance, for the objective function in section **??** a sample analog is

$$\widehat{Q}(b) = \sum_{l=1}^{d} sgn(x^l b) sgn(\hat{P}^l - 0.5).$$

Let $\widehat{B}^*$ be the set of maximizers of $Q(b)$:

$$\widehat{B}^* = \{(b_2, \ldots, b_k) \,:\, (1, b_2, \ldots, b_k) \in \operatorname{Arg}\max_{b:b_1=1} \widehat{Q}(b)\}.$$

If $\widehat{B} \neq \emptyset$, then $\widehat{B} = \widehat{B}^*$. All results in the current section remain true if $\widehat{B}$ is substituted for $\widehat{B}^*$. Other objective functions $\widehat{Q}(\cdot)$ also can be considered. Because the model is well specified, the only condition required for $\widehat{Q}(\cdot)$ is that the set of maximizers of its population analog $Q(\cdot)$ coincides with $B$. In general, it is preferable to consider objective functions that reflect some of the model's properties.

If $\widehat{B} = \emptyset$, either sampling errors in conditional probabilities estimates or misspecification could be at fault. Though it would be interesting to develop tests to distinguish between those two cases, I leave this task to future research.

The maximum score estimates obtained from a random sample maximize the function

$$\widehat{S}^{ms}(b) = \frac{1}{N}\sum_{i=1}^{N}(2y_i - 1)sgn(x_i b),$$

which can be equivalently written as

$$\widehat{S}^{ms}(b) = 2\sum_{l=1}^{d}\hat{q}^l(\hat{P}^l - 0.5)sgn(x^l b),$$

where $\hat{P}^l$ is defined as in (6.5), and $\hat{q}^l = \frac{1}{N}\sum_{i=1}^{N} 1(x_i = x^l)$.

**Proposition 6.5.** *Suppose that model (BRM) is well specified. Let*

$$\widehat{B}^{ms} = \{(b_2, \ldots, b_k) : (1, b_2, \ldots, b_k) \in Arg \max_{b:b_1=1} \widehat{S}^{ms}(b)\}.$$

*If* $P^l \neq 0.5$ *for any* $x^l \in S(X)$*, then* $\exists\,(0 < \rho < 1)$

$$Pr(\widehat{B}^{ms} \neq B) = o(\rho^N),$$
$$Pr(R(\widehat{B}^{ms}) \neq R(B)) = o(\rho^N).$$

*as* $N \to \infty$*.*

*If* $P^l = 0.5$ *for some* $x^l \in S(X)$*, then*

$$\exists(p_0 > 0) \quad Pr(\widehat{B}^{ms} \neq B) \geq p_0 \text{ as } N \to \infty.$$

## 6.2 Statistical inference

This section outlines a possible approach to building confidence regions for both $B$ and $R(B)$. This method is based on using normal approximations of conditional probabilities $P^l$. Because of the problem's discrete nature, it is not always possible to achieve an exact nominal confidence level; in many cases, the true confidence coefficient is greater than the stated level. The

Given a random sample of size $N$, the objective is to construct regions $\widehat{B}$ that asymptotically cover $B$ with probability $1 - \alpha$, where $\alpha$ is a prespecified value between 0 and 1:

$$\lim_{N\to\infty} P(\widehat{B} \supset B) \geq 1 - \alpha. \tag{6.6}$$

If such regions are found, then rectangles $R(\widehat{B})$ asymptotically cover $R(B)$ with probability $1 - \alpha$ as well:

$$\lim_{N\to\infty} P(R(\widehat{B}) \supset R(B)) \geq \lim_{N\to\infty} P(\widehat{B} \supset B) \geq 1 - \alpha.$$

For any $x^l \in S(X)$, let $\hat{P}^l$ be the frequency estimator of $P^l$ as in (6.5). Asymptotically,

$$\sqrt{N}\left(\hat{P}^l - P^l\right) \to \mathcal{N}(0, (\sigma^l)^2), \quad \text{where} \quad (\sigma^l)^2 = \frac{P^l(1 - P^l q^l)}{q^l}.$$

Substitute $\sigma^l$ with its estimate $\hat{\sigma}^l$:

$$\sqrt{N}\frac{\hat{P}^l - P^l}{\hat{\sigma}^l} \to \mathcal{N}(0,1), \quad \text{where} \quad \hat{\sigma}^l = \sqrt{\frac{\hat{P}^l(1 - \hat{P}^l \hat{q}^l)}{\hat{q}^l}}, \quad \hat{q}^l = \frac{1}{N}\sum 1(x_i = x^l).$$

Choose numbers $\{\gamma_l\}_{l=1}^d$ such that $\gamma_l \geq 0$, $l = 1, \ldots, d$, and $\sum_{l=1}^d \gamma_l = \alpha$ ($d$ continues to denote the number of points in $S(X)$). Let $\zeta_{\gamma_l}$ denote the $1 - \gamma_l$ quantile of the standard normal distribution. Construct a system of linear inequalities in the following way:
if for a given $x^l$

$$\hat{P}^l - \zeta_{\gamma_l}\frac{\hat{\sigma}^l}{\sqrt{N}} > 0.5,$$

then add $x^l b \geq 0$ to the system. If

$$\hat{P}^l + \zeta_{\gamma_l}\frac{\hat{\sigma}^l}{\sqrt{N}} < 0.5,$$

then add $-x^l b > 0$ to the system. If the interval $\left[\hat{P}^l - \zeta_{\gamma_l}\frac{\hat{\sigma}^l}{\sqrt{N}},\, \hat{P}^l + \zeta_{\gamma_l}\frac{\hat{\sigma}^l}{\sqrt{N}}\right]$ contains 0.5, then no inequalities corresponding to $x^l$ are added to the system. The solution set $\widehat{B}$ for the system constructed according to this method has the property

$$P(\widehat{B} \supset B) \to 1 \text{ as } N \to \infty,$$

which, in its turn, implies that

$$P(R(\widehat{B}) \supset R(B)) \to 1 \text{ as } N \to \infty.$$

Indeed,

$$P(\widehat{B} \not\supset B) \leq \sum_{l=1}^{d}\left(P\left(\hat{P}^l - \zeta_{\gamma_l}\frac{\hat{\sigma}^l}{\sqrt{N}} > 0.5 | P^l < 0.5\right) + P\left(\hat{P}^l - \zeta_{\gamma_l}\frac{\hat{\sigma}^l}{\sqrt{N}} < 0.5 | P^l > 0.5\right)\right).$$

Note that

$$P\left(\hat{P}^l - \zeta_{\gamma_l}\frac{\hat{\sigma}^l}{\sqrt{N}} > 0.5 | P^l < 0.5\right) = 1 - \Phi\left((0.5 - P^l)\frac{\sqrt{N}}{\hat{\sigma}^l} - \zeta_{\gamma_l}\right) \to 0 \text{ as } N \to \infty$$

and, similarly,

$$P\left(\hat{P}^l - \zeta_{\gamma_l}\frac{\hat{\sigma}^l}{\sqrt{N}} < 0.5|P^l > 0.5\right) \to 0 \text{ as } N \to \infty.$$

Clearly,

$$\begin{aligned} P(\widehat{B} \supset B) = 1 - P(\widehat{B} \not\supset B) \to 1, \\ P(R(\widehat{B}) \supset R(B)) \to 1 \end{aligned}$$

as $N \to \infty$.

As can be sees, confidence region $\widehat{B}$ is overly conservative: Its actual coverage probability is 1. In this sense, it is inaccurate. Furthermore, in practice $\widehat{B}$ can be empty. In this case, instead of $\widehat{B}$, set $\widehat{B}^*$ that solves the following optimization problem can be considered:

$$\max_{b:b_1=1}\sum_{l=1}^{d} sgn(x^l b \geq 0)\left(1\left(\hat{P}^l - \zeta_{\gamma_l}\frac{\hat{\sigma}^l}{\sqrt{N}} > 0.5\right) - 1\left(\hat{P}^l + \zeta_{\gamma_l}\frac{\hat{\sigma}^l}{\sqrt{N}} < 0.5\right)\right).$$

If $\widehat{B}$ is not empty, $\widehat{B} = \widehat{B}^*$. Using the technique employed for region $\widehat{B}$, it can be proven that $Pr(\widehat{B}^* \supset B) \to 1$ and $Pr(R(\widehat{B}^*) \supset R(B)) \to 1$ as $N \to \infty$.

