A sufficient and necessary condition for identification of binary choice models with fixed effects

Yinchu Zhu*

Department of Economics,
Brandeis University
September 23, 2022

Abstract

We study the identification of binary choice models with fixed effects. We provide a condition called sign saturation and show that this condition is sufficient for the identification of the model. In particular, we can guarantee identification even with bounded regressors. We also show that without this condition, the model is never identified unless the error distribution belongs to a small class. A test is provided to check the sign saturation condition and can be implemented using existing algorithms for the maximum score estimator. We also discuss the practical implication of our results.

Key words: identification, panel model, binary choice, fixed effects

*Email: yinchuzhu@brandeis.edu. I am extremely grateful to Whitney Newey for introducing me to the literature on fixed effects in nonlinear models and for having many inspiring discussions with me. I thank Roy Allen, Whitney Newey and Martin Mugnier for pointing out mistakes in previous versions and for providing helpful comments on the paper. All the errors are my own.
1 Introduction

This paper considers panel models with binary outcomes in the presence of fixed effects. These models are extremely useful in economic analysis as it allows for fairly general unobserved individual heterogeneity. The nonlinear nature of the binary choice models makes it difficult to eliminate the fixed effects by differencing. Then one would need to view the fixed effects or their conditional distribution as a nuisance parameter. As a result, the identification of the parameter of interest is tricky. In this paper, we provide new results and insights on what drives the identification of the coefficients on the covariates and what this means for applied work.

We observe independent and identically distributed (i.i.d) copies of \((Y, X)\), where \(Y = (Y_0, Y_1) \in \{0, 1\} \times \{0, 1\}\) and \(X = (X_0, X_1) \in \mathcal{X}_0 \times \mathcal{X}_1\). The data generating process is

\[
Y_t = 1\{X_t'\beta + \alpha \geq u_t\} \quad t \in \{0, 1\},
\]

where the fixed effects are represented by the scalar variable \(\alpha\). The identification of \(\beta\) is typically based on the distributional stationarity condition of Manski (1987):

**Assumption 1.** The distribution of \(u_t \mid (X, \alpha)\) does not depend on \(t \in \{0, 1\}\). Let the cumulative distribution function (c.d.f) of this common distribution be \(F(\cdot \mid x, \alpha)\). Assume that for any \((x, \alpha)\), \(F(\cdot \mid x, \alpha)\) is a strictly increasing function on \(\mathbb{R}\).

Without any restriction on the magnitude of \(\alpha\) and \(u_t\) in (1), it is impossible to identify the magnitude of \(\beta\) but the following result by Manski (1987) gives the identification of the “direction” of \(\beta\); in other words, \(\beta\) is identified up to scaling.

**Lemma 1.** Let Assumption 1 hold. Then with probability one,

\[
\text{sgn} \left( E(Y_1 - Y_0 \mid X) \right) = \text{sgn} (W'\beta),
\]

where \(W = X_1 - X_0\) and \(\text{sgn}(\cdot)\) is defined as \(\text{sgn}(t) = 1\) for \(t > 0\), \(\text{sgn}(t) = -1\) for \(t < 0\) and \(\text{sgn}(0) = 0\).

Under some regularity conditions, the identification result in Manski (1987) finds that \(b \neq \mu \beta\) for any \(\mu > 0\) if \(R(b) > 0\), where for any value \(b\),

\[
R(b) = P(\text{sgn}(W'\beta) \neq \text{sgn}(W'b)).
\]

(2)
In this paper, we provide new perspectives on what drives the identification of $\beta$. First, what is the role of the distribution of $X$? A critical assumption in the existing literature is that $X$ needs to be unbounded. For example, Assumption 2 in Manski (1987) requires the unboundedness of at least one component of $W$: for $W = (W_1, ..., W_K)'$, there exists $k$ such that $\beta_k \neq 0$ and the conditional distribution $W_k | (W_1, ..., W_{k-1}, W_{k+1}, ..., W_K)$ has support equal to $\mathbb{R}$ almost surely. Theorem 1 of Chamberlain (2010) goes even further: for bounded $X$, the identification fails in certain regions of the parameter space if no further restrictions are imposed on the distribution of the error term $u_t$.

We show that a sufficient condition for identification is what we refer to as sign saturation (Assumption 2), which states that $P(E(Y_1 - Y_0 | X) > 0)$ and $P(E(Y_1 - Y_0 | X) < 0)$ are both strictly positive. This condition can hold with bounded regressors $X$. Moreover, we show that this is also a necessary condition for identification unless the distribution of $u_t$ is in a special class. Therefore, although we know (from Chamberlain (2010)) that for bounded regressors the identification fails in some regions of the parameter space, our results pinpoint these regions: the identification fails exactly in regions where the sign saturation fails. Therefore, whether the regressors are bounded is not what really drives the identification. The identifiability is more closely related to the sign saturation condition.

Second, what is the role of the distribution of the error term $u_t$? In the current literature, the logistic link function is a special case, where $u_t | (X_t, \alpha)$ follows a logistic distribution. Andersen (1970) considers the estimation and inference in this case. The logistic link function is also singled out in Chamberlain (2010) as the “nice” link function: the identification does not require unbounded regressors and the Fisher information bound is strictly positive. Identification also holds under the logistic link function via conditional maximum likelihood, see e.g., Rasch (1960), Chamberlain (1980) and Davezies et al. (2021) among others; some works also extend these results on logit models to settings with more than 2 time periods, such as Honoré and Weidner (2020), Davezies et al. (2020), Honoré et al. (2021) and the references therein. This suggests that imposing a special class of parametric structure on the error distribution is a useful strategy of achieving identification and this special class is the logistic functions. We provide new insight on this. We show that this special class also includes those such that $\hat{G}(\cdot)$ is periodic, where $\hat{G}(\cdot)$ is the derivative of $\ln \frac{F(\cdot)}{1-F(\cdot)}$ and $F(\cdot)$ is the distribution function of $u_t$. If $F(\cdot)$ is logit,
then \( \dot{G}(\cdot) \) is a constant function, which is clearly periodic. However, we find that identification is possible for any periodic \( \dot{G}(\cdot) \). Although Chamberlain (2010) tells us that for non-logistic distributions, the identification fails in an open neighborhood, we show that for any period \( \dot{G}(\cdot) \), the identification must hold in another open neighborhood. Therefore, it seems to us that the truly hopeless distributions are those with non-periodic \( \dot{G}(\cdot) \). Indeed, outside this special class of periodic \( \dot{G}(\cdot) \), we show that identification is impossible at every point if the sign saturation condition is not satisfied. Thus, for generic distribution functions, sign saturation is a necessary condition for identification. We summarize the relation between identifiability and error distribution in Table 1.

Third, we propose a test that can check the sign saturation condition from the data. For bounded regressors, the existing literature points out the danger of identification failure. These results conclude that the identification possibly (not definitely) fails. However, such possibility can be intimidating enough to limit the usage of fixed effect models in practice. We provide a more accurate diagnosis. Using the proposed test, the applied researcher can simply check the identifiability for the problem at hand. It turns out that the computation is already familiar to applied researchers in this area as it only involves computing the usual maximum score estimator and its bootstrapping.

### 1.1 A simple example

Before we establish results for a general case, we illustrate the identification in a simple model with two bounded regressors. The identification strategy is still using \( R(\cdot) \) defined in (2).

**Example.** Recall that \( W = X_1 - X_0 \). Suppose that \( W = (W_1, W_2)' \in \mathbb{R}^2 \), where \( W_1, W_2 \sim \text{uniform}(-1,1) \) and are independent. We normalize \( \beta = (1, \beta_2)' \in \mathbb{R}^2 \) with \( \beta_2 \neq 0 \). We now show that if \( b = (1, b_2)' \in \mathbb{R}^2 \) satisfies that \( \beta_2 \neq b_2 \) and \( b_2 \neq 0 \), then \( R(b) > 0 \).

**Proof.** We prove the result assuming \( b_2 > \beta_2 \); the case for \( b_2 < \beta_2 \) is analogous. For any \( a > 0 \), define \( f(a) := E \left[ 1 \{ 0 < W_2 < a \} W_2 \right] = \frac{1}{2} \int_0^a wdw = \frac{1}{4} a^2 \). Clearly, \( f(a) > 0 \)

\(^1\)The normalization is not without loss of generality and assumes that the first component of \( \beta \) is strictly positive. The assumption of \( b_2 \neq 0 \) and \( \beta_2 \neq 0 \) is also restrictive. We adopt them to illustrate the idea in the simplest case. A general theoretical result will be provided later.
whenever $a > 0$. Notice that

\[ R(b) = P (W_1 + W_2 \beta_2 < 0 < W_1 + W_2 b_2) + P (W_1 + W_2 b_2 < 0 < W_1 + W_2 \beta_2). \]

Let $\mathcal{L}$ denote the Lebesgue measure. Then

\[
P (W_1 + W_2 \beta_2 < 0 < W_1 + W_2 b_2)
= P (-W_2 b_2 < W_1 < -W_2 \beta_2)
= P (W_1 \in (-W_2 b_2, -W_2 \beta_2))
= E \{ P (W_1 \in (-W_2 b_2, -W_2 \beta_2) \mid W_2) \}
\overset{(i)}{=} \frac{1}{2} \cdot E \left\{ 1 \{ -W_2 b_2 < -W_2 \beta_2 \} \cdot \mathcal{L} \left( (-W_2 b_2, -W_2 \beta_2) \cap (-1, 1) \right) \right\}
= \frac{1}{2} \cdot E \left\{ 1 \{ 0 < W_2 (b_2 - \beta_2) \} \cdot \mathcal{L} \left( (-W_2 b_2, -W_2 \beta_2) \cap (-1, 1) \right) \right\}
\overset{(ii)}{=} \frac{1}{2} \cdot E \left\{ 1 \{ W_2 > 0 \} \cdot \mathcal{L} \left( (-W_2 b_2, -W_2 \beta_2) \cap (-1, 1) \right) \right\},
\]

where (i) follows by the fact that $W_1$ has density $1/2$ on $(-1, 1)$ and (ii) follows by $b_2 > \beta_2$. Now we consider the event

\[
\left\{ 0 < W_2 < \min \left\{ 1, \frac{1}{|b_2|}, \frac{1}{|\beta_2|} \right\} \right\}.
\]

On this event, we clearly have $|W_2 b_2| < 1$ and $|W_2 \beta_2| < 1$, which means that $(-W_2 b_2, -W_2 \beta_2) \subset (-1, 1)$ and thus $\mathcal{L}((-W_2 b_2, -W_2 \beta_2) \cap (-1, 1)) = W_2 (b_2 - \beta_2)$. It follows that

\[
E \left\{ 1 \{ W_2 > 0 \} \cdot \mathcal{L} \left( (-W_2 b_2, -W_2 \beta_2) \cap (-1, 1) \right) \right\}
\geq E \left[ 1 \left\{ 0 < W_2 < \min \left\{ 1, \frac{1}{|b_2|}, \frac{1}{|\beta_2|} \right\} \right\} \cdot \mathcal{L} \left( (-W_2 b_2, -W_2 \beta_2) \cap (-1, 1) \right) \right]
= E \left[ 1 \left\{ 0 < W_2 < \min \left\{ 1, \frac{1}{|b_2|}, \frac{1}{|\beta_2|} \right\} \right\} \cdot W_2 (b_2 - \beta_2) \right]
= (b_2 - \beta_2) \cdot f \left( \min \left\{ 1, \frac{1}{|b_2|}, \frac{1}{|\beta_2|} \right\} \right).
\]

Since $b_2 - \beta_2 > 0$ and $|b_2|, |\beta_2| < \infty$, the above is strictly positive. Thus, $R(b) > 0$. \qed
1.2 How does Theorem 1 of Chamberlain (2010) fit into this example?

Suppose that \( F(\cdot|x,\alpha) \) does not depend on \((x,\alpha)\). We can just write \( F(\cdot) \). Assume that \( F(\cdot) \) is known. Theorem 1 of Chamberlain (2010) concludes that the identification of \( \beta \) fails for bounded regressors unless \( F \) is the logistic function. The natural question is: why, in the previous example, can we have that the regressors are allowed to be bounded and \( \beta \) is identified (after normalization)?

To answer this question, revisit the setting in Chamberlain (2010). Under his setting, the previous example would assume that \( W = (W_1, W_2)' \) with \( W_2 = 1 \), i.e., \( X_t = (X_{t,1}, X_{t,2})' \) with \( X_{t,2} = 1 \{t = 1\} \). In this case, we recall (from the above calculation) that in the case of \( b_2 > \beta_2 \),

\[
P(W_1 + W_2\beta_2 < 0 < W_1 + W_2b_2) = \frac{1}{2} \cdot E \left\{ \mathbf{1}\{W_2 > 0\} \cdot \mathcal{L} \left( (-W_2b_2, -W_2\beta_2) \cap (-1, 1) \right) \right\}
\]

\[
= \frac{1}{2} \cdot E \left\{ \mathcal{L} \left( (-b_2, -\beta_2) \cap (-1, 1) \right) \right\}.
\]

(3)

When \( |\beta_2| \) is large and \( b_2 \) is in a small neighborhood of \( \beta_2 \), \((-b_2, -\beta_2) \cap (-1, 1)\) is empty. Thus, \( R(b) = 0 \) for \( b \neq \beta \) if \( \beta \) and \( b \) are both large in magnitude with the same sign. In contrast, the example in Section 1.1 uses the fact that \((-W_2b_2, -W_2\beta_2) \cap (-1, 1)\) cannot always be empty if the support of \( W_2 \) contains \((0,1)\).

We can see that whether we can identify \( \beta \) (up to normalization) depends on whether \( W \) has enough variation. From the above discussion, the criterion for this is not simply whether the support of \( W \) is bounded. We will consider the following sign saturation condition.

**Assumption 2** (Sign Saturation). Both \( P(E(Y_1 - Y_0 | X) > 0) \) and \( P(E(Y_1 - Y_0 | X) < 0) \) are strictly positive.

