Identification in Binary Response Models

James J. Heckman
Source: Manski, 1985, 1988; Matzkin, 1992; Cosslett, 1983

February 3, 2004
$D$ determined by $x \in X \subseteq R^K$
$u \in U \subseteq R^l$

$D = 1(x\beta + u \geq 0)$ i.e. $= 1$ if $x\beta + u \geq 0$
$= 0$ otherwise

$$\Pr(D = 1|x) = \int 1[x\beta + u \geq 0] \, dF_{u|x}(u)$$

Observe that all info we have is $(x, D) \Rightarrow \Pr(D = 1|x) = \int 1[x\beta + u] \, dF_{u|x}(u)$. 

1
Seek information about

(a) $\beta$
(b) distribution of $U$ (joint with dist. of $x$).

We will assume $F_{u|x}$ is strictly increasing in $U$ a.e. $x$.

Model is *always* indistinguishable from one with $m(x\beta,u) \geq 0$ whenever $x\beta + u \geq 0$ (sign preserving)
Examples of Equivalent Models

(i) \( m_\gamma(x\beta + u) - m_\gamma(0); \) \( m_\gamma \) is a strictly increasing function
(ii) \(-m_\gamma(-x\beta) + m_\gamma(u)\)
(iii) \( m_\gamma(x\beta) + m_\gamma(u) \) \( \text{for } m_\gamma \text{ antisymmetric} \)

(i) is obviously true;
(ii) \( x\beta + u \geq 0 \iff u \geq -x\beta \)
\( \iff m_\gamma(u) \geq m_\gamma(-x\beta) \)
\( \Rightarrow -m_\gamma(-x\beta) + m_\gamma(u) \geq 0 \)
(iii) antisymmetry \( \Rightarrow -m_\gamma(-x\beta) = m_\gamma(x\beta) \) and result follows.

All of these models are indistinguishable using the \( (D, X) \) data.

In this sense \( \exists \) a fundamental nonidentification problem
Issues in identification

(i) distribution of the $X$, $F_x$ known, as well as $\Pr(D = 1|x)$
(ii) other info may be available on $(\beta, F_{u|x})$ (positive parameter values)

Identification:

A Formal Definition

$\Phi$ is space of all possible probability distributions

$\Phi^X$ : Cartesian Product of $\Phi$ over the space $X$
It is assumed that

\[(\beta, F_{u|x}) \in \Omega \subset R^K \times \Phi^X \quad \text{(true values)}\]

Consider other parameter configurations. For each \((b, G_{u|x}) \equiv (x \in X : \Pr(D = 1|x))\) in \(R^K \times \Phi^X\), define

\[X(b, G_{u|x}) \equiv \{x \in X : \Pr(D = 1|x) \neq \int 1\{xb + u \geq 0\} \ dG_{u|x}\}\]

Then \(\beta, F_{u|x}\) is identified relative to \((b, G_{u|x})\) if \((b, G_{u|x}) \notin \Omega\) (so it is ruled out \textit{a priori}) or else

\[\Pr(x \in X(b, G_{u|x})) > 0\]

Now for a subset \(\Omega_0 \subset \Omega\), \(\exists\)for \((b, G_{u|x}) \in \Omega_0\) \((b, F_{u|x})\) not identified, let \(C(\cdot, \cdot)\) be a function from \(R^K \times \Phi^X\) \textit{onto} some space \(C\). For any \(e \in C\), \(C(\beta, F_{u|x})\) is identified relative to \(e\) if

\[\exists \text{ no } (b, G_{u|x}) \in \Omega_0 \text{ such that } C(b, G_{u|x}) = e\]
**Theorem: Conventional Case:**

If $U \perp \perp X$ and $F_0$ a known (continuously increasing) distribution, then $\beta$ is identified relative to $b$ iff $\Pr(x\beta \neq xb) > 0$

**Proof:**
Let $x \in X$. Given that $F_0$ is continuous and strictly increasing,
\[
\Pr(D = 1|x) = \int 1(U \geq -x\beta)dF_0 = \int 1(U \geq -xb)dF_0
\]
\[
\Leftrightarrow x\beta = xb \text{ result follow}
\]
Now
\[
\Pr(x(\beta - b) \neq 0) > 0 \forall b \neq \beta \text{ iff } \exists \text{no proper linear subspace of } \mathbb{R}^k
\]
with probability 1 under $F_x$ (rules out degeneracies *i.e.* multicollinearity)
**Quantile Independence:**

For a given $\alpha \in (0, 1)$ \(\Pr(U < 0|x) = \alpha\) \(\forall x \in X\),

\[\alpha = \frac{1}{2}\] (median independence) \(\text{median}(y|x) = x\beta\)

For any quantile, define for \(b \in R^k\),

\[Q_b = \{x \in X : xb < 0 \leq x\beta \cup x\beta < 0 \leq xb\}.