Recent studies on the construction of confidence sets for partially identified parameters include Imbens and Manski (2004), Chernozhukov, Hong and Tamer (2007) and Rosen (2006), among others. Imbens and Manski (2004) propose confidence intervals that cover the true value of the parameter rather than the entire identification region. Chernozhukov, Hong and Tamer (2007) consider models in which the identification region is the set of the minimizers for a criterion function. They build confidence regions with a specified probability by using a suggested subsampling procedure. Rosen (2006) examines models defined by a finite number of moment inequalities and constructs confidence sets through pointwise testing. In recent years, there has also been increasing interest in finite-sample methods of inference. For instance, for binary choice and multinomial choice models, Manski (2007) develops confidence sets that are valid for all sample sizes.

# 7 Monte Carlo experiment and an empirical example

## 7.1 Monte Carlo experiment

The design of the Monte Carlo experiment is based on Example 3.1. The outcome data are generated as follows:

$$Y = 1(X_1 + 1.25X_2 - 0.5X_3 + U \geq 0).$$

$X_1$, $X_2$ and $X_3$ take the values described in Example 3.1 ($X_2$ is the constant term). Let the distribution of the error term be

$$U|x \sim \frac{x_1 1(x_1 < 0)}{\sqrt{2x_1^2 + 2x_3^2 + 0.001}} Z + 0.1 x_3 1(x_1 \geq 0) V,$$

where random variable $Z$ has a standard normal distribution, random variable $V$ is distributed uniformly on $[-1, 1]$, and $Z$ and $V$ are independent. Clearly, the conditional median independence assumption is satisfied. I report results for a sample of size $N = 5,000$.[5]

For comparison, I apply several estimation procedures. There are 87 points in the sample's support. From the sample I calculate frequency estimates of conditional probabilities. Using these estimates I construct a system of linear inequalities by the rule

$$\hat{P}^l \geq 0.5 \quad \Leftrightarrow \quad x_1^l + b_2 + b_3 x_3^l \geq 0$$

and apply the recursive procedure to find bounds on $\beta_2$ and $\beta_3$. Because the system has solutions, its set of maximum score estimates coincides with its solution set. See Table 1 for the estimation results. Observe that the set of maximum rank correlation estimates for $\beta_3$ does not contain value -0.5 used to design the experiment, although this value is very close to its border.

I want to emphasize that the results presented for the recursive procedure and the maximum score method are identification intervals for each individual component of parameter; the identification set for $(\beta_2, \beta_3)$ is smaller than rectangle $(1, 1.6) \times (-0.6, -0.42587)$ (see Figure 2).

I also report probit and logit estimates with 95% confidence intervals for each parameter, as well as normalized probit and logit estimates (ratios $\hat{\beta}_2/\hat{\beta}_1$ and $\hat{\beta}_3/\hat{\beta}_1$). As we can see, the normalized probit and logit estimates belong to the identification intervals, but they are far from the parameter values used to generate the outcome data.

## 7.2 Women's labor force participation

In this section I present an empirical application based on MROZ data regarding married women's labor force participation (WORK). Let $WORK = 1$ if a woman participates in the labor force; otherwise, let $WORK = 0$. The variables we use to explain labor force participation are education (EDUC), experience (EXPER), age (AGE) and number of children under six years old (KIDS). The descriptive statistics for these variables are presented in Table 2.

[5] $X_1$ and $X_3$ are simulated independently. $X_1$ is generated as a random variable that takes 11 values -5, -4, ..., 4, 5 with probabilities 0.22, 0.11, 0.07, 0.05, 0.03, 0.01, 0.03, 0.05, 0.07, 0.11, 0.25, respectively. $X_3$ is generated as a random variable that takes 8 values 0, 1, ..., 7 with probabilities 0.04, 0.08, 0.18, 0.20, 0.20, 0.18, 0.08, 0.04, respectively.

| | $X_1$ | $X_2 = CONST$ | $X_3$ |
|---|---|---|---|
| Recursive procedure | 1 | **(1, 1.6)** | **(-0.6, -0.42587)** |
| Set of maximum score estimates | 1 | **(1, 1.6)** | **(-0.6, -0.42587)** |
| MRC | 1 | | **(-0.5714, -0.5001)** |
| Probit | 2.9152<br>(2.3794 ,3.4510) | 4.3893<br>(3.4141,5.3646) | -1.5922<br>(-1.9173, -1.2672) |
| Probit (r) | 1 | **1.5057**<br>(1.1711, 1.8402) | **-0.5462**<br>(-0.6577, -0.4347) |
| Logit | 5.4318<br>(4.3201, 6.5435) | 8.0505<br>(6.0992, 10.0019) | -2.9515<br>(-3.6126, -2.2904) |
| Logit (r) | 1 | **1.4821**<br>(1.1229, 1.8414) | **-0.5434**<br>(-0.6651, -0.4217) |

Table 1. Estimation results for the Monte Carlo experiment

| | EDUC | EXPER | AGE | KIDS |
|---|---|---|---|---|
| Mean | 12.287 | 10.631 | 42.538 | 0.238 |
| SD | 2.280 | 8.069 | 8.073 | 0.524 |
| Median | 12 | 9 | 43 | 0 |
| Min. | 5 | 0 | 30 | 0 |
| Max. | 17 | 45 | 60 | 3 |

Table 2. Descriptive statistics for MROZ data

Thus, I estimate the binary response model

$$WORK_i = 1(EDUC_i + \beta_0 + \beta_{EXPER}EXPER_i + \beta_{AGE}AGE_i + \beta_{KIDS}KIDS_i + u_i \geq 0),$$

where I normalize the coefficient corresponding to EDUC. For comparison, I apply several estimation procedures. Table 3 contains the results of these estimations.

There are $N = 753$ observations. After calculating frequency estimates $\hat{P}^l$ and combining data for women with identical characteristics, I obtain 670 points in the support. Based on $\hat{P}^l$, I construct a system of inequalities as usual. This system has no solution, so I employ the methods suggested in section 5.2.1 for dealing with misspecification. A possible cause of model misspecification is the omission of variables that may affect women's labor force participation, such as husband' wages or husbands' attitudes toward women working outside the home. Sampling errors may be another reason why the system has no solution.

| | CONST | EXPER | AGE | KIDS |
|---|---|---|---|---|
| MS | (2.5972, 2.7714) | (0.92361, 0.9433) | (-0.48571, -0.47917) | (-5.1429, -4.9931) |
| MNCE | (7.3514, 7.8447) | (1.0412, 1.0651) | (-0.608, -0.59513) | (-12.536, -8.7) |
| MGCE | (6.9065, 6.9224) | (0.63395, 0.63427) | (-0.51237, -0.51201) | (-7.5439, -7.5389) |
| MRC | | (0.83343, 0.85997) | (-0.59998, -0.58065) | (-8.7897, -8.5005) |
| Probit(r) | 7.68301 | 0.65254 | -0.54431 | -7.86264 |
| | (0.0189, 15.3471) | (0.5247, 0.7804) | (-0.6799, -0.4088) | (-9.8915, -5.8338) |
| Logit(r) | 6.96944 | 0.65890 | -0.53036 | -7.68174 |
| | (-0.6815, 14.6203) | (0.5215, 0.7963) | (-0.6686, -0.3921) | (-9.7471, -5.6164) |
| OLS(r) | 23.47147 | 0.68967 | -0.57315 | -8.20569 |
| | (15.6655, 31.2775) | (0.5134, 0.7311) | (-0.6355, -0.3987) | (-9.1492, -5.6577) |
| LAD(r) | 26.53767 | 0.78495 | -0.68817 | -10.40862 |
| | (17.1571, 35.9182) | (0.6402, 0.9297) | (-0.8457, -0.5306 ) | (-12.7282, -8.089) |

Table 3. Estimation of labor force participation

MS stands for the maximum score estimation method. The set of the maximum score estimates of $\beta = (\beta_0, \beta_{EXPER}, \beta_{AGE}, \beta_{KIDS})$ is the union of several disjoint convex polyhedra, and thus, the set of the maximum score estimates for an individual coefficient is a union of disjoint intervals. Bounds reported in Table 3 are the sharp bounds on such a union of disjoint intervals for each individual coefficient. MNCE stands for the method of minimal number of classification errors described in section 5.2.1; the set of the MNCE estimates of $\beta$ is the union of several disjoint convex polyhedra, and the interpretation of the reported interval bounds for each coefficient is analogous to that in the MS method. MGCE stands for the minimal general classification error method outlined in section 5.2.2. The set of the MGCE estimates of $\beta$ is a convex polyhedron, and thus, the set of the MGCE estimates for each individual coefficient is a connected interval. MRC stands for the maximum rank correlation method. The set of the maximum rank correlation estimates is the union of several disjoint convex polyhedra (in $\Re^3$), and the interpretation of the MRC interval bounds shown in Table 3 is similar to that in the MS method.