Let us see how Assumption 2 fits into the calculation in (3). We observe that if \(|\beta_2| \geq 1\), then \((-b_2, -\beta_2) \cap (-1, 1) = \emptyset \) for any \( b_2 > \beta_2 \). On the other hand, if \(|\beta_2| < 1\), then \( \mathcal{L} \left((-b_2, -\beta_2) \cap (-1, 1)\right) > 0 \) for any \( b_2 > \beta_2 \). To see this, notice that if \(|b_2| \leq 1 \) and \( b_2 \neq \beta_2 \), then \((-b_2, -\beta_2) \cap (-1, 1) = (-b_2, -\beta_2) \), which has positive

\[2\text{Notice that we only need to consider the case of } b_2 > \beta_2 \text{ as stated in the analysis of the example in Section 1.1. If } b_2 < \beta_2, \text{ we apply a similar argument to the other term: } P(W_1 + W_2b_2 < 0 < W_1 + W_2\beta_2).]
Lebesgue measure; if \( b_2 > 1 \), then \((-b_2, -\beta_2) \cap (-1, 1) = (-1, -\beta_2)\), which also has positive Lebesgue measure. Hence, we can identify \( \beta_2 \) whenever \(|\beta_2| < 1\).

Notice that in this case, the condition of \(|\beta_2| < 1\) corresponds to the sign saturation in Assumption 2. By Lemma 1, the sign of \( E(Y_1 - Y_0 | X) \) is equal to the sign of \( W' \beta = W_1 + W_2 \beta_2 = W_1 + \beta_2 \). Since \( W_1 \) has support \([-1, 1]\), the condition of \( P(W_1 + \beta_2 > 0) > 0 \) and \( P(W_1 + \beta_2 < 0) > 0 \) is equivalent to \(|\beta_2| < 1\). Therefore, in this example, identification condition \(|\beta_2| < 1\) is the same as the sign saturation condition.

Let us point out the subtleties here with respect to Chamberlain (2010). Theorem 1 of Chamberlain (2010) says that for bounded \( X \), the identification fails in some open neighborhood of the parameter space. The above example, under his notation and assumptions, would have a parameter space \( B_1 \times B_2 \) for \( (\beta_1, \beta_2) \) and \( B_1 \) contains all \( \beta_1 \) with \(|\beta_1| \) sufficiently small. Since we normalize \( \beta_1 = 1 \), this means that \( B_2 \) contains numbers that are sufficiently large. Indeed, we have seen that if \( \beta_2 \) is too large in magnitude (i.e., \(|\beta_2| \geq 1\)), we would have \((-b_2, -\beta_2) \cap (-1, 1) = \emptyset\). Hence, we agree with Theorem 1 of Chamberlain (2010) in that for bounded \( X \), the identification fails in some regions of the parameter space (e.g., \(|\beta_2| \geq 1\) in the above example). However, the contribution here is to point out that in some regions of the parameter space, we still have identification and these regions exactly correspond to the sign saturation condition.

## 2 Sign saturation and identification

In this section, we will show that in more general settings, the sign saturation condition plays a central role. Assumption 2 is enough to guarantee the identification. If Assumption 2 does not hold, then the identification is impossible even under the “nice” logistic distribution.

### 2.1 The sign saturation condition is sufficient for identification

Following Chamberlain (2010), we assume that one of the regressors is binary in that it is equal to zero at \( t = 0 \) and is equal to one at \( t = 1 \); in other words, one component of \( W = X_1 - X_0 \) is one. Thus, without loss of generality, we can partition \( W = (Z', 1)' \in \mathbb{R}^K \). If all the regressors are continuous, then the result can still be
applied because the coefficient for the binary regressor is allowed to be zero. The first main result of this paper is to show that the sign saturation condition (together with additional weak assumptions) is enough to guarantee identification up to scaling. To state the formal result, we first recall the definition of the support of a random variable (or its corresponding probability measure): the support of probability measure $\lambda$ is the smallest closed set $A$ such that $\lambda(A) = 1$, e.g., page 227 of Dudley (2002).

**Theorem 1.** Let Assumption 1 hold. Partition $W = (Z', 1)' \in \mathbb{R}^K$. Suppose that the support of $Z$ is convex and has non-empty interior. Let Assumption 2 hold.

Then $b = \mu \beta$ for some $\mu > 0$ if and only if $\min\{P(W'b < 0), P(W'b > 0)\} > 0$ and $R(b) = 0$.

By Theorem 1, to check whether $b$ is a rescaled version of the true $\beta$, we only need to check two quantities from the distribution of the observed data: $R(b)$ and $\min\{P(W'b < 0), P(W'b > 0)\}$.

The requirement on the distribution of $Z$ is mild. Let $Z$ be the support of $Z$. First, we do not require $Z$ to be unbounded. Second, the distribution of $Z$ does not have to admit a density and “atoms” (or point masses) are allowed. As long as $Z$ has variations that cover an open set, this will provide enough information for identification. Theorem 1 imposes this via the convexity and non-empty interior of $Z$. When $Z$ is not convex, the result is still useful: as long as $Z$ contains a convex subset with non-empty interior, we can apply the result on this subset by restricting the sample, i.e., the sign saturation condition holds on the restricted sample. By inspecting the proof, we can see that the key is to ensure that $W'\beta$ has strictly positive probability mass on $(-\varepsilon, \varepsilon)$ for any $\varepsilon > 0$.

### 2.2 Is the sign saturation condition also necessary?

It turns out that in the many situations, sign saturation largely characterizes identifiability. To illustrate this point, we consider the setting studied by Chamberlain (2010) and assume that $u_t$ is independent of $(X, \alpha)$ and is from a known distribution. We show that without sign saturation, identification fails at every point unless the distribution of $u_t$ is from a special class.

Suppose that in time period $t \in \{0, 1\}$, we observe $(Y_t, X_t)$, where $X_t = (X_{t,1}', X_{t,2})'$ with $X_{t,1} \in \mathbb{R}^{K-1}$ and $X_{t,2} = 1\{t = 1\}$. Suppose that $u_t$ is independent of $(X, \alpha)$ and
has c.d.f \( F(\cdot) \). Then from the data, the distribution of \( Y \) given \((X, \alpha)\) is determined by the vector

\[
L(X; \beta, \alpha) = \begin{pmatrix}
F(X_{0,1}'\beta_1 + \alpha) \\
F(X_{1,1}'\beta_1 + \beta_2 + \alpha) \\
F(X_{0,1}'\beta_1 + \alpha) \cdot F(X_{1,1}'\beta_1 + \beta_2 + \alpha)
\end{pmatrix}
\]

where \( \beta = (\beta_1, \beta_2) \) is partitioned as \( \beta_1 \in \mathbb{R}^{K-1} \) and \( \beta_2 \in \mathbb{R} \), and \( W := X_1 - X_0 = (Z', 1)' \) with \( Z = X_{1,1} - X_{0,1} \). As pointed out in Chamberlain (2010), here we should aim for identification of \( \beta \), not just up to scaling, because “our scale normalization is built in to the given specification for the \( u_t \) distribution”. Let \( Z \) be the support of \( Z \), i.e., again the smallest closed set with the full probability mass. Since \( W = (Z', 1)' \), the support of \( W \) is \( W = Z \times \{1\} \).

The special class of functions is best described in terms of a transformation of \( F(\cdot) \). Define the function \( G(\cdot) = \ln \frac{F(\cdot)}{1 - F(\cdot)} \). Let \( \hat{G}(\cdot) \) denote the derivative of \( G(\cdot) \). If \( F(\cdot) \) is the logistic function, then \( G(\cdot) \) is an affine function and \( \hat{G}(\cdot) \) is a constant function. It turns out that if we are willing to rule out cases with \( \hat{G}(\cdot) \) being a periodic function,\(^3\) then sign saturation is a necessary condition for identification. This special class includes the logistic function, whose \( \hat{G}(\cdot) \) is a constant function and is clearly periodic. (Any real number is a period of a constant function.)

We now introduce the notations for describing identification. Let \( \Pi \) denote the set of probability measures on \( \mathbb{R} \). The distribution of \( Y \mid X = x \) is \( \int L(x; \beta, \alpha) d\pi_x(\alpha) \) for some \( \pi_x \in \Pi \). Here, \( \pi_x \) denotes the distribution \( \alpha \mid X = x \). We now recall the notation of identification failure discussed in Chamberlain (2010).

**Definition 1** (Identification failure at \( \beta \)). We say that the identification fails at \( \beta \) if there exists \( b \neq \beta \) such that for any \( x \) in the support, there exist \( \pi_x, \tilde{\pi}_x \in \Pi \) depending on \( x \) such that \( \int L(x; \beta, \alpha) d\pi_x(\alpha) = \int L(x; b, \alpha) d\tilde{\pi}_x(\alpha) \).

This definition says that there exist two “mixing distributions” \( G_x \) and \( \tilde{G}_x \) in \( \mathcal{G} \) such

\(^3\)A function \( h(\cdot) \) on \( \mathbb{R} \) is a periodic function if there exists \( c \neq 0 \) such that \( h(c + t) = h(t) \) for any \( t \in \mathbb{R} \). In this case, \( c \) is said to be a period of \( h(\cdot) \). Notice that we assume that a period has to be non-zero; otherwise, every function would be a periodic function with a period being zero.
that the two mixtures (representing the distribution $Y \mid X = x$) $\int L(x; \beta, \alpha)d\pi_x(\alpha)$ and $\int L(x; b, \alpha)d\tilde{\pi}_x(\alpha)$ are identical. Following Chamberlain (2010), we can state identification in terms of convex hulls of $L(x; \beta, \alpha)$. The identification fails at $\beta$ if there exists $b \neq \beta$ such that \( \text{conv}\{L(x; \beta, \alpha) : \alpha \in \mathbb{R}\} \cap \text{conv}\{L(x; b, \alpha) : \alpha \in \mathbb{R}\} \neq \emptyset \) for any $x$, where \( \text{conv} \) denotes the convex hull.

Notice that $X_0' \beta_1$ can be absorbed by $\alpha$ since the distribution of $\alpha$ is allowed to have arbitrary dependence on $x$. Without loss of generality, we view the entire $X_0' \beta_1 + \alpha$ as the fixed effects to simplify notations. Then we can restate Definition 1 as follows.

**Definition 2.** We say that the identification fails at $\beta$ if there exists $b \neq \beta$ such that $\mathcal{A}(w'\beta) \cap \mathcal{A}(w'b) \neq \emptyset$ for any $w \in \mathcal{W}$, where for any $t \in \mathbb{R}$, $\mathcal{A}(t) = \text{conv}\{p(t, \alpha) : \alpha \in \mathbb{R}\}$ and

$$p(t, \alpha) := \begin{pmatrix} F(\alpha) \\ F(t + \alpha) \\ F(\alpha) \cdot F(t + \alpha) \end{pmatrix}.$$

Perhaps the simplest way to see the equivalence between the two definitions is to notice that \( \text{conv}\{L(x; \beta, \alpha) : \alpha \in \mathbb{R}\} = \mathcal{A}(w'\beta) \), where $w = x_1 - x_0$. We now define the parameter space in which the sign saturation fails: $\mathcal{B}_+ = \{ \beta \in \mathbb{R}^K : w'\beta > 0 \ \forall w \in \mathcal{W}\}$. The main result for the necessity is the following.

**Theorem 2.** Suppose that $\mathcal{Z}$ is bounded. Suppose that $F(\cdot)$ is continuously differentiable. If $\dot{G}(\cdot)$ is not a periodic function, then the identification fails at every point in $\mathcal{B}_+$.

We compare this with Theorem 1 of Chamberlain (2010), which states that if $F(\cdot)$ is outside a special class, identification fails in an open neighborhood. Theorem 2 complements this result by showing that if $F(\cdot)$ is outside a special class, identification fails at every point, not just in an open neighborhood. The special class in Chamberlain (2010) only includes logistic functions, whereas here our special class is larger and includes all functions with periodic $\dot{G}(\cdot)$. This enlargement of the special class cannot be avoided because there are instances of non-logistic $F(\cdot)$ for which identification holds.\(^4\)

\(^4\)This enlargement is not restrictive in practice since distributions with $\dot{G}(t) = 2 + \cos(t)$ do not seem to be of great economic importance.
**Theorem 3.** Suppose that \( \dot{G}(\cdot) \) is a continuous periodic function that is non-constant. Let \( \beta = (\beta_1', \beta_2') \in \mathcal{B}_+ \). Then

1. there exists a minimal positive period \( \eta > 0 \) for \( \dot{G}(\cdot) \), i.e., the set \( \{a > 0 : \dot{G}(a + t) = \dot{G}(t) \ \forall \ t \in \mathbb{R} \} \) has a smallest element.
2. If \( (z' \beta_1 + \beta_2)/\eta \) is an integer for some \( z \) in the interior of \( \mathcal{Z} \), then \( \beta \) is identified; moreover, for any \( b \in \mathcal{B}_+ \) with \( b \neq \beta \), \( P(\mathcal{A}(W'\beta) \cap \mathcal{A}(W'b) = \emptyset) > 0 \), where \( W = (Z', 1)' \).

Hence, as long as \( \dot{G}(\cdot) \) is periodic and \( z' \beta_1 + \beta_2 \) is in \( \eta \cdot \mathbb{N} \) for some interior point \( z \) (\( \mathbb{N} \) denotes the set of positive integers), we still have identification. This second condition is always satisfied by some \( \beta \in \mathcal{B}_+ \). The reason is that for any \( \eta > 0 \), any \( q \in \mathcal{B}_+ \) and for \( z \), we can always find \( \alpha > 0 \) such that \( \alpha(z', 1)'q = \eta \) and clearly \( \beta = \alpha q \in \mathcal{B}_+ \). Moreover, if \( (z' \beta_1 + \beta_2)/\eta \) is an integer for some \( z \) in the interior of \( \mathcal{Z} \), then for any \( b = (b_1', b_2') \) close enough to \( \beta \), \( (z' b_1 + b_2)/\eta \) is also an integer for some \( z \) in the interior of \( \mathcal{Z} \).

Hence, Theorem 3 complements Theorem 1 of Chamberlain (2010) in an interesting way. For non-logistic \( F(\cdot) \), it is true that the identification fails at every point in an open set. On the other hand, we show that the identification also holds at every point in another open set for periodic \( \dot{G}(\cdot) \). It is worth noting that in Theorem 3, the identification of \( \beta \) is robust in the sense that it is not based on a small number of “unimportant points” in \( \mathcal{W} \); the points in \( \mathcal{W} \) that allow us to identify \( \beta \) have strictly positive probability mass. Therefore, the truly hopeless cases for identification are those with non-periodic \( \dot{G}(\cdot) \). Finally, we summarize our results and the existing literature in Table 1.

