NONPARAMETRIC AND DISTRIBUTION-FREE ESTIMATION OF THE BINARY THRESHOLD CROSSING AND THE BINARY CHOICE MODELS

BY ROSA L. MATZKIN

In this paper, it is shown that it is possible to identify binary threshold crossing models and binary choice models without imposing any parametric structure either on the systematic function of observable exogenous variables or on the distribution of the random term. This identification result is employed to develop a fully nonparametric maximum likelihood estimator for both the function of observable exogenous variables and the distribution of the random term. The estimator is shown to be strongly consistent, and a two step procedure for its calculation is developed. The paper also includes examples of economic models that satisfy the conditions that are necessary to apply the results.

KEYWORDS: Nonparametric, distribution-free, binary threshold crossing model, binary choice model, concavity, homogeneity of degree one, monotonicity, identification, consistency.

1. INTRODUCTION

In recent years, there has been increasing interest in the study of binary threshold crossing models and binary choice models. These models have been employed to study a wide variety of problems in economics, biology, marketing, and other fields. The applications in economics have included labor force participation, choice of mode of transportation, and choice between education and labor force participation. The estimation of these models has been originally parametric and most recently semiparametric. This paper introduces a fully nonparametric estimation analysis.

In binary threshold crossing models, the dependent variable is a binary indicator, whose value depends on a function, $h^*$, of observable exogenous variables and an unobservable random term, $\eta$. In binary choice models, $h^*$ is the difference between two functions, $V_1^*$ and $V_2^*$, and $\eta$ is the difference between two unobservable random terms, $\xi_1$ and $\xi_2$.

In the past, the estimation of these models has proceeded by specifying parametric structures for the function $h^*$ (resp. $V_1^*$ and $V_2^*$) and for the distribution $F^*$ of $\eta$. Unlike the estimation of linear models, however, an erroneous specification of the distribution $F^*$ may cause the estimator of the parameters of the correctly specified function $h^*$ to be inconsistent. In pioneering work, Manski (1975, 1985) proved that it is possible to estimate the

---

1 The support of the National Science Foundation through Grants No. SES-8720596 and SES-8900291 is gratefully acknowledged. I am grateful to James Heckman, Charles Manski, Whitney Newey, and Vassilis Hajivassilou for suggestions and discussions. Comments by participants in seminars at Bonn, Brown, Columbia, Harvard/MIT, and Yale Universities, the 1988 European and Summer Meetings of the Econometric Society, and the 1988 Canadian Econometrics Study Group are appreciated. The suggestions and comments of three referees and the co-editor improved considerably the previous versions of this paper.
parameters of $h^*$ (resp. $V_1^*$ and $V_2^*$) consistently without requiring the distribution of $\eta$ to be parametric. Recently, many other distribution-free methods of estimating $h^*$ (resp. $V_1^*$ and $V_2^*$) or $(h^*, F^*)$ (resp. $(V_1^*, V_2^*, F^*)$) have been developed. The maximum likelihood distribution-free estimator of Cossett (1983) for binary choice models, the maximum rank correlation estimator of Han (1987) for generalized regression models, Ichimura's (1987) estimator for single index models, Klein and Spady's (1990) estimator for discrete choice models, and Stoker's (1986) average derivative estimator are some of the new distribution-free estimators that apply to the qualitative dependent variable models studied in this paper. All these methods still rely on a parametric structure for $h^*$.

Matzkin (1991a) presented a method of estimating monotone and concave subutility functions, $V_i^*$ ($i = 1, \ldots, J$), in polychotomous choice models that does not require that $V_i^*$ ($i = 1, \ldots, J$) possess a parametric structure. This estimation method provided a strongly consistent estimator for the $V_i^*$ functions. The method required, however, that the distribution of the unobservable random terms $\epsilon_i$ ($i = 1, \ldots, J$) belong to a parametric family. The parameters of the distribution of $(\epsilon_1, \ldots, \epsilon_J)$ were also consistently estimated. Another estimation method in this semiparametric vein is the flexible form methods of Gallant (1981, 1982).

All these previously developed methods assume that either $F^*$ or $h^*$ (resp. $V_i^*$; $i = 1, \ldots, J$) is known up to a finite dimensional parameter vector. These assumptions are, however, almost never justified. Economic theory does not impose any restrictions on the parametric structure of functions.

In this paper, we introduce a strongly consistent method of estimating binary threshold crossing models and binary choice models that does not require specifying a parametric structure either for the distribution $F^*$ or for the function $h^*$.

This new method exploits the knowledge economists possess about properties of the function $h^*$ (resp. $V_i^*$); for example, $h^*$ (resp. $V_i^*$) may be known to belong to the set of monotone increasing, concave, and homogeneous of degree one functions. Instead of estimating $h^*$ (resp. $V_i^*$) from a parametric family of functions, the new method obtains an estimator for $h^*$ (resp. $V_i^*$) from a subset of functions possessing this particular set of properties. The requirements on the distribution function $F^*$ are very weak. Since economic theory does not in general imply any properties for $F^*$, it is desirable to develop methods that do not require $F^*$ to satisfy any specific conditions.

Section 2 defines formally the binary threshold crossing model and studies its nonparametric identification. This section also discusses the application of the identification results to the development of various new nonparametric distribution-free methods. Section 3 describes a particular strongly consistent estimator for the pair $(h^*, F^*)$. The estimator is obtained by maximizing a likelihood function over a set of pairs of nonparametric functions. In Section 4, the identification and strong consistency results are applied to binary choice models. Section 5 presents examples of sets of nonparametric functions that satisfy the
2. IDENTIFICATION IN BINARY THRESHOLD CROSSING MODELS

The use of binary threshold crossing models is wide spread in many fields, economics in particular. See, for example, the books of Cox (1970), Finney (1971), and Maddala (1983) for an extensive list of empirical applications of these models.

In binary threshold crossing models, the value of an observable dichotomous variable \( y \) is determined by

\[
(1) \quad y = 1[h^*(r) - \eta \geq 0],
\]

where \( r \in R^K \) denotes a vector of observable exogenous variables, \( h^*: T_i \rightarrow R \), \( T_i \subset R^K \), and \( \eta \) is an unobservable random variable that possesses a cumulative distribution function, \( F^* \); \([\cdot]\) is the logical operator that equals 1 when \([\cdot]\) is true and equals zero otherwise. The vector \( r \) possesses an unknown probability density function \( g \). We will denote by \( G \) the probability measure induced by \( g \) and by \( S_G \) the support of \( G \). The set \( T_i \cap S_G \) will be denoted by \( T \). Although binary threshold crossing models do not necessarily impose this restriction, we will assume that the random variable \( \eta \) is independent of \( r \).

For instance, in a model of labor participation, the value of \( y \) for an individual equals one if he participates in the labor force and \( y \) equals zero otherwise; \( r \) denotes a vector of relevant characteristics of the individual, such as his income, education, and work history; and \( h^*(r) - \eta \) denotes the willingness of the individual to participate in the labor force. The individual is assumed to participate in the labor force if his willingness to participate is above a given threshold.

Our objective is to develop strongly consistent methods of estimating binary threshold crossing models that do not require specification of a parametric structure either for \( h^* \) or for \( F^* \). The main advantage of the new methods over existing ones is that the new methods may be less susceptible to misspecifications.

To develop these new methods, we first need to show that it is possible to identify the pair \((h^*, F^*)\) without specifying a parametric structure either for \( h^* \) or for \( F^* \). For this, we first note that by (1) the probability, \( \Pr[y = 1| r; h^*, F^*] \), that \( y \) equals one when the vector of observable exogenous variables is \( r \) equals the value of the distribution function \( F^* \) at \( h^*(r) \):

\[
(2) \quad \Pr[y = 1| r; h^*, F^*] = F^*(h^*(r)).
\]
The most we can obtain from the data is the value of a function, \( p(r) \), which equals \( \Pr[y = 1|r; h^*, F^*] \) at a.e. \( [G] \) value of \( r \) on \( T \). In other words, \( p(r) \) assigns to almost every \( [G] \) value of \( r \) on \( T \) the frequency with which \( y = 1 \) given \( r \). To recover the functions \( F^* \) and \( h^* \) from \( p \), enough conditions need to be satisfied by \( h^* \) and \( F^* \) that will allow us to separate the influence of \( F^* \) from the influence of \( h^* \) on the observed frequencies. The pair \( (h^*, F^*) \) will then be identified from any other pair \((h, F)\) that belongs to a set \((W \times \Gamma)\) of pairs of functions satisfying the same conditions.

As usual (see Manski (1988)), we say that \((h^*, F^*)\) is identified in a set \((W \times \Gamma)\) to which \((h^*, F^*)\) belongs if any other pair \((h, F)\) in \((W \times \Gamma)\) induces probability densities of \( y \), conditional on the observable exogenous variables, that are different from the true conditional probability density of \( y \). Formally we have the following definition.

**Definition:** The pair \((h^*, F^*)\) is identified in the set \((W \times \Gamma)\) if \((h^*, F^*) \in (W \times \Gamma)\) and for any pair \((h, F)\) in \((W \times \Gamma)\) such that \((h, F) \neq (h^*, F^*)\) there exists a set \( D \subseteq T \) such that \( G[D] > 0 \) and for all \( r \in D \)

\[
\Pr[y = 1|r; h, F] \neq \Pr[y = 1|r; h^*, F^*].
\]

We will say that two distribution functions \( F \) and \( F' \) are different if they attain different values on a subset of \( h^*(T) \) that possesses positive Lebesgue measure. And we will say that two functions \( h \) and \( h' \) are different if they attain different values on a subset of \( T \) that possesses positive probability measure with respect to \( G \).

We next describe a set of assumptions that will allow us to identify \((h^*, F^*)\) within a set \((W \times \Gamma)\).

**Assumptions on the Set of Functions \( W \):**

W.1. \( W \) is a set of real valued, continuous functions with domain \( T_r \).

W.2. There exists a subset \( \bar{T} \) of \( T \) such that (i) for all \( h, h' \in W \) and all \( r \in \bar{T} \), \( h(r) = h'(r) \); and (ii) for all \( h \in W \) and all \( t \) in \( h^*(T) \), there exists \( r \in \bar{T} \) such that \( h(r) = t \).

W.3. \( h^* \in W \).

W.4. \( h^* \) is strictly increasing in the \( K \)th coordinate of \( r \).

Section 4 presents examples of sets of nonparametric functions that satisfy these assumptions.

Assumption W.1 states that the functions in \( W \) are real valued and continuous. The former property implicitly guarantees that the domain of \( F^* \) is in the real line. The latter property implies that to know the values on \( T \) of any function \( h \) in \( W \) it suffices to know the values of \( h \) on a dense subset of \( T \).

Assumption W.2(ii) states that all functions in \( W \) attain the same values on a subset \( \bar{T} \) of \( T_r \). This assumption allows us to separate the influence of \( F^* \) from the influence of \( h^* \), when the probability that \( y = 1 \) given \( r \) is known for all
$r \in \bar{T}$, since, given that all functions in $\mathcal{W}$ attain the same values on $\bar{T}$, a difference in the probability of $y$, given $r \in \bar{T}$ and $(h, F)$, can only be accounted for by a difference between $F$ and $F^*$. Assumption W.2(ii) implies that the set of values that $h^*$ attains over $T$ equals the set of values that $h^*$ attains over $\bar{T}$. This guarantees that to know the values of $F^*$ on $h^*(T)$, it suffices to know the values of $F^*$ over $\bar{T}$.

Assumption W.3 says that $h^*$ possesses the same properties that all functions in $\mathcal{W}$ possess.

Assumption W.4 is necessary when $F^*$ is not assumed to belong to a set of continuous functions. This assumption guarantees that no subset possessing positive Lebesgue measure in $T_i$ will be mapped by $h^*$ into a point of discontinuity of $F^*$. Without Assumption W.4, we may not be able to recover the values of $h^*$ on this subset, since, because of the weak assumptions we are making, we cannot guarantee that the values of $F^*$, at its points of discontinuity, are uniquely determined from the knowledge of $\Pr[y = 1| r; h^*, F^*]$ at a.e. $[G]$ value of $r$. If $F^*$ is assumed to belong to a set of continuous functions, Assumption W.4 is no longer necessary to guarantee the identification of $(h^*, F^*)$.

**ASSUMPTIONS ON THE SET $\Gamma$:**

1. $\Gamma$ is the set of all monotone increasing functions on $R$ with values in $[0, 1]$.
2. $F^* \in \Gamma$.
3. $F^*$ is strictly increasing on $h^*(T)$.

Assumption $\Gamma.1$ implies that $\Gamma$ contains all possible distribution functions of $\eta$ that are independent of $r$. Assumption $\Gamma.2$ guarantees that $F^*$ satisfies the properties that all functions in $\Gamma$ satisfy. Assumption $\Gamma.3$ insures that variations in the value of $h^*$ induce variations in the value of the observed frequencies, $\Pr[y = 1| r; h^*, F^*]$. Without this latter assumption, we may not be able to recover the values of $h^*$.

**ASSUMPTIONS ON THE PROBABILITY DENSITY $g$:**

1. For any $r \in T$ and any $\delta > 0$, $G[B(r, \delta) \cap T] > 0$, where $B(r, \delta) = \{r' \in R^K : \|r - r'\| < \delta\}$.
2. The $k$th coordinate of $r$ possesses on $T$ a Lebesgue density conditional on the other components of $r$.

Assumption G.1 guarantees that whenever two continuous functions attain different values at some $r \in T$, they attain different values on a subset of $T$ that possesses positive probability with respect to $G$.

Assumption G.2 implies that any $r = (r_1, \ldots, r_{K-1}, r_K) \in T_i$ belongs to some one dimensional interval $[\{r_1, \ldots, r_{K-1}, r_{K-1}, r_K\}, (r_1, \ldots, r_{K-1}, r_K)]$ in $T_i$ and the probability of this interval, conditional on $(r_1, \ldots, r_{K-1})$, is positive. These properties are employed, with Assumption W.4, to show the denseness of a subset of $T$ at which the values of $h^*$ are points of continuity of $F^*$. Without
the existence of this dense set, we may not be able to recover all values of \( h^* \) when \( F^* \) possesses points of discontinuity. When \( I' \) is assumed to possess only continuous functions, this assumption is no longer necessary to guarantee the identification of \((h^*, F^*)\).

Our main result in this section is the following theorem, whose proof is given in Appendix A.²

**Theorem 1** (Identification of the Threshold Crossing Model): Suppose that the binary threshold crossing model satisfies Assumptions W.1–W.4, I.1–I.3, and G.1–G.2. Then, \((h^*, F^*)\) is identified within \((W \times I')\).

When stronger assumptions are imposed, it is easy to show why this theorem holds. Suppose that the assumptions guarantee that the values of \( \Pr[y = 1|r; h^*, F^*] \) are known everywhere on \( T \). It then follows that \( F^* \) can be recovered from the values of \( \Pr[y = 1|r; h^*, F^*] \) on \( T \). To see this, for each \( t \in h^*(T) \) let \( r^t \in T \) be such that \( h^*(r^t) = t \). (The existence of \( r^t \) is guaranteed by Assumption W.2.) Then \( F^*(t) = F^*(h^*(r^t)) = \Pr[y = 1|r; h^*, F^*] \). Hence, we can recover \( F^* \). Next, the values of \( h^* \) can be recovered from \( \Pr[y = 1|r; h^*, F^*] \) and \( F^* \), since, for each \( r \in T \), \( h^*(r) = (F^*)^{-1}(\Pr[y = 1|r; h^*, F^*]) \). Hence, we can recover \( h^* \).

The proof of Theorem 1, which is presented in Appendix A, is more involved than the above argument because \( \Pr[y = 1|r; h^*, F^*] \) may not be known on a set that possesses zero \( G \)-probability and \( F^* \) may be discontinuous on a countable set.

If all the functions in the set \( I' \) are assumed to be continuous, then \( \Pr[y = 1|r; h^*, F^*] \) will be continuous, and knowledge of \( \Pr[y = 1|r; h^*, F^*] \) for all \( r \in T \) except perhaps on a set of zero probability implies knowledge of \( \Pr[y = 1|r; h^*, F^*] \) everywhere on \( T \). In this situation, the argument given above is a precise argument. (See the proof of Theorem 1 for more detail.) Moreover, when the functions in \( I' \) are continuous, the conclusions of Theorem 1 are stronger. The values of the function \( F^* \) are identified at every \( t \) in \( h^*(T) \) —not just at a.e. \( t \), with respect to the Lebesgue measure. In addition, as has already been mentioned, it is not necessary in this case to use either Assumption W.4 or Assumption G.2.

The identification result of Theorem 1 can be employed to develop strongly consistent estimation methods for \((h^*, F^*)\) or for \( h^* \). In particular, existing semiparametric methods, which require that either \( h^* \) or \( F^* \) be parametric, can be combined and modified to obtain methods that are nonparametric in both \( h^* \) and \( F^* \). Or, estimators for \((h^*, F^*)\) can be derived from nonparametric estimators for the frequency function \( \Pr[y = 1|r; h^*, F^*] \). The most critical step in these procedures requires being able to restrict \( h^* \) to satisfy Assumption W.2. Theorem 1 can be employed to insure the identification of \((h^*, F^*)\). Particular

²A recent paper by Matzkin (1990b) also studies the fully nonparametric identification (and estimation) of binary threshold crossing models.
sets of restrictions on $W$, $I$, and $G$ will typically be needed to insure the strong consistency of the particular estimation method employed.

In the next section, we describe a strongly consistent estimator for $(h^*, F^*)$ that combines two semiparametric methods. This estimator, which is obtained by maximizing a likelihood function over a set of pairs of nonparametric functions, is based upon Coslett's (1983) distribution-free semiparametric method and Matzkin's (1991a) semiparametric method of estimating concave and monotone functions. Another example of a distribution-free semiparametric estimator that could be combined with Matzkin's semiparametric estimator to obtain strongly consistent fully nonparametric estimators for $(h^*, F^*)$ is the estimator of Klein and Spady (1988) (see Matzkin (1990d)).

A strongly consistent estimator for $(h^*, F^*)$ that is derived directly from a nonparametric estimator for $\Pr\{y = 1|r, h^*, F^*\}$ is presented in Matzkin (1990c). This method proceeds by first finding an estimator, $\hat{h}_N$, for the conditional expectation of $y$ given $r$ by using the method of kernels (Nadaraya (1965), Watson (1964)). Next, $\hat{h}_N$ is employed to derive estimators for $h^*$ and $F^*$ in a way similar to that used to recover $h^*$ and $F^*$ from $\Pr\{y = 1|r, h^*, F^*\}$ in the paragraph immediately after the statement of Theorem 1. Various other nonparametric methods of estimating the conditional expectation of $y$ given $r$ also could be employed, instead of the method of kernels (see Prakasa Rao (1983)).

The basic identification results obtained in this section also can be applied to develop strongly consistent estimators for a nonparametric function $h^*$ of observable exogenous variables in generalized regression models. An illustration of such an application is presented in Matzkin (1991b).

3. A CONSISTENT ESTIMATOR FOR THE BINARY THRESHOLD CROSSING MODEL

In this section we introduce a maximum likelihood estimator for the pair of functions $(h^*, F^*)$ and show its strong consistency. To define this estimator, we let $x^{(N)} = \{y^i, r^i\}_{i=1}^N$ denote $N$ independent observations on the dependent variable $y$ and the vector of observable exogenous variables $r$. The conditional log-likelihood function is then

$\mathcal{L}(x^{(N)}, h, F) = \sum_{i=1}^N \left[ y^i \log \left( F(h(r^i)) \right) + (1 - y^i) \log \left( 1 - F(h(r^i)) \right) \right].$  

We define our maximum likelihood estimator for $(h^*, F^*)$ to be a pair $(\hat{h}_N, \hat{F}_N)$ that maximizes $\mathcal{L}(x^{(N)}, h, F)$ over the set $(W \times I')$. In particular applications, the functions in $W$ will be restricted to satisfy properties that are assumed to be satisfied by $h^*$.

3 I am indebted to one of the referees for suggesting the development of this kind of estimator.
We will establish the strong consistency of this estimator with respect to the metric, \( m \), defined by

\[
m[(h, F), (h', F')] = d_F(F, F') + d_W(h, h'),
\]

where \( d_F \) is a metric on \( \Gamma \) and \( d_W \) is a metric on \( W \). The metric \( d_F : \Gamma \times \Gamma \to R \) is defined by

\[
d_F(F, F') = \int_{h^*(T)} |F(t) - F'(t)| e^{-||t||} \, dt,
\]

where the integration is with respect to the Lebesgue measure. We define the metric \( d_W : W \times W \to R \) by

\[
d_W(h, h') = \int_T \left| h(r) - h'(r) \right| e^{-||r||} \, dG(r).
\]

As will be discussed below, other metrics on \( W \) can also be used.

To prove the strong consistency of the maximum likelihood estimator, we will use the assumptions stated in Section 2 and the following additional assumptions on \( W \) and \( G \).

**Assumptions on the Set of Functions \( W \):**

W.5. \( W \) is compact with respect to the metric \( d_W \).

W.6. Convergence in \( W \) with respect to \( d_W \) implies pointwise convergence on \( T \); i.e., if \( \{h_n\} \subset W \), \( h \in W \), and \( d_W(h_n, h) \to 0 \) as \( n \to \infty \), then for all \( r \in T \), \( h_n(r) \to h(r) \).

The compactness assumption is made to guarantee that \( W \) is separable and it can be covered, for any \( \varepsilon > 0 \), by a finite number of \( \varepsilon \)-neighborhoods. The separability of \( W \) is used to prove the measurability of the supremum of probability densities over neighborhoods in the space \((W \times \Gamma)\). Assumption W.6 is employed to prove the continuity in \((h, F)\) of the probability density of the observations and the measurability of the supremum of probability densities over neighborhoods in \((W \times \Gamma)\). Assumptions W.1, W.4, and G.2 of the previous section are also used to prove the continuity in \((h, F)\) of the probability density of the observations. Assumption W.6 is satisfied, for example, when \( T_i \) is an open set and all functions in \( W \) are monotone and continuous on \( T_i \).

**Assumptions on the Probability Density \( g \):**

G.3. \( \int_T |\log g(r)| \, dG(r) < \infty \).

This assumption is needed to prove the integrability of several auxiliary functions. Many probability densities satisfy this assumption; in particular, when \( T \) is a bounded set any bounded density satisfies Assumption G.3.

The consistency of the maximum likelihood estimator is stated in the following theorem, which is proved in Appendix A.
THEOREM 2 (Consistency of the Maximum Likelihood Estimator): Suppose that the binary threshold crossing model satisfies Assumptions W.1–W.6, Γ.1–Γ.3, and G.1–G.3. If for each \( N (N = 1, 2, \ldots) (\hat{h}_N^M, \hat{F}_N^M) \in (W \times \Gamma) \) maximizes the likelihood of \( N \) independent observations \( x^{(N)} \) over \( (W \times \Gamma) \), then
\[
\Pr \left( \lim_{N \to \infty} m \left( \left( \hat{h}_N^M, \hat{F}_N^M \right), (h^*, F^*) \right) = 0 \right) = 1.
\]

From the proof of Theorem 2 it is easy to see that the only property used about the metric \( d_W \) is that the pair \((W, d_W)\) satisfies Assumptions W.5–W.6. Hence, other metrics on \( W \) for which these assumptions are satisfied can also be used. Suppose, for example, that \( W \) is compact with respect to the essential supremum metric defined by \( G \) on \( T \). Then, under Assumptions W.1 and G.2, Assumption W.6 is satisfied, and the maximum likelihood estimator of \( h^* \) is strongly consistent with respect to this stronger metric.

The computation of the maximum likelihood estimator defined above requires maximizing the conditional log-likelihood function (3) over a set of nonparametric functions. A procedure to solve this maximization problem is described in Section 6.

4. THE BINARY CHOICE MODEL

Many choices made by economic agents can be modeled as a binary choice. The choice between buying or renting a house, buying or not buying a commodity, accepting or rejecting a loan application, and commuting by car or bus are just few examples of binary choices.

In the binary choice model, a typical individual is assumed to choose one of two alternatives. The dependent observable variable, \( y \), equals one when the first alternative is chosen and it equals zero otherwise. The value of \( y \) is assumed to be determined by
\[
y = 1 \left[ V_1^*(s, z_1) + \epsilon_1 \geq V_2^*(s, z_2) + \epsilon_2 \right],
\]
where \( s \in S \subset R^L \) denotes a vector of observable socioeconomic characteristics of the individual, \( z_1 \in Z_1 \subset R^{K_1} \) and \( z_2 \in Z_2 \subset R^{K_2} \) denote vectors of observable attributes of the first and second alternative, respectively; \( V_1^* : (S \times Z_1) \to R \) and \( V_2^* : (S \times Z_2) \to R \) are unknown functions; \( \epsilon_1 \) and \( \epsilon_2 \) are unobservable random terms representing the values of unknown functions of unobservable attributes of the first and second alternative, respectively.

For example, in a model of choice between commuting by car or commuting by bus, \( s \) denotes income and car availability of the commuter, \( z_1 \) denotes the time and cost of commuting by car, and \( z_2 \) denotes the time and cost of commuting by bus. \( V_1^*(s, \cdot) \) and \( V_2^*(s, \cdot) \) denote the utility that a commuter with income and car availability \( s \) derives from time and cost when he takes the

\[\text{4 The essential supremum metric, } \hat{d}_W, \text{ defined by } G \text{ on } T \text{ is}
\[
\hat{d}_W(h, h') = \inf \{ t | G \{ r \in T | h(r) - h' (r) > t \} = 0 \}.
\]
car and when he takes the bus, respectively; \( \varepsilon_1 \) and \( \varepsilon_2 \) denote the utility the typical commuter derives from unobservable attributes, such as comfort, of the car and the bus, respectively.

We will denote by \( V^* \) the pair of functions \( (V_1^*, V_2^*) \). The vector \((s, z_1, z_2)\) will be assumed to possess a probability density, \( g \), that induces a probability measure \( G \). Let \( \eta \) denote \(-(\varepsilon_1 - \varepsilon_2)\). Assume that the random variable \( \eta \) is distributed independently of \((s, z_1, z_2)\) with an unknown cumulative distribution function, \( F^* : R \to [0, 1] \). The probability that \( y = 1 \) given \((s, z_1, z_2)\) is then

\[
\Pr \{ y = 1 | s, z_1, z_2; V^*, F^* \} = F^* (V_1^* (s, z_1) - V_2^* (s, z_2)).
\]

This is a particular case of the binary threshold crossing model discussed in Section 2. Here, \( T_{i} = S \times Z_1 \times Z_2 \), \( h^*(s, z_1, z_2) = V_1^*(s, z_1) - V_2^*(s, z_2) \), and \( \eta = \varepsilon_1 - \varepsilon_2 \). In particular,

\[
h^*(T_i) = \{ t \in R | t = V_1^*(s, z_1) - V_2^*(s, z_2) \}
\]

for some \((s, z_1, z_2) \in (S \times Z_1 \times Z_2)\).

We will let \( T_e \) denote the set \( S \times Z_1 \times Z_2 \) and \( h_e^* \) denote the real valued function on \( T_e \) defined by

\[
h_e^*(s, z_1, z_2) = V_1^*(s, z_1) - V_2^*(s, z_2).
\]

We will denote the support of \( G \) by \( S^G \times S^G \times S^G \subset R^l \times R^{k1} \times R^{k2} \) and the sets \( S \cap S^G, Z_1 \cap S^G, \) and \( Z_2 \cap S^G \) by \( T_S \), \( T_{Z_1} \), and \( T_{Z_2} \), respectively. The set \( T_S \times T_{Z_1} \times T_{Z_2} \) will be denoted by \( \tilde{T} \).

### 4.1. Identification in the Binary Choice Model

Similarly to the definition in Section 2, we say that the pair \((V^*, F^*)\) is identified in a set \((U \times \Gamma)\) if \((V^*, F^*) \in (U \times \Gamma)\) and for any pair \((V, F)\) in \((U \times \Gamma)\) such that \((V, F) \neq (V^*, F^*)\) there exists a set \( D \subset T \) such that \( G[D] > 0 \) and for all \( r \in D \)

\[
\Pr \{ y = 1 | s, z_1, z_2; V, F \} \neq \Pr \{ y = 1 | s, z_1, z_2; V^*, F^* \}.
\]

In this definition, \( F \neq F^* \) if \( F \) and \( F^* \) attain different values on a subset of \( h_e^*(T) \) that possesses positive Lebesgue measure and \( V \neq V^* \) if they attain different values on a subset of \( T \) that possesses positive probability with respect to \( G \).

We will show that, by applying the identification result in Section 2, the functions \( V^* \) and \( F^* \) can be identified also in this model, within a set \((U \times \Gamma)\).

For this result we require the following assumptions:

**Assumptions on \( U \):**

- **U.1.** \( U \) is the set \( U_1 \times U_2 \), where \( U_1 \) is a set of real valued, continuous functions on \( S \times Z_1 \) and \( U_2 \) is a set of real valued, continuous functions on \( S \times Z_2 \).

- **U.2.** There exists \( \bar{z}_2 \in T_{Z_2} \) and \( \gamma \in R \) such that for all \( s \in T_S \) and all \( z_2 \in U_2 \),

\[
V_2^*(s, z_2) = \gamma.
\]

- **U.3.** There exists a subset \( \bar{T} \) of \( T \) such that (i) for all \((V_1, V_2, V_1', V_2') \in U \) and all \((s, z_1, z_2) \in \bar{T} \),

\[
V_1(s, z_1) - V_2(s, z_2) = V_1'(s, z_1) - V_2'(s, z_2),
\]

and (ii) for all
\((V_1, V_2) \in U \text{ and all } t \in h_c^\circ(T), \text{ there exists } (s, z_1, z_2) \in \tilde{T} \text{ such that } V_1(s, z_1) - V_2(s, z_2) = t.\)

U.4. \(V^* \in U.\)

U.5. There exists a coordinate of the vector \((s, z_1, z_2)\) such that \(h_c^*\) is strictly increasing with respect to that coordinate.

Let \(W_c\) denote the set of all functions \(h_c^*: T_c \rightarrow R\) such that for some \(V \in U\) and all \((s, z_1, z_2) \in T_c,\)

\[ h_c^*(s, z_1, z_2) = V_1(s, z_1) - V_2(s, z_2). \]

Assumption U.1, which is analogous to Assumption W.1 in Section 2, implies that all functions in \(W_c\) are continuous. This will be employed to show that \(h_c^*\) can be recovered from \(\Pr[y = 1|s, z_1, z_2; V^*, F^*].\)

Assumption U.2 states that all functions in \(U_2\) attain a value \(\gamma\) at every point in \(T_1 \times \tilde{T}_2.\) This is analogous to the assumption that one of the subutility functions of the alternatives is independent of the vector of socioeconomic characteristics, which is made in parametric discrete choice models where the utility of the observable attributes are assumed to be different across alternatives. This assumption allows us to recover \(V^*_1\) and \(V^*_2\) from \(h_c^*\).

Assumptions U.3 and U.4 are analogous, respectively, to Assumptions W.2 and W.3 in Section 2. When Assumption U.2 holds, Assumption U.3 is satisfied, for example, when there exists a subset \(T_1\) of \(T_1 \times \tilde{T}_2\) such that (i) for all \(V_1, V_1' \in U_1\) and all \((s, z_1) \in \tilde{T}_1, V_1(s, z_1) = V_1'(s, z_1),\) and (ii) for all \(V_1 \in U_1\) and all \(t \in h_c^\circ(T),\) there exists \((s, z_1) \in \tilde{T}_1\) such that \(V_1(s, z_1) = t + \gamma.\)

Assumption U.5, which is analogous to Assumption W.4 in Section 2, is unnecessary when the distribution function \(F^*\) is assumed to belong to a set of continuous functions. Assumption U.5 will be satisfied, for example, when \(V^*_1\) is strictly increasing in one coordinate of \(Z_1,\) or when \(V^*_2\) is strictly decreasing in one coordinate of \(Z_2,\) or when \(V^*_1\) is strictly increasing in one coordinate of \(s\) and \(V^*_2\) is either independent or decreasing in that same coordinate of \(s.\)

**Assumptions on \(G:\)**

G.1'. For any \((s, z_1, z_2) \in T \text{ and any } \delta > 0, G[B((s, z_1, z_2), \delta) \cap T] > 0.\)

G.2'. At least one of the coordinates of \((s, z_1, z_2)\) in which \(h_c^*\) is strictly increasing possesses a Lebesgue density, conditional on the other coordinates of \((s, z_1, z_2).\)

Assumptions G.1' and G.2' are analogous, respectively, to the Assumptions G.1 and G.2, which were made in Section 2.

The identification of \((V^*, F^*)\) in binary choice models is established in the next theorem.\(^5\)

\(^5\)Matrkin (1990b) also studies the identification (and estimation) of binary choice models. Independent results by Yellott (1977) and Strauss (1979) provide a different set of conditions for the identification of \((V^*, F^*).\)
THEOREM 3 (Identification of the Binary Choice Model): Suppose that the binary choice model satisfies Assumptions U.1–U.5, Γ.1–Γ.3, and G.1–G.2'. Then, \((V^*, F^*)\) is identified within \((U \times \Gamma)\).

In the proof of this theorem, which is presented in Appendix A, we first employ Theorem 1 to show that the function \(h^*_\eta\) and the distribution, \(F^*\), of \(\eta\) can be recovered when the value of \(\Pr[y = 1|s, z_1, z_2; V^*, F^*]\) is known at a.e. \([G]\) point in \(T\). We then show that \(V^*_1\) and \(V^*_2\) can be recovered from \(h^*_\eta\).

We next describe an estimator that uses the above identification result.

4.2. Maximum Likelihood Estimation

Let \(x^{(N)} = \{y^i, s^i, z_1^i, z_2^i\}_{i=1}^N\) denote a sample of \(N\) independent observations. By (7), the conditional log-likelihood function of these observations is

\[
\mathcal{L}(x^{(N)}, V, F) = \sum_{i=1}^N \left[ y^i \log \left( F \left[ V_1(s^i, z_1^i) - V_2(s^i, z_2^i) \right] \right) 
+ (1 - y^i) \log \left( 1 - F \left[ V_1(s^i, z_1^i) - V_2(s^i, z_2^i) \right] \right) \right].
\]

We define our maximum likelihood estimator for \((V^*, F^*)\) to be a pair \((V_{ML}^{c,N}, F_{ML}^{c,N})\) that maximizes (8) over the set \((U \times \Gamma)\).

We will establish the strong consistency of this estimator with respect to the metric \(m\) defined by

\[
m\left[ (V, F), (V', F') \right] = d_F(F, F') + d_U(V, V')
\]

where \(d_F\) is as defined in (4) (Section 3.1) and \(d_U\) is a metric defined by

\[
d_U(V, V') = \sum_{j=1}^2 \int_T |V_j'(s, z_j) - V'_j(s, z_j)| e^{-\|G(s, z_1, z_2)\|} dG(s, z_1, z_2).
\]

Our consistency result will require the following assumptions, in addition to those used in Theorem 3.

ASSUMPTIONS ON \(U\):

U.6. \(U\) is compact with respect to the metric \(d_U\).
U.7. Convergence in \(U\) with respect to \(d_U\) implies pointwise convergence on \(T\).

ASSUMPTIONS ON \(G\):

G.3'. \(\int_T |\log g(s, z_1, z_2)| dG(s, z_1, z_2) < \infty\).

These assumptions will play the same role as Assumptions W.5–W.6 and G.3 played in the proof of Theorem 2 in Section 3.1.

THEOREM 4 (Consistency of the Maximum Likelihood Estimator for Binary Choice Models): Suppose that the binary choice model satisfies Assumptions...
U.1–U.7, Γ.1–Γ.3, and G.1–G.3. If for each \( N \) (\( N = 1, 2, \ldots \)) \((\hat{V}_{c,N}^{ML}, \hat{F}_{c,N}^{ML}) \in (U \times \Gamma)\) maximizes the likelihood of \( N \) independent observations \( x^{(N)} \) over \((U \times \Gamma)\), then

\[
\Pr \left\{ \lim_{N \to \infty} m \left[ \left( \hat{V}_{c,N}^{ML}, \hat{F}_{c,N}^{ML}, (V^*, F^*) \right) = 0 \right] \right\} = 1.
\]

Since the proof of this theorem is very similar to the proof of Theorem 2, we omit its proof.

Metrics on \( U \) other than \( d_U \) could also be used. If, for example, \( U \) is compact with respect to the stronger metric \( \bar{d}_U \) defined by

\[
\bar{d}_U(V, V') = \inf \{ t \mid G \left[ \left\{ (s, z_1, z_2) \in T \mid \left| V_j(s, z_j) - V'_j(s, z_j) \right| > t \right\} \right] = 0 \}
\]

then the consistency of the estimator can be obtained with respect to this stronger metric.

In Section 6, we discuss the calculation of the maximum likelihood estimator studied in this subsection.

5. EXAMPLES

In this section, we first present several examples of sets of nonparametric functions that satisfy the assumptions made about the sets \( W \) and \( U \) in the previous sections. Then, in Subsection 5.2, we describe examples of economic models where the functions \( h^* \) and \( V^* \) belong to some of these sets of functions.

5.1. Examples of Sets of Functions

Example 1: A set \( W \) of continuous, monotone increasing, concave, and homogeneous of degree one functions\(^6\) for the threshold crossing model. Suppose that \( T_r \), the domain of the functions in \( W \), is \( R^K_{\geq 0} \), the set of all \( r \in R^K \) with strictly positive coordinates. Let \( r^* \in T_r \), \( \alpha > 0 \), and \( B > 0 \) be given. Let \( W \) be the set of all continuous, monotone increasing, concave, and homogeneous of degree one functions, \( h: T_r \to R \), such that for each \( h \in W \), \( h(r^*) = \alpha \) and for each \( r \in T_r \) the \( K \)th coordinate of a subgradient\(^7\) of \( h \) at \( r \) is bounded below by \( B \). Assume that \( h^* \) belongs to \( W \) and \( r \) possesses an absolutely continuous distribution whose support is \( R^K_{\geq 0} \). Then, \( W \) satisfies Assumptions W.1–W.6.

To see this, note that since \( h^* \in W \) and all functions in \( W \) are continuous and strictly increasing with respect to the \( K \)th coordinate of \( r \), Assumptions W.1, W.3, and W.4 are satisfied. Let \( \overline{T} \) be the ray \( \{ r \in T_r \mid r = \lambda r^* \text{ for some } \lambda > 0 \} \).

\(^6\)A function \( h: T_r \to R \) is homogeneous of degree one if, for all \( r \in T_r \) and all \( \lambda > 0 \), \( h(\lambda r) = \lambda h(r) \).

\(^7\)A vector \( D \in R^K \) is a subgradient of a concave function \( h: T_r \to R \) at \( r \in T_r \) if, for all \( r' \in T_r \), \( h(r') \leq h(r) + D \cdot (r' - r) \).
Then, $W$ satisfies W.2(i) because the homogeneity of degree one of the functions in $W$ implies that for all $h \in W$ and all $r$ in $\bar{T}$,

$$h(r) = h(\lambda r^*) = \lambda h(r^*) = \lambda a$$

for $\lambda \in R$ such that $r = \lambda r^*$. $W$ satisfies W.2(ii) because the monotonicity and homogeneity of degree one of the functions in $W$ implies, together with the fact that $T = R^k_{+}$, that each function $h$ in $W$ maps $\bar{T}$ and $T$ onto $(0, \infty)$. Thus, all the requirements on $W$ necessary for the identification result of Theorem 1 are satisfied.

The additional requirements on $W$ necessary for the consistency result of Theorem 2 are also satisfied. Assumption W.5 follows by Lemma B.1 in Appendix B and Assumption W.6 follows by the definition of $T_i$ and the continuity and monotonicity of the functions in $W$.

Subsection 5.2 presents an example of an economic model in which $h^*$ belongs to the set $W$ described above.

The assumption that $W$ is a set of homogeneous of degree one functions implies that each function in $W$ can be determined by one level set and that the value of the function increases linearly across level sets. The rate of increase in the value of the functions in $W$ is determined by the specified value $a$ at $r^*$. Different functions in $W$ attain the same value at $r^*$ and at the ray $\bar{T}$ that passes through $r^*$, but differ in their level sets. (See Figure 1.)

Semiparametric, distribution-free methods of estimating threshold crossing models typically assume that $h^*(r) = r \cdot \beta^*$ for some parameter $\beta^*$, which

![Figure 1](image-url)

Figure 1
belongs to the set

\[(9) \quad \{ \beta \in R^K \mid \| \beta \| = 1 \}. \]

Fixing the norm of the parameter vector is analogous to our assumption that \( h(r^*) = \alpha \); both fix the rate of increase of the functions. Linearity of the functions in \( r \) implies not only that the functions are homogeneous of degree one but also that their level sets are hyperplanes. (See Figure 2.)

**Example 2:** A set of functions, for the binary threshold crossing model, that are additively separable into a continuous and monotone increasing function and a continuous, monotone increasing, concave, and homogeneous of degree one function. Denote each \( r \in R^K \) by \( r = (r_1, r_2) \), where \( r_1 \) and \( r_2 \) are subvectors of dimension \( K_1 \) and \( K_2 \), respectively. Let \( T_i \), the domain of the functions in \( W \), be \( T_i^1 \times T_i^2 \), where \( T_i^1 \subset R^{K_1} \) and \( T_i^2 = R^{K_2}_+ \). Let \( r_i^* \in T_i^1 \), \( r_i^2 \in T_i^2 \), \( \alpha > 0 \), \( \beta \in R \), and \( B > 0 \) be given. Assume that, conditional on \( r_1 \), \( r_2 \) possesses a conditional distribution that is absolutely continuous and its support is \( R^{K_2}_+ \). Let \( T = T_i \cap S_G \), where \( S_G \) denotes the support of the probability measure, \( G \), of \( r \). Assume that, for all \( r \in T \) and all \( \delta > 0 \), \( G(B(r, \delta) \cap T) > 0 \) and that, for all \( r_2 \in T_i^2 \), \( (r_i^*, r_2) \in T \). For any two real valued functions \( u_1, u_1' \) on \( T_i^1 \) let \( d_{u}(u_1, u_1') = \int_T |u_1(r_1) - u_1'(r_1)| e^{-\|r_2\|_B} dG(r) \).

Let \( W_1 \) be a compact set (with respect to \( d_{u_1} \)) of continuous and monotone increasing functions \( u_i^* : T_i^1 \rightarrow \mathbb{R} \) such that \( u_i^*(r_i^*) = \beta \) and \( u_i(r_i) \geq \beta \) for all \( r_1 \in T_i^1 \). Let \( W_2 \) be a set of continuous, monotone increasing, concave, and
homogeneous of degree one functions, \( v_2: T_i^2 \to R \), such that for each \( v_2 \in W_2, v_2(r^*_2) = \alpha \) and for each \( r_2 \in T_i^2 \) the \( K \)th coordinate of a subgradient of \( v_2 \) at \( r_2 \) is bounded below by \( B \). For any two functions \( v_2, v'_2 \) in \( W_2 \) let
\[
d_{w_2}(v_2, v'_2) = \int_T |v_2(r_2) - v'_2(r_2)| e^{-\|r_2\|} dG(r).
\]

Let \( W \) be the set of functions, \( h: T_i \to R \), such that for some \( v_1 \in W_1 \), some \( v_2 \in W_2 \), and all \( (r_1, r_2) \in T_i \),
\[
h(r_1, r_2) = v_1(r_1) + v_2(r_2).
\]
Assume that \( h^* \) belongs to \( W \). Then, \( W \) satisfies Assumptions W.1–W.6.

That Assumptions W.1, W.3–W.4, and W.6 are satisfied can be easily verified.

Let \( \bar{T} \) be the set defined by
\[
\bar{T} = \{(r_1, r_2) \in T_i | r_1 = r^*_1, r_2 = \lambda r^*_2 \text{ for some } \lambda > 0\}.
\]

Then, since for all \( h \in W \) there exist \( v_1 \in W_1 \) and \( v_2 \in W_2 \) such that for all \( r \in \bar{T}_i \),
\[
h(r_1, r_2) = v_1(r_1) + v_2(r_2) = v_1(r^*_1) + v_2(\lambda r^*_2) = \beta + \lambda \alpha,
\]

where \( \lambda \) is such that \( r_2 = \lambda r^*_2 \), \( W \) satisfies Assumption W.2(i). And since for all \( h \in W \), \( h(T_i) = [\beta, \infty) = h(\bar{T}_i) \), \( W \) satisfies W.2(ii). To show that Assumption W.5 is satisfied, we note that by a straightforward modification of Lemma B.1, \( W_2 \) is compact with respect to \( d_{w_2} \). Assumption W.5 then follows from Lemma B.2 in Appendix B, by the compactness of \( W_1 \) and \( W_2 \).

The set of functions described here is useful, for example, when \( h^* \) is known to depend on some exogenous variables, \( r_1 \), and little is known regarding the properties of \( h^* \) with respect to those variables. Note, however, that in most cases, requiring that \( h^* \) be additive separable will be a restrictive assumption.

**Example 3:** A set of functions, for the binary threshold crossing model, that are additively separable into the value of one variable and a continuous, monotone increasing function of the remaining variables. Let \( E \) be a subset of \( R^{k-1} \) such that \( 0 \in E \). Assume that the support of \( G \) is \( E \times R \) and \( r_K \) is distributed on \( E \times R \), conditional on \((r_1, \ldots, r_{K-1})\), with a Lebesgue density. For any two functions \( t: E \to R \) and \( t': E \to R \) define
\[
d_{w_2}(t, t') = \int_{E \times R} |t(r_1) - t'(r_1)| e^{-\|r\|} dG(r).
\]

Let \( W_1 \) be a compact set (w.r.t. \( d_{w_2} \)) of continuous and monotone increasing functions \( t: E \to R \) such that \( t(0) = 0 \). Let \( W \) be the set of all functions \( h: (E \times R) \to R \) for which there exists a function \( t \in W_1 \) such that
\[
h(r_1, r_2, \ldots, r_K) = t(r_1, \ldots, r_{K-1}) + r_K \quad \text{for all } \quad r \in (E \times R).
\]
Assume that \( h^* \in W \). Then, \( W \) satisfies Assumptions W.1–W.6.

As in the above examples, it is easy to verify that Assumptions W.1 and W.3–W.6 are satisfied. Let \( \bar{T} \) be the ray
\[
\{r = (r_1, \ldots, r_K) \in R^K | r_k = 0, k = 1, \ldots, K - 1\}.
\]
Then, since for each function \( t \) in \( W_1 \), \( t(0) = 0 \), it follows from (10) that all functions in \( W \) attain the same value at each point in \( \bar{T} \) and they map \( \bar{T} \) onto \( R \). Hence, Assumption W.2 is satisfied.

Note that, except for the separability structure in (10), the conditions on the set of functions \( W \) are very weak. In particular, when our assumptions on the conditional distribution of \( r_K \) are satisfied, the requirements on \( W \) necessary for the identification result follow from (10), the assumption that for all \( t \in W_1 \), \( t(0) = 0 \), and the continuity of the functions in \( W_1 \).

**Example 4:** A set of continuous, concave, monotone increasing, and homogeneous of degree one functions for the binary choice model. Suppose that \( Z_1 = R^{K_1} \) and \( Z_2 = R^{K_2} \). Let \( z_1^* \in Z_1 \), \( z_2^* \in Z_2 \), \( \alpha_1 > 0 \), \( \alpha_2 > 0 \), and \( B > 0 \) be given. Let \( U_1 \) be the set of all continuous, monotone increasing, concave, and homogeneous of degree one functions, \( V_i: Z_1 \to R \), such that for each \( V_i \in U_1 \), \( V_i(z_1^*) = \alpha_1 \) and for each \( z_1 \in Z_1 \), the \( K \)th coordinate of a subgradient of \( V_i \) at \( z_1 \) is bounded below by \( B \). Let \( U_2 \) be the set of all continuous, monotone increasing, concave, and homogeneous of degree one functions, \( V_2: Z_2 \to R \), such that \( V_2(z_2^*) = \alpha_2 \). Let \( U = U_1 \times U_2 \). Assume that \( \nu^* \in U \), the support of \( G \) is \( R^{K_1} \times R^{K_2} \), and the \( K \)th coordinate of \( z_1 \) possesses, conditional on the other coordinates of \( (z_1, z_2) \), a Lebesgue density. Then, \( U \) satisfies Assumptions U.1–U.6.

As in the examples above, Assumptions U.1, U.4, U.5, and U.7 can be easily verified. Assumption U.2 is satisfied by letting \( \tilde{z}_2 = z_2^* \) and \( \gamma = \alpha_2 \). Let

\[
\bar{T} = \{(z_1, z_2) | z_1 = \lambda_1 z_1^*, z_2 = \lambda_2 z_2^*, \text{ for some } \lambda_1 > 0 \text{ and } \lambda_2 > 0 \}.
\]

Then, since for all \( (V_1, V_2) \in U \) and all \( (z_1, z_2) \in \bar{T} \),

\[
V_1(z_1) - V_2(z_2) = \lambda_1 \alpha_1 - \lambda_2 \alpha_2
\]

for \( \lambda_1 \) and \( \lambda_2 \) such that \( z_1 = \lambda_1 z_1^* \) and \( z_2 = \lambda_2 z_2^* \), \( U \) satisfies Assumption U.3(i); and since for any \( (V_1, V_2) \in U \) the values of \( V_1(z_1) - V_2(z_2) \) range from \(-\infty \) to \( \infty \) when \( (z_1, z_2) \) varies over \( Z_1 \times Z_2 \), \( U \) satisfies Assumption U.3(ii). Finally, Assumption U.6 follows by a simple modification of the argument in Lemma B.1.

Hence, \( U \) satisfies all the assumptions needed to establish the results of Theorems 3 and 4.

**Example 5:** A set of additively separable functions for the binary choice model. Suppose that \( Z \subset R^{K_1} \) contains only a finite number of points, \( 0, q_1, \ldots, q_C \). Let \( \beta_L \in R \) and \( \beta_U \in R \) be given and be such that \( \beta_L < \beta_U \). Let \( U_a \) be the set of all functions \( w: Z \to R \) such that \( w(0) = 0 \) and for all \( j = 1, \ldots, C, \beta_L \leq w(q_j) \leq \beta_U \).

Suppose that \( S = R \), and let \( U \) be the set of all pairs \( (V_1, V_2) \) of functions \( V_i: S \times Z \to R \) and \( V_2: S \times Z \to R \) such that for some function \( w \in U_a \) and all \( (s, z_1, z_2) \in S \times Z \times Z \),

\[
V_1(s, z_1) = s + w(z_1) \quad \text{ and } \quad V_2(s, z_2) = w(z_2).
\]

Assume that the support of \( G \) is \( R \times Z \times Z \), \( V^* \in U \), and \( s \) possesses a
Lebesgue density conditional on \((z_1, z_2)\). Then, it is again easy to verify that \(U\) satisfies Assumptions U.1–U.7.

5.2. Application to Economic Models

We now describe some economic examples in which the functions \(h^*\) and \(V^*\) belong to some of the sets of nonparametric functions described in Subsection 5.1.

Suppose that in the binary threshold crossing model described in Section 2, \(h^*\) denotes the minimum cost to a typical firm of performing a particular project, \(r\) denotes the vector of prices of the inputs necessary to perform the project, and \(\eta\) denotes the revenue to the firm if it performs the project. Assume that the distribution of \(r\) is absolute continuous and possesses support \(R^\times\). The firm decides on whether or not to undertake the project according to whether the value of \(h^*(r) - \eta\) is, respectively, below or above zero. Although for the firm \(r, h^*, \) and \(\eta\) are known, the econometrician only observes \(r\) and the decision of the firm.

We next show that in this model one can identify and strongly consistently estimate \(h^*\) and the c.d.f. \(F^*\) of the revenue \(\eta\). For this, we first note that the definition of \(h^*\) and standard assumptions about the production technology of the firm imply that \(h^*\) is continuous, concave, strictly increasing in all coordinates of \(r\), and homogenous of degree one. Normalize the rate of increase of the function \(h^*\) by assuming that the value of \(h^*\) at some given \(r^*\) is known to be some given value \(\alpha^*\), and strengthen the strict monotonicity of \(h^*\) with respect to the \(K\)th coordinate by requiring that for each \(r\) the \(K\)th coordinate of a subgradient of \(h^*\) at \(r\) be bounded below by \(B > 0\). Then, \(h^*\) belongs to the set of functions \(W^\times\) defined in Example 1. If we further assume that \(F^*\) is strictly increasing on \(R^\times\) and Assumption G.3 is satisfied, it follows by the discussion in Example 1 and the results in Sections 2 and 3 that \((h^*, F^*)\) can be identified and strongly consistently estimated imposing no further restrictions.

Next suppose that in the above economic example, the cost to the typical firm depends also on some characteristics of the firm but that little is known about the properties of \(h^*\) with respect to those characteristics. To estimate \(h^*\), we could then impose the restriction that \(h^*\) be additively separable into a function, \(v^*_1\), of the characteristics of the firm, denoted by \(r_1\), and a function, \(v^*_2\), of the prices of the inputs, denoted by \(r_2\). By the arguments given in the above paragraph, \(v^*_2\) can be assumed to belong to the set of functions \(W^2\) defined in Example 2. Then, if the function \(v^*_1\) is restricted to be continuous, monotone increasing, and to satisfy \(v^*_1(r^*_1) = \beta\) for some given \(r^*_1\) and some given \(\beta > 0\), and \(v^*_1(r_1) \geq \beta\) for all \(r_1\) in the domain of \(V^*_1\), it will follow by the discussion in Example 2 that the identification and strong consistency results presented in Sections 2 and 3 can also be obtained for this example.

For a similar economic example of a binary choice model, suppose that a typical firm faces two alternative projects and it is restricted to perform only one of them. The two projects are performed with technologies that
require different sets of inputs. The revenues from each project are $e_1$ and $e_2$ respectively. The firm chooses to perform the project that yields the highest profit. Let $V^*_1$ and $V^*_2$ denote the minimum cost of performing the first and second alternative projects, respectively. And let $z_1$ and $z_2$ denote the price vectors of the inputs necessary to perform the first and second projects, respectively. Then, standard assumptions about the production technology of the firm imply that $V^*_1$ and $V^*_2$ are continuous, monotone increasing, concave, and homogeneous of degree one functions. Normalize the rates of increase of the functions $V^*_1$ and $V^*_2$ by assuming that $V^*_1(z^*_1) = \alpha_1$ and $V^*_2(z^*_2) = \alpha_2$ for some $z^*_1 \in R_{+1}^{K_1}$, $z^*_2 \in R_{+2}^{K_2}$, $\alpha_1 > 0$, and $\alpha_2 > 0$. Assume that for each $z_1$, $V^*_1$ possesses a subgradient whose $K_1$th coordinate is bounded below by $B > 0$. Assume also that $(z_1, z_2)$ possesses an absolutely continuous probability measure whose support is $R_{+1}^{K_1} \times R_{+2}^{K_2}$. Then, $V^*$ belongs to the set of functions described in Example 4. The set of functions $U$ described in that example can then be employed to estimate $(V^*_1, V^*_2)$ in this economic model.

Summing up, in this section we have described several sets of nonparametric functions that satisfy the assumptions made in the theorems of Sections 2–4. We have also presented some economic examples in which $h^*$ and $V^*$ belong to some of these sets of nonparametric functions.

6. COMPUTATION OF THE MAXIMUM LIKELIHOOD ESTIMATOR

In Section 3 we stated that to calculate the maximum likelihood estimator for $(h^*, F^*)$ we need to compute a solution to the maximization of the likelihood function over a set of nonparametric functions. In this section we describe a procedure to calculate this solution.

The procedure is based upon transforming the maximization over a set of nonparametric functions into a constrained maximization problem over a finite dimensional vector. This procedure follows techniques similar to those described in Cosslett (1983) and Matzkin (1991a).

We will describe the procedure for a binary threshold crossing model. At the end of this section we illustrate with an example the extension of the method to a binary choice model.

To describe the procedure, we first note that given any finite number of observations, the value of the likelihood $L$ at any function $h \in W$ depends on $h$ only through the values that $h$ attains at the vectors $r^1, \ldots, r^N$. Hence, the maximization of $L$ over $W$ can be transformed into the maximization of $L$ over the set of all finite dimensional vectors $(h^1, \ldots, h^N)$ for which there exists a function $h \in W$ with $h(r^i) = h^i$ ($i = 1, \ldots, N$). Similarly, the value of $L$ at any $F \in \Gamma$ depends on $F$ only through the values that $F$ attains at $h^1, \ldots, h^N$. Hence, the maximization of $L$ over $\Gamma$ can be transformed into the maximization of $L$ over the set of all vectors $(F^1, \ldots, F^N)$ for which there exists a function $F \in \Gamma$ with $F(h^i) = F^i$ ($i = 1, \ldots, N$).

The characterization of the set of vectors $(F^1, \ldots, F^N)$ for which there exists a function $F \in \Gamma$ with $F(h^i) = F^i$ ($i = 1, \ldots, N$) can be found in Cosslett (1983).
The characterization of the set of vectors \((h^1, \ldots, h^N)\) for which there exists a function \(h\), within a certain class of nonparametric functions, with \(h^i = h(r^i)\) \((i = 1, \ldots, N)\), can be found in Matzkin (1991a). The sets of nonparametric functions studied in Matzkin (1991a) are sets of concave (or convex) functions that may possess, in addition to concavity, other properties such as monotonicity, homogeneity of degree one, and additive separability.

Combining these characterizations, we can obtain a constrained maximization problem that, after interpolating its solution, is equivalent to the maximization of the conditional likelihood function over the set \((W \times \Gamma)\). Suppose, for example, that \(W\) is the set of functions described in Example 1 of Section 5.1. Let \(r^{0} = 0\) and \(r^{N+1} = \tau^{*}\). Then, the corresponding constrained optimization problem is the following:

\[
\begin{align*}
\text{(11.1)} & \quad \text{Maximize} \quad \sum_{i=1}^{N} \left\{ y^i \log F^i + (1 - y^i) \log (1 - F^i) \right\} & \text{subject to} \\
(11.2) & \quad h^i = D^i \cdot r^i & (i = 0, 1, \ldots, N, N + 1), \\
(11.3) & \quad D^i \cdot r^j \leq D^j \cdot r^i & (i, j = 0, 1, \ldots, N, N + 1), \\
(11.4) & \quad D^{N+1} \cdot r^{N+1} = \alpha, \\
(11.5) & \quad 0 \leq D^i & (i = 0, 1, \ldots, N, N + 1), \\
(11.6) & \quad B \leq D^i & (i = 0, 1, \ldots, N, N + 1), \\
(11.7) & \quad 0 \leq F^i \leq 1 & (i = 1, \ldots, N), \\
(11.8) & \quad F^i \leq F^j & \text{if} \ h^i \leq h^j & (i, j = 1, \ldots, N), \\
(11.9) & \quad F^i = F^j & \text{if} \ h^i = h^j & (i, j = 1, \ldots, N).
\end{align*}
\]

This maximization is performed with respect to \(\{F^i\}_{i=1}^{N}\), \(\{h^i\}_{i=0}^{N+1}\), and \(\{D^i\}_{i=0}^{N+1}\). The values \(\{F^i\}_{i=1}^{N}\) are interpreted as the values of a distribution function, while the values \(\{h^i\}_{i=0}^{N+1}\) and the vectors \(\{D^i\}_{i=0}^{N+1}\) are interpreted as the values and subgradients at the points \(\{r^i\}_{i=0}^{N+1}\) of a systematic function.

The function in (11.1) is the conditional log-likelihood function in (3) with \(F(h(r^i))\) replaced by \(F^i\). The constraints in (11.7)–(11.9) characterize vectors \((F^1, \ldots, F^N)\) for which there exists \(F \in \Gamma\) with \(F^i = F(h^i)\) \((i = 1, \ldots, N)\). In Lemma B.3 in Appendix B, we show that constraints (11.2)–(11.6) characterize vectors \((h^0, \ldots, h^{N+1})\) for which there exists a function \(h\) in \(W\) with \(h(r^i) = h^i\) \((i = 0, \ldots, N + 1)\). After using (11.2), the vector \((h^0, \ldots, h^{N+1})\) can be eliminated, so the above problem can be solved for vectors \((D^0, \ldots, D^{N+1})\) and \((F^1, \ldots, F^N)\).

Hence, when \(W\) is the set of functions described in Example 1, one can obtain a solution to the maximization of (3) over \((W \times \Gamma)\) by finding a solution \((\hat{D}^0, \ldots, \hat{D}^{N+1}), \hat{F}^1, \ldots, \hat{F}^N)\) to the maximization of (11.1) subject to the constraints (11.2)–(11.9). Once this solution vector is obtained, one can derive a particular solution to the maximization of the likelihood function (3) over \((W \times \Gamma)\) by interpolating between the obtained values \((\hat{D}^0, \ldots, \hat{D}^{N+1})\) and
(\hat{E}^1, \ldots, \hat{E}^N). For example, we may interpolate linearly between \hat{E}^1, \ldots, \hat{E}^N and employ the following interpolation\footnote{This interpolation is based upon Afriat (1972).} for \hat{D}^0, \ldots, \hat{D}^{N+1}:

\hat{h}^N_M(r) = \min \{\hat{D}^i \cdot r | i = 0, 1, \ldots, N, N + 1\}.

The above interpolated function \(\hat{h}^N_M\) is a piecewise linear function that belongs to \(W\). Figure 3 illustrates the features of a typical function \(\hat{h}^N_M\) in \(R^2\) resulting from this interpolation. For comparison, Figure 4 shows the graph for an estimate of \(h^*\) that is restricted to be linear.

The maximization problem presented in (11.1)–(11.9) can be solved in two steps, analogous to the two steps suggested by Coslett (1983) to calculate his semiparametric distribution-free estimator. The first step proceeds by finding, for any given vector \((D^0, \ldots, D^{N+1})\), the vector \(\hat{F}^1, \ldots, \hat{F}^N\) that solves the following problem:

\[
\begin{align*}
(12) \quad \text{Maximize} \quad & \sum_{i=1}^{N} \left\{ y^i \log F^i + (1 - y^i) \log (1 - F^i) \right\} \\
\text{subject to} \quad & 0 \leq F^i \leq 1 \quad (i = 1, \ldots, N), \\
& F^i \leq F^j \quad \text{if} \quad D^i \cdot r^i \leq D^j \cdot r^j \quad (i, j = 1, \ldots, N), \\
& F^i = F^j \quad \text{if} \quad D^i \cdot r^i = D^j \cdot r^j \quad (i, j = 1, \ldots, N).
\end{align*}
\]

This can be done by employing the algorithm introduced by Asher et al. (1955).
(A description of this algorithm can be found also in Cosslett (1983) and Amemiya (1985).) The optimal value of the objective function of (12) that is obtained from this algorithm depends only on \( D^0, \ldots, D^{N+1} \). Denote this value by \( L(D^0, \ldots, D^{N+1}) \). In the second step, the function \( L(\cdot) \) is maximized over all vectors \( (D^0, \ldots, D^{N+1}) \) that satisfy (11.3)–(11.6).

To find a solution to the constrained optimization problem of the second step, one can employ a modification of the procedure used by Manski (1975) to maximize a discontinuous function over a finite dimensional vector (see also Cosslett (1983)). The modification is necessary because the values of the finite dimensional vector in our problem must satisfy the constraints of the problem.

To describe this modified algorithm, we let \( L(\cdot) \) denote, as before, the objective function of the second step maximization and \( Y \) denote the vector over which the maximization is performed. In the above example, \( Y = (D^0, \ldots, D^{N+1}) \).

The algorithm proceeds as follows. First, start from a vector \( Y^{(0)} \) inside the constraint set. Second, randomly draw a vector \( Y^{(1)} \). Third, determine the values \( \bar{b}, \bar{b} \in R \) such that the segment \([Y^{(0)} + \bar{b}Y^{(1)}, Y^{(0)} + \bar{b}Y^{(1)}] \) is the intersection of the line \( \{Y^{(0)} + \lambda Y^{(1)} \mid \lambda \in R \} \) with the constraint set. Fourth, using a line search procedure, find a vector \( Y^{(*)} \) at which \( L(\cdot) \) attains its maximum over the segment \([Y^{(0)} + \bar{b}Y^{(1)}, Y^{(0)} + \bar{b}Y^{(1)}] \). If \( L(Y^{(*)}) > L(Y^{(0)}) \), set \( Y^{(0)} = Y^{(*)} \). Fifth, go back to the second step. The algorithm stops, if for each of a prespecified number of consecutive times, the value of \( Y^{(0)} \) is the same when the second step starts. The value of \( Y^{(0)} \) is then taken to be the maximizer of the problem.

Matzkin (1991c) presents several techniques that increase the efficiency of this algorithm. These techniques take advantage of (i) the convexity of the
constraint set, (ii) the sparsity of the constraint set, and (iii) the fact that the vectors of subgradients of parametric functions that are concave and homogeneous of degree one belong to the constraint set.

When the set of functions in \( W \) possess properties different from those assumed in Example 1, constraints (11.2)–(11.6) need to be modified accordingly. (See Matzkin (1987) for the characterizations of vectors \( (h_1^*, \ldots, h_N^*) \) whose coordinates are the values of concave functions that possess various other properties.)

To indicate how the method works when the observations are from a binary choice model, we consider a binary choice model in which the set of functions \( U \) is as described in Example 5. The maximization of \( \mathcal{L} \) over \( (U \times \Gamma) \) can be solved also in this case by solving a constrained optimization problem. This constrained optimization problem is the following:

\[
(13.1) \quad \text{Maximize } \sum_{i=1}^{N} \left\{ y^i \log F^i + (1 - y^i) \log (1 - F^i) \right\} \quad \text{subject to }
\]

\[
(13.2) \quad \beta_j \leq w^i_j \leq \beta_U \quad (i = 1, \ldots, N; \ j = 1, 2),
\]

\[
(13.3) \quad w^i_j = w^i_k \quad \text{if } z^i_j = z^i_k \quad (i, r = 1, \ldots, N; \ j, k = 1, 2),
\]

\[
(13.4) \quad w^i_j = 0 \quad \text{if } z^i_j = 0 \quad (i = 1, \ldots, N; \ j = 1, 2),
\]

\[
(13.5) \quad F^i \leq F^j \quad \text{if } \left(s^i + w^i_1 - w^i_2\right) < \left(s^j + w^j_1 - w^j_2\right) \quad (i, j = 1, \ldots, N),
\]

\[
(13.6) \quad F^i = F^j \quad \text{if } \left(s^i + w^i_1 - w^i_2\right) = \left(s^j + w^j_1 - w^j_2\right) \quad (i, j = 1, \ldots, N).
\]

The calculation of this problem can proceed in two steps, like the calculation of the problem described by (11.1)–(11.9).

In conclusion, in this section we have shown how the problem of maximizing a likelihood function over a set of pairs of nonparametric functions can be transformed into a feasible problem.

7. SUMMARY

We showed that it is possible to identify and consistently estimate binary threshold crossing models and binary choice models without assuming any parametric structure either for the systematic functions of observable exogenous variables or for the distribution of the unobservable random term. The most critical assumptions used to prove these results were shown to be often implied by economic theory. In particular, in binary threshold crossing models, the systematic function and the distribution of the random term can be identified within a set \( (W \times \Gamma) \), where \( W \) is the set of continuous, monotone increasing, and homogeneous of degree one functions on \( \mathbb{R}_+^k \), that attain a value \( \alpha \) at a point \( r^* \) of their domain and possess subgradients that are uniformly bounded away from zero; \( \Gamma \) denotes the set of monotone increasing functions on \( \mathbb{R} \) with values in \([0, 1]\). The assumption that all functions in \( W \) attain a common value at a common point was made to fix the scale of the functions.
The identification results were employed to develop a strongly consistent estimator for each model. The estimators are obtained by maximizing a likelihood function over a set of nonparametric functions and distributions. A two step procedure to calculate these estimators was presented.

The estimators presented in this paper constitute only one example of the applicability of our identification results. Various other estimators that apply these identification results can also be constructed (see, for example, Matzkin (1990c, 1991b)).

**Cowles Foundation for Research in Economics, Department of Economics, Yale University, P.O. Box 2125 Yale Station, New Haven, CT 06520-2125, U.S.A.**

*Manuscript received November, 1988; final revision received June, 1991.*

**APPENDIX A**

**Proof of Theorem 1.** Let \( p^*: T \rightarrow R \) be defined by, for all \( r \in T \),

\[
(1.1) \quad p^*(r) = F^*(h^*(r)).
\]

and let \( p: T \rightarrow R \) be such that for some \( (h, F) \in (W \times F) \), and all \( r \in T \),

\[
(1.2) \quad p(r) = F(h(r)).
\]

Suppose that

\[
(1.3) \quad p(r) = p^*(r) \quad \text{a.s. } [G].
\]

Theorem 1 will be proved if we show that

\[
(1.4) \quad F(t) = F^*(t) \quad \text{for all } \quad t \in h^*(T),
\]

except perhaps for a subset that possesses zero Lebesgue measure, and

\[
(1.5) \quad h(r) = h^*(r) \quad \text{for all } \quad r \in T,
\]

except perhaps for a subset that possesses zero probability w.r.t. \( G \).

To show (1.4) and (1.5), let \( N \subset T \) be such that, for all \( r \in N \), \( p(r) = p^*(r) \). Then, by (1.3), \( G(N^c) = 0 \), where \( N^c \) denotes the complement of \( N \).

Let \( C = \{ t \in h^*(T) | t \text{ is a point of continuity of both } F \text{ and } F^* \} \).

We will show that our assumptions imply that

\[
(1.6) \quad \text{for all } t \in C \quad F^*(t) = F(t).
\]

To show (1.6), let \( t \in C \). Then, by Assumption W.2, there exists \( r' \in \overline{T} \subset T \) such that

\[
(1.7) \quad h(r') = h^*(r') = t.
\]

Since \( G(N^c) = 0 \) and \( r' \in T \) it follows from Assumption G.1 that in any neighborhood of \( r' \) there exists \( r' \in N \). This establishes the existence of a sequence \( \{ r_k \}_{k=1}^\infty \subset N \) such that \( r_k \rightarrow r' \). By (1.2), the definition of \( N \), and (1.1), it follows that

\[
(1.8) \quad F(h(r_k)) = p(r_k) = p^*(r_k) = F^*(h^*(r_k)) \quad \text{for } \quad k = 1, 2, \ldots.
\]

Since \( r_k \rightarrow r' \) and both \( h \) and \( h^* \) are continuous at \( r' \), the continuity of \( F \) and \( F^* \) at \( t \in C \) imply that

\[
(1.9) \quad F(h(r_k)) \rightarrow F(h(r')) \quad \text{and} \quad F^*(h^*(r_k)) \rightarrow F^*(h^*(r')).
\]
Hence, by (1.8) and (1.9),
\[ F^*(h^s(r')) = \lim_{k \to \infty} F^*(h^s(r_k)) = \lim_{k \to \infty} F(h(r_k)) = F(h(r')). \]
Thus, by (1.7),
\[ F^*(t) = F(t). \]
Since \( t \in C \), this shows that (1.6) holds. (If \( r' \in N \), then (1.6) holds without any continuity argument.)

Now (1.4) follows immediately from (1.6) because the monotonicity of \( F^* \) and \( F \) implies that \( C \) possesses zero Lebesgue measure.

To show that (1.5) is satisfied, we let
\[ NC = \{ r \in N \mid h^s(r) \text{ and } h(r) \text{ are points of continuity of both } F \text{ and } F^* \}. \]
We will show that
\[ (1.10) \quad \text{for all } r \in NC \quad h(r) = h^s(r). \]
To show (1.10), let \( r \in NC \). Then,
\[ (1.11) \quad p(r) = p^*(r). \]
By (1.1), (1.11), (1.2), (1.6), and the fact that \( h^s(r) \) and \( h(r) \) are points of continuity of both \( F \) and \( F^* \),
\[ (1.12) \quad F^*(h^s(r)) = p^*(r) = p(r) = F(h(r)) = F^*(h(r)). \]
Since \( F^* \) is strictly increasing, (1.12) implies that \( h^s(r) = h(r) \).
This proves (1.10).

We next show that \( NC \) is dense in \( T \). This, together with (1.10) and the continuity of \( h \) and \( h^s \), will imply that (1.4) is satisfied.
Let \( r \in T \). By Assumption G.2, there exists \( \delta^* > 0 \) such that either \( [r, r + \delta^*] \subset T \) or \( [r - \delta^*, r] \subset T \), where \( \delta^*_k \) denotes the vector with 0's in all coordinates other than \( K \) and with a 1 in the \( K \)th coordinate. Suppose w.l.o.g. that \( [r, r + \delta^*_1] \subset T \) and let \( A = [r, r + \delta^*_1] \). Since \( h^s \) is continuous and strictly increasing with respect to its \( K \)th coordinate, there exists \( (r_k) \subset A \) and \( \delta_k \subset R_+ \) such that
\[ (i) \quad r_k \to r, \]
\[ (ii) \quad \delta_k \to 0, \quad \text{and} \]
\[ (iii) \quad \text{for all } r_k^{(k)} \in B(r_k, \delta_k) \cap T, r_{k+1}^{(k)} \in B(r_k, \delta_k) \cap T \]
\[ h^s(r) < \cdots < h^s(r_k^{(k)}) < h^s(r_{k+1}) < \cdots. \]
Since \( r_k \in T \) and \( G(N^c) = 0 \), it follows by Assumptions G.1–G.2 that for all \( k = 1, 2, \ldots \) there exists \( r_k^{(k)} \in B(r_k, \delta_k) \) such that the set
\[ \{ r_k^{(k)} + \delta_k | - \infty < \delta < \infty \} \cap B(r_k, \delta_k) \cap N \]
is uncountable. Let
\[ C(r_k^{(k)}) = \{ r_k^{(k)} + \delta_k | - \infty < \delta < \infty \} \cap B(r_k, \delta_k) \cap N. \]
Since by Assumption W.4 \( h^s \) is strictly increasing on \( C(r_k^{(k)}) \), \( h \) is also strictly increasing on \( C(r_k^{(k)}) \). To see this, let \( r_k, r_k \in C(r_k^{(k)}) \) be such that \( r_k < r_k \). Then, the strict monotonicity of \( F^* \) on \( h^s(T) \) and the strict monotonicity of \( h^s \) on \( C(r_k^{(k)}) \) imply that \( p^*(r_k) < p^*(r_k^{(k)}) \). Thus, the definition of \( C(r_k^{(k)}) \), (1.1), and (1.2) imply that \( F(h(r_k)) = p(r) = p^*(r_k) < p^*(r_k^{(k)}) = F(h(r_k^{(k)})) \). Hence, since Assumption I.3 and (1.4) imply that \( F \) is strictly increasing on \( h^s(T) \), \( h(r_k) < h(r_k^{(k)}) \). Therefore, \( h \) is also strictly increasing on \( C(r_k^{(k)}) \).

Hence, since both \( h^s \) and \( h \) are strictly increasing on the uncountable set \( C(r_k^{(k)}) \), they each map \( C(r_k^{(k)}) \) onto an uncountable set. Since the set of points that are points of discontinuity of either \( F \) or \( F^* \) is a countable set, it follows that, for some \( r_k \in C(r_k^{(k)}) \), \( h(r_k) \) and \( h(r_k) \) are points of continuity of \( F \) and \( F^* \). This shows that for all \( k = 1, 2, \ldots \) there exists \( r_k \in B(r_k, \delta_k) \cap NC \). Moreover, \( r_k \to r \), since \( r_k \to r \) and \( \delta_k \to 0 \). Hence, \( NC \) is dense in \( T \).
Since for all \( k = 1, 2, \ldots \) and all \( \tilde{r}_k, h^*(\tilde{r}_k) = h(\tilde{r}_k) \) and since \( h \) and \( h^* \) are continuous, it follows that
\[
h^*(r) = \lim_{k \to \infty} h^*(\tilde{r}_k) = \lim_{k \to \infty} h(\tilde{r}_k) = h(r).
\]
Hence, for all \( r \in T, h^*(r) = h(r) \). This shows (1.4).

Q.E.D.

Proof of Theorem 2: To prove Theorem 2, we will employ an adaptation of the result of Wald (1949), as presented in Matzkin (1991a) (see also Kiefer and Wolfowitz (1956)). To employ this result we first need to show some lemmas. Most of these lemmas correspond to assumptions made in Wald (1949).

We let \( x \) denote the vector of observable variables \((y, r)\) and \( f \) denote the probability density of \( x \). Then, for any \( x = (y, r) \in [0, 1] \times T \) and any \((h, F) \in (W \times \Gamma)\),
\[
f(x; h, F) = \tilde{g}(r) \left[ F(h(r)) \right]^{1 - F(h(r))},
\]
where \( \tilde{g}(r) \) is the conditional probability density of \( r \), conditional on \( r \in T \). Let \( \tilde{G} \) denote the probability measure induced by \( \tilde{g} \). We will denote the support of \( f \) by \( X \) and the probability measure induced by \( f : h^*, F^* \) by \( P^* \). We define functions \( f' \) and \( f^* \) by
\[
f'(x; h, F, \varepsilon) = \sup \{ f(x; h', F') \mid (h', F') \in (W \times \Gamma) \} \quad \text{and}
\]
\[
f^*(x; h, F, \varepsilon) = \begin{cases} f'(x; h, F, \varepsilon) & \text{if } f'(x; h, F, \varepsilon) > 1, \\ 1 & \text{otherwise}. \end{cases}
\]
where \( \varepsilon > 0 \).

We next present the Lemmas.

Lemma 1 (Compactness): If Assumptions W.5 and \( \Gamma.1 \) are satisfied, then \((W \times \Gamma)\) is compact with respect to \( m \).

Proof: Since \( \Gamma \) is compact with respect to \( d_{\varepsilon} \) (Cossett (1983)) and \( W \) is compact with respect to \( d_{\varepsilon_{\infty}} \), \((W \times \Gamma)\) is compact with respect to \( m \). Q.E.D.

Lemma 2 (Assumption 3 in Wald (1949)): Suppose that Assumptions W.1, W.4, W.6, G.2, and \( \Gamma.1 \) are satisfied. If \((h_1, F_1)_{\varepsilon=0}^{\infty} \) is a sequence in \((W \times \Gamma)\) such that \((h_1, F_1) \to (h, F) \in (W \times \Gamma)\), then \( f(x; h_1, F_1) \to f(x; h, F) \) for all \( x \in X \), except possibly on a subset of \( X \) of probability measure zero \( \{P^*\} \).

Proof: Since, by Assumption W.6, convergence of \( h_1 \) to \( h \) implies that \( \forall r \in T, h_1(r) \to h(r) \), it follows (see Cossett (1983)) that if \( h(r) \) is a point of continuity of \( f, f(x; h_1, F_1) \to f(x; h, F) \). By Assumptions W.1, W.4, and G.2, \( h \) possesses an absolutely continuous distribution with respect to \( G \). Hence, since \( F \) has at most a countable number of discontinuities, the subset of \( T \) at which convergence is guaranteed has unit probability measure \( \{P^*\} \). Q.E.D.

Lemma 3 (Assumption 8 in Wald (1949)): Suppose that Assumptions W.1, W.5, W.6, and \( \Gamma.1 \) are satisfied. Then, for any \((h, F) \in (W \times \Gamma)\) and any small enough \( \varepsilon > 0 \), the function \( f'(x; h, F, \varepsilon) \) is measurable in \( x \).

Proof: Since the set \( W \) is compact, there exists a countable dense subset \( Q \subset W \). Let \( \Gamma' \subset \Gamma \) be the countable set of all step functions in \( \Gamma ' \) whose (finite number of) jumps are rational and occur at

\[\text{Lemma 1 in Appendix B of Cossett (1983) states that if } \{F_n\} \text{ is a sequence in } \Gamma \text{ such that } F_n \to F \in \Gamma, \{\eta_n\} \text{ a sequence of real numbers such that } \eta_n \to \eta, \text{ and } \eta \text{ is a point of continuity of } F, \text{ then } F(\eta_n) \to F(\eta).\]
rational points. We will show that, for all \(x \in X\),

\[(L.3.1) \quad t = \sup \{ f(x; h', F') \mid (h', F') \in (W \times I), \ m((h, F), (h', F')) < \varepsilon \} \]

\[= \sup \{ f(x; h', F') \mid (h', F') \in (Q \times I), \ m((h', F'), (h, F)) < \varepsilon \} = u. \]

Clearly, \(t \geq u\). Suppose \(t > u\); then, for some \((h', F') \in (W \times I)\), some \(\alpha \in R\), and \(W(h', F') \in (Q \times I^*)\), either

\[(L.3.2) \quad F'(h'(r)) > \alpha > F'(h(r)), \quad \text{or} \]

\[(L.3.3) \quad F'(h'(r)) < \alpha < F'(h(r)). \]

Suppose w.l.o.g. that (L.3.2) holds. Since \(Q\) is dense in \(W\), there exists a sequence \(\{h^i\} \subset Q\) such that \(h^i \to h\) with respect to \(d_y\) or, hence, by Assumption W.6, \(h^i(r) \to h(r)\). Then, there exists a sequence \(\{F^i\} \subset I^*\) and a subsequence \(\{h^i\}\) of \(\{h^i\}\) such that \(F^i(h^i(r)) \to F(h(r))\) (Cossett (1983)).\(^10\) This contradicts (L.3.2); hence \(t = u\).

Since \(h^i\) is continuous and \(F^i\) is increasing, \(F(h(r))\) is measurable on \(T\). Hence, \(f(x; h^i, F^i)\) is measurable on \(X\). Consequently, the supremum is measurable too. By (L.3.1), \(f^*\) is measurable in \(x\).

**Lemma 4 (Assumption 6 in Wald (1949)):** If Assumptions G.3 and \(\Gamma.2\) are satisfied,

\[\int_X |\log f(x; h^*_x, F^*_x)| dP^*_x(x) < \infty. \]

**Proof:** By the functional structure of \(f(x; h^*_x, F^*_x)\)

\[\int_X |\log f(x; h^*_x, F^*_x)| dP^*_x(x) \leq \int_T |\log \tilde{g}(r)| d\tilde{G}(r)\]

\[+ \int_T |\log F^*(h^*_x(r))| [F^*(h^*_x(r))] d\tilde{G}(r)\]

\[+ \int_T |\log [1 - F^*(h^*_x(r))]| [1 - F^*(h^*_x(r))] d\tilde{G}(r). \]

Since the ranges of \(F^*_x\) and \(1 - F^*_x\) are included in the interval \([0, 1]\), and since the function \(q(y) = y \log(y)\) has a bounded range on that interval, the last two integrals are bounded. Since the first integral is bounded by Assumption G.3, it then follows that

\[\int_X |\log f(x; h^*_x, F^*_x)| dP^*_x(x) < \infty. \quad Q.E.D. \]

**Lemma 5 (Assumption 2 in Wald (1949)):** Suppose that Assumptions W.1, W.5, W.6, \(\Gamma.1\), \(\Gamma.2\), and G.3 are satisfied. Then, for sufficiently small \(\varepsilon > 0,\)

\[\int_X \log f^*(x; h, F, \varepsilon) dP^*_x(x) \quad \text{is finite}. \]

**Proof:** Let

\[C = \{x \in X| f^*(x; h, F, \varepsilon) < 1\}, \]

\[D = \{x \in X| f^*(x; h, F, \varepsilon) \geq 1\}, \quad \text{and} \]

\[E = \{r \in T| \tilde{g}(r) > 1\}. \]

From the definitions of \(f^*_x\), \(f'_x\), and \(\Gamma\) it follows that for all \(x = (y, r) \in D, \tilde{g}(r) > 1\) and

\(^{10}\) Lemma 2 in Cossett (1983) states that if \(F \in \Gamma, \eta \in R, \) and \((\eta_i)\) is a sequence converging to \(\eta,\)

there exists a sequence \((F_i) \subset \Gamma^*\) and a subsequence \((\eta_{i(k)})\) such that \(F_i(\eta_{i(k)}) \to F(\eta).\)
\[ 0 \leq \log f^*(x; h, F, \varepsilon) \leq \log \bar{g}(r). \] Hence, since by Lemma 3 \( f^*(x; h, F, \varepsilon) \) is measurable,
\[
\int_{\mathcal{X}} \log \left( f^*(x; h, F, \varepsilon) \right) dP^*(x)
\]
\[
- \int_{\mathcal{C}} \log \left( f^*(x; h, F, \varepsilon) \right) dP^*(x) + \int_{\mathcal{D}} \log \left( f^*(x; h, F, \varepsilon) \right) dP^*(x)
\]
\[
\leq \int_{\mathcal{D}} \log \left( f^*(x; h, F, \varepsilon) \right) dP^*(x)
\]
\[
\leq \int_{\mathcal{E}} \log (\bar{g}(r)) d\bar{G}(r).
\]

Since this last integral is bounded, by Assumption G.3, it follows that
\[
\int_{\mathcal{X}} \log f^*(x; h, F, \varepsilon) dP^*(x)
\]

is finite. \( \text{Q.E.D.} \)

Theorem 2 can now be proved from these lemmas as in Matzkin (1991a).

**Proof of Theorem 3**: Assumptions U.1, U.3–U.5, F.1–F.3, and G.1'–G.2' imply by Theorem 1 that \((h^*_e, F^*)\) is identified within \( (W_e \times \Gamma) \).

It then remains to show that if \( V \in U \) is such that for all \( (s, z_1, z_2) \in T \)
\begin{equation}
(3.1)
\quad h^*_e(s, z_1, z_2) = V_l(s, z_1) - V_2(s, z_2),
\end{equation}
then \( (V_1, V_2) = (V^*_1, V^*_2) \). To show this, we note that by Assumptions U.2 and U.4, for all \( (s, z_1) \in T_3 \times T_2 \), \( V_l(s, z_1) = h^*_e(s, z_1, z_2) + \gamma = V^*_1(s, z_1) \). Hence,
\begin{equation}
(3.2)
\quad V_1 = V^*_1.
\end{equation}

Next, by (3.1) and (3.2), for all \( (s, z_2) \in T_3 \times T_2 \), and any \( z_1 \in T_2 \),
\[
V_2(s, z_2) = V_l(s, z_1) - h^*_e(s, z_1, z_2) - V^*_1(s, z_1) - h^*_e(s, z_1, z_2) = V^*_2(s, z_2).
\]

Hence,
\begin{equation}
(3.3)
\quad V_2 = V^*_2.
\end{equation}

So, \( V^* \) can be uniquely recovered from \( h^*_e \). Since \((h^*_e, F^*)\) is identified within \( (W_e \times \Gamma) \), this implies that \((V^*, F^*)\) is identified within \( (U \times \Gamma) \). \( \text{Q.E.D.} \)

**Appendix B**

**Lemma B.1**: Suppose that \( T_e = R_+^k \). Let \( r^* \in T_e \), \( \alpha > 0 \), and \( B > 0 \) be given. Let \( W \) be the set of all continuous, monotone increasing, concave, and homogeneous of degree one functions, \( h: T_e \to R \), such that for each \( h \in W \), \( h(r^*) = \alpha \) and for each \( r \in T_e \), the Kth coordinate of a subgradient of \( h \) at \( r \) is bounded below by \( B \). Suppose that \( r \) is distributed with an absolutely continuous distribution whose support is \( R_+^k \). Then \( W \) is compact with respect to \( d_W \).

**Proof**: Let \( (h_k)_{k=1}^\infty \) be a sequence in \( W \). We need to show that \( (h_k)_{k=1}^\infty \) has a subsequence that converges to an element of \( W \).

Define the function \( h: T_e \to R \) by \( h(r) = \inf\{T_\alpha | r \leq T_\alpha^*\} \). Since for any \( h \in W \) and any \( r \in T_e \), \( 0 \leq h(r) \leq h(r^*) = \alpha \), where \( \gamma = \inf\{T_\gamma | r \leq T_\gamma^*\} \), it follows that for any \( h \in W \) and any \( r \in T_e \), \( h(r) \leq h(r) \). Hence, the functions in \( W \) are pointwise bounded.
Let \( Q \) denote a countable dense subset in \( T_r \). By the usual diagonalization process (see, e.g., the proof of Helly's Theorem), there exists a subsequence \( \{ h_{k_n} \} \) of \( \{ h_k \} \) and a function \( h: Q \to \mathbb{R} \) such that for all \( q \) in \( Q \), \( h_{k_n}(q) \to h(q) \). By Rockafellar (1972, Theorem 10.8) there then exists a concave function \( h: T_r \to \mathbb{R} \) such that for all \( r \in T_r \), \( h_{k_n}(r) \to h(r) \). Since \( h \) is the pointwise limit of monotone increasing and homogeneous of degree one functions that attain the value \( a \) at \( r^n \), \( h \) is monotone increasing, homogeneous of degree one, and it attains the value \( a \) at \( r^n \). Since \( h \) is concave on \( T_r \), \( h \) is continuous.

To prove that \( h \) belongs to \( W \), it only remains to show that, for each \( r \in T_r \), \( h \) possesses a subgradient whose \( K \)th coordinate is bounded below by \( B \). To show this we will first prove

(B.1.1) \[ \text{for all } r \in T_r \text{ there exists } \gamma > 0 \text{ such that for all } h \in W \text{ and all subgradients } D h(r) \text{ of } h \text{ at } r, \|D h(r)\| \leq b(\gamma r). \]

Let \( r \in T_r \), \( h \in W \), and \( D h(r) \) be a subgradient of \( h \) at \( r \). Then, for all \( \lambda > 0 \) one has that

\[
h'(\lambda r) = h'(r) + D h(r) \cdot (\lambda r - r),
\]

which by the homogeneity of degree one of \( h' \) implies that

\[
(\lambda - 1) h'(r) \leq (\lambda - 1) D h(r) \cdot r.
\]

Letting \( \lambda \) be first bigger than 1 and next smaller than 1, it follows that

(B.1.2) \[ h'(r) = D h(r) \cdot r. \]

This implies, in particular, that for all \( \lambda > 0 \) one has that \( D h(r) \) is a subgradient of \( h' \) at \( \lambda r \). To see this, note that for any \( s \in T_r \),

\[
h'(s) = h'(r) + D h(r) \cdot (s - r) = D h(r) \cdot s = h'(\lambda r) + D h(r) \cdot (s - \lambda r).
\]

Hence,

(B.1.3) \[ \text{for all } \lambda > 0, D h(r) \text{ is a subgradient of } h' \text{ at } \lambda r. \]

Let \( \gamma > 0 \) be such that \( \gamma r > 1 \), where 1 denotes the vector \((1, 1, \ldots, 1) \in \mathbb{R}^K \). Since \( h' \) is monotone increasing, for each \( k \in \{1, \ldots, K\} \), \( D_k h(r) > 0 \), where \( D_k h(r) \) denotes the \( k \)th coordinate of \( D h(r) \). Hence, by the definition of \( \gamma \) it follows that

\[
D h(r) \cdot 1 \leq D h(r) \cdot \gamma r.
\]

By (B.1.2), (B.1.3), and the definition of \( b(\cdot) \) it then follows that

\[
\|D h(r)\| \leq \left( \sum_k D_k h(r)^2 \right)^{1/2} \leq D h(r) \cdot \gamma r = h'(\gamma r) \leq b(\gamma r).
\]

This proves (B.1.1).

Now, to prove that the \( K \)th coordinate of a subgradient of \( h' \) at \( r \) is bounded below by \( B \), we let \( D h_k(r) \) denote, for each function \( h_k \), the subgradient of \( h_k \) at \( r \) such that the \( K \)th coordinate of \( D h_k(r) \) is bounded below by \( B \). By (B.1.1) the sequence \( D h_k(r) \) is bounded. Hence, there exists a convergence subsequence \( \{ D h_{k_n}(r) \} \) of \( \{ D h_k(r) \} \). Let \( D(r) \) denote the limit of this subsequence. Then, \( D_k(r) \geq B \) and since \( h_k(r) \) converges to \( h(r) \) it follows that \( D(r) \) is a subgradient of \( h \).

This completes the proof that \( h \in W \).

It remains to show that \( h_{k_n} \to h \) with respect to \( d_B \). Let \( \varepsilon > 0 \) be given, and let \( K \) be a large enough compact set in \( T_r \) such that

\[
\int_K |D h(r)| \exp(-|r|d) dG(r) < \varepsilon/2.
\]
Since the functions in $\mathcal{W}$ are bounded by zero and the function $b$, for all $k$,

$$(B.1.4) \quad \int_{K} |h_{k}(r) - h(r)| \exp(-\|r\|) dG(r)$$

$$\leq \int_{K} |b(r)| \exp(-\|r\|) dG(r)$$

$$< \epsilon/2.$$  

Let $\zeta \in R$, be such that $\forall r \in K$, $r < \zeta r^*$. Since the functions $h_{k}$ and $h$ are monotone increasing, they are uniformly bounded on $K$ by $\zeta a$. By Lebesgue's Dominated Convergence Theorem (Billingsley (1986, pp. 213)), it then follows that for all large enough $u$,

$$(B.1.5) \quad \int_{K} |h_{k}(r) - h(r)| \exp(-\|r\|) dG(r) < \epsilon/2.$$  

From (B.1.4)-(B.1.5) and the definition of $d_{W}$, it then follows that $d_{W}(h_{k}, h) \to 0$ as $u \to \infty$. Hence, $\{h_{k}\}$ has a subsequence that converges to an element of $\mathcal{W}$.

**LEMMA B.2:** Let $T_{1} = T_{1}^1 \times T_{1}^2 \subset R^{K}$, where $T_{1}^1 \subset R^{K_{1}}$ and $T_{1}^2 \subset R^{K_{2}}$. Let $W_{1}$ be a set of continuous functions $v: T_{1} \to R$, let $W_{2}$ be a set of continuous functions $w: T_{2} \to R$, and let $W$ be the set of functions $V: T \to R$ for which there exist functions $v \in W_{1}$ and $w \in W_{2}$ such that for all $r = (r_{1}, r_{2}) \in (T_{1}^1 \times T_{1}^2)$, $V(r_{1}, r_{2}) = v(r_{1}) + w(r_{2})$. Let $T = T_{1} \cap S_{G}$, where $S_{G}$ is the support of the probability measure, $G$, of $r$. For any $u, u' \in W_{1}$ and $w, w' \in W_{2}$, let

$$d_{W}(u, u') = \int_{T} |v(r_{1}) - u'(r_{1})| e^{-\|r\|} dG(r)$$

and

$$d_{W}(w, w') = \int_{T} |w(r_{2}) - w'(r_{2})| e^{-\|r\|} dG(r).$$

Suppose that $G$ is such that for all $r \in T$ and all $\delta > 0$, $G(B(r, \delta) \cap T) > 0$. Then, if $W_{1}$ and $W_{2}$ are compact, respectively, with respect to the metrics $d_{W_{1}}$ and $d_{W_{2}}$, then $W$ is compact with respect to the metric $d_{W}$.

**PROOF:** Let $\{V_{n}\}_{n=1}^{\infty}$ be a sequence in $W$. We need to show that $\{V_{n}\}_{n=1}^{\infty}$ possesses a subsequence that converges to a function in $W$. Since $V_{n} \in W$, there exists $u_{n} \in W_{1}$ and $w_{n} \in W_{2}$ such that for all $r = (r_{1}, r_{2}) \in T$, $V_{n}(r_{1}, r_{2}) = u_{n}(r_{1}) + w_{n}(r_{2})$ for $n = 1, 2, \ldots$. Since $W_{1}$ is compact with respect to $d_{W_{1}}$, there exists a subsequence $\{\hat{u}_{n}\}$ of $\{u_{n}\}$ and a function $v \in W_{1}$ such that $\hat{u}_{n} \to v$ with respect to $d_{W_{1}}$. Since $W_{2}$ is compact with respect to $d_{W_{2}}$, there exists a subsequence $\{\hat{w}_{n}\}$ of $\{w_{n}\}$ and a function $w \in W_{2}$ such that $\hat{w}_{n} \to w$ with respect to $d_{W_{2}}$. Since $\hat{u}_{n} \to v$, $\hat{w}_{n} \to w$, let $V: T \to R$ be defined by $V(r_{1}, r_{2}) = u_{n}(r_{1}) + w_{n}(r_{2})$ for all $(r_{1}, r_{2}) \in T$. Then, $V \in W$. Moreover, since

$$d_{W}(V_{n}, V) = \int |V_{n}(r_{1}) + w_{n}(r_{2}) - v(r_{1}) - w(r_{2})| e^{-\|r\|} dG(r)$$

$$< \int |u_{n}(r_{1}) - v(r_{1})| e^{-\|r\|} dG(r) + \int |w_{n}(r_{2}) - w(r_{2})| e^{-\|r\|} dG(r),$$

and $u_{n} \to v$ and $w_{n} \to w$ with respect to $d_{W_{1}}$ and $d_{W_{2}}$ respectively, it follows that $V_{n} \to V$ with respect to $d_{W}$. Hence, $\{V_{n}\}$ possesses a subsequence that converges to a function in $W$. Since $(W, d_{W})$ is a metric space, this implies that $W$ is compact with respect to $d_{W}$.

**LEMMA B.3:** Suppose that $r^{1}, \ldots, r^{N}$ belong to $R^{K}$. Let $r^{0} = 0$ and $r^{N+1} = r^{*}$. Then, the set of all vectors $(h^{0}, h^{1}, \ldots, h^{N}, h^{N+1})$, for which there exists a continuous, monotone increasing, concave, and homogeneous of degree one function $h: R^{K} \to R$ such that $h(r^{*}) = a$, $h(r^{0}) = h^{0}$, and, for each $i$, $h^{i}$ possesses a subgradient whose $i$th coordinate is bounded below by $B$, is the set of all vectors $(h^{0}, h^{1}, \ldots, h^{N}, h^{N+1})$ satisfying (11.2)-(11.6) in Section 5.

---

**Note:** This lemma is a modification of Afriat's (1972) results.
Proof: We will first prove that if \( h \) is continuous, monotone increasing, concave, homogeneous of degree one, \( h(r^*) = \alpha \), and for each \( r \) the \( K \)th coordinate of a subgradient of \( h \) at \( r \) is bounded below by \( B \), then the values of \( h \) at \( r^0, r^1, \ldots, r^N, r^{N+1} \) satisfy \((11.2)-(11.6)\) for some vectors \( D^0, \ldots, D^{N+1} \).

Let \( D^0, \ldots, D^{N+1} \subset R^K \) denote, respectively, the subgradients of \( h \) at \( 0, r^1, \ldots, r^N, r^{N+1} \) such that the \( K \)th coordinate of \( D^i \) \((i = 0, 1, \ldots, N + 1) \) is bounded below by \( B \). Since \( h \) is concave,

\[
h(r^i) \leq h(r^i) + D^i \cdot (r^i - r^i) \quad (i, t = 0, 1, \ldots, N, N + 1).
\]

Since \( h \) is homogeneous of degree one (see the proof of \((11.2)\) in Lemma B.1),

\[
h(r^i) = D^i \cdot r^i \quad (i = 0, 1, \ldots, N, N + 1).
\]

Since \( h \) is monotone increasing,

\[
D^i \geq 0 \quad (i = 0, 1, \ldots, N, N + 1).
\]

And since the \( K \)th coordinates of \( D^0, \ldots, D^{N+1} \) are bounded below by \( B \),

\[
D^i_K \geq B \quad (i = 0, 1, \ldots, N, N + 1).
\]

By \((11.4)\) and \((11.4)\),

\[
D^i \cdot r^i \leq D^i \cdot r^i \quad (i, t = 0, 1, \ldots, N, N + 1).
\]

Hence, \( h(r^0), h(r^1), \ldots, h(r^N), h(r^{N+1}) \) satisfy \((11.2)-(11.6)\) for some vector \( D^0, \ldots, D^{N+1} \). Define the function \( h: R^N \rightarrow R \) by

\[
h(r) = \min \{ D^i \cdot r | i = 0, \ldots, N + 1 \}.
\]

Then, this piecewise linear function is continuous, monotone increasing, concave, and homogeneous of degree one on \( R^N \). Moreover,

\[
h(r^*) = D^{N+1} \cdot r^{N+1} - h^{N+1} = \alpha,
\]

and, since for each \( r \) a subgradient of \( h \) at \( r \) is some \( D^i \), \( h \) possesses at each \( r \) a subgradient whose \( K \)th coordinate is bounded below by \( B \).

\[Q.E.D.\]

REFERENCES