I also include normalized probit, logit, OLS and LAD estimates (ratios $\hat{\beta}_0/\hat{\beta}_{EDUC}$, $\hat{\beta}_{EXPER}/\hat{\beta}_{EDUC}$, $\hat{\beta}_{AGE}/\hat{\beta}_{EDUC}$ and $\hat{\beta}_{KIDS}/\hat{\beta}_{EDUC}$).

As we can see, the results produced by methods MS, MNCE, MGCE and MRC are in certain sense consistent with each other. For each regressor, the method MGCE provides shorter intervals than methods MS, MNCE and MRC. This does not come as a surprise, because, first of all, MGCE finds only a subset of separating hyperplanes, and, second, it can be shown that this subset always lies in a hyperplane in the space $\Re^{k-1}$ (in our case, in the space $R^4$).

# 8 Conclusion

In this paper, I examine binary response models when the regressors have discrete support. Ignoring the continuity conditions sufficient for point identification can lead to unsound and misleading inference results on the parameter of interest.

Given these concerns, it is critical to seek a complete characterization of the parameters that fit the model. This paper provides such a characterization for semiparametric binary response models. I offer a recursive procedure to find the sharp bounds on the parameter's identification set. A big advantage of this procedure is the ease of implementation. Moreover, it allows us to explore other aspects of identification, such as the extrapolation problem or changes in the identification set when one regressor's support becomes increasingly dense. Furthermore, the procedure can be used in single-index models with a monotone link function and in ordered-response models.

I go beyond the identification issue by investigating the estimation of the identification region and examining model's misspecification, which I approach in several different ways and provide insight into its possible causes and consequences. I also present an empirical application that compares several estimation techniques and argue that the results critically depend on our preferences for a certain estimation approach.

Several unresolved issues would benefit from future research. It is interesting to look deeper into model's misspecification. Studies that develop tests for misspecification would be particularly useful. When the identification set estimated from a random sample is empty, for instance, we would like to have a test that would allow us to determine whether misspecification or random sampling are behind this problem. Another worthwhile extension would be to learn how to construct finite-sample confidence sets for the identification region.

Despite the several issues that remain to be explored, this paper enhances our understanding of the structure and properties of the identification region in binary response models with discrete regressors. It also provides empirical economists with another avenue for using semiparametric methods when data do not satisfy the sufficient conditions for point identification.

# 9 Appendix

## 9.1 Examples of the recursive procedure

**Example 9.1.** Consider the following system of inequalities with three unknown variables:

$$\begin{aligned} -b_2 + 3b_3 - 4b_4 &\geq 0 \\ 4 - b_2 \qquad\qquad &\geq 0 \\ 2 + b_2 - 2b_3 + 6b_4 &\geq 0 \\ b_2 \qquad + 2b_4 &\geq 0 \\ -1 - b_2 \qquad - 5b_4 &\geq 0. \end{aligned} \tag{9.1}$$

To eliminate variable $b_2$ from this system, rewrite it as

$$\begin{aligned} 3b_3 - 4b_4 &\geq b_2 \\ 4 &\geq b_2 \\ b_2 &\geq -2 + 2b_3 - 6b_4 \\ b_2 &\geq -2b_4 \\ -1 - 5b_4 &\geq b_2 \end{aligned}$$

and obtain that

$$\begin{aligned} 3b_3 - 4b_4 &\geq -2 + 2b_3 - 6b_4 \\ 3b_3 - 4b_4 &\geq -2b_4 \\ 4 &\geq -2 + 2b_3 - 6b_4 \\ 4 &\geq -2b_4 \\ -1 - 5b_4 &\geq -2 + 2b_3 - 6b_4 \\ -1 - 5b_4 &\geq -2b_4, \end{aligned}$$

or, equivalently,

$$\begin{aligned} 2 + b_3 + 2b_4 &\geq 0 \\ 6 - 2b_3 + 6b_4 &\geq 0 \\ 1 - 2b_3 + b_4 &\geq 0 \\ 3b_3 - 2b_4 &\geq 0 \\ 4 \qquad + 2b_4 &\geq 0 \\ -1 \qquad - 3b_4 &\geq 0. \end{aligned} \tag{9.2}$$

**Example 9.2.** Exclude $b_3$ from (9.2) and obtain the following system:

$$\begin{aligned} 10 + 10b_4 &\geq 0 \\ 5 + 5b_4 &\geq 0 \\ 18 + 14b_4 &\geq 0 \\ 3 - b_4 &\geq 0 \\ 4 + 2b_4 &\geq 0 \\ -1 - 3b_4 &\geq 0. \end{aligned} \tag{9.3}$$

From system (9.3), find that $\underline{b}_4 = -1$ and $\overline{b}_4 = -1/3$. Similarly, find sharp bounds on $b_2$ by excluding $b_3$ and $b_4$ from the system: $\underline{b}_2 = 2/3$ and $\overline{b}_2 = 4$. For $b_3$, find that $\underline{b}_3 = -1/2$ and $\overline{b}_3 = 1/3$.

**Example 9.3.** Add to system (9.1) one more inequality:

$$-6 - b_2 + 4b_3 + 10b_4 \geq 0.$$

Then there will be two more inequalities in system (9.2):

$$\begin{aligned} -4 + 2b_3 + 16b_4 &\geq 0 \\ -6 + 4b_3 + 5b_4 &\geq 0, \end{aligned}$$

After eliminating $b_3$ obtain system (9.3) plus two more inequalities

$$\begin{aligned} 2 + 22b_4 &\geq 0 \\ -2 + 7b_4 &\geq 0, \end{aligned}$$

and find that $\underline{b}_4 = 2/7$ and $\overline{b}_4 = -1/3$. Because $\underline{b}_4 > \overline{b}_4$, the system has no solution.

**Example 9.4.** Add to system (9.1) one more inequality:

$$-5 - b_2 + 2b_3 - 6b_4 \geq 0.$$

Then there will be two more inequalities in system (9.2):

$$\begin{aligned} -3 &\geq 0 \\ -6 + 4b_3 + 12b_4 &\geq 0. \end{aligned}$$

There is an obvious contradiction $-3 \geq 0$ in the system, so it does not have solutions.

**Example 9.5.** Drawing from system (9.1), we modify two inequalities and consider the following system:

$$\begin{aligned} -b_2 + 3b_3 - 4b_4 &\geq 0 \\ 4 - b_2 + 4b_3 \qquad &\geq 0 \\ 2 + b_2 - 2b_3 + 6b_4 &\geq 0 \\ b_2 \qquad + 2b_4 &\geq 0 \\ -1 - b_2 + 5b_3 - 5b_4 &\geq 0. \end{aligned} \tag{9.4}$$

Eliminating variable $b_2$ from (9.4), we obtain

$$\begin{aligned} 2 + b_3 + 2b_4 &\geq 0 \\ 3b_3 - 2b_4 &\geq 0 \\ 6 + 2b_3 + 6b_4 &\geq 0 \\ 4 + 4b_3 + 2b_4 &\geq 0 \\ 1 + 3b_3 + b_4 &\geq 0 \\ -1 + 5b_3 - 3b_4 &\geq 0. \end{aligned}$$

All coefficients corresponding to $b_3$ are positive, indicating not only that the system has solutions but also that values of variable $b_3$ are not bounded from above. Although $b_3$ cannot be eliminated from the system, $b_4$ has coefficients of both signs and therefore can be excluded to obtain information about $b_3$. After eliminating $b_4$, we will obtain eight inequalities, from which we will find that $\underline{b}_3 = -1/7$.

If we excluded $b_3$ (9.4) at the first step, we would obtain the following system:

$$\begin{aligned} 6 + 3b_2 + 10b_4 &\geq 0 \\ 8 + b_2 + 12b_4 &\geq 0 \\ 8 + 3b_2 + 20b_4 &\geq 0 \\ b_2 + 2b_4 &\geq 0. \end{aligned}$$

Because all coefficients corresponding to $b_2$ and $b_3$ would be positive, we would conclude that $b_2$ and $b_3$ are bounded from neither below nor above.

## 9.2 Proofs

**Proof of Proposition 3.1.**

Even though an analog of this proposition is established in Solodovnikov (1977), it is worth briefly proving it here.

First, it is obvious that if $(b_2^*, b_3^*, \ldots, b_k^*)$ is a solution to$(S_1)$, then $(b_3^*, \ldots, b_k^*)$ is a solution to $(S_2)$. This fact, applied at each step of the recursive process, implies the first statement of the proposition.

To prove the second statement of the proposition, I first establish that if $(b_3^*, \ldots, b_k^*)$ is a solution to $(S_2)$, then there exists $b_2$ such that $(b_2, b_3^*, \ldots, b_k^*)$ is a solution to $(S_1)$.

Indeed, let $(b_3^*, \ldots, b_k^*)$ be a solution of $(S_2)$. Plug these numbers into $D_i$ and $N_j$ and obtain numbers $D_i^*$ and $N_j^*$ such that

$$N_j^* \geq D_i^*, \quad i = 1, \ldots, I, \quad j = 1, \ldots, J.$$

Take any $b_2$ such that

$$\min_{j=1,\ldots,J} N_j^* \geq b_2 \geq \max_{i=1,\ldots,I} D_i^*.$$

Then $(b_2, b_3^*, \ldots, b_k^*)$ is a solution to $(S_1)$. This result, applied at each step of the recursive procedure, proves the second statement of the proposition.

**Proof of Proposition 3.2**

In the system

$$\begin{aligned} z_{11} + z_{12}b_2 + z_{13}b_3 &\geq 0 \\ &\ldots \\ z_{d1} + z_{d2}b_2 + z_{d3}b_3 &\geq 0, \end{aligned}$$

consider any inequality

$$z_{j1} + z_{j2}b_2 + z_{j3}b_3 \geq 0$$

with $z_{j2} < 0$. This inequality is equivalent to

$$-\frac{z_{j1}}{z_{j2}} - \frac{z_{j3}}{z_{j2}}b_3 \geq b_2.$$

Now consider any inequality

$$z_{i1} + z_{i2}b_2 + z_{i3}b_3 \geq 0,$$

with $z_{i2} > 0$ and rewrite it as

$$b_2 \geq -\frac{z_{i1}}{z_{i2}} - \frac{z_{i3}}{z_{i2}}b_3.$$

Necessarily,

$$-\frac{z_{j1}}{z_{j2}} - \frac{z_{j3}}{z_{j2}}b_3 \geq -\frac{z_{i1}}{z_{i2}} - \frac{z_{i3}}{z_{i2}}b_3;$$

that is,

$$\frac{z_{i1}}{z_{i2}} - \frac{z_{j1}}{z_{j2}} \geq \left(\frac{z_{j3}}{z_{j2}} - \frac{z_{i3}}{z_{i2}}\right)b_3.$$

If

$$\frac{z_{j3}}{z_{j2}} - \frac{z_{i3}}{z_{i2}} > 0,$$

then

$$b_3 \leq \frac{\frac{z_{i1}}{z_{i2}} - \frac{z_{j1}}{z_{j2}}}{\frac{z_{j3}}{z_{j2}} - \frac{z_{i3}}{z_{i2}}} = \frac{z_{i1}z_{j2} - z_{j1}z_{i2}}{z_{j3}z_{i2} - z_{i3}z_{j2}} = -\frac{\begin{vmatrix} z_{j1} & z_{j2} \\ z_{i1} & z_{i2} \end{vmatrix}}{\begin{vmatrix} z_{j3} & z_{j2} \\ z_{i3} & z_{i2} \end{vmatrix}}, \tag{9.5}$$

where

$$\begin{vmatrix} z_{j3} & z_{j2} \\ z_{i3} & z_{i2} \end{vmatrix} = z_{i2}z_{j2}\left(\frac{z_{j3}}{z_{j2}} - \frac{z_{i3}}{z_{i2}}\right) < 0. \tag{9.6}$$

Because (9.5) holds for an arbitrary $i$ and $j$ such that $z_{j2} < 0$, $z_{i2} > 0$ and (9.6) are satisfied, then

$$b_3 \leq \min_{i,j}\left\{ -\frac{\begin{vmatrix} z_{j1} & z_{j2} \\ z_{i1} & z_{i2} \end{vmatrix}}{\begin{vmatrix} z_{j3} & z_{j2} < 0 \\ z_{i3} & z_{i2} > 0 \end{vmatrix} < 0} \right\}.$$

Similarly, prove that

$$b_3 \geq -\frac{\begin{vmatrix} z_{j1} & z_{j2} \\ z_{i1} & z_{i2} \end{vmatrix}}{\begin{vmatrix} z_{j3} & z_{j2} \\ z_{i3} & z_{i2} \end{vmatrix}}$$

for any $i$ and $j$ such that $z_{j2} < 0$, $z_{i2} > 0$ and

$$\begin{vmatrix} z_{j3} & z_{j2} \\ z_{i3} & z_{i2} \end{vmatrix} = z_{i2}z_{j2}\left(\frac{z_{j3}}{z_{j2}} - \frac{z_{i3}}{z_{i2}}\right) > 0;$$

that is,

$$b_3 \geq \max_{i,j}\left\{ -\frac{\begin{vmatrix} z_{j1} & z_{j2} \\ z_{i1} & z_{i2} \end{vmatrix}}{\begin{vmatrix} z_{j3} & z_{j2} < 0 \\ z_{i3} & z_{i2} > 0 \end{vmatrix} > 0} \right\}.$$

**Proof of Proposition 3.3**

The proof proceeds by induction on $k$. As has been proved above, this proposition holds for $k = 3$. Suppose that it also holds for some value $k$. For this case, let us prove that it holds for $k + 1$ as well. Consider system

$$\begin{aligned} z_{11} + z_{12}b_2 + \ldots + z_{1k}b_k + z_{1,k+1}b_{k+1} &\geq 0 \\ z_{21} + z_{22}b_2 + \ldots + z_{2k}b_k + z_{2,k+1}b_{k+1} &\geq 0 \\ &\ldots \\ z_{n1} + z_{n2}b_2 + \ldots + z_{nk}b_k + z_{n,k+1}b_{k+1} &\geq 0, \end{aligned}$$

and apply the recursive algorithm to exclude $b_2$ from the system. The new system consists of inequalities of the form

$$\left(\frac{z_{i1}}{z_{i2}} - \frac{z_{j1}}{z_{j2}}\right) + \left(\frac{z_{i3}}{z_{i2}} - \frac{z_{j3}}{z_{j2}}\right) b_3 + \ldots + \left(\frac{z_{i,k+1}}{z_{i2}} - \frac{z_{j,k+1}}{z_{j2}}\right) b_{k+1} \geq 0,$$

where $z_{i2} > 0$ and $z_{j2} < 0$. Let us write this system as

$$r_{l1} + r_{l3}b_3 + \ldots + r_{l,k+1}b_{k+1} \geq 0, \quad l = 1, \ldots, n_1.$$

Let $\tilde{A}_d$, $d \geq 1$ stand for the determinants corresponding to this new system and $A_d$ stand for the determinants corresponding to the original system. Let us show that $\tilde{A}_d$ is determined by $2^{d+1}$ indices $i_1$, $j_1$, $\ldots$, $i_{2^d}$, $j_{2^d}$ and that

$$\tilde{A}_d(m, i_1, j_1, \ldots, i_{2^d}, j_{2^d}) = \frac{1}{z_{i_1 2} z_{j_1 2} \ldots z_{i_{2^d} 2} z_{j_{2^d} 2}} A_{d+1}(m+1, i_1, j_1, \ldots, i_{2^d}, j_{2^d}).$$

To prove this, use the induction method. Consider $d = 1$:

$$\tilde{A}_1(m, l_1, l_2) = \begin{vmatrix} r_{l_2,m+1} & r_{l_2,3} \\ r_{l_1,m+1} & r_{l_1,3} \end{vmatrix}.$$

Inequality $l_1$ was obtained from some inequalities $i_1$ and $j_1$ of the original system. Similarly, inequality $l_2$ has some corresponding inequalities $i_2$ and $j_2$. Then

$$\tilde{A}_1(m, l_1, l_2) = \tilde{A}_1(m, i_1, j_1, i_2, j_2) = \begin{vmatrix} r_{l_2,m+1} & r_{l_2,3} \\ r_{l_1,m+1} & r_{l_1,3} \end{vmatrix} =$$

$$= \frac{1}{z_{i_1 2} z_{j_1 2} z_{i_2 2} z_{j_2 2}} \begin{vmatrix} \begin{vmatrix} z_{j_1 m+1} & z_{j_1 2} \\ z_{i_1 m+1} & z_{i_1 2} \end{vmatrix} & \begin{vmatrix} z_{j_1 3} & z_{j_1 2} \\ z_{i_1 3} & z_{i_1 2} \end{vmatrix} \\ & \\ \begin{vmatrix} z_{j_2 m+1} & z_{j_2 2} \\ z_{i_2 m+1} & z_{i_2 2} \end{vmatrix} & \begin{vmatrix} z_{j_2 3} & z_{j_2 2} \\ z_{i_2 3} & z_{i_2 2} \end{vmatrix} \end{vmatrix} =$$

$$= \frac{1}{z_{i_1 2} z_{j_1 2} z_{i_2 2} z_{j_2 2}} \begin{vmatrix} A_1(m+1, i_1, j_1) & A_1(3, i_1, j_1) \\ A_1(m+1, i_2, j_2) & A_1(3, i_2, j_2) \end{vmatrix}.$$

Thus, for $d = 1$ the statement is true. Suppose that it is also true for some $d-1$. Let us prove that in this case, it is also true for $d$. Because $\tilde{A}_{d-1}$ depends on $2^{d-1}$ indices, then

$$\tilde{A}_d(m, \ldots, ) = \begin{vmatrix} \tilde{A}_{d-1}(m, i_1, \ldots, j_{2^{d-1}}) & \tilde{A}_{d-1}(k+1, i_1, \ldots, j_{2^{d-1}}) < 0 \\ \tilde{A}_{d-1}(m, i_{2^{d-1}+1}, \ldots, j_{2^d}) & \tilde{A}_{d-1}(k+1, i_{2^{d-1}+1}, \ldots, j_{2^d}) > 0 \end{vmatrix}$$

depends on $2^d$ indices. For $d-1$, the statement of the lemma is true. Therefore,

$$\tilde{A}_d(m, i_1, \ldots, j_{2^d}) =$$

$$= \frac{1}{z_{i_1 2} \ldots z_{j_{2^{d-1}} 2} z_{i_{2^{d-1}+1} 2} \ldots z_{j_{2^d} 2}} \begin{vmatrix} A_d(m+1, i_1, \ldots, j_{2^{d-1}}) & A_d(d+2, i_1, \ldots, j_{2^{d-1}}) < 0 \\ A_d(m+1, i_{2^{d-1}+1}, \ldots, j_{2^d}) & A_d(d+2, i_{2^{d-1}+1}, \ldots, j_{2^d}) > 0 \end{vmatrix} =$$

$$= \frac{1}{z_{i_1 2} z_{j_1 2} \ldots z_{i_{2^d} 2} z_{j_{2^d} 2}} A_{d+1}(m+1, i_1, j_1, \ldots, i_{2^d}, j_{2^d}).$$

Because

$$\frac{\tilde{A}_{k-2}(1, i_1, \ldots, j_{2^{k-2}})}{\tilde{A}_{k-2}(k, i_1, \ldots, j_{2^{k-2}})} = = \frac{A_{k-1}(1, i_1, \ldots, j_{2^{k-2}})}{A_{k-1}(k+1, i_1, \ldots, j_{2^{k-2}})},$$

then we conclude that the formula is true for $b_{k+1}$.

**Proof of Proposition 4.1**

This proof is based on the symmetrical property of the formulas for $b_3^l$ and $b_3^u$ in Proposition 3.2. According to these formulas,

$$b_{3U} \leq -\frac{\begin{vmatrix} z_{j_2,1} & z_{j_2,2} \\ z_{i_1,1} & z_{i_1,2} \end{vmatrix}}{\begin{vmatrix} z_{j_2,3} & z_{j_2,2} \\ z_{i_1,3} & z_{i_1,2} \end{vmatrix}} = \frac{\begin{vmatrix} z_{j_2,1} & -z_{i_2,2} \\ z_{i_1,1} & z_{i_1,2} \end{vmatrix}}{\begin{vmatrix} z_{i_2,3} & z_{i_2,2} \\ z_{i_1,3} & z_{i_1,2} \end{vmatrix}},$$

$$b_{3L} \geq -\frac{\begin{vmatrix} z_{j_1,1} & z_{j_1,2} \\ z_{i_2,1} & z_{i_2,2} \end{vmatrix}}{\begin{vmatrix} z_{j_1,3} & z_{j_1,2} \\ z_{i_2,3} & z_{i_2,2} \end{vmatrix}} = -\frac{\begin{vmatrix} z_{j_1,1} & -z_{i_1,2} \\ z_{i_2,1} & z_{i_2,2} \end{vmatrix}}{\begin{vmatrix} z_{i_2,3} & z_{i_2,2} \\ z_{i_1,3} & z_{i_1,2} \end{vmatrix}},$$

and, hence,

$$b_{3U} - b_{3L} \leq \frac{\begin{vmatrix} z_{j_1,1} & -z_{i_1,2} \\ z_{i_2,1} & z_{i_2,2} \end{vmatrix} + \begin{vmatrix} z_{j_2,1} & -z_{i_2,2} \\ z_{i_1,1} & z_{i_1,2} \end{vmatrix}}{\begin{vmatrix} z_{i_2,3} & z_{i_2,2} \\ z_{i_1,3} & z_{i_1,2} \end{vmatrix}} = \frac{z_{i_2,2}(z_{j_1,1} + z_{i_1,1}) + z_{i_1,2}(z_{j_2,1} + z_{i_2,1})}{\begin{vmatrix} z_{i_2,3} & z_{i_2,2} \\ z_{i_1,3} & z_{i_1,2} \end{vmatrix}} \leq$$

$$\leq \frac{\Delta(z_{i_2,2} + z_{i_1,2})}{\begin{vmatrix} z_{i_2,3} & z_{i_2,2} \\ z_{i_1,3} & z_{i_1,2} \end{vmatrix}} = \Delta \frac{\begin{vmatrix} 1 & -z_{i_2,2} \\ 1 & z_{i_1,2} \end{vmatrix}}{\begin{vmatrix} z_{i_2,3} & z_{i_2,2} \\ z_{i_1,3} & z_{i_1,2} \end{vmatrix}}.$$

**Proof of Corollary 4.2**

Suppose that $x_2 x_3^* - x_2^* x_3 > 0$, and consider the following four cases.

**Case 1:** $x_2 > 0$**,** $x_2^* > 0$. Define

$$(z_{i_1,1}, z_{i_1,2}, z_{i_1,3}) = (x_1, x_2, x_3), \quad (z_{j_1,1}, z_{j_1,2}, z_{j_1,3}) = (-\tilde{x}_1, -x_2, -x_3),$$
$$(z_{i_2,1}, z_{i_2,2}, z_{i_2,3}) = (x_1^*, x_2^*, x_3^*), \quad (z_{j_2,1}, z_{j_2,2}, z_{j_2,3}) = (-\tilde{x}_1^*, -x_2^*, -x_3^*).$$

Then all conditions in Proposition 4.1 are satisfied. Therefore,

$$b_{3U} - b_{3L} \leq \Delta \frac{x_2 + x_2^*}{x_2 x_3^* - x_2^* x_3}.$$

**Case 2:** $x_2 > 0$, $x_2^* < 0$. Define

$$(z_{i_1,1}, z_{i_1,2}, z_{i_1,3}) = (-\tilde{x}_1^*, -x_2^*, -x_3^*), \quad (z_{j_1,1}, z_{j_1,2}, z_{j_1,3}) = (x_1^*, x_2^*, x_3^*)$$
$$(z_{i_2,1}, z_{i_2,2}, z_{i_2,3}) = (x_1, x_2, x_3), \quad (z_{j_2,1}, z_{j_2,2}, z_{j_2,3}) = (-\tilde{x}_1, -x_2, -x_3)$$

Then all condition in Proposition 4.1 are satisfied. Therefore,

$$b_{3U} - b_{3L} \leq \Delta \frac{x_2 - x_2^*}{x_2 x_3^* - x_2^* x_3}.$$

**Case 3:** $x_2 < 0$, $x_2^* > 0$. Define

$$(z_{i_1,1}, z_{i_1,2}, z_{i_1,3}) = (x_1^*, x_2^*, x_3^*), \quad (z_{j_1,1}, z_{j_1,2}, z_{j_1,3}) = (-\tilde{x}_1^*, -x_2^*, -x_3^*),$$
$$(z_{i_2,1}, z_{i_2,2}, z_{i_2,3}) = (-\tilde{x}_1, -x_2, -x_3), \quad (z_{j_2,1}, z_{j_2,2}, z_{j_2,3}) = (x_1, x_2, x_3).$$

Then all condition in Proposition 4.1 are satisfied. Therefore,

$$b_{3U} - b_{3L} \leq \Delta \frac{-x_2 + x_2^*}{x_2 x_3^* - x_2^* x_3}$$

**Case 3:** $x_2 < 0$, $x_2^* < 0$. Define

$$(z_{i_1,1}, z_{i_2,2}, z_{i_2,3}) = (-\tilde{x}_1, -x_2, -x_3), \quad (z_{j_1,1}, z_{j_1,2}, z_{j_1,3}) = (x_1, x_2, x_3),$$
$$(z_{i_2,1}, z_{i_2,2}, z_{i_2,3}) = (-\tilde{x}_1^*, -x_2^*, -x_3^*), (z_{j_2,1}, z_{j_2,2}, z_{j_2,3}) = (x_1^*, x_2^*, x_3^*).$$

Then all condition in Proposition 4.1 are satisfied. Therefore,

$$b_{3U} - b_{3L} \leq \Delta \frac{-x_2 - x_2^*}{x_2 x_3^* - x_2^* x_3}.$$

The case in which $x_2 x_3^* - x_2^* x_3 < 0$ can be considered in a similar way.

**Proof of Proposition 4.3**

The proof of Proposition 4.3 is based on the symmetrical property of the formulas for $b_{4L}$ and $b_{4U}$. According to the formulas in Proposition 3.3,

$$b_{4L} \geq -\frac{C_1}{D}, \quad b_{4U} \leq \frac{C_2}{D},$$

where

$$C_1 = \begin{vmatrix} \begin{vmatrix} z_{j_2}1 & z_{j_2}2 \\ z_{i_1}1 & z_{i_1}2 \end{vmatrix} & \begin{vmatrix} z_{j_2}3 & z_{j_2}2 \\ z_{i_1}3 & z_{i_1}2 \end{vmatrix} \\ & \\ \begin{vmatrix} z_{j_3}1 & z_{j_3}2 \\ z_{i_4}1 & z_{i_4}2 \end{vmatrix} & \begin{vmatrix} z_{j_3}3 & z_{j_3}2 \\ z_{i_4}3 & z_{i_4}2 \end{vmatrix} \end{vmatrix}, \quad C_2 = \begin{vmatrix} \begin{vmatrix} z_{j_4}1 & z_{j_4}2 \\ z_{i_3}1 & z_{i_3}2 \end{vmatrix} & \begin{vmatrix} z_{j_4}3 & z_{j_4}2 \\ z_{i_3}3 & z_{i_3}2 \end{vmatrix} \\ & \\ \begin{vmatrix} z_{j_1}1 & z_{j_1}2 \\ z_{i_2}1 & z_{i_2}2 \end{vmatrix} & \begin{vmatrix} z_{j_1}3 & z_{j_1}2 \\ z_{i_2}3 & z_{i_2}2 \end{vmatrix} \end{vmatrix}.$$

Then

$$b_{4U}-b_{4L}\le\frac{C_1}{D}+\frac{C_2}{D}=\frac{1}{D}\begin{vmatrix} z_{j_4}1 & z_{j_4}2\\ z_{i_3}1 & z_{i_3}2\end{vmatrix}\cdot\begin{vmatrix} z_{i_2}2 & z_{i_2}3\\ z_{j_1}2 & z_{j_1}3\end{vmatrix}-\frac{1}{D}\begin{vmatrix} z_{i_2}1 & z_{i_2}2\\ z_{j_1}1 & z_{j_1}2\end{vmatrix}\cdot\begin{vmatrix} z_{j_4}2 & z_{j_4}3\\ z_{i_3}2 & z_{i_3}3\end{vmatrix}+$$

$$+\frac{1}{D}\begin{vmatrix} z_{j_2}1 & z_{j_2}2\\ z_{i_1}1 & z_{i_1}2\end{vmatrix}\cdot\begin{vmatrix} z_{i_4}2 & z_{i_4}3\\ z_{j_3}2 & z_{j_3}3\end{vmatrix}-\frac{1}{D}\begin{vmatrix} z_{i_4}1 & z_{i_4}2\\ z_{j_3}1 & z_{j_3}2\end{vmatrix}\cdot\begin{vmatrix} z_{j_2}2 & z_{j_2}3\\ z_{i_1}2 & z_{i_1}3\end{vmatrix}=$$

$$=\frac{1}{D}(z_{i_1}2(z_{j_2}1+z_{i_2}1)+z_{i_2}2(z_{j_1}1+z_{i_1}1))\begin{vmatrix} z_{i_4}2 & z_{i_4}3\\ z_{j_3}2 & z_{j_3}3\end{vmatrix}+\frac{1}{D}(z_{i_3}2(z_{j_4}1+z_{i_4}1)+z_{i_4}2(z_{j_3}1+z_{i_3}1))\begin{vmatrix} z_{j_2}2 & z_{j_2}3\\ z_{i_1}2 & z_{i_1}3\end{vmatrix}\le$$

$$\le\frac{\Delta}{D}(z_{i_1}2+z_{i_2}2)\begin{vmatrix} z_{i_4}2 & z_{i_4}3\\ z_{j_3}2 & z_{j_3}3\end{vmatrix}+\frac{\Delta}{D}(z_{i_3}2+z_{i_4}2)\begin{vmatrix} z_{j_2}2 & z_{j_2}3\\ z_{i_1}2 & z_{i_1}3\end{vmatrix}=\frac{\Delta}{D}\left|\begin{array}{ccc}\begin{vmatrix} 1 & -z_{i_4,2}\\ 1 & z_{i_3,2}\end{vmatrix} & - & \begin{vmatrix} z_{i_4,3} & z_{i_4,2}\\ z_{i_3,3} & z_{i_3,2}\end{vmatrix}\\ \\ \begin{vmatrix} 1 & -z_{i_2,2}\\ 1 & z_{i_1,2}\end{vmatrix} & & \begin{vmatrix} z_{i_2,3} & z_{i_2,2}\\ z_{i_1,3} & z_{i_1,2}\end{vmatrix}\end{array}\right|.$$

**Proof of Proposition 5.1**

The maximal possible value of $S^{ms}$ is $2\sum_{x^l\in S(X)}q^l|P^l-0.5|$. Evidently, this value is attained on set $B$; that is, $B\subseteq B^{ms}$. On the other hand, if $P^l\neq 0.5$ for any $x^l\in S(X)$, then $\max_{b:b_1=1}S^{ms}(b)=2\sum_{x^l\in S(X)}q^l|P^l-0.5|$ implies that there exists $b$ that solves the system of linear inequalities constructed according to the rule

$$P^l\ge 0.5\quad\Leftrightarrow\quad x^lb\ge 0,\quad l=1,\dots,d,$$

which, in turn, defines set $B$. Thus, $B^{ms}\subseteq B$, and, therefore, $B=B^{ms}$.

Now suppose that $P^l=0.5$ for some $l$. If, for instance, $P^1=0.5$, then any $b$ satisfying the system of inequalities

$$P^l\ge 0.5\quad\Leftrightarrow\quad x^lb\ge 0,\quad l=2,\dots,d$$

also gives a maximal value to $S^{ms}$. So, in this case, set $B^{ms}$ is larger than $B$; that is, $B\subset B^{ms}$.

**Proof of Theorem 6.1**

$$\widehat{B}\neq B\quad\Rightarrow\quad\exists(x^l\in S(X))\quad sgn(\hat{P}^l-0.5)\neq sgn(P^l-0.5),$$

therefore,

$$Pr(\widehat{B}\neq B)\le\sum_{x^l\in S(X)}P(sgn(\hat{P}^l-0.5)\neq sgn(P^l-0.5)).$$

Because

$$\begin{aligned}P^l>0.5\quad\text{and}\quad sgn(\hat{P}^l-0.5)\neq sgn(P^l-0.5)\quad&\Rightarrow\quad P^l-\hat{P}^l>P^l-0.5,\\ P^l<0.5\quad\text{and}\quad sgn(\hat{P}^l-0.5)\neq sgn(P^l-0.5)\quad&\Rightarrow\quad \hat{P}^l-P^l>0.5-P^l,\end{aligned}$$

then

$$\forall (x^l \in S(X)) \quad Pr(sgn(\hat{P}^l - 0.5) \neq sgn(P^l - 0.5)) \leq Pr(|\hat{P}^l - P^l| > |0.5 - P^l|).$$

The consistency property (6.1) and the fact that $|0.5 - P^l| > 0$ for any $x^l \in S(X)$ imply

$$\forall (x^l \in S(X)) \quad Pr(|\hat{P}^l - P^l| > |0.5 - P^l|) \to 0 \text{ as } N \to \infty.$$

It is evident now that

$$Pr(\widehat{B} \neq B) \to 0 \text{ as } N \to \infty.$$

Since

$$R(\widehat{B}) \neq R(B) \quad \Rightarrow \quad \widehat{B} \neq B,$$

then

$$Pr(R(\widehat{B}) \neq R(B)) \leq Pr(\widehat{B} \neq B) \to 0 \text{ as } N \to \infty.$$

**Proof of Theorem 6.2**

$$Pr(\widehat{B} \neq B) \leq \sum_{x^l \in S(X)} P(sgn(\hat{P}^l - 0.5 + \epsilon_N) \neq sgn(P^l - 0.5)).$$

If $P^l > 0.5$, then

$$Pr(sgn(\hat{P}^l - 0.5 + \epsilon_N) \neq sgn(P^l - 0.5)) = Pr(P^l - \hat{P}^l > P^l - 0.5 + \epsilon_N) \leq$$
$$\leq Pr(P^l - \hat{P}^l > P^l - 0.5) \to 0 \text{ as } N \to \infty.$$

Let $P^l < 0.5$. Convergence $\epsilon_N \to 0$ implies that, when $N$ is large enough, $0.5 - P^l - \epsilon_N > \delta$ for some $\delta > 0$, and, consequently,

$$Pr(sgn(\hat{P}^l - 0.5 + \epsilon_N) \neq sgn(P^l - 0.5)) = Pr(\hat{P}^l - P^l > 0.5 - P^l - \epsilon_N) \leq$$
$$\leq Pr(\hat{P}^l - P^l > \delta) \to 0 \quad \text{as} \quad N \to \infty.$$

If $P^l = 0.5$, then

$$Pr(sgn(\hat{P}^l - 0.5 + \epsilon_N) \neq sgn(P^l - 0.5)) = Pr(P^l - \hat{P}^l > \epsilon_N) = Pr(\epsilon_N^{-1}(P^l - \hat{P}^l) > 1).$$

(6.3) and (6.4) imply that

$$\epsilon_N^{-1}(P^l - \hat{P}^l) = (\epsilon_N \tau_N)^{-1} \tau_N (P^l - \hat{P}^l) \xrightarrow{p} 0 \text{ as } N \to \infty$$

and, thus,

$$Pr(\epsilon_N^{-1}(P^l - \hat{P}^l) > 1) \to 0 \quad \text{as} \quad N \to \infty.$$

Finally,

$$Pr(R(\widehat{B}) \neq R(B)) \leq Pr(\widehat{B} \neq B) \to 0 \text{ as } N \to \infty.$$

**Proof of Corollary 6.3.**

Suppose the conditions of Theorem 6.1 or Theorem 6.2 hold. For any $\epsilon > 0$,

$$r_N H(\widehat{B}, B) \geq \epsilon \quad \Rightarrow \quad H(\widehat{B}, B) \neq 0 \quad \Rightarrow \quad \widehat{B} \neq B.$$

Therefore,

$$Pr(r_N H(\widehat{B}, B) \geq \epsilon) \leq Pr(\widehat{B} \neq B) \to 0 \quad \text{as} \quad N \to \infty.$$

**Proof of Theorem 6.4**

$$Pr(\widehat{B} \neq B) \leq \sum_{x^l \in S(X)} Pr(sgn(\hat{P}^l - 0.5) \neq sgn(P^l - 0.5))$$

Denote

$$V_N(x^l) = \sum_{i=1}^{N} (2y_i - 1)1(x_i = x^l).$$

Then

$$\hat{P}^l \geq 0.5 \quad \Leftrightarrow \quad V_N(x^l) \geq 0.$$

Note that random variable $(2y_i - 1)1(x_i = x^l)$ takes values 1, 0 and -1 with probabilities $P^l q^l$, $1 - q^l$ and $(1 - P^l)q^l$, respectively. Its expected value is $(2P^l - 1)q^l$.

Let $P^l > 0.5$. Then $Pr(sgn(\hat{P}^l - 0.5) \neq sgn(P^l - 0.5)) = Pr(V_N(x^l) < 0)$. By Hoeffding's inequality,

$$Pr(V_N(x^l) < 0) = Pr(V_N(x^l) - N(2P^l - 1)q^l < -N(2P^l - 1)q^l) \leq e^{-N((2P^l-1)q^l)^2/2}.$$

If $P^l < 0.5$, then $Pr(sgn(\hat{P}^l - 0.5) \neq sgn(P^l - 0.5)) = Pr(V_N(x^l) \geq 0)$. By Hoeffding's inequality

$$Pr(V_N(x^l) \geq 0) = Pr(V_N(x^l) - N(2P^l - 1)q^l \geq N(1 - 2P^l)q^l) \leq e^{-N((2P^l-1)q^l)^2/2}.$$

Thus, if $P^l \neq 0.5$,

$$Pr(sgn(\hat{P}^l - 0.5) \neq sgn(P^l - 0.5)) \leq e^{-N((2P^l-1)q^l)^2/2}.$$

Let $\rho < 1$ be such that

$$\rho > \max_{l=1,\ldots,d} e^{-((2P^l-1)q^l)^2/2}.$$

Then, if $P^l \neq 0.5$ for any $x^l$,

$$Pr(\widehat{B} \neq B) = o(\rho^N) \text{ as } N \to \infty,$$

and since $Pr(R(\widehat{B}) \neq R(B)) \leq Pr(\widehat{B} \neq B)$,

$$Pr(R(\widehat{B}) \neq R(B)) = o(\rho^N) \text{ as } N \to \infty$$

too. This proves the first part of the theorem.

Now suppose that there is $x^l$ such that $P^l = 0.5$. Because

$$P^l = 0.5 \text{ and } \hat{P}^l < 0.5 \quad \Rightarrow \quad \widehat{B} \neq B,$$

then

$$Pr(\widehat{B} \neq B) \geq Pr(V_N(x^l) < 0).$$

Note that

$$P^l = 0.5 \quad \Rightarrow \quad Pr(V_N(x^l) < 0) = 0.5(1 - Pr(V_N(x^l) = 0)).$$

If we will find a bound on $Pr(V_N(x^l) = 0)$ from above, we will find a bound on $Pr(V_N(x^l) < 0)$ from below.

$$Pr(V_N(x^l) = 0) = \sum_{j=0}^{[\frac{N}{2}]} C_j^N C_j^{N-j} 0.5^{2j} (q^l)^{2j} (1 - q^l)^{N-2j} = \sum_{j=0}^{[\frac{N}{2}]} C_{2j}^N C_j^{2j} 0.5^{2j} (q^l)^{2j} (1 - q^l)^{N-2j} =$$

$$= (1 - q^l)^N + \sum_{j=1}^{[\frac{N}{2}]} C_{2j}^N C_j^{2j} 0.5^{2j} (q^l)^{2j} (1 - q^l)^{N-2j},$$

where $C_{k_1}^{k_2} = \frac{k_2!}{k_1!(k_2-k_1)!}$. Use the fact that for $j \geq 1$,

$$C_j^{2j} 0.5^{2j} = (-1)^j \frac{(-0.5)(-0.5-1)\dots(-0.5-j+1)}{j!} \leq 0.5$$

to obtain

$$Pr(V_N(x^l) = 0) \leq (1 - q^l)^N + 0.5 \sum_{j=1}^{[\frac{N}{2}]} C_{2j}^N (q^l)^{2j} (1 - q^l)^{N-2j} \leq (1 - q^l)^N + 0.5.$$

Then

$$Pr(V_N(x^l) < 0) = 0.5(1 - Pr(V_N(x^l) = 0)) \geq 0.5(1 - (1 - q^l)^N - 0.5)$$

and

$$Pr(\widehat{B} \neq B) \geq 0.5(1 - (1 - q^l)^N - 0.5) \ \to \ 0.25 \ \text{ as } N \to \infty.$$

# References

[1] Balinski, M.L. (1961). An Algorithm for Finding All Vertices of Convex Polyhedral Sets, *J. Soc. Indust. and Appl. Math.*, 9, 72-88.

[2] Bierens, H.J., and J. Hartog (1988). Non-Linear Regression with Explanatory Variables, with an Application to the Earnings Function, *Journal of Econometrics*, 38, 269-299.

[3] Cavanagh, C., and R.P. Sherman (1998). Rank Estimators for Monotonic Index Models, *Journal of Econometrics*, 84, 351-381.

[4] Chernikova, N.V. (1965). Algorithm for Finding a General Formula for the Nonnegative Solutions of a System of Linear Inequalities, *U.S.S.R. Computational Mathematics and Mathematical Physics*, 5, 228-233.

[5] Chernikova, N.V. (1965). Algorithm for Finding a General Formula for the Nonnegative Solutions of a System of Linear Inequalities, *U.S.S.R. Computational Mathematics and Mathematical Physics*, 5, 228-233.

[6] Chernozhukov, V., H. Hong, and E. Tamer (2007). Estimation and Confidence Regions for Parameter Sets in Econometric Models, *Econometrica*, 75 (5), 1243-1284.

[7] Cortes, C., and V. Vapnik (1995). Support Vector Networks, *Machine Learning*, 20, 1-25.

[8] Cosslett, S.R. (1983). Distribution-Free Maximum Likelihood Estimator of the Binary Choice Model, *Econometrica*, 51, 765-782.

[9] Han, A.K. (1987). Non-parametric Analysis of a Generalized Regression Model, *Journal of Econometrics*, 35, 303-316.

[10] Honore, Bo E., and E. Tamer (2006). Bounds on Parameters in Panel Dynamic Discrete Choice Models, *Econometrica*, 74 (3), 611-629.

[11] Horowitz, J.L. (1992). A Smoothed Maximum Score Estimator for the Binary Response Model, *Econometrica*, 60, 505-531.

[12] Horowitz, J.L. (1998). *Semiparametric Methods in Econometrics*, Lecture Notes in Statistics, Vol. 131: Springer-Verlag New York, Inc.

[13] Ichimura, H. (1993). Semiparametric Least Squares (SLS) and Weighted SLS Estimator of Single-Index Models, *Journal of Econometrics*, 58, 71-120.

[14] Imbens, G.W., and C.F. Manski (2004). Confidence Intervals for Partially Identified Parameters, *Econometrica*, 72 (6), 1845-1857.

[15] Kim, J., and D. Pollard (1990). Cube Root Asymptotics, *Annals of Statistics*, 18, 191-219.

[16] Klein, R.W., and R.H. Spady (1993). An Efficient Semiparametric Estimator of Binary Response Models, *Econometrica*, 61, 387-421.

[17] Kuhn, H. W. (1956). Solvability and Consistency for Linear Equations and Inequalities, *The American Mathematical Monthly*, 63 (4), 217-232.

[18] Kuhn, H.W. and A.W. Tucker (1956). *Linear Inequalities and Related Systems*, Annals of Mathematics Studies 38, Princeton U. Press, Princeton.

[19] Lin, Y. (2000). On the Support Vector Machine, Technical Report No. 1029, University of Wisconsin.

[20] Lin, Y. (2002). Support Vector Machines and the Bayes Rule in Classification, *Data Miniing and Knowledge Discovery*, 6 (3), 259-275.

[21] Magnac, T., and E. Maurin (2005). Partial Identification in Monotone Binary Models: Discrete Regressors and Interval Data, Working Paper, University of Toulouse.

[22] Manas, M., and J. Nedoma (1968). Finding All Vertices of a Convex Polyhedron, *Numerische Mathematik*, 12, 226-229.

[23] Manski, C.F. (1975). Maximum Score Estimation of the Stochastic Utility Model of Choice, *Journal of Econometrics*, 3, 205-228.

[24] Manski, C.F. (1985). Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator, *Journal of Econometrics*, **27**, 313-334.

[25] Manski, C.F. (1988). Identification of Binary Response Models, *Journal of the American Statistical Association*, **83**, No. 403, 729-738.

[26] Manski, C.F. (1990). Nonparametric Bounds on Treatment Effects, *The American Economic Review*, 80 (2), 319-323.

[27] Manski, C.F. (1995). *Identification Problems in the Social Sciences*, Cambridge: Harvard University press.

[28] Manski, C.F. (2003). *Partial Identification of Probability Distributions*, Springer Series in Statistics: Springer-Verlag New York, Inc.

[29] Manski, C.F. (2007). Partial Identification of Counterfactual Choice Probabilities, *International Economic Review*, forthcoming.

[30] Manski, C.F., and E. Tamer (2002). Inference on regressions with Interval Data on a Regressor or Outcome, *Econometrica*, 70, 519-547.

[31] Manski, C.F., and T.S. Thompson (1986). Operational Characteristics of Maximum Score Estimation, *Journal of Econometrics*, 32, 85-108.

[32] Manski, C.F., and T.S. Thompson (1989). Estimation of Best Predictors of Binary Response, *Journal of Econometrics*, 40, 97-123.

[33] Matheiss, T.H. (1973). An Algorithm for the Determination of Irrelevant Constraints and All Vertices in Systems of Linear Inequalities, *Operations Research*, 21 (1), 247-260.

[34] Matheiss, T.H., and David S. Rubin (1980). A Survey and Comparison of Methods for Finding All Vertices of Convex Polyhedral Sets, *Mathematics of Operations Research*, 5 (2), 167-185.

[35] Motzkin, T.S., H. Raiffa, G.L. Thompson, and R.M. Thrall (1953). The Double D Survey and Comparison of Methods for finding All Vertices of Convex Polyhedral Sets, *Mathematics of Operations Research*, 5 (2), 167-185.

[36] Padberg, M. (1999). *Linear Optimization and Extensions*, Springer-Verlag Berlin Heidelberg.

[37] Rockafellar, R.T. (1972). *Convex Analysis*, Princeton University Press.

[38] Rosen, A.M. (2006). Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities, Working Paper CWP25/06, The Institute for Fiscal Studies and Department of Economics, University College London.

[39] Solodovnikov, A.S. (1977). *Sistemy Linejnyh Neravenstv*, Populyarnye lektcii po matematike, Vypusk 48: Izadatelstvo "Nauka", Moskva

[40] Vapnik, V.N. (2000). *The Nature of Statistical Learning Theory*, Statistics for Engineering and Information Science: Springer-Verlag New York, Inc.