### 3 Testing for the sign saturation condition

We have seen that the sign saturation condition is sufficient and necessary for the identification. The conditional mean function \( \phi(X) = E(Y_1 - Y_0 \mid X) \) is identified. Therefore, ideally the sign saturation condition can be checked in data. Here, we provide a simple check that does not involve nonparametric estimation of \( \phi \).

By Lemma 1, \( P(\phi(X) \leq 0) = 1 \) if and only if \( P(W'\beta \leq 0) = 1 \). Hence, we define \( \rho(q) = E1\{W'q \geq 0\}(Y_1 - Y_0) \) for \( q \in \mathbb{R}^K \). Here, we notice that we can replace \( \mathbb{R}^K \) with \([{-1, 1}]^K \) or any set that contains an open neighborhood of zero. We now state the condition of \( P(\phi(X) \leq 0) = 1 \) in terms of \( \rho(\cdot) \), which can be easily estimated.
Table 1: Identification and sign saturation

For bounded and convex $Z$ with non-empty interior:

| $\dot{G}$ constant (logistic $F$) | Sign saturation holds | Sign saturation does not hold |
|----------------------------------|-----------------------|-----------------------------|
| ID holds at every point (Chamberlain (1980) Davezies et al. (2021)) | ID holds at every point (Chamberlain (1980) Davezies et al. (2021)) |

| $\dot{G}$ periodic but non-constant | ID holds at every point (Theorem 1) | ID holds in an open set but fails in another open set (Theorem 3 and Chamberlain (2010)) |

| $\dot{G}$ non-periodic | ID holds at every point (Theorem 1) | ID fails at every point (Theorem 2) |

**Lemma 2.** Let Assumption 1 hold. Suppose that Then

1. $E(Y_1 - Y_0 \mid X) \leq 0$ almost surely if and only if $\sup_{q \in \mathbb{R}^K} \rho(q) \leq 0$.
2. $E(Y_1 - Y_0 \mid X) \geq 0$ almost surely if and only if $\inf_{q \in \mathbb{R}^K} \rho(q) \geq 0$.

By Lemma 2, we only need to construct one-sided confidence intervals for $\sup_{q \in \mathbb{R}^K} \rho(q)$ and $\inf_{q \in \mathbb{R}^K} \rho(q)$. We define

$$\hat{\rho}_n(q) = n^{-1} \sum_{i=1}^n 1\{W_i'q \geq 0\}(Y_{i,1} - Y_{i,0}).$$

Notice that computing $\sup_{q \in \mathbb{R}^K} \hat{\rho}_n(q)$ is equivalent to computing the maximum score estimator and all the existing computational algorithms and software for the maximum score estimator can be used. Finding $\inf_{q \in \mathbb{R}^K} \hat{\rho}_n(q)$ also reduces to computing the maximum score estimator once we swap $Y_1$ and $Y_0$.

Consider the following null hypothesis

$$H_0: E(Y_1 - Y_0 \mid X) \leq 0 \text{ almost surely.} \quad (4)$$

Define $S_n(q) = \sqrt{n}(\hat{\rho}_n(q) - \rho(q))$. Notice that the class of mappings $W \mapsto 1\{W'q \geq 0\}$ indexed by $q$ has VC-dimension at most $K + 2$ (Lemmas 2.6.15 and 2.6.18 of van der Vaart and Wellner (1996)). Clearly, the class $1\{W'q \geq 0\}(Y_1 - Y_0)$ indexed
by $q$ is bounded. By Theorem 2.5.2 of van der Vaart and Wellner (1996), $S_n$ converges to a mean-zero Gaussian process $S_*$. By Theorem 3.6.1 of van der Vaart and Wellner (1996), we can approximate the distribution of $S_*$ via the nonparametric bootstrap.

Of course, we need to check that $S_*$ is not degenerate under $H_0$. By Lemma 1, under $H_0$, $P(W'\beta \leq 0) = 1$,

$$\rho(-\beta) = E1\{W'\beta \leq 0\}(Y_1 - Y_0) = E(Y_1 - Y_0)$$

and thus

$$E(S_n(-\beta))^2 = E1\{W'\beta \leq 0\}(Y_1 - Y_0)^2 - (\rho(\beta))^2$$

$$= E(Y_1 - Y_0)^2 - (E(Y_1 - Y_0))^2 = \text{Var}(Y_1 - Y_0).$$

Therefore, $1\{W' \beta < 0\}(Y_1 - Y_0)$ is not degenerate if $Y_1 - Y_0$ is non-zero variance. We now derive a test for $H_0$. Under this null hypothesis, we have $\sup_{q \in \mathbb{R}^K} \rho(q) \leq 0$ (due to Lemma 2) and thus

$$\sqrt{n} \sup_{q \in \mathbb{R}^K} \hat{\rho}_n(q) = \sqrt{n} \sup_{q \in \mathbb{R}^K} (\rho(q) + n^{-1/2}S_n(q)) \leq \sqrt{n} \sup_{q \in \mathbb{R}^K} \rho(q) + \sup_{q \in \mathbb{R}^K} S_n(q)$$

$$\leq \sup_{q \in \mathbb{R}^K} S_n(q) \rightarrow^d \sup_{q \in \mathbb{R}^K} S_*,$$

where the convergence in distribution follows by the continuous mapping theorem. Under the alternative, Lemma 2 implies that $\sup_{q \in \mathbb{R}^K} \rho(q) > 0$ and thus

$$\sqrt{n} \sup_{q \in \mathbb{R}^K} \hat{\rho}_n(q) = \sqrt{n} \sup_{q \in \mathbb{R}^K} (\rho(q) + n^{-1/2}S_n(q)) \geq \sqrt{n} \sup_{q \in \mathbb{R}^K} \rho(q) + \inf_{q \in \mathbb{R}^K} S_n(q)$$

$$= \sqrt{n} \sup_{q \in \mathbb{R}^K} \rho(q) + O_P(1) \rightarrow \infty.$$

We have proved the following result.

**Theorem 4.** Let Assumption 1 hold. Suppose that we have i.i.d data. Assume that $Y_1 - Y_0$ is not degenerate. Then $S_n$ converges to a centered non-degenerate Gaussian process $S_*$, which can be approximated by nonparametric bootstrap. Moreover, under $H_0$,

$$\limsup_{n \to \infty} P\left(\sqrt{n} \sup_{q \in \mathbb{R}^K} \hat{\rho}_n(q) > c_{1-\alpha}\right) \leq \alpha,$$

13
where \( c_{1-\alpha} \) is the \((1-\alpha)\) quantile of \( \sup_{q \in \mathbb{R}^K} S_* \). If \( H_0 \) does not hold, then

\[
\limsup_{n \to \infty} P \left( \sqrt{n} \sup_{q \in \mathbb{R}^K} \hat{\rho}_n(q) > c_{1-\alpha} \right) = 1.
\]

Theorem 4 establishes the asymptotic validity of following test for \( H_0 \). Note that the cubic root asymptotics typically associated with the maximum score estimator (see e.g., Kim and Pollard (1990) and Cattaneo et al. (2020)) does not arise here. The reason is that here we focus on the max of \( \hat{\rho}_n(\cdot) \), rather than the argmax. The cubic root asymptotics of the maximum score estimator is due to the non-standard rate of some terms in the expansion for analyzing the argmax. Fortunately, we do not have to deal with such terms for our purpose.

**Algorithm.** Implement the following procedure for testing \( H_0 \) with nominal size \( \alpha \in (0,1) \)

1. Collect data \( \{(W_i, Y_{i,1}, Y_{i,0})\}_{i=1}^n \).

2. Compute the test statistic \( T_n = \sqrt{n} \sup_{q \in \mathbb{R}^K} \hat{\rho}_n(q) \).

3. Draw a random sample \( \{\tilde{W}_i, \tilde{Y}_{i,1}, \tilde{Y}_{i,0}\}_{i=1}^n \) with replacement from the data and compute \( \sup_{q \in \mathbb{R}^K} \tilde{S}_n(q) \), where \( \tilde{S}_n(q) = \sqrt{n} \left( n^{-1} \sum 1\{\tilde{W}_i'q > 0\}(\tilde{Y}_{i,1} - \tilde{Y}_{i,0}) - \hat{\rho}_n(q) \right) \).

4. Repeat the previous step many times and compute \( c_{1-\alpha} \), the \((1-\alpha)\) quantile of \( \sup_{q \in \mathbb{R}^K} \tilde{S}_n(q) \).

5. Reject \( H_0 \) if \( T_n > c_{1-\alpha} \).

Similarly, we can test \( H_0 : E(Y_1 - Y_0 \mid X) \geq 0 \) almost surely: reject \( H_0 \) if \( \sqrt{n} \inf_{q \in \mathbb{R}^K} \hat{\rho}_n(q) < -c_{1-\alpha} \). Here, we can use the same \( c_{1-\alpha} \) because centered Gaussian processes are symmetric around zero.

For applied researchers, results in this paper suggest the following steps in practice. We can start with blindly computing the maximum score estimator and bootstrapping it without worrying about whether the identification fails. Then this allows us to check the identification by verifying the sign saturation condition using the above algorithm; in essence, the test statistic and the critical values are simply the objective function values of the maximum score estimator and its bootstrapped version. Once
we are confident that the model is identified, the estimate we already computed is
econometrically justified and the bootstrapped results we computed at the beginning
can be used for inference.

A Appendix: proofs of theoretical results

A.1 Proof of Theorem 1

We define the following notations that will be used throughout this subsection (proof
of Theorem 1). For any set $A$, $\overline{A}$ denotes the closure of $A$ and $A^0$ denotes the interior
of $A$. Let $Z$ be the support of $Z$. We partition $\beta = (\beta_1', \beta_2')$ and $b = (b_1', b_2')$ with
$\beta_1, b_1 \in \mathbb{R}^{K-1}$. Then $W'\beta = Z'\beta_1 + \beta_2$. Define the following sets: $A_1 = \{z \in \mathbb{R}^{K-1}: z'\beta_1 + \beta_2 > 0\}$, $A_2 = \{z \in \mathbb{R}^{K-1}: z'b_1 + b_2 < 0\}$, $B_1 = \{z \in \mathbb{R}^{K-1}: z'\beta_1 + \beta_2 < 0\}$ and $B_2 = \{z \in \mathbb{R}^{K-1}: z'b_1 + b_2 > 0\}$. Before we prove Theorem 1, we establish some
auxiliary results.

**Lemma 3.** Suppose that $A$ is an open set (in $\mathbb{R}^{K-1}$) such that $A \cap Z \neq \emptyset$. Then $P(Z \in A) > 0$.

**Proof.** Recall that $Z$ is the smallest closed set such that $P(Z \in Z) = 1$, see the
definition of the support of a measure on page 227 of Dudley (2002).

Suppose that $P(Z \in A) = 0$. Then $P(Z \in Z \setminus A) = 1$. Since $A$ is open, $Z \setminus A$ is
a closed set. Notice that $Z \setminus A$ is a proper subset of $Z$ because $A \cap Z$ is non-empty.
Then $Z \setminus A$ is a closed set with full probability mass and is a smaller set than $Z$,
contradicting the assumption that $Z$ is the support of $Z$. □

**Lemma 4.** Suppose that $B$ is an open set (in $\mathbb{R}^{K-1}$) such that $\overline{B} \cap Z^0 \neq \emptyset$. Then $P(Z \in B) > 0$.

**Proof.** Let $A = B \cap Z^0$. By the openness of $B$ and $Z^0$, $A$ is open. The desired result
follows by Lemma 3 once we show that $A$ is non-empty.

We do so by contradiction. Suppose $B \cap Z^0 = \emptyset$. Then $B \subseteq (Z^0)^c$. Since $Z^0$
is open, $(Z^0)^c$ is closed. Thus, $(Z^0)^c$ is a closed set containing $B$. Recall that $\overline{B}$ is
the smallest closed set containing $B$. It follows that $\overline{B} \subseteq (Z^0)^c$, which means that
$\overline{B} \cap Z^0 = \emptyset$. This contradicts the assumption of $\overline{B} \cap Z^0 \neq \emptyset$. Hence, $A = B \cap Z^0$ is
non-empty. □
Lemma 5. Suppose that \( b \neq \mu \beta \) for any \( \mu > 0 \). If \( P(Z \in A_1), P(Z \in A_2), P(Z \in B_1) \) and \( P(Z \in B_2) \) are all strictly positive, then \( \overline{A_1 \cap A_2} = \overline{A_1 \cap \overline{A_2}} \) and \( B_1 \cap \overline{B_2} = \overline{B_1 \cap \overline{B_2}} \).

Proof. We prove the first claim; the second claims follows by an analogous argument.

Since \( P(Z \in A_1) \) and \( P(Z \in B_1) \) are both strictly positive, we have that \( \beta_1 \neq 0 \). To see this, observe that if \( \beta_1 = 0 \), then the two events \( \{Z \in A_1\} = \{\beta_2 > 0\} \) and \( \{Z \in B_1\} = \{\beta_2 < 0\} \) cannot both have strictly positive probability. Similarly, \( P(Z \in A_2) > 0 \) and \( P(Z \in B_2) \) yield yield \( b_1 \neq 0 \).

First observe that \( \overline{A_1 \cap A_2} \subseteq \overline{A_1} \cap \overline{A_2} \) (because \( A_1 \cap A_2 \subseteq \overline{A_1} \cap \overline{A_2} \) and \( \overline{A_1} \cap \overline{A_2} \) is closed).

To show the other direction, we fix an arbitrary \( z_* \in \overline{A_1} \cap \overline{A_2} \). We need to show that \( z_* \in \overline{A_1 \cap A_2} \).

Notice that \( z'_* \beta_1 + \beta_2 \geq 0 \) (due to \( z_* \in \overline{A_1} \)) and \( z'_* b_1 + b_2 \leq 0 \) (due to \( z_* \in \overline{A_2} \)). If \( z'_* \beta_1 + \beta_2 > 0 \) and \( z'_* b_1 + b_2 < 0 \), then \( z_* \in A_1 \cap A_2 \), which means that \( z_* \in \overline{A_1 \cap A_2} \).

Therefore, we only need to show that \( z_* \in \overline{A_1 \cap A_2} \) in the following three cases:

(i) \( z'_* \beta_1 + \beta_2 = 0 \) and \( z'_* b_1 + b_2 < 0 \)

(ii) \( z'_* \beta_1 + \beta_2 > 0 \) and \( z'_* b_1 + b_2 = 0 \)

(iii) \( z'_* \beta_1 + \beta_2 = 0 \) and \( z'_* b_1 + b_2 = 0 \).

Consider Case (i). Let \( z_j = z_* + cj^{-1} \beta_1 \) for \( j \in \{1, 2, \ldots\} \). Since \( z'_* b_1 + b_2 < 0 \), we can choose a constant \( c > 0 \) to be small enough such that \( cb'_1 \beta_1 < |z'_* b_1 + b_2|/2 \), which would imply that \( z'_j b_1 + b_2 < (z'_* b_1 + b_2)/2 < 0 \). On the other hand, notice that \( z'_j \beta_1 + \beta_2 = (z'_* \beta_1 + \beta_2) + cj^{-1} \|\beta_1\|_2^2 = cj^{-1} \|\beta_1\|_2^2 > 0 \) due to \( \beta_1 \neq 0 \). Thus, we have constructed a sequence \( \{z_j\}_{j=1}^{\infty} \) in \( A_1 \cap A_2 \) such that \( \lim_{j \to \infty} z_j = z_* \). Hence, \( z_* \) is a limit point of \( A_1 \cap A_2 \), which means that \( z_* \in \overline{A_1 \cap A_2} \).

Similarly, we can show that \( z_* \in \overline{A_1 \cap A_2} \) in Case (ii).

We now consider Case (iii). We first show that there exists \( \Delta \in \mathbb{R}^{K-1} \) such that \( \Delta' \beta_1 > 0 \) and \( \Delta' b_1 < 0 \). We proceed by contradiction. Suppose that there does not exist \( \Delta \in \mathbb{R}^{K-1} \) with \( \Delta' \beta_1 > 0 \) and \( \Delta' b_1 < 0 \). Then for any \( \Delta \in \mathbb{R}^{K-1} \), \( \Delta' \beta_1 > 0 \) implies \( \Delta' b_1 \geq 0 \). By the continuity of \( \Delta \mapsto \Delta' \beta_1, \Delta' \beta_1 \geq 0 \) implies \( \Delta' b_1 \geq 0 \). By Farkas’s lemma (e.g., Corollary 22.3.1 of Rockafellar (1970)), there exists \( \xi \geq 0 \) such that \( b_1 = \xi \beta_1 \). Since we have proved \( b_1 \neq 0 \), it follows that \( \xi \neq 0 \), which means that \( \xi > 0 \). Now we have \( z'_* b_1 + b_2 = \xi z'_* \beta_1 + b_2 = \xi(z'_* \beta_1 + b_2 \xi^{-1}) \). Since \( z'_* \beta_1 + \beta_2 = 0 \) and
Thus, we can approximate $z^*$ would imply $\lambda$ for any $\Delta' > 0$ and $\Delta'b < 0$.

Now we choose $\Delta \in \mathbb{R}^{K-1}$ such that $\Delta' \beta_1 > 0$ and $\Delta'b < 0$. Define $\tilde{z}_j = z^* + j^{-1}\Delta$ for $j \in \{1, 2, \ldots\}$. Then $\tilde{z}_j'\beta_1 + \beta_2 = (z^*\beta_1 + \beta_2) + j^{-1}\Delta' \beta_1 > 0$ and $z^*b_1 + b_2 = (z^*b_1 + b_2) + j^{-1}\Delta'b < 0$. Again, we have constructed a sequence $\{\tilde{z}_j\}_{j=1}^{\infty}$ in $A_1 \cap A_2$ such that $\lim_{j \to \infty} \tilde{z}_j = z^*$. Hence, $z^*$ is a limit point of $A_1 \cap A_2$, which means that $z^* \in \overline{A_1 \cap A_2}$.

\[\square\]

**Lemma 6.** Suppose that $Z$ is convex and $Z^0$ is non-empty. Assume that $W'\beta$ has strictly positive probability mass on $(0, \varepsilon)$ and $(-\varepsilon, 0)$ for any $\varepsilon > 0$. Moreover, assume that $P(Z \in A_1)$ and $P(Z \in A_2)$ are both strictly positive. If $b \neq \mu \beta$ for any $\mu > 0$, then at least one of $\overline{A_1 \cap A_2} \cap Z^0$ and $\overline{B_1 \cap B_2} \cap Z^0$ is not empty.

**Proof.** We proceed by contradiction. Suppose that $\overline{A_1 \cap A_2} \cap Z^0 = \emptyset$ and $\overline{B_1 \cap B_2} \cap Z^0 = \emptyset$. Clearly, $\overline{A_1} = \{z \in \mathbb{R}^{K-1}: z'\beta_1 + \beta_2 \geq 0\}$ and $\overline{A_2} = \{z \in \mathbb{R}^{K-1}: z'\beta_1 + \beta_2 \leq 0\}$ as well as $\overline{B_1} = \{z \in \mathbb{R}^{K-1}: z'\beta_1 + \beta_2 \leq 0\}$ and $\overline{B_2} = \{z \in \mathbb{R}^{K-1}: z'\beta_1 + \beta_2 \geq 0\}$.

Since $P(Z \in A_1)$ and $P(Z \in A_2)$ are strictly positive, $A_1$ and $A_2$ are both non-empty. Then there exist $\delta_1, \delta_2 > 0$ and $z_1, z_2 \in Z$ such that $z_1'\beta_1 + \beta_2 = \delta_1$ and $z_2'\beta_1 + \beta_2 = -\delta_2$. Since $Z$ is convex and $Z^0$ is non-empty, we have that the interior of $Z$ is dense in $Z$, i.e., $\overline{Z^0} = Z$; see Lemma 5.28 of Aliprantis and Border (2006). Thus, we can approximate $z_1$ and $z_2$ by points in $Z^0$. Hence, we can change $\delta_1, \delta_2 > 0$ slightly such that $z_1'\beta_1 + \beta_2 = \delta_1$ and $z_2'\beta_1 + \beta_2 = -\delta_2$ with $z_1, z_2 \in Z^0$.

By $\overline{A_1 \cap A_2} \cap Z^0 = \emptyset$, we have $z_1'\beta_1 + \beta_2 > 0$; since $z_1 \in \overline{A_1 \cap Z^0}$, $z_1'\beta_1 + \beta_2 \leq 0$ would imply $z_1 \in \overline{A_1 \cap A_2} \cap Z^0$ and contradict $\overline{A_1 \cap A_2} \cap Z^0 = \emptyset$. Similarly, $\overline{A_1 \cap A_2} \cap Z^0 = \emptyset$ implies $z_2'\beta_1 + \beta_2 < 0$; since $z_2 \in \overline{A_2 \cap Z^0}$, $z_2'\beta_1 + \beta_2 \geq 0$ would imply $z_2 \in \overline{A_1 \cap A_2} \cap Z^0$.

Now define $z_\ast(\lambda) = \lambda z_1 + (1-\lambda) z_2$ for $\lambda \in (0, 1)$. Since $Z$ is convex, we have that $Z^0$ is also convex by Lemma 5.27 of Aliprantis and Border (2006). Thus, $z_\ast(\lambda) \in Z^0$ for any $\lambda \in (0, 1)$. By $\overline{A_1 \cap A_2} \cap Z^0 = \emptyset$ and $\overline{B_1 \cap B_2} \cap Z^0 = \emptyset$, we have that $z_\ast(\lambda) \notin A_1 \cap A_2$ and $z_\ast(\lambda) \notin B_1 \cap B_2$ for any $\lambda \in (0, 1)$. This means that $z_\ast(\lambda)'\beta_1 + \beta_2$ and $z_\ast(\lambda)'b_1 + b_2$ have the same sign. In other words, we have that $f(\lambda) \geq 0$ for any $\lambda \in (0, 1)$, where

$$f(\lambda) = [z_\ast(\lambda)'\beta_1 + \beta_2] \cdot [z_\ast(\lambda)'b_1 + b_2].$$
By straight-forward algebra using \( z'_1\beta_1 + \beta_2 = \delta_1 \) and \( z'_2b_1 + b_2 = -\delta_2 \), we have

\[
z_\ast(\lambda)'\beta_1 + \beta_2 = \lambda\delta_1 + (1 - \lambda)(z'_2\beta_1 + \beta_2) = [\delta_1 - (z'_2\beta_1 + \beta_2)]\lambda + (z'_2\beta_1 + \beta_2) \quad (5)
\]

and

\[
z_\ast(\lambda)'b_1 + b_2 = \lambda(z'_1b_1 + b_2) - (1 - \lambda)\delta_2 = [(z'_1b_1 + b_2) + \delta_2]\lambda - \delta_2. \quad (6)
\]

Hence, \( f(\lambda) \) is a quadratic function of \( \lambda \). By \( \delta_1, \delta_2 > 0 \), \( z'_1b_1 + b_2 > 0 \) and \( z'_2\beta_1 + \beta_2 < 0 \), the coefficient of the \( \lambda^2 \) term in \( f(\lambda) \) satisfies

\[ [\delta_1 - (z'_2\beta_1 + \beta_2)] \cdot [(z'_1b_1 + b_2) + \delta_2] > \delta_1\delta_2 > 0. \]

Therefore, \( f(\cdot) \) is strictly convex. By (5) and (6), we notice that \( f(\lambda_1) = f(\lambda_2) = 0 \), where

\[
\lambda_1 = -\frac{z'_2\beta_1 + \beta_2}{\delta_1 - (z'_2\beta_1 + \beta_2)} \quad \text{and} \quad \lambda_2 = -\frac{\delta_2}{(z'_1b_1 + b_2) + \delta_2}.
\]

Since \( \delta_1, \delta_2 > 0 \), \( z'_1b_1 + b_2 > 0 \) and \( z'_2\beta_1 + \beta_2 < 0 \), we have \( \lambda_1 \in (0, 1) \) and \( \lambda_2 \in (0, 1) \). We notice that we must have \( \lambda_1 = \lambda_2 \). Otherwise, there would exist \( \lambda \) that is strictly between \( \lambda_1 \) and \( \lambda_2 \), i.e., \( \min\{\lambda_1, \lambda_2\} < \lambda < \max\{\lambda_1, \lambda_2\} \). Since \( f(\cdot) \) is strictly convex and \( f(\lambda_1) = f(\lambda_2) = 0 \), we have that \( f(\lambda) < 0 \) for any \( \lambda \) that is strictly between \( \lambda_1 \) and \( \lambda_2 \), contradicting \( f(\lambda) \geq 0 \) for all \( \lambda \in (0, 1) \). Hence, we must have that \( \lambda_1 = \lambda_2 \).

Let \( \lambda_0 \) denote this common value of \( \lambda_1 = \lambda_2 \). Then \( z_\ast(\lambda_0)'\beta_1 + \beta_2 = 0 \) and \( z_\ast(\lambda_0)'b_1 + b_2 = 0 \) from (5) and (6). Therefore, we have found \( z_0 := z_\ast(\lambda_0) \in \mathcal{Z}^\circ \) such that \( z'_0\beta_1 + \beta_2 = 0 \) and \( z'_0b_1 + b_2 = 0 \). In other words, \( z_0 \in \overline{\mathcal{A}_1 \cap \mathcal{A}_2} \cap \mathcal{Z}^\circ \) and \( z_0 \in \overline{\mathcal{B}_1 \cap \mathcal{B}_2} \cap \mathcal{Z}^\circ \). This contradicts the assumption that \( \overline{\mathcal{A}_1 \cap \mathcal{A}_2} \cap \mathcal{Z}^\circ = \emptyset \) and \( \overline{\mathcal{B}_1 \cap \mathcal{B}_2} \cap \mathcal{Z}^\circ = \emptyset \). \( \square \)

Proof of Theorem 1. The proof proceeds in three steps. In Step 1, we show a preliminary result. In Steps 2 and 3, we prove the “if” and the “only if” parts, respectively.

Step 1: show that \( P(W'\beta \in (0, \varepsilon)) > 0 \) and \( P(W'\beta \in (-\varepsilon, 0)) > 0 \) for any \( \varepsilon > 0 \).

Notice that \( \{ z \in \mathbb{R}^{K-1} : 0 < z'_1\beta_1 + \beta_2 < \varepsilon \} \) and \( \{ z \in \mathbb{R}^{K-1} : -\varepsilon < z'_1\beta_1 + \beta_2 < 0 \} \) are open sets because they are inverse images of open sets of continuous functions, i.e., \( f^{-1}((0, \varepsilon)) \) and \( f^{-1}((-\varepsilon, 0)) \) with \( f(z) = z'_1\beta_1 + \beta_2 \). By Lemma 3, we only need to show that both \( \{ z \in \mathbb{R}^{K-1} : 0 < z'_1\beta_1 + \beta_2 < \varepsilon \} \cap \mathcal{Z} \) and \( \{ z \in \mathbb{R}^{K-1} : -\varepsilon < z'_1\beta_1 + \beta_2 < 0 \} \cap \mathcal{Z} \) are non-empty.
By Lemma 1, Assumption 2 (i.e., \( P(E(Y_1 - Y_0 \mid X) > 0) > 0 \) and \( P(E(Y_1 - Y_0 \mid X) < 0) > 0 \)) implies that \( P(Z'\beta_1 + \beta_2 > 0) > 0 \) and \( P(Z'\beta_1 + \beta_2 < 0) > 0 \).

Hence, there exist \( z_1, z_2 \in \mathcal{Z} \) and \( \delta_1, \delta_2 > 0 \) such that \( z'_1\beta_1 + \beta_2 = \delta_1 \) and \( z'_2\beta_1 + \beta_2 = -\delta_2 \). Let \( \delta = \min\{\delta_1, \delta_2\} \). Since \( \delta/2 \in (-\delta_2, \delta_1) \), there exists \( \lambda \in (0, 1) \) such that \( z'\beta_1 + \beta_2 = \delta/2 \), where \( z_0 = \lambda z_1 + (1 - \lambda)z_2 \). Since \( \mathcal{Z} \) is convex, \( z_0 \in \mathcal{Z} \). Since \( \delta/2 < \varepsilon \), \( z_0 \in \{ z \in \mathcal{Z} : 0 < z'_1\beta_1 + \beta_2 < \varepsilon \} \), which means that \( \{ z \in \mathcal{Z} : 0 < z'_1\beta_1 + \beta_2 < \varepsilon \} \) is non-empty.

Similarly, \( -\delta/2 \in (-\delta_2, \delta_1) \) and thus we can find \( \bar{z}_0 \in \{ z \in \mathcal{Z} : -\varepsilon < z'_1\beta_1 + \beta_2 < 0 \} \) by interpolating between \( z_1 \) and \( z_2 \). Hence, \( \{ z \in \mathcal{Z} : -\varepsilon < z'_1\beta_1 + \beta_2 < 0 \} \) is non-empty.

**Step 2:** show that if \( \min\{P(W'b < 0), P(W'b > 0)\} > 0 \) and \( R(b) = 0 \), then \( b = \mu \beta \) for some \( \mu > 0 \).

Equivalently, we need to show that if \( b \neq \mu \beta \) for any \( \mu > 0 \), then either \( \min\{P(W'b < 0), P(W'b > 0)\} = 0 \) or \( R(b) > 0 \). In other words, we need to show that if \( b \neq \mu \beta \) for any \( \mu > 0 \) and \( \min\{P(W'b < 0), P(W'b > 0)\} > 0 \), then \( R(b) > 0 \). As argued in Step 1, \( P(Z'\beta_1 + \beta_2 > 0) > 0 \) and \( P(Z'\beta_1 + \beta_2 < 0) > 0 \), i.e., \( P(Z \in A_1) > 0 \) and \( P(Z \in B_1) > 0 \). By \( \min\{P(W'b < 0), P(W'b > 0)\} > 0 \), both \( P(W'b < 0) \) and \( P(W'b > 0) \) are strictly positive, i.e., \( P(Z \in A_2) > 0 \) and \( P(Z \in B_2) > 0 \).

Since both \( P(Z \in A_1) \) and \( P(Z \in A_2) \) are strictly positive, Step 1 and Lemma 6 imply that at least one of \( \bar{A}_1 \cap \bar{A}_2 \cap \mathcal{Z}^\circ \) and \( \bar{B}_1 \cap \bar{B}_2 \cap \mathcal{Z}^\circ \) is non-empty. We discuss these two cases separately.

Suppose that \( \bar{A}_1 \cap \bar{A}_2 \cap \mathcal{Z}^\circ \) is non-empty. By the openness of \( A_1 \) and \( A_2 \), \( A_1 \cap A_2 \) is open. We have seen that \( P(Z \in A_1) \), \( P(Z \in A_2) \), \( P(Z \in B_1) \) and \( P(Z \in B_2) \) are all strictly positive. By Lemma 5, we also have that \( \bar{A}_1 \cap \bar{A}_2 = \bar{A}_1 \cap \bar{A}_2 \). Thus, \( \bar{A}_1 \cap \bar{A}_2 \cap \mathcal{Z}^\circ = \bar{A}_1 \cap \bar{A}_2 \cap \mathcal{Z}^\circ \neq \emptyset \). Hence, we apply Lemma 4 with \( B = A_1 \cap A_2 \) and obtain \( P(Z \in A_1 \cap A_2) > 0 \). By an analogous argument, we have \( P(Z \in B_1 \cap B_2) > 0 \) if \( \bar{B}_1 \cap \bar{B}_2 \cap \mathcal{Z}^\circ \) is non-empty.

Therefore, we have proved that if \( b \neq \mu \beta \) for any \( \mu \in \mathbb{R} \) and \( \min\{P(W'b < 0), P(W'b > 0)\} > 0 \), then at least one of \( P(Z \in A_1 \cap A_2) \) and \( P(Z \in B_1 \cap B_2) \) is strictly positive. Since \( R(b) \geq P(A_1 \cap A_2) + P(B_1 \cap B_2) \), this means that \( R(b) > 0 \).

**Step 3:** show that if \( b = \mu \beta \) for some \( \mu > 0 \), then \( \min\{P(W'b < 0), P(W'b > 0)\} > 0 \) and \( R(b) = 0 \).
Since $\mu > 0$, we have
\[
\min\{P(W'b < 0), P(W'b > 0)\} = \min\{P(\mu W'\beta < 0), P(\mu W'\beta > 0)\}
\]
\[
= \min\{P(W'\beta < 0), P(W'\beta > 0)\} (i) > 0,
\]
where (i) follows by Step 1.

Since $b = \mu \beta$ with $\mu > 0$, $\text{sgn}(W'b) = \text{sgn}(W'\beta)$ almost surely, which means that $R(b) = 0$.

\section*{A.2 Proof of Theorem 2}

\begin{lemma}
Let $s, t \in \mathbb{R}$. If for any $v \neq (0, 0)'$, $\sup_{\alpha \in \mathbb{R}} v'p(s, \alpha) > \inf_{\alpha \in \mathbb{R}} v'p(t, \alpha)$, then $A(s) \cap A(t) \neq \emptyset$.
\end{lemma}

\begin{proof}
We proceed by contradiction. Suppose that $A(s) \cap A(t) = \emptyset$. Notice that both $A(s)$ and $A(t)$ are convex and non-empty. By the separating hyperplane theorem (e.g., Theorem 7.30 of Aliprantis and Border (2006)), there exists $v_0 \neq (0, 0)'$ and $r \in \mathbb{R}$ such that
\[
v_0'a \leq r \quad \forall a \in A(s)
\]
and
\[
v_0'b \geq r \quad \forall b \in A(t).
\]

In other words,
\[
\int v_0'p(s, \alpha)d\pi_1(\alpha) \leq r \quad \forall \pi_1 \in \Pi
\]
and
\[
\int v_0'p(t, \alpha)d\pi_2(\alpha) \geq r \quad \forall \pi_2 \in \Pi.
\]

By assumption and $v_0 \neq 0$, $\sup_{\alpha} v_0'p(s, \alpha) > \inf_{\alpha} v_0'p(t, \alpha)$. Thus, there exists $\varepsilon > 0$ such that
\[
\sup_{\alpha} v_0'p(s, \alpha) > \varepsilon + \inf_{\alpha} v_0'p(t, \alpha). \quad (7)
\]

We now choose $\alpha_*$ and $\xi_*$ such that $v_0'p(s, \alpha_*) \geq \sup_{\alpha} v_0'p(s, \alpha) - \varepsilon/4$ and $v_0'p(t, \xi_*) \leq \inf_{\xi} v_0'p(t, \xi) + \varepsilon/4$. Choose $\pi_{1,*}$ to be the probability measure that puts all the mass on $\alpha_*$. Similarly, choose $\pi_{2,*}$ to be the probability measure that puts all the mass on $\xi_*$. Then $v_0'p(s, \alpha_*) = \int v_0'p(s, \alpha)d\pi_{1,*}(\alpha) \leq r$ and
\[ v_0'(t, \xi_*) = \int v_0'(t, \alpha) d\pi_{2,*}(\alpha) \geq r. \] Hence,

\[ v_0'(t, \xi_*) \geq v_0'(s, \alpha_*). \]  

(8)

By (7) and the definitions of \( \alpha_* \) and \( \xi_* \), we have

\[ v_0'(s, \alpha_*) + \frac{\varepsilon}{4} \geq \sup_\alpha v_0'(s, \alpha) > \varepsilon + \inf_\alpha v_0'(t, \alpha) \geq \varepsilon + v_0'(t, \xi_*) - \frac{\varepsilon}{4}. \]

This means that \( v_0'(s, \alpha_*) \geq \varepsilon/2 + v_0'(t, \xi_*) \). Since \( \varepsilon > 0 \), this contradicts (8). \( \square \)

**Lemma 8.** Suppose that \( s > t \). Then for any \( v \in \mathbb{R}^3 \), \( \sup_{\alpha \in \mathbb{R}} v_0'(s, \alpha) \geq \inf_{\alpha \in \mathbb{R}} v_0'(t, \alpha) \).

**Proof.** Suppose not. Then there exists \( v \in \mathbb{R}^3 \) and \( \varepsilon > 0 \) such that

\[ \varepsilon + \sup_{\alpha \in \mathbb{R}} v_0'(s, \alpha) < \inf_{\xi \in \mathbb{R}} v_0'(t, \xi). \]

Let \( \delta = s - t \). By assumption, \( \delta > 0 \). The above display implies

\[ v'[p(t + \delta, \alpha) - p(t, \xi)] \leq -\varepsilon \quad \forall (\alpha, \xi) \in \mathbb{R}^2. \]

Let \( v = (v_1, v_2, v_3)' \). Then for any \( \alpha, \xi \in \mathbb{R} \),

\[ v_1[F(\alpha) - F(\xi)] + v_2[F(t + \delta + \alpha) - F(t + \xi)] + v_3[F(\alpha)F(t + \delta + \alpha) - F(\xi)F(t + \xi)] \leq -\varepsilon. \]

(9)

Taking \( \xi = \alpha \), we obtain that for any \( \alpha \in \mathbb{R} \),

\[ v_2[F(t + \delta + \alpha) - F(t + \alpha)] + v_3F(\alpha)[F(t + \delta + \alpha) - F(t + \alpha)] \leq -\varepsilon. \]

This means that for any \( \alpha \in \mathbb{R} \),

\[ v_2 \leq -v_3F(\alpha) - \frac{\varepsilon}{F(t + \delta + \alpha) - F(t + \alpha)}. \]

Notice that

\[ \inf_{\alpha \in \mathbb{R}} \left[ -v_3F(\alpha) - \frac{\varepsilon}{F(t + \delta + \alpha) - F(t + \alpha)} \right] \leq \inf_{\alpha \in \mathbb{R}} \left[ v_3 - \frac{\varepsilon}{F(t + \delta + \alpha) - F(t + \alpha)} \right] \]

21
\[
\leq \lim_{\alpha \to -\infty} \left[ (v_3) - \frac{\varepsilon}{F(t + \delta + \alpha) - F(t + \alpha)} \right] = -\infty.
\]

Thus, \( v_2 \leq -\infty \). This contradicts \( v \in \mathbb{R}^3 \).

**Lemma 9.** Define the function \( G(a) = \ln \frac{F(a)}{1 - F(a)} \). Assume that \( s > t > 0 \). If \( \sup_{\alpha \in \mathbb{R}} v'p(s, \alpha) = \inf_{\alpha \in \mathbb{R}} v'p(t, \alpha) \) for some \( v \neq (0, 0, 0)' \), then

\[
\inf_{\alpha \in \mathbb{R}} [G(s + \alpha) - G(\alpha)] > \sup_{\xi \in \mathbb{R}} [G(t + \xi) - G(\xi)].
\]

**Proof.** Let \( v = (v_1, v_2, v_3)' \). Without loss of generality, we can normalize and assume \( v_3 \in \{0, 1, -1\} \). Notice that \( \lim_{\alpha \to \infty} v'p(t, \alpha) = 0 \) and \( \lim_{\alpha \to \infty} v'p(s, \alpha) = 0 \). Thus, \( \inf_{\alpha \in \mathbb{R}} v'p(t, \alpha) \leq 0 \leq \sup_{\alpha \in \mathbb{R}} v'p(s, \alpha) \). By \( \sup_{\alpha \in \mathbb{R}} v'p(s, \alpha) = \inf_{\alpha \in \mathbb{R}} v'p(t, \alpha) \), it follows that

\[
\sup_{\alpha \in \mathbb{R}} v'p(s, \alpha) = \inf_{\alpha \in \mathbb{R}} v'p(t, \alpha) = 0.
\]

Let \( \delta = s - t \). By assumption, \( \delta > 0 \) and \( t > 0 \). Then for any \( \alpha, \xi \in \mathbb{R} \), \( v'p(s, \alpha) \leq 0 \leq v'p(t, \xi) \). In other words, for any \( \alpha, \xi \in \mathbb{R} \),

\[
v_1 F(\alpha) + v_2 F(t + \delta + \alpha) + v_3 F(\alpha) F(t + \delta + \alpha) \leq 0
\]

\[
\leq v_1 F(\xi) + v_2 F(t + \xi) + v_3 F(\xi) F(t + \xi). \tag{10}
\]

We now proceed in three steps.

**Step 1:** rule out \( v_3 = 0 \).

Suppose that \( v_3 = 0 \). By (10), we have that for any \( \alpha, \xi \in \mathbb{R} \),

\[
v_1 [F(\alpha) - F(\xi)] + v_2 [F(t + \delta + \alpha) - F(t + \xi)] \leq 0. \tag{11}
\]

Take \( \alpha = \xi - \delta \) and obtain \( v_1 [F(\xi - \delta) - F(\xi)] \leq 0 \) for any \( \xi \in \mathbb{R} \). By \( \delta > 0 \), this means that \( v_1 \geq 0 \).

It is easy to verify \( v_1 + v_2 = 0 \): taking \( \alpha = -\infty \) in (11), we have that \( -v_1 F(\xi) - v_2 F(t + \xi) \leq 0 \) for any \( \xi \in \mathbb{R} \), which means that \( v_1 + v_2 \geq 0 \) (take \( \xi \to \infty \)); taking \( \alpha = \infty \) in (11), we have that \( v_1 (1 - F(\xi)) + v_2 (1 - F(t + \xi)) \leq 0 \) for any \( \xi \in \mathbb{R} \), which means that \( v_1 + v_2 \leq 0 \) (take \( \xi \to -\infty \)).

Hence, \( v = (v_1, -v_1, 0) \). By \( v \neq (0, 0, 0)' \) and \( v_1 \geq 0 \), we have \( v_1 > 0 \). By (11), we
have that for any $\alpha, \xi \in \mathbb{R}$,

$$v_1[F(\alpha) - F(\xi)] - v_1[F(t + \delta + \alpha) - F(t + \xi)] \leq 0.$$ 

Since $v_1 > 0$, it follows that for any $\alpha, \xi \in \mathbb{R}$,

$$F(\alpha) - F(\xi) \leq F(t + \delta + \alpha) - F(t + \xi).$$

Taking $\alpha = \infty$, we have $F(\xi) \geq F(t + \xi)$, which is impossible because $t > 0$. Thus, $v_3 \neq 0$.

**Step 2:** rule out $v_3 = 1$.

Suppose that $v_3 = 1$. Then (10) implies that for any $\alpha, \xi \in \mathbb{R}$,

$$v_1[F(\alpha) - F(\xi)] + v_2[F(t + \delta + \alpha) - F(t + \xi)] + F(\alpha)F(t + \delta + \alpha) - F(\xi)F(t + \xi) \leq 0. \quad (12)$$

Taking $\alpha \to \infty$ and $\xi \to -\infty$, we have $v_1 + v_2 + 1 \leq 0$. Taking $\alpha \to -\infty$ and $\xi \to \infty$, we have $-v_1 - v_2 - 1 \leq 0$, which means $v_1 + v_2 + 1 \geq 0$. Hence, $v_1 + v_2 + 1 = 0$.

Take $\alpha = \xi - \delta$ in (12) and obtain that for any $\xi \in \mathbb{R}$,

$$v_1[F(\xi - \delta) - F(\xi)] + [F(\xi - \delta) - F(\xi)]F(t + \xi) \leq 0.$$

Since $\delta > 0$, we have $v_1 + F(t + \xi) \geq 0$ for any $\xi \in \mathbb{R}$. Taking $\xi \to -\infty$, we have $v_1 \geq 0$.

By $v_3 = 1$ and $v_1 + v_2 + 1 = 0$, the right-hand side of (10) implies that for any $\xi \in \mathbb{R}$,

$$v_1F(\xi) + (-1 - v_1)F(t + \xi) + F(\xi)F(t + \xi) \geq 0.$$

Taking $\xi = 0$, we have $[F(0) - 1]F(t) \geq v_1[F(t) - F(0)]$. Since $t > 0$, we have

$$v_1 \leq \frac{[F(0) - 1]F(t)}{F(t) - F(0)} < 0.$$

This contradicts $v_1 \geq 0$. Thus, $v_3 \neq 1$.

**Step 3:** show the final result.

From the previous two steps, we ruled out $v_3 \in \{0, 1\}$. Hence, $v_3 = -1$. Therefore,
(10) implies that for any \( \alpha, \xi \in \mathbb{R} \),

\[
[v_1 - F(t + \delta + \alpha)][F(\alpha) - F(\xi)] + [v_2 - F(\xi)][F(t + \delta + \alpha) - F(t + \xi)] \leq 0
\] (13)

Taking \( \alpha = \xi - \delta \) gives \( v_1 - F(t + \xi) \geq 0 \) (due to \( \delta > 0 \)) for any \( \xi \in \mathbb{R} \), which means that \( v_1 \geq 1 \).

By the right-hand side of (10) and \( v_3 = -1 \), \( F(\xi)v_1 + v_2F(t + \xi) - F(\xi)F(t + \xi) \geq 0 \) for any \( \xi \in \mathbb{R} \). Taking \( \xi = \infty \) gives \( v_1 + v_2 \geq 1 \). On the other hand, the left-hand side of (10) and \( v_3 = -1 \) yield \( v_1F(\alpha) + v_2F(t + \delta + \alpha) - F(\alpha)F(t + \delta + \alpha) \leq 0 \) for any \( \alpha \in \mathbb{R} \), which (by taking \( \alpha \to \infty \)) yields \( v_1 + v_2 \leq 1 \). Hence, \( v_1 + v_2 = 1 \), which means that \( v_2 = 1 - v_1 \) and \( v = (v_1, 1 - v_1, -1)' \).

Using the left-hand side of (10) and \( v = (v_1, 1 - v_1, -1)' \), we have that for any \( \alpha \in \mathbb{R} \),

\[
v_1F(\alpha) + (1 - v_1)F(t + \delta + \alpha) - F(\alpha)F(t + \delta + \alpha) \leq 0.
\]

Since \( F(t + \delta + \alpha) > F(\alpha) \) (due to \( t + \delta > 0 \)), we have that for any \( \alpha \in \mathbb{R} \),

\[
v_1 \geq \frac{F(t + \delta + \alpha)[1 - F(\alpha)]}{F(t + \delta + \alpha) - F(\alpha)}.
\]

Hence,

\[
v_1 \geq \sup_{\alpha \in \mathbb{R}} \frac{F(t + \delta + \alpha)[1 - F(\alpha)]}{F(t + \delta + \alpha) - F(\alpha)}.
\] (14)

Similarly, the right-hand side of (10) and \( v = (v_1, 1 - v_1, -1)' \), we have that for any \( \xi \in \mathbb{R} \), \( v_1F(\xi) + (1 - v_1)F(t + \xi) - F(\xi)F(t + \xi) \geq 0 \). Hence,

\[
v_1 \leq \inf_{\xi \in \mathbb{R}} \frac{F(t + \xi)[1 - F(\xi)]}{F(t + \xi) - F(\xi)}.
\] (15)

Combining (14) and (15), we have

\[
\sup_{\alpha \in \mathbb{R}} \frac{F(t + \delta + \alpha)[1 - F(\alpha)]}{F(t + \delta + \alpha) - F(\alpha)} \leq \inf_{\xi \in \mathbb{R}} \frac{F(t + \xi)[1 - F(\xi)]}{F(t + \xi) - F(\xi)}.
\] (16)

Recall \( G(\alpha) = \ln \frac{F(\alpha)}{1 - F(\alpha)} \). Then with straightforward algebra, (16) becomes \( \inf_{\alpha \in \mathbb{R}} [G(t + \delta + \alpha) - G(\alpha)] \geq \sup_{\xi \in \mathbb{R}} [G(t + \xi) - G(\xi)] \). \( \square \)

**Lemma 10.** Let \( T \subset (0, \infty) \) be a compact set. Assume that \( G(\cdot) \) is continuously
differentiable. Suppose that for any $\delta > 0$, there exists $t \in T$ such that

\[
\inf_{\alpha \in \mathbb{R}} [G(t + \delta + \alpha) - G(\alpha)] \geq \sup_{\xi \in \mathbb{R}} [G(t + \xi) - G(\xi)].
\]

Then $\dot{G}(\cdot)$ is a periodic function with a period in $T$, where $\dot{G}(\cdot)$ is the derivative of $G(\cdot)$. In other words, there exists $t \in T$ such that $G(t + \xi) = \dot{G}(\xi)$ for any $\xi \in \mathbb{R}$.

**Proof.** For any $b \in \mathbb{R}$, define $r_b(\cdot)$ by $r_b(a) = G(b + a) - G(a)$. We also define $c(b) = \sup_{\xi \in \mathbb{R}} r_b(\xi)$. Let $K > 0$ be arbitrary. Then by assumption, for any $\delta > 0$, there exists $t \in T$ (depending only on $\delta$) such that $\inf_{\alpha \in \mathbb{R}} [r_t(\delta + \alpha) + r_\delta(\alpha)] \geq \sup_{\xi \in \mathbb{R}} r_t(\xi) = c(t)$, which means that $\inf_{\alpha + \delta \in [-K,K]} [r_t(\delta + \alpha) + r_\delta(\alpha)] \geq c(t)$ and thus

\[
c(t) \leq \inf_{|\delta + \alpha| \leq K} r_t(\delta + \alpha) + \sup_{|\delta + \alpha| \leq K} r_\delta(\alpha) = \inf_{|\xi| \leq K} r_t(\xi) + \sup_{|\delta + \alpha| \leq K} r_\delta(\alpha).
\]

Notice that $\sup_{|\delta + \alpha| \leq K} r_\delta(\alpha) = \sup_{|\delta + \alpha| \leq K} [G(\delta + \alpha) - G(\alpha)] = \sup_{|\xi| \leq K} [G(\xi) - G(\xi - \delta)]$. We have that for any $\delta, K > 0$, there exists $t \in T$ depending only on $\delta$ such that

\[
\sup_{|\xi| \leq K} [G(\xi) - G(\xi - \delta)] \geq c(t) - \inf_{|\xi| \leq K} r_t(\xi) = \sup_{\xi \in \mathbb{R}} r_t(\xi) - \inf_{|\xi| \leq K} r_t(\xi) \geq 0.
\]

Now we choose $\delta = 1/n$. Then for any $n > 1$, there exists $t_n$ (not depending on $K$) such that

\[
\sup_{|\xi| \leq K} [G(\xi) - G(\xi - 1/n)] \geq c(t_n) - \inf_{|\xi| \leq K} r_{t_n}(\xi) \geq 0.
\]

Let $q_n(K) := c(t_n) - \inf_{|\xi| \leq K} r_{t_n}(\xi)$. Since $G(\cdot)$ is continuous, it is uniformly continuous on compact sets (Heine-Cantor theorem). Hence, $\limsup_{n \to \infty} \sup_{|\xi| \leq K} [G(\xi) - G(\xi - 1/n)] = 0$. The above display implies $\limsup_{n \to \infty} q_n(K) = 0$.

Since $t_n$ is in a compact set $T$, there exists a subsequence $t_{n_j}$ and $t_\star \in T$ such that $t_{n_j} \to t_\star$ and $t_\star$ does not depend on $K$. We can further extract a subsequence that is either increasing or decreasing. With an abuse of notation, we write $t_n$ rather than $t_{n_j}$ in the rest of the proof. We show the following claim in two cases ($t_{n_j} \uparrow t_\star$ and $t_{n_j} \downarrow t_\star$):

\[
\sup_{|\xi| \leq K} r_{t_\star}(\xi) = \inf_{|\xi| \leq K} r_{t_\star}(\xi).
\]
Step 1: show (17) assuming $t_n \uparrow t_*$. 

Since $t_n \leq t_*$, we have $r_{t_n}(\xi) \leq r_{t_*}(\xi)$ for any $\xi$ and thus

$$c(t_n) = q_n + \inf_{|\xi| \leq K} r_{t_n}(\xi) \leq q_n + \inf_{|\xi| \leq K} r_{t_*}(\xi).$$

Hence, by $\lim\sup_{n \to \infty} q_n(K) = 0$, we have

$$\lim\sup_{n \to \infty} c(t_n) \leq \inf_{|\xi| \leq K} r_{t_*}(\xi). \quad (18)$$

Let $R = \max_{t \in T} |t|$. Since $T$ is compact, $R$ is bounded. Notice that $G(\cdot)$ is continuous. Again, by the Heine-Cantor theorem, it is uniformly continuous on the compact set $[-K-R, K+R]$. Notice that $t_n + \xi \in [-K-R, K+R]$ for any $\xi \in [-K, K]$ (due to $|t_n| \leq R$). Thus, $\sup_{|\xi| \leq K} |r_{t_n}(\xi) - r_{t_*}(\xi)| = \sup_{|\xi| \leq K} |G(t_n + \xi) - G(t_* + \xi)| \to 0$. It follows that

$$\lim\sup_{n \to \infty} \sup_{|\xi| \leq K} r_{t_n}(\xi) = \sup_{|\xi| \leq K} r_{t_*}(\xi).$$

Now by (18) and the above display, we have that

$$\inf_{|\xi| \leq K} r_{t_*}(\xi) \geq \lim\sup_{n \to \infty} c(t_n) = \lim\sup_{n \to \infty} \sup_{\xi \in \mathbb{R}} r_{t_n}(\xi) \geq \lim\sup_{n \to \infty} \inf_{|\xi| \leq K} r_{t_n}(\xi) = \inf_{|\xi| \leq K} r_{t_*}(\xi).$$

On the other hand, we have $\inf_{|\xi| \leq K} r_{t_*}(\xi) \leq \sup_{|\xi| \leq K} r_{t_*}(\xi)$. This proves (17).

Step 2: show (17) assuming $t_n \downarrow t_*$. 

Since the argument is similar to Step 1, we only provide an outline here. By $t_n \geq t_*$, we have

$$q_n + \inf_{|\xi| \leq K} r_{t_n}(\xi) = c(t_n) \geq c(t_*) = \sup_{\xi \in \mathbb{R}} r_{t_*}(\xi) \geq \sup_{|\xi| \leq K} r_{t_*}(\xi).$$

Notice that $r_{t_*}(\xi) = G(t_* + \xi) - G(\xi)$ and $r_{t_n}(\xi) = G(t_n + \xi) - G(\xi)$. Again, by the Heine-Cantor theorem and the continuity of $G(\cdot)$, we have $\sup_{|\xi| \leq K} |r_{t_n}(\xi) - r_{t_*}(\xi)| = \sup_{|\xi| \leq K} |G(t_n + \xi) - G(t_* + \xi)| \to 0$ due to $t_n \to t_*$. Thus,

$$\lim\inf_{n \to \infty} \inf_{|\xi| \leq K} r_{t_n}(\xi) = \inf_{|\xi| \leq K} r_{t_*}(\xi).$$

The above two displays and $q_n \to 0$ imply $\inf_{|\xi| \leq K} r_{t_*}(\xi) \geq \sup_{|\xi| \leq K} r_{t_*}(\xi)$. This gives (17).
Step 3: show the final result.

Notice that we have verified (17) for an arbitrary \( K > 0 \). Since \( t_* \) does not depend on \( K \), \( r_{t_*} \) is a constant function on \( \mathbb{R} \). This means that the derivative of \( r_{t_*} \) is zero on \( \mathbb{R} \). Recalling that \( r_{t_*}(\xi) = G(t_* + \xi) - G(\xi) \), we have \( \dot{G}(t_* + \xi) = \dot{G}(\xi) \) for any \( \xi \in \mathbb{R} \).

**Proof of Theorem 2.** Fix an arbitrary \( \beta = (\beta'_1, \beta'_2)' \in B'_+ \). Then \( w' \beta = z' \beta_1 + \beta_2 > 0 \) for any \( z \in \mathcal{Z} \). Let \( \delta > 0 \) be an arbitrary number. Define \( b = (\beta'_1, \beta'_2 + \delta)' \). Clearly, \( w' b = z' \beta_1 + \beta_2 + \delta > 0 \) for any \( z \in \mathcal{Z} \). Thus, \( b \in B_+ \). Define \( T \) to be the closure of \( \{z' \beta_1 + \beta_2 : z \in \mathcal{Z}\} \). Since \( \mathcal{Z} \) is bounded, \( T \) is also bound. By the closedness of \( T \), \( T \) is also compact. We also notice that \( 0 \notin T \) and thus \( T \subset (0, \infty) \). (To see this, notice that \( \inf_{z \in \mathcal{Z}} (z', 1)' \beta \) is achieved by some \( z_* \in \mathcal{Z} \) (because \( \mathcal{Z} \) is closed and bounded). On the other hand, since for any \( z \in \mathcal{Z} \), \( (z', 1)' \beta > 0 \), we have that \( (z'_*, 1)' \beta > 0 \).

We now show that there exists \( \delta > 0 \) such that for any \( z \in \mathcal{Z} \), \( \mathcal{A}(w' \beta) \cap \mathcal{A}(w' b) \neq \emptyset \), where \( w = (z', 1)' \). We proceed by contradiction. Suppose that for any \( \delta > 0 \), there exists \( z_* \in \mathcal{Z} \) such that \( \mathcal{A}(w' b) \cap \mathcal{A}(w'_* \beta) = \emptyset \), where \( w_* = (z'_*, 1)' \). By Lemma 7, there exists \( v_\delta \neq 0 \) such that \( \sup_{\alpha \in \mathbb{R}} v_\delta p(w'_* b, \alpha) \leq \inf_{\alpha \in \mathbb{R}} v_\delta p(w'_* \beta, \alpha) \). By Lemma 8 and \( w'_* b - w'_* \beta = \delta > 0 \), we have that \( \sup_{\alpha \in \mathbb{R}} v_\delta p(w'_* b, \alpha) > \inf_{\alpha \in \mathbb{R}} v_\delta p(w'_* \beta, \alpha) \).

Hence,

\[
\sup_{\alpha \in \mathbb{R}} v_\delta p(w'_* \beta + \delta, \alpha) = \sup_{\alpha \in \mathbb{R}} v_\delta p(w'_* b, \alpha) = \inf_{\alpha \in \mathbb{R}} v_\delta p(w'_* \beta, \alpha).
\]

By Lemma 9 (with \( t = w'_* \beta \)),

\[
\inf_{\alpha \in \mathbb{R}} [G(w'_* \beta + \delta + \alpha) - G(\alpha)] > \sup_{\xi \in \mathbb{R}} [G(t_* + \xi) - G(\xi)].
\]

Notice that \( w'_* \beta \in T \). Therefore, we have shown that for any \( \delta > 0 \), there exists \( t_* \in T \) such that

\[
\inf_{\alpha \in \mathbb{R}} [G(t_* + \delta + \alpha) - G(\alpha)] > \sup_{\xi \in \mathbb{R}} [G(t_* + \xi) - G(\xi)].
\]

By Lemma 10, \( \dot{G} \) is a periodic function. However, this contradicts the assumption that \( \dot{G} \) is not a periodic function. \(\square\)
A.3 Proof of Theorem 3

Lemma 11. Suppose that $h(\cdot)$ is a continuous and non-constant function on $\mathbb{R}$. If $h(\cdot)$ is a periodic function, then $h(\cdot)$ has a minimal positive period, i.e., the set $\{a > 0 : h(a + x) = h(x) \forall x \in \mathbb{R}\}$ has a smallest element.

Proof. Let $K = \{a > 0 : h(a + x) = h(x) \forall x \in \mathbb{R}\}$.

We first show that $\inf K > 0$. Suppose that this is not true. Then $\inf K = 0$. Thus, there exists $a_n \in K$ such that $a_n \downarrow 0$. Let $x, y \in \mathbb{R}$ such that $x < y$. For any $n$, define $J_n$ to be the integer part of $(y - x)/a_n$, i.e., $J_n$ is the integer satisfying $J_n a_n \leq y - x < (J_n + 1)a_n$. Since $a_n$ is a period for $h(\cdot)$ and $J_n$ is an integer, $J_n a_n$ is also a period of $h(\cdot)$. Thus, $h(y) = h(y - x + x) = h(y - x - J_n a_n + x)$.

Notice that $0 \leq y - x - J_n a_n < a_n \to 0$. Thus, $(y - x - J_n a_n) + x \to x$. By the continuity of $h(\cdot)$, we have that $h(y - x - J_n a_n + x) \to h(x)$.

Hence, we have proved that $h(y) = h(x)$. Since $x, y$ are two arbitrary numbers with $x < y$, this means that $h(\cdot)$ is a constant function on $\mathbb{R}$. This would contradict the assumption that $h(\cdot)$ is non-constant. Hence, $\inf K > 0$.

Let $\eta = \inf K$. Since $\eta > 0$ and $\eta < a$ for any $a \in K$, we only need to show that $\eta \in K$. By definition, there exists $a_n \in K$ with $a_n \to \eta$. For any $x$, we apply the continuity of $h(\cdot)$ and obtain that $h(x) = h(x + a_n) \to h(x + \eta)$, which means that $h(x) = h(x + \eta)$ and thus $\eta \in K$. Therefore, $\eta$ is the smallest element of $K$. \hfill \Box

Lemma 12. Define the function $G(a) = \ln \frac{F(a)}{1 - F(a)}.$ Assume that $s > t > 0$. If

$$\inf_{a \in \mathbb{R}} [G(s + \alpha) - G(\alpha)] > \sup_{\xi \in \mathbb{R}} [G(t + \xi) - G(\xi)] > 0,$$

then $\mathcal{A}(s) \cap \mathcal{A}(t) = \emptyset$.

Proof. Let $\delta = s - t$. By $s > t$, $\delta > 0$. We proceed in two steps.

**Step 1:** show that there exists $v \in \mathbb{R}^3$ such that $v'p(s, \alpha) < 0 < v'p(t, \xi)$ for any $\alpha, \xi \in \mathbb{R}$.

Consider $v = (v_1, 1 - v_1, -1)'$. We would like to choose $v_1$ such that for any $\alpha, \xi \in \mathbb{R}$

$$v_1 F(\alpha) + (1 - v_1) F(s + \alpha) - F(\alpha) F(s + \alpha) < 0 \quad (19)$$

and

$$v_1 F(\xi) + (1 - v_1) F(t + \xi) - F(\alpha) F(t + \xi) > 0. \quad (20)$$
Since \( s > t > 0 \), it suffices to choose any \( v_1 \) such that

\[
\sup_{\alpha \in \mathbb{R}} \frac{F(s + \alpha)[1 - F(\alpha)]}{F(s + \alpha) - F(\alpha)} < v_1 < \inf_{\xi \in \mathbb{R}} \frac{F(t + \xi)[1 - F(\xi)]}{F(t + \xi) - F(\xi)}.
\]

By \( F(\cdot) = [\exp(-G(\cdot)) + 1]^{-1} \), this means that we can choose any \( v_1 \) such that

\[
\frac{1}{1 - \sup_{\alpha \in \mathbb{R}} \exp[G(\alpha) - G(s + \alpha)]} < v_1 < \frac{1}{1 - \inf_{\xi \in \mathbb{R}} \exp[G(\xi) - G(t + \xi)]}.
\]

Since \( \inf_{\alpha \in \mathbb{R}}[G(s + \alpha) - G(\alpha)] > \sup_{\xi \in \mathbb{R}}[G(t + \xi) - G(\xi)] > 0 \), we have \( \sup_{\alpha \in \mathbb{R}}[G(\alpha) - G(s + \alpha)] < \inf_{\xi \in \mathbb{R}}[G(\xi) - G(t + \xi)] < 0 \) and thus a choice of \( v_1 \) in the above display is clearly possible. Therefore, by \( v = (v_1, 1 - v_1, -1)' \), we have found \( v \in \mathbb{R}^3 \) such that (19) and (20) hold.

**Step 2:** show the final result.

We proceed by contradiction. Suppose that \( \mathcal{A}(s) \cap \mathcal{A}(t) \neq \emptyset \). Then there exists \( r \in \mathcal{A}(s) \cap \mathcal{A}(t) \). Since \( r \in \mathcal{A}(s) \subset \mathbb{R}^3 \), by Carathéodory’s theorem (e.g., Theorem 5.32 in Aliprantis and Border (2006)), there exist \( \alpha_1, \alpha_2, \alpha_3, \alpha_4 \in \mathbb{R} \) and \( \lambda_1, ..., \lambda_4 \geq 0 \) such that \( r = \sum_{j=1}^{4} \lambda_j p(s, \alpha_j) \) and \( \sum_{j=1}^{4} \lambda_j = 1 \). Similarly, \( r = \sum_{j=1}^{4} \rho_j p(t, \xi_j) \), where \( \xi_1, ..., \xi_4 \in \mathbb{R} \) and \( \rho_1, ..., \rho_4 \geq 0 \) satisfy \( \sum_{j=1}^{4} \rho_j = 1 \). Therefore, for any \( v \in \mathbb{R}^3 \),

\[
\max_{1 \leq j \leq 4} p(s, \alpha_j)'v = \sum_{j=1}^{4} \lambda_j p(s, \alpha_j)'v = r'v = \sum_{j=1}^{4} \rho_j p(t, \xi_j)'v \geq \min_{1 \leq j \leq 4} p(t, \xi_j)'v.
\]

However, in Step 1, we have shown that there exists \( v \neq (0, 0, 0)' \) such that \( v'p(s, \alpha) < 0 < v'p(t, \xi) \) for any \( \alpha, \xi \in \mathbb{R} \). This is a contradiction. \( \square \)

**Lemma 13.** Assume that \( s > t > 0 \). Suppose that \( \hat{G} \) is a periodic function with a positive period \( \eta > 0 \). If either \( s/\eta \) or \( t/\eta \) is an integer, then there exists \( \varepsilon \in (0, (s - t)/\eta) \) such that

\[
\inf_{\alpha \in \mathbb{R}} [G(s - \varepsilon + \alpha) - G(\alpha)] > \sup_{\xi \in \mathbb{R}} [G(t + \varepsilon + \xi) - G(\xi)] > 0.
\]

**Proof.** Since \( \hat{G}(l \cdot \eta + \xi) - \hat{G}(\xi) = 0 \) for any \( \xi \in \mathbb{R} \) and for any \( l \in \mathbb{Z} \) (\( \mathbb{Z} \) denotes the set of all integers), it follows that \( G(l \cdot \eta + \xi) - G(\xi) \) does not depend on \( \xi \), which
means that \(G(l \cdot \eta + \xi) - G(\xi) = G(l \cdot \eta + 0) - G(0)\). We notice that

\[
G(l \cdot \eta) = G(0) + \int_0^{l \cdot \eta} \dot{G}(t) dt = G(0) + \sum_{j=1}^{l} \int_{\eta(j-1)}^{\eta j} \dot{G}(t) dt = G(0) + \sum_{j=1}^{l} \int_0^\eta \dot{G}(t) dt = lq_0 + G(0),
\]

where \(q_0 = G(\eta) - G(0)\). Thus, we have that for any \(\xi \in \mathbb{R}\) and for any \(l \in \mathbb{Z}\),

\[
G(l \cdot \eta + \xi) = G(\xi) + lq_0.
\] (21)

Notice that we can represent all the real numbers as \((k + r)\eta\) with \(k \in \mathbb{Z}\) and \(r \in [0, 1)\). Then it suffices to show that for some \(\varepsilon > 0\),

\[
\inf_{(k_1, r_1) \in \mathbb{Z} \times [0, 1)} [G(s - \varepsilon + (k_1 + r_1)\eta) - G((k_1 + r_1)\eta)] > \sup_{(k_2, r_2) \in \mathbb{Z} \times [0, 1)} [G(t + \varepsilon + (k_2 + r_2)\eta) - G((k_2 + r_2)\eta)] > 0.
\]

By (21), \(G(s - \varepsilon + (k_1 + r_1)\eta) - G((k_1 + r_1)\eta) = G(s - \varepsilon + r_1\eta) - G(r_1\eta)\), which does not depend on \(k_1\); similarly, \(G(t + \varepsilon + (k_2 + r_2)\eta) - G((k_2 + r_2)\eta) = G(t + \varepsilon + r_2\eta) - G(r_2\eta)\). Thus, it suffices to show that for some \(\varepsilon > 0\),

\[
\inf_{r_1 \in [0, 1)} [G(s - \varepsilon + r_1\eta) - G(r_1\eta)] > \sup_{r_2 \in [0, 1)} [G(t + \varepsilon + r_2\eta) - G(r_2\eta)].
\] (22)

We now construct \(\varepsilon > 0\) that satisfies (22). We consider two cases: \(s/\eta \in \mathbb{Z}\) or \(t/\eta \in \mathbb{Z}\).

**Case 1**: suppose \(s/\eta \in \mathbb{Z}\). Let \(k_1 = s/\eta\). Then by (21), we have \(G(s + r_1\eta) - G(r_1\eta) = k_1q_0\) and \(G(t + r_2\eta) = G(t - s + r_2\eta) + k_1q_0\). Hence,

\[
\inf_{r_1 \in [0, 1)} [G(s + r_1\eta) - G(r_1\eta)] - \sup_{r_2 \in [0, 1)} [G(t + r_2\eta) - G(r_2\eta)]
\]

\[
= k_1q_0 - \sup_{r_2 \in [0, 1)} [G(t - s + r_2\eta) + k_1q_0 - G(r_2\eta)]
\]

\[
= \inf_{r_2 \in [0, 1)} [G(r_2\eta - G(r_2\eta - (s - t))] \geq \inf_{r_2 \in [0, 1)} [G(r_2\eta) - G(r_2\eta - (s - t))].
\]

Since \(r_2 \mapsto G(t + r_2\eta) - G(t - s + r_2\eta)\) is a continuous function and \([0, 1]\) is a compact set, the above infimum is achieved at point in \([0, 1]\). Since \(G\) is strictly
increasing and $s - t > 0$, we have that

$$\inf_{r_2 \in [0,1]} [G(r_2\eta) - G(r_2\eta - (s - t))] > 0.$$  

(Of otherwise, there would exist $r_{2,*} \in [0,1]$ such that $r_{2,*}\eta = r_{2,*}\eta - (s - t)$.) Let $\Delta_1 = \inf_{r_2 \in [0,1]}[G(r_2\eta) - G(r_2\eta - (s - t))]$. We now choose $\varepsilon \in (0, \min\{1, (s - t)/4\}]$ such that

$$\sup_{r_1 \in [0,1]} [G(r_1\eta) - G(r_1\eta - \varepsilon)] \leq \Delta_1/4$$

and

$$\sup_{r_2 \in [0,1]} [G(r_2\eta - (s - t) + \varepsilon) - G(r_2\eta - (s - t))] \leq \Delta_1/4$$

To see that this is possible, notice that $G(\cdot)$ is continuous and thus is uniformly continuous on the compact set $[-1, \eta] \cup [(s - t), \eta + 1 - (s - t)]$. Thus, there exists a constant $\kappa_1 > 0$ such that $\sup_{r_1 \in [0,1]}[G(r_1\eta) - G(r_1\eta - \varepsilon)] \leq \kappa_1 \varepsilon$ and $\sup_{r_2 \in [0,1]}[G(r_2\eta - (s - t) + \varepsilon) - G(r_2\eta - (s - t))] \leq \kappa_1 \varepsilon$. Hence, we can simply choose $\varepsilon = \min\{1, (s - t)/4, \Delta_1/(4\kappa_1)\}$. Therefore,

$$\inf_{r_1 \in [0,1]} [G(s - \varepsilon + r_1\eta) - G(r_1\eta)] - \sup_{r_2 \in [0,1]} [G(t + \varepsilon + r_2\eta) - G(r_2\eta)]$$

$$= \inf_{r_1 \in [0,1]} [G(-\varepsilon + r_1\eta) + k_1q_0 - G(r_1\eta)] - \sup_{r_2 \in [0,1]} [G(t - s + \varepsilon + r_2\eta) + k_1q_0 - G(r_2\eta)]$$

$$= \inf_{r_1 \in [0,1]} [G(-\varepsilon + r_1\eta) - G(r_1\eta)] - \sup_{r_2 \in [0,1]} [G(t - s + \varepsilon + r_2\eta) - G(r_2\eta)]$$

$$= -\sup_{r_1 \in [0,1]} [G(r_1\eta) - G(r_1\eta - \varepsilon)] + \inf_{r_2 \in [0,1]} [G(r_2\eta) - G(r_2\eta - (s - t) + \varepsilon)]$$

$$\geq -\sup_{r_1 \in [0,1]} [G(r_1\eta) - G(r_1\eta - \varepsilon)] + \inf_{r_2 \in [0,1]} [G(r_2\eta) - G(r_2\eta - (s - t))]$$

$$+ \inf_{r_2 \in [0,1]} [G(r_2\eta - (s - t)) - G(r_2\eta - (s - t) + \varepsilon)]$$

$$= -\sup_{r_1 \in [0,1]} [G(r_1\eta) - G(r_1\eta - \varepsilon)] + \inf_{r_2 \in [0,1]} [G(r_2\eta) - G(r_2\eta - (s - t))]$$

$$- \sup_{r_2 \in [0,1]} [G(r_2\eta - (s - t) + \varepsilon) - G(r_2\eta - (s - t))]$$

$$(i) \geq -\Delta_1/4 + \inf_{r_2 \in [0,1]} [G(r_2\eta) - G(r_2\eta - (s - t))] - \Delta_1/4 \geq -\Delta_1/4 + \Delta_1 - \Delta_1/4 > 0,$$

where (i) follows by the construction of $\varepsilon$ and (ii) follows by the definition of $\Delta_1$.  

31
**Case 2:** suppose $t/\eta \in \mathbb{Z}$. Let $k_2 = t/\eta$. Then by (21), we have $G(t + r_2 \eta) - G(r_2 \eta) = k_2 q_0$ and $G(s + r_1 \eta) = G(s - t + r_1 \eta) + k_2 q_0$. Hence,

$$\inf_{r_1 \in [0,1]} [G(s + r_1 \eta) - G(r_1 \eta)] - \sup_{r_2 \in [0,1]} [G(t + r_2 \eta) - G(r_2 \eta)]$$

$$= \inf_{r_1 \in [0,1]} [G(s - t + r_1 \eta) + k_2 q_0 - G(r_1 \eta)] - k_2 q_0$$

$$= \inf_{r_1 \in [0,1]} [G(s - t + r_1 \eta) - G(r_1 \eta)] \geq \inf_{r_1 \in [0,1]} [G(s - t + r_1 \eta) - G(r_1 \eta)].$$

Since $r_1 \mapsto G(s - t + r_1 \eta) - G(r_1 \eta)$ is a continuous function and $[0, 1]$ is a compact set, the above infimum is achieved at point in $[0, 1]$. Since $G$ is strictly increasing and $s - t > 0$, we have

$$\Delta_2 := \inf_{r_1 \in [0,1]} [G(s - t + r_1 \eta) - G(r_1 \eta)] > 0.$$

We now choose $\varepsilon \in (0, \min\{1, (s - t)/4\})$ such that

$$\sup_{r_2 \in [0,1]} [G(\varepsilon + r_2 \eta) - G(r_2 \eta)] \leq \Delta_2/4$$

and

$$\sup_{r_1 \in [0,1]} [G(s - t + r_1 \eta) - G(s - \varepsilon - t + r_1 \eta)] \leq \Delta_2/4$$

To see that this is possible, again notice that $G(\cdot)$ is uniformly continuous on the compact set $[0, \eta + 1] \bigcup [(s - t), \eta + 1 + (s - t)]$. Thus, there exists a constant $\kappa_2 > 0$ such that $\sup_{r_2 \in [0,1]} [G(\varepsilon + r_2 \eta) - G(r_2 \eta)] \leq \kappa_2 \varepsilon$ and $\sup_{r_1 \in [0,1]} [G(s - t + r_1 \eta) - G(s - \varepsilon - t + r_1 \eta)] \leq \kappa_2 \varepsilon$. Hence, we can simply choose $\varepsilon = \min\{1, (s - t)/4, \Delta_2/(4\kappa_2)\}$. Therefore,

$$\inf_{r_1 \in [0,1]} [G(s - \varepsilon + r_1 \eta) - G(r_1 \eta)] - \sup_{r_2 \in [0,1]} [G(t + \varepsilon + r_2 \eta) - G(r_2 \eta)]$$

$$= \inf_{r_1 \in [0,1]} [G(s - \varepsilon - t + r_1 \eta) + k_2 q_0 - G(r_1 \eta)] - \sup_{r_2 \in [0,1]} [G(\varepsilon + r_2 \eta) + k_2 q_0 - G(r_2 \eta)]$$

$$= \inf_{r_1 \in [0,1]} [G(s - \varepsilon - t + r_1 \eta) - G(r_1 \eta)] - \sup_{r_2 \in [0,1]} [G(\varepsilon + r_2 \eta) - G(r_2 \eta)]$$

$$\geq \inf_{r_1 \in [0,1]} [G(s - \varepsilon - t + r_1 \eta) - G(s - t + r_1 \eta)] + \inf_{r_1 \in [0,1]} [G(s - t + r_1 \eta) - G(r_1 \eta)]$$

$$- \sup_{r_2 \in [0,1]} [G(\varepsilon + r_2 \eta) - G(r_2 \eta)]$$
\[
\begin{align*}
&= - \sup_{r_1 \in [0,1)} [G(s - t + r_1 \eta) - G(s - \varepsilon - t + r_1 \eta)] + \inf_{r_1 \in [0,1)} [G(s - t + r_1 \eta) - G(r_1 \eta)] \\
&\quad - \sup_{r_2 \in [0,1)} [G(\varepsilon + r_2 \eta) - G(r_2 \eta)] \\
&\geq -\Delta_2/4 + \inf_{r_1 \in [0,1)} [G(s - t + r_1 \eta) - G(r_1 \eta)] - \Delta_2/4 \geq -\Delta_2/4 + \Delta_2 - \Delta_2/4 > 0,
\end{align*}
\]

where (i) follows by the construction of \( \varepsilon \) and (ii) follows by the definition of \( \Delta_2 \).

Thus, we have verified (22) and thus, \( \inf_{\alpha \in \mathbb{R}}[G(s - \varepsilon + \alpha) - G(\alpha)] > \sup_{\xi \in \mathbb{R}}[G(t + \varepsilon + \xi) - G(\xi)] \) for some \( \varepsilon > 0 \). Clearly \( \sup_{\xi \in \mathbb{R}}[G(t + \varepsilon + \xi) - G(\xi)] > 0 \) since \( t, \varepsilon > 0 \) and \( G \) is strictly increasing (just choose \( \xi = 0 \)).

**Proof of Theorem 3.** The first claim follows by Lemma 11. It remains to prove the second claim.

We partition \( \beta = (\beta'_1, \beta_2)' \). By assumption, there exists \( z_0 \in \mathbb{Z}^0 \) such that \( k = (z'_0 \beta_1 + \beta_2)/\eta \) is an integer. Clearly \( k > 0 \) since \( \beta \in \mathcal{B}_+ \) (i.e., \( z' \beta_1 + \beta_2 > 0 \) for any \( z \in \mathcal{Z} \)). Consider any \( b \in \mathcal{B}_+ \) such that \( b = (b'_1, b_2) \neq \beta \). We proceed in two steps.

**Step 1:** show that there exists \( z_* \in \mathbb{Z}^0 \) such that \((z'_*, 1)\beta \neq (z'_*, 1)b \) and one of \((z'_*, 1)\beta/\eta \) and \((z'_*, 1)b/\eta \) is an integer.

Suppose that \( z'_0b_1 + b_2 \neq k\eta \). Then we take \( z_* = z_0 \) and have \((z'_*, 1)\beta = k\eta \neq (z'_*, 1)b \). Clearly, one of \((z'_*, 1)\beta/\eta \) and \((z'_*, 1)b/\eta \) is an integer.

Suppose that \( z'_0b_1 + b_2 = k\eta \). We notice that \( \beta_1 \neq b_1 \). To see this, suppose \( \beta_1 = b_1 \). Then \( z'_0b_1 = z'_0\beta_1 \). Notice that \( z'_0b_1 + b_2 = k\eta = z'_0\beta_1 + b_2 \). We have \( b_2 = \beta_2 \). By \( \beta_1 = b_1 \), this would imply \( \beta = b \), contradicting \( \beta \neq b \). Hence, \( \beta_1 \neq b_1 \) and thus at least one of \( \beta_1 \) and \( b_1 \) is not zero. Then we only need to discuss three cases.

**Case 1:** suppose \( \beta_1 \neq 0 \) and \( b_1 \neq 0 \). Since \( z_0 \in \mathbb{Z}^0 \) and \( \beta_1 \neq 0 \), \( z_1 = z_0 + (I - \beta_1 \beta_1/\|\beta_1\|^2)\Delta \in \mathbb{Z}^0 \) when \( \|\Delta\|_2 \) is small enough. We choose \( \Delta \) such that \( b'_1(I - \beta_1 \beta_1/\|\beta_1\|^2)\Delta > 0 \); this is possible because \( b_1 \neq \beta_1 \), \( \beta_1 \neq 0 \) and \( b_1 \neq 0 \). Then \( z'_1\beta_1 + \beta_2 = z'_0\beta_1 + \beta_2 = k\eta \) and \( z'_1b_1 + b_2 = z'_0b_1 + b_2 + (z_1 - z_0)'b_1 = k\eta + b'_1(I - \beta_1 \beta_1/\|\beta_1\|^2)\Delta > k\eta \). We now define \( t = z'_1\beta_1 + \beta_2 \) and \( s = z'_1b_1 + b_2 \). Thus \( s = k\eta > t > 0 \). In other words, we can take \( z_* = z_1 \).

**Case 2:** suppose \( \beta_1 \neq 0 \) and \( b_1 = 0 \). Since \( z'_0b_1 + b_2 = k\eta \) and \( b_1 = 0 \), we have \( b_2 = k\eta \). Since \( \mathbb{Z}^0 \) contains an open neighborhood of \( z_0 \) and \( \beta_1 \neq 0 \), we can find \( z_1 \) close enough to \( z_0 \) such that \((z_1 - z_0)'\beta_1 > 0 \). This means that \( z'_1\beta_1 + \beta_2 = (z_1 - z_0)'\beta_1 + (z'_0\beta_1 + \beta_2) > k\eta \) and \( z'_1b_1 + b_2 = b_2 = k\eta \). We now define \( s = z'_1\beta_1 + \beta_2 \) and \( t = z'_1b_1 + b_2 \). Thus, \( s > t = k\eta > 0 \). In other words, we can take \( z_* = z_1 \).
Case 3: suppose $\beta_1 = 0$ and $b_1 \neq 0$. The argument is analogous to the above case. We repeat it here for completeness. Since $z_0 \beta_1 + \beta_2 = k\eta$ and $\beta_1 = 0$, we have $\beta_2 = k\eta$. Since $\mathcal{Z}^\circ$ contains an open neighborhood of $z_0$ and $b_1 \neq 0$, we can find $z_1$ close enough to $z_0$ such that $(z_1 - z_0)b_1 > 0$. This means that $z_1'b_1 + b_2 = (z_1 - z_0)b_1 + (z_0'b_1 + b_2) > k\eta$ and $z_1'\beta_1 + \beta_2 = \beta_2 = k\eta$. We now define $s = z_1'b_1 + b_2$ and $t = z_1'\beta_1 + \beta_2$. Thus, $s > t = k\eta > 0$. In other words, we can take $z_* = z_1$.

**Step 2:** show the final result.

We now prove the result in two cases.

**Case 1:** suppose $(z'_*, 1)\beta > (z'_*, 1)b$. Define $s = (z'_*, 1)\beta$ and $t = (z'_*, 1)b$. Then $s > t > 0$ and either $s/\eta$ or $t/\eta$ is an integer. By Lemma 13, there exists $\varepsilon \in (0, (s - t)/4]$ such that

$$\inf_{\alpha \in \mathbb{R}} [G(s - \varepsilon + \alpha) - G(\alpha)] > \sup_{\xi \in \mathbb{R}} [G(t + \varepsilon + \xi) - G(\xi)] > 0.$$  

Let $\mathbb{B} = \{v \in \mathbb{R}^{K-1} : \|v\|_2 \leq 1\}$ be the unit ball. Notice that $z_* \in \mathcal{Z}^\circ$. Clearly, there exists $\tau > 0$ such that $z_* + \tau \mathbb{B} \subset \mathcal{Z}^\circ$, $\tau|v'\beta_1| \leq \varepsilon$ and $\tau|v'b_1| \leq \varepsilon$ for any $v \in \mathbb{B}$. Define $\mathcal{Z}_\tau = z_* + \tau \mathbb{B}$. By construction, $\mathcal{Z}_\tau$ is an open set contained in $\mathcal{Z}^\circ$. Notice that for any $z \in \mathcal{Z}_\tau$, $(z', 1)\beta = (z'_*, 1)\beta + (z - z_*)\beta_1 \geq (z'_*, 1)\beta - \varepsilon = s - \varepsilon$ and $(z', 1)b = (z'_*, 1)b + (z - z_*)\beta_1 \leq (z'_*, 1)b + \varepsilon = t + \varepsilon$. Thus, the above display implies that for any $z \in \mathcal{Z}_\tau$, $(z', 1)\beta \geq s - \varepsilon > t + \varepsilon \geq (z'_*, 1)b$ and thus

$$\inf_{\alpha \in \mathbb{R}} [G((z', 1)\beta + \alpha) - G(\alpha)] \geq \inