NORTHWESTERN UNIVERSITY

Essays on Identification in Econometric Models

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Economics Sciences

By

Tatiana Komarova

EVANSTON, ILLINOIS

December 2008

Abstract

Essays on Identification in Econometric Models

Tatiana Komarova

This dissertation consists of three essays on the identification analysis of econometric models.

The first essay explores the identification question in semiparametric binary response models when all regressors have discrete support. I suggest a recursive procedure that finds sharp bounds on the parameter of interest and can be applied to the analysis of identification sets in other single-index models. Furthermore, I investigate asymptotic properties of estimators of the identification set. I also propose three 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.

The second essay proposes an approach to proving nonparametric identification for the distributions of bidders' values in asymmetric second-price auctions. I consider the case where bidders have independent private values, and the only available data pertain to the winner's identity and to the transaction price. I provide conditions on observable data sufficient to guarantee point identification. I demonstrate how the techniques of the identification proof can be used to obtain identification in generalized competing risks models. Finally, I provide a sieve minimum distance estimator that consistently estimates the underlying valuation distribution of interest.

The third essay analyzes identification in second-price auction within the private values framework when bidders' values are *not* independent. When only the winner's identity and the winning price are observed, neither the joint nor the marginal distribution of bidders' values are point identified but I derive tight bounds on the distribution of values for any subset of bidders. In addition, I investigate how these bounds change when data on other elements of the model become available. Finally, I use the representation of joint distributions through copulas and prove point identification for certain classes of copulas.

Acknowledgement

I would like to express my sincere gratitude to all those who gave me the possibility to complete this thesis.

I am deeply indebted to my supervisor Elie Tamer whose guidance, stimulating suggestions and encouragement invaluably helped me throughout my academic program, and I would not be where I am without him. He was always accessible and willing to help me with my research. As a result, research life became smooth and rewarding for me.

Professors Joel L. Horowitz, Charles F. Manski and Rosa L. Matzkin deserve special thanks as my committee members and advisers. They have offered many important suggestions and comments that substantially improved my work. I also benefited greatly from conversations with Xiaohong Chen, Sokbae Lee and Robert Porter.

I warmly thank Maria Ponomareva and Ija Trapeznikova for their kind support and assistance in preparing this thesis. I am grateful to Maxim Ivanov, Denis Nekipelov and Eugene Orlov for their encouragement, especially during the job market period. I appreciate feedback and comments from Jeffrey Ely, Paul Grieco, Aviv Nevo, Benjamin Passty, Zahra Siddique and Basit Zafar.

I also wish to thank the secretaries in the Economics Department for assisting me in many different ways.

It is difficult to overstate my gratitude to the New Economic School Community. Without them, my ambition to study abroad and my academic career goals could hardly be realized.

My sincere thanks are due to Boris Nikolaevich Sadovsky who taught me mathematics and mentored me for many years.

Last but not least, I want to thank my family who constantly supported me throughout my academic career. They have been there for me every step of the way.

Table of Contents

List of Figures

List of Tables

Chapter 1

Set Identification in Binary Response Models with Discrete Regressors

1.1 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 chapter examines the question of identification in semiparametric binary response models in the absence of continuity. Consider

$$Y = 1(X\beta + U \geq 0), \qquad (BR)$$

where Y is an observable binary outcome, U is an unobservable, real-valued, scalar random variable, β 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):

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

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

I show 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 chapter is to provide a recursive procedure that allows us to easily find sharp bounds on the identification region. Although this approach 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 parameters, which prove useful in analyzing cases when the support of one regressor can become increasingly dense. 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 chapter 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 (M) 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 chapter 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. The third, a modification of a soft margin hyperplane approach in SVMs, lets us determine the extent of misspecification by determining the minimal size of a general classification error. Each method features a crucial property: The set of solutions coincides with the identification set when the model is well specified.

An interesting element of this chapter is that it compares the identification region with sets of solutions given by other methods, for instance, the maximum rank correlation method. In the discussion of single-index models I demonstrate that the maximum rank correlation method may give unreliable results.

Another contribution of this chapter is to explore the estimation of the identification region. Although this chapter focuses on identification, it is of interest to analyze cases where conditional probabilities $P(Y = 1|X = x)$ are not known, but their estimates $\hat{P}(Y = 1|X = x)$ are available. In this situation, we can find estimators of identification sets from a system of linear inequalities that uses $\hat{P}(Y = 1|X = x)$ instead of $P(Y = 1|X = x)$. I show that when the model is well specified, such set estimators converge to the true identification set 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 show that because of the discrete nature of the problem, they are usually conservative.

The last contribution of this chapter is an empirical application. The empirical portion of this chapter consists of two parts. The first 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 estimator of the identification set obtained from the system of inequalities that uses estimated conditional probabilities and the set of maximum score estimates coincides with the identification set. For parameters corresponding to non-constant regressors, I find the set of maximum rank correlation estimates, which turn out to lie inside the identification set but form a much smaller set. I also present normalized probit and logit estimates. Though these estimates are located inside the identification set, they are far from the value of the parameter, which was used to generate the model.

The second empirical part 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. We 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, we 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 chapter 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 recent 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 ϕ_θ 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 chapter is organized as follows. Section 1.2 explains the problem and defines the identification set. Section 1.3 contains the mathematical apparatus, describes the recursive procedure and its applications and draws an analogy to SVMs. Section 2.4 considers misspecification issues and suggests techniques for dealing with them. Section 1.5 analyzes the case in which the discrete support of regressors grows increasingly dense. Section 1.6 discusses identification in single-index models with a monotone link function and compares identification sets in those models with sets of maximum rank correlation estimates. It also describes the identification set in ordered response models. Section 1.7 considers the estimation of the identification set from a sample and statistical inference. Section 1.8 is an empirical section that contains the results of estimations in a Monte Carlo experiment and an MROZ data application. Section 1.9 concludes and outlines ideas for future research.

1.2 Identification set

We 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 (M), the following conditions on the support of X guarantee that β 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_h, $h \in \{1, \ldots, k\}$ with $\beta_h \neq 0$ such that for almost every $\widetilde{x} = (x_1, \ldots, x_{h-1}, x_{h+1}, \ldots, x_k)$, the distribution of X_h 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 β is identified.

The smoothed maximum score method described in Horowitz (1992) and the maximum rank correlation method presented in Han (1987) require these conditions. Klein and Spady's (1993) approach, on the other hand, imposes a stronger assumption: At least one component of X must be a continuous random variable.

It is worth mentioning that Manski (1988) presents other identification results. For instance, under certain conditions, even when β is not identified, the signs of $\beta_1,\ldots,\beta_k$ can be identified. Horowitz (1998) contains a thorough review of identification results for binary response models.

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

$$S(X) = \{x^1, \ldots, x^d\}. \tag{1.2.1}$$

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

$$Pr(Y = 1|X = x) = Pr(U \geq -x\beta|X = x) = 1 - Pr(U < -x\beta|X = x),$$

the median condition implies that

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

Thus, model (BR) together with (M) is equivalent to model (BRM), and the identification problem comes down to solving a system of inequalities. Manski and Thompson (1989) interpret condition (BRM) as a single-crossing condition for response probabilities. Because $S(X)$ contains a finite number of points, the number of inequalities in the system is also finite. If $Pr(Y = 1|X = x^l) \geq 0.5$, 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 $Pr(Y = 1|X = x^l) < 0.5$, 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, we will write it as encompassing 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, we will use normalization $\beta_1 = 1$, along with the notations x^l, which will denote points in the support, and d, which will stand for the number of these points, as in (1.2.1). We will assume that all points in the support are different. Furthermore, we will use q^l to signify the probability of x^l in the population and P^l to indicate conditional probabilities $P(Y = 1|X = x^l)$, assuming that $0 < q^l < 1$ for any l. The parameter's identification set will be denoted as B.

N will stand for the number of observations in a sample, $\hat{q}^l_N$ will denote the sample frequency estimate of q^l and $\hat{P}^l_N$ will signify the estimate of P^l. The sample estimate of B will be denoted as B_N. 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 $z_l = sgn(\hat{P}_N^l - 0.5)x^l$. These cases will be clear from the context.

Theorem 1.1. *If X has support (1.2.1), then parameter β in (BRM) is identified up to a k_0-dimensional convex polyhedron, where $k_0 \leq k-1$.*

Corollary 1.1. *If X has support (1.2.1), then each β_m, $m \neq 1$, in (BRM) is identified up to a connected interval.*

Theorem 1.1 does not need any proof: It is a direct consequence of the definition of convex polyhedra. To prove Corollary 1.1, we note that the identification interval for β_m, $m \neq 1$, is the projection of the identification set on axis x_l. Due to the convexity of B, this interval is a connected set.

A few words about convex polyhedra are in order. By definition, a set $B \subset \Re^{k-1}$ is a convex polyhedron if and only if it is the intersection of a finite number of closed half-spaces in $\Re^{k-1}$ space. Formally, a set $B \subset \Re^{k-1}$ is a convex polyhedron if and only if there is an $m \times (k-1)$ matrix H and a vector h of m real numbers such that $B = \{b \in \Re^{k-1} : \ Hb \leq h\}$.

According to this formal definition, the identification set in our model is not a standard convex polyhedron because the system of linear inequalities defining B contains strict and non-strict inequalities. For our purposes, however, B can be considered a convex polyhedron, keeping in mind that it does not contain some surface points.

The solution set of a finite system of linear inequalities in a finite number of variables is not the only representation of a convex polyhedron. Because every bounded polyhedron has a double description, it can also be described pointwise as the convex hull of a finite number

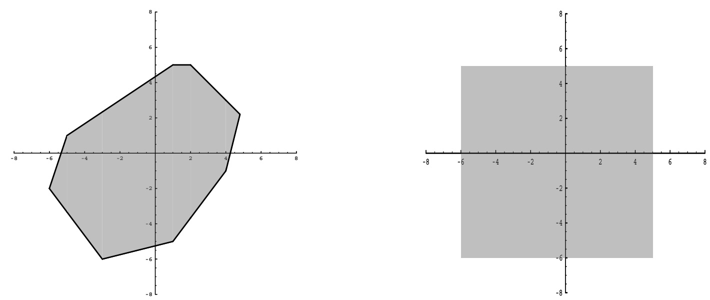


Figure 1.1. Convex polyhedron and its smallest rectangular superset

of points. (The minimal set of such points is called the 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)).

Several authors provide algorithms for finding all the vertices of convex polyhedra when they are described by systems of inequalities. Examples of such works are Motzkin et. al. (1953), Balinski (1961), Chernikova (1965), Manas and Nedoma (1968), Matheiss (1973) and Matheiss and Rubin (1980). Though applicable in practice, these 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 β_m, $m \neq 1$. Its dimension can be smaller than $k-1$ if some β_m, $m \neq 1$, are point identified.

Figure 1.1 shows an identification set on the left and its smallest rectangular superset on the right.

1.3 Mathematical tools

1.3.1 Recursive method

In this section, we describe a recursive procedure that finds identification intervals for β_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), we supplement it by discussing cases in which the system has an unbounded or empty set of solutions and by deriving formulas for parametric bounds.

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 $z_{i2} < 0$, then it is equivalent to

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

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

$$b_2 \geq D_i, \quad i = 1, \ldots, I,$$

$$N_j \geq b_2, \quad j = 1, \ldots, J,$$

$$Z_m \geq 0, \quad m = 1, \ldots, M,$$

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.

Let us illustrate this first step in the following example.

Example 1.3.1. *Consider the following system of inequalities with three unknown variables:*

$$\begin{aligned} -b_2 + 3b_3 - 4b_4 &\geq 0 \\ 4 - b_2 \qquad\qquad &\geq 0 \\ 2 + b_2 - 2b_3 + 6b_4 &\geq 0 \\ b_2 \qquad + 2b_4 &\geq 0 \\ -1 - b_2 \qquad - 5b_4 &\geq 0. \end{aligned} \tag{1.3.1}$$

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,

$$\begin{aligned} 2 + b_3 + 2b_4 &\geq 0 \\ 6 - 2b_3 + 6b_4 &\geq 0 \\ 1 - 2b_3 + b_4 &\geq 0 \\ 3b_3 - 2b_4 &\geq 0 \\ 4 \qquad + 2b_4 &\geq 0 \\ -1 \qquad - 3b_4 &\geq 0. \end{aligned} \tag{1.3.2}$$

The next proposition establishes a relationship between the solutions of (S_1) and (S_2).

Proposition 1.2. *If $(b_2^*, b_3^*, \ldots, b_k^*)$ is a solution of (S_1), then $(b_3^*, \ldots, b_k^*)$ is a solution of (S_2). If $(b_3^*, \ldots, b_k^*)$ is a solution of (S_2), then there exists b_2 such that $(b_2, b_3^*, \ldots, b_k^*)$ is a solution of (S_1).*

We repeat the process above and exclude a variable from (S_2) (for example, b_3) to obtain a system (S_3) of linear inequalities with unknown variables $(b_4, \ldots, b_k)$. We continue repeating this procedure, removing another variable each time, until we obtain a system with only one unknown variable b_k:

$$A_s + B_s b_k \geq 0, \quad s = 1, \ldots, S, \qquad (S_{k-1})$$

$$C_q \geq 0, \quad q = 1, \ldots, Q,$$

where $B_s \neq 0$, $s = 1, \ldots, S$. The statement of Proposition 1.2, applied at each step of the recursive process, implies the following fact.

Proposition 1.3. *If $(b_2^*, b_3^*, \ldots, b_k^*)$ is a solution of (S_1), then b_k^* is a solution of (S_{k-1}). If b_k^* is a solution of (S_{k-1}) , then there exists $(b_2, \ldots, b_{k-1})$ such that $(b_2, b_3, \ldots, b_{k-1}, b_k^*)$ is a solution of (S_1).*

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

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

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

If $\overline{b}_k < \underline{b}_k$ or $C_q < 0$ for some q, then system (S_{k-1}) and, therefore, (S_1) do not have solutions. Otherwise, the set of solutions for (S_{k-1}) is $\left[\underline{b}_k, \overline{b}_k\right]$. From Proposition 1.3, we conclude that $\underline{b}_k$ and $\overline{b}_k$ are sharp bounds for b_k.

Example 1.3.1 (continued) After b_3 is excluded from (1.3.2), we obtain the following system:

$$\begin{aligned} 10 + 10b_4 &\geq 0 \\ 5 + 5b_4 &\geq 0 \\ 18 + 14b_4 &\geq 0 \\ 3 - b_4 &\geq 0 \\ 4 + 2b_4 &\geq 0 \\ -1 - 3b_4 &\geq 0. \end{aligned} \tag{1.3.3}$$

From system (1.3.3), we find that $\underline{b}_4 = -1$ and $\overline{b}_4 = -1/3$. Similarly, we can 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, we find that $\underline{b}_3 = -1/2$ and $\overline{b}_3 = 1/3$.

Evidently, to obtain a system with only variable b_k in the last step, we can exclude variables $b_2, \ldots, b_{k-1}$ in an order different from the one described here. For instance, we could exclude variables in the order $b_{k-1}, b_{k-2}, \ldots, b_2$. So, if the set of solutions for (S_1) is nonempty and bounded, we can eliminate variables in any order.

Now we turn to cases where the system either does not have solutions or has a set of unbounded solutions. First, we consider what happens when (S_1) has no solution, focusing particularly on situations in which it is easy to detect that the solution set is empty. In this instance, it is intuitive that the recursive procedure will break at some point. Though we have implicitly assumed until now that (S_1) has solutions, it is likely that systems of inequalities derived from econometric models will not have solutions due to model misspecification. At some step in this situation, 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 $C_q \geq 0$ where C_q is a negative constant or $\overline{b}_k < \underline{b}_k$.

Example 1.3.2. *Add to system (1.3.1) one more inequality:*

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

Then there will be two more inequalities in system (1.3.2):

$$-4 + 2b_3 + 16b_4 \geq 0$$

$$-6 + 4b_3 + 5b_4 \geq 0,$$

After eliminating b_3 we will obtain system (1.3.3) plus two more inequalities

$$2 + 22b_4 \geq 0$$

$$-2 + 7b_4 \geq 0,$$

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

Example 1.3.3. *Add to system (1.3.1) one more inequality:*

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

Then there will be two more inequalities in system (1.3.2):

$$-3 \geq 0$$

$$-6 + 4b_3 + 12b_4 \geq 0.$$

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

Second, we consider the case in which (S_1) has an unbounded set of solutions, a condition that can be easy to spot if at one step in the process we notice that all the coefficients corresponding to some variable b_i have the same strict sign, then we can conclude right away that the solution set is unbounded. Nevertheless, we may still be able to continue the procedure if there is a variable with coefficients of both signs. We stop when no variables in the system have both negative and positive signs. At this point, if at least one variable has all coefficients of the same strict sign, then the set of solutions is unbounded; if each variable has several zero coefficients, then the system needs further investigation because its solution set may be either unbounded or empty. Consider the following example.

Example 1.3.4. *Drawing from system (1.3.1), we modify two inequalities and consider the* *following system:*

$$
\begin{aligned}
-b_2 + 3b_3 - 4b_4 &\geq 0 \\
4 - b_2 + 4b_3 \qquad\quad &\geq 0 \\
2 + b_2 - 2b_3 + 6b_4 &\geq 0 \\
b_2 \qquad\quad + 2b_4 &\geq 0 \\
-1 - b_2 + 5b_3 - 5b_4 &\geq 0.
\end{aligned}
\tag{1.3.4}
$$

Eliminating variable b_2 *from (1.3.4), 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 (1.3.4) at the first step, we would obtain the following system:

$$6 + 3b_2 + 10b_4 \geq 0$$

$$8 + b_2 + 12b_4 \geq 0$$

$$8 + 3b_2 + 20b_4 \geq 0$$

$$b_2 + 2b_4 \geq 0.$$

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.

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 $b_4, \ldots, b_k$ in the manner described.

1.3.2 Examples to illustrate working of the recursive procedure

Example 1. Consider the binary response model

$$Y = 1(X_1 + \beta_2 X_2 + \beta_3 X_3 + U \geq 0), \tag{1.3.5}$$

where random variable X_1 takes values from $\{2, 0\}$, random variable X_2 takes values from $\{0.5, 1.5, 2.5, 3.5\}$ and random variable X_3 takes values from $\{0, 1, 2, 3, 4\}$. Thus, the support of $X = (X_1, X_2, X_3)$ consists of 40 points. Conditional probabilities P^l, $l = 1, \ldots, 40$, are determined by the rule

$$x_1^l + 1.5x_2^l - 2x_3^l \geq 0 \quad \Rightarrow \quad P^l \geq 0.5$$

$$x_1^l + 1.5x_2^l - 2x_3^l < 0 \quad \Rightarrow \quad P^l < 0.5.$$

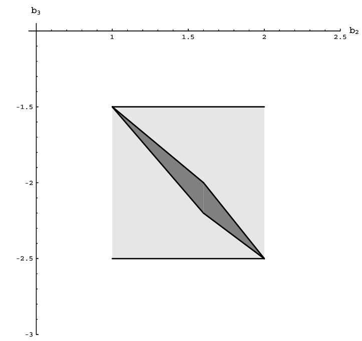


Figure 1.2. Example 1

Because all P^l are known, we can construct a system of linear inequalities that defines the identification set B:

$$P^l \geq 0.5 \quad \Leftrightarrow \quad x_1^l + x_2^l b_2 + x_3^l b_3 \geq 0, \quad l = 1, \ldots, 40.$$

The goal is to find the smallest rectangular superset of B. The recursive method yields the identification intervals $\beta_2 \in (1, 2)$, $\beta_3 \in (-2.5, -1.5)$; thus, the smallest rectangle is $(1, 2) \times (-2.5, -1.5)$. Figure **??** shows the identification region (the dark gray area) and its smallest rectangular superset.

Example 2. Consider a binary response 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. Conditional

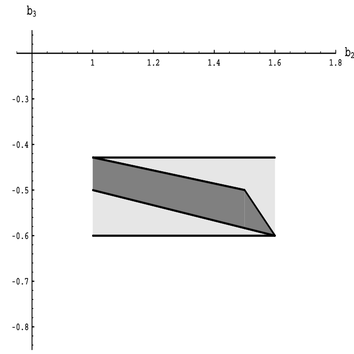


Figure 1.3. Example 2

probabilities P^l, $l = 1, \ldots, 88$, are determined by the rule

$$x_1^l + 1.25 - 0.5x_3^l \geq 0 \quad \Rightarrow \quad P^l \geq 0.5$$

$$x_1^l + 1.25 - 0.5x_3^l < 0 \quad \Rightarrow \quad P^l < 0.5.$$

Because all P^l are known, we can construct a 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 for β_2 and β_3: $\beta_2 \in (1, 1.6)$, $\beta_3 \in (-0.6, -0.42877)$. Figure 1.3 shows both the identification set and bounds.

1.3.3 Formulas for bounds on β_m, $m \neq 1$

In this section, we use the recursive procedure to derive formulas for the bounds on β_m, $m \neq 1$, in model (BRM). These bounds are expressed in terms of $z_i = sgn(P^i - 0.5)x^i$,

$i = 1, \ldots, d$. For clarity, we first show formulas for the cases $k = 3$ and $k = 4$ and then present them in general.

Proposition 1.4. *Let $k = 3$. Suppose the solution set of (S_1) is non-empty and bounded. Then*

$$ {b_3}^l \leq \beta_3 \leq {b_3}^u, \qquad \textit{where} $$

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

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

We assume there are sets of indices i, j such that the conditions in the definition of ${b_3}^u$ and ${b_3}^l$ are satisfied. Bounds ${b_3}^u$ and ${b_3}^l$ are not necessarily sharp.

Let us explain the last statement in this proposition. According to the recursive procedure, if in some inequalities the coefficients corresponding to b_2 are 0, then we carry over those inequalities to the next step without any changes. The formulas for ${b_3}^u$ and ${b_3}^l$, however, ignore situations in which some coefficients corresponding to b_2 can be 0. Therefore, they do not necessarily describe sharp bounds. Sharp bounds $\tilde{b}_3^u$ and $\tilde{b}_3^l$ that account for these cases can be written as follows:

$$\tilde{b}_3^u = \min\left\{b_3^u, \min_i\{-\frac{z_{i1}}{z_{i3}} : z_{i2} = 0,\, z_{i3} < 0\}\right\},\ \ \tilde{b}_3^l = \min\left\{b_3^l, \min_i\{-\frac{z_{i1}}{z_{i3}} : z_{i2} = 0,\, z_{i3} > 0\}\right\}.$$

If there is a constant term among regressors, these formulas are simpler. Without loss of generality, $x_{i2} = 1$, $i = 1, \ldots, d$. 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:

$$b_3{}^u = \min_{i,j}\left\{-\frac{x_{i1} - x_{j1}}{x_{i3} - x_{j3}} : x_{i3} - x_{j3} < 0\right\},\ \ b_3{}^l = \max_{i,j}\left\{-\frac{x_{i1} - x_{j1}}{x_{i3} - x_{j3}} : x_{i3} - x_{j3} > 0\right\}.$$

It is worth noticing the symmetry of the formulas for b_3^l and b_3^u. To find the lower (upper) bound b_3^l or (b_3^u), we choose inequalities for which the determinant in the denominator is positive (negative). This symmetry will hold in the general case as well and it will allow us to formulate results in section 1.5, when we consider the situation in which the support of X becomes increasingly dense.

The next proposition considers the case of any number of parameters. First, however, let us us introduce some notations. 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 us 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 1.4,

$$b_3{}^u = \min_{i,j}\left\{-\frac{A_1(1,i,j)}{A_1(3,i,j)<0}\right\},\quad b_3{}^l = \max_{i,j}\left\{-\frac{A_1(1,i,j)}{A_1(3,i,j)>0}\right\}.$$

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 $A_0(m,j) = z_{jm}$. Let

$$A_2(m,i_1,j_1,i_2,j_2) = \begin{vmatrix} A_1(m,i_1,j_1) & A_1(3,i_1,j_1)<0 \\ A_1(m,i_2,j_2) & A_1(3,i_2,j_2)>0 \end{vmatrix}.$$

Then $b_4{}^u$, $b_4{}^l$ can be written as

$$b_4{}^u = \min_{i_1,j_1,i_2,j_2}\left\{-\frac{A_2(1,i_1,j_1,i_2,j_2)}{A_2(4,i_1,j_1,i_2,j_2)<0}\right\},\quad b_4{}^l = \max_{i_1,j_1,i_2,j_2}\left\{-\frac{A_2(1,i_1,j_1,i_2,j_2)}{A_2(4,i_1,j_1,i_2,j_2)>0}\right\}.$$

Now we can formulate a general result.

Proposition 1.5. *Let $k \geq 3$. Suppose the solution set of (S_1) is non-empty and bounded. Then*

$$b_k{}^l \leq \beta_k \leq b_k{}^u, \qquad \textit{where}$$

$$b_k{}^u = \min_{i_1,\ldots,j_{2^{k-3}}}\left\{-\frac{A_{k-2}(1,i_1,\ldots,j_{2^{k-3}})}{A_{k-2}(k,i_1,\ldots,j_{2^{k-3}})<0}\right\},$$

$$b_k{}^l = \max_{i_1,\ldots,j_{2^{k-3}}}\left\{-\frac{A_{k-2}(1,i_1,\ldots,j_{2^{k-3}})}{A_{k-2}(k,i_1,\ldots,j_{2^{k-3}})>0}\right\},$$

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

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

We assume there are sets of indices $i_1, \ldots, i_{2^{k-3}}, j_1, \ldots, j_{2^{k-3}}$ such that the conditions in the definition of $b_k{}^u$ and $b_k{}^l$ are satisfied. Bounds $b_k{}^u$ and $b_k{}^l$ are not necessarily sharp.

The reason why $b_k{}^u$ and $b_k{}^l$ are not necessarily sharp is similar to the explanation described earlier, when we considered 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 becomes difficult to keep track of these inequalities at each step of the procedure; that is why we do not give formulas for sharp bounds for $k > 3$.

1.3.4 Applications of the recursive method

The recursive method's effectiveness extends well beyond identification intervals. It can come in handy, for instance, when we are interested in some specific properties of B or other related questions. This section presents three applications of the method.

A rectangular subset with the largest perimeter

Because not every point in the smallest rectangular superset of B belongs to B, it may be of interest to find subsets of B that have certain properties. The recursive method allows us to easily find a rectangular subset of B with the largest possible perimeter. For simplicity, consider the case where $k = 3$. The problem that has to be solved is

$$a_2 + a_3 \rightarrow \max_{b_2, b_3, a_2, a_3,}$$

subject to

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

$$z_{i1} + z_{i2}(b_2 + a_2) + z_{i3}b_3 \geq 0$$

$$z_{i1} + z_{i2}b_2 + z_{i3}(b_3 + a_3) \geq 0$$

$$z_{i1} + z_{i2}(b_2 + a_2) + z_{i3}(b_3 + a_3) \geq 0$$

$$i = 1, \ldots, d,$$

$$a_2 \geq 0, \quad a_3 \geq 0.$$

Denote $a = a_2 + a_3$. Substitute expression $a_3 = a - a_2$ into the system, and obtain

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

$$z_{i1} + z_{i2}b_2 + z_{i3}b_3 + z_{i2}a_2 \geq 0$$

$$z_{i1} + z_{i2}b_2 + z_{i3}b_3 - z_{i3}a_2 + z_{i3}a \geq 0$$

$$z_{i1} + z_{i2}b_2 + z_{i3}b_3 + (z_{i2} - z_{i3})a_2 + z_{i3} \geq 0 \tag{1.3.6}$$

$$i = 1, \ldots, d,$$

$$a_2 \geq 0, \quad a \geq 0.$$

Find the upper bound on a from system (1.3.6), and denote it as a^u. Let (b_2^*, b_3^*, a_2^*) be any solution of (1.3.6) when a is fixed at its highest value: $a = a^u$. Value $2a^u$ is the largest possible perimeter, and points (b_2^*, b_3^*), $(b_2^* + a_2^*, b_3^*)$, $(b_2^*, b_3^* + a^u - a_2^*)$ and $(b_2^* + a_2, b_3^* + a^u - a_2^*)$ describe a rectangular subset with this perimeter.

An cubic subset with the largest volume

The recursive procedure can also determine a cubic subset of B with the largest volume. Again for simplicity assume $k = 3$. In this instance, we look for a square with the largest area. Denote the side of a square as a and find the upper bound a^u on a in the system

$$\begin{aligned} z_{i1} + z_{i2}b_2 + z_{i3}b_3 &\geq 0 \\ z_{i1} + z_{i2}b_2 + z_{i3}b_3 + z_{i2}a &\geq 0 \\ z_{i1} + z_{i2}b_2 + z_{i3}b_3 + z_{i3}a &\geq 0 \\ z_{i1} + z_{i2}b_2 + z_{i3}b_3 + (z_{i2} + z_{i3})a &\geq 0 \\ i = 1, \ldots, d& \\ a &\geq 0. \end{aligned} \tag{1.3.7}$$

Let (b_2^*, b_3^*) be any solution of (1.3.7) when a is fixed at its largest value: $a = a^u$. Value $(a^u)^2$ is the largest possible area of a square subset, and points (b_2^*, b_3^*), $(b_2^* + a^u, b_3^*)$, $(b_2^*, b_3^* + a^u)$ and $(b_2^* + a^u, b_3^* + a^u)$ describe a square subset with this area.

Extrapolation

Suppose we have a point $x^* = (x_1^*, \ldots, x_k^*) \notin S(X)$ and would like to learn the value of $P(Y = 1|X = x^*)$. Clearly, this value is not identified, but we may be able to determine the sign of $P(Y = 1|X = x^*) - 0.5$.

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

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

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 we cannot draw any conclusions about the value of $sgn(P(Y = 1|X = x^*) - 0.5)$.

Let us show how we can find these bounds. Introduce a new variable b' which is the value of the linear function:

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

Assume that at least one of x_i^*, $i = 2, \ldots, k$, is different from 0. Without a loss of generality, $x_k^* \neq 0$. Then b_k can be expressed through $b_2, \ldots, b_k, b'$ as follows:

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

Substitute this expression into the system that defines the identification set:

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

Then obtain a system of linear inequalities with unknown variables $b_2, \ldots, b_{k-1}, b'$:

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

Using the recursive procedure, we can find the sharp bounds for b' - that is, the sharp bounds on the values of the linear function.

1.3.5 Support vector machines

In this section, we show a connection between the identification problem in binary response models and support vector machines in statistical learning theory.

To describe this relationship we 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 the same as 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 1.4 shows optimal and a non-optimal separating hyperplanes for two classes of points.

Let us briefly present the mathematical description of SVMs (for more details, see Vapnik (2000)). For a better comparison with existeing literature, we 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

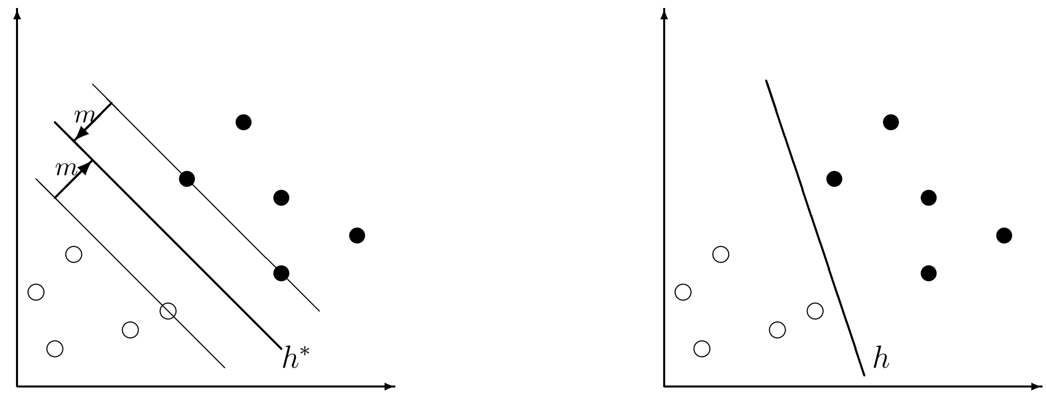


Figure 1.4. The optimal hyperplane h^* with maximal margin m (left) and a non-optimal separating hyperplane h (right)

hyperplane $w_1x_1+\ldots+w_kx_k-a=0$. To find the optimal hyperplane, we solve the quadratic programming problem

$$\min_{w_1,\ldots,w_k,a} \sum_{j=1}^{k} w_j^2 \qquad (1.3.8)$$

subject to

$$y_i[w_1x_{i,1}+\ldots+w_kx_{i,k}-a] \geq 1, \quad i=1,\ldots,n. \qquad (1.3.9)$$

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

Let us 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:

$$R(\eta) = P(\eta(X) \neq Y) = \int 0.5|y-\eta(x)|dP, \qquad (1.3.10)$$

where η 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 (M) 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 (1.3.10).

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

1.4 Misspecification

If (BRM) is misspecified, then system (S_1) is very likely to have no solutions. One way to handle this situation is to consider an optimization problem with respect to b and find its solution set. This problem has to satisfy two requirements: first, it should be meaningful, allowing us to interpret its set of solutions in terms of the properties of model (BR). Second, its solution set should coincide with the identification set when (BRM) is well specified.

We suggest several optimization techniques to apply when a model is misspecified and discuss their sets of solutions. There is no best method among them, and preferences for a certain approach usually change depending on the situation. Of course, procedures described below are just a few of the possible options.

1.4.1 Maximum score method

An attractive feature of the maximum score method is that it always produces solutions, even when the model is misspecified. Manski (1985) defines a maximum score estimator (on the population level) as an estimator the minimizes the expected absolute distance between the conditional expectation of Y conditional on X and $1(Xb \geq 0)$ (if the median condition (M) holds, then $1(Xb \geq 0) = M(Y|X)$):

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

Equivalently, a maximum score estimator minimizes the expected squared distance between between $E(Y|X)$ and $1(Xb \geq 0)$:

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

First, let us show that when (BRM) is well specified, the set of maximum score estimates coincides with the identification set. On the population level, the maximum score method maximizes the function

$$S^{ms}(b) = 2\sum_{l=1}^{d} q^l(P^l - 0.5)sgn(x^l b), \tag{1.4.1}$$

where, we remind, $q^l = P(X = x^l)$, $P^l = P(Y = 1|X = x^l)$. Let B^{ms} stand for the set of $b \in \Re^k$ with a normalization $b_1 = 1$ that maximizes S^{ms}:

$$B^{ms} = \text{Argmax}_{b\in\Re^k:b_1=1} S^{ms}(b).$$

The next proposition compares sets B^{ms} and B.

Proposition 1.6. *Suppose that model (BRM) is well specified. Then* B^{ms} *is a convex polyhedron and*

$$\forall(x^l \in S(X))\ P^l \neq 0.5 \quad \Rightarrow \quad B = B^{ms}$$

$$\exists(x^l \in S(X))\ P^l = 0.5 \quad \Rightarrow \quad B \subset B^{ms}$$

with a strict inclusion.

If (BRM) is misspecified, then B^{ms} may be not a convex polyhedron. However, it will always be a finite union of several disjoint convex polyhedra. We are especially likely to get several polyhedra as solutions if many identical values exist among $(P^l - 0.5)q^l$, $l = 1, \ldots, d$.

1.4.2 Minimal number of classification errors

As we already mentioned, 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 that class; otherwise, it is assigned to the other class. If the model is misspecified and $B = \emptyset$, then some classification errors have been made. In order to find the values of b that minimize the number of classification errors, consider the following optimization problem:

$$\max_{b:b_1=1} Q(b), \quad \text{where} \quad Q(b) = \sum_{l=1}^{d} sgn(P^l - 0.5) sgn(x^l b). \tag{1.4.2}$$

If (BRM) is well specified, the set of solutions for (1.4.2) coincides with B. Because the value of $Q(b)$ is d minus the number of classification errors, the maximization of $Q(b)$ minimizes the number of these errors. We can modify problem (1.4.2) and consider

$$\max_{b:b_1=1} Q(b), \quad \text{where} \quad Q(b) = \sum_{l=1}^{d} sgn(P^l - 0.5) sgn(x^l b) q^l. \tag{1.4.3}$$

It is of interest to note that problem (1.4.3) is equivalent to the minimization of the expected absolute distance between $M(Y|X)$ and $1(Xb \geq 0)$:

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

The solution sets of (1.4.2) and (1.4.3) are finite unions of disjoint convex polyhedra.

1.4.3 Minimal general classification error

In this section, we discuss a geometric approach to misspecification. It is based on the soft margin hyperplane method in SVMs, 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. We describe this approach in more detail in the end of the section.

The method in SVMs chooses only one soft margin hyperplane: the optimal one. We are interested in finding a set of these hyperplanes, however, rather than just one. In fact, we 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 1.3.5, we first must assume that there is a constant term among regressors. Without 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 us 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:

$$\exists((b_1,\ldots,b_k)\in\tilde{B})\exists(\delta>0)\forall(l=1,\ldots,d)\quad sgn(P^l-0.5)(x_1^lb_1+x_2^lb_2+\ldots+x_{k-1}^lb_{k-1}+b_k)\geq\delta. \tag{1.4.4}$$

Note that in this assertion, we have incorporated the finite support of X and the constant term among regressors. Because for any $b\in\tilde{B}$, vector αb, $\alpha>0$, defines the same hyperplane as b, δ can be any positive number. Without a loss of generality, we suppose that $\delta=1$. Let h stand for the separating hyperplane $x_1^lb_1+x_2^lb_2+\ldots+x_{k-1}^lb_{k-1}+b_k=0$. Let h_1 denote the hyperplane $x_1^lb_1+x_2^lb_2+\ldots+x_{k-1}^lb_{k-1}+b_k=1$ and h_2 denote the hyperplane $x_1^lb_1+x_2^lb_2+\ldots+x_{k-1}^lb_{k-1}+b_k=-1$. Then, according to (1.4.4), 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 may be not linearly divisible. In this case, we introduce non-negative slack variables to allow for some error in separation and to find a soft margin hyperplane.

Let $v_i\geq 0$, $i=1,\ldots,d$, be slack variables. Consider the following linear programming problem:

$$\min_{b_1,\ldots,b_k,\{v_l\}_{l=1}^d} Q(b,v)=\sum_{l=1}^d v_l \tag{1.4.5}$$

subject to

$$sgn(P^l-0.5)(x_1^lb_1+x_2^lb_2+\ldots+x_{k-1}^lb_{k-1}+b_k)\geq 1-v_l,$$

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

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 $(b^*, v^*) \in D^*$, a 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 1.5, 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, they do not have any classification errors. 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 1.5, is a soft margin hyperplane for the two classes of points.

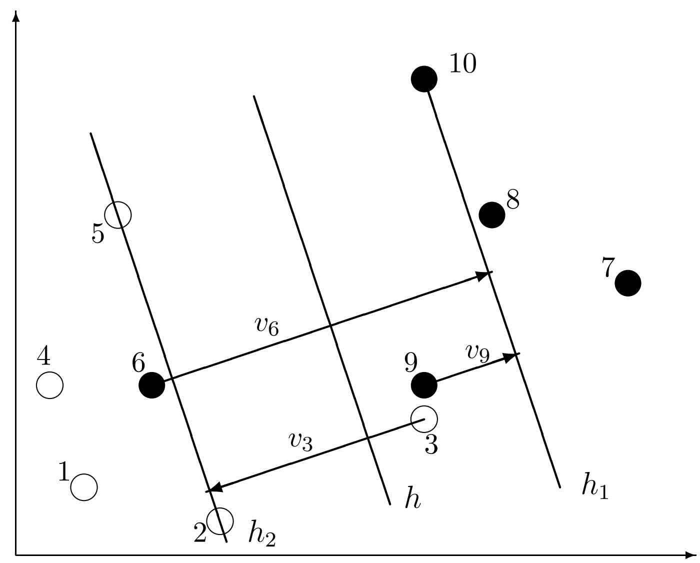


Figure 1.5. A soft margin separating hyperplane (h) and classification errors

Let us now return to the general problem. If instead of $\delta = 1$ we took a different $\delta > 0$ and solved (1.4.5) 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 δ: $D^*(\delta) = \delta D^*$. Given this scale effect, we are interested in the set

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

rather than set B^* itself. (At this stage, we assume that $b_1^* > 0$ for any $b^* \in B^*$.)

Finding bounds on 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 bounds 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 $(b_2, \ldots, b_k)$ of the following system:

$$sgn(P^l - 0.5)(x_1^l + x_2^l b_2 + \ldots + x_{k-1}^l b_{k-1} + b_k) \geq \frac{1 - v_l^*}{b_1^*}, \quad l = 1, \ldots, d. \tag{1.4.6}$$

Clearly, $B_{1,s}^* \subset B_1^*$, and system (1.4.6) can be rewritten in the form of (S_1). Therefore, we can find the bounds on $B_{1,s}^*$ with the recursive procedure.

Let us discuss some modifications of the soft margin hyperplane method. In the general classification error in (1.4.5), 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 $\sum_{l=1}^d \lambda_l = 1$, $\lambda_l \geq 0$, $l = 1, \ldots, d$. 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 λ_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 (1.3.8), 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

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

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

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 ξ. When σ 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, then the parameter value belongs to the identification set.

1.5 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 1.7 formalizes this suggestion.

Proposition 1.7. *Consider system* (S_1) *with* $k = 3$. *Suppose that its set of solutions is non-empty and bounded. Also, suppose that the system contains four inequalities*

$$z_{i_1,1} + z_{i_1,2}b_2 + z_{i_1,3}b_3 \geq 0$$

$$z_{i_2,1} + z_{i_2,2}b_2 + z_{i_2,3}b_3 \geq 0$$

$$z_{j_1,1} + z_{j_1,2}b_2 + z_{j_1,3}b_3 \geq 0$$

$$z_{j_2,1} + z_{j_2,2}b_2 + z_{j_2,3}b_3 \geq 0$$

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*

$$b_3^u - b_3^l \leq \frac{\begin{vmatrix} \Delta & -z_{i_2,2} \\ \Delta & z_{i_1,2} \end{vmatrix}}{\begin{vmatrix} z_{i_2,3} & z_{i_2,2} \\ z_{i_1,3} & z_{i_1,2} \end{vmatrix}} = \Delta \frac{\begin{vmatrix} 1 & -z_{i_2,2} \\ 1 & z_{i_1,2} \end{vmatrix}}{\begin{vmatrix} z_{i_2,3} & z_{i_2,2} \\ z_{i_1,3} & z_{i_1,2} \end{vmatrix}},$$

where b_3^u, b_3^l *are defined as in Proposition 1.4.*

This result is obtained using the symmetry of the formulas for b_3^l and b_3^u. If we take four other inequalities that satisfy the conditions of the proposition, then we obtain a different bound for $b_3^u - b_3^l$, and we can choose the lower of two.

The role of Proposition 1.4 may be better appreciated if we formulate an analogous result in terms of the properties of the support.

Corollary 1.2. *Let B be non-empty and bounded. Suppose that there exist (x_2, x_3) and (x_2^*, x_3^*) such that*

$$x_2 x_3^* - x_2^* x_3 \neq 0,\ x_2 \neq 0,\ x_2^* \neq 0.$$

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

$$b_3^u - b_3^l \leq \Delta \frac{|x_2| + |x_2^*|}{|x_2 x_3^* - x_2^* x_3|}. \tag{1.5.1}$$

The results for $k = 3$ can be generalized for the case of any k.

1.6 Single-index models

1.6.1 Identification set

Consider a single-index model

$$E(Y|X = x) = G(x\beta) \tag{1.6.1}$$

with an increasing function G. The identification set for this model is

$$B^M = \{b \in \Re^{k-1} : \forall (x^l, x^m \in S(X)) \quad E(Y|X = x^l) > E(Y|X = x^m) \;\Rightarrow\; x^l b > x^m b\}.$$

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 a system of $d - 1$ inequalities:

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

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

$$x^1 b > x^m b, \quad l = 1, \ldots, d/2, \quad m = d/2 + 1, \ldots, d.$$

Note that the inequalities that describe B^M do not have a constant term.

1.6.2 Maximum rank correlation estimator

One of the common approaches to estimating models like (1.6.1) is the maximum rank correlation method. On the population level, it maximizes the following objective function:

$$Q^{MRC}(b) = E\left[1(y_1 > y_2)1(x_1 b > x_2 b) + 1(y_1 < y_2)1(x_1 b < x_2 b)\right].$$

Let B^{MRC} stand for the maximand of Q^{MRC}: $B^{MRC} = \text{Argmax}_{b\in\Re^k:b_1=1}Q^{MRC}(b)$. We compare sets B^{MRC} and B^M in two instances of (1.6.1). In the first case

$$y = G(x\beta) + \epsilon, \tag{1.6.2}$$

where ϵ is independent of X and has a strictly monotone distribution function, and $E(\epsilon) = 0$. In the second case, we introduce a heterogeneity of errors and consider

$$y = G(x\beta) + \sigma(x)\epsilon, \tag{1.6.3}$$

where ϵ is independent of X and has a strictly monotone distribution function, and $E(\epsilon) = 0$.

For (1.6.2)

$$E\left[1(y_1 > y_2)1(x_1b > x_2b)|x_1 = x^l, x_2 = x^m\right] = Pr(y_1 > y_2|x_1 = x^l, x_2 = x^m)1(x^lb > x^mb) =$$

$$= Pr(\epsilon_2 - \epsilon_1 < G(x^l\beta) - G(x^m\beta))1(x^lb > x^mb) = \tilde{F}(G(x^l\beta) - G(x^m\beta))1(x^lb > x^mb),$$

where $\tilde{F}$ is the distribution function of $\epsilon_2 - \epsilon_1$. So, clearly,

$$\begin{aligned} Q^{MRC}(b) = \sum_{l,m:m<l} [\tilde{F}(G(x^l\beta) - G(x^m\beta))1(x^lb > x^mb)+ \\ +\tilde{F}(G(x^m\beta) - G(x^l\beta))1(x^lb < x^mb)]q(x^l)q(x^m). \end{aligned} \tag{1.6.4}$$

Note that $\tilde{F}$ is strictly increasing, and $\tilde{F}(0) = 0.5$. It is easy to show that

$$\forall(l, m : l \neq m) \quad G(x^l\beta) \neq G(x^m\beta) \qquad \Rightarrow \quad B^{MRC} = B^M. \tag{1.6.5}$$

In the heterogeneity case (1.6.3), the maximum rank correlation method may yield a set of estimators that is different from the identification set B^M. The example below describes a case where $B^{MRC} \cap B = \emptyset$.

Example 1.8. *Consider (1.6.3), where $x\beta = x_1 + x_2\beta$ and*

$$S(X) = \{x^l\}_{l=1}^{L}, \quad x^l = (l, l+1).$$

Suppose $E(Y|X = x^l) > E(Y|X = x^m)$ when $l > m$ and $\sigma(x^l)/\sigma(x^m) = 4^{l-m}$. It is straightforward to find the identification set:

$$B^M = \{b : l + (l+1)b > m + (m+1)b, l > m\} = (-1, +\infty).$$

Let ϵ be distributed as

$$F_\epsilon(t) = \begin{cases} \frac{3}{4}e^t, & t \le 0 \\ 1 - \frac{1}{4}e^{-\frac{1}{3}t}, & t > 0. \end{cases}$$

It is easy to verify that $E(\epsilon) = 0$. The CDF of $\sigma(x^m)\epsilon_2 - \sigma(x^l)\epsilon_1$ is

$$\tilde{F}_{lm}(\xi) = \int_{-\infty}^{+\infty} F_\epsilon\left(\frac{t+\xi}{\sigma(x^m)}\right)\frac{1}{\sigma(x^l)}f_\epsilon\left(\frac{t}{\sigma(x^l)}\right)dt,$$

where f_ϵ is the density of ϵ:

$$f_\epsilon(t) = \begin{cases} \frac{3}{4}e^t, & t < 0 \\ \frac{1}{12}e^{-\frac{1}{3}t}, & t > 0. \end{cases}$$

Let us find the value of $\tilde{F}_{lm}$ for 0:

$$\tilde{F}_{lm}(0) = \int_{-\infty}^{+\infty} F_\epsilon\left(\frac{t}{\sigma(x^m)}\right)\frac{1}{\sigma(x^l)}f_\epsilon\left(\frac{t}{\sigma(x^l)}\right)dt = \int_{-\infty}^{0}\frac{3}{4}e^{t/\sigma(x^m)}\frac{1}{\sigma(x^l)}\frac{3}{4}e^{t/\sigma(x^l)}dt +$$
$$+\int_0^{+\infty}\left(1 - \frac{1}{4}e^{-\frac{t}{3\sigma(x^m)}}\right)\frac{1}{\sigma(x^l)}\frac{1}{12}e^{-\frac{t}{3\sigma(x^l)}}dt = \frac{1}{4} + \frac{5}{8}\frac{1}{1+\sigma(x^l)/\sigma(x^m)}.$$

When $l > m$, $\tilde{F}_{lm}(0) \le 3/8$. Because $\tilde{F}_{lm}$ is continuous and there is a finite number of pairs (l, m), then

$$\exists(\delta > 0)\forall(\xi \in (0,\delta))\forall(l, m : l > m) \quad \tilde{F}_{lm}(\xi) < 0.5.$$

Now assume $E(Y|X = x^L) - E(Y|X = x^1) < \xi$. *In (1.6.4), every term in the sum attains its maximum possible value when* $x^l b < x^m b$ *for* $l > m$. *In other words,*

$$B^{MRC} = \{b : l + (l+1)b < m + (m+1)b, l > m\} = (-\infty, -1).$$

As we can see, $B^{MRC} \cap B^M = \emptyset$.

Cavanagh and Sherman (1998) recognized this problem with the maximum rank correlation method and suggested a slightly altered approach. Their technique maximizes a modified function Q^{MRC} and gives a set of solutions B^{MRC} that always satisfies property (1.6.5).

1.6.3 Ordered response models: Median independence

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

$$Y = \sum_{j=1}^{J+1} j1(\alpha_{j-1} < Y^* \leq \alpha_j),$$

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

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

The set of identified parameters is described by a system of linear inequalities with a maximum of $2d + J - 1$ 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 $P(Y \leq J|x = x^l) < 0.5$, or

$P(Y \leq 1|X = x^l) < 0.5$ and $P(Y \leq J|x = x^l) \geq 0.5$. In the first case, x^l contributes the inequality

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

In the second case, it provides

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

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

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

Then x^l contributes two inequalities

$$\alpha_{j(x^l)-1} - x^l b < 0, \quad \alpha_{j(x^l)} - x^l b \geq 0. \tag{1.6.6}$$

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

$$\alpha_j - \alpha_{j-1} > 0, \quad j = 2, \ldots, J.$$

Some of these additional inequalities will be excessive because, for example, (1.6.6) implies that $\alpha_{j(x^l)-1} < \alpha_{j(x^l)}$.

After we normalize b by setting $b_1 = 1$, we get a system of at most $2d + J - 1$ linear inequalities with $k - 1 + J$ unknown variables $(b_2, \ldots, b_k, \alpha_1, \ldots, \alpha_J)$. This system contains both strict and non-strict inequalities.

1.7 Asymptotic properties

1.7.1 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. Usually, only estimates of conditional probabilities are available, so a natural question is, how close does an estimated identification set get to the true set as the sample size increases.

In this section, we assume that model (BRM) is well specified, and we distinguish two cases: one in which all conditional probabilities are different from 0.5 and one in which some of them are 0.5. We differentiate these two situations because 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 a sample size is.

Theorem 1.9. *Suppose that*

$$\hat{P}^l_N \xrightarrow{p} P^l \text{ as } N \to \infty, \tag{1.7.1}$$

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

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

Then

$$Pr(B_N \neq B) \to 0 \quad as \quad N \to \infty.$$

Theorem 1.9 allows us to formulate properties of convergence in terms of Hausdorff distances.

Corollary 1.3. *Under the conditions of Theorem 1.9,*

$$Pr(H(B_N, B) \neq 0) \to 0 \quad as \quad N \to \infty.$$

Corollary 1.4. *Under the conditions of Theorem 1.9,*

$$\tau_N H(B_N, B) \xrightarrow{p} 0 \quad as \quad N \to \infty$$

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

The statement of Theorem 1.9 does not hold if there exist P^l equal to 0.5. Nevertheless, in this case we can derive a consistent estimator of B by introducing slack variables ϵ_N.

Theorem 1.10. *Suppose that*

$$\tau_N(\hat{P}_N^l - P^l), \quad l = 1, \ldots, d, \tag{1.7.2}$$

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

$$\epsilon_N > 0 \text{ and } \epsilon_N \to 0, \quad \epsilon_N \tau_N \to \infty \quad as \quad N \to \infty. \tag{1.7.3}$$

Let B_N *be a solution set for the system of linear inequalities derived by the rule*

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

Then

$$Pr(B_N \neq B) \to 0 \quad as \quad N \to \infty.$$

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

$$\hat{P}_N^l = \frac{\sum y_i 1(x_i = x^l)}{\sum 1(x_i = x^l)}, \tag{1.7.4}$$

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

Theorem 1.11. *Let $\hat{P}_N^l$ be defined as in (1.7.4) and B_N be a solution set for the system of linear inequalities derived by the rule*

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

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

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

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

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

It is important to note that random sampling errors in estimated response probabilities may cause the system of inequalities based on $\hat{P}_N^l$ to have no solution. In this case, B_N would be empty. One way to address this problem is to consider a sample analog of the objective function in section 1.4.2:

$$Q_N(b) = \sum_{l=1}^{d} sgn(x^l b) sgn(P_N^l - 0.5).$$

Let B_N^* be the set of maximizers of $Q_N(b)$:

$$B_N^* = \text{Argmax}_{b \in \Re^k : b_1 = 1} \sum_{l=1}^{d} sgn(x^l b) sgn(P_N^l - 0.5).$$

If $B_N \neq \emptyset$, then $B_N = B_N^*$. All results in the current section remain true if we substitute B_N for B_N^*. Other objective functions $Q_N(\cdot)$ also can be considered. Because the model is assumed to be well specified, the only condition required for $Q_N(\cdot)$ is that the set of maximizers of its population analog $Q(\cdot)$ coincides with B. Maximum score estimation approach, for instance, has this property. We discuss this approach below. In general, we would prefer to consider objective functions that reflect some of the model's properties.

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

Let us analyze the behavior of maximum score estimates obtained from a random sample. In the sample, maximum score estimates maximize the function

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

which can be equivalently written as

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

where $\hat{P}_N^l$ is defined as in (1.7.4).

Proposition 1.12. *Suppose that model (BRM) is well specified. Let*

$$B_N^{ms} = Argmax_{b \in \Re^k : b_1 = 1} S_N^{ms}(b).$$

Then

$$\forall(x^l \in S(X))\ P^l \neq 0.5 \quad \Rightarrow \quad \exists(0<\rho<1) \quad Pr(B_N^{ms} \neq B) = o(\rho^N)\ \text{as}\ N \to \infty,$$

$$\exists(x^l \in S(X))\ P^l = 0.5 \quad \Rightarrow \quad \exists(p_0>0) \quad Pr(B_N^{ms} \neq B) \geq p_0\ \text{as}\ N \to \infty.$$

1.7.2 Statistical inference

We now describe two methods of building confidence regions for the identification set. Both are based on the normal approximations of conditional probabilities P^l. Because of the problem's discrete nature, we cannot always achieve an exact nominal confidence level; in many cases, the true confidence coefficient is greater than the stated level.

Given a random sample of size N, our objective is to construct regions B_N that asymptotically cover B with probability $1-\alpha$, where α is some prespecified value between 0 and 1:

$$\lim_{N\to\infty} P(B_N \supset B) \geq 1-\alpha. \tag{1.7.5}$$

Let $\hat{P}_N^l$ be a frequency estimator of P^l as in (1.7.4). Asymptotically,

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

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

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

In the first method, we choose numbers $\{\gamma_l\}_{l=1}^d$ such that $\gamma_l \geq 0$, $l=1,\ldots,d$, and $\sum_{l=1}^d \gamma_l = \alpha$. Let ζ_{γ_l} denote the $1-\gamma_l$ quantile of the standard normal distribution, then construct a system of linear inequalities in the following way:

if for a given x^l

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

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

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

then add the inequality $-x^l b > 0$ to the system. If the interval $\left[\hat{P}^l_N - \zeta_{\gamma_l}\frac{\hat{\sigma}^l_N}{\sqrt{N}},\ \hat{P}^l_N + \zeta_{\gamma_l}\frac{\hat{\sigma}^l_N}{\sqrt{N}}\right]$ contains 0.5, then no inequalities in the system correspond to x^l. We claim that the solution set B_N for the system constructed according to this method has the property

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

Indeed,

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

We find that

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

and, similarly,

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

Clearly,

$$P(B_N \supset B) = 1 - P(B_N \not\supset B) \to 1 \text{ as } N \to \infty.$$

As we can see, confidence region B_N is overly conservative: Its actual coverage probability is 1. In this sense, it is inaccurate.

Furthermore, in practice set B_N can be empty. In this case, instead of B_N, we can find a set B_N^* such that

$$B_N^* = \text{Argmax}_{b \in \Re^k : b_1 = 1} \sum_{l=1}^{d} sgn(x^l b \geq 0) \left(1(\hat{P}_N^l - \zeta_{\gamma_l} \frac{\hat{\sigma}_N^l}{\sqrt{N}} > 0.5) - 1(\hat{P}_N^l + \zeta_{\gamma_l} \frac{\hat{\sigma}_N^l}{\sqrt{N}} < 0.5) \right).$$

If B_N is not empty, $B_N = B_N^*$. Using the technique we employed for region B_N, we can prove that $Pr(B_N^* \supset B) \to 1$ as $N \to \infty$.

The method of finding B_N^* is equivalent to the following approach. Let

$$I = \prod_{l=1}^{d} \left[\hat{P}_N^l - \zeta_{\gamma_l} \frac{\hat{\sigma}_N^l}{\sqrt{N}}, \hat{P}_N^l + \zeta_{\gamma_l} \frac{\hat{\sigma}_N^l}{\sqrt{N}} \right].$$

Set I is an asymptotic $1-\alpha$ confidence set for $P = (P^1, \ldots, P^d)$. If for any $\tilde{P} = (\tilde{P}^1, \ldots, \tilde{P}^d) \in I$ we construct a system of inequalities by the rule

$$\tilde{P}_l \geq 0.5 \quad \Leftrightarrow \quad x^l b \geq 0, \tag{1.7.6}$$

and find its solution $B_{N,\tilde{P}}$, then we can show that

$$B_N^* = \cup_{\tilde{P} \in I} B_{N,\tilde{P}}.$$

(Of course, many of sets $B_{N,\tilde{P}}$ will be empty.)

Let us describe another method for building confidence sets. Note that asymptotically

$$\sum_{l=1}^{d} N \frac{(\hat{P}_N^l - P^l)^2}{(\hat{\sigma}_N^l)^2} \to \chi^2(d-1) \quad \text{as} \quad N \to \infty.$$

Instead of I, consider ellipsoid E, which is a $1-\alpha$ confidence set for P:

$$E = \left\{ P = (\tilde{P}^1, \ldots, \tilde{P}^d) : \sum_{l=1}^{d} \frac{(\hat{P}_N^l - \tilde{P}^l)^2}{(\hat{\sigma}_N^l)^2} \leq \frac{\chi^2_{1-\alpha}(d-1)}{N} \right\}.$$

If for any $\tilde{P} = (\tilde{P}^1, \ldots, \tilde{P}^d) \in E$ we construct a system of inequalities according to (1.7.6) and find its solution set $B_{N,\tilde{P}}$, then region

$$B^*_{N,E} = \cup_{\tilde{P} \in E} B_{N,\tilde{P}}$$

satisfies (1.7.5).

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.

1.8 Empirical example

1.8.1 Monte Carlo simulations

The design of the Monte Carlo experiment is based on Example 2, which was described in section 3.3. 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 outlined in Example 2. We specify distributions for X_1, X_2 and X_3. Let the distribution of the error term be

$$U|x \sim \frac{x_1 1(x_1 < 0)}{\sqrt{2x_1^2 + 2x_2^2 + 0.001}} Z + 0.1x_2 1(x_1 \geq 0)V,$$

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

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

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

and apply the recursive procedure to find bounds on β_2 and β_3. Because the system has solutions, its set of maximum score estimates coincides with its solution set. See Table 1.1

for the estimation results. Observe that the set of maximum rank correlation estimates for β_3 does not contain value -0.5 used to design the experiment, although this value is very close to its border.

We want to emphasize that the results presented for the recursive procedure and the maximum score method are identification intervals for each individual parameter; the identification set for (β_2, β_3) is smaller than rectangle $(1, 1.6) \times (-0.6, -0.42587)$ (see Figure 1.3).

We 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.

1.8.2 Women's labor force participation

In this section we 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 1.2.

Thus, we estimate the binary response model

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

Table 1.1. Estimation results for the Monte Carlo experiment

	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	4.3893	-1.5922
	(2.3794 ,3.4510)	(3.4141,5.3646)	(-1.9173, -1.2672)
Probit (r)	1	**1.5057**	**-0.5462**
		(1.1711, 1.8402)	(-0.6577, -0.4347)
Logit	5.4318	8.0505	-2.9515
	(4.3201, 6.5435)	(6.0992, 10.0019)	(-3.6126, -2.2904)
Logit (r)	1	**1.4821**	**-0.5434**
		(1.1229, 1.8414)	(-0.6651, -0.4217)

Table 1.2. Descriptive statistics for MROZ data

	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

where we normalize the coefficient corresponding to EDUC. For comparison, we apply several estimation procedures. Table 1.3 contains the results of these estimations.

There are $N = 753$ observations. After calculating frequency estimates $\hat{P}_N^l$ and combining data for women with identical characteristics, we obtain 670 points in the support. Based on $\hat{P}_N^l$, we construct a system of inequalities as usual. This system has no solution, so we employ the methods suggested in section 2.4 for dealing with misspecification.

MS stands for the maximum score estimation methods; the set of maximum score estimates is the union of several disjoint convex polyhedra, and the reported bounds are the sharp bounds for this union. MNCE stands for the method of minimal number of classification errors described in section 1.4.2; the set of MNCE estimates is the union of several disjoint convex polyhedra, and the reported bounds are the sharp bounds for that union. MGCE

Table 1.3. Estimation of labor force participation

	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)
MGCE1	(6.9065, 6.9224)	(0.63395, 0.63427)	(-0.51237, -0.51201)	(-7.5439, -7.5389)
MGCE2	(4.1547, 4.1805)	(0.58402, 0.58466)	(-0.45484, -0.45419)	(-6.8742, -6.8653)
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)

stands for the minimal general classification error method outlined in section 1.4.3. MGCE1 minimizes the general classification error defined in (1.4.5), whereas MGCE2 minimizes the weighted general classification error

$$\sum_{l=1}^{d} |P^l - 0.5| q^l v_l.$$

Method MGCE1 treats all points in the support as being equally important. Method MGCE2 accounts for two factors, one of which is a higher probability of occurring of the points, and the other one is the distance of conditional probabilities to 0.5. The sets of MGCE1 and MGCE2 estimates are convex polyhedra. MRC stands for the maximum rank correlation method; the set of maximum rank correlation estimates is the union of several disjoint convex polyhedra (in $\Re^3$), and the bounds shown in Table 1.3 are the sharp bounds for this union.

We 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, MGCE1, MGCE2 and MRC are consistent with each other. For each regressor, methods MGCE1 and MGCE2 provide shorter intervals than methods MS, MNCE and MRC. This does not come as a surprise, because, first of all, MGCE1 and MGCE2 find only a subset of separating hyperplanes, and, second, we can show that this subset always lies in a hyperplane in the space $\Re^{k-1}$ (in our case, in the space R^4).

1.9 Conclusion

In this paper, we 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. We 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.

We go beyond the identification issue by investigating the estimation of the identification region and examining model's misspecification, which we approach in several different ways and provide insight into its possible causes and consequences. We 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.

1.10 Appendix

Proof of Proposition 1.2.

In the proof of Proposition 3.1, the first part is intuitive, so we focus on the second claim. Let $(b_3^*, \ldots, b_k^*)$ be a solution of (S_2). If we plug these numbers into D_i, N_j, Z_m, we will obtain numbers D_i^*, N_j^*, Z_m^* such that

$$N_j^* \geq D_i^*$$

$$Z_m^* \geq 0.$$

If we take any b_2 such that

$$N_j^* \geq b_2 \geq D_i^*,$$

then $(b_2, b_3^*, \ldots, b_k^*)$ is a solution of (S_1).

Proof of Proposition 1.4

In the system

$$z_{11} + z_{12}b_2 + z_{13}b_3 \geq 0$$

$$\dots$$

$$z_{d1} + z_{d2}b_2 + z_{d3}b_3 \geq 0,$$

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

$$b_3 \leq \frac{\frac{z_{i1}}{z_{i2}} - \frac{z_{j1}}{z_{j2}}}{\frac{z_{j3}}{z_{j2}} - \frac{z_{i3}}{z_{i2}}} = \frac{z_{i1}z_{j2} - z_{j1}z_{i2}}{z_{j3}z_{i2} - z_{i3}z_{j2}} = -\frac{\begin{vmatrix} z_{j1} & z_{j2} \\ z_{i1} & z_{i2} \end{vmatrix}}{\begin{vmatrix} z_{j3} & z_{j2} \\ z_{i3} & z_{i2} \end{vmatrix}}, \tag{1.10.1}$$

where

$$\begin{vmatrix} z_{j3} & z_{j2} \\ z_{i3} & z_{i2} \end{vmatrix} = z_{i2}z_{j2}\left(\frac{z_{j3}}{z_{j2}} - \frac{z_{i3}}{z_{i2}}\right) < 0. \tag{1.10.2}$$

Because (1.10.1) holds for an arbitrary i and j such that $z_{j2} < 0$, $z_{i2} > 0$ and (1.10.2) 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, we 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 1.5

Our proof proceeds by induction on k. As has been proved above, Proposition 1.5 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

$$z_{11} + z_{12}b_2 + \ldots + z_{1k}b_k + z_{1,k+1}b_{k+1} \geq 0$$

$$z_{21} + z_{22}b_2 + \ldots + z_{2k}b_k + z_{2,k+1}b_{k+1} \geq 0$$

$$\ldots$$

$$z_{n1} + z_{n2}b_2 + \ldots + z_{nk}b_k + z_{n,k+1}b_{k+1} \geq 0,$$

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, ..., 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, we 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 1.6

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

$$P^l \geq 0.5 \quad \Leftrightarrow \quad x^l b \geq 0, \quad l = 1, \ldots, d,$$

which, in turn, defines set B. Thus, in this case, $B = B^{ms}$.

If, for instance, $P^1 = 0.5$, then any b satisfying the system of inequalities

$$P^l \geq 0.5 \quad \Leftrightarrow \quad x^l b \geq 0, \quad l = 2, \ldots, d$$

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

Proof of Proposition 1.7

This proof is based on the symmetrical property of the formulas for b_3^l and b_3^u. According to the formulas in Proposition 1.4,

$$b_3^u \leq -\frac{\begin{vmatrix} z_{j_2,1} & z_{j_2,2} \\ z_{i_1,1} & z_{i_1,2} \end{vmatrix}}{\begin{vmatrix} z_{j_2,3} & z_{j_2,2} \\ z_{i_1,3} & z_{i_1,2} \end{vmatrix}} = \frac{\begin{vmatrix} z_{j_2,1} & -z_{i_2,2} \\ z_{i_1,1} & z_{i_1,2} \end{vmatrix}}{\begin{vmatrix} z_{i_2,3} & z_{i_2,2} \\ z_{i_1,3} & z_{i_1,2} \end{vmatrix}},$$

$$b_3^l \geq -\frac{\begin{vmatrix} z_{j_1,1} & z_{j_1,2} \\ z_{i_2,1} & z_{i_2,2} \end{vmatrix}}{\begin{vmatrix} z_{j_1,3} & z_{j_1,2} \\ z_{i_2,3} & z_{i_2,2} \end{vmatrix}} = -\frac{\begin{vmatrix} z_{j_1,1} & -z_{i_1,2} \\ z_{i_2,1} & z_{i_2,2} \end{vmatrix}}{\begin{vmatrix} z_{i_2,3} & z_{i_2,2} \\ z_{i_1,3} & z_{i_1,2} \end{vmatrix}},$$

and, hence,

$$b_3^u - b_3^l \leq \frac{\begin{vmatrix} z_{j_1,1} & -z_{i_1,2} \\ z_{i_2,1} & z_{i_2,2} \end{vmatrix} + \begin{vmatrix} z_{j_2,1} & -z_{i_2,2} \\ z_{i_1,1} & z_{i_1,2} \end{vmatrix}}{\begin{vmatrix} z_{i_2,3} & z_{i_2,2} \\ z_{i_1,3} & z_{i_1,2} \end{vmatrix}} = \frac{z_{i_2,2}(z_{j_1,1} + z_{i_1,1}) + z_{i_1,2}(z_{j_2,1} + z_{i_2,1})}{\begin{vmatrix} z_{i_2,3} & z_{i_2,2} \\ z_{i_1,3} & z_{i_1,2} \end{vmatrix}} \leq$$

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

Proof of Corollary 1.2

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 1.7 are satisfied. Therefore,

$$b_3^u - b_3^l \leq \Delta \frac{x_2 + x_2^*}{x_2 x_3^* - x_2^* x_3}.$$

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 1.7 are satisfied. Therefore,

$$b_3^u - b_3^l \leq \Delta \frac{x_2 - x_2^*}{x_2 x_3^* - x_2^* x_3}.$$

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 1.7 are satisfied. Therefore,

$$b_3^u - b_3^l \leq \Delta \frac{-x_2 + x_2^*}{x_2 x_3^* - x_2^* x_3}$$

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 1.7 are satisfied. Therefore,

$$b_3^u - b_3^l \leq \Delta \frac{-x_2 - x_2^*}{x_2 x_3^* - x_2^* x_3}.$$

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

Proof of Theorem 1.9

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

therefore,

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

Because $P^l > 0.5$,

$$P^l > 0.5 \quad \text{and} \quad sgn(\hat{P}_N^l - 0.5) \neq sgn(P^l - 0.5) \quad \Rightarrow \quad P^l - \hat{P}_N^l > P^l - 0.5$$

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

then

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

The consistency property (1.7.1) and the fact that $|0.5 - P^l| > 0$ for any $x^l \in S(X)$ imply

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

It is evident now that

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

Proof of Corollary 1.3.

$$H(B_N, B) \neq 0 \quad \Rightarrow \quad B_N \neq B.$$

Therefore,

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

Proof of Corollary 1.4.

For any $\epsilon > 0$,

$$\tau_N H(B_N, B) \geq \epsilon \quad \Rightarrow \quad H(B_N, B) \neq 0.$$

Therefore,

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

Proof of Theorem 1.10

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

If $P^l > 0.5$, then

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

$$\leq Pr(P^l - \hat{P}_N^l > P^l - 0.5) \to 0 \text{ as } N \to \infty.$$

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,

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

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

If $P^l = 0.5$, then

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

(1.7.2) and (1.7.3) imply that

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

and, thus,

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

Proof of Theorem 1.11

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

Denote

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

Then

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

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 $Pr(sgn(\hat{P}_N^l - 0.5) \neq sgn(P^l - 0.5)) = Pr(V_N(x^l) < 0)$. 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 $Pr(sgn(\hat{P}_N^l - 0.5) \neq sgn(P^l - 0.5)) = Pr(V_N(x^l) \geq 0)$. 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 $P^l \neq 0.5$,

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

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 $x^l \in S(X)$,

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

This proves the first part of the theorem.

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

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

then

$$Pr(B_N \neq B) \geq Pr(V_N(x^l) < 0).$$

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.

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

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

Use the fact that for $j \geq 1$,

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

to obtain

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

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

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

Proof of Proposition 1.12

Suppose that $P^l \neq 0.5$ for any $x^l \in S(X)$. Then

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

and the proof proceeds in the same way as the proof of the first part of Proposition 1.11.

The proof of the second part, when $P^l = 0.5$ for some $x^l \in S(X)$, is the same as the proof of the second part of Proposition 1.11.

Chapter 2

Nonparametric Identification in Asymmetric Second-Price Auctions: A New Approach

2.1 Introduction

In auctions, researchers are often interested in learning models' economic primitives, particularly the joint distribution of bidders' values. Because this underlying distribution is not known a priori, it must be learned from the data. To obtain credible estimation results, a researcher must first study the identification question, which includes whether the distribution of interest is point identified. The importance of this issue has generated many methodological papers on identification in auction models. This chapter contributes to that literature.

I examine the nonparametric identification of the distributions of bidders' values in asymmetric second-price auctions. The identification analysis cannot be conducted without (a) imposing conditions on the joint distribution of bidders' signals and (b) specifying what data are available from the auctions' outcomes. In this chapter, I assume that bidders have private values and that the only available data pertain to the winner's identity and the transaction price. I also suppose that the data are obtained from repeated auctions and that at each auction, the bidders' values are independent draws from the same joint distribution.

It is well known that in second-price auctions within the private-values framework, a weakly dominant strategy for bidders entails submitting their true value.[1] I consider an equilibrium where bidders employ this strategy. In this case, even though the submitted bids directly reveal bidders' values, the joint distribution of these values cannot be identified nonparametrically without making further assumptions.[2] The point identification of the parameter of interest usually requires strengthening the model's assumptions. I show that in our problem, it suffices to assume that bidders' values are independent. There are three main issues to address in obtaining this result. First, I must identify the distribution functions nonparametrically to avoid incorrect assumptions about their form. Second, I must overcome the challenge posed by the asymmetry of the bidders participating in the auction. Finally, given that the transaction price is the value of the second-highest bid, I must base the

[1] See, for example, Vickrey (1961) or Krishna (2002).

[2] Athey and Haile (2002) show that in second-price auctions within the private-values paradigm, the distributions of bidders' values are not identified if some bids are not observed.

identification proof on the second-order statistic. (Identification would be relatively easy to prove if the first-order statistic were available.)

One of the main contributions of this chapter is to provide conditions on observable data sufficient to guarantee point identification. First, I present the conditions on the observables that are necessary and sufficient for the existence of a solution to the model; thus, we always know with certainty whether the model has a solution. I then show that these conditions, together with an additional condition on the observables, are sufficient to show the uniqueness of a solution and therefore to ensure the identification of distribution functions.

A methodological contribution of this chapter is to suggest a new approach to proving identification in analyzed auction models. The idea behind my method is to relate unknown underlying distribution functions to the observable data using a system of differential equations, then to establish the existence and uniqueness of a solution to this system. This strategy includes two major steps. First, I show that the system has a unique solution on a subinterval of the support; this is what I call a local solution. Second, I demonstrate that this local solution can be extended to the whole support. This two-step approach is constructive and enables us to conduct a thorough qualitative analysis of the identification problem.

Furthermore, the techniques I develop allow for two generalizations of the auction setting. One relaxes the support conditions and permits distributions to have different upper support points as well as holes in the support. The other considers second-price auctions in which the number of actual bidders is unknown and varies exogenously. Using the case of three bidders, I outline the specifics of proving identification in these models.

Another contribution of this chapter is to uncover conditions sufficient to detect misspecification. These conditions do not pertain to the observable data and can be verified only after we determine the solution to the system of differential equations. To address this issue, I present an example of a misspecified model that proves the possibility of finding well-defined observable functions such that the functions in the solution corresponding to them are not all monotone − that is, not all of them possess the properties of distribution functions.

In discussing the estimation issue, I propose a sieve minimum distance estimator of unknown distribution functions. I explore the properties of the operator that maps these functions into the observable data and find that it is continuous in the uniform metric. The properties of this operator allow us to show the consistency of the sieve estimator. In addition, I prove that under weak conditions on the set of underlying distribution functions, its inverse operator is continuous too.

Because assuming the independence of bidders' values may seem dubious in some applications, it is worthwhile to analyze auctions with affiliated private values. Though the marginal distribution functions of bidders' values are not identified in this situation, I derive tight point-wise bounds on them and outline how these bounds change when we acquire data on other bids.

Within the private-values framework, second-price auctions are equivalent to ascending auctions. For proofs of identification in these two types of auctions, particularly when the data indicate only the winner's identity and the transaction price, many researchers have referred to results in the statistical literature that examines identification in generalized competing risks models. Athey and Haile (2002) were first to observe that analyzed auctions can be considered a special case of these models. In generalized competing risks models, an object that consists of different components fails as a result of the cumulative failure of several of its elements.[3] Though the main identification result for these cases was obtained by Meilijson (1981), his proofs lack some essential details, most importantly, conditions on the observables that guarantee identification. I show that my method, on the other hand, provides an exhaustive proof of identification in generalized competing risks models. For any of these models, I provide conditions on the observables that guarantee that the model cannot have more than one solution. I also explain why the existence of a solution cannot be proved in general and must be assumed. For a special class of generalized competing risks models (one that encompasses our auction models), I present necessary and sufficient conditions for existence. Moreover, I discuss why the methods from the generalized competing risks literature cannot be applied to auctions with an unknown number of bidders, whereas my techniques can.

For a thorough overview of nonparametric identification in auctions, see Athey and Haile (2002, 2005, 2006) and references therein. The authors obtain numerous nonparametric identification results for various auctions settings, and some of the point identification results rely on the work of Meilijson (1981). Brendstrup and Paarsch (2006) deal specifically with asymmetric ascending auctions within the independent-private-values framework, consider-

[3] In classical competing risks models, an object fails as soon as one of its components reaches a failure state.

ing both single-unit and multi-unit settings. For nonparametric identification in single-unit auctions, they also refer to Meilijson (1981). They then apply their methods to data from fish auctions in Denmark and estimate the distribution functions of bidders' values by employing the quasi-maximum likelihood approach. Finally, they use these estimates to conduct a policy experiment that compares the performance of the ascending auction and the Dutch auction for that particular fish market. Banerji and Meenakshi (2004) and Meenakshi and Banerji (2005) also consider asymmetric ascending auctions within the independent-private-values framework by examining wheat markets in India. Similar to Brendstrup and Paarsch (2006), they cite Meilijson (1981) to show identification. They use the result to uncover collusion among a set of buyers and estimate its impact on market prices and other factors.

Another thread of the literature related to this chapter applies the techniques of the theory of differential equations to identification problems. In auctions, examples of such papers are Campo, Perrigne and Vuong (2003); Guerre, Perrigne and Vuong (2007) and Lebrun (1999). Campo, Perrigne and Vuong (2003) prove nonparametric identification for asymmetric first-price auctions with affiliated private values. Guerre, Perrigne and Vuong (2007) address the nonparametric identification of utility functions for bidders in first-price auctions, specifically when the bidders are risk averse and have private values. Lebrun (1999) analyzes first-price auctions with independent private values and characterizes a Bayesian equilibrium as a solution to a system of non-linear differential equations. He then refers to results in the theory of differential equations to show that an equilibrium exists and that in some special models, it is unique. In a related area of classical competing risks, Buera (2006) uses the theory of partial differential equations to prove identification in a certain class of Roy models.

The rest of this chapter is organized as follows. Section 2.2 states the identification problem in second-price auctions and provides a preview of the chapter's main results. It also introduces generalized competing risks models and explains their connection to second-price auctions. Finally, the section presents a mathematical description of the identification problem for the case of three bidders. Section 2.3, the central part of this chapter, provides a detailed proof of identification and indicates conditions on the observables that are used to obtain this result. Section 2.4 discusses misspecification and describes an example of a misspecified model. Section 2.5 contains two extensions of the results obtained in section 2.3: One concerns second-price auctions with any number of bidders, and the other relates to distributions with weaker support conditions. Section 2.6 proposes a sieve minimum distance estimator of distribution functions and shows its consistency. Section 2.7 presents two applications. The first considers second-price auctions in which the number of actual bidders is unknown and varies exogenously; the second deals with identification in generalized competing risks models. Section 2.8 concludes with a summary of contributions. The Appendix provides detailed proofs that do not appear in the text.

2.2 Statement of identification problem and preview of results

In this section, I first state the identification problem in second-price auctions. Next, I introduce generalized competing risks models and show their connection to these auctions. Finally, I summarize this chapter's main results.

2.2.1 Statement of identification problem

I start by reviewing the auction model. A single object is up for sale, and d buyers are bidding on it. The bids are submitted in sealed envelopes. The highest bidder wins and pays the value of the second-highest bid; thus, in these auctions, the second-highest bid is the transaction price. Suppose that all bidders are aware of their value. It is known that in this setting, a weakly dominant strategy for bidders is to submit their true value − and this is an equilibrium that I consider later.

Denote bidders' private values as X_i, $i = 1, \ldots, d$. Assume that these values are independent and have absolutely continuous distributions on a common support $[t_0, T]$. Also assume that bidders' values at each auction are independent draws from the same joint distribution. We aim to learn this distribution from the available data. Note that in the equilibrium, the bids' joint distribution coincides with the distribution of the bidders' private values. Therefore, if all the bids are observed, then the distribution of values can be clearly identified. If some of them are not observed, however, then neither the joint nor the marginal value distributions can be identified, as shown in Athey and Haile (2002). Given that our knowledge is often limited to the second-highest bid, I show that when the only available data pertain to the bid and the winner's identity, the marginal distributions of bidders' values can be identified if these values are independent.

It is worth mentioning that within the private-values framework, second-price auctions are equivalent to open ascending auctions. One form of ascending auctions is a "button auction," in which bidders hold down a button as the auctioneer raises the price. When the price gets too high for a bidder, she drops out by releasing the button. The auction ends when only one bidder remains. This person wins the object and pays the price at which the auction stopped.

Second-price auctions and generalized competing risks

Now I turn to a brief description of generalized competing risks models. To clarify the connection between these models and second-price auctions, I use the equivalence of second-price and ascending auctions within the private-values paradigm.

Consider a button auction, as described above, with d bidders, and suppose that only the winner's identity and the transaction price are observed. Notice that observing the identity of the winner is equivalent to observing the identities of the bidders who dropped out.[4]

Compare this auction framework to the following model from reliability theory. Consider a machine that consists of d elements, and assume that it works as long as at least two of its elements are functioning; in other words, the machine breaks once $d-1$ of its elements are

[4] I suppose that the set of all bidders is known.

dead. The set of these $d-1$ elements that caused the machine's failure is called a fatal set. (Clearly, the breakdown of other $d-1$ components would also be fatal.) Now suppose that when the machine dies, only its lifetime and the set of elements that caused its breakdown are recorded. This model is equivalent to the auction setting: A fatal set is an analog of the set of bidders that dropped out, and the machine's lifetime is an analog of the transaction price.

The examples of the button auction and the machine breakdown represent special cases of ***generalized competing risks models***. Classical competing risks models are another special case. The classical models correspond to a situation in which a machine breaks down as soon as one of its components fails; the data available after the breakdown are the machine's lifetime and the element that caused the failure. In economics, the widely known Roy model is isomorphic to classical competing risks models. In the Roy model, a person chooses from a finite set of occupational alternatives to obtain the highest income, and the outcomes of the choice (occupation and income) are observed. More details about generalized competing risks and their connection to the analyzed auction model is given in section 2.7.2.

Notation

Throughout this chapter, I use the following notations. A bid submitted by player i is denoted as b_i. Symbol M^{tr} represents the transpose of matrix M, and $\|\|_1$ stands for a norm in $\Re^d$ defined as $\|x\|_1 = \sum_{i=1}^{d} |x_i|$. The distribution functions of X_i are denoted as F_i, $i = 1, \ldots, d$. Function F_i is called positive (negative) if $F_i(t) > 0$ for $t > t_0$. A vector-valued function $F = (F_1, \ldots, F_d)^{tr}$ on $[t_0, T]$ is called positive (negative) if each of its components F_i is a positive (negative) function. F is referred to as strictly increasing if each F_i is strictly increasing on $[t_0, T]$.

For simplicity, I first consider the case of three bidders, then generalize the results to any number of bidders. Next, I turn to stating the identification problem. I assume that only the winner's identity and the transaction price are observed in an auction's outcome. Under this assumption, the probability of an event $\{\text{price} \leq t,\, i \text{ wins}\}$ is known for any $t \in [t_0, T]$ and any $i = 1, 2, 3$. So, for each bidder i, we observe the following function G_i on $[t_0, T]$:

$$G_i(t) = Pr(\text{price} \leq t,\, i \text{ wins}), \quad i = 1, 2, 3.$$

The identification problem is to determine whether there is only one collection of private values distribution functions F_1, F_2 and F_3 that rationalize observable functions G_1, G_2 and G_3.

2.2.2 Statement and discussion of main results

In this section, I formulate the main identification result of this chapter as well as some related conclusions.

I will say that *the model is misspecified* if at least one of the following conditions fails to hold: 1) bidders submit their true values; 2) bidders have independent private values; 3) bidders' values have absolutely continuous distributions; 4) bidders' values are distributed on $[t_0, T]$.

The next proposition indicates ***necessary conditions on observable*** G_i implied by the model.

Proposition 2.1. *If the model is well specified, then the following conditions hold:*

Necessary conditions (I)

1. $G_i(t_0) = 0$, $i = 1, 2, 3$
2. G_i *is absolutely continuous on* $[t_0, T]$, $i = 1, 2, 3$
3. G_i *is strictly increasing on* $[t_0, T]$, $i = 1, 2, 3$

Proof. By assumption, the distributions of private values X_i are absolutely continuous. This implies, in particular, that players submit bids equal to t_0 with probability 0. Also, t_0 is the lower support point for all distributions. These two facts imply Condition 1. Condition 2 follows from the absolute continuity of the distributions of X_i. Condition 3 is true because the support of each X_i is the connected interval $[t_0, T]$, without any holes in it. □

Even though these conditions are simple, it is worth indicating them because they are useful in the proof of identification. As we can see, all the properties of the private values distributions, except for the assumption of independence and the boundary conditions $F_i(T) = 1$, $i = 1, 2, 3$, are used in establishing Proposition 2.1. The independence assumption, combined with necessary conditions (I), allows us to obtain the following result.

Proposition 2.2. *Suppose that the model is well specified. Let* F *be a solution to the model. Then*

$$\lim_{t \downarrow t_0} \frac{F_1}{\sqrt{\frac{G_2 G_3}{G_1}}}(t) = 1, \quad \lim_{t \downarrow t_0} \frac{F_2}{\sqrt{\frac{G_1 G_3}{G_2}}}(t) = 1, \quad \lim_{t \downarrow t_0} \frac{F_3}{\sqrt{\frac{G_1 G_2}{G_3}}}(t) = 1. \tag{2.2.1}$$

Conditions (2.2.1) are formulated in terms of both observable and unobservable functions. Note that they characterize a solution F to the model only in a neighborhood of t_0. To be more precise, they find the rate of convergence of unknown distribution functions F_i at t_0 in terms of observable functions G_i. These conditions will be essential for proving identification.

Proposition 2.2 implies the following properties of the observable functions.

Corollary 2.1. *Suppose that the model is well specified. Then the following conditions hold:*

Necessary conditions (II)

$$\lim_{t \downarrow t_0} \frac{G_2 G_3}{G_1}(t) = 0, \quad \lim_{t \downarrow t_0} \frac{G_1 G_3}{G_2}(t) = 0, \quad \lim_{t \downarrow t_0} \frac{G_1 G_2}{G_3}(t) = 0. \qquad (2.2.2)$$

These conditions play an important role in identification. The reasoning behind them is that, no matter how different the underlying distributions are, the observable functions do not have considerably different rates of convergence at t_0.

Now that I have presented necessary conditions on observables, I turn to describing the mathematical model of identification and explain how necessary conditions (I) and (II) are employed in the identification proof.

Mathematical model of identification problem

Assuming the independence of bidders' values, functions G_i can be expressed through F_i as follows. Let b_i, $i = 1, 2, 3$, indicate the submitted bids. Then

$$G_1(t) = Pr(\max\{b_2, b_3\} < b_1, \max\{b_2, b_3\} \leq t) = \int_{t_0}^{t} (F_2 F_3)'(1 - F_1) ds,$$

where integration is to be understood in the sense of Lebesgue. Functions G_2 and G_3 have similar expressions. Therefore, unknown distribution functions F_i are related to observable functions G_i by means of this system of integral-differential equations:

$$\begin{aligned} G_1(t) &= \int_{t_0}^{t} (F_2F_3)^{'}(1-F_1)ds \\ G_2(t) &= \int_{t_0}^{t} (F_1F_3)^{'}(1-F_2)ds \\ G_3(t) &= \int_{t_0}^{t} (F_1F_2)^{'}(1-F_3)ds. \end{aligned} \tag{2.2.3}$$

Notice that the left-hand and right-hand sides of the equations in (2.2.3) are absolutely continuous functions, allowing us to differentiate them and obtain the following system of differential equations almost everywhere (a.e.) on $[t_0, T]$:

$$\textit{Main system} \qquad \begin{aligned} g_1 &= (F_2F_3)^{'}(1-F_1) \\ g_2 &= (F_1F_3)^{'}(1-F_2) \\ g_3 &= (F_1F_2)^{'}(1-F_3), \end{aligned} \tag{DE}$$

where g_i stands for the a.e. derivative of G_i. I will refer to system (DE) as the main system. Distribution functions F_i in this system must satisfy the following initial conditions:

$$\textit{Initial conditions} \qquad F_i(t_0) = 0, \quad i = 1, 2, 3. \tag{IC}$$

I will refer to problem (DE)-(IC) as the main problem. The definition below explains the meaning of a solution to (DE)-(IC).

Definition 2.1. *Function $F = (F_1, F_2, F_3)^{tr}$ is a solution to problem (DE)-(IC) on an interval $[t_0, t_0 + a]$ if F_i, $i = 1, 2, 3$, are absolutely continuous on $[t_0, t_0 + a]$, satisfy equations (DE) a.e. on $[t_0, t_0 + a]$ and satisfy (IC).*

The system of differential equations (DE) is a convenient tool because identifying functions F_i is equivalent to proving that problem (DE)-(IC) has a unique positive solution F on $[t_0, T]$.

Notice that I did not mention anything about the monotonicity of the solution. There are two reasons for that. First, as I will show in section 2.4, functions F_i in a solution to (DE)-(IC) may be non-monotone. Second, it is not clear whether one can find conditions on G_i that guarantee the monotonicity of all F_i. Therefore, the monotonicity of a solution to problem (DE)-(IC) will be assumed.

The theorems below formulate the existence result for problem (DE)-(IC).

Theorem 2.3. ***(Existence of a solution)*** *Let observable functions G_i satisfy conditions (I) and (II). Then problem (DE)-(IC) has a solution on $[t_0, T]$.*

Remember that all conditions on G_i required in this proposition are necessary conditions implied by the model. Therefore, conditions (I) and (II) are necessary and sufficient conditions for the existence of a solution to the model. This is a powerful result because from the available data, we know with certainty whether the model has a solution. If even one of the conditions in (I) and (II) fails to hold, it can be immediately concluded that the model is misspecified.

The next theorem implies that the underlying distribution functions F_i are identified.

Theorem 2.4. ***(Uniqueness of a solution)*** *Let observable functions G_i satisfy conditions (I), (II) and*

Sufficient condition (III): the function

$$\left(\frac{g_1}{G_1}+\frac{g_2}{G_2}+\frac{g_3}{G_3}\right)\left(\sqrt{\frac{G_2G_3}{G_1}}+\sqrt{\frac{G_1G_3}{G_2}}+\sqrt{\frac{G_1G_2}{G_3}}\right) \tag{2.2.4}$$

is Lebesgue integrable − that is, belongs to the class L^1 *− in a neighborhood of* t_0.

Then problem (DE)-(IC) has a unique solution on $[t_0, T]$.

Theorems 2.3 and 2.4 will follow from results obtained in section 2.3.

The most important element in obtaining sufficient condition (III) is the result of Proposition 2.2. To acquire a better understanding of this condition, I write it in terms of distribution functions F_i.

Remark. *If* F_i *are absolutely continuous distribution functions with support* $[t_0, T]$*, then condition (III) is equivalent to the following condition:*

The function

$$\left(\frac{F_1'}{F_1}+\frac{F_2'}{F_2}+\frac{F_3'}{F_3}\right)(F_1+F_2+F_3) \tag{2.2.5}$$

is Lebesgue integrable in a neighborhood of t_0.

Now it is intuitive that the reasoning behind this condition is that the underlying distribution functions F_1, F_2 and F_3 are not too different in a certain sense. For instance, if the underlying distribution functions are $F_1 = t$, $F_2 = t$ and $F_3 = \exp^{1-\frac{1}{t^2}}$ on $[0, 1]$, then the corresponding observable functions G_i do not satisfy condition (III) in a neighborhood of 0. Figure 2.1 depicts these distribution functions F_i. As we can see, the value distribution for the third bidder has a very small mass around point 0, whereas values for the first and second bidders are distributed uniformly on $[0,1]$. Figure 2.2 shows a non-integrable function that has the same behavior in a neighborhood of point 0 as the observable function in (2.2.5).

It is worth mentioning that if F_i represent common parametric distributions, such as uniform, gamma, normal (truncated) distributions, et cetera, then the corresponding observable functions G_i satisfy condition (III).

2.3 Identification

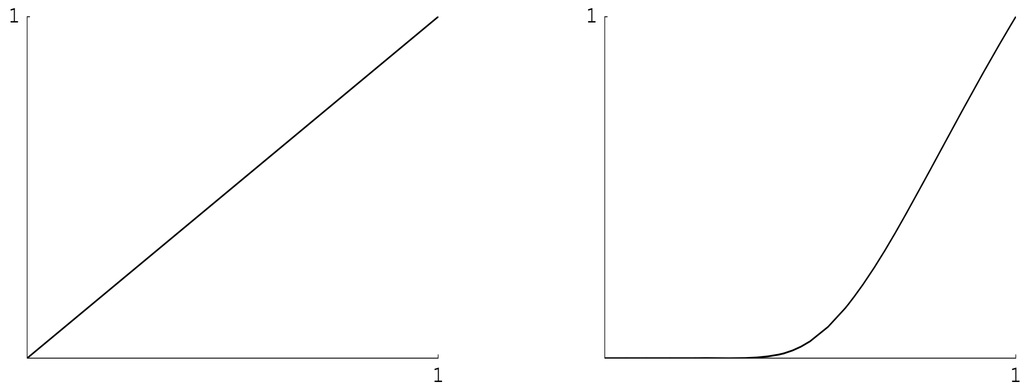
This is the main section of the chapter. I first describe the strategy to proving identification. Then I implement this strategy and establish the identification result. Detailed proofs of theorems and lemmas that do not appear in the text are collected in the appendix.

2.3.1 Identification strategy

I start with a brief discussion of system (DE). I will say that functions F_i are locally identified if problem (DE)-(IC) has a unique positive solution F in a small neighborhood of t_0; this solution is what I call a local solution. I will say that functions F_i are globally identified if problem (DE)-(IC) has a unique positive solution F on the whole support $[t_0, T]$; this solution is what I call a global solution.[5] I introduce these concepts because my strategy for proving identification consists of two logical steps: first establishing local identification, then global identification.

It can be shown that (DE)-(IC) always has a negative local solution as well as a positive

[5] Clearly, global identification is simply the identification of F_i.



$F_1 = t,\ F_2 = t$ $\qquad\qquad$ $F_3 = \exp^{1-\frac{1}{t^2}}$

Figure 2.1. Underlying distribution functions.

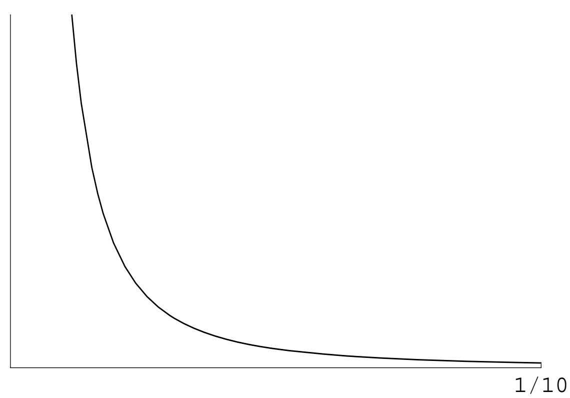


Figure 2.2. Function that has the same behavior in a neighborhood of point 0 as the observable function in (2.2.5).

local solution.[6] Conditions for uniqueness in the theory of differential equations do not let us control the sign of solutions. Therefore, even though I am interested only in a positive solution and can neglect a negative one, sufficient conditions that guarantee uniqueness of a positive local solution cannot be derived from system (DE). To tackle this problem, I use auxiliary tools.

Auxiliary tools

I transform (DE) into a new system by introducing auxiliary functions H_1, H_2, H_3:

$$H_1 = F_2F_3, \quad H_2 = F_1F_3, \quad H_3 = F_1F_2.$$

Clearly, these functions are the distribution functions of $\max\{X_2, X_3\}$, $\max\{X_1, X_3\}$ and $\max\{X_1, X_2\}$, respectively. Functions F_i are expressed through H_i as $F_1^2 = \frac{H_2H_3}{H_1}$, $F_2^2 = \frac{H_1H_3}{H_2}$, $F_3^2 = \frac{H_1H_2}{H_3}$. Taking into account that F_i must be positive, I obtain

$$F_1 = \sqrt{\frac{H_2H_3}{H_1}}, \quad F_2 = \sqrt{\frac{H_1H_3}{H_2}}, \quad F_3 = \sqrt{\frac{H_1H_2}{H_3}}. \tag{2.3.1}$$

Thus, for any point $t > t_0$, system (DE) can be written as

$$H_1' = \frac{g_1}{1 - \sqrt{\frac{H_2H_3}{H_1}}}$$

$$H_2' = \frac{g_2}{1 - \sqrt{\frac{H_1H_3}{H_2}}}$$

$$H_3' = \frac{g_3}{1 - \sqrt{\frac{H_1H_2}{H_3}}}.$$

[6] See section 2.3.2 for further explanation.

Note that initial conditions $H_i(t_0) = 0$ cannot be imposed because the right-hand sides of the equations in this system are undefined when $H_i = 0$. Instead, I can set conditions on the upper limit of H_i at t_0:

$$\lim_{t \downarrow t_0} H_i(t) = 0, \quad i = 1, 2, 3. \tag{IC_H}$$

The right-hand side of the last system is a vector-valued function that depends on t, H_1, H_2 and H_3. Denote it as $J(t, H)$:

$$J(t,H) = \left(\frac{g_1(t)}{1 - \sqrt{\frac{H_2 H_3}{H_1}}}, \frac{g_2(t)}{1 - \sqrt{\frac{H_1 H_3}{H_2}}}, \frac{g_3(t)}{1 - \sqrt{\frac{H_1 H_2}{H_3}}} \right)^{tr}, \tag{2.3.2}$$

and rewrite the last system as

$$H'(t) = J(t, H(t)). \tag{DE_H}$$

I will refer to system (DE_H) as an auxiliary system and to problem (DE_H)-(IC_H) as an auxiliary problem.

Definition 2.2. *Function $H = (H_1, H_2, H_3)^{tr}$ is a solution to (DE_H)-(IC_H) on an interval $(t_0, t_0 + a]$ if H_i are absolutely continuous on $(t_0, t_0 + a]$, satisfy (DE_H) a.e. on $(t_0, t_0 + a]$ and also satisfy (IC_H).*

Proof roadmap

I now outline an approach to the identification proof. It comprises two major parts: establishing the local identification result and the global identification result.

The local identification result is proved in steps. In the first step, I show that conditions (I) and (II) are sufficient to guarantee that problem (DE_H)-(IC_H), which is the auxiliary problem, has a local solution. In the second step, I use formulas (2.3.1) to find F_i from H_i and show that these F_i constitute a local solution to the main problem. Lastly, for the auxiliary problem, I establish that its local solution that was found in the first step is unique. This implies that for the main problem, its local solution that was found in the second step is unique.

The global identification result is obtained from the local identification result by showing how the unique local solution to (DE)-(IC) can be extended to the unique local solution on the whole support. The idea is to extend this local solution to small intervals progressively farther to the right until the upper support point T is reached.

2.3.2 Local identification

Proving local identification is the most challenging part of the identification proof. I show that to establish the existence of a local solution, I only need conditions (I) and (II). To obtain local uniqueness, I use condition (III) as well as (I) and (II).

Existence of a local solution

I now turn to conducting the first and the second step in the identification strategy. I start by finding an interval on which a local solution to the auxiliary problem (DE_H)-(IC_H) and a local solution to the main problem (DE)-(IC) exist. Then I prove local existence for (DE_H)-(IC_H) and use this result to establish local existence for (DE)-(IC).

Before moving on, I must introduce some notations and carry out preliminary technical work. First of all, I have to indicate the domain of function $J(t,H)$. Take into account formulas (2.3.1), which express F through H, and note that for the auxiliary problem, we want to prove not only that there is a local solution but also that this solution is such that functions $\frac{H_2H_3}{H_1}$, $\frac{H_1H_3}{H_2}$, $\frac{H_1H_2}{H_3}$ take values less than 1 and the following conditions hold:

$$\lim_{t\downarrow t_0}\frac{H_2H_3}{H_1}(t)=0,\quad \lim_{t\downarrow t_0}\frac{H_1H_3}{H_2}(t)=0,\quad \lim_{t\downarrow t_0}\frac{H_1H_2}{H_3}(t)=0.$$

This accords with the fact that for function $J(t,H)$ to be well defined, the denominators in this function must be separated from 0. To do that, choose any $\delta\in(0,1)$ and allow H to take values only in the following sets:

$$\bar{H}_0(\delta)=(0,\infty)^3\cap\{(h_1,h_2,h_3)^{tr}:h_2h_3\le\delta h_1,h_1h_3\le\delta h_2,h_2h_3\le\delta h_1\}.$$

Let $\bar{D}_0(\delta)=[t_0,T]\times\bar{H}_0(\delta)$ be the domain of $J(t,H)$ (a.e. with respect to t). As we can see, δ guarantees that the denominators in $J(t,H)$ are separated from 0 by the value $1-\sqrt{\delta}$.

To determine an interval of existence for a local solution, I use conditions (II). Choose $\gamma>0$ such that $\gamma/(1-\sqrt{\delta})^2\le\delta$. Let t_0+a, $a>0$, be a point from $[t_0,T]$ such that

$$\forall(t\in[t_0,t_0+a])\quad \frac{G_2G_3}{G_1}(t)\le\gamma,\quad \frac{G_1G_3}{G_2}(t)\le\gamma,\quad \frac{G_1G_2}{G_3}(t)\le\gamma. \tag{2.3.3}$$

Conditions (II) guarantee that such t_0+a exists. Interval $[t_0,t_0+a]$ is an interval on which a solution to problem (DE_H)-(IC_H) exists.

The next proposition formulates the local existence result for the auxiliary problem.

Proposition 2.5. *Let observable functions G_i satisfy conditions (I) and (II). Let $J(t,H)$ be defined on $\bar{D}_0(\delta)$. Then (DE_H)-(IC_H) has a solution on $(t_0,t_0+a]$.*

It is remarkable that this existence result does not require any assumptions on observable G_i besides necessary conditions, which are satisfied in the model.

The proof of this proposition, which is in the appendix, implies that if we take a solution H to (DE_H)-(IC_H) on $(t_0, t_0 + a]$ and define the function for t_0 as $H(t_0) = (0, 0, 0)^{tr}$, then this extended function is absolutely continuous on $[t_0, t_0 + a]$ and clearly satisfies (DE_H)-(IC_H) a.e. on $[t_0, t_0 + a]$. In other words, a solution H can be extended from $(t_0, t_0 + a]$ to $[t_0, t_0 + a]$.

The following explanation shows why I cannot use standard existence theorems to prove Proposition 2.5. A general form of a system of differential equations is

$$x'(t) = v(t, x(t)),$$

where x and v are vector-valued functions. Let the initial condition be

$$x(t_0) = x_0.$$

In our problem, x is function H, and $v(t, x)$ is $J(t, H)$.[7] Existence theorems are usually proved for the situation in which the domain of v is $[t_0 - h, t_0 + h] \times B(x_0)$ or $[t_0, t_0 + h] \times B(x_0)$, where $B(x_0)$ is an open ball with the center in x_0.[8] This property implies, for example, that x_0 is an interior point in the domain of v with respect to x. Existence theorems are also proved for some more general cases, but all require, at the very least, x_0 to be an interior point in the domain of v with respect to x, and this domain must satisfy certain properties.

[7]Even though initial conditions (IC_H) characterize the limit at t_0 rather than the value at t_0, this does not matter because, as I mentioned above, solution H can be extended from $(t_0, t_0 + a]$ to $[t_0, t_0 + a]$.

[8]For systems with discontinuous right-hand sides, this result is illustrated in Filipov (1988).

Because of the specificity of set $\bar{D}_0$ and the fact that the point of the initial conditions is on the border of $\bar{D}_0$, I cannot apply any of those results.

Now that I have established the local existence result for the auxiliary problem (DE_H)-(IC_H), I can turn to proving that the main problem (DE)-(IC) has a local solution. This result is easy to obtain if we recall how H and F are related in formulas (2.3.1).

Theorem 2.6. *Let observable functions G_i satisfy conditions (I) and (II). Then (DE)-(IC) has a solution on $[t_0, t_0 + a]$.*

Proof. Let H be a solution to (DE_H)-(IC_H) on $(t_0, t_0 + a]$. For $t > t_0$, define F_i according to formulas (2.3.1), and let $F_i(t_0) = 0$, $i = 1, 2, 3$. It follows from (DE_H) that the ratios $\frac{H_i}{G_i}$ have finite positive limits when $t \to t_0$. Therefore, functions F_i are continuous at t_0 because for some constant C,

$$\lim_{t \downarrow t_0} F_1(t) = \lim_{t \downarrow t_0} \sqrt{\frac{H_2 H_3}{H_1}}(t) = C \lim_{t \downarrow t_0} \sqrt{\frac{G_2 G_3}{G_1}}(t) = 0$$

and, similarly, $\lim_{t \downarrow t_0} F_2(t) = 0$, $\lim_{t \downarrow t_0} F_3(t) = 0$. Because functions F_i are absolutely continuous on $[t_0 + \Delta, t_0 + a]$ for any $\Delta \in (0, a)$, and continuous at point t_0, they are absolutely continuous $[t_0, t_0 + a]$. It is evident that F_i solve equations (DE) a.e. on $[t_0, t_0 + a]$. □

Observe that because $J(t, H)$ is defined on $\bar{D}_0(\delta)$ and therefore a solution H to (DE_H)-(IC_H) takes values only in $\bar{H}_0(\delta)$, the values of the corresponding functions F_i belong to $[0, \sqrt{\delta}]$ only. The goal, however, is to identify F_i for all values in $[0, 1]$. This will be possible because δ can be arbitrarily close to 1.

The last thing that I want to mention here concerns the remark made in section **??** about the existence of a negative function F that satisfies (DE) a.e. in a neighborhood of t_0 and also satisfies (IC). Note that functions F_i are expressed through H_i as $F_1^2 = \frac{H_2 H_3}{H_1}$, $F_2^2 = \frac{H_1 H_3}{H_2}$, $F_3^2 = \frac{H_1 H_2}{H_3}$, as follows from the definition of functions H_i. Taking into account that F_i are positive, I obtained (2.3.1) and substituted these formulas into (DE) to obtain the auxiliary system (DE_H). However, if I were looking for negative solutions, I would substitute formulas

$$F_1 = -\sqrt{\frac{H_2 H_3}{H_1}}, \quad F_2 = -\sqrt{\frac{H_1 H_3}{H_2}}, \quad F_3 = -\sqrt{\frac{H_1 H_2}{H_3}}$$

into (DE) and obtain a different form of the auxiliary system:

$$H_1' = \frac{g_1}{1 + \sqrt{\frac{H_2 H_3}{H_1}}}$$

$$H_2' = \frac{g_2}{1 + \sqrt{\frac{H_1 H_3}{H_2}}}$$

$$H_3' = \frac{g_3}{1 + \sqrt{\frac{H_1 H_2}{H_3}}}.$$

Using the techniques of this section, it can be shown that this system with initial conditions (IC_H) has a local solution H. This implies there is a negative function F that solves (DE) a.e. in a neighborhood of t_0.

Uniqueness of a local solution

The next step in the proof of local identification is to show that (DE)-(IC) has only one local solution. Local existence was proved without imposing any assumptions on G_i besides necessary conditions (I) and (II). To establish local uniqueness, I will assume that condition (III) is also satisfied. In fact, condition (III) is key to proving local uniqueness.

I start by stating the local uniqueness result.

Theorem 2.7. *Let observable functions G_i satisfy conditions (I), (II) and (III). Then (DE)-(IC) has only one solution in a neighborhood of t_0.*

The idea of the proof of this theorem is to take two local solutions to problem (DE)-(IC) and show that they coincide on their common interval of existence.

Suppose that F and $\tilde{F}$ are two local solutions to (DE)-(IC) with a common interval of existence $[t_0, t_0 + c]$, $c > 0$. Let H_i and $\tilde{H}_i$ be corresponding to them auxiliary functions:

$$H_1 = F_2 F_3, \quad H_2 = F_1 F_3, \quad H_3 = F_1 F_2,$$

$$\tilde{H}_1 = \tilde{F}_2 \tilde{F}_3, \quad \tilde{H}_2 = \tilde{F}_1 \tilde{F}_3, \quad \tilde{H}_3 = \tilde{F}_1 \tilde{F}_2.$$

Clearly, if functions H and $\tilde{H}$ are identical, then F and $\tilde{F}$ coincide.

The lemma below is key to proving that functions H and $\tilde{H}$ are identical.

Lemma 2.8. *Functions H and $\tilde{H}$ satisfy the following inequality a.e. on $[t_0, t_0 + c]$:*

$$\|H'(t) - \tilde{H}'(t)\|_1 \leq \Gamma_0(t) \|H(t) - \tilde{H}(t)\|_1, \tag{2.3.4}$$

where

$$\Gamma_0(t) = C \left(\frac{g_1}{G_1}(t) + \frac{g_2}{G_2}(t) + \frac{g_3}{G_3}(t) \right) \left(\sqrt{\frac{G_2 G_3}{G_1}(t)} + \sqrt{\frac{G_1 G_3}{G_2}(t)} + \sqrt{\frac{G_1 G_2}{G_3}(t)} \right).$$

Establishing inequality (2.3.4) is the most challenging part of proving local uniqueness. This result relies mostly on conditions (2.2.1), which find the rate of convergence of F_i at t_0 in terms of observable functions G_i.

Notice that because H and $\tilde{H}$ solve the auxiliary problem (DE_H)-(IC_H), then a.e. on $(t_0, t_0 + c]$

$$H'(t) = J(t, H(t)),$$

$$\tilde{H}'(t) = J(t, \tilde{H}(t)).$$

Therefore, inequality (2.3.4) can be rewritten as

$$\|J(t, H(t)) - J(t, \tilde{H}(t))\|_1 \leq \Gamma_0(t)\|H(t) - \tilde{H}(t)\|_1.$$

This last inequality is a generalized local Lipschitz condition for function $J(t, H)$ with respect to variable H.

The following two lemmas prove that inequality (2.3.4) together with condition (III) imply that H and $\tilde{H}$ are identical functions.

Lemma 2.9. *Let $z : [\tau, \xi] \to \Re^n$ be an absolutely continuous function. Then $\|z\|_1$ has the right derivative $D_R\|z\|_1$ a.e. on $[\tau, \xi]$, and*

$$D_R\|z(t)\|_1 \leq \|z'(t)\|_1 \quad \text{a.e. on } [\tau, \xi].$$

Lemma 2.10. *Let function $v : [\tau, \xi] \to \Re$ be absolutely continuous. Suppose that $v(\tau) = 0$, and a.e. on $[\tau, \xi]$*

$$D_R v(t) \leq \Gamma(t)v(t), \quad \text{where} \quad \Gamma \in L^1[\tau, \xi].$$

Then

$$v(t) \leq 0, \quad t \in [\tau, \xi].$$

Let me explain how these two lemmas imply that functions H and $\tilde{H}$ coincide on $[t_0, t_0+c]$. Consider $[\tau, \xi] = [t_0, t_0 + c]$. In the first lemma, take $z(t) = H(t) - \tilde{H}(t)$ and use inequality (2.3.4) to obtain

$$D_R \|H(t) - \tilde{H}(t)\|_1 \leq \Gamma_0(t) \|H(t) - \tilde{H}(t)\|_1.$$

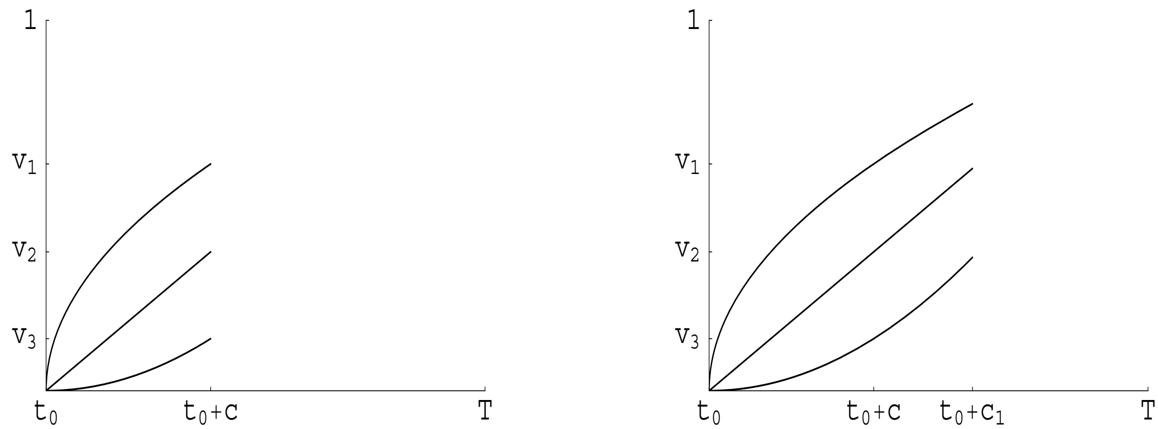
In the second lemma, let $v(t) = \|H(t) - \tilde{H}(t)\|_1$ and $\Gamma(t) = \Gamma_0(t)$. Because condition (III) holds, then according to this lemma, $\|H(t) - \tilde{H}(t)\|_1 \leq 0$, $t \in [t_0, t_0 + c]$. This means that $\|H(t) - \tilde{H}(t)\|_1 = 0$, $t \in [t_0, t_0 + c]$, or, in other words, functions H and $\tilde{H}$ coincide on $[t_0, t_0 + c]$. This implies that functions F and $\tilde{F}$ coincide on $[t_0, t_0 + c]$ too.

To summarize, I have shown that, given conditions (I), (II) and (III) on observable functions G_i, problem (DE)-(IC) has the unique solution F in a neighborhood of t_0. As mentioned in section 2.2.2, this solution is assumed to be monotone.

2.3.3 Global identification

In this section, I establish that (DE)-(IC) has a unique solution on the whole support $[t_0, T]$. To obtain this result, I show that the local solution to (DE)-(IC) can be extended to a solution on the entire interval $[t_0, T]$, and that such extension is unique.

To gain intuition, consider Figure 2.3. The picture on the left shows the local solution F found on some interval $[t_0, t_0 + c]$. The idea of constructing a global solution is to extend this solution F to the right at least to a small interval $(t_0 + c, t_0 + c_1]$, $c_1 > c$, in such a



F_1, F_2, F_3 on $[t_0, t_0 + c]$ F_1, F_2, F_3 on $[t_0, t_0 + c_1]$

Figure 2.3. Solution to the main problem on $[t_0, t_0 + c]$ (left) and extended solution to the main problem on $[t_0, t_0 + c_1]$ (right).

way that the extended solution solves (DE)-(IC) on $[t_0, t_0 + c_1]$. The picture on the right in Figure 2.3 shows this extended solution. Then this solution is extended even farther to the right and so on. This process continues until the upper support point T is reached.

Consider Figure 2.3 and the local solution F on $[t_0, t_0 + c]$ depicted on the left in this figure. Notice that all functions F_i take positive values at $t_0 + c$ and these values are known. Denote them as $v_i = F_i(t_0 + c)$, $v_i > 0$. To extend the local solution to the right, I need to solve system (DE) in a right-hand side neighborhood of $t_0 + c$ given that functions F_i in a solution to this system take values v_i at $t_0 + c$. Clearly, results in section 2.3.2 cannot be used for this problem because the methods in that section were developed for the situation when all initial values of F_i are 0. Therefore, to carry out the extension process I first need to prove the local existence and uniqueness result for the case when all initial values of F_i

are positive.

Positive initial values

Let $t_1 \in (t_0, T)$ and functions F_i satisfy initial conditions

$$F_i(t_1) = v_i, \quad i = 1, 2, 3, \tag{2.3.5}$$

where v_i are known, $0 < v_i < 1$. Notice that the values of $G_i(t_1)$ are known. Similar to section 2.3.2, I first consider the auxiliary system (DE_H). The initial conditions on functions H_i are obviously

$$H_1(t_1) = v_2 v_3, \quad H_2(t_1) = v_1 v_3, \quad H_3(t_1) = v_1 v_2. \tag{2.3.6}$$

The proposition below establishes the local existence result for the auxiliary problem.

Proposition 2.11. *Let observable functions G_i satisfy conditions (I). Then (DE_H)- (2.3.6) has a solution in a right-hand neighborhood of t_1.*

The existence result of Proposition 2.5 also required G_i to satisfy conditions (II). Note that because the values of the underlying distribution functions F_i at t_1 are separated from 0, then the result of Proposition 2.11 does not require any conditions on the behavior of G_i around t_1.

The next proposition establishes the local existence and uniqueness result for problem (DE)- (2.3.5). It is noteworthy that the uniqueness result holds without any additional conditions on observable functions G_i.

Theorem 2.12. *Let observable functions G_i satisfy conditions (I). Then (DE)- (2.3.5) has only one solution in a right-hand neighborhood of t_1.*

The proof of the uniqueness part of this theorem is much easier than the proof for problem (DE)-(IC). Indeed, for (DE)-(IC), the difficulty of proving uniqueness for stemmed from the fact that all F_i had values 0 at t_0. Now all $F_i(t_1)$ are positive and, as shown in the Appendix, for some constant C, a generalized local Lipschitz condition will be as follows:

$$\|H'(t) - \tilde{H}'(t)\|_1 \leq C(g_1(t) + g_2(t) + g_3(t))\|H(t) - \tilde{H}(t)\|_1$$

a.e. in a right-hand neighborhood of t_1. Because $g_i \in L^1$, lemmas 3.12 and 3.13 imply the uniqueness of a local solution to (DE)- (2.3.5).

Extension of the local solution to the whole support

Now I turn to the final element of the identification proof. I demonstrate how the unique local solution to (DE)-(IC) can be uniquely extended to a solution on the whole support. Throughout this section, I assume that functions F_i obtained from H_i are strictly monotone – that is, the ratios $\frac{H_2 H_3}{H_1}$, $\frac{H_1 H_3}{H_2}$, $\frac{H_1 H_2}{H_3}$ are strictly increasing.

To begin, recall that in the proof of the existence result in section 2.3.2, function $J(t, H)$ was defined on $\bar{D}_0(\delta)$ and the values of function H were restricted to set $\bar{H}(\delta)$ for a chosen $0 < \delta < 1$:

$$\bar{H}_0(\delta) = (0, \infty)^3 \cap \{(h_1, h_2, h_3)^{tr} : h_2 h_3 \leq \delta h_1, h_1 h_3 \leq \delta h_2, h_2 h_3 \leq \delta h_1\}.$$

Because the local solution to the auxiliary problem takes values only in this set, the functions F_i in the corresponding local solution to the main problem (DE_H)-(IC_H) take values in

$[0, \sqrt{\delta}]$ only. However, we also want to identify F_i when these functions take values above $\sqrt{\delta}$. Notice that $\delta < 1$ could be chosen arbitrarily close to 1, allowing us to extend the local solution to the whole support.

Fix δ, $0 < \delta < 1$, and let the domain of $J(t, H)$ be $\bar{D}_0(\delta) = [t_0, T] \times \bar{H}_0(\delta)$ (a.e. with respect to t). Theorem 2.7 proved that given conditions (I), (II) and (III), system (DE_H) with initial conditions (IC_H) has the unique solution $H = (H_1, H_2, H_3)$ on some interval $[t_0, t_0 + c]$. Denote $t_1 = t_0 + c$, and calculate

$$x_{i1} = H_i(t_1), \quad i = 1, 2, 3.$$

Because H_i are strictly increasing functions, then $x_{i1} > 0$. Note that $H(t_1) \in \bar{H}_0(\delta)$. If $H(t_1)$ is an interior point in $\bar{H}_0(\delta)$ − that is, if

$$\frac{x_{11}x_{21}}{x_{31}} < \delta, \quad \frac{x_{11}x_{31}}{x_{21}} < \delta, \quad \frac{x_{21}x_{31}}{x_{11}} < \delta \; -$$

then $(t_1, H(t_1))$ is an interior point of $\bar{D}_0(\delta)$, therefore $J(t, H)$ is defined in a neighborhood of this point. This means that the auxiliary system (DE_H), considered for $t \geq t_1$, with initial conditions

$$H_i(t_1) = x_{i1}, \quad i = 1, 2, 3,$$

is a well-defined problem. In light of results in section 2.3.3, this problem has a unique solution H on some interval $[t_1, t_1 + \mu]$, $\mu > 0$. Thus, I can uniquely extend the local solution found on $[t_0, t_1]$ to a solution on the interval $[t_0, t_1 + \mu]$. Note that the value of $H(t_1 + \mu)$ belongs to $\bar{H}_0(\delta)$. If this value is in the interior of set $\bar{H}_0(\delta)$, I can extend the solution even farther to the right and continue this process until I reach a point in which the value of function H becomes located on the border of set $\bar{H}_0(\delta)$. This point determines the solution's right maximal interval of existence for the given value of δ.

Definition 2.3. *An interval $[t_0, T_1]$ is the maximal interval of existence of solution H to (DE_H)-(IC_H) if there does not exist an extension of H over an interval $[t_0, T_1 + \eta]$ such that $\eta > 0$ and H remains a solution to (DE_H)-(IC_H).*

In the case that I am currently considering, the solution's maximal interval of existence is determined by the value of δ that was chosen to define set $\bar{H}_0(\delta)$. The proposition below yields an explicit formula for this interval.

Proposition 2.13. *Let function $J(t, H)$ be defined on $\bar{D}_0(\delta)$. Assume that all conditions on G_i that guarantee existence and uniqueness of a local solution to (DE_H)-(IC_H) are satisfied. The maximal interval of existence of solution H to (DE_H)-(IC_H) is $[t_0, T_1]$, where T_1 is such that*

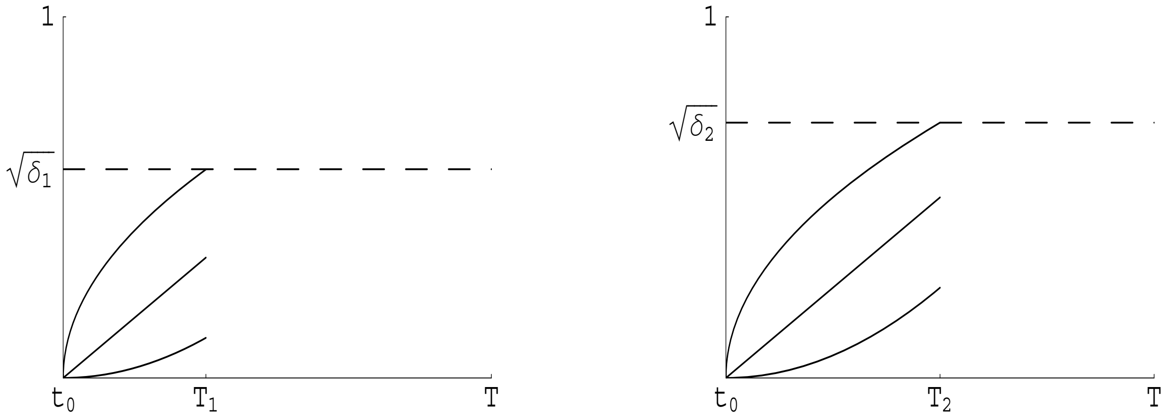
$$\max\left\{\frac{H_2(T_1)H_3(T_1)}{H_1(T_1)},\ \frac{H_1(T_1)H_3(T_1)}{H_2(T_1)},\ \frac{H_1(T_1)H_2(T_1)}{H_3(T_1)}\right\} = \delta.$$

This proposition follows from the discussion above and therefore it is left without a proof.

Proposition 2.13 implies that for the given δ, $[t_0, T_1]$ is the maximal interval of existence of a corresponding solution F to problem (DE)-(IC). Also, the values of functions F_i on $[t_0, T_1]$ belong to $[0, \sqrt{\delta}]$, and for point T_1,

$$\max\left\{F_1(T_1),\ F_2(T_1),\ F_3(T_1)\right\} = \sqrt{\delta}.$$

Figure 2.4 depicts maximal intervals of existence of a solution F for values δ_1 and δ_2, where $\delta_2 > \delta_1$. Maximal interval $[t_0, T_1]$ corresponds to δ_1, and maximal interval $[t_0, T_2]$



F_1, F_2, F_3 on $[t_0, T_1]$ $\qquad$ F_1, F_2, F_3 on $[t_0, T_2]$

Figure 2.4. Maximal intervals of existence of a solution to the main problem: $[t_0, T_1]$ corresponds to δ_1 (left), $[t_0, T_2]$ corresponds to δ_2, where $\delta_2 > \delta_1$ (right).

corresponds to δ_2. Because functions F_i are strictly increasing, then $T_2 > T_1$. Intuitively, if δ approaches 1, then the maximal interval of existence approaches support $[t_0, T]$. The proposition below establishes this fact.

Theorem 2.14. *Consider a strictly increasing sequence δ_n, $n \geq 1$, such that $\delta_n < 1$, and $\delta_n \to 1$ as $n \to \infty$. Let $J(t, H)$ be defined on $\bar{D}_0(\delta_n)$. Assume that all conditions on G_i that guarantee the existence and uniqueness of a local solution to problem (DE_H)-(IC_H) are satisfied. Let $[t_0, T_n]$ be the maximal interval of existence for the solution to (DE_H)-(IC_H). Then T_n is determined from the equation*

$$\max\left\{\frac{H_2(T_n)H_3(T_n)}{H_1(T_n)},\ \frac{H_1(T_n)H_3(T_n)}{H_2(T_n)},\ \frac{H_1(T_n)H_2(T_n)}{H_3(T_n)}\right\} = \delta_n, \tag{2.3.7}$$

and T_n is a strictly increasing sequence. If

$$F_i(T) = 1, \quad i = 1, 2, 3, \tag{2.3.8}$$

then $T_n \to T$ *as* $n \to \infty$.

Taking into account that $F_1^2 = \frac{H_2 H_3}{H_1}$, $F_2^2 = \frac{H_1 H_3}{H_2}$, $F_3^2 = \frac{H_1 H_2}{H_3}$, we can see that Proposition 2.14 guarantees that by choosing δ arbitrarily close to 1, we will identify F_i on the whole support $[t_0, T]$. This completes the proof of identification.

To summarize, I have proved that if observable functions G_i satisfy conditions (I), (II) and (III), then problem (DE)-(IC) has the unique solution F on $[t_0, T]$. In other words, distribution functions of bidders' private values are identified.

2.4 Misspecification

In this section, I illustrate that it is possible to indicate a system of well-defined functions G_i whose corresponding F_i, which constitute the unique solution to (DE)-(IC), are not all monotone. Under these circumstances, we can conclude that the auction model is misspecified.

This non-monotonicity result carries important implications for generalized competing risks models and contrasts sharply with results for classical competing risks models. In classical models, every dependent risks model has a unique, observationally equivalent independent risks model, as shown in Tsiatis (1975). Mathematically, a system of differential equations that frames a classical model with independent risks can be transformed into a system of linear differential equations in which every equation contains only one unknown function. Thus, it is straightforward to find a closed-form solution to this system and to show that this solution is unique: Functions F_i in this solution are always strictly increasing.

Note that classical competing risks models are isomorphic to second-price auctions when only the winner's identity and bid are observed. We therefore can better understanding Tsiatis' result by considering second-price auctions with these observable data. In this case, the data provide direct knowledge of the winner's value. Intuitively, it should be easy to identify the distributions of bidders' private values when these values are independent.

Because the data about the winner's identity and bid are available, the following functions are known:

$$R_i(t) = P(b_i \leq t,\ i \text{ wins}) = P(b_i \leq t,\ \max_{j \neq i} b_j < b_i), \quad i = 1, \ldots, d.$$

Given the private values' independence, the distribution functions F_i satisfy this system of differential equations:

$$R_i' = (F_1 \ldots F_{i-1} F_{i+1} \ldots F_d) F_i^{'}, \quad i = 1, \ldots, d. \tag{2.4.1}$$

Add all the equations in this system and obtain

$$R_1 + \ldots + R_d = F_1 \ldots F_d.$$

Use the last equation to rewrite system (2.4.1) equivalently as a system of linear differential equations:

$$F_i' = \frac{R_i'}{R_1 + \ldots + R_d} F_i, \quad i = 1, \ldots, d.$$

System (2.4.1) considered for $t \leq T$ and with the boundary conditions $F_i(T) = 1$ possesses a unique solution on $[t_0, T]$, which has a closed form:

$$F_i(t) = \exp^{-\int_t^T \frac{R_i'}{R_1 + \ldots + R_d} ds}, \quad i = 1, \ldots, d.$$

Given that all R_i are strictly increasing, functions F_i in this solution are strictly increasing too. Thus, F_i have the properties of distribution functions and therefore can be the distribution functions of bidders' values. To summarize, because the highest bid is observable, researchers can easily determine the value distribution for each bidder.

Now consider the original situation in which only the winner's identity and the transaction price are observed. In this case, the auction data provide direct knowledge regarding groups of $d-1$ bidders. Nevertheless, to learn each bidder' values separately, this information has to be disentangled further. The complexity of auction data can allow us to detect misspecification. Indeed, as I demonstrate below, it is possible to obtain a solution in which some functions F_i are not monotone, an indicator of misspecification.

Functions H_i in a solution to the auxiliary problem are always strictly increasing because observable G_i are strictly increasing. This property does not imply that the corresponding functions F_i in a solution to the main problem are monotone, however. In the following example, I consider monotone functions F_2 and F_3 and a non-monotone F_1. I show that the auxiliary functions H_i are strictly increasing and the corresponding G_i satisfy conditions (I) and (II). I also give an example of the joint, absolutely continuous distribution of bidders' values that rationalizes G_i.

Example 2.4.1. *This example is illustrated in Figure 2.5.*

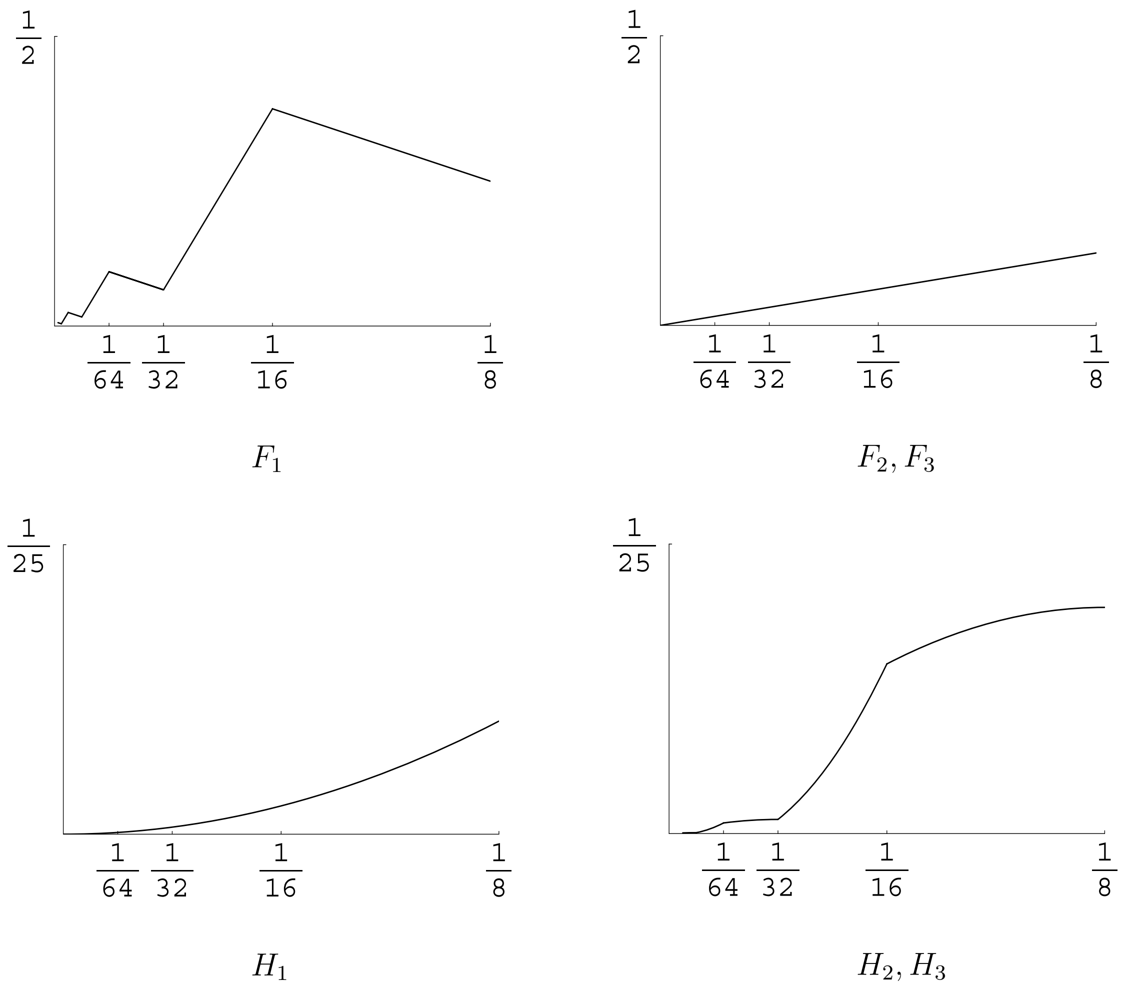


Figure 2.5. F_1, F_2, F_3 and H_1, H_2, H_3 on $[0, \frac{1}{8}]$ (Example 2.4.1).

On $[0,1]$ *consider* $F_2(t) = t$, $F_3(t) = t$, *and define function* F_1 *in the following way:*

$$F_1(t) = -2t + \frac{1}{2^{2m-1}}, \quad t \in \left[\frac{1}{2^{2m+2}}, \frac{1}{2^{2m+1}}\right], \quad m \geq 1,$$

$$F_1(t) = 10t - \frac{1}{2^{2m-2}}, \quad t \in \left[\frac{1}{2^{2m+1}}, \frac{1}{2^{2m}}\right], \quad m \geq 2,$$

$$F_1(t) = \frac{6}{7}t + \frac{1}{7}, \quad t \in \left[\frac{1}{8}, 1\right].$$

Function F_1 *is Lipschitz and therefore absolutely continuous. It is strictly decreasing on intervals* $\left[\frac{1}{2^{2m+2}}, \frac{1}{2^{2m+1}}\right]$, $m \geq 1$, *and strictly increasing on other intervals. In particular,* F_1 *is not increasing in any small neighborhood of* t_0.

I now demonstrate that functions H_i *are strictly increasing. Clearly,* $H_1 = F_2F_3 = t^2$ *is strictly increasing on* $[0,1]$. *Consider* H_2 *on an interval* $\left[\frac{1}{2^{2m+2}}, \frac{1}{2^{2m+1}}\right]$, $m \geq 1$:

$$H_2(t) = t(-2t + \frac{1}{2^{2m-1}}).$$

Function $t\left(-2t + \frac{1}{2^{2m-1}}\right)$ *is quadratic; it strictly increases until point* $\frac{1}{2^{2m+1}}$, *then it strictly decreases. So,* H_2 *is strictly increasing on* $\left[\frac{1}{2^{2m+2}}, \frac{1}{2^{2m+1}}\right]$.

Now consider H_2 *on an interval* $\left[\frac{1}{2^{2m+1}}, \frac{1}{2^{2m}}\right]$, $m \geq 2$:

$$H_2(t) = t\left(10t - \frac{1}{2^{2m-2}}\right).$$

Quadratic function $t\left(10t - \frac{1}{2^{2m-2}}\right)$ *strictly decreases until point* $\frac{1}{5\cdot 2^{2m}}$ *and strictly increases afterward. Because* $\frac{1}{2^{2m+1}} > \frac{1}{5\cdot 2^{2m}}$, *then* H_2 *strictly increases on* $\left[\frac{1}{2^{2m+1}}, \frac{1}{2^{2m}}\right]$. *Obviously, on* $\left[\frac{1}{8}, 1\right]$ *function*

$$H_2(t) = t\left(\frac{6}{7}t + \frac{1}{7}\right)$$

strictly increases. Thus, H_2 *and, consequently,* H_3 *are strictly increasing on* $[0,1]$.

Now find functions G_i according to integral-differential equations (2.2.3). They clearly satisfy conditions (I), and it can be shown that they also satisfy conditions (II). Hence, we can find well-defined, observable G_i whose corresponding F_i are not all monotone.

Finally, I present an example of a joint absolutely continuous distribution that rationalizes G_i. The properties of functions G_i guarantee that we can find functions Φ_i, Ψ_i, Ξ_i, $i = 1, 2, 3$, that satisfy the following conditions:

1. *They are defined on $[0, 1]$, non-negative, increasing and absolutely continuous on $[0, 1]$.*
2. *$\Phi_i(0) = 0$, $\Psi_i(0) = 0$ and $\Xi_i(0) = 0$, $i = 1, 2, 3$.*
3.

$$
\begin{aligned}
G_1(t) &= \int_{t_0}^{t} (\Phi_2\Phi_3)'(1 - \Phi_1)ds \\
G_2(t) &= \int_{t_0}^{t} (\Psi_1\Psi_3)'(1 - \Psi_2)ds \\
G_3(t) &= \int_{t_0}^{t} (\Xi_1\Xi_2)'(1 - \Xi_3)ds.
\end{aligned}
\tag{2.4.2}
$$

Denote $\phi_i = \Phi'$, $\psi_i = \Psi'$ and $\xi_i = \Xi'$. Consider the following function defined on $[0, 1]^3$:

$$
\begin{aligned}
f(x_1, x_2, x_3) = \phi_1(x_1)\phi_2(x_2)\phi_3(x_3)1(x_1 \geq x_2)1(x_1 \geq x_3) + \\
+ \psi_1(x_1)\psi_2(x_2)\psi_3(x_3)1(x_2 > x_1)1(x_2 \geq x_3) + \\
+ \xi_1(x_1)\xi_2(x_2)\xi_3(x_3)1(x_3 > x_1)1(x_3 > x_2).
\end{aligned}
\tag{2.4.3}
$$

Function f is an example of a joint density of the distribution of bidders' values that rationalizes observable G_i. Note that because we can choose functions Φ_i, Ψ_i, Ξ_i in many different ways, there are many observationally equivalent densities of this type.

To summarize, I have shown that in second-price auctions and, hence, in generalized competing risks models, we can detect misspecification in some situations, including those in which not all functions in the solution are monotone. It is important to mention, however, that non-monotonicity is a sufficient but not a necessary condition for misspecification. In other words, for misspecified models, it is possible to obtain a solution in which all functions are distribution functions.

In auctions, misspecification can occur for a variety of reasons. If bidders' private values are not independent, for instance, or if bidders do not behave rationally, misspecification results. In generalized competing risks models in reliability theory, however, the non-monotonicity of at least one function F_i means that the risks are dependent. Also, the non-monotonicity finding of this chapter runs contrary to the results of classical competing risks models (Roy model), in which dependent and independent risks cannot be distinguished nonparametrically.

2.5 Generalizations

2.5.1 Auctions with any number of bidders

In this section, I show how the identification result for auctions with three bidders can be generalized to auctions with any number of bidders. I state main results but do not discuss them because their interpretations and intuitiveness are similar to those in the case of three bidders. A more detailed discussion is given in the appendix.

Mathematical model

I first present the system of integral-differential equations that describes relationships between observable functions G_i and unknown distribution functions F_i. By definition,

$$G_i(t) = Pr(\text{price} \leq t, i \text{ wins}) = Pr(\max_{j \neq i} b_j \leq t, \max_{j \neq i} b_j < b_1), \quad i = 1, \dots, d.$$

The assumption of the independence of private values yields

$$G_i(t) = \int_{t_0}^{t} (F_1 \dots F_{i-1} F_{i+1} \dots F_d)' (1 - F_i) ds, \quad i = 1, \dots, d.$$

The differentiation of both sides of these equations yields a system of differential equations

$$g_i = (F_1 \dots F_{i-1} F_{i+1} \dots F_d)' (1 - F_i), \quad i = 1, \dots, d. \tag{2.5.1}$$

Functions F_i in this system must satisfy initial conditions

$$F_i(t_0) = 0, \quad i = 1, \dots, d. \tag{2.5.2}$$

Identification result

The theorem below states the local existence result.

Theorem 2.15. *Let observable functions G_i satisfy conditions*

1. $G_i(t_0) = 0$, $i = 1, \dots, d$

2. G_i is absolutely continuous on $[t_0, T]$, $i = 1, \dots, d$

3. G_i is strictly increasing on $[t_0, T]$, $i = 1, \dots, d$

Then (2.5.1)-(2.5.2) has a solution in a neighborhood of t_0.

The next theorem states the local identification result,

Theorem 2.16. *Suppose that all conditions on G_i in Theorem 2.15 are satisfied. If, in addition,*

$$\sum_{i=1}^{d}\left(\frac{G_1G_2\ldots G_{i-1}G_{i+1}\ldots G_d}{G_i^{d-1}}\right)^{\frac{1}{d-2}}\cdot\sum_{i=1}^{d}\frac{g_i}{G_i}\in L^1 \tag{2.5.3}$$

in a neighborhood of t_0, then (2.5.1)-(2.5.2) has a unique solution in a neighborhood of t_0.

Condition (2.5.3) is essential for proving identification.

Once the local identification result is established, techniques from section 2.3.3 can be employed to extend the local solution to the whole support $[t_0, T]$ in a unique way, and therefore show global identification.

2.5.2 Identification with weaker support conditions

So far, I have assumed that bidders' values are distributed on the same support $[t_0, T]$. In practice, however, these distributions may have different supports as well as supports that are not a connected interval. It is of interest therefore to explore whether identification holds in these circumstances. In this section, I examine two ways to relax the support condition. One is to permit distributions to have different upper support points. The other is to allow for holes in the supports. I start by considering a case of three bidders and analyzing identification when their distributions' upper support points differ from each other. I then generalize this analysis for auctions with any number of bidders. Finally, I briefly discuss what happens when distributions have holes in their supports.

Auctions with three bidders

Let τ_1, τ_2, τ_3 be the upper support points of the distributions of bidders' valuations. Without a loss of generality, assume that $\tau_1 \leq \tau_2 \leq \tau_3$. Given these inequalities, there are four possibilities for the locations of τ_1, τ_2, τ_3 with respect to each other; they are illustrated in figures 2.6 and 2.7.

Case 1: $\tau_1 = \tau_2 = \tau_3$

This is the case analyzed in this paper. Clearly, functions G_i are defined on $[t_0, \tau_1]$ and satisfy the boundary condition

$$G_1(\tau_1) + G_2(\tau_1) + G_3(\tau_1) = 1$$

because $G_1(\tau_1) + G_2(\tau_1) + G_3(\tau_1) = Pr(\text{price} \leq \tau_1) = 1$. In section 2.3, it was established that given G_1, G_2 and G_3, distribution functions F_1, F_2 and F_3 are identified on $[t_0, \tau_1]$. This case is illustrated in figure 2.6.

Case 2: $\tau_1 < \tau_2 = \tau_3$

Because the first player never submits bids higher than τ_1, function G_1 is defined on $[t_0, \tau_1]$. The second and the third players have positive probability of submitting bids in $[\tau_1, \tau_2]$, so G_2 and G_3 are defined on $[t_0, \tau_2]$. The definitions of G_1, G_2 and G_3 imply the boundary condition

$$G_1(\tau_1) + G_2(\tau_2) + G_3(\tau_2) = 1.$$

Identification is obtained in the following way. On $[t_0, \tau_1]$, functions F_i must solve (DE)-(IC). First, I find the unique solution to (DE)-(IC) in a small neighborhood of t_0. I then use methods from section 2.3.3 to extend it farther to the right until one of the functions F_1, F_2, F_3 reaches value 1. Because by assumption $\tau_1 < \tau_2 = \tau_3$, function F_1 will be the first one to reach value 1, and that will happen at τ_1. Thus, all F_i can be identified on $[t_0, \tau_1]$, and I can find values $x_2 = F_2(\tau_1) > 0$ and $x_3 = F_3(\tau_1) > 0$. The next step is to identify functions F_2 and F_3 on $(\tau_1, \tau_2]$. On this interval, functions F_2 and F_3 must solve the system

$$g_2 = F_3'(1 - F_2) \tag{2.5.4}$$

$$g_3 = F_2'(1 - F_3)$$

and satisfy initial conditions

$$F_2(\tau_1) = x_2, \quad F_3(\tau_1) = x_3.$$

Applying techniques from section 2.3.3, I can show that this problem has the unique solution in a right-hand neighborhood of τ_1. Employing extension methods from section 2.3.3, I can demonstrate that this local solution can be extended to the interval $[\tau_1, \tau_2]$ and that such extension is unique.[9] This case is illustrated in figure 2.7.

Case 3: $\tau_1 = \tau_2 < \tau_3$

In this case, there are no observable data on $(\tau_1, \tau_3]$. So, functions G_i are defined on $[t_0, \tau_1]$ and satisfy the boundary condition

$$G_1(\tau_1) + G_2(\tau_1) + G_3(\tau_1) = 1.$$

Given G_1, G_2 and G_3, I find the unique solution F to (DE)-(IC) on $[t_0, \tau_1]$. Hence, F_1 and F_2 are identified. As for F_3, nothing can be learned about this function for $t > \tau_1$

[9] The methodology presented here can be used for auctions with any number of bidders. In this particular case, however, system (2.5.4) can be handled much more easily once it is rewritten as a system of two linear equations: $F_2' = \frac{g_3(1-F_2)}{1-G_2-G_3}$ and $F_3' = \frac{g_2(1-F_3)}{1-G_2-G_3}$. This system has a closed-form solution: $F_2(t) = 1 - (1 - x_2)\exp(-\int_{\tau_1}^{t} \frac{g_3(s)}{1-G_2(s)-G_3(s)} ds)$, $F_3(t) = 1 - (1 - x_3)\exp(-\int_{\tau_1}^{t} \frac{g_2(s)}{1-G_2(s)-G_3(s)} ds)$.

because there are no observations corresponding to those t, so this function is only partially identified. This case is illustrated in figure 2.6.

Case 4: $\tau_1 < \tau_2 < \tau_3$

This situation is similar to case 3. Function G_1 is defined on $[t_0, \tau_1]$, G_2 is defined on $[t_0, \tau_2]$, and because a transaction price never exceeds τ_2, G_3 is defined only on $[t_0, \tau_2]$. Functions G_1, G_2 and G_3 satisfy the boundary condition

$$G_1(\tau_1) + G_2(\tau_2) + G_3(\tau_2) = 1.$$

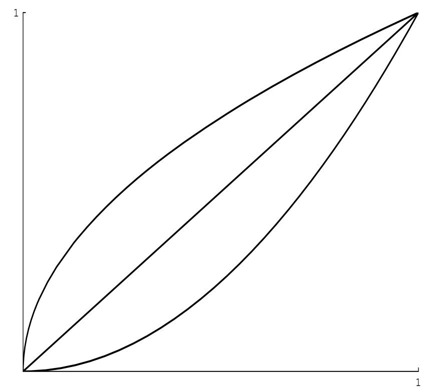
To identify F_1 and F_2 as well as partially identify F_3, I proceed as follows. On $[t_0, \tau_1]$, functions F_1, F_2 and F_3 must solve problem (DE)-(IC). On $[\tau_1, \tau_2]$, functions F_2 and F_3 must solve system (2.5.4) and satisfy certain initial conditions at τ_1. Therefore, F_1 and F_2 are identified. Because there are no observed data for $t > \tau_2$, function F_3 cannot be learned for $t > \tau_2$; that is, F_3 is only partially identified. This case is illustrated in figure 2.7.

Thus, in auctions with three bidders, two distribution functions F_1 and F_2 are always identified. In two cases, the third distribution function F_3 is identified too. In two other cases, it is identified only on a subinterval of the corresponding support.

Auctions with any number of bidders

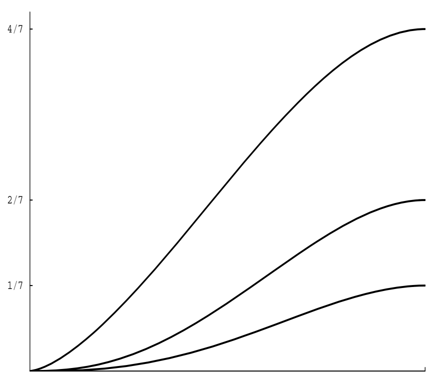
Now I consider auctions with any number of bidders and briefly discuss identification in this case. Suppose I found the unique local solution to (DE)-(IC). I extend it farther and farther to the right until I reach a point where one of the functions F_i hits the value of 1. Beyond this point, system (DE) does not contain the equation corresponding to bidder i. Thus, $d-1$ equations (or possibly even fewer equations, if several functions F_i hit value 1 at the

Case 1: $\tau_1 = \tau_2 = \tau_3$

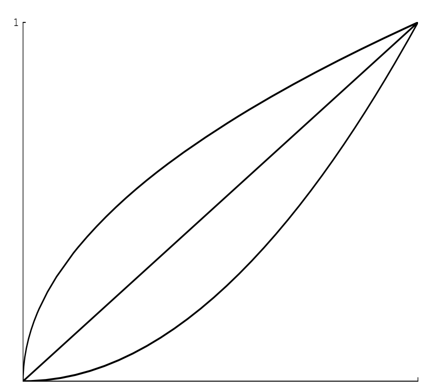


Underlying distribution functions

$$F_1 = \sqrt{t}, \quad F_2 = t^2, \quad F_3 = t$$

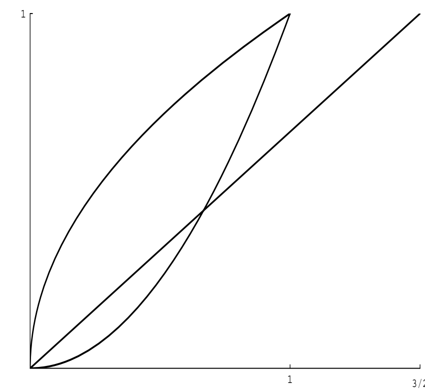


Observed functions G_1, G_2, G_3



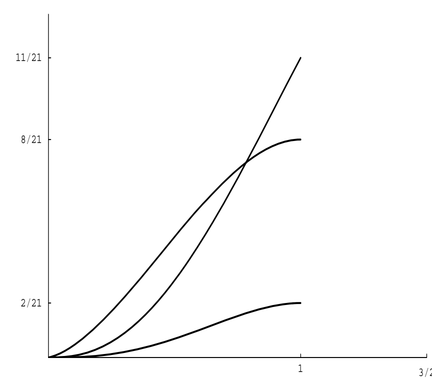
Identified functions

Case 3: $\tau_1 = \tau_2 < \tau_3$

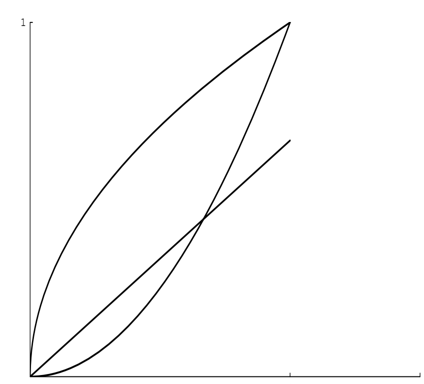


Underlying distribution functions

$$F_1 = \sqrt{t}, \quad F_2 = t^2, \quad F_3 = \frac{2}{3}t$$



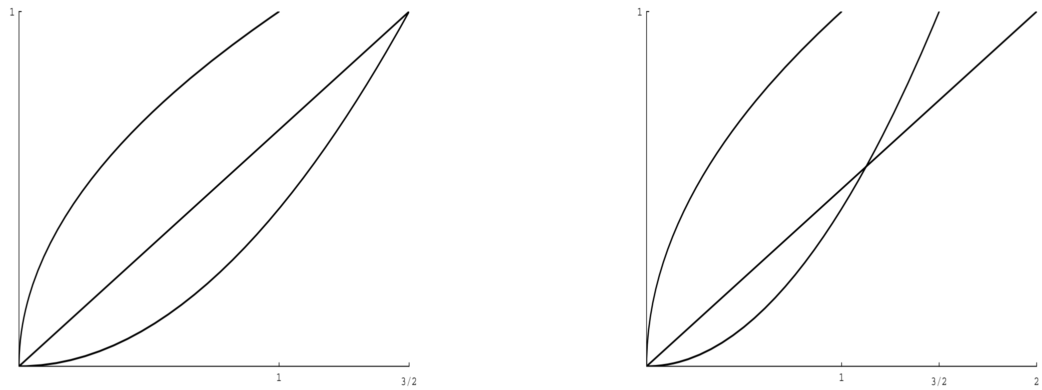
Observed functions G_1, G_2, G_3



Identified functions

Figure 2.6. Supports for cases 1 and 3.

Case 2: $\tau_1 < \tau_2 = \tau_3$ Case 4: $\tau_1 < \tau_2 < \tau_3$

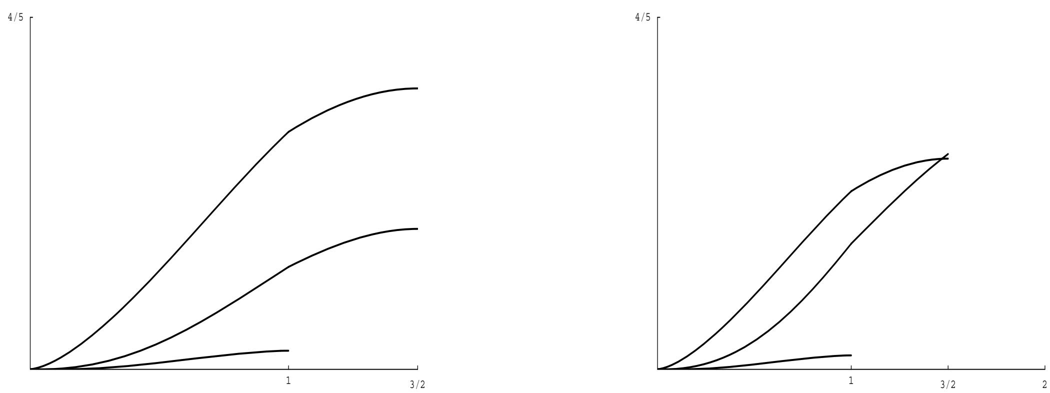


Underlying distribution functions

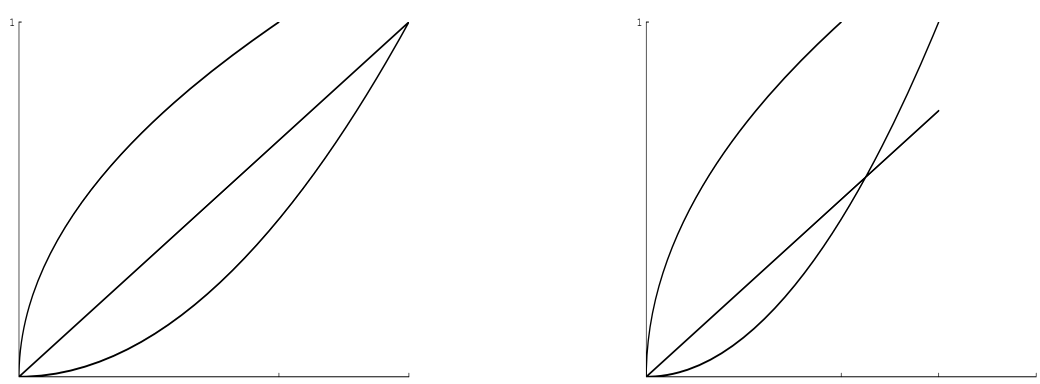
$F_1 = \sqrt{t}, \quad F_2 = \frac{4}{9}t^2, \quad F_3 = \frac{2}{3}t$

Underlying distribution functions

$F_1 = \sqrt{t}, \quad F_2 = \frac{4}{9}t^2, \quad F_3 = \frac{1}{2}t$



Observed functions G_1, G_2, G_3 Observed functions G_1, G_2, G_3



Identified functions Identified functions

Figure 2.7. Supports for cases 2 and 4

same point) remain in system (DE). Next, I find the unique local solution to the reduced system and extend this solution to the right until I reach a point where one of the remaining functions F_i hits value 1. Beyond this point, the number of equations in the system decreases again. Proceeding in this way, I eventually come to a final system such that all functions F_i in its solution, with the possible exception of one function, reach value 1 at the same point.

The result of the next proposition is intuitive from case of the three bidders. I do not prove this result because it follows from the local existence and local uniqueness results, as well as the extension techniques.

Proposition 2.17. *Suppose that d bidders are participating in the auction, and the distributions of their values are locally identified. Let $[t_0, \tau_i]$ stand for the support of bidder i, $i = 1, \ldots, d$. Without a loss of generality, assume that*

$$\tau_1 \leq \tau_2 \leq \ldots \leq \tau_{d-1} \leq \tau_d.$$

If $\tau_{d-1} = \tau_d$, then all distribution functions $F_1, \ldots, F_d$ are identified. If $\tau_{d-1} < \tau_d$, then $d-1$ functions $F_1, \ldots, F_{d-1}$ are identified, and F_d is identified only on $[t_0, \tau_{d-1}]$.

Holes in the support

Finally, I want to informally discuss identification in a situation where the distributions of bidders' values can have holes in the supports. In this case, distribution functions F_i can have intervals on which they are constant. If all distributions have the same lower support point t_0, then $F_i(t) > 0$ for $t > t_0$ and therefore local identification can be shown by using techniques from section 2.3.2. Extension to a global solution depends on distributions' upper support points, as explained in Proposition 2.17.

2.6 Sieve estimation of distribution functions

This section presents an approach to estimating the distribution functions of private values from a random sample. First, I define an operator A that maps unknown distribution functions F_1, F_2, F_3 to observable functions G_1, G_2, G_3. I show that this operator is Lipschitz and that under weak conditions on the set of $F = (F_1, F_2, F_3)$, its inverse operator A^{-1} is continuous. I then derive sieve estimators of F_i and use the properties of A to show their consistency.

2.6.1 Operator A

For an absolutely continuous function $F = (F_1, F_2, F_3)^{tr}$ define

$$A(F) = (A(F)_1, A(F)_2, A(F)_3)^{tr} \text{ as}$$

$$A(F)_1(t) = \int_{t_0}^{t} (F_2F_3)'(1 - F_1)ds$$
$$A(F)_2(t) = \int_{t_0}^{t} (F_1F_3)'(1 - F_2)ds$$
$$A(F)_3(t) = \int_{t_0}^{t} (F_1F_2)'(1 - F_3)ds.$$

Let Λ be the set of functions $F = (F_1, F_2, F_3)^{tr}$ defined on $[t_0, T]$ and satisfying the following conditions:

Conditions C1.

1. *F_i are absolutely continuous on $[t_0, T]$.*
2. *F_i are strictly increasing on $[t_0, T]$.*

3. $F_i(t_0) = 0,\ F_i(T) = 1,\ i = 1, 2, 3.$

4. *Function $G = A(F)$ satisfies the condition*

$$\left(\frac{g_1}{G_1} + \frac{g_2}{G_2} + \frac{g_3}{G_3}\right)\left(\sqrt{\frac{G_2G_3}{G_1}} + \sqrt{\frac{G_1G_3}{G_2}} + \sqrt{\frac{G_1G_2}{G_3}}\right) \in L^1$$

in a small neighborhood of t_0.

Let A be defined on Λ. Properties of the image $A(\Lambda)$ are described in Proposition 2.1: G_i are absolutely continuous on $[t_0, T]$, strictly increasing on $[t_0, T]$, $G_i(t_0) = 0$ and $G_1(T) + G_2(T) + G_3(T) = 1$. As shown in this chapter, there exist the inverse operator $A^{-1} : A(\Lambda) \to \Lambda$.

Endow both domain Λ and its image $A(\Lambda)$ with the following uniform metric:

$$d(F, \tilde{F}) = \sup_{t\in[0,1]} \sqrt{(F(t) - \tilde{F}(t))^{tr}(F(t) - \tilde{F}(t))}$$

$$d(G, \tilde{G}) = \sup_{t\in[0,1]} \sqrt{(G(t) - \tilde{G}(t))^{tr}(G(t) - \tilde{G}(t))}.$$

The proposition below implies that A is continuous in this metric.

Proposition 2.18. *For any $F, \tilde{F} \in \Lambda$,*

$$d(A(F), A(\tilde{F})) \leq 9\sqrt{3}d(F, \tilde{F});$$

that is, operator A is Lipschitz on Λ.

The properties of A are important for proving the consistency of the estimators of F_i. Usually, it is easier to establish consistency when the space of functions is compact. I compactify Λ by bounding the densities of F_i by the same function in L^1:

Condition C2.

$$F_i'(t) \le \phi'(t) \quad a.e. \quad [t_0, T], \quad i = 1, 2, 3,$$

where ϕ *is some absolutely continuous function on* $[t_0, T]$.

Let Λ_ϕ be a subset of Λ such that all functions F from Λ_ϕ satisfy condition *C2*. This condition guarantees that Λ_ϕ is relatively compact under the uniform metric. Indeed, for any $F \in \Lambda_\phi$ and $t, \tau \in [t_0, T]$,

$$|F_i(t) - F_i(\tau)| = \left| \int_\tau^t F_i'(s) ds \right| \le |\phi(t) - \phi(\tau)|, \quad i = 1, 2, 3.$$

Because ϕ is absolutely continuous, the last inequality implies that the set Λ_ϕ is equicontinuous. It is also uniformly bounded because the values of F_i do not exceed 1. According to the Arzela-Ascoli theorem, Λ_ϕ is relatively compact in metric $d(\cdot, \cdot)$. Note that if $F \in \Lambda_\phi$, then function $G = A(F)$ satisfies this condition too:

$$g_i(t) \le \phi'(t) \quad a.e. \quad [t_0, T], \quad i = 1, 2, 3.$$

Let $\overline{\Lambda}_\phi$ stand for the closure of Λ_ϕ under metric $d(\cdot, \cdot)$. Because Λ_ϕ is relatively compact, $\overline{\Lambda}_\phi$ is a compact set. To consider operator A on $\overline{\Lambda}_\phi$, I first have to show that A is defined for functions in this set that do not belong to Λ_ϕ. The proposition below establishes that all functions in $\overline{\Lambda}_\phi$ satisfy conditions 1, 3 and a modified condition 2 in *C1*, and also satisfy condition *C2*.

Proposition 2.19. *If* $F = (F_1, F_2, F_3)^{tr} \in \overline{\Lambda}_\phi$, *then functions* F_i *are absolutely continuous, increasing and satisfy* $F_i(t_0) = 0$. *Also,* $F_i'(t) \le \phi'(t)$ *a.e. on* $[t_0, T]$, $i = 1, 2, 3$.

Because all functions in $\overline{\Lambda}_\phi$ are absolutely continuous, operator A can be extended from Λ_ϕ to $\overline{\Lambda}_\phi$. The next proposition implies that A is continuous on $\overline{\Lambda}_\phi$.

Proposition 2.20. *For any $F, \tilde{F} \in \overline{\Lambda}_\phi$,*

$$d(A(F), A(\tilde{F})) \leq C_0 d(F, \tilde{F}), \tag{2.6.1}$$

where $C_0 = 3\sqrt{3}(1 + 3\phi(T) - 3\phi(t_0))$; that is, A is Lipschitz on $\overline{\Lambda}_\phi$.

Finally, I establish the continuity of A^{-1} on $A(\Lambda_\phi)$.

Proposition 2.21. *A^{-1} is continuous on $A(\Lambda_\phi)$.*

This proposition follows from the fact that if a continuous operator is defined on a compact set and the inverse operator is defined on the image of that set, then the inverse operator is continuous. The inverse operator A^{-1} is clearly defined $A(\Lambda_\phi)$ but I have not shown that it is defined on the set $A(\overline{\Lambda}_\phi)$. However, this does not affect the result (see explanation in the appendix).

2.6.2 Consistency

In this section, I define sieve estimators of the distribution functions F_i and prove their consistency.

Without a loss of generality, assume that the distributions have support $[0, 1]$. Denote the true distribution functions as F_i^* and the corresponding observable functions as G_i^*. That is, $G^* = A(F^*)$, where $F^* = (F_1^*, F_2^*, F_3^*)^{tr}$ and $G^* = (G_1^*, G_2^*, G_3^*)^{tr}$. Let $F^* \in \Lambda_\phi$. The next lemma introduces a function Q on $\overline{\Lambda}_\phi$ that is uniquely minimized by F^*.

Lemma 2.22. *F^* is the unique minimizer of*

$$Q(F) = E(G^* - A(F))^{tr}(G^* - A(F))$$

on $\overline{\Lambda}_\phi$.

The idea of sieve estimation is to use a sample analog of Q and approximate $\overline{\Lambda}_\phi$ with finite-dimensional spaces. For each k, choose base functions $p_{1,k}, \ldots, p_{m(k),k}$ (for example, B-splines with uniform knots or Bernstein polynomials) and introduce the set of linear combinations of these functions:

$$M_k = \{(F_1, F_2, F_3)^{tr} : F_i(t) = \sum_{l=1}^{m(k)} \alpha_l^i p_{l,k}(t),\ t \in [0,1]\}.$$

In this set of functions, consider only those functions that are in Λ_ϕ:

$$\Sigma_k = \Lambda_\phi \cap M_k.$$

Set Σ_k consists of functions from M_k with certain restrictions on coefficients α_l^i. It is relatively compact and, hence, it closure $\overline{\Sigma}_k$ is compact, and $\overline{\Sigma}_k \subset \overline{\Lambda}_\phi$.

Consider a random sample of n observations $(t_i, w_i)_{i=1}^n$, where t_i is the observed price and w_i is the winner's identity in i's auction. Without a loss of generality, assume that $t_i < t_{i+1}$, $i = 1, \ldots, n-1$. From the sample, find consistent estimators $\hat{G}_{i,n}$ of G_i, for instance, analogs of empirical distribution functions.

Let the sample objective function be

$$\hat{Q}_n(F) = \frac{1}{n}\sum_{i=1}^{n}(\hat{G}_n(t_i) - A(F)(t_i))^{tr}(\hat{G}_n(t_i) - A(F)(t_i)).$$

Also let $k = k(n)$, and define the following estimator of F^*:

$$\hat{F}_n = \text{argmin}_{F \in \overline{\Sigma}_{k(n)}} \hat{Q}_n(F).$$

The theorem below establishes the estimator consistency when sets $\overline{\Sigma}_k$ well approximate set $\overline{\Lambda}_\phi$.

Theorem 2.23. *If*

$$\forall (F \in \overline{\Lambda}_\phi) \exists (\tilde{F} \in \overline{\Sigma}_k) \quad d(F, \tilde{F}) \xrightarrow{p} 0 \text{ as } k \to \infty, \tag{2.6.2}$$

then estimator $\hat{F}_n$ *is consistent:*

$$d(\hat{F}_n, F^*) \xrightarrow{p} 0 \text{ as } n \to \infty.$$

Condition (2.6.2) holds if approximating sets are chosen properly − for instance, if base functions $p_{1,k}, \ldots, p_{m(k),k}$ are B-splines with uniform knots, Bernstein polynomials or truncated power series.

2.7 Applications

The main purpose of this section is to show how the techniques described so far can be used in more general settings. I present two applications. One relates to identification issues for second-price auctions with stochastic number of bidders. The other pertains to generalized competing risks models. I explain why identification in auctions with stochastic number of bidders cannot be proved by using techniques of generalized competing risks models, but can be established by applying the methods of this chapter.

2.7.1 Auctions with stochastic number of bidders

As before, I consider second-price auctions with independent private values, but now I allow for exogenous variation in the number of bidders. I assume that the number of potential buyers is known and does not change, but the number of actual bidders is unknown and varies exogenously. For instance, this may happen because of entrance fees or the different costs of acquiring information. In this setting, I do not aim to present a complete analysis of identification in general. Rather, I want to illustrate how the methods developed in this chapter allow us to approach the identification problem. To gain some insight while keeping the problem simple, I return to the case of three buyers.

Suppose that the number of bidders and their identities are determined by chance and the process through which bidders are selected is taken to be exogenous, embodied in probabilities p_A, $A \subset \{1, 2, 3\}$. I assume that at least two bidders are participating in the auction:

$$p_{12} + p_{13} + p_{23} + p_{123} = 1, \tag{2.7.1}$$

and each buyer has a positive probability of participation:

$$\sum_{j \neq i} p_{ij} + p_{123} > 0, \quad i = 1, 2, 3. \tag{2.7.2}$$

Suppose that buyers' private values are absolutely continuous and distributed on a common support $[t_0, T]$. If the available data tells us the winner's identity and the transaction price, then we observe functions $G_i(t) = Pr(\text{price} \leq t,\ i \text{ wins})$, $i = 1, 2, 3$. Using the law of total probability, it can be found that

$$G_1(t) = P(\text{ price} \le t, 1 \text{ wins } |\{1,2\})p_{12} + P(\text{ price} \le t, 1 \text{ wins } |\{1,3\})p_{13} +$$

$$+P(\text{ price} \le t, 1 \text{ wins } |\{1,2,3\})p_{123} = \int_{t_0}^{t} (p_{12}F_2^{'} + p_{13}F_3^{'} + p_{123}(F_2F_3)^{'})(1-F_1)ds.$$

Likewise,

$$G_2(t) = \int_{t_0}^{t} (p_{12}F_1^{'} + p_{23}F_3^{'} + p_{123}(F_1F_3)^{'})(1-F_2)ds$$

$$G_3(t) = \int_{t_0}^{t} (p_{13}F_1^{'} + p_{23}F_2^{'} + p_{123}(F_1F_2)^{'})(1-F_3)ds.$$

The differentiation of these equations yields that a.e. on $[t_0, T]$

$$g_1 = (p_{12}F_2^{'} + p_{13}F_3^{'} + p_{123}(F_2F_3)^{'})(1-F_1) \tag{2.7.3}$$

$$g_2 = (p_{12}F_1^{'} + p_{23}F_3^{'} + p_{123}(F_1F_3)^{'})(1-F_2)$$

$$g_3 = (p_{13}F_1^{'} + p_{23}F_2^{'} + p_{123}(F_1F_2)^{'})(1-F_3).$$

To prove identification, I have to show that system (2.7.3) with initial conditions

$$F(t_0) = 0, \quad i = 1,2,3, \tag{2.7.4}$$

has a unique positive solution on $[t_0, T]$. My approach is to construct an auxiliary system by introducing new functions

$$H_1 = p_{12}F_2 + p_{13}F_3 + p_{123}F_2F_3,$$

$$H_2 = p_{12}F_1 + p_{23}F_3 + p_{123}F_1F_3,$$

$$H_3 = p_{13}F_1 + p_{23}F_2 + p_{123}F_1F_2.$$

Below I demonstrate that, in general, functions F_i have a unique representation in terms of H_i. Let $F_i = q_i(H)$, $i = 1, 2, 3$. Then (2.7.3) can be written as the system of differential equations

$$H_1' = \frac{g_1}{1 - q_1(H)}$$
$$H_2' = \frac{g_2}{1 - q_2(H)}$$
$$H_3' = \frac{g_3}{1 - q_3(H)}.$$

The initial conditions on H_i are

$$\lim_{t \downarrow t_0} H_i(t) = 0, \quad i = 1, 2, 3.$$

The existence of a local solution to the auxiliary problem can be proved by applying techniques from section 2.3.2. First, I would find necessary conditions on G_i. Assuming these conditions I would use the Tonelli approximations method to prove the local existence of a solution H to the auxiliary problem. Then, I would find a solution F to (2.7.3)-(2.7.4) from H by using formulas $F_i = q_i(H)$, $i = 1, 2, 3$. The extension techniques outlined in section 2.3.3 would be used to show global identification.

Now I demonstrate that F_i have unique representations through H_i. I consider two cases: one with $p_{12} > 0$, $p_{13} > 0$, $p_{23} > 0$, $p_{123} > 0$, and the other where $p_{12} > 0$, $p_{13} > 0$, $p_{23} > 0$, $p_{123} = 0$. In both cases, the only conditions required for uniqueness are $g_i \in L^1$ in a small neighborhood of t_0; these conditions are obviously satisfied. Note that the example in which $p_{123} = 1$ constitutes the chapter's original problem.

Case $p_{12} > 0$**,** $p_{13} > 0$**,** $p_{23} > 0$**,** $p_{123} > 0$

Observe that

$$H_1 = p_{123}\left(F_2 + \frac{p_{13}}{p_{123}}\right)\left(F_3 + \frac{p_{12}}{p_{123}}\right) - \frac{p_{12}p_{13}}{p_{123}}$$
$$H_2 = p_{123}\left(F_1 + \frac{p_{23}}{p_{123}}\right)\left(F_3 + \frac{p_{12}}{p_{123}}\right) - \frac{p_{12}p_{23}}{p_{123}}$$
$$H_3 = p_{123}\left(F_1 + \frac{p_{23}}{p_{123}}\right)\left(F_2 + \frac{p_{13}}{p_{123}}\right) - \frac{p_{13}p_{23}}{p_{123}}.$$

Taking into account that F_i are positive for $t > t_0$, I derive the following formulas:

$$F_1 = -\frac{p_{23}}{p_{123}} + \frac{1}{p_{123}}\sqrt{\frac{(p_{123}H_2 + p_{12}p_{23})(p_{123}H_3 + p_{13}p_{23})}{p_{123}H_1 + p_{12}p_{13}}}$$
$$F_2 = -\frac{p_{13}}{p_{123}} + \frac{1}{p_{123}}\sqrt{\frac{(p_{123}H_1 + p_{12}p_{13})(p_{123}H_3 + p_{13}p_{23})}{p_{123}H_2 + p_{12}p_{23}}}$$
$$F_3 = -\frac{p_{12}}{p_{123}} + \frac{1}{p_{123}}\sqrt{\frac{(p_{123}H_1 + p_{12}p_{13})(p_{123}H_2 + p_{12}p_{23})}{p_{123}H_3 + p_{13}p_{23}}}.$$

The expressions on the right-hand sides of these equations are $q_1(H)$, $q_2(H)$ and $q_3(H)$, respectively.

Case $p_{12} > 0$**,** $p_{13} > 0$**,** $p_{23} > 0$**,** $p_{123} = 0$

Because

$$H_1 = p_{12}F_2 + p_{13}F_3$$

$$H_2 = p_{12}F_1 + p_{23}F_3$$

$$H_3 = p_{13}F_1 + p_{23}F_2,$$

F_i can be expressed through H_i by inverting matrix

$$B = \begin{pmatrix} 0 & p_{12} & p_{13} \\ p_{12} & 0 & p_{23} \\ p_{13} & p_{23} & 0 \end{pmatrix}.$$

Because

$$B^{-1} = \begin{pmatrix} -\frac{p_{23}}{2p_{13}p_{12}} & \frac{1}{2p_{12}} & \frac{1}{2p_{13}} \\ \frac{1}{2p_{12}} & -\frac{p_{13}}{2p_{12}p_{23}} & \frac{1}{2p_{23}} \\ \frac{1}{2p_{13}} & \frac{1}{2p_{23}} & -\frac{p_{12}}{2p_{13}p_{23}} \end{pmatrix},$$

then

$$\begin{aligned} q_1(H) &= -\frac{p_{23}}{2p_{13}p_{12}}H_1 + \frac{1}{2p_{12}}H_2 + \frac{1}{2p_{13}}H_3 \\ q_2(H) &= \frac{1}{2p_{12}}H_1 - \frac{p_{13}}{2p_{12}p_{23}}H_2 + \frac{1}{2p_{23}}H_3 \\ q_3(H) &= \frac{1}{2p_{13}}H_1 + \frac{1}{2p_{23}}H_2 - \frac{p_{12}}{2p_{13}p_{23}}H_3. \end{aligned}$$

The expressions on the right-hand sides of these equations are $q_1(H)$, $q_2(H)$ and $q_3(H)$, respectively.

As we can see, in both cases F_i are uniquely expressed in terms of H_i. This completes the analysis of the second-price auctions with exogenous variation in the number of bidders.

Several papers have explored other types of auctions with exogenous variation in the number of bidders. For instance, McAfee and McMillan (1987) allow the number of actual bidders to be stochastic in first-price, sealed-bid auctions with independent private values. They investigate how bidders' uncertainty about the number of actual rivals affects their equilibrium behavior, not to mention the seller's expected revenue and other issues. In another study, Harstad, Kagel and Levin (1990) consider symmetric first-price and second-price auctions with an uncertain number of actual bidders. They show that first-price and second-price auctions, each with the number of bidders known or uncertain, and English auctions are revenue-equivalent.

2.7.2 Identification in generalized competing risks models

I now describe generalized competing risks models in detail. The main purpose of this section is to present conditions on observables sufficient to guarantee identification in generalized competing risks and to prove the identification result.

In section 2.2.1, I gave two examples of generalized competing risks models. First, I explained why we can consider second-price auctions to be a special case of these models. In the other example, I considered widely used classical competing risks models. In reliability theory, a classical competing risks model is a situation in which a machine breaks down as soon as one of its components reaches a failure state; the observed data pertain to the machine's lifetime and the component that caused the failure. One example in economics is duration models. Also, the Roy model is isomorphic to classical competing risks. In biometrics, the death of an individual because of a particular disease when that person is also facing several other diseases presents a classical competing risks model, based on a fundamental assumption that a single cause is behind every death. Generalized competing risks models relax this assumption and consider cases in which an object fails because of the cumulative failure of some of its elements rather than a single one.

I now proceed to a more detailed description of generalized competing risks models. For convenience, I use the terminology of reliability theory, which refers to generalized competing risks models as coherent systems.[10] Essentially, a coherent system is a system that collapses because several of its elements fail.

Suppose that a machine with a coherent structure consists of d elements. Denote the elements' lifetimes as $X_1,\ldots,X_d$ and the machine's lifetime as Z; the lifetime Z is a function of $X_1,\ldots, X_d$. Conveniently, Z can be characterized by fatal sets. A *fatal set* is a subset of parts such that the failure of all the parts in the subset causes the failure of the machine; in other words, it is a set of the elements that failed before the machine broke down. Even more conveniently, Z can be characterized by the collection $I_1, \ldots, I_m$ of *minimal fatal sets*, which are fatal sets that do not encompass other fatal sets.

The examples below clarify the structure of a coherent system. To guarantee that the probability of the simultaneous failure of several elements is 0, I suppose that the joint distribution of $X_1,\ldots, X_d$ is absolutely continuous.

Example 2.7.1. *Consider a classical competing risks model in which the collection of minimal fatal sets is* $I_1 = \{1\}, \ldots, I_d = \{d\}$*, and the machine's lifetime is*

$$Z = \min\{X_1, \ldots, X_d\}.$$

Clearly, the number of minimal fatal sets coincides with the number of elements. Furthermore, there are no fatal sets other than sets I_i*. Take, for instance, set* $\{1, 2\}$*. Although it is a superset of fatal sets* $\{1\}$ *and* $\{2\}$*, it is not fatal itself. Indeed, the death of these two*

[10] The concept of a coherent system was introduced in Barlow and Proschan (1975).

elements could not cause the machine's failure because the death of either of them would have led to failure earlier.

Example 2.7.2. *Consider a button auction with three bidders who have private values. In this case, the fatal sets are the sets of bidders who dropped out before the auction ended. The collection of minimal fatal sets is* $I_1 = \{2, 3\}$*,* $I_2 = \{1, 3\}$ *and* $I_3 = \{1, 2\}$*. Here, element lifetimes* X_1*,* X_2 *and* X_3 *are bidders' private values, and the lifetime* Z *is the transaction price. Notice that the number of minimal fatal sets is the same as the number of bidders, and there are no fatal sets besides* I_i*.*

Example 2.7.3. *Consider a machine with five parts. Let the collection of minimal fatal sets be* $I_1 = \{1, 2, 3\}$*,* $I_2 = \{1, 2, 4\}$*,* $I_3 = \{1, 3, 4\}$*,* $I_4 = \{2, 3, 4\}$*,* $I_5 = \{1, 3, 5\}$ *and* $I_6 = \{2, 3, 5\}$*. An example of a fatal set that is not a minimal fatal set is* $\{1, 2, 3, 5\}$*: It causes the death of the machine when, for instance, the machine's elements break in the order of 5, 1, 2 and 3. Set* $\{1, 2, 3, 4\}$*, on the other hand, is not fatal because all its three-element subsets are minimal fatal sets.*

For coherent systems, we want to learn the distributions of element lifetimes X_i from observed "autopsy" data, which comprise the machine's lifetime Z and a fatal set I that is responsible for the machine's failure. These autopsy models are defined in Meilijson (1981), which raises the question of whether the joint distribution of (Z, I) identifies the marginal distributions of lifetimes X_i.

In second-price auctions, observing the winner's identity and the transaction price is analogous to observing the fatal set and the lifetime of the machine in a coherent system.

Therefore, the problem of identifying the distributions of bidders' private values can be considered a special case of the identification problem for generalized competing risks models.

Meilijson (1981) claims that, under certain restrictions on a coherent system's structure, the distributions of the components' lifetimes are identified if the lifetimes are independent. To formulate the identification result, he introduces an incidence matrix constructed in the following way. Given a collection of minimal fatal sets, the coherent system's incidence matrix is a matrix M such that $M(i,j) = 1$ if $j \in I_i$, and $M(i,j) = 0$ otherwise, $i = 1, \ldots, m$, $j = 1, \ldots, d$.

For example, in the three-bidder auctions considered in Example 2.7.2, the incidence matrix is

$$M = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

In classical competing risks models, on the other hand, the incidence matrix is the $d \times d$ identity matrix.

The main result of Meilijson (1981) is that if X_1, X_2, ..., X_d are non-atomic and independent and possess the same essential infimum and supremum, and if the rank of M is d, then the joint distribution of Z and I uniquely determines the distribution of each X_j, $j = 1, \ldots, d$.

The idea behind Meilijson's proof is (a) to use data only from those cases where set I is a minimal fatal set and (b) to obtain integral equations that relate the distribution functions of components' lifetimes to to observable functions, and then apply to them a fixed point theorem for multidimensional functional spaces. Though Meilijson (1981) made important contributions, including the observation that only the data corresponding to minimal fatal sets can be considered, as well as and the rank condition on the incidence matrix, the proofs lack some essential details. First, the author does not discuss necessary conditions on observable data besides mentioning them as a prospect for future research. As we have seen in the auction model, however, such conditions are crucial for obtaining the existence and uniqueness result. Second, he does not explore the existence of underlying distributions that rationalize the observables. A possible reason for this omission is the fact that in the majority of generalized competing risks models, existence cannot be proved and must be assumed, as I explain below. Nevertheless, I show that existence can be established for a special class of competing risks models, and I present conditions on observables that are necessary and sufficient for existence. Another important piece missing from Meilijson's proof is conditions on observables sufficient to guarantee the uniqueness of underlying distributions consistent with the data. I provide these conditions for any generalized competing risks model. Finally, although the author mentions that the locally identified distribution functions can be extended to the whole support, he does not present a proof of this result. As in the auction, such a proof would require the identification result for the case in which all distribution functions have positive values at the initial point.

In this chapter, I suggest a new approach to identification in generalized competing risks models that offers a complete transparent proof of the identification result. I assume that the components' lifetimes have absolutely continuous distributions, even though Meilijson

(1981) obtains his result under the weaker assumption that the lifetimes' distributions are merely continuous. The idea behind my method is similar to the case of the auction; namely, I derive a system of non-linear differential equations that relates the underlying distribution functions to observable functions, then examine the existence and uniqueness issues for this system. I use the incidence matrix and assume the rank condition as in Meilijson.

In this section, I state the main results for generalized competing risks models. The identification proof and an outline of Meilijson's method are in the appendix. I assume that whenever the machine breaks, we observe its lifetime and the fatal set that caused the failure. Therefore, for any fatal set D, there is a corresponding observable function G_D:

$$G_D(t) = P(Z \leq t, D \text{ caused the failure}).$$

Because lifetimes X_i are independent, then

$$G_D(t) = \int_{t_0}^{t} (\prod_{j \in C_D} F_j(s))' \prod_{j \in D^c} (1 - F_j(s)) \prod_{j \in D \setminus C_D} F_j(s) ds, \tag{2.7.5}$$

where F_j is the distribution function of X_j, C_D is the intersection of all minimal fatal sets contained in D, $D^c = \{1, \ldots, d\} \setminus D$.

Let G_i be an observable function corresponding to the minimal fatal set I_i, $i = 1, \ldots, m$:

$$G_i(t) = \int_{t_0}^{t} (\prod_{j \in I_i} F_j(s))' \prod_{j \in I_i^c} (1 - F_j(s)) ds, \quad i = 1, \ldots, m. \tag{2.7.6}$$

System (2.7.6) of integral-differential equations is an analog of system (2.2.3). The differentiation of the equations in (2.7.6) yields the following system of non-linear differential equations:

$$(\prod_{j \in I_i} F_j)' = \frac{g_i}{\prod_{j \in I_i^c} (1 - F_j)}, \quad i = 1, \ldots, m. \tag{2.7.7}$$

I analyze this system with initial conditions

$$F_i(t_0) = 0, \quad i = 1, \ldots, d. \tag{2.7.8}$$

First, I consider the case in which the number of minimal fatal sets coincides with the number of the machine's components − that is $m = d$. In this instance, the matrix M is quadratic. Let b_{ij} stand for the (i, j) element of the inverse matrix M^{-1}.

The next theorem formulates the local existence result for problem (2.7.7)-(2.7.8) and describes the conditions on G_i that guarantee it.

Theorem 2.24. [11] *Let $m = d$. Suppose that functions G_i satisfy the following conditions:*

1. $G_i(t_0) = 0$, $i = 1, \ldots, d$
2. *Absolutely continuous in a neighborhood of* t_0, $i = 1, \ldots, d$
3. *Strictly increasing in a neighborhood of* t_0
4. $\lim_{t \downarrow t_0} \prod_{j=1}^{d} G_j^{b_{ij}}(t) = 0$, $i = 1, \ldots, d$

Then problem (2.7.7)-(2.7.8) has a solution F in a neighborhood of t_0.[12]

Notice that, from the model, conditions 1-4 in this proposition are necessary on G_i. Indeed, 1-3 follow directly from the definition of functions G_i. Given that conditions 1-3 hold, condition 4 can be obtained from (2.7.7). The interpretation of these conditions is similar to that of conditions (I) and (II) in the auction model.

An important difference between this case and the auction, however, is that even if problem (2.7.7)-(2.7.8) possesses a solution F and all F_i in this solution have the properties

[11] The proof of this theorem is available upon request.

[12] In the solution I refer to, F_i are positive for $t > t_0$.

of distribution functions, the existence of a solution to the model is not guaranteed. Indeed, to satisfy the model, F must solve equation (2.7.5) for any fatal set D. System (2.7.7), however, accounts only for the minimal fatal sets. Therefore, after finding a solution to (2.7.7)-(2.7.8), we have to substitute it into (2.7.5) to verify that it solves this equation for any D. Because it is difficult (and perhaps impossible) to find conditions on functions G_D under which the model has a solution, it is common in reliability theory to assume existence. The only situation in which the conditions in Theorem 2.24 guarantee existence is when $m = d$ and the only fatal sets in the model are minimal fatal sets. Notice that this is the case in the auction model analyzed in this chapter.

The next theorem provides the conditions on G_i that are sufficient for the uniqueness of a solution to (2.7.7)-(2.7.8).

Theorem 2.25. *Let $m = d$. Suppose that all conditions on G_i in Theorem 2.24 are satisfied. Denote*

$$\Gamma_i(t) = g_i \sum_{l \in I_i^c} \sum_{h=1}^{d} |b_{lh}| (\prod_{j \neq h} G_j^{lj}) G_h^{b_{lh}-1}.$$

If for any $i = 1, \ldots, d$

$$\Gamma_i \in L^1 \tag{2.7.9}$$

in a small neighborhood of t_0, then problem (2.7.7)-(2.7.8) has a unique solution in a neighborhood of t_0.

Because problem (2.7.7)-(2.7.8) has a unique solution, the model cannot have more than one solution. Therefore, the following corollary holds.

Corollary 2.2. *Let $m = d$. Suppose that all conditions in Theorem 2.25 are satisfied. Then a solution to the model, if it exists, is unique.*

When the number of minimal fatal sets exceeds d – that is, $m > d$ – the existence of a solution to the model is always assumed. It is easy, however, to indicate conditions on observable functions that guarantee the uniqueness of a solution to the model when one exists. Consider any $d \times d$ full-rank submatrix of M. Without a loss of generality, suppose that this submatrix is formed by the first d rows in M. The subsystem of (2.7.7) that comprises the differential equations corresponding to the first d rows in M has at most one solution if G_i satisfy the conditions in Theorem 2.25. Consequently, the model has at most one solution. We can find other sufficient conditions by choosing different submatrices of M.

This completes the discussion of local identification. The methods of section 2.3.3 allow us to show that the local solution to the model can be extended to the whole support and that this extension is unique. Furthermore, drawing on the results from section 2.4, we know that in some situations we can determine that the components' lifetimes are dependent. More specifically, the independence assumption is rejected when not all functions in the solution to the model are increasing.

2.8 Conclusion

This chapter has examined identification in second-price auctions within the independent-private-values framework when the only available data pertain to the winner's identity and the second-highest bid. The keys to identification are the independence of private values, as well as the conditions on the observable data obtained in this chapter. Recall that I have indicated conditions on observables that are necessary and sufficient for the existence of a solution to the model, and I have established restrictions on observables under which the model is point identified.

The chapter has provided methodological contributions by presenting a new way of proving identification in analyzed auction models. This approach, which employs the techniques of the theory of differential equations, is based on establishing the existence and uniqueness of a solution to the system of non-linear differential equations that relates the underlying unknown distribution functions to the observable data. This method is constructive and provides new insight by looking at identification from a fresh perspective. Though it allows us to explore identification in more general auction settings, this approach is not limited to auctions only. As I have demonstrated, it can be applied to prove identification in a wide class of generalized competing risks models used in statistical research. For these models, I have shown that if the model is misspecified, then it is possible that not all functions in its solution are distribution functions. For auctions with affiliated private values, I have presented point-wise tight bounds on the marginal distribution functions of interest and investigated how they change when we are supplied with data on other bids. Finally, I have addressed the estimation issue and suggested a consistent sieve minimum distance estimator of the relevant underlying distribution.

2.9 Appendix

2.9.1 Proofs of the results in section 2.2

Proof of Proposition 2.2. It suffices to show that $\lim_{t\downarrow t_0} \frac{F_1}{\sqrt{\frac{G_2G_3}{G_1}}}(t) = 1$. Let $t_1 > t_0$ be very close to t_0 and let $0 < M < 1$ be such that $F_i(t) \leq M$ for any $t \in (t_0, t_1)$, $i = 1, 2, 3$. Consider the first equation in system (2.2.3):

$$G_1(t) = \int_{t_0}^{t} (F_2F_3)^{'}(1 - F_1)ds.$$

Use it to obtain that

$$G_1(t_1) \geq \int_{t_0}^{t_1} (F_2F_3)^{'}(1 - M)ds = (1 - M)F_2(t_1)F_3(t_1),$$

$$G_1(t_1) \leq F_2(t_1)F_3(t_1).$$

Likewise,

$$(1 - M)F_1(t_1)F_3(t_1) \leq G_2(t_1) \leq F_1(t_1)F_3(t_1),$$

$$(1 - M)F_1(t_1)F_2(t_1) \leq G_3(t_1) \leq F_1(t_1)F_2(t_1).$$

Because $F_1 = \sqrt{\frac{F_1F_2F_1F_3}{F_2F_3}}$, then

$$F_1(t_1) \leq \frac{1}{1 - M}\sqrt{\frac{G_2G_3}{G_1}},$$
$$F_1(t_1) \geq \sqrt{1 - M}\sqrt{\frac{G_2G_3}{G_1}}.$$

Because $F_1(t_0) = 0$ and t_1 can be chosen arbitrarily close to t_0, then M can be arbitrarily close to 0. This implies that $\lim_{t\downarrow t_0} \frac{F_1}{\sqrt{\frac{G_2G_3}{G_1}}}(t) = 1$.

Proof of Corollary 2.1. Conditions (2.2.2) follow from Proposition 2.2 and the fact that $\lim_{t\downarrow t_0} F_i(t) = 0$, $i = 1, 2, 3$.

Theorem 2.3 and **Theorem 2.4** follow from the results in section 2.3.

2.9.2 Proofs of the results in section 2.3

Proof of Proposition 2.5

Auxiliary system with ϵ

The right-hand side $J(t, H)$ of the auxiliary system (DE_H) has singularities in H when $H_1 = 0$ or $H_2 = 0$ or $H_3 = 0$. These singularities can be handled by using a very small $\epsilon > 0$ and considering an auxiliary system with $\epsilon > 0$:

$$H_1' = \frac{g_1}{1 - \sqrt{\frac{H_2 H_3}{H_1 + \epsilon}}}$$

$$H_2' = \frac{g_2}{1 - \sqrt{\frac{H_1 H_3}{H_2 + \epsilon}}}$$

$$H_3' = \frac{g_3}{1 - \sqrt{\frac{H_1 H_2}{H_3 + \epsilon}}},$$

together with initial conditions

$$H_i(t_0) = 0, \quad i = 1, 2, 3. \qquad (IC_{H,\epsilon})$$

Denote

$$J^\epsilon(t, H) = \left(\frac{g_1(t)}{1 - \sqrt{\frac{H_2 H_3}{H_1 + \epsilon}}}, \frac{g_2(t)}{1 - \sqrt{\frac{H_1 H_3}{H_2 + \epsilon}}}, \frac{g_3(t)}{1 - \sqrt{\frac{H_1 H_2}{H_3 + \epsilon}}} \right)^{tr}$$

and rewrite the system with ϵ as

$$H'(t) = J^\epsilon(t, H(t)). \qquad (DE_{H,\epsilon})$$

The definition of a solution to $(DE_{H,\epsilon})$-$(IC_{H,\epsilon})$ is analogous to Definition 2.2 and defines a solution on $[t_0, t_0 + a]$ instead of $(t_0, t_0 + a]$.

Introduce

$$\bar{H}(\delta) = [0, \infty)^3 \cap \{(h_1, h_2, h_3)^{tr} : h_2 h_3 \leq \delta h_1, h_1 h_3 \leq \delta h_2, h_2 h_3 \leq \delta h_1\}$$

and let $\bar{D}(\delta) = [t_0, T] \times \bar{H}(\delta)$ be the domain of $J^\epsilon(t, H)$ (a.e. with respect to t). The difference between $\bar{H}(\delta)$ and $\bar{H}_0(\delta)$ is that $\bar{H}(\delta)$ allows H_i to take value 0.

Lemma 2.26. *Let observable functions G_i satisfy conditions (I) and (II). Let $J^\epsilon(t, H)$ be defined on $\bar{D}(\delta)$. Then ($DE_{H,\epsilon}$)-($IC_{H,\epsilon}$) has a solution on $[t_0, t_0 + a]$.*

Proof. To prove this result, I use a Tonelli approximation approach, which builds special approximations of a solution on very small intervals. These approximations have an important property − when the lengths of the intervals go to zero, the sequence of approximations has a subsequence converging to a solution to $(DE_{H,\epsilon})$-$(IC_{H,\epsilon})$.

Tonelli approximations are constructed step by step according to a specified rule. Consider, for example, intervals $[t_0, t_0 + \frac{1}{k}]$, $[t_0 + \frac{1}{k}, t_0 + \frac{2}{k}]$, ..., $[t_0 + \frac{r}{k}, t_0 + a]$, where $a \le \frac{r+1}{k}$, and k is very large. For these intervals an approximation is built in the following way. First, an approximation is found on $[t_0, t_0 + \frac{1}{k}]$, then it is extended to interval $(t_0 + \frac{1}{k}, t_0 + \frac{2}{k}]$. Next, the approximation is extended to $(t_0 + \frac{2}{k}, t_0 + \frac{3}{k}]$ and so on. This process is continued until the approximation is constructed on the whole interval $[t_0, t_0 + a]$.

A special feature of the Tonelli approach is that the extension of the approximation to $(t_0 + \frac{i}{k}, t_0 + \frac{i+1}{k}]$ is completely determined by the values of the approximating function on $[t_0 + \frac{i-1}{k}, t_0 + \frac{i}{k}]$ and therefore does not require any knowledge about the approximation on $[t_0, t_0 + \frac{i-1}{k})$.

Now I turn to describing the rule of constructing approximations. The integration of both sides in $(DE_{H,\epsilon})$ yields $H(t) = \int_{t_0}^{t} J^\epsilon(s, H)ds$. For a given k, denote a corresponding Tonelli approximation as $H^k = (H_1^k, H_2^k, H_3^k)$. Function H^k is defined according to the following rule:

$$H^k(t) = \int_{t_0}^{t} J^\epsilon\left(s, H^k\left(s - \frac{1}{k}\right)\right) ds, \quad t \in [t_0, t_0 + a]. \tag{2.9.1}$$

Choose a k that is large enough. To carry out the first step of constructing an approximation on $[t_0, t_0 + \frac{1}{k}]$, let

$$H_i^k(t) = 0, \quad t \in [t_0 - 1, t_0], \quad i = 1, 2, 3.$$

Let me show that formula (2.9.1) is meaningful. In the first step, it defines $H^k(t)$ for $t \in [t_0, t_0 + \min\{\frac{1}{k}, a\}]$. Because $J^\epsilon(s, H^k(s - \frac{1}{k})) = (g_1(s), g_2(s), g_3(s))^{tr}$ for any $s \in [t_0, t_0 + \min\{\frac{1}{k}, a\}]$ and $g_i \in L^1[t_0, t_0+a]$, then the integral on the right-hand side exists. For the next step to be well defined, I have to check that for $t \in [t_0, t_0 + \min\{\frac{1}{k}, a\}]$, the values of the constructed function $H^k(t) = (H_1^k, H_2^k, H_3^k)^{tr}$ belong to $\bar{H}(\delta)$. Indeed, $H_i^k(t) = G_i(t)$. Properties $\frac{H_2^k(t)H_3^k(t)}{H_1^k(t)} \leq \delta$, $\frac{H_1^k(t)H_3^k(t)}{H_2^k(t)} \leq \delta$ and $\frac{H_1^k(t)H_2^k(t)}{H_3^k(t)} \leq \delta$ follow from (2.3.3) and the fact that $\gamma < \delta$. Therefore, $H^k(t) \in \bar{H}(\delta)$.

In the second step, formula (2.9.1) defines H^k on $[t_0 + \frac{1}{k}, t_0 + \min\{\frac{2}{k}, a\}]$. For $t \in [t_0 + \frac{1}{k}, t_0 + \min\{\frac{2}{k}, a\}]$, the Lebesgue integral on the right-hand side exists because function $J^\epsilon\left(s, H^k\left(s - \frac{1}{k}\right)\right) ds$ is evidently measurable and bounded by a Lebesgue integrable function:

$$\left| J_i^\epsilon\left(s, H^k\left(s - \frac{1}{k}\right)\right) ds \right| \leq \frac{g_i(s)}{1 - \sqrt{\delta}} \in L^1[t_0, t_0 + a], \quad s \in [t_0, t_0 + \min\{\frac{2}{k}, a\}].$$

Clearly, $H_i^k(t) > 0$. Because $H_2^k(t) \leq \frac{G_2(t)}{1-\sqrt{\delta}}$, $H_3^k(t) \leq \frac{G_3(t)}{1-\sqrt{\delta}}$ and $H_1^k(t) \geq G_1(t)$, then

$$\frac{H_2^k(t)H_3^k(t)}{H_1^k(t)} \leq \frac{G_2(t)G_3(t)}{(1-\sqrt{\delta})^2 G_1(t)} \leq \frac{\gamma}{(1-\sqrt{\delta})^2} \leq \delta.$$

Likewise,

$$\frac{H_1^k(t)H_3^k(t)}{H_2^k(t)} \leq \frac{G_1(t)G_3(t)}{(1-\sqrt{\delta})^2 G_2(t)} \leq \frac{\gamma}{(1-\sqrt{\delta})^2} \leq \delta,$$
$$\frac{H_1^k(t)H_2^k(t)}{H_3^k(t)} \leq \frac{G_1(t)G_2(t)}{(1-\sqrt{\delta})^2 G_3(t)} \leq \frac{\gamma}{(1-\sqrt{\delta})^2} \leq \delta.$$

Therefore, $H^k(t) \in \bar{H}(\delta)$ for $t \in [t_0 + \frac{1}{k}, t_0 + \min\{\frac{2}{k}, a\}]$.

All subsequent steps are similar to the second step. By continuing to construct approximations in this manner, I can eventually define function H^k on the whole interval $[t_0, t_0 + a]$.

I take progressively smaller intervals and obtain a sequence of approximations $\{H^k\}$. Because for any k

$$\|H^k(t)\|_1 \leq \frac{G_1(t) + G_2(t) + G_3(t)}{1 - \sqrt{\delta}} \leq \frac{G_1(t_0 + a) + G_2(t_0 + a) + G_3(t_0 + a)}{1 - \sqrt{\delta}}, \tag{2.9.2}$$

functions H^k in this sequence are uniformly bounded. Moreover, sequence $\{H^k\}$ is equicontinuous, a property that is implied by inequality (2.9.3) and the absolute continuity of G_i on $[t_0, t_0 + a]$:

$$\|H^k(t) - H^k(\tau)\|_1 \leq \frac{\|G(t) - G(\tau)\|_1}{1 - \sqrt{\delta}}, \quad t, \tau \in [t_0, t_0 + a]. \tag{2.9.3}$$

According to the Arzela-Ascoli theorem, sequence $\{H^k\}$ is relatively compact in $C([t_0, t_0 + a], \bar{H})$, so it contains a subsequence $\{H^{k_m}\}$ such that for some function H

$$\sup_{t \in [t_0, t_0 + a]} \|H(t) - H^{k_m}(t)\|_1 \to 0$$

as $m \to \infty$. Because

$$J^{\epsilon}\left(t, H^{k_m}\left(t - \frac{1}{k_m}\right)\right) \to J^{\epsilon}(t, H(t)) \quad a.e. \quad [t_0, t_0 + a]$$

as $m \to \infty$, and a.e. on $[t_0, t_0 + a]$

$$\left\| J^{\epsilon}\left(t, H^{k_m}\left(t - \frac{1}{k_m}\right)\right) \right\|_1 \leq \frac{g_1(t) + g_2(t) + g_3(t)}{1 - \sqrt{\delta}} \in L^1[t_0, t_0 + a],$$

then according to the Lebesgue dominated convergence theorem, H solves

$$H(t) = \int_{t_0}^{t} J^{\epsilon}(s, H(s)) ds, \quad t \in [t_0, t_0 + a].$$

The last equation implies that H is absolutely continuous and solves $(DE_{H,\epsilon})$-$(IC_{H,\epsilon})$ a.e. on $[t_0, t_0 + a]$. □

Proof of Proposition 2.5. Choose a sequence ϵ_m such that $\epsilon_m \to 0$ as $m \to \infty$. For every ϵ_m, denote a solution constructed under Proposition 2.26 for this ϵ_m as H^{ϵ_m}. As I proved, for every ϵ_m, function H^{ϵ_m} is absolutely continuous on $[t_0, t_0 + a]$ and $H_i^{\epsilon_m}(t) > 0$, $t \in (t_0, t_0 + a]$.

Notice that the bounds in (2.9.2) and (2.9.3) do not depend on the value of ϵ, therefore,

$$\|H^{\epsilon_m}(t)\|_1 \leq \frac{\|G(t_0 + a)\|_1}{1 - \sqrt{\delta}}, \quad t \in [t_0, t_0 + a],$$

and

$$\|H^{\epsilon_m}(t) - H^{\epsilon_m}(\tau)\|_1 \leq \frac{\|G(t) - G(\tau)\|_1}{1 - \sqrt{\delta}}, \quad t, \tau \in [t_0, t_0 + a].$$

The last two inequalities and the Arzela-Ascoli theorem imply that sequence $\{H^m\}$ is relatively compact in $C([t_0, t_0 + a], \bar{H})$. Consequently, it has a subsequence H^{m_l} such that for some function H

$$\sup_{t \in [t_0, t_0 + a]} \|H(t) - H^{\epsilon_m}(t)\|_1 \to 0$$

as $m \to \infty$. Because

$$J^{\epsilon}(t, H^{\epsilon_m}(t)) \to J(t, H(t)) \;\; a.e. \;\; [t_0, t_0 + a]$$

as $m \to \infty$, and a.e. on $[t_0, t_0 + a]$

$$\|J^{\epsilon_m}(t, H^m(t))\|_1 \leq \frac{g_1(t) + g_2(t) + g_3(t)}{1 - \sqrt{\delta}} \in L^1[t_0, t_0 + a],$$

the Lebesgue dominated convergence theorem yields

$$H(t) = \int_{t_0}^{t} J(s, H(s))ds, \quad t \in [t_0, t_0 + a].$$

From the last equation, I conclude that H_i are absolutely continuous on $[t_0, t_0 + a]$ and constitute a solution to (DE_H)-(IC_H) on $(t_0, t_0 + a]$.

Remark. Take a solution H to (DE_H)-(IC_H) on $(t_0, t_0 + a]$ and define $H(t_0) = (0,0,0)^{tr}$. The proof of Proposition 2.5 implies that this extended function is absolutely continuous on $[t_0, t_0 + a]$. It clearly satisfies (DE_H)-(IC_H) a.e. on $[t_0, t_0 + a]$. Therefore, the solution to the auxiliary problem can be extended from $(t_0, t_0 + a]$ to $[t_0, t_0 + a]$.

Proof of Theorem 2.7

Proof of Lemma 2.8. From (DE_H) obtain

$$\begin{aligned} H_1' - \tilde{H}_1{}' &= \frac{g_1(F_1 - \tilde{F}_1)}{(1 - F_1)(1 - \tilde{F}_1)} \\ H_2' - \tilde{H}_2{}' &= \frac{g_2(F_2 - \tilde{F}_2)}{(1 - F_2)(1 - \tilde{F}_2)} \\ H_3' - \tilde{H}_3{}' &= \frac{g_3(F_3 - \tilde{F}_3)}{(1 - F_3)(1 - \tilde{F}_3)}. \end{aligned} \tag{2.9.4}$$

From equalities

$$\begin{aligned} H_1 - \tilde{H}_1 &= F_2(F_3 - \tilde{F}_3) + \tilde{F}_3(F_2 - \tilde{F}_2) \\ H_2 - \tilde{H}_2 &= F_1(F_3 - \tilde{F}_3) + \tilde{F}_3(F_1 - \tilde{F}_1) \\ H_3 - \tilde{H}_3 &= F_1(F_2 - \tilde{F}_2) + \tilde{F}_2(F_1 - \tilde{F}_1), \end{aligned}$$

find that on $(t_0, t_0 + c]$

$$F_1 - \tilde{F}_1 = -\frac{F_1}{\tilde{F}_3(F_2 + \tilde{F}_2)}(H_1 - \tilde{H}_1) + \frac{F_2}{\tilde{F}_3(F_2 + \tilde{F}_2)}(H_2 - \tilde{H}_2) + \frac{1}{F_2 + \tilde{F}_2}(H_3 - \tilde{H}_3) \tag{2.9.5}$$

$$F_2 - \tilde{F}_2 = \frac{\tilde{F}_2}{\tilde{F}_3(F_2 + \tilde{F}_2)}(H_1 - \tilde{H}_1) - \frac{F_2\tilde{F}_2}{F_1\tilde{F}_3(F_2 + \tilde{F}_2)}(H_2 - \tilde{H}_2) + \frac{F_2}{F_1(F_2 + \tilde{F}_2)}(H_3 - \tilde{H}_3)$$

$$F_3 - \tilde{F}_3 = \frac{1}{F_2 + \tilde{F}_2}(H_1 - \tilde{H}_1) + \frac{\tilde{F}_2}{(F_2 + \tilde{F}_2)F_1}(H_2 - \tilde{H}_2) - \frac{\tilde{F}_3}{(F_2 + \tilde{F}_2)F_1}(H_3 - \tilde{H}_3).$$

According to (2.2.1), there exist constants $C_1 > 0$, $C_2 > 0$ such that on $(t_0, t_0 + c]$

$$C_1 \le \frac{F_1}{\sqrt{\frac{G_2G_3}{G_1}}} \le C_2, \quad C_1 \le \frac{F_2}{\sqrt{\frac{G_1G_3}{G_2}}} \le C_2, \quad C_1 \le \frac{F_3}{\sqrt{\frac{G_1G_2}{G_3}}} \le C_2$$

and

$$C_1 \le \frac{\tilde{F}_1}{\sqrt{\frac{G_2G_3}{G_1}}} \le C_2, \quad C_1 \le \frac{\tilde{F}_2}{\sqrt{\frac{G_1G_3}{G_2}}} \le C_2, \quad C_1 \le \frac{\tilde{F}_3}{\sqrt{\frac{G_1G_2}{G_3}}} \le C_2$$

($t_0 + c$ can be taken close enough to t_0). Then on $(t_0, t_0 + c]$,

$$\begin{aligned}
|F_1 - \tilde{F}_1| &\le K\frac{1}{G_1}\sqrt{\frac{G_2G_3}{G_1}}|H_1 - \tilde{H}_1| + K\sqrt{\frac{G_3}{G_1G_2}}|H_2 - \tilde{H}_2| + K\sqrt{\frac{G_2}{G_1G_3}}|H_3 - \tilde{H}_3| \\
|F_2 - \tilde{F}_2| &\le K\sqrt{\frac{G_3}{G_1G_2}}|H_1 - \tilde{H}_1| + K\frac{1}{G_2}\sqrt{\frac{G_1G_3}{G_2}}|H_2 - \tilde{H}_2| + K\sqrt{\frac{G_1}{G_2G_3}}|H_3 - \tilde{H}_3| \qquad (2.9.6) \\
|F_3 - \tilde{F}_3| &\le K\sqrt{\frac{G_2}{G_1G_3}}|H_1 - \tilde{H}_1| + K\sqrt{\frac{G_1}{G_2G_3}}|H_2 - \tilde{H}_2| + K\frac{1}{G_3}\sqrt{\frac{G_1G_2}{G_3}}|H_3 - \tilde{H}_3|,
\end{aligned}$$

where $K > 0$ is a constant expressed in terms of C_1 and C_2. Let $B > 0$ be a constant that bounds F_i and $\tilde{F}_i$ from above on $[t_0, t_0 + c]$. Denote $C = \frac{K}{(1-B)^2}$. Inequalities (2.9.6) and equations (2.9.4) imply that a.e. on $[t_0, t_0 + c]$

$$\|H' - \tilde{H}'\|_1 \le C\left(\frac{g_1}{G_1} + \frac{g_2}{G_2} + \frac{g_3}{G_3}\right)\left(\sqrt{\frac{G_2G_3}{G_1}} + \sqrt{\frac{G_1G_3}{G_2}} + \sqrt{\frac{G_1G_2}{G_3}}\right)\|H - \tilde{H}\|_1.$$

Proof of Lemma 3.12. Hartman (1964) proves a similar lemma for smooth functions for the maxnorm and the euclidian norm. First, for any fixed i consider function $|z_i|$. Since z_i is absolutely continuous, $|z_i|$ is absolutely continuous too. $D_R|z_i(t)|$ then exists a.e. on $[\tau, \xi]$.

Let $t \in [\tau, \xi]$ be a point in which z_i has derivative. Use the definition of the right derivative:

$$D_R|z_i(t)| = \lim_{h \to +0} \frac{|z_i(t+h)| - |z_i(t)|}{h}$$

to conclude that $D_R|z_i(t)| = z_i'(t)$ if $z_i(t) > 0$ and $D_R|z_i(t)| = -z_i'(t)$ if $z_i(t) < 0$. Indeed, if $z_i(t) > 0$, then $z_i(t+h) > 0$ for small enough h, and $D_R|z_i(t)| = z_i'(t)$. In a similar way we consider the case $z_i(t) < 0$. If $z_i(t) = 0$, then

$$D_R|z_i(t)| = \lim_{h\to 0+} \frac{|z_i(t+h)|}{h} = |\lim_{h\to 0+} \frac{z_i(t+h)}{h}| = |z_i'(t)|.$$

In all three cases $D_R|z_i(t)| \leq |z_i'(t)|$.

Function $\|z\|_1$ is the sum of absolutely continuous function and, hence, absolutely continuous. Then a.e. on $[\tau, \xi]$

$$D_R\|z(t)\|_1 = D_R(\sum_{i=1}^{n} |z_i(t)|) = \sum_{i=1}^{n} D_R|z_i(t)| \leq \sum_{i=1}^{n} |z_i'(t)| = \|z'(t)\|_1.$$

Proof of Lemma 3.13. Results similar to the one in Lemma 3.13 have been obtained by researchers on a more general level. However, it is easier to prove this lemma directly than to show how it follows from more general results.

Function $\phi(t) = v(t)e^{-\int_\tau^t \Gamma(s)ds}$ is absolutely continuous as the product of two absolutely continuous function and

$$D_R\phi(t) = D_R(v(t))e^{-\int_\tau^t \Gamma(s)ds} - \Gamma(t)v(t)e^{-\int_\tau^t \Gamma(s)ds} \leq 0 \quad a.e.\ [\tau, \xi].$$

Szarski (1965) uses Zygmund's lemma to show that if ϕ is absolutely continuous and $D_R\phi(t) \leq 0$ a.e. on $[\tau, \xi]$, then ϕ is non-increasing on $[\tau, \xi]$. Since $\phi(\tau) = 0$, then $\phi(t) \leq 0$ on $[\tau, \xi]$ and, hence, $v(t) \leq 0$ on $[\tau, \xi]$.

Proofs of Proposition 2.11 and Theorem 2.12

Proof of Proposition 2.11. The proof uses the Tonelli approximations approach. It is similar to the proof of Proposition 2.26 and differs from it by technical details.

Let me first specify the domain of the right-hand side $J(t, H)$ of the auxiliary system (DE_H) and find a solution's interval of existence. Let $\Delta > 0$ be any number such that $\Delta < \min\{1 - v_1, 1 - v_2, 1 - v_3\}$. Define set

$$\bar{H}(\Delta) = [0, \infty)^3 \cap \{(h_1, h_2, h_3)^{tr} : h_2 h_3 \leq (v_1 + \Delta)^2 h_1, h_1 h_3 \leq (v_2 + \Delta)^2 h_2, h_2 h_3 \leq (v_3 + \Delta)^2 h_1\}.$$

Let the domain of $J(t, H)$ be $\bar{D}(\Delta) = [t_1, T] \times \bar{H}$. For a given Δ, I can always choose a $\gamma > 0$ small enough so that

$$(1 + \gamma)^2 v_1^2 \leq (v_1 + \Delta)^2, \quad (1 + \gamma)^2 v_2^2 \leq (v_2 + \Delta)^2, \quad (1 + \gamma)^2 v_3^2 \leq (v_3 + \Delta)^2.$$

Because $\lim_{t \downarrow t_1} G_i(t) = G_i(t_1)$, there exists a point $t_1 + a_1$, $a_1 > 0$, from $[t_1, T]$ such that

$$G_1(t_1 + a_1) - G_1(t_1) \leq \gamma v_2 v_3 (1 - v_1 - \Delta)$$

$$G_2(t_1 + a_1) - G_2(t_1) \leq \gamma v_1 v_3 (1 - v_2 - \Delta)$$

$$G_3(t_1 + a_1) - G_3(t_1) \leq \gamma v_1 v_2 (1 - v_3 - \Delta).$$

Interval $[t_1, t_1 + a_1]$ is an interval on which a local solution exists.

Now I construct Tonelli approximations. For any natural number k let

$$H_1^k(t) = v_2 v_3, \quad H_2^k(t) = v_1 v_3, \quad H_3^k(t) = v_1 v_2$$

for $t \in [t_1 - 1, t_1]$. Denote $v_0 = (v_2 v_3, v_1 v_3, v_1 v_2)^{tr}$ and let v_0^i be the i's coordinate of v_0, $i = 1, 2, 3$. Define function

$$H^k(t) = v_0 + \int_{t_1}^{t} J\left(s, H^k\left(s - \frac{1}{k}\right)\right) ds, \quad t \in [t_1, t_1 + a_1]. \tag{2.9.7}$$

This formula is meaningful. In the first step it defines H on $[t_1, t_1 + \min\{\frac{1}{k}, a_1\}]$. For t from this interval the Lebesgue integral on the right-hand side exists because the integrand is bounded from above by functions from $L^1[t_1, t_1 + a_1]$:

$$\left| J_i \left(s, H^k \left(s - \frac{1}{k} \right) \right) \right| \le \frac{g_i(s)}{1 - v_i}, \quad s \in [t_1, t_1 + \min\{\frac{1}{k}, a_1\}].$$

Evidently, for $t \in [t_1, t_1 + \min\{\frac{1}{k}, a_1\}]$

$$H_1^k(t) = v_2 v_3 + \frac{G_1(t) - G_1(t_1)}{1 - v_1}$$
$$H_2^k(t) = v_1 v_3 + \frac{G_2(t) - G_2(t_1)}{1 - v_2}$$
$$H_3^k(t) = v_1 v_2 + \frac{G_3(t) - G_3(t_1)}{1 - v_3}.$$

Let me show that $H^k(t) \in \bar{H}$ for $t \in [t_1, t_1 + \min\{\frac{1}{k}, a_1\}]$. Consider, for instance, $\frac{H_2^k H_3^k}{H_1^k}$. Because

$$H_2^k(t) \le v_1 v_3 + \frac{G_2(t_1 + a_1) - G_2(t_1)}{1 - v_2} \le v_1 v_3 + \frac{\gamma v_1 v_3 (1 - v_2 - \Delta)}{1 - v_2} \le (1 + \gamma) v_1 v_3,$$

$$H_3^k(t) \le (1 + \gamma) v_1 v_2,$$

$$H_1^k(t) \ge v_2 v_3,$$

then

$$\frac{H_2^k(t) H_3^k(t)}{H_1^k(t)} \le (1 + \gamma)^2 v_1^2 \le (v_1 + \Delta)^2.$$

Likewise,

$$\frac{H_1^k(t) H_3^k(t)}{H_2^k(t)} \le (v_2 + \Delta)^2, \quad \frac{H_1^k(t) H_2^k(t)}{H_3^k(t)} \le (v_3 + \Delta)^2.$$

In the second step formula (2.9.7) defines H on $[t_1 + \frac{1}{k}, t_1 + \min\{\frac{2}{k}, a_1\}]$. For t from this interval the Lebesgue integral on the right-hand side exists because

$$\left| J_i \left(s, H^k \left(s - \frac{1}{k} \right) \right) \right| \le \frac{g_i(s)}{1 - v_i - \Delta} \in L^1[t_1, t_1 + a_1], \quad s \in [t_1, t_1 + \min\{\frac{2}{k}, a_1\}].$$

Note that $H^k(t) \in \bar{H}$ for $t \in [t_1 + \frac{1}{k}, t_1 + \min\{\frac{2}{k}, a_1\}]$. Indeed,

$$H_2^k(t) \le v_1 v_3 + \frac{G_2(t_1 + a_1) - G_2(t_1)}{1 - v_2 - \Delta} \le v_1 v_3 + \gamma v_1 v_3 = (1 + \gamma) v_1 v_3,$$

$$H_3^k(t) \le (1+\gamma)v_1v_2,$$

$$H_1^k(t) \ge v_2v_3.$$

Therefore,

$$\frac{H_2^k(t)H_3^k(t)}{H_1^k(t)} \le (1+\gamma)^2v_1^2 \le (v_1+\Delta)^2.$$

In a similar way I can show that for $t \in [t_1+\frac{1}{k}, t_1+\min\{\frac{2}{k}, a_1\}]$

$$\frac{H_1^k(t)H_3^k(t)}{H_2^k(t)} \le (v_2+\Delta)^2, \quad \frac{H_1^k(t)H_2^k(t)}{H_3^k(t)} \le (v_3+\Delta)^2.$$

This process continues and defines function H^k on the whole interval $[t_1, t_1+a_1]$.

Now let me obtain the properties of sequence $\{H^k\}$. Inequality

$$\|H^k(t)\|_1 \le (1+\gamma)(v_2v_3+v_1v_3+v_1v_2)$$

for all $t \in [t_1, t_1+a_1]$ implies that sequence $\{H^k\}$ is uniformly bounded.

Because for any $t, \tau \in [t_1, t_1+a_1]$

$$\|H^k(t)-H^k(\tau)\|_1 \le \frac{|G_1(t)-G_1(\tau)|}{1-v_1-\Delta} + \frac{|G_2(t)-G_2(\tau)|}{1-v_2-\Delta} + \frac{|G_3(t)-G_3(\tau)|}{1-v_3-\Delta} \le$$
$$\le \frac{\|G(t)-G(\tau)\|_1}{1-\max\{v_1+\Delta, v_2+\Delta, v_3+\Delta\}},$$

and G_i are absolutely continuous on $[t_1, t_1+a_1]$, then sequence $\{H^k\}$ is equicontinuous. According to the Arzela-Ascoli theorem, $\{H^k\}$ is relatively compact in $C([t_1, t_1+a_1], \bar{H})$. Hence, it contains a subsequence H^{k_m} such that for some function H,

$$\sup_{[t_1, t_1+a_1]} \|H(t)-H^{k_m}(t)\|_1 \to 0 \text{ as } m \to \infty.$$

Because

$$J\left(t, H^{k_m}\left(t-\frac{1}{k_m}\right)\right) \to J(t, H(t)) \quad \text{a.e. on } [t_1, t_1+a_1]$$

and a.e. on $[t_1, t_1 + a_1]$

$$\left| J\left(t, H^{k_m}\left(t - \frac{1}{k_m}\right)\right)\right| \leq \frac{g_1(t) + g_2(t) + g_3(t)}{1 - \max\{v_1 + \Delta, v_2 + \Delta, v_3 + \Delta\}} \in L^1[t_1, t_1 + a],$$

then by the Lebesgue dominated convergence theorem, $H(t)$ solves

$$H(t) = v_0 + \int_{t_1}^{t} J(s, H(s))ds, \quad t \in [t_1, t_1 + a_1].$$

The last equation implies that H_i are absolutely continuous and solve (DE_H)- (2.3.6) on $[t_1, t_1 + a_1]$.

Proof of Theorem 2.12. According to Proposition 2.11, problem (DE_H)- (2.3.6) has a solution H on $[t_1, t_1 + a_1]$, $a_1 > 0$. Use this solution to find functions

$$F_1 = \sqrt{\frac{H_2 H_3}{H_1}}, \quad F_2 = \sqrt{\frac{H_1 H_3}{H_2}}, \quad F_3 = \sqrt{\frac{H_1 H_2}{H_3}}.$$

Clearly, $F = (F_1, F_2, F_3)^{tr}$ is absolutely continuous and solves (DE)- (2.3.5) on $[t_1, t_1 + a_1]$.

The uniqueness proof is based on obtaining a generalized local Lipschitz condition (2.3.4). Let F and $\tilde{F}$ be two local solutions of (DE)- (2.3.5). Without a loss of generality, assume that $[t_1, t_1 + a_1]$ is their common interval of existence. Let H and $\tilde{H}$ be their corresponding auxiliary functions:

$$H_1 = F_2 F_3, \quad H_2 = F_1 F_3, \quad H_3 = F_1 F_2,$$

$$\tilde{H}_1 = \tilde{F}_2 \tilde{F}_3, \quad \tilde{H}_2 = \tilde{F}_1 \tilde{F}_3, \quad \tilde{H}_3 = \tilde{F}_1 \tilde{F}_2.$$

Functions H and $\tilde{H}$ solve the auxiliary system (DE_H) a.e. on $[t_1, t_1 + a_1]$. Use (2.9.5) and the fact that F_i are separated from 0 in a neighborhood of t_1 (without a loss of generality, a_1 is small enough) to obtain

$$|F_i - \tilde{F}_i| \leq K \|H - \tilde{H}\|_1$$

on $[t_1, t_1 + a_1]$ for some constant K. Exploit (2.9.4) and establish that for some constant C,

$$\|H'(t) - \tilde{H}'(t)\|_1 \leq C(g_1(t) + g_2(t) + g_3(t))\|H(t) - \tilde{H}(t)\|_1.$$

a.e. on $[t_1, t_1 + a_1]$. Because $g_i \in L^1[t_1, t_1 + a_1]$, then lemmas 3.12 and 3.13 imply that H and $\tilde{H}$ coincide.

Proof of Theorem 2.14

Proposition 2.13 clearly implies equation (2.3.7). Because functions $\frac{H_2 H_3}{H_1}$, $\frac{H_1 H_3}{H_2}$, $\frac{H_1 H_2}{H_3}$ and sequence δ_n are strictly increasing, equation (2.3.7) implies that sequence T_n is strictly increasing. Because T_n increases and is bounded from above by T, it converges to some point $\bar{T} \le T$. If $\bar{T} < T$, then we get a contradiction with the condition $\delta_n \to 1$ and conditions (2.3.8). Thus, $\bar{T} = T$.

2.9.3 Auctions with any number of bidders: Discussion and proofs of results

First of all, it can be shown that all conditions in Theorem 2.15 are necessary conditions on G_i from the model. Theorem 2.15 states that they are also sufficient for the existence of a local solution.

The following conditions are analogous to (2.2.1):

$$\lim_{t \downarrow t_0} \frac{F_i}{\left(\frac{G_1 \ldots G_{i-1} G_{i+1} \ldots G_d}{G_i^{d-2}}\right)^{\frac{1}{d-1}}}(t) = 1, \quad i = 1, \ldots, d. \tag{2.9.8}$$

These conditions play an important role in the local identification result. The relations between the underlying distribution functions and observable functions at the limit enable us to obtain a condition on observable functions G_i sufficient to guarantee the uniqueness of a local solution. Conditions

$$\lim_{t \downarrow t_0} \frac{G_1 \ldots G_{i-1} G_{i+1} \ldots G_d}{G_i^{d-2}}(t) = 0, \quad i = 1, \ldots, d$$

are generalizations of (2.2.2).

As in the case of three buyers, system (2.5.1) can be rewritten in a convenient form by introducing d auxiliary functions $H_1,\ H_2,\ \ldots,\ H_d$ that stand for the distribution functions of $\max\{v_2, v_3, \ldots, v_d\}$, $\max\{v_1, v_3, \ldots, v_d\}$, …, $\max\{v_1, v_2, \ldots, v_{d-1}\}$, respectively:

$$H_1 = F_2F_3\ldots F_d, \quad H_2 = F_1F_3\ldots F_d, \quad \ldots, \quad H_d = F_1F_2\ldots F_{d-1}.$$

For $t > t_0$ functions F_i can be expressed through H_i as

$$F_1 = \left(\frac{H_2H_3\ldots H_d}{H_1^{d-2}}\right)^{\frac{1}{d-1}},\ F_2 = \left(\frac{H_1H_3\ldots H_d}{H_2^{d-2}}\right)^{\frac{1}{d-1}},\ \ldots,\ F_d = \left(\frac{H_1H_2\ldots H_{d-1}}{H_d^{d-2}}\right)^{\frac{1}{d-1}}, \quad (2.9.9)$$

therefore, system (2.5.1) can be rewritten in the following way:

$$H_i' = \frac{g_i}{1 - \left(\frac{H_1\ldots H_{i-1}H_{i+1}\ldots H_d}{H_i^{d-2}}\right)^{\frac{1}{d-1}}}, \quad i = 1, \ldots, d. \quad (2.9.10)$$

This system together with initial conditions

$$\lim_{t\downarrow t_0} H_i(t) = 0, \quad i = 1, \ldots, d.$$

constitutes the auxiliary problem. To deal with discontinuities in H on the right-hand side in (2.9.10), I introduce a very small number $\epsilon > 0$ and obtain an auxiliary system with ϵ:

$$H_i' = \frac{g_i}{1 - \left(\frac{H_1\ldots H_{i-1}H_{i+1}\ldots H_d}{H_i^{d-2}+\epsilon}\right)^{\frac{1}{d-1}}}, \quad i = 1, \ldots, d.$$

To prove Theorem 2.15, which is the local existence result for (2.5.1)-(2.5.2), I can use the same approach as in the case of three bidders. First, I can establish local existence for the auxiliary system with ϵ. Then I can show existence of local solution to the auxiliary problem by letting $\epsilon \to 0$. After that, I can use formulas (2.9.9), which express F through H, to prove that the main problem (2.5.1)-(2.5.2) has a local solution.[13]

[13] A detailed proof of Theorem 2.15 is available upon request.

Proof of Theorem 2.16. The existence part of this theorem follows from Theorem 2.15. To prove the uniqueness part, let F and $\tilde{F}$ be two solutions of (DE)-(IC) with a common interval of existence $[t_0, t_0 + c]$, $c > 0$. Let

$$H_i = F_1 \dots F_{i-1} F_{i+1} \dots F_d, \quad \tilde{H}_i = \tilde{F}_1 \dots \tilde{F}_{i-1} \tilde{F}_{i+1} \dots \tilde{F}_d, \quad i = 1, \dots, d.$$

The idea is to derive an inequality similar to (2.3.4). Use (2.9.9) and (2.9.10) to obtain that a.e. on $[t_0, t_0 + c]$

$$H_i' - \tilde{H}_i{}' = \frac{g_i(F_i - \tilde{F}_i)}{(1 - F_i)(1 - \tilde{F}_i)}. \tag{2.9.11}$$

The definitions of H and $\tilde{H}$ allow to express $H - \tilde{H}$ through $F - \tilde{F}$ as follows:

$$H - \tilde{H} = B(F, \tilde{F})(F - \tilde{F}),$$

where a $d \times d$ matrix B depends on F and $\tilde{F}$ in this way:

$$B(F, \tilde{F}) = \begin{pmatrix} 0 & F_3F_4 \dots F_d & \tilde{F}_2 F_4 \dots F_d & \tilde{F}_2 \tilde{F}_3 F_5 \dots F_d & \dots & \tilde{F}_2 \tilde{F}_3 \dots \tilde{F}_{d-1} \\ F_3 F_4 \dots F_d & 0 & \tilde{F}_1 F_4 \dots F_d & \tilde{F}_1 \tilde{F}_3 F_5 \dots F_d & \dots & \tilde{F}_1 \tilde{F}_3 \dots \tilde{F}_{d-1} \\ \dots & \dots & \dots & \dots & & \\ F_2 F_3 \dots F_{d-1} & \tilde{F}_1 F_3 \dots F_{d-1} & \tilde{F}_1 \tilde{F}_2 F_4 \dots F_d & \tilde{F}_1 \tilde{F}_2 \tilde{F}_3 \dots F_d & \dots & 0 \end{pmatrix}.$$

Conditions (2.9.8) imply that $\lim_{t \downarrow t_0} \frac{F_i}{\tilde{F}_i}(t) = 1$. Therefore, for a t close enough to t_0 (without a loss of generality, I can assume that $t_0 + c$ is close enough to t_0), matrix B can be written as

$$B(F, \tilde{F}) = (I + M_{o(1)}(F, \tilde{F})) B_0(F),$$

where I is the $d \times d$ identity matrix, $M_{o(1)}(F, \tilde{F})$ is a $d \times d$ matrix such that each of its elements is

$o(1)$ as $t \to t_0$, and $B_0(F) = B(F, F)$:

$$B_0(F) = \begin{pmatrix} 0 & F_3F_4\dots F_d & F_2F_4\dots F_d & F_2F_3F_5\dots F_d & \dots & F_2F_3\dots F_{d-1} \\ F_3F_4\dots F_d & 0 & F_1F_4\dots F_d & F_1F_3F_5\dots F_d & \dots & F_1F_3\dots F_{d-1} \\ \dots & \dots & \dots & \dots & & \\ F_2F_3\dots F_{d-1} & F_1F_3\dots F_{d-1} & F_1F_2F_4\dots F_d & F_1F_2F_3\dots F_d & \dots & 0 \end{pmatrix}.$$

Matrix $B_0(F)$ is symmetric and invertible at any point $t \neq t_0$. The inverse matrix is

$$B_0^{-1}(F) = \frac{1}{(d-1)F_1F_2\dots F_d} \begin{pmatrix} -(d-2)F_1^2 & F_1F_2 & F_1F_3 & F_1F_4 & \dots & F_1F_d \\ F_1F_2 & -(d-2)F_2^2 & F_2F_3 & F_2F_4 & \dots & F_2F_d \\ \dots & \dots & \dots & \dots & & \\ F_1F_d & F_2F_d & F_3F_d & F_4F_d & \dots & -(d-2)F_d^2 \end{pmatrix}.$$

Thus, $F - \tilde{F}$ can be expressed through $H - \tilde{H}$ as

$$F - \tilde{F} = B_0^{-1}(F)(I + M_{o(1)}(F, \tilde{F}))^{-1}(H - \tilde{H}). \tag{2.9.12}$$

The next step is to bound on $[t_0, t_0 + c]$ the absolute values of the elements in $B_0^{-1}(F)$ by observable functions (without a loss of generality, c can be assumed to be small enough). This is achieved by using conditions (2.9.8). Take, for instance, the element $B_0^{-1}(F)_{11}$ in the first row and the first column:

$$|B_0^{-1}(F)_{11}| = \left|\frac{(d-2)F_1}{(d-1)F_2\dots F_d}\right| \leq K_{11} \frac{\left(\frac{G_2\dots G_d}{G_1^{d-2}}\right)^{\frac{1}{d-1}}}{\prod_{i=2}^d \left(\frac{G_1\dots G_{i-1}G_{i+1}\dots G_d}{G_i^{d-2}}\right)^{\frac{1}{d-1}}}$$

for some constant K_{11}. Consider another cell in $B_0^{-1}(F)$, for example, the element $B_0^{-1}(F)_{12}$ in the first row and the second column:

$$|B_0^{-1}(F)_{12}| = \left|\frac{1}{(d-1)F_3\dots F_d}\right| \leq K_{12} \prod_{i=3}^d \left(\frac{G_1\dots G_{i-1}G_{i+1}\dots G_d}{G_i^{d-2}}\right)^{-\frac{1}{d-1}}$$

for some constant K_{12}. For the other elements, bounds are found in a similar way. Then equations (2.9.11) and (2.9.12) yield that a.e. on $[t_0, t_0 + c]$

$$\|H' - \tilde{H}'\|_1 \leq C \sum_{i=1}^{d} \left(\frac{G_1 G_2 \ldots G_{i-1} G_{i+1} \ldots G_d}{G_i^{d-1}} \right)^{\frac{1}{d-2}} \cdot \sum_{i=1}^{d} \frac{g_i}{G_i} \|H - \tilde{H}\|_1$$

for some constant C.

The last inequality and lemmas 3.12 and 3.12 prove Theorem 2.16.

2.9.4 Proofs of results in section 2.6

Proof of Proposition 2.18. Let F, $\tilde{F} \in \Lambda$, and $G = A(F)$, $\tilde{G} = A(\tilde{F})$. For convenience, I temporarily use the following metric:

$$d_1(F, \tilde{F}) = \sup_{t \in [0,1]} \sum_{j=1}^{3} |F_j(t) - \tilde{F}_j(t)|$$

$$d_1(G, \tilde{G}) = \sup_{t \in [0,1]} \sum_{j=1}^{3} |G_j(t) - \tilde{G}_j(t)|.$$

From the definition of A,

$$G_1(t) - \tilde{G}_1(t) = \int_{t_0}^{t} (F_2 F_3)'(1 - F_1) ds - \int_{t_0}^{t} (\tilde{F}_2 \tilde{F}_3)'(1 - \tilde{F}_1) ds =$$

$$= F_2 F_3 - \tilde{F}_2 \tilde{F}_3 - \int_{t_0}^{t} (F_2 F_3)'(F_1 - \tilde{F}_1) ds + \int_{t_0}^{t} \tilde{F}_1((\tilde{F}_2 \tilde{F}_3)' - (F_2 F_3)') ds.$$

Integration by parts yields

$$G_1(t) - \tilde{G}_1(t) = (F_2 F_3 - \tilde{F}_2 \tilde{F}_3)(1 - \tilde{F}_1) - \int_{t_0}^{t} (F_2 F_3)'(F_1 - \tilde{F}_1) ds + \int_{t_0}^{t} \tilde{F}_1'(F_2 F_3 - \tilde{F}_2 \tilde{F}_3) ds.$$

Knowing that $\tilde{F}_1$ and $F_2 F_3$ are distribution functions, obtain that for any $t \in [t_0, T]$,

$$|G_1(t) - \tilde{G}_1(t)| \leq 2 \sup_{[t_0,T]} |F_2 F_3 - \tilde{F}_2 \tilde{F}_3| + \sup_{[t_0,T]} |F_1 - \tilde{F}_1| \leq 3 d_1(F, \tilde{F}).$$

After arriving at similar inequalities for $G_2 - \tilde{G}_2$ and $G_3 - \tilde{G}_3$,

$$\sum_{j=1}^{3} |G_j(t) - \tilde{G}_j(t)| \leq 9d_1(F, \tilde{F}), \quad t \in [t_0, T],$$

and, hence,

$$d_1(G, \tilde{G}) \leq 9d_1(F, \tilde{F}).$$

Because

$$d_1(F, \tilde{F}) \leq \sqrt{3}d(F, \tilde{F}) \text{ and } d_1(G, \tilde{G}) \geq d(G, \tilde{G}), \tag{2.9.13}$$

then

$$d(G, \tilde{G}) \leq 9\sqrt{3}d(F, \tilde{F}),$$

or, equivalently,

$$d(A(F), A(\tilde{F})) \leq 9\sqrt{3}d(F, \tilde{F}).$$

Proof of Proposition 2.19. Because $F \in \overline{\Lambda}_\phi$, then there exists a sequence $F_n \in \Lambda_\phi$ such that $d(F_n, F) \to 0$ as $n \to \infty$. Take any two points $t_1, t_2 \in [t_0, T]$ and any F_i, $i = 1, 2, 3$. Convergence in metric d implies point-wise convergence. Therefore,

$$|F_j(t_1) - F_j(t_2)| = \lim_{n\to\infty} |F_{n,j}(t_1) - F_{n,j}(t_2)| \leq |\phi(t_1) - \phi(t_2)|.$$

The last inequality and the absolute continuity of ϕ imply that F_i is absolutely continuous. Functions $F_{n,i}$ are strictly increasing and converge to F_i point-wise, so F_i are increasing. $F_{n,i}(t_0)$ converge to $F_i(t_0)$. Hence, $F_i(t_0) = 0$. In a similar way, it can proved that $F_i(T) = 1$.

Because F_i are absolutely continuous, they can differentiated a.e. on $[t_0, T]$. Let t be a point at which both F_i and ϕ have derivatives. For any fixed h,

$$\frac{F_i(t+h) - F_i(t)}{h} = \lim_{n\to\infty} \frac{F_{n,i}(t+h) - F_{n,i}(t)}{h} \leq \frac{\phi(t+h) - \phi(t)}{h}.$$

Taking the limit as $h \to 0$, we obtain that $F_i'(t) \le \phi'(t)$.

Proof of Proposition 2.20. This proof is similar to the proof of Proposition 2.18. Let F, $\tilde{F} \in \overline{\Lambda}_\phi$ and $G = A(F)$, $\tilde{G} = A(\tilde{F})$. Integration by parts yields

$$G_1(t) - \tilde{G}_1(t) = (F_2F_3 - \tilde{F}_2\tilde{F}_3)(1 - \tilde{F}_1) - \int_{t_0}^{t} (F_2F_3)'(F_1 - \tilde{F}_1)ds + \int_{t_0}^{t} \tilde{F}_1'(F_2F_3 - \tilde{F}_2\tilde{F}_3)ds.$$

Therefore, for any $t \in [t_0, T]$,

$$|G_1(t) - \tilde{G}_1(t)| \le (1 + \phi(T) - \phi(t_0)) \sup_{[t_0,T]} |F_2F_3 - \tilde{F}_2\tilde{F}_3| +$$

$$+ 2(\phi(T) - \phi(t_0)) \sup_{[t_0,T]} |F_1 - \tilde{F}_1| \le (1 + 3\phi(T) - 3\phi(t_0))d_1(F, \tilde{F}).$$

Similar inequalities for $G_2 - \tilde{G}_2$ and $G_3 - \tilde{G}_3$ imply that

$$d_1(G, \tilde{G}) \le 3(1 + 3\phi(T) - 3\phi(t_0))d_1(F, \tilde{F}).$$

Taking into account (2.9.13),

$$d(A(F), A(\tilde{F})) \le C_0 d(F, \tilde{F}), \quad \text{where} \quad C_0 = 3\sqrt{3}(1 + 3\phi(T) - 3\phi(t_0)).$$

Proof of Proposition 2.21. Let $G_0 \in A(\Lambda_\phi)$ and $d(G_n, G_0) \to 0$ as $n \to \infty$ for $G_n \in A(\Lambda_\phi)$. Denote $F_0 = A^{-1}G_0$, $F_n = A^{-1}G_n$. Clearly, F_0, $F_n \in \Lambda_\phi$. I want to show that $d(F_n, F_0) \to 0$ as $n \to \infty$. Notice that the sequence F_n is equicontinuous, as all functions in the sequence are bounded and

$$|F_n(t_1) - F_n(t_2)| \le |\phi(t_1) - \phi(t_2)|$$

for any $t_1, t_2 \in [t_0, T]$. According to the Arzela-Ascoli theorem, there is a convergent subsequence F_{n_k}. Let F^* be the limit of F_{n_k}. Because $F^* \in \overline{\Lambda}_\phi$ and A is continuous on $\overline{\Lambda}_\phi$,

$$d(AF_{n_k}, AF^*) \to 0.$$

Thus, $AF^* = G_0$. Given that on $A(\Lambda_\phi)$ inverse A^{-1} is defined, $F^* = F_0$.

Proof of Lemma 2.22. $Q(F^*) = 0$. Because the inverse operator A^{-1} exists on $A(\Lambda_\phi)$, then $A(F) \neq G^*$ and, hence, $Q(F) > 0$ for any $F \in \Lambda_\phi$, $F \neq F^*$. Now consider $F \in \overline{\Lambda}_\phi \backslash \Lambda_\phi$. Taking into account the result of Proposition 2.19, conclude that there is a function F_i in F that is constant on some interval in $[t_0, T]$. There are two possible cases for F: when (a) $F_i(t) > 0$ for $t > t_0$, $i = 1, 2, 3$, and (b) some F_i takes value 0 in a right-hand side neighborhood of t_0. In the first case, $A(F) \neq G^*$ because the uniqueness result in section 2.3.2 was proved without the assumption of the strict monotonicity of F_i (see also discussion in section 2.5.2). In the second case, without a loss of generality assume that $F_1(t) = 0$, $t \in [t_0, t_0 + \omega)$. Then $G_2(t) = 0$ and $G_3(t) = 0$, $t \in [t_0, t_0 + \omega)$, for the corresponding $G = A(F)$. Because $G_i^*(t) > 0$ for $t > t_0$, $i = 1, 2, 3$, then obviously $A(F) \neq G^*$.

Proof of Theorem 2.23. To prove this theorem, I use lemmas A1 and A2 from Newey and Powell (2003). Consistency will hold if all conditions in Lemma A1 are satisfied. I divide these conditions into three groups, as in Newey and Powell (2003).

(i) According to Lemma 2.22, F^* is the unique minimizer of Q on $\overline{\Lambda}_\phi$.

(ii) Set $\overline{\Lambda}_\phi$ is compact. Let me show that Q and $\hat{Q}_n$ are continuous on $\overline{\Lambda}_\phi$ and

$$\sup_{F \in \overline{\Lambda}_\phi} |\hat{Q}_n(F) - Q(F)| \xrightarrow{p} 0. \tag{2.9.14}$$

The continuity of Q and $\hat{Q}_n$ will follow from the properties of A on $\overline{\Lambda}_\phi$. First, consider Q. For any F, $\tilde{F} \in \overline{\Lambda}_\phi$

$$|Q(F) - Q(\tilde{F})| = |E\{(G^* - A(F))^{tr}(G^* - A(F)) - (G^* - A(\tilde{F}))^{tr}(G^* - A(\tilde{F}))\}| =$$
$$= |E \sum_{j=1}^{3} (A(\tilde{F})_j - A(F)_j)(2G_j^* - A(F)_j - A(\tilde{F})_j)|.$$

For any $t \in [t_0, T]$, $A(F)_j(t) \leq 1$ and $G_j^*(t) \leq 1$, $j = 1, 2, 3$, therefore

$$|Q(F) - Q(\tilde{F})| \leq 4E \sum_{j=1}^{3} |A(\tilde{F})_j - A(F)_j|.$$

Applying the Cauchy-Schwartz inequality and (2.6.1),

$$|Q(F) - Q(\tilde{F})| \leq 4\sqrt{3}E\sqrt{(A(\tilde{F}) - A(F))^{tr}(A(\tilde{F}) - A(F))} \leq 4\sqrt{3}d(A(F), A(\tilde{F})) \leq$$
$$\leq 4\sqrt{3}C_0 d(F, \tilde{F}).$$

Thus, function Q is Lipschitz and therefore continuous.

Now consider function $\hat{Q}_n$. Similar to the methods described above,

$$\begin{aligned}
|\hat{Q}_n(F) - \hat{Q}_n(\tilde{F})| &\leq \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{3} |(\hat{G}_{n,j}(t_i) - A(F)_j(t_i))^2 - (\hat{G}_{n,j}(t_i) - A(\tilde{F})_j(t_i))^2| = \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{3} |(A(\tilde{F})_j(t_i) - A(F)_j(t_i))(2\hat{G}_{n,j}(t_i) - A(\tilde{F})_j(t_i) - A(F)_j(t_i))| \leq \\
&\leq \frac{4\sqrt{3}}{n}\sum_{i=1}^{n} \sqrt{(A(\tilde{F})(t_i) - A(F)(t_i))^{tr}(A(\tilde{F})(t_i) - A(F)(t_i))} \leq 4\sqrt{3}d(A(F), A(\tilde{F})) \leq \\
&\leq 4\sqrt{3}C_0 d(F, \tilde{F}).
\end{aligned} \tag{2.9.15}$$

Property (2.9.14) will follow from Lemma A2 in Newey and Powell (2003). Indeed, it is clear that

$$\forall (F \in \overline{\Lambda}_\phi) \quad \hat{Q}_n(F) \xrightarrow{p} Q(F).$$

This fact combined with (2.9.15) implies (2.9.14).

(iii) This condition follows from assumption (2.6.2).

Conditions (i)-(iii) imply the consistency property (2.23).

2.9.5 Discussion of generalized competing risks and proof of identification

First, I outline Meilijson's approach. From (2.7.7), Meilijson obtains a system of integral equations that do not contain the derivatives of F_j:

$$F(t) = \exp\left\{\tilde{T}\log\int_{t_0}^{t}\exp\{-\bar{M}\log(1-F(s))dG(s)\}\right\},$$

where matrix $\bar{M}$ is such that $\bar{M}(i,j) = 1 - M(i,j)$ and $\tilde{T} = (M^{tr}M)^{-1}M^{tr}$. He suggests applying to these equations a fixed point theorem for multidimensional functional spaces. As I mentioned, however, his proofs miss important parts.

I now turn to describing my method. The rank condition implies that $m \geq d$ – that is, there are at least as many minimal fatal sets as the number of the elements in a coherent system. First, I consider case $m = d$ and assume that the rank condition for the incidence matrix M holds – that is, M is invertible. Introduce auxiliary functions

$$H_i = \prod_{j\in I_i} F_j, \quad i = 1,\ldots,d,$$

and denote $H = (H_1,\ldots,H_d)^{tr}$. The rank condition guarantees that functions F_i, $i = 1,\ldots,d$, are uniquely expressed through functions H_i, $i = 1,\ldots,d$, via multiplication, division and taking a rational root. Indeed,

$$\log H_i = \sum_{j\in I_i}\log F_j, \quad i = 1,\ldots,d.$$

These equations can be rewritten as $\log H = M\log F$, therefore $F = \exp\{M^{-1}\log H\}$. Let b_{ij} stand for the (i,j) element of inverse matrix M^{-1}. (Clearly, numbers b_{ij} are rational.) Then

$$F_i = \prod_{j=1}^{d} H_j^{b_{ij}}, \quad i = 1,\ldots,d. \tag{2.9.16}$$

Similar to the auction problem, I obtain an auxiliary system of differential equations by rewriting (2.7.7) in terms of H:

$$H_i' = \frac{g_i}{\prod_{j \in I_i^c} \left(1 - \prod_{l=1}^d H_l^{b_{jl}}\right)}, \quad i = 1, \ldots, d. \tag{2.9.17}$$

As with the auction, the existence of a local solution to (2.7.7)-(2.7.8) can be proved in several steps. First, to avoid discontinuities in H, I can modify the auxiliary system (2.9.17) by introducing a very small number ϵ when necessary. Using Tonelli approximations, I can establish existence of a local solution for the auxiliary system with ϵ. After that, I can take the limit as $\epsilon \to 0$ and show the existence of a local solution for (2.9.17). Finally, I can use formulas (2.9.16) to prove existence of a local solution to problem (2.7.7)-(2.7.8).

The proof of local existence in Theorem 2.24 is omitted. It is available upon request. Below I prove the uniqueness part of Theorem 2.25.

Proof of Theorem 2.25. The idea of the proof is similar to the auctions case. Let F and $\tilde{F}$ be two local solutions to (2.7.7)-(2.7.8) with a common interval of existence $[t_0, t_0 + c]$. Let H and $\tilde{H}$ be the corresponding auxiliary functions. Then H and $\tilde{H}$ solve auxiliary system (2.9.17) a.e. on $(t_0, t_0 + c]$. Denote the right-hand side of (2.9.17) as

$$J(t, H) = \left(\frac{g_1(t)}{\prod_{j \in I_1^c} \left(1 - \prod_{l=1}^d H_l^{b_{jl}}\right)}, \ldots, \frac{g_d(t)}{\prod_{j \in I_d^c} \left(1 - \prod_{l=1}^d H_l^{b_{jl}}\right)} \right).$$

A plan is to derive a generalized local Lipschitz condition and then use lemmas 3.12 and 3.13 to establish that H and $\tilde{H}$ coincide. This will imply that F and $\tilde{F}$ coincide. Consider $H_i - \tilde{H}_i$ for any

i and let $|I_i^c|$ for the number of elements in I_i^c. Then a.e. on $[t_0, t_0+c]$

$$|H_i' - \tilde{H}_i'| = \frac{g_i}{\prod_{j\in I_i^c}(1-F_j)} - \frac{g_i}{\prod_{j\in I_i^c}(1-\tilde{F}_j)} =$$

$$= \frac{g_i}{\prod_{j\in I_i^c}(1-F_j)\prod_{j\in I_i^c}(1-\tilde{F}_j)}\left(\prod_{j\in I_i^c}(1-F_j) - \prod_{j\in I_i^c}(1-F_j+(F_j-\tilde{F}_j))\right) \leq$$

$$\leq \frac{g_i}{\prod_{j\in I_i^c}(1-F_j)\prod_{j\in I_i^c}(1-\tilde{F}_j)} 2^{|I_i^c|-1}\sum_{j\in I_i^c}|F_j-\tilde{F}_j| \leq C_i g_i \sum_{j\in I_i^c}|F_j-\tilde{F}_j|$$

for some constant C_i. Let us bound differences $|F_j - \tilde{F}_j|$ from above by expressions of $|H_j - \tilde{H}_j|$. According to (2.9.16), for $t > t_0$,

$$F_j - \tilde{F}_j = \prod_{l=1}^{d} H_l^{b_{jl}} - \prod_{l=1}^{d} \tilde{H}_l^{b_{jl}},$$

therefore

$$F_j - \tilde{F}_j = \sum_{h=1}^{d}\prod_{l<h} H_l^{b_{jl}} \prod_{m>h} \tilde{H}_m^{b_{jm}} (H_h^{b_{jh}} - \tilde{H}_h^{b_{jh}})$$

For $x_1, x_2 > 0$, by the mean value theorem

$$x_l^\alpha - x_2^\alpha = \alpha(\theta x_1 + (1-\theta)x_2)^{\alpha-1}(x_1 - x_2),$$

where $\theta = \theta(x_1, x_2) \in [0, 1]$. If $\alpha \geq 1$, then

$$|x_l^\alpha - x_2^\alpha| \leq \alpha(\max\{x_1, x_2\})^{\alpha-1}|x_1 - x_2|.$$

If $\alpha < 1$, then

$$|x_l^\alpha - x_2^\alpha| \leq |\alpha|(\min\{x_1, x_2\})^{\alpha-1}|x_1 - x_2|.$$

Because $H_h(t), \tilde{H}_h(t) > 0$ for $t > t_0$, then for $t > t_0$,

$$|H_h^{b_{jh}}(t) - \tilde{H}_h^{b_{jh}}(t)| \leq W_{jh}(t)|H_h(t) - \tilde{H}_h(t)|,$$

where

$$W_{jh}(t) = \Big(1(b_{jh} \geq 1)\max\{H_h(t), \tilde{H}_h(t)\} + 1(b_{jh} < 1)\min\{H_h(t), \tilde{H}_h(t)\}\Big)^{b_{jh}-1}.$$

Because

$$\lim_{t \downarrow t_0} \frac{H_h(t)}{G_h(t)} = 1 \text{ and } \lim_{t \downarrow t_0} \frac{\tilde{H}_h(t)}{G_h(t)} = 1,$$

then

$$\lim_{t \downarrow t_0} \frac{W_{jh}(t)}{G_h^{b_{jh}-1}(t)} = 1.$$

Thus, for $t > t_0$,

$$|F_j - \tilde{F}_j| \leq L_j \sum_{h=1}^{d} (\prod_{l \neq h} G_l^{b_{jl}}) G_h^{b_{jh}-1} |b_{jh}| |H_h - \tilde{H}_h|$$

for some constants $L_j > 0$. Thus, a.e. on $[t_0, t_0 + c]$

$$|H_i'(t) - \tilde{H}_i'(t)| \leq D_i g_i \sum_{j \in I_i^c} \sum_{h=1}^{d} (\prod_{l \neq h} G_l^{b_{jl}}(t)) G_h^{b_{jh}-1}(t) |b_{jh}| |H_h(t) - \tilde{H}_h(t)|$$

for some constants $D_i > 0$ and, hence,

$$||H'(t) - \tilde{H}'(t)||_1 \leq C(\Gamma_1(t) + \ldots \Gamma_d(t)) ||H(t) - \tilde{H}(t)||_1$$

for some constant $C > 0$. This inequality and lemmas 3.12 and 3.13 imply that $H(t) = \tilde{H}(t)$, $t \in [t_0, t_0 + c]$.

Chapter 3

Identification in Dependent Generalized Competing Risks Models

3.1 Introduction

This chapter analyzes identification in second-price auctions when bidders have private values and auction's outcomes provide data only on the winner's identity and the transaction price. Chapter 2 proves that in this framework, the distribution of bidders' values is identified if these values are independent and the observables satisfy certain conditions. In many situations, however, the independence assumption is dubious and cannot be substantiated. The main goal of this chapter is to investigate the identification issue in the absence of independence. Many of the results obtained in the paper can be extended to dependent generalized competing risks models.

For affiliated private values, Athey and Haile (2002) show that when only a subset of bids is observed, then the joint distribution of bidders' values is not identified without any additional assumptions. In particular, even if all the bids except for the highest bid are known, there are many distributions consistent with the data.

Even though the distribution of values is not identified, the data are informative and allow us to find bounds on the distribution. These bounds can exploited in the analysis of counterfactuals and other applications. This chapter considers two observational schemes. In section 3.2.1, I examine the first scheme, which is the case when only the winner's identity and the transaction price are observed. For this scheme, I derive bounds on the joint distribution of values for any subset of bidders. It is of interest to analyze how these bounds change when more data become available − data on other identities or other bids. That is why in section 3.2.2, I consider the second scheme, which is the situation when all the identities and all the bids except for the highest bid are known. For this scheme,I present bounds on the joint distribution of values for the set of all bidders and bounds on the marginal distributions. All the bounds in section 3.2 are derived for any type of dependence, not only when private values are affiliated.

The results on bounds in section 3.2 rely on the assumption that bidders play their weakly dominant strategy by submitting their true values. For sealed-bid auctions, this assumption is plausible. For ascending auctions, however, researchers are often reluctant to associate a non-winning bid with a value of the corresponding bidder. I investigate what happens when the equilibrium condition is relaxed and substituted with two weaker assumptions. These assumptions were first introduced in Haile and Tamer (2003) who used them to obtain bounds on distributions. The first assumption supposes that bidders do not bid more than they are willing to pay. The other one suggests that bidders do not allow an opponent to win at a price they are willing to beat. I show that for the first observational scheme, the bounds remain the same. I also discuss why this is not so for the second scheme.

Suppose now that some information about the properties or a form of the joint distribution of values is available. Then a natural question is whether this additional knowledge can help to identify the distribution. To be more specific, imagine that we know how the joint distribution depends on the marginal distributions. Then the model implies relationships between the marginal distributions and the observables. If the marginal distributions are uniquely determined from these relationships then the joint distribution is identified as well. I take on this problem for the first observational scheme.

It is known that any distribution can be uniquely represented through the marginal distributions in the form of a copula. My goal is to indicate conditions on copulas sufficient to guarantee the identification of the marginal distributions and, hence, the identification of the joint distribution. I present such conditions for the case of two bidders in section 3.3.1. In section 3.3.2, I analyze the case of an arbitrary number of bidders and prove identification for Archimedean copulas in the situation when there are only two types of bidders.

The Appendix collects the proofs of the theorems, propositions and lemmas.

3.2 Bounds on distributions

Because the auction problem is related to generalized competing risks models, I start by reviewing the competing risks literature regarding partial identification. The study of partial identification in classical competing risks models was initiated by Peterson (1976), who obtained tight point-wise bounds on the joint and marginal survival functions. Crowder (1991) and Bedford and Meilijson (1997) obtained new results on bounds for those functions. We also can look to Manski (1990), who examined partial identification for self-selection models, of which competing risks models are a subset. For generalized competing risks models, several results are established by Deshpande and Karia (1997).

Throughout this section I suppose that d bidders are participating in a second-price auction, and their private values $X_1, \ldots, X_d$ have continuous marginal distributions on the same support $[t_0, T]$. I also assume that $P(X_i = X_j) = 0$, $i \neq j$, so that the probability of a tie is 0.

Suppose that $D = \{i_1, \ldots, i_r\}$ is a subset that consists of bidders $i_1, \ldots, i_r$. Define Q_D as the distribution function of valuations of bidders in this subset:

$$Q_D(t_{i_1}, \ldots, t_{i_r}) = P(X_{i_1} \leq t_{i_1}, \ldots, X_{i_r} \leq t_{i_r})$$

For $D = \{1, \ldots, d\}$, I denote Q_D as Q. For $D = \{1, \ldots, m-1, m+1, \ldots, d\}$, I denote Q_D as Q_{-m}. For $D = \{j\}$, I denote Q_D as F_j.

3.2.1 Identification when only the winner's identity and the winning price are observed

In this section, I explore the situation when only the winner's identity and the transaction price are observed. In other words, we know functions $G_1, \ldots, G_d$:

$$G_i(t) = P(i \text{ wins, price } \le t).$$

I suppose that bidders play their weakly dominant strategies by submitting their true values.

To obtain the lower bound on Q_D, I use the fact that if bidder $j \notin D = \{i_1, \ldots, i_r\}$ wins and the price does not exceed t, then all the values $X_{i_1}, \ldots, X_{i_r}$ do not exceed t either. In other words, functions G_j, $j \notin D$, provide information about the lower bound on Q_D. On the other hand, if bidder i_k wins, then it is not known how large the value X_{i_k} is and, consequently, G_{i_k} is not helpful in finding the lower bound on Q_D.

To obtain the upper bound on Q_D, I exploit the fact that if we know an upper bound on value X_{i_k}, then we know an upper bound on the price when bidder i_k wins. If we know upper bounds on values $X_{i_1}, \ldots, X_{i_r}$, and bidder $j \notin D = \{i_1, \ldots, i_r\}$ wins, then in general no conclusion can be made about the price. In other words, only functions $G_{i_1}, \ldots, G_{i_r}$ determine the upper bound on Q_D.

The theorem below formalizes this discussion and presents bounds on distribution functions Q_D.

Theorem 3.1. *Suppose that bidders play their weakly dominant strategy by submitting their true values. Also suppose that only the winner's identity and the transaction price are ob-*

served.

(a) Then Q_D is bounded from below as follows:

$$Q_D(t_{i_1}, \ldots, t_{i_r}) \geq \sum_{l=1}^{r} 1(t_{i_l} = T) G_{i_l}(\min_{k=1,\ldots,r} t_{i_k}) + \sum_{j \in CD} G_j(\min_{k=1,\ldots,r} t_{i_k}).$$

(b) Function Q is bounded from above as follows:

$$Q(t_1, \ldots, t_d) \leq 1(\min_i t_i > t_0) \sum_{j=1}^{d} G_j(\min\{t_j, \max_{i \neq j} t_i\}).$$

For any $m = 1, \ldots, d$, distribution function Q_{-m} is tightly bounded from above as follows:

$$Q_{-m}(t_1, \ldots, t_{m-1}, t_{m+1}, \ldots, t_d) \leq 1(\min_{i \neq m} t_i > t_0) \left(\sum_{j \neq m} G_j(t_j) + G_m(\max_{i \neq m} t_i) \right).$$

If D contains at most $d - 2$ elements, then Q_D is bounded from above as follows:

$$Q_D(t_{i_1}, \ldots, t_{i_r}) \leq 1(\min_{i_k \in D} t_{i_k} > t_0) \left(\sum_{i_k \in D} G_{i_k}(t_{i_k}) + \sum_{j \in CD} G_j(T) \right).$$

It is useful to demonstrate tight bounds on the marginal distribution functions. By theorem 3.1, for any $j = 1, \ldots, d$,

$$1(t = T) G_j(T) + \sum_{i \neq j} G_i(t) \leq F_j(t) \leq 1(t > t_0) \left(G_j(t) + 1 - G_j(T) \right).$$

Theorem 3.1 relies on the fact that bidders submit their true values. However, this condition can be relaxed. Consider the following two assumptions.

Assumption 1 (A1). Bidders do not bid more than they are willing to pay.

Assumption 2 (A2). Bidders do not allow an opponent to win at a price they are willing to beat.

These assumptions were introduced in Haile and Tamer (2003). The authors were among the first ones to relax equilibrium conditions and allow other types of bidders' behavior. One of their contributions is the construction of bounds on distributions for this limited structure.

The proposition below shows that in the framework of this paper, the bounds in theorem 3.1 are correct when the equilibrium condition is replaced with assumptions 1 and 2.

Proposition 3.2. *Suppose that only the winner's identity and the transaction price are observed.*

(a) If A1 holds, then Q_D are tightly bounded from above as in theorem 3.1.

(b) If A2 holds, then Q_D are tightly bounded from below as in theorem 3.1.

3.2.2 Identification when all the identities and all the bids except for the highest bid are observed

Let Π_d denote the set of all the permutations of set $\{1, \ldots, d\}$ and $\rho \in \Pi_d$. Let $\rho(i)$ stand for the ith element of permutation ρ. In the auction context, bidder $\rho(i)$ is the ith highest bidder.

The following $d!$ functions are observed:

$$W_\rho(s_2, \ldots, s_d) = P(\cap_{i \neq 1}(b_{\rho(i)} \leq s_i), b_{\rho(1)} > b_{\rho(2)} > \ldots > b_{\rho(d)}).$$

Notice that

$$G_j(t) = \sum_{\rho \in \Pi_d : \rho(1) = j} W_\rho(t, \ldots, t).$$

Also introduce the following $d(d-1)$ functions:

$$G_{jh}(t) = P(\max_{i\neq j, i\neq h} b_i < b_h, b_h < b_j, b_h \leq t), \quad h \neq j, \;\; j = 1, \ldots, d, \;\; h = 1, \ldots, d.$$

Note that

$$G_{jh}(t) = \sum_{\rho\in\Pi_d:\rho(1)=j,\rho(2)=h} W_\rho(t, \ldots, t).$$

Value $G_{jh}(t)$ is the probability that bidder j wins, bidder h submits the second-highest bid and this bid does not exceed t.

For simplicity, I obtain bounds only for the joint distribution function Q and marginal distribution functions F_j, $j = 1, \ldots, d$.

Theorem 3.3. *Suppose that bidders play their weakly dominant strategy by submitting their true values. Also suppose all the identities and all the bids except for the highest bid are observed.*

(a) Then function Q is bounded above and below as follows:

$$\prod_{i=1}^{d} 1(t_i = T) \leq Q(t_1, \ldots, t_d) \leq \sum_{\rho\in\Pi_d} W_\rho(\min\{t_{\rho(1)}, t_{\rho(2)}\}, \ldots, \min_{m=1,\ldots,l} t_{\rho(m)}, \ldots, \min_{i=1,\ldots,d} t_i).$$

(b) The marginal distribution functions F_j, $j = 1, \ldots, d$, are bounded above and below as follows:

$$\sum_{i\neq j} G_i(t) + 1(t = T)G_j(T) \leq F_j(t) \leq \sum_{i\neq j} G_{ji}(t) + \sum_{i\neq j} G_{ij}(t) + 1(t > t_0) \sum_{i\neq j} \sum_{h\neq i, h\neq j} G_{ih}(T).$$

Part (b) of theorem 3.3 states that in order to obtain bounds on the marginal distribution functions, it is enough to know the identities of the winner and the second-highest bidder, and also the transaction price. In other words, for these bounds, all the information about the bids and the identities of the bidders whose bids are lower than the transaction price, is irrelevant.

The lower bounds on F_j from theorems 3.1 and 3.3 coincide. The upper bound, on the other hand, in general is tighter in the second case versus the first.

I illustrate the difference between bounds in theorems 3.1 and 3.3 in the following example, where I show the bounds on the marginal distributions of bidders' values.

Example 3.4. *Consider the auction with three buyers. Let* $\tilde{X}_1$, $\tilde{X}_2$, $\tilde{X}_3$ *and* A *be independent random variables distributed on* $[0, 1]$ *with distribution functions* $\tilde{F}_1(t) = t$, $\tilde{F}_2(t) = t^2$, $\tilde{F}_3(t) = \sqrt{t}$ *and* $\tilde{F}_A(t) = t$. *Let private values* X_1, X_2 *and* X_3 *of the buyers be*

$$X_1 = 0.25\tilde{X}_1 + 0.75A$$

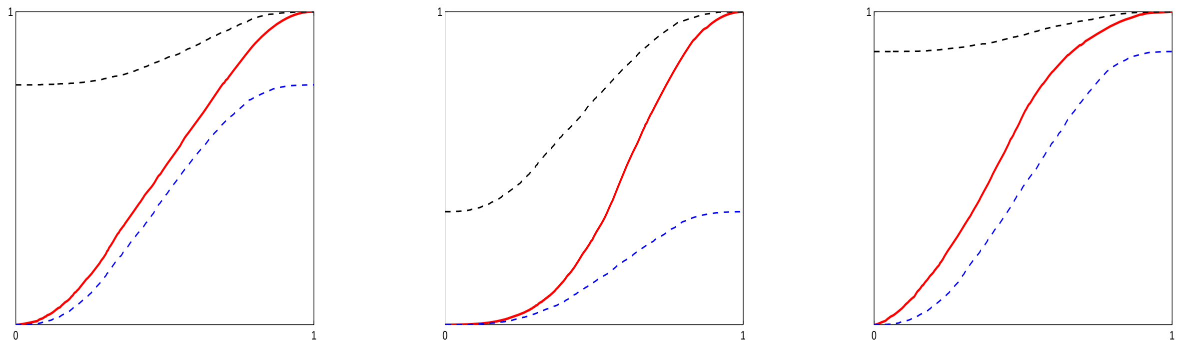
$$X_2 = 0.6\tilde{X}_2 + 0.4A$$

$$X_3 = 0.5\tilde{X}_3 + 0.5A.$$

Figure 3.1 shows the bounds on F_1, F_2 *and* F_3 *from theorem 3.1. Figure 3.2 shows the bounds from theorem 3.3.*

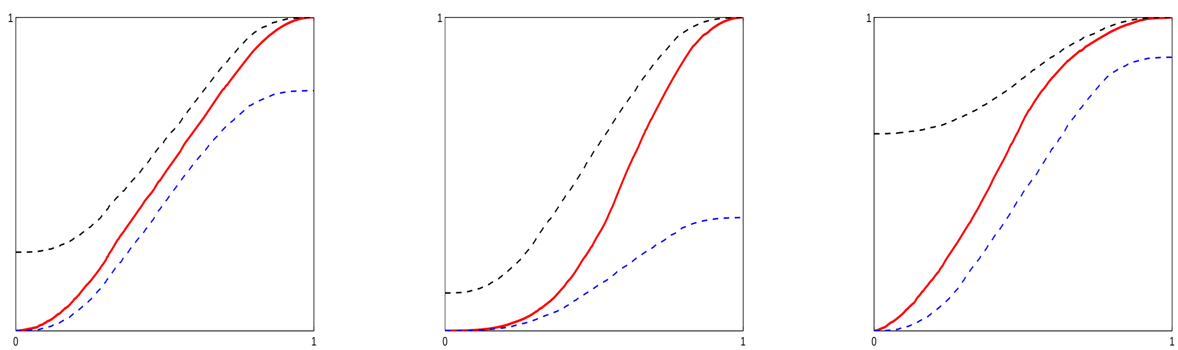
3.3 Identification using copulas

In this section, I show that if the form of dependence is specified, then it may be possible to identify the joint distribution. Any joint distribution can be represented in terms of the marginal distributions by means of a copula. Then the question is to explore whether the



F_1 F_3 F_3

Figure 3.1. Bounds on the marginal distribution functions from theorem 3.1.



F_1 F_2 F_3

Figure 3.2. Bounds on the marginal distribution functions from theorem 3.3.

joint distribution is identified if its copula is known. I examine identification for various classes of copulas.

Once the form of a copula is specified, similarly to the case of independent values in Komarova (2007) I can write a system of differential equations that describes relationships between the unknown marginal distributions of bidders' values and observables. The goal is to show that for given initial conditions, this system can have at most one solution. In order to show this, I use methods developed in Komarova (2007).

First, I present identification results for two bidders. I provide two different sufficient conditions for identification. Both conditions are based on bounding from above the difference in the derivatives of two possible solutions. The case of two bidders is easier to analyze than the general case because we always know that the transaction price is the value of the bidder who lost. Nevertheless, the techniques that I use for this case, prove to be helpful for any number of bidders in the situation when bidders are of two types. For an arbitrary number of bidders of several types, I discuss the difficulties of establishing identification and give a detailed analysis for the class of Archimedean copulas.

Throughout this section I assume that a solution to the model exists and concentrate on proving the uniqueness of a solution. The major part of the identification result is to establish uniqueness in a small neighborhood of the initial support point. Then it is easy to show that this local solution can be uniquely extended along the whole support. I omit the proof of the extension result because it is very similar to the proof of an analogous result in Komarova (2007).

3.3.1 Two bidders

I start by reviewing the definition of copulas. Let $I = [0, 1]$.

Definition 3.1. *A copula is a function C from I^2 to I with the following properties:*

1. *For every u_1, u_2 from I,*

$$C(u_1, 0) = 0 = C(0, u_2),$$

$$C(u_1, 1) = u_1, \quad C(1, u_2) = u_2.$$

2. *For every u_1, u_2, $\tilde{u}_1$,$\tilde{u}_2$ from I such that $u_1 \leq \tilde{u}_1$ and $u_2 \leq \tilde{u}_2$,*

$$C(\tilde{u}_1, \tilde{u}_2) + C(u_1, u_2) \geq C(\tilde{u}_1, u_2) + C(u_1, \tilde{u}_2).$$

I consider copulas with the following properties:

$$\begin{aligned} &\frac{\partial C(u_1, u_2)}{\partial u_1} \text{ is absolutely continuous with respect to } u_2 \\ &\frac{\partial C(u_1, u_2)}{\partial u_2} \text{ is absolutely continuous with respect to } u_1 \end{aligned} \tag{3.3.1}$$

and

$$\frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2} = \frac{\partial^2 C(u_1, u_2)}{\partial u_2 \partial u_1} \text{ a.e. on } I^2. \tag{3.3.2}$$

Let bidders' values X_1 and X_2 be absolutely continuous random variables. Then distribution functions F_1 and F_2 are absolutely continuous. The random vector (X_1, X_2) has the following joint distribution function:

$$Q(t_1, t_2) = C(F_1(t_1), F_2(t_2)).$$

Properties (3.3.1) and (3.3.2) and the absolute continuity of F_1 and F_2 guarantee that this distribution has density

$$q(t_1, t_2) = \frac{\partial^2 C(F_1(t_1), F_2(t_2))}{\partial u_1 \partial u_2} f_1(t_1) f_2(t_2),$$

where $f_1 = F_1'$, $f_2 = F_2'$.

Now let me construct the system of differential equations that allows us to determine F_1 and F_2 from observables. Because

$$G_1(t) = P(\text{price} < t, 1 \text{ wins}) = P(b_2 < t, b_1 \geq t) + P(b_2 < b_1, b_1 \leq t) =$$
$$= \int_t^T \left(\int_{t_0}^t q(t_1, t_2) dt_2 \right) dt_1 + \int_{t_0}^t \left(\int_{t_0}^{t_1} q(t_1, t_2) dt_2 \right) dt_1$$

and G_1 is absolutely continuous on $[t_0, T]$, then a.e. on $[t_0, T]$

$$g_1(t) = \int_t^T f(t_1, t) dt_1 = \int_t^T \frac{\partial^2 C(F_1(t_1), F_2(t))}{\partial u_1 \partial u_2} f_1(t_1) f_2(t) dt_1 =$$
$$= f_2(t) \frac{\partial C(F_1(T), F_2(t))}{\partial u_2} - f_2(t) \frac{\partial C(F_1(t), F_2(t))}{\partial u_2} = f_2(t) \left(1 - \frac{\partial C(F_1(t), F_2(t))}{\partial u_2} \right),$$

where $g_1 = G_1'$. Similarly, a.e. on $[t_0, T]$

$$g_2(t) = f_1(t) \left(1 - \frac{\partial C(F_1(t), F_2(t))}{\partial u_1} \right),$$

where $g_2 = G_2'$. Therefore, the system of interest is

$$F_1' = \frac{g_2}{1 - \frac{\partial C(F_1, F_2)}{\partial u_1}} \qquad (DE_1)$$

$$F_2' = \frac{g_1}{1 - \frac{\partial C(F_1, F_2)}{\partial u_2}}.$$

I want to show that (DE_1) together with initial conditions

$$F_1(t_0) = F_2(t_0) = 0 \qquad (IC_1)$$

can have at most one solution (F_1, F_2) on $[t_0, T]$.[1] Two theorems below present conditions on copulas sufficient to guarantee the uniqueness of a solution.

Theorem 3.5. *Suppose that* $\frac{\partial C(u_1,u_2)}{\partial u_1}$ *is decreasing in* u_1 *and* $\frac{\partial C(u_1,u_2)}{\partial u_2}$ *is decreasing in* u_2*. Then* (DE_1)*-*(IC_1) *has only one solution on* $[t_0, T]$*.*

Theorem 3.6. *Suppose that* C *has second-order derivatives* $\frac{\partial^2 C(u_1,u_2)}{\partial u_1^2}$ *and* $\frac{\partial^2 C(u_1,u_2)}{\partial u_1^2}$ *and these derivatives are bounded for small* u_1 *and* u_2*. Suppose that* $\frac{\partial^2 C(u_1,u_2)}{\partial u_1 \partial u_2}$ *is also bounded for small* u_1 *and* u_2*. Then* (DE_1)*-*(IC_1) *can have only one solution in a neighborhood of* t_0*.*

To prove theorem 3.5, I show that if problem (DE_1)-(IC_1) has two local solutions $F = (F_1, F_2)$ and $\tilde{F} = (\tilde{F}_1, \tilde{F}_2)$, then a.e. in a neighborhood of t_0

$$\frac{d\|F - \tilde{F}\|^2}{dt} \leq 0,$$

where $\|F - \tilde{F}\|$ stands for the Euclidean norm of $F - \tilde{F}$. This inequality and initial conditions (IC_1) imply that F and $\tilde{F}$ coincide on their common interval of existence.

To prove theorem 3.6, I demonstrate that for two local solutions F and $\tilde{F}$ there exists a constant $K > 0$ such that a.e. in a neighborhood of t_0

$$\frac{d\|F - \tilde{F}\|_1}{dt} \leq K(g_1 + g_2)\|F - \tilde{F}\|_1,$$

where $\|F - \tilde{F}\|_1$ is the norm $|F_1 - \tilde{F}_1| + |F_2 - \tilde{F}_2|$. Because $K(g_1 + g_2)$ is Lebesgue integrable in a neighborhood of t_0, then this inequality and initial conditions (IC_1) guarantee that F and $\tilde{F}$ are identical on their common interval of existence.

[1] I do not present to proof of the existence of a solution and focus on proving uniqueness.

Example 3.7. ***Fairlie-Gumbel-Morgenstern (FGM) copulas****.*

$$C(u_1, u_2) = u_1 u_2 (1 + \theta(1 - u_1)(1 - u_2)), \quad \theta \in [-1, 1].$$

Notice that all second-order derivatives

$$\frac{\partial^2 C(u_1, u_2)}{\partial u_1^2} = -2\theta u_2(1 - u_2),$$
$$\frac{\partial^2 C(u_1, u_2)}{\partial u_2^2} = -2\theta u_1(1 - u_1),$$
$$\frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2} = 1 + \theta - 2\theta u_1 - 2\theta u_2 + 4\theta u_1 u_2$$

are bounded around $(0, 0)$*, therefore according to theorem 3.6, the joint distribution is point identified.*

Example 3.8. ***Ali-Mikhail-Haq (AMH) copulas****.*

$$C(u_1, u_2) = \frac{u_1 u_2}{1 + \theta(1 - u_1)(1 - u_2)}, \quad \theta \in [-1, 1].$$

If $\theta \in (-1, 0]$*, then either of theorems 3.5 and 3.6 can be applied to show identification. If* $\theta \in (0, 1]$*, then theorem 3.6 should be used. If* $\theta = -1$*, then theorem 3.5 should be used.*

Now I want to discuss *Archimedean copulas*. These copulas have the following form:

$$C(u_1, u_2) = \psi^{-1} \left(\psi(u_1) + \psi(u_2) \right),$$

where function ψ is defined on $(0, 1]$ and

$$\psi(1) = 0, \quad \lim_{x \to 0} \psi(x) = \infty, \quad \psi'(x) < 0, \quad \psi''(x) > 0.$$

The proposition below is a corollary from theorem 3.5 and it presents a condition on function ψ sufficient to guarantee identification.

Proposition 3.9. *If function* $\frac{\psi''(x)}{(\psi'(x))^2}$ *is increasing, then problem* (DE_1)*-*(IC_1) *has only one solution on* $[t_0, T]$.

Example 3.10. ***Archimedean copulas***.

Clayton copulas: $\psi(x) = \frac{1}{\theta}(x^{-\theta} - 1)$, $\theta \in (0, \infty)$. *Identification holds because function* $\frac{\psi''(x)}{(\psi'(x))^2} = (\theta + 1)x^{\theta}$ *is increasing. It is interesting to note that theorem 3.6 cannot be applied to these copulas because, for instance,* $\frac{\partial^2 C(u_1,u_2)}{\partial u_1^2}$ *is not bounded around* $(0, 0)$.

Gumbel copulas: $\psi(x) = (-\ln x)^{\theta}$, $\theta \in [1, \infty)$. *The joint distribution is identified because* $\frac{\psi''(x)}{(\psi'(x))^2} = \frac{\theta - 1 - \ln x}{\theta(-\ln x)^{\theta}}$ *is an increasing function.*

Frank copulas:

$$\psi(x) = -\ln \frac{e^{-\theta x} - 1}{e^{-\theta} - 1}, \quad \theta \in (-\infty, \infty)/\{0\},$$

$$\psi(x) = -\ln x, \quad \theta = 0.$$

On the one hand, $\frac{\psi''(x)}{(\psi'(x))^2} = e^{\theta x}$ *and, therefore, proposition 3.9 can be applied to prove identification when* $\theta \geq 0$. *On the other hand, for any* θ, *all second-order derivatives are bounded around* $(0, 0)$ *and, hence, theorem 3.6 can be used to show identification.*

Joe copulas: $\psi(x) = -\ln(1 - (1 - x)^{\theta})$, $\theta \in [1, \infty)$. *Because function* $\frac{\psi''(x)}{(\psi'(x))^2} = (1 - x)^{2-2\theta}$ *is increasing, then the joint distribution is identified.*

Ali-Mikhail-Haq (AMH) copulas. *These copulas were considered in example 3.8. For* $\theta \in (-1, 1]$ *these copulas are Archimedean copulas with*

$$\psi(x) = \ln \frac{1 + \theta(1 - x)^{-\theta}}{x}.$$

Function $\frac{\psi''(x)}{(\psi'(x))^2} = 1 - \frac{2\theta}{\theta+1}x$ *is increasing for* $\theta \in (-1, 0]$*, therefore, for such* θ *identification follows from proposition 3.9. In example 3.8 I explained why the joint distribution is identified for* $\theta \in (0, 1]$ *as well. Even though for* $\theta = -1$ *the AMH copulas are not Archimedean, I showed in example 3.8 that the joint distribution is identified for this* θ *too.*

3.3.2 Any number of bidders

Mathematical model

Let me start by giving the definition of a d-dimensional copula.

Definition 3.2. *Let* $a = (a_1, \dots, a_d)$ *and* $b = (b_1, \dots, b_d)$*. Consider a d-box with* 2^d *vertices* $c = (c_1, \dots, c_d)$*, where each* c_i *is either* a_i *or* b_i*,* $i = 1, \dots, d$*. Denote it as* $[\boldsymbol{a,b}]$*. Consider a function* H *defined on this box. Denote* $\boldsymbol{t} = (t_1, \dots, t_d)$ *The H-volume of this box is defined as*

$$V_H([\boldsymbol{a,b}]) = \Delta_{a_d}^{b_d} \Delta_{a_{d-1}}^{b_{d-1}} \cdots \Delta_{a_2}^{b_2} \Delta_{a_1}^{b_1} H(\boldsymbol{t}),$$

where

$$\Delta_{a_i}^{b_i} H(\boldsymbol{t}) = H(t_1, \dots, t_{i-1}, b_i, t_{i+1}, \dots, t_d) - H(t_1, \dots, t_{i-1}, a_i, t_{i+1}, \dots, t_d).$$

Definition 3.3. *A d-dimensional copula* C *is a function from* I^d *to* I *with the following properties:*

1. *For every* $a, b \in I^d$ *such that* $a \leq b$*,*

$$C(\boldsymbol{u}) = 0 \quad \textit{if at least one coordinate of } \boldsymbol{u} \textit{ is } 0.$$

2. *If all coordinates of* $\boldsymbol{u}$ *are 1 except for* u_i, *then* $C(\boldsymbol{u}) = u_i$.

3. *For every* $a, b \in I^d$ *such that* $a \leq b$,

$$V_C([\boldsymbol{a},\boldsymbol{b}]) \geq 0.$$

For simplicity, I first consider the case of three bidders. Let $C(u_1, u_2, u_3)$ be a copula such that

(a) all mixed second-order derivatives $\frac{\partial^2 C(u_1,u_2,u_3)}{\partial u_i \partial u_j}$, $i \neq j$, are absolutely continuous with respect to u_k, $k \neq i$, $k \neq j$,

(b) all mixed third-order derivatives $\frac{\partial^2 C(u_1,u_2,u_3)}{\partial u_i \partial u_j \partial u_k}$, $i \neq j$, $i \neq k$, $k \neq j$, coincide a.e. on I^3.

Consider

$$G_1(t) = Pr(\max\{b_2, b_3\} < b_1, \max\{b_2, b_3\} \leq t) = P(\max\{b_2, b_3\} < b_1, b_1 \leq t)+$$
$$+P(\max\{b_2, b_3\} \leq t, b_1 > t) = \int_{t_0}^{t} \frac{\partial C}{\partial u_1}(F_1(s_1), F_2(s_1), F_3(s_1)) f_1(s_1) ds_1 +$$
$$+\int_{t}^{T} \frac{\partial C}{\partial u_1}(F_1(s_1), F_2(t), F_3(t)) f_1(s_1) ds_1 = \int_{t_0}^{t} \frac{\partial C}{\partial u_1}(F_1(s_1), F_2(s_1), F_3(s_1)) f_1(s_1) ds_1 +$$
$$+C(1, F_2(t), F_3(t)) - C(F_1(t), F_2(t), F_3(t))$$

I rewrite the last equation in the following way.

$$g_1 = \frac{\partial C}{\partial u_1}(F_1, F_2, F_3)F_1' + \frac{\partial C}{\partial u_2}(1, F_2, F_3)F_2' + \frac{\partial C}{\partial u_3}(1, F_2, F_3)F_3' - \frac{\partial C}{\partial u_1}(F_1, F_2, F_3)F_1' -$$
$$-\frac{\partial C}{\partial u_2}(F_1, F_2, F_3)F_2' - \frac{\partial C}{\partial u_3}(F_1, F_2, F_3)F_3' = F_2'\left(\frac{\partial C}{\partial u_2}(1, F_2, F_3) - \frac{\partial C}{\partial u_2}(F_1, F_2, F_3)\right) +$$
$$+F_3'\left(\frac{\partial C}{\partial u_3}(1, F_2, F_3) - \frac{\partial C}{\partial u_3}(F_1, F_2, F_3)\right).$$

Thus, the marginal distribution functions (F_1, F_2, F_3) constitute a solution to the system of differential equations

$$g_1 = F_2'\left(\frac{\partial C}{\partial u_2}(1, F_2, F_3) - \frac{\partial C}{\partial u_2}(F_1, F_2, F_3)\right) + F_3'\left(\frac{\partial C}{\partial u_3}(1, F_2, F_3) - \frac{\partial C}{\partial u_3}(F_1, F_2, F_3)\right)$$
$$g_2 = F_1'\left(\frac{\partial C}{\partial u_1}(F_1, 1, F_3) - \frac{\partial C}{\partial u_1}(F_1, F_2, F_3)\right) + F_3'\left(\frac{\partial C}{\partial u_3}(F_1, 1, F_3) - \frac{\partial C}{\partial u_3}(F_1, F_2, F_3)\right)$$
$$g_3 = F_1'\left(\frac{\partial C}{\partial u_1}(F_1, F_2, 1) - \frac{\partial C}{\partial u_1}(F_1, F_2, F_3)\right) + F_2'\left(\frac{\partial C}{\partial u_2}(F_1, F_2, 1) - \frac{\partial C}{\partial u_2}(F_1, F_2, F_3)\right)$$

with initial conditions

$$F_1(t_0) = F_2(t_0) = F_3(t_0) = 0.$$

This system can be resolved with respect to the derivatives F_1', F_2', F_3':

$$F_1' = -\frac{D_{23}D_{32}}{D}G_1' + \frac{D_{13}D_{32}}{D}G_2' + \frac{D_{12}D_{23}}{D}G_3'$$
$$F_2' = \frac{D_{23}D_{31}}{D}G_1' - \frac{D_{13}D_{31}}{D}G_2' + \frac{D_{13}D_{21}}{D}G_3'$$
$$F_3' = \frac{D_{21}D_{32}}{D}G_1' + \frac{D_{12}D_{31}}{D}G_2' - \frac{D_{12}D_{21}}{D}G_3'$$

where

$$D_{12} = \frac{\partial C}{\partial u_2}(1, F_2, F_3) - \frac{\partial C}{\partial u_2}(F_1, F_2, F_3), \quad D_{13} = \frac{\partial C}{\partial u_3}(1, F_2, F_3) - \frac{\partial C}{\partial u_3}(F_1, F_2, F_3),$$
$$D_{21} = \frac{\partial C}{\partial u_1}(F_1, 1, F_3) - \frac{\partial C}{\partial u_1}(F_1, F_2, F_3), \quad D_{23} = \frac{\partial C}{\partial u_3}(F_1, 1, F_3) - \frac{\partial C}{\partial u_3}(F_1, F_2, F_3),$$
$$D_{31} = \frac{\partial C}{\partial u_1}(F_1, F_2, 1) - \frac{\partial C}{\partial u_1}(F_1, F_2, F_3), \quad D_{32} = \frac{\partial C}{\partial u_2}(F_1, F_2, 1) - \frac{\partial C}{\partial u_2}(F_1, F_2, F_3),$$

$$D = D_{12}D_{23}D_{31} + D_{13}D_{21}D_{32}.$$

However, often this form however is not convenient. As I explained for the case of independent private values, that is when $C(u_1, u_2, u_3) = u_1u_2u_3$, the system in this form is difficult to handle. Moreover, I argued that this system does not allow us to derive conditions on observables sufficient for point identification because it always has a non-positive solution as well as a non-negative one. Similar problems are likely to arise for other copulas too.

What I have done in the case of independence, I introduced new functions H_1, H_2 and H_3 that had a one-to-one relationship with functions F_1, F_2 and F_3. Then I proved that H_i are determined uniquely and this implied the uniqueness of F_i. A similar strategy can be used to show identification for Archimedean copulas.

Archimedean copulas

Introduce the following functions:

$$\Sigma_1(t) = C(1, F_2(t), F_3(t)), \quad \Sigma_2(t) = C(F_1(t), 1, F_3(t)), \quad \Sigma_3(t) = C(F_1(t), F_2(t), 1),$$

$$\Sigma(t) = C(F_1(t), F_2(t), F_3(t)).$$

Then

$$\begin{aligned}
G_1' &= \Sigma_1' - F_2'\frac{\partial C}{\partial u_2}(F_1, F_2, F_3) - F_3'\frac{\partial C}{\partial u_3}(F_1, F_2, F_3) \\
G_2' &= \Sigma_2' - F_1'\frac{\partial C}{\partial u_1}(F_1, F_2, F_3) - F_3'\frac{\partial C}{\partial u_3}(F_1, F_2, F_3) \\
G_3' &= \Sigma_3' - F_1'\frac{\partial C}{\partial u_1}(F_1, F_2, F_3) - F_2'\frac{\partial C}{\partial u_2}(F_1, F_2, F_3).
\end{aligned}$$

I want to express $F_2'\frac{\partial C}{\partial u_2}(F_1, F_2, F_3) + F_3'\frac{\partial C}{\partial u_3}(F_1, F_2, F_3)$, $F_1'\frac{\partial C}{\partial u_1}(F_1, F_2, F_3) + F_3'\frac{\partial C}{\partial u_3}(F_1, F_2, F_3)$ and $F_1'\frac{\partial C}{\partial u_1}(F_1, F_2, F_3) + F_2'\frac{\partial C}{\partial u_2}(F_1, F_2, F_3)$ through functions Σ_i and their derivatives Σ_i'.

Because Archimedean copulas have the form

$$C(u_1, u_2, u_3) = \psi^{-1}\left(\psi(u_1) + \psi(u_2) + \psi(u_3)\right),$$

then

$$\Sigma(t) = \psi^{-1}\left(\psi(F_1(t)) + \psi(F_2(t)) + \psi(F_3(t))\right)$$

$$\Sigma_1(t) = \psi^{-1}\left(\psi(F_2(t)) + \psi(F_3(t))\right)$$

$$\Sigma_2(t) = \psi^{-1}\left(\psi(F_1(t)) + \psi(F_3(t))\right)$$

$$\Sigma_3(t) = \psi^{-1}\left(\psi(F_1(t)) + \psi(F_2(t))\right)$$

and, consequently,

$$\Sigma_1'(t) = \frac{\psi'(F_2(t))F_2'(t) + \psi'(F_3(t))F_3'(t)}{\psi'(\Sigma_1(t))}$$
$$\Sigma_2'(t) = \frac{\psi'(F_1(t))F_1'(t) + \psi'(F_3(t))F_3'(t)}{\psi'(\Sigma_2(t))}$$
$$\Sigma_3'(t) = \frac{\psi'(F_1(t))F_1'(t) + \psi'(F_2(t))F_2'(t)}{\psi'(\Sigma_3(t))}.$$

Now notice that

$$F_2'\frac{\partial C}{\partial u_2}(F_1, F_2, F_3) + F_3'\frac{\partial C}{\partial u_3}(F_1, F_2, F_3) = \frac{\psi'(F_2(t))F_2'(t) + \psi'(F_3(t))F_3'(t)}{\psi'(\Sigma(t))} = \Sigma_1'(t)\frac{\psi'(\Sigma_1(t))}{\psi'(\Sigma(t))}$$
$$F_1'\frac{\partial C}{\partial u_1}(F_1, F_2, F_3) + F_3'\frac{\partial C}{\partial u_3}(F_1, F_2, F_3) = \frac{\psi'(F_1(t))F_1'(t) + \psi'(F_3(t))F_3'(t)}{\psi'(\Sigma(t))} = \Sigma_2'(t)\frac{\psi'(\Sigma_2(t))}{\psi'(\Sigma(t))}$$
$$F_1'\frac{\partial C}{\partial u_1}(F_1, F_2, F_3) + F_2'\frac{\partial C}{\partial u_2}(F_1, F_2, F_3) = \frac{\psi'(F_1(t))F_1'(t) + \psi'(F_2(t))F_2'(t)}{\psi'(\Sigma(t))} = \Sigma_3'(t)\frac{\psi'(\Sigma_3(t))}{\psi'(\Sigma(t))}.$$

The final step is to express Σ through Σ_1, Σ_2 and Σ_3:

$$\Sigma(t) = \psi^{-1}\left(0.5\psi(\Sigma_1(t)) + 0.5\psi(\Sigma_2(t)) + 0.5\psi(\Sigma_3(t))\right).$$

To summarize, the system of differential equations can be written in terms of unknown functions Σ_1, Σ_2, Σ_3:

$$\begin{aligned}
\Sigma_1' &= \frac{g_1}{1 - \frac{\psi'(\Sigma_1(t))}{\psi'(\psi^{-1}(0.5\psi(\Sigma_1(t))+0.5\psi(\Sigma_2(t))+0.5\psi(\Sigma_3(t))))}} \\
\Sigma_2' &= \frac{g_2}{1 - \frac{\psi'(\Sigma_2(t))}{\psi'(\psi^{-1}(0.5\psi(\Sigma_1(t))+0.5\psi(\Sigma_2(t))+0.5\psi(\Sigma_3(t))))}} \\
\Sigma_3' &= \frac{g_3}{1 - \frac{\psi'(\Sigma_3(t))}{\psi'(\psi^{-1}(0.5\psi(\Sigma_1(t))+0.5\psi(\Sigma_2(t))+0.5\psi(\Sigma_3(t))))}}.
\end{aligned} \tag{3.3.3}$$

There is a one-to-one relationship between function Σ_i and F_i because

$$F_1 = \psi^{-1}(0.5\psi(\Sigma_2) + 0.5\psi(\Sigma_3) - 0.5\psi(\Sigma_1))$$

$$F_2 = \psi^{-1}(0.5\psi(\Sigma_1) + 0.5\psi(\Sigma_3) - 0.5\psi(\Sigma_2))$$

$$F_3 = \psi^{-1}(0.5\psi(\Sigma_1) + 0.5\psi(\Sigma_2) - 0.5\psi(\Sigma_3)).$$

These formulas take into account the fact that F_i are non-negative functions. For d bidders, system (3.3.3) has the following form:

$$\Sigma_i' = \frac{g_i}{1 - \frac{\psi'(\Sigma_i(t))}{\psi'(\psi^{-1}\left(\frac{1}{d-1}\sum_{j=1}^{d}\psi(\Sigma_j(t))\right)}}, \quad i = 1, \ldots, d.$$

This system is quite difficult to analyze. I consider the case when there are only two types of bidders and present a condition sufficient to guarantee identification in this situation.

Suppose that the first k bidders are of type I and the rest $d-k$ bidders are of type II. Let F_1 be the marginal distribution function of a type I bidder and F_2 be the marginal distribution function of a type II bidder. Denote

$$G_1(t) = P(\text{bidder of type I wins, price } \le t),$$

$$G_2(t) = P(\text{bidder of type II wins, price } \le t),$$

$$\Sigma_1 = C(1, \underbrace{F_1, \ldots, F_1}_{k-1}, \underbrace{F_2, \ldots, F_2}_{d-k}) = \psi^{-1}\left((k-1)\psi(F_1) + (d-k)\psi(F_2)\right),$$
$$\Sigma_2 = C(\underbrace{F_1, \ldots, F_1}_{k}, \underbrace{F_2, \ldots, F_2}_{d-k-1}, 1) = \psi^{-1}\left(k\psi(F_1) + (d-k-1)\psi(F_2)\right).$$

The system of differential equations that determines Σ_1 and Σ_2 is

$$\Sigma_1' = \frac{g_1}{1 - \frac{\psi'(\Sigma_1)}{\psi'(\psi^{-1}(\frac{k}{d-1}\psi(\Sigma_1)+\frac{d-k}{d-1}\psi(\Sigma_2)))}} \tag{3.3.4}$$
$$\Sigma_2' = \frac{g_2}{1 - \frac{\psi'(\Sigma_2)}{\psi'(\psi^{-1}(\frac{k}{d-1}\psi(\Sigma_1)+\frac{d-k}{d-1}\psi(\Sigma_2)))}}.$$

I analyze it together with initial conditions

$$\Sigma_1(t_0) = \Sigma_2(t_0) = 0. \tag{3.3.5}$$

Theorem 3.11. *If function $\frac{\psi''(x)}{(\psi'(x))^2}$ is increasing, then problem (3.3.4)-(3.3.5) can have only one solution on $[t_0, T]$.*

The proof of this theorem uses the same logic as the proof of theorem 3.5.

3.4 Conclusion

This chapter considered inference in auctions when bidders' values are not independent. I showed that even though the joint distribution of values is not point identified, data can help to find bounds on the distribution. Moreover, I explained how these bounds change when more data become available. In addition, I demonstrated that the joint distribution may be point-identified if the copula that represents this distribution through marginal distributions is known.

3.5 Appendix

3.5.1 Bounds on distributions: Proof of results

Proof of Theorem 3.1.

(a) First, I prove the result for the lower bound. Suppose that $\max_{k=1,\dots,r} t_{i_k} < T$. Then

$$Q_D(t_{i_1},\dots,t_{i_r}) = P(\cap_{k=1}^r (X_{i_k} \le t_{i_k})) = P(\cap_{k=1}^r (b_{i_k} \le t_{i_k})) =$$

$$= \sum_{m=1}^r P(\cap_{k=1}^r (b_{i_k} \le t_{i_k}), i_m \text{ wins}) + \sum_{j\in CD} P(\cap_{k=1}^r (b_{i_k} \le t_{i_k}), j \text{ wins}) \ge$$

$$\ge \sum_{j\in CD} P(\text{price } \le \min_{k=1,\dots,r} t_{i_k}, j \text{ wins}) = \sum_{j\in CD} G_j(\min_{k=1,\dots,r} t_{i_k}).$$

Now consider the case when at least one of t_{i_k} takes value T. It is enough to consider the case when $t_{i_1} = T$ and $\max_{k=2,\dots,r} t_{i_k} < T$. Denote $\tilde{D} = \{i_2,\dots,i_r\}$. From what I have shown above, it follows that

$$Q_D(T,\dots,t_{i_r}) = Q_{\tilde{D}}(t_{i_2},\dots,t_{i_r}) \ge \sum_{j\in C\tilde{D}} G_j(\min_{k=2,\dots,r} t_{i_k}) = G_{i_1}(\min_{k=2,\dots,r} t_{i_k}) +$$

$$+ \sum_{j\in CD} G_j(\min_{k=2,\dots,r} t_{i_k}) = G_{i_1}(\min_{k=1,\dots,r} t_{i_k}) + \sum_{j\in CD} G_j(\min_{k=1,\dots,r} t_{i_k}) =$$

$$= \sum_{l=1}^r 1(t_{i_l} = T) G_{i_l}(\min_{k=1,\dots,r} t_{i_k}) + \sum_{j\in CD} G_j(\min_{k=1,\dots,r} t_{i_k}).$$

(b) For any $D = \{t_{i_1},\dots,t_{i_r}\}$, if $\min_{k=1,\dots,r} t_{i_k} = t_0$ then $Q_D(t_{i_1},\dots,t_{i_r}) = 0$ because by assumption all marginal distributions X_i are continuous and, therefore, do not have mass points. Below I suppose that $\min_{k=1,\dots,r} t_{i_k} > t_0$.

For any $m = 1, \ldots, d$,

$$Q_{-m}(t_1, \ldots, t_{m-1}, t_{m+1}, \ldots, t_d) = P(\cap_{i \neq m}(X_i \leq t_i)) = P(\cap_{i \neq m}(b_i \leq t_i)) =$$

$$= \sum_{j \neq m} P(\cap_{i \neq m}(b_i \leq t_i), j \text{ wins}) + P(\cap_{i \neq m}(b_i \leq t_i), m \text{ wins}) \leq$$

$$\leq \sum_{j \neq m} P(\text{price } \leq t_j, j \text{ wins}) + P(\text{price } \leq \max_{i \neq m} t_i, m \text{ wins}) = \sum_{j \neq m} G_j(t_j) + G_m(\max_{i \neq m} t_i).$$

Let $D = \{1, \ldots, d\}$.

$$Q(t_1, \ldots, t_d) = P(\cap_{i=1}^{d}(X_i \leq t_i)) = P(\cap_{i=1}^{d}(b_i \leq t_i)) =$$

$$= \sum_j P(\cap_{i \neq j}(b_i \leq t_i), b_j \leq t_j, j \text{ wins}) \leq \sum_j P(\text{price } \leq \min\{\max_{i \neq j} t_i, t_j\}, j \text{ wins}) =$$

$$= \sum_j G_j(\min\{\max_{i \neq j} t_i, t_j\}).$$

Let $D = \{t_{i_1}, \ldots, t_{i_r}\}$ contain at most $d-2$ elements.

$$Q_D(t_{i_1}, \ldots, t_{i_r}) = P(\cap_{k=1}^{r}(X_{i_k} \leq t_{i_k})) = P(\cap_{k=1}^{r}(b_{i_k} \leq t_{i_k})) =$$

$$= \sum_{m=1}^{r} P(\cap_{k=1}^{r}(b_{i_k} \leq t_{i_k}), i_m \text{ wins}) + \sum_{j \in CD} P(\cap_{k=1}^{r}(b_{i_k} \leq t_{i_k}), j \text{ wins}) \leq$$

$$\leq \sum_{m=1}^{r} P(\text{price } \leq t_{i_m}, i_m \text{ wins}) + \sum_{j \in CD} P(j \text{ wins}) = \sum_{i_m \in D} G_{i_m}(t_{i_m}) + \sum_{j \in CD} G_j(T).$$

Proof of Proposition 3.2 .

(a) According to A1, for any $D = \{t_{i_1}, \ldots, t_{i_r}\}$, the event $\{\cap_{k=1}^{r}(X_{i_k} \leq t_{i_k})\}$ implies the event $\{\cap_{k=1}^{r}(b_{i_k} \leq t_{i_k})\}$. Therefore,

$$Q_D(t_{i_1}, \ldots, t_{i_r}) = P(\cap_{k=1}^{r}(X_{i_k} \leq t_{i_k})) \leq P(\cap_{k=1}^{r}(b_{i_k} \leq t_{i_k})).$$

The rest of the proof for the upper bounds is the same as in theorem 3.1.

(b) Suppose that $\max_{k=1,\ldots,r} t_{i_k} < T$. Then

$$Q_D(t_{i_1},\ldots,t_{i_r}) = P(\cap_{k=1}^r (X_{i_k} \le t_{i_k})) = \sum_{k=1}^r P(\cap_{k=1}^r (X_{i_k} \le t_{i_k}), i_k \text{ wins})+$$

$$+\sum_{j\in CD} P(\cap_{k=1}^r (X_{i_k} \le t_{i_k}), j \text{ wins}) \ge \sum_{j\in CD} P(\cap_{k=1}^r (X_{i_k} \le t_{i_k}), j \text{ wins}).$$

According to A2, for $j \in CD$, the event $\{\text{price } \le \min_{k=1,\ldots,r} t_{i_k}, j \text{ wins}\}$ implies the event $\{\cap_{k=1}^r (X_{i_k} \le t_{i_k}), j \text{ wins}\}$. Indeed, if some X_{i_k} was larger than t_{i_k}, then bidder i_k would not allow bidder j to win at a price less or equal than $\min_{k=1,\ldots,r} t_{i_k}$. Therefore,

$$Q_D(t_{i_1},\ldots,t_{i_r}) \ge \sum_{j\in CD} P(\text{price } \le \min_{k=1,\ldots,r} t_{i_k}, j \text{ wins}).$$

The rest of the proof for the lower bounds is the same as in theorem 3.1.

Proof of Theorem 3.3 .

(a) The lower bound is obvious. Let me obtain the upper bound.

$$Q(t_1,\ldots,t_d) = \sum_{\rho\in\Pi_d} P(\cap_{i=1,\ldots d}(b_i \le t_i), b_{\rho(1)} > b_{\rho(2)} > \ldots > b_{\rho(d)}) \le$$

$$\le \sum_{\rho\in\Pi_d} P(\cap_{l=2,\ldots,d}(b_{\rho(l)} \le \min_{m=1,\ldots,l} t_{\rho(m)}), b_{\rho(1)} > b_{\rho(2)} > \ldots > b_{\rho(d)}) =$$

$$= \sum_{\rho\in\Pi_d} W_\rho(\min\{t_{\rho(1)}, t_{\rho(2)}\},\ldots,\min_{m=1,\ldots,l} t_{\rho(m)},\ldots,\min_{i=1,\ldots,d} t_i).$$

(b) Without a loss of generality, consider function F_1.

$$F_1(t) = P(X_1 \le t) = P(b_1 \le t) = \sum_i \sum_{h\ne i} P(\max_{l\ne h, l\ne i} b_l < b_h, b_h < b_i, b_1 \le t) =$$

$$= \sum_{h\ne 1} P(\max_{l\ne h, l\ne 1} b_l < b_h, b_h < b_1, b_1 \le t) + \sum_{i\ne 1} P(\max_{l\ne 1, l\ne i} b_l < b_1, b_1 < b_i, b_1 \le t)+$$

$$+\sum_{i\ne 1}\sum_{h\ne i, h\ne 1} P(\max_{l\ne h, l\ne i} b_l < b_h, b_h < b_i, b_1 \le t).$$

Let $t \in (t_0, T]$. For the upper bound,

$$P(b_1 \leq t) \leq \sum_{h \neq 1} P(\max_{l \neq h, l \neq 1} b_l < b_h, b_h < b_1, b_h \leq t) + \sum_{i \neq 1} P(\max_{l \neq 1, l \neq i} b_l < b_1, b_1 < b_i, b_1 \leq t) +$$

$$+ \sum_{i \neq 1} \sum_{h \neq i, h \neq 1} P(\max_{l \neq h, l \neq i} b_l < b_h, b_h < b_i) = \sum_{h \neq 1} G_{1h}(t) + \sum_{i \neq 1} G_{i1}(t) + \sum_{i \neq 1} \sum_{h \neq i, h \neq 1} G_{ih}(T).$$

Let $t \in [t_0, T)$. For the lower bound,

$$P(b_1 \leq t) \geq 0 + \sum_{i \neq 1} P(\max_{l \neq 1, l \neq i} b_l < b_1, b_1 < b_i, b_1 \leq t) +$$

$$+ \sum_{i \neq 1} \sum_{h \neq i, j \neq 1} P(\max_{l \neq h, l \neq i} b_l < b_h, b_h < b_i, b_h \leq t) = \sum_{i \neq 1} G_{i1}(t) + \sum_{i \neq 1} \sum_{h \neq i, h \neq 1} G_{ih}(t) =$$

$$= \sum_{i \neq 1} \left(G_{i1}(t) + \sum_{h \neq i, h \neq 1} G_{ih}(t) \right) = \sum_{i \neq 1} \sum_{h \neq i} G_{ih}(t) = \sum_{i \neq 1} G_i(t).$$

Evidently, $F_1(t_0) = 0$ and $F_1(T) = 1$.

As we can see, the upper bound on F_1 is tighter than the upper bound on F_1 from theorem 3.1. The lower bound does not change.

3.5.2 Copulas: Proof of results

Before I prove results of section 3.3, I formulate two lemmas that help me to establish these results. The proof of these lemmas can be found in Komarova (2008).

Lemma 3.12. *Let $z : [\tau, \xi] \to \Re^n$ be an absolutely continuous function. Then $\|z\|_1$ has the right derivative $D_R\|z\|_1$ a.e. on $[\tau, \xi]$, and*

$$D_R\|z(t)\|_1 \leq \|z'(t)\|_1 \quad a.e. \text{ on } [\tau, \xi].$$

Lemma 3.13. *Let function $v : [\tau, \xi] \to \Re$ be absolutely continuous. Suppose that $v(\tau) = 0$, and a.e. on $[\tau, \xi]$*

$$D_R v(t) \leq \Gamma(t)v(t), \quad \textit{where} \quad \Gamma \in L^1[\tau, \xi].$$

Then

$$v(t) \leq 0, \quad t \in [\tau, \xi].$$

Proof of Theorem 3.5. The most important step is to show that (DE_1)-(IC_1) has only one solution in a neighborhood of t_0. Then similar to the case of independent values, it can be shown that this local solution will have the unique extension along the whole support $[t_0, T]$.

Suppose that (DE_1)-(IC_1) has two local solutions (F_1, F_2) and $(\tilde{F}_1, \tilde{F}_2)$ with a common interval of existence $[t_0, t_0 + a]$. Let $F = (F_1, F_2)^{tr}$, $\tilde{F} = (\tilde{F}_1, \tilde{F}_2)^{tr}$ The idea of the proof is to establish that

$$(F - \tilde{F}, F' - \tilde{F}') \leq 0 \text{ a.e. on } [t_0, t_0 + a]. \tag{3.5.1}$$

Inequality (3.5.1) implies that

$$\frac{d\|F - \tilde{F}\|^2}{dt} \leq 0 \text{ a.e. on } [t_0, t_0 + a],$$

where $\|F - \tilde{F}\|$ stand for the Euclidean norm of $F - \tilde{F}$, and therefore, F and $\tilde{F}$ coincide on $[t_0, t_0 + a]$.

Let me show that (3.5.1) holds. First of all, notice that

$$F_1 + F_2 - C(F_1, F_2) = G_1 + G_2,$$

$$\tilde{F}_1 + \tilde{F}_2 - C(\tilde{F}_1, \tilde{F}_2) = G_1 + G_2.$$

Hence,

$$F_1 - \tilde{F}_1 = \tilde{F}_2 - F_2 + C(F_1, F_2) - C(\tilde{F}_1, \tilde{F}_2) = \tilde{F}_2 - F_2 + C(F_1, F_2) - C(\tilde{F}_1, F_2) +$$

$$+ C(\tilde{F}_1, F_2) - C(\tilde{F}_1, \tilde{F}_2) = \tilde{F}_2 - F_2 + \frac{\partial C(F_1^*, F_2)}{\partial u_1}(F_1 - \tilde{F}_1) + \frac{\partial C(\tilde{F}_1, F_2^*)}{\partial u_2}(F_2 - \tilde{F}_2),$$

where $F_1^* = \alpha F_1 + (1-\alpha)\tilde{F}_1$ for some $\alpha = \alpha(F_1, \tilde{F}_1, F_2) \in [0,1]$, and $F_2^* = \beta F_2 + (1-\beta)\tilde{F}_2$ for some $\beta = \beta(\tilde{F}_1, F_2, \tilde{F}_2) \in [0,1]$. Rewrite the last equation as

$$\left(1 - \frac{\partial C(F_1^*, F_2)}{\partial u_1}\right)(F_1 - \tilde{F}_1) = \left(1 - \frac{\partial C(\tilde{F}_1, F_2^*)}{\partial u_2}\right)(\tilde{F}_2 - F_2).$$

Because expressions before $F_1 - \tilde{F}_1$ and $\tilde{F}_2 - F_2$ are non-negative, then

$$F_1 \geq \tilde{F}_1 \text{ iff } F_2 \leq \tilde{F}_2.$$

Now consider $(F_1 - \tilde{F}_1)(F_1' - \tilde{F}_1')$.

$$(F_1 - \tilde{F}_1)(F_1' - \tilde{F}_1') =$$

$$= \frac{g_2}{\left(1 - \frac{\partial C(F_1, F_2)}{\partial u_1}\right)\left(1 - \frac{\partial C(\tilde{F}_1, \tilde{F}_2)}{\partial u_1}\right)}(F_1 - \tilde{F}_1)\left(\frac{\partial C(F_1, F_2)}{\partial u_1} - \frac{\partial C(\tilde{F}_1, \tilde{F}_2)}{\partial u_1}\right).$$

Without a loss of generality suppose that $F_1 \geq \tilde{F}_1$. Then $F_2 \leq \tilde{F}_2$. From the definition of copulas, it follows that $\frac{\partial C(u_1, u_2)}{\partial u_1}$ is increasing in u_2. Therefore,

$$\frac{\partial C(F_1, F_2)}{\partial u_1} \leq \frac{\partial C(F_1, \tilde{F}_2)}{\partial u_1}.$$

From the condition of the theorem,

$$\frac{\partial C(F_1, \tilde{F}_2)}{\partial u_1} \leq \frac{\partial C(\tilde{F}_1, \tilde{F}_2)}{\partial u_1},$$

therefore,

$$(F_1 - \tilde{F}_1)(F_1' - \tilde{F}_1') \leq 0.$$

Similarly,

$$(F_2 - \tilde{F}_2)(F_2' - \tilde{F}_2') \leq 0.$$

To summarize,

$$(F - \tilde{F})(F' - \tilde{F}') \leq 0.$$

Therefore, (3.5.1) is established and, hence, the theorem is proved.

Proof of Theorem 3.6 . Suppose that (DE_1)-(IC_1) has two local solutions $F = (F_1, F_2)$ and $\tilde{F} = (\tilde{F}_1, \tilde{F}_2)$ with a common interval of existence $[t_0, t_0 + a]$. Consider $F_1' - \tilde{F}_1'$.

$$F_1' - \tilde{F}_1' = \frac{g_2}{\left(1 - \frac{\partial C(F_1,F_2)}{\partial u_1}\right)\left(1 - \frac{\partial C(\tilde{F}_1,\tilde{F}_2)}{\partial u_1}\right)}\left(\frac{\partial C(F_1, F_2)}{\partial u_1} - \frac{\partial C(\tilde{F}_1, \tilde{F}_2)}{\partial u_1}\right) =$$

$$= \frac{g_2}{\left(1 - \frac{\partial C(F_1,F_2)}{\partial u_1}\right)\left(1 - \frac{\partial C(\tilde{F}_1,\tilde{F}_2)}{\partial u_1}\right)}\left(\frac{\partial C(F_1, F_2)}{\partial u_1} - \frac{\partial C(\tilde{F}_1, F_2)}{\partial u_1}\right) +$$

$$+ \frac{g_2}{\left(1 - \frac{\partial C(F_1,F_2)}{\partial u_1}\right)\left(1 - \frac{\partial C(\tilde{F}_1,\tilde{F}_2)}{\partial u_1}\right)}\left(\frac{\partial C(\tilde{F}_1, F_2)}{\partial u_1} - \frac{\partial C(\tilde{F}_1, \tilde{F}_2)}{\partial u_1}\right) =$$

$$= \frac{g_2}{\left(1 - \frac{\partial C(F_1,F_2)}{\partial u_1}\right)\left(1 - \frac{\partial C(\tilde{F}_1,\tilde{F}_2)}{\partial u_1}\right)}\left(\frac{\partial^2 C(F_1^{**}, F_2)}{\partial u_1^2}(F_1 - \tilde{F}_1) + \frac{\partial^2 C(\tilde{F}_1, F_2^{**})}{\partial u_1 \partial u_2}(F_2 - \tilde{F}_2)\right),$$

where $F_1^{**} = \alpha^* F_1 + (1 - \alpha^*)\tilde{F}_1$, $\alpha^* = \alpha^*(F_1, \tilde{F}_1, F_2) \in [0, 1]$, and $F_2^{**} = \beta^* F_2 + (1 - \beta^*)\tilde{F}_2$, $\beta = \beta(\tilde{F}_1, F_2, \tilde{F}_2) \in [0, 1]$. According to the conditions of the theorem, for some constant $K_1 > 0$,

$$|F_1' - \tilde{F}_1'| \leq K_1 g_2(|F_1 - \tilde{F}_1| + |F_2 - \tilde{F}_2|).$$

Similarly, for some constant $K_2 > 0$,

$$|F_2' - \tilde{F}_2'| \leq K_2 g_1(|F_1 - \tilde{F}_1| + |F_2 - \tilde{F}_2|).$$

To summarize,

$$\|F' - \tilde{F}'\|_1 \leq (K_1 + K_2)(g_1 + g_2)\|F - \tilde{F}\|_1.$$

According to lemmas 3.12 and 3.13, the last inequality and initial conditions (IC_1) imply that F and $\tilde{F}$ coincide on $[t_0, t_0 + a]$.

Proof of Proposition 3.9.

$$\frac{\partial C(u_1, u_2)}{\partial u_1} = \frac{\psi'(u_1)}{\psi'(\psi^{-1}(\psi(u_1) + \psi(u_2)))},$$

$$\frac{\partial^2 C(u_1, u_2)}{\partial u_1^2} = \frac{\psi''(u_1)(\psi'(\psi^{-1}(\psi(u_1) + \psi(u_2))))^2 - (\psi'(u_1))^2 \psi''(\psi^{-1}(\psi(u_1) + \psi(u_2)))}{(\psi'(\psi^{-1}(\psi(u_1) + \psi(u_2))))^3}.$$

Notice that $\frac{\partial^2 C(u_1,u_2)}{\partial u_1^2} \leq 0$ iff the numerator of the last ratio is non-negative, or equivalently,

$$\frac{\psi''(u_1)}{(\psi'(u_1))^2} \geq \frac{\psi''(\psi^{-1}(\psi(u_1) + \psi(u_2)))}{(\psi'(\psi^{-1}(\psi(u_1) + \psi(u_2))))^2}. \tag{3.5.2}$$

Because $u_1 \geq \psi^{-1}(\psi(u_1) + \psi(u_2))$ and function $\frac{\psi''(x)}{(\psi'(x))^2}$ is increasing, then inequality (3.5.2) holds, and therefore, the conditions in theorem 3.5 are satisfied.

Proof of Theorem 3.11. It is enough to show that problem (3.3.4)-(3.3.5) can have only one solution in a neighborhood of t_0. Then the extension of this solution along the whole support will be unique too.

System (3.3.4) implies that

$$k\Sigma_1 + (d-k)\Sigma_2 - (d-1)\psi^{-1}\left(\frac{k}{d-1}\psi(\Sigma_1) + \frac{d-k}{d-1}\psi(\Sigma_2)\right) = kG_1 + (d-k)G_2.$$

Suppose that problem (3.3.4)-(3.3.5) has two solutions (Σ_1, Σ_2) and $(\tilde{\Sigma}_1, \tilde{\Sigma}_2)$ with a common interval of existence $[t_0, t_0 + a]$. I want to show that for any $t \in [t_0, t_0 + a]$, $\Sigma_1(t) \geq \tilde{\Sigma}_1(t)$ iff $\Sigma_2(t) \leq \tilde{\Sigma}_2(t)$. Fix $(t_0, t_0 + a]$ From equation

$$k\Sigma_1 + (d-k)\Sigma_2 - (d-1)\psi^{-1}\left(\frac{k}{d-1}\psi(\Sigma_1) + \frac{d-k}{d-1}\psi(\Sigma_2)\right) =$$
$$= k\tilde{\Sigma}_1 + (d-k)\tilde{\Sigma}_2 - (d-1)\psi^{-1}\left(\frac{k}{d-1}\psi(\tilde{\Sigma}_1) + \frac{d-k}{d-1}\psi(\tilde{\Sigma}_2)\right)$$

obtain that

$$\begin{aligned} &k\left(1 - \frac{\psi'(\Sigma_1^*)}{\psi'\left(\psi^{-1}\left(\frac{k}{d-1}\psi(\Sigma_1^*) + \frac{d-k}{d-1}\psi(\Sigma_2)\right)\right)}\right)(\Sigma_1 - \tilde{\Sigma}_1) = \\ &= (d-k)\left(1 - \frac{\psi'(\Sigma_2^*)}{\psi'\left(\psi^{-1}\left(\frac{k}{d-1}\psi(\tilde{\Sigma}_1) + \frac{d-k}{d-1}\psi(\Sigma_2^*)\right)\right)}\right)(\tilde{\Sigma}_2 - \Sigma_2), \end{aligned} \tag{3.5.3}$$

where $\Sigma_1^* = \alpha\Sigma_1 + (1-\alpha)\tilde{\Sigma}_1$ for some $\alpha = \alpha(\Sigma_1(t), \tilde{\Sigma}_1(t), \Sigma_2(t)) \in [0,1]$, and $\Sigma_2^* = \beta\Sigma_2 + (1-\beta)\tilde{\Sigma}_2$ for some $\beta = \beta(\tilde{\Sigma}_1(t), \Sigma_2(t), \tilde{\Sigma}_2(t)) \in [0,1]$. Note that for $t < T$,

$$\frac{\psi'(\Sigma_1)}{\psi'(\psi^{-1}(\frac{k}{d-1}\psi(\Sigma_1) + \frac{d-k}{d-1}\psi(\Sigma_2)))} < 1, \quad \frac{\psi'(\tilde{\Sigma}_1)}{\psi'(\psi^{-1}(\frac{k}{d-1}\psi(\tilde{\Sigma}_1) + \frac{d-k}{d-1}.\psi(\tilde{\Sigma}_2)))} < 1.$$

Because $\frac{\tilde{\Sigma}_1}{\Sigma_1} \to 1$, $\frac{\tilde{\Sigma}_2}{\Sigma_2} \to 1$ as $t \to t_0$, then for t close enough to t_0,

$$\frac{\psi'(\Sigma_1^*)}{\psi'\left(\psi^{-1}\left(\frac{k}{d-1}\psi(\Sigma_1^*) + \frac{d-k}{d-1}\psi(\Sigma_2)\right)\right)} < 1, \quad \frac{\psi'(\Sigma_2^*)}{\psi'\left(\psi^{-1}\left(\frac{k}{d-1}\psi(\tilde{\Sigma}_1) + \frac{d-k}{d-1}\psi(\Sigma_2^*)\right)\right)} < 1.$$

Therefore, from (3.5.3) I obtain that $\Sigma_1(t) \geq \tilde{\Sigma}_1(t)$ iff $\Sigma_2(t) \leq \tilde{\Sigma}_2(t)$. This fact and the fact that function $\frac{\psi''(x)}{(\psi'(x))^2}$ is increasing allow me to show that

$$(\Sigma_1' - \tilde{\Sigma}_1', \Sigma_1 - \tilde{\Sigma}_1) + (\Sigma_2' - \tilde{\Sigma}_2', \Sigma_2 - \tilde{\Sigma}_2) \leq 0 \quad a.e.\ [t_0, t_0 + a],$$

or, equivalently,

$$\frac{1}{2}\frac{d}{dt}\|(\Sigma_1, \Sigma_2) - (\tilde{\Sigma}_1, \tilde{\Sigma}_2)\|^2 \leq 0 \quad a.e.\ [t_0, t_0 + a].$$

The last inequality implies that (Σ_1, Σ_2) and $(\tilde{\Sigma}_1, \tilde{\Sigma}_2)$ coincide on $[t_0, t_0 + a]$.

References

Akhmerov, R.R., and B.N. Sadovskij (2002). *Ocherki po teorii obyknovennykh differentcial'nykh uravnenij*, Unpublished manuscript.

Athey, S., and P.A. Haile (2002). Identification of Standard Auction Models, *Econometrica*, 70 (6), 2107-2140.

Athey, S., and P. Haile (2005). Nonparametric Approaches to Auctions, forthcoming in *Handbook of Econometrics* , Vol. 6, eds J. Heckman and E. Leamer.

Athey, S., and P. Haile (2006). Empirical Models of Auctions, Unpublished Manuscript.

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

Banerji, A., and J.V. Meenakshi (2004a). Buyer Collusion and Efficiency of Government Intervention in Wheat Markets in Northern India: An Asymmetric Structural Auctions Analysis, *American Journal of Agricultural Economics*, 86 (1), 236-253.

Barlow, R.E., and F. Proschan (1975). *Statistical Theory of Reliability and Life Testing: Probability Models*. New York: Holt, Rinehart and Winston.

Bedford, T., and I. Meilijson (1997). A Characterization of Marginal Distributions of (Possibly Dependent) Lifetime Variables which Right Censor each other, *The Annals of Statistics*, 25 (4), 1622-1645.

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.

Brendstrup, B., and H. Paarsch (2003). Nonparametric Estimation of Dutch and First-Price, Sealed-Bid Auction Models with Asymmetric Bidders, Working Paper, University of Iowa.

Brendstrup, B., and H. Paarsch (2004). Nonparametric Identification and Estimation of Multi- Unit, Sequential, Oral Ascending-Price Auctions with Asymmetric Bidders, Working Paper, University of Iowa.

Brendstrup, B., and H.J. Paarsch (2006). Identification and Estimation in Sequential, Asymmetric, English Auctions, *Journal of Econometrics*, 134, 69-94.

Buera, F. (2006). Non-parametric Identification and Testable Implications of the Roy Model, Working Paper, Northwestern University.

Campo, S., I. Perrigne, and Q. Vuong (2003), Asymmetry in First-Price Auctions with Affiliated Private Values, *Journal of Applied Econometrics*, 18, 179207.

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

Chen Chen, X. (2006). Large Sample Sieve Estimation of Semi-Nonparametric Models, forthcoming in *Handbook of Econometrics* , Vol. 6, eds J. Heckman and E. Leamer.

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.

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.

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

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

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

Crowder, M. (1991). On the Identifiability Crises in Competing Risks Analysis, *Scand. J. Statis.*, 18, 223-233.

Deshpande, J.V., and S.R. Karia (1997). Bounds for the Joint Survival and Incidence Functions through Coherent System Data, *Advances in Applied Probability*, 29 (2), 478-497.

Filippov, A.F. (1988). *Differential Equations with Discontinuous Righthand Sides*. Dordrecht: Kluwer Academic Publishers.

Guerre, E., I. Perrigne, and Q. Vuong (2007). Nonparametric Identification of Risk Aversion in First-Price Auctions Under Exclusion Restrictions, Unpublished Manuscript.

Haile, P. and E. Tamer (2003). Inference with an Incomplete Model of English Auctions, *Journal of Political Economy*, 111, 152.

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

Harstad, R.M., J.H. Kagel, and D. Levin (1990). Equilibrium Bid Functions for Auctions with an Uncertain Number of Bidders, *Economics Letters*, 33, 35-40.

Hartman, P. (1964). *Ordinary Differential Equations.* New York: John Wiley & Sons, Inc.

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

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

Horowitz, J.L. and E. Mammen (2005). Nonparametric Estimation of an Additive Model with a Link Function, *Annals of Statistics*, 32, 2412-2443.

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

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

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

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

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

Komarova Komarova, T. (2007). Nonparametric Identification in Asymmetric Second-Price Auctions: A New Approach, Working Paper, Northwestern University.

Krishna, V. (2002). *Auction Theory.* San Diego: Academic Press.

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

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

Lebrun Lebrun, B. (1999). First-Price Auctions and the Asymmetric N Bidders Case, *International Economic Review*, 40, 125-142.

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

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

McAfee, R.P., and J. McMillan (1987). Auctions with a Stochastic Number of Bidders, *Journal of Economic Theory*, 43, 1-19.

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

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

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

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

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

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

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

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

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

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

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

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.

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.

Meenakshi, J.V., and A. Banerji (2005). The Unsupportable Support Price: An Analysis of Collusion and Government Intervention in Paddy Auction Markets in North India, *Journal of Development Economics*, 76, 377403.

Meilijson, I. (1981). Estimation of the Lifetime Distribution of the Parts from the Autopsy Statistics of the Machine, *Journal of Applied Probability*, 18, 829-838.

Milgrom, P., and R. Weber (2000). A Theory of Auctions and Competitive Bidding, *Econometrica*, 50, 1089-1122.

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.

Nelsen Nelsen, R.B. (1999). *An Introduction to copulas.* New York: Springer.

Newey, W.K., and J.L. Powell (2003). Instrumental Variable Estimation of Nonparametric Models, *Econometrica*, 71 (5), 1565-1578.

Nowik, S. (1990). Identifiability Problems in Coherent Systems, *Journal of Applied Probability*, 28, 862-872.

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

Peterson, A.V.Jr. (1976). Bounds for a Joint Distribution Function with Fixed Sub-Distribution Functions, *Proceedings of the National Academy of Sciences, USA*, 73, 11-13.

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.

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

Szarski, J. (1965). *Differential Inequalities.* Warszawa: Polish Scientific Publishers.

Tsiatis Tsiatis, A.A. (1975). A Nonidentifiability Aspect of the Problem of Competing Risks, *Proceedings of the National Academy of Sciences, USA*, 72, 20-22.

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

Vickrey Vickrey, W. (1961). Counterspeculation, Auctions, and Competitive Sealed Tenders, *Journal of Finance*, 16 (1), 8-37.