Then $\beta$ is identified relative to $b$ iff

\[\Pr(x \in Q_b) > 0\]

(Idea: need for each admissible $b$ some set of values that will classify $D$ in different ways (relative to $\beta$)),

7
Proof:

**Sufficiency.**

\[
\Pr(D = 1|x) \geq 1 - \alpha \iff \Pr(U \geq -x\beta|x) \geq 1 - \alpha
\]

Quantile independence \( \Rightarrow \)

\[
\eta \leq 0 \iff \Pr(U \geq \eta|x) \geq 1 - \alpha
\]

\[\therefore x\beta \geq 0 \iff \Pr(D = 1|x) \geq 1 - \alpha\]

For each \( x \in Q_b \)

\[
\begin{align*}
xb &< 0 \cap \Pr(D = 1|x) \geq 1 - \alpha \\
xb &\geq 0 \cap \Pr(D = 1|x) < 1 - \alpha;
\end{align*}
\]

Thus \( \exists \) no \( G_{u|x} \) such that \( \int 1[u \geq 0] \, dG(u|x) = 1 - \alpha \) and \( \int 1[u \geq -xb] \, dG(u|x) = \Pr(D = 1|x) \)

\[\therefore \beta \text{ is identified relative to all } b \text{ if } \Pr(x \in Q_b) > 0\]
**Necessity.** Consider \( x \in X \setminus Q_b \),

Define \( G_{u|x} \) to be a strictly increasing distribution such that

\[
\int 1(u \geq 0) dG(u|x) = 1 - \alpha
\]

\[
\int 1(u \geq -xb)dG(u|x) = \Pr(D = 1|x)
\]

\( \therefore (b, G_{u|x}) \) produces same responses as the true model \((\beta, F_{u|x})\)

\( \therefore \) If \( \Pr(x \in X \setminus X_b) = 1 \), \( \beta \) not identified relative to \((b, G_{u|x})\)
Observe some features of these models:

Need a normalization for $b = a\beta$, $a$ is scalar $Q_b = 0$

One such normalization: $\frac{\beta}{\|\beta\|}$

Other possible (e.g. $Var(U) = 1$, etc.)
Observe that if the data (X) don’t exhibit sufficient variation, model not identified:

\[ e.g. \text{ suppose that} \]
\[
\exists \mu > 0 \exists \Pr(|x_k| < \mu) = 1, k = 1, ..., K;
\]
\[
\exists \lambda > 0, \exists |x\beta| \geq \lambda(a.e \ F_x)
\]

Then model not identified: suppose b space covers \(\beta\) (i.e. there are admissible \(b\) in \(\text{nbd of } \beta\)).
Let $\delta_k = b_k - \beta_k$ and consider a $b$ such that

$$|\delta_k| < \frac{\lambda}{K\mu}, k = 1, \ldots, K.$$ 

For all $x \ni x\beta \geq \lambda$

$$xb = x\beta + \sum_{k=1}^{K} x_k \delta_k \geq \lambda - \sum_{k=1}^{K} |x_k| |\delta_k| > 0$$

For all $x \ni x\beta \geq -\lambda$

$$xb \leq -\lambda + \sum_{k=1}^{K} |x_k| |\delta_k| < 0$$

$\therefore Q(b) = 0 \therefore$ model not identified against these $b$.

Note that if the model is outside the neighborhood of $\beta$, might still be identified ($\therefore$ in a fundamental sense, model depends on support of regressors)
Corollary:

(Use normalization)

Under

1. Quantile independence

2. \( \exists \) no proper linear subspace of \( \mathbb{R}^K \) having probability 1 under \( F_X \)

3. \( \exists \) at least one \( k \in [2, \ldots K] \) \( \beta_k \neq 0 \) and for almost all values of \( w = (x_1, x_{k-1}, x_{k+1}, x_K) \), \( \Pr(x_k \in (a_1, a_2) | w) > 0 \) for all open intervals \( (a_1, a_2) \in \mathbb{R}^1 \). Choose \( k = K \), then model fully identified.
Pf: Let $\beta_K > 0$. For any $b \in R^K$,

\[
\tilde{b} = (b_1, \ldots, b_{K-1}); \quad \tilde{\beta} = (\beta_1, \ldots, \beta_{K-1}); \quad w = (x_1, \ldots, x_{K-1}).
\]

Define $R(b) = \Pr(xb < 0 < x\beta) + \Pr(x\beta < 0 \leq xb)$.

Need to show that $R(b) \neq 0$ for all $b$ not scalar multiples of $\beta$

(a) Let $b_K < 0$.

\[
\Pr(xb < 0 < x\beta) = \Pr\left(\frac{\tilde{w}\tilde{b}}{b_K} < x_K; \quad -\frac{\tilde{w}\tilde{\beta}}{\beta_K} < x_K\right)
\]

Conditioning on $w$, rhs is a tail probability for $x_K$ with positive probability (a.e) $w$.

$\therefore$ the prob positive and $\Pr Q_b > 0$ and $\beta_k$ identified.

(b) $b_K = 0$. $\tilde{b}$ is $b$ deleting $K$th coordinate; $\tilde{\beta}$ is $\beta$ deleting $K$th coordinate.

(*) $\Pr(xb < 0 < x\beta) = \Pr\left(\frac{\tilde{w}\tilde{b}}{\beta_K} < x_K\right)$

(**) $\Pr(x\beta < 0 < xb) = \Pr\left(0 < \frac{\tilde{w}\tilde{\beta}}{b_K}; \quad x_K < -\frac{\tilde{w}\tilde{\beta}}{\beta_K}\right)$

Now

if $\Pr(\tilde{w}\tilde{b} < 0) > 0$, (*) is positive by previous argument

if $\Pr(\tilde{w}\tilde{b} > 0) > 0$ (***) is positive

(Full rank condition (ii) makes $\Pr\left(\tilde{w}\tilde{\beta} = 0\right) < 1 \therefore$ identified)
(iii) $b_K > 0$

$\text{Prob}(xb < 0 < x\beta) = \text{Pr}(\frac{-w\beta}{\beta K} < x_K < \frac{-wb}{b_K})$

$\text{Prob}(x\beta < 0 < xb) = \text{Pr}(\frac{-wb}{b_K} < x_K < \frac{-w\beta}{\beta K})$

Consider any value of $x \exists \frac{w\beta}{\beta K} \neq -\frac{wb}{b_K}$

For almost all such $\tilde{w}$ one of the intervals $\left(\frac{-w\beta}{\beta K}, \frac{-wb}{b_K}\right)$, $\left(\frac{-wb}{b_K}, \frac{-w\beta}{\beta K}\right)$ exists and contains positive $x_K$ probabilities;

$\therefore \text{Pr}(Q(b)) > 0$ if $\text{Pr}\left(\frac{w\beta}{\beta K} = \frac{wb}{b_K}\right) < 1$

(Fails if scalar multiples not ruled out)

Also ruled out by assumption (ii) $\tilde{w}$ doesn’t lie in a probability one subspace

$\therefore \beta$ identified
Note:

Need more than this Full Rank condition.

Need “Full Support” condition.

Not needed if the distribution $F$ is known.

“Full Support” enables us to trace out distribution of $F_{u|x,K}$
Identifying sign of $\beta$

Under quantile independence, For $x \in X$, let $w \equiv (x_1, ..., x_{K-1})$

Let $W$ be the domain of $w$.

For $b \in R^K$, define

$$W_b \equiv \{w \in W : \text{Prob}(xb < 0|w) > 0 \land \text{Prob}(xb > 0|w) > 0\}$$

Suppose $\beta_K \neq 0$, $\text{sgn}(\beta_K)$ identified if Pr($w \in w_b$) > 0.

If $\beta_K = 0$, $\beta_K$ identified if Pr($w \in W_b$) > 0 for all $b \in R^K$ $\ni b_K \neq 0$ example: $x = (1, x_2)$

sign $\beta_2$ identified if

$$\text{Pr}(x_2 < -\frac{\beta_1}{\beta_2}) > 0 \text{ and } \text{Pr}(x_2 > -\frac{\beta_1}{\beta_2}) > 0$$

If $\beta_2 = 0$, need

$$\text{Pr}(x_2 < \eta) > 0 \text{ and } \text{Pr}(x_2 > \eta) > 0 \quad \forall \eta \in (-\infty, \infty)$$
Proof: (see Manski, 1988)

Manski also shows that symmetry and median independence have the same identifying power.

**Statistical Independence:** ⇒ Quantile $\alpha$ independence (full rank and full support) enable us to identify $F_{u|x} = F_u$ and $\beta$ up to normalization

(These are sufficient conditions)

**Necessary and sufficient conditions**

In addition to $Q_b$ define

$$R_b = \{(x, \varepsilon) \in \mathcal{X} \times \mathcal{X} : (x - \varepsilon)b < 0 \leq (x - \varepsilon)\beta \cup (x - \varepsilon)\beta < 0 \leq (x - \varepsilon)b\}$$

$\beta$ identified relative to $b$ iff

$$\Pr(x \in Q_b) + \Pr((x, \varepsilon) \in R_b) > 0$$
Proof: With statistical independence, \( \Pr[(x, \varepsilon) \in R_b] > 0 \) sufficient to identify model. Why?

\[
x\beta \geq \varepsilon\beta \iff \Pr(D = 1|x) \geq \Pr(D = 1|\varepsilon)
\forall (x, \varepsilon) \in \mathcal{X} \times \mathcal{X}
\]

This is a consequence of statistical independence and not just single quantile independence.
Let \((x, \varepsilon) \in R_b\), then either

\[
xb < \varepsilon b \cap \Pr(D = 1|x) \geq \Pr(D = 1|\varepsilon)
\]

or

\[
xb \geq \varepsilon b \cap \Pr(D = 1|x) < \Pr(D = 1|\varepsilon)
\]

\[
\therefore \text{ it follows that}
\]

\[
(\Pr(D = 1|x), \Pr(D = 1|\varepsilon) \neq \left( \int 1(u \geq -xb)dG, \int 1(u \geq \varepsilon b)dG \right)
\]

19
For all probability distributions $G$
\[ \therefore R \text{ is identified relative to } b \]
\[ \Pr(x \in Q_b) = \Pr((x, \varepsilon) \in R_b) = 0 \]

Define $S_b \equiv \{(x, \varepsilon) \in X \times X : x \notin Q_b \cap \varepsilon \notin Q_b \cap (x, \varepsilon) \notin R_b \}$
\[ (x\beta \geq 0 \Leftrightarrow xb \geq 0) \cap (\varepsilon\beta \geq 0 \Leftrightarrow \varepsilon b \geq 0) \cap (x\beta \geq \varepsilon \beta \Leftrightarrow xb \geq \varepsilon b) \forall (x, \varepsilon) \in S_b \]

Possible to find a continuous strictly increasing distribution $G$ such that
\[
\int 1[U \geq u] \, dG = .5 \text{ and } (\Pr(D = 1|x)), \Pr(D = 1|\varepsilon)
\]
\[
= \left( \int 1[u \geq -xb] \, dG, \int 1[u \geq -\varepsilon b] \, dG \right) \forall (x, \varepsilon) \in S_b
\]
\[ \therefore (\beta, F_0) \text{ not identified relative to } (b, G) \text{ if } \Pr((x, \varepsilon) \in S_b) = 1 \]
Some corollaries:
Let $X$ be multinomial with positive probability on each of the $M$ elements:

\[ X^M = (x^1, \ldots, x^M) \subset \mathcal{X} \]
\[ x^m \beta \neq 0, \ m = 1, \ldots, M \]

Define:

\[ \lambda_1 \equiv \min \{ |x^m \beta| : m = 1, \ldots, M \} \]
\[ \lambda_2 \equiv \min \{ |(x^m - x^n) \beta| : m, n = 1, \ldots, M, m \neq n \} \]
\[ \mu_1 \equiv \max \{ |x^m_k| : m = 1, \ldots, M; k = 1, \ldots, K \} \]
\[ \mu_2 \equiv \max \{ |x^m_k - x^n_k| : m, n = 1, \ldots, M \} \]
**Thm:**

\( \beta \) not identified relative to any \( b \supset \)

\[ |b_k - \beta_k| < \min \left[ \frac{\lambda_1}{k\mu_1}, \ldots, \frac{\lambda_K}{k\mu_K} \right], \ k = 1, \ldots, K \]

Proof: we already established that

\[ \Pr(x \in Q_b) = 0 \]

For all \( b \supset |b_k - \beta_k| < \frac{\lambda_i}{K\mu_1} \)

Thus we need to consider

\[ \Pr((x, \varepsilon) \in \mathcal{R}_b) \]

(This is just for multinomial)

\[ \therefore \text{weaker than non-identification} \]
Corollary: Suppose we have

(i) statistical independence

(ii) rank condition

\[ \exists \text{ at least one } k \in [2, ..., K] \in \beta_k \neq 0 \text{ and such that for almost all values of } W, \]

\[ \exists r_w \subset R^1 \text{ such that} \]

either \[ \Pr[x_k \in (a_1, a_2)|w] > 0 \text{ for all open intervals } (a_1, a_2) \subset (-\infty, r_w) \]

or \[ \Pr[x_k \in (a_1, a_2)|w] > 0 \text{ for all open intervals } (a_1, a_2) \subset (r_w, \infty) \]

Then \((\beta_2, ..., \beta_K)/||\beta||\) is identified.
PF: Let $b \in \mathbb{R}^K$ for $(x, \varepsilon) \in \mathcal{X} \times \mathcal{X}$ let $y \equiv (x_2 - \varepsilon_2, \ldots, x_K - \varepsilon_K)$ let $\mathcal{Y}$ be the domain of $y$. Let $\gamma \equiv (\beta_2, \ldots, \beta_K)$ and $c \equiv (b_2, \ldots, b_K)$

Define

$\mathcal{Y}_b \equiv \{ y \in \mathcal{Y} : (yc < 0 \leq yr) \cup (yr < 0 \leq yc) \}$

$(x_1 = \varepsilon_1 = 1)$

$\therefore (x, \varepsilon) \in R_b \iff y \in \mathcal{Y}_b$

We now establish that

$$\Pr(y \in \mathcal{Y}_b) > 0 \text{ for all } b \text{ such that } \frac{c}{||c||} \neq \frac{\gamma}{||\gamma||}$$

condition (iii)

$\Rightarrow$ conditional on $(y_1, \ldots, y_{K-2}), y_{K-1}$ has positive probability mass in every open interval of $\mathbb{R}^1$.

condition (ii) $\Rightarrow F_y$ does not place probability one on any proper subspace of $\mathbb{R}^K$.

$\therefore$ Distribution of $Y$ satisfies condition previously used in Quantile independence.
Corollary: Under Statistical Independence, for 
\((x, \varepsilon) \in \mathcal{X} \times \mathcal{X}\)
Let
\[
\begin{align*}
  w &\equiv (x_1, \ldots, x_{K-1}) \\
  z &\equiv ((x_1 - \varepsilon_1, \ldots, x_{K-1} - \varepsilon_{K-1})
\end{align*}
\]
Let \(W\) and \(Z\) be domain of \(W\) and \(Z\)
For \(b \in R^K\) define
\[
\begin{align*}
  W_b &\equiv \{w \in W : \Pr (xb < 0 \mid w) > 0 \cap \Pr (xb > 0 \mid w > 0)\} \\
  Z_b &\equiv \{z \in B : \Pr ((x - \varepsilon)b < 0 \mid z) > 0 \cap \Pr ((x - \varepsilon)b > 0 \mid z) > 0\}
\end{align*}
\]
Then
if \(\beta_K \neq 0\), sign \((\beta_K)\) is identified if
\[
\Pr (w \in W_b) + \Pr (w \in Z_b) > 0
\]
If \(\beta_K = 0\), \(\beta_K\) is identified if \(\Pr (w \in W_b) + \Pr (d \in Z_b > 0)\) for all \(b \in R^K \ni b_K \neq 0\).
Other Types of Information:

Index Sufficiency:

a) $F_{u|x} = F_{u|x|x}$ \quad $\forall x \in \mathcal{X}$

b) For each $x \in \mathcal{X}$, the distribution of $F_{u|x|x}$ is absolutely continuous wrt Lebesgue measure.

$\varphi (\cdot | v)$ is conditional density

And for all $U \in \mathbb{R}^1$,

$$\frac{\partial \int_{v}^{\infty} \varphi (u | v) du}{\partial v} = \int_{v}^{\infty} \frac{\partial \varphi (u | v)}{\partial v} du$$
In Sample Selection Models if \( x \beta \) is selection parameters, we have that

\[
Y = x \beta + u \\
I = z \gamma + v \\
D = 1(Z \gamma + v)
\]

\[
E(Y|X, Z, D = 1) = E(x \beta + v|X, Z, D = 1) = x \beta + E(v|X, Z, D = 1) = x \beta + E(v|x, z, v \geq -z \gamma)
\]

Now if joint density exists (\( = f(u, v) \))

\[
f(u, v|x, z)
\]

\[
E(u|x, z, v \geq -z \gamma) = \frac{\int u \int_{-z \gamma}^{\infty} f(u, v|x, z)dv}{\int_{-\infty}^{\infty} \int_{-z \gamma}^{\infty} f(u, v|x, z)dudv}
\]

density depends only on \( z \gamma \).
Observe in Linear Regression model we have

\[ Y = X\beta + U \]

\[ E(U|X) = 0 \] identifies \( \beta \) given rank (subspace) condition,

Not so here. (Why?) Let \( Y \) be latent variable

\[ E(Y|X) = X\beta \]

implies no relationship between

\( 1(Y \geq 0) \) and \( X \)
**Procedures:**

a) Powell: Stratify data on $\Pr(D = 1|X)$ and look for variation within strata. Differences out $E(U|D = 1, X))$

b) Newey-Andrews: Take Polynomial Expression then $\Pr(D = 1 | X)$ (exploit index)

c) Ichimura and Lee: Use Kernel methods to difference out $E(U | D = 1, X)$

d) (Smoothness; Gallant and Nychka (1984))
Panel Data Version (Manski, 1987): Quantile Independence:
Consider Two Periods of Panel Data

\[(y_t, x_t, u_t, t = 0, 1, c)\]

\[y_t = x_t \beta + c + u_t \quad t = 0, 1\]
\[D_t = 1(y_t > 0)\]
\[\beta \in \mathbb{R}^K\]

\(c\) correlated with \(x\)

\[u \equiv (u_0, u_1) \quad x \equiv (x_0, x_1) \quad D = (d_0, d_1)\]

\[F_{c|x} = \text{distribution of } c \text{ given } x.\]
Assumptions

1(a) $F_{u_1|x,c} = F_{u_0|x,c}$ (stationarity) all $(x,c)$

1(b) Support of $F_{u_0|x,c} = F_{u_1|x,c}$ is $R^1$ for all $x,c$.
     (guarantees $\Pr(D_1 \neq D_0) > 0$)

2 Let $w = (x_1 - x_0)$ and $F_w$ is distribution of $w$,

   (a) Support of $F_w$ is not contained in any proper linear subspace of $R^X$.

   (b) $\exists$ at least one $k \in [1, ..., K] \ni \beta_k \neq 0$ and such that for almost every value of
       $\tilde{w} = (w_1, ..., w_{k-1}, ..., w_{k+1}, w_K)$ the scalar random variable has everywhere positive
       Lebesgue density condition on $\tilde{w}$. 

31
**Proof of Identification:**

Lemma:

\[ x_{1\beta} > x_{0\beta} \iff E(D_1|x) > E(D_0|x) \]
\[ x_{1\beta} = x_{0\beta} \iff E(D_1|x) = E(D_0|x) \]
\[ x_{1\beta} < x_{0\beta} \iff E(D_1|x) < E(D_0|x) \]

Proof:

\[
\Pr (y_1 \geq 0|x, c)) = \int_{-x_{1\beta} - c}^{\infty} dF_{u_1|x,c}
\]
\[
\Pr (y_0 \geq 0|x, c) = \int_{-x_{0\beta} - c}^{\infty} dF_{u_0|x,c}
\]

Note: this condition is on bivariate \(x\)

\(:. \) for all \(c\)

\[
x_{1\beta} \geq x_{0\beta} \iff \Pr (y_1 \geq 0 \mid x, c) \geq \Pr (y_0 \geq 0 \mid x, c)
\]

and result follows.
Another way to state it:

\[ w_\beta > 0 \iff E(D_1 - D_0|x) > 0 \]
\[ w_\beta = 0 \iff E(D_1 - D_0|x) = 0 \]
\[ w_\beta < 0 \iff E(D_1 - D_0|x) < 0. \]

Now observe that

\[ \text{Median} (D_1 - D_0 \mid x, D_1 \neq D_0) = sgn(w_\beta) \]

Why?

\[
\Pr(D_1 - D_0 = 1 \mid x, D_1 \neq D_0) = \frac{\Pr(D_1 = 1, D_0 = 0 \mid x)}{\Pr(D_1 \neq D_0 \mid x)}
\]
\[
\Pr(D_1 - D_0 = -1 \mid x, D_1 \neq D_0) = \frac{\Pr(D_1 = 0, D_0 = 1 \mid x)}{\Pr(D_1 \neq D_0 \mid x)}
\]
Therefore,

\[\text{Median}(D_1 - D_0 | x, D_1 \neq D_0) = sgn \left[ \Pr(D_1 = 1, D_0 = 0 | x) - \Pr(D_1 = 0, D_0 = 1 | x) \right]\]

Now

\[\Pr(D_1 = 1, D_0 = 0 | x) = \Pr(D_1 = 1 | x) - \Pr(D_1 = 1, D_0 = 1 | x)\]

\[\Pr(D_1 = 0, D_0 = 1 | x) = \Pr(D_1 = 1 | x) - \Pr(D_1 = 1, D_0 = 1 | x)\]

Therefore

\[\text{Median}[(D_1 - D_0 | x, D_1 \neq D_0) = sgn[\Pr(D_1 = 1 | x) - \Pr(D_0 = 1 | x)]\]

So using previous Lemma,

\[sgn \Pr(D_1 = 1 | x) - \Pr(D_0 = 1 | x) = sgn(w\beta) \quad \text{QED}\]
Let

\[ W_b = \{ w \in \mathbb{R}^k : \text{sgn}(wb) \neq \text{sgn}(w\beta) \} \, . \]

\( \beta \) identified relative to \( b \) if

\[ \Pr(w \in W_b) > 0 \]

Normalize and you are done.
\[
y = 1 \left( h^* (r) \geq y \right)
\]
\[
r \in R^K
\]
\[
h^* : T_i \to R
\]
\[
T_i \subset R^K
\]
\[
T_i \text{ in domain of } h
\]
\[
F^* \text{ cdf of } \eta
\]
\[
r \sim G(r) \text{ with support } S_G
\]
\[
T = S_G \cap T_i
\]
\[
\eta \perp \perp r
\]
\[
\Pr (y = 1 | r, h^*, F^*) = F^* (h^* (r))
\]
\[
= P (r) \text{ what we get from the raw data}
\]
\[
(h^*, F^*) \in (W \times \Gamma)
\]
Def: \((h^*, F^*)\) identified in \((W \times \Gamma)\) if \((h^*, F^*) \in (W \times \Gamma)\) and for any pair \((h, F) \ni (h, F) \neq (h^*, F^*)\) a set \(D \subset T \ni G(D) > 0\) and for all \(r \in D\)

\[
\Pr(y = 1| r, h, F) \neq \Pr(y = 1| r, h^*, F^*)
\]

Assumptions on \(W\).

W1: \(W\) is real valued, continuous functions with domain \(T_i\).

W2 (all functions satisfy this for subset): \(\exists\) subset \(\bar{T}\) of \(T\) such that

(i) For all \(h, h' \in W, and all r \in \bar{T}\)

\(h(r) = h'(r)\)

(ii) For all \(h \in W\) and all \(t \in h^*(T) \ni r \in \bar{T} \ni h(r) = t\)

W3: \(h^* \in W\)

W4: \(h^*\) is strictly increasing in \(K\)th coordinate of \(r\). (No subset with positive Lebesgue measure in \(T_i\) will be mapped by \(h^*\) into a point of discontinuity of \(F^*\))
Assumptions of $\Gamma$

$\Gamma_1$: $\Gamma$ is set of all monotone increasing functions on $R$ with range $[0,1)$

$\Gamma_2$: $F^* \in \Gamma$

$\Gamma_3$: $F^*$ strictly increasing on $h^*(T)$.

Assumptions on $g$, density of $r$

$G_1$: For any $r \in T$ and any $\delta > 0$, $G(B(r, \delta) \cap T) > 0$

$$B(r, \delta) = \{ r' \in R^k | \parallel r - r' \parallel < \delta \}$$

$G_2$: $K$th coordinate of $r$ possesses a density conditional on other components of $r$.

Theorem: Suppose these conditions satisfied the $(h^*, F^*)$ identified.

Proof (Stronger case):

Suppose values of $\Pr(y = 1|r, h^*, F^*)$ known on $T$

$F^*$ can be recovered. Why?

For each $t \in h^*(T)$ let $r^t \in \bar{T}$ be such that $h^*(r^t) = t$ (given by assumption W2)

Therefore $F^*(t) = F^*(h^*(r^t)) = \Pr(y = 1|r, h^*, F^*)$

Therefore we recover $F$. Therefore $h^* = (F^*)^{-1}(P(r))$
Examples

$W$ monotone increase homogenous of degree 1 functions for $h$

Let $r^t \in T_t, \alpha > 0$

\[
\begin{align*}
  h : T_t &\to R \ni h \in W, \\
  h(r^*) &= \alpha 
\end{align*}
\]

For each $r \in T_t$, $K$th coordinate of subgradient bounded by $B$ is subgradient if

\[
  h(r') < h(r) + h(r' - r)
\]

Assume $h^*$ belongs to $W$ and $r$ is absolutely continuous.

Another example:

\[
  h(r) = t(r_1, ..., r_{K-1}) + r_K
\]

Another example:

\[
  h(r) = \nu(r_1) + \nu(r_2)
\]
Ichimura and Thompson (1998)

\[ y = 1(x\beta \geq 0) \]

(includes the intercept term)

\[ \beta_1 \sim F_0 \in \mathcal{F} \]

\[ H(x) = \{ b | x'b \geq 0 \} \]

\[ P(x, F) = \int_{H(x)} dF \text{ for } F \in \mathcal{F} \]

The distribution of \( F_0 \) identified relative to \( \mathcal{F} \) iff

\[ \Pr \{ P(X_i, F) = P(X_i, F_0) \} \implies F = F_0 \]

Assume \( \Pr(\| B \| = 0) = 0 \); \( B = \{ \beta \| \beta \| = 1 \} \)

**Thm:** If

1. \( y_i = 1(x_i\beta_i \geq 0) \)
2. \( x_i \perp \perp \beta_i \)
3. \( \exists c \in B \ni \Pr \{ c'\beta_i > 0 \} = 1 \)
4. \( \Pr(\tilde{x}_i \in E) > 0 \) for each open set \( E \subset R^{k-1} \)

Then \( F_0 \) identified.

See JOE 1998 (Uses Cramer Wold Device)
**Binary Prediction as a Decision Problem**

Let $D$ be a binary outcome and $D^*$ a prediction of $D$. Let $g(D, D^*)$ be the loss realized when the state of nature is $D$ and the prediction is $D^*$. Due to the discreteness $g$ only takes on four values.

Expected loss is

$$L(D^*) = \int g(D, D^*) \, dF(D),$$

$$= g(1, D^*) p + g(0, D^*)(1 - p),$$

where $p = \Pr(D = 1)$

So

$$L(D^*) = \begin{cases} 
  g(1, 0) p + g(0, 0)(1 - p) & \text{if } D^* = 0 \\
  g(1, 1) p + g(0, 1)(1 - p) & \text{if } D^* = 1 
\end{cases}$$
Then

\[ D^* = 0 \text{ is optimal} \iff g(1, 0)p + g(0, 1)(1 - p) < g(1, 1)p + g(0, 1)(1 - p) \]

which implies

\[ D^* = 0 \text{ is optimal} \iff p < \frac{g(0, 1) - g(0, 0)}{g(1, 0) - g(0, 0) - g(1, 1) + g(0, 1)} \]

otherwise \( D^* = 1 \) is optimal.

Note that for symmetric loss functions: \( g(1, 0) = g(0, 1), g(0, 0) = g(1, 1) \)
(this covers squared error loss and absolute error loss) we get

\[ D^* = 0 \text{ is optimal} \iff p < 1/2 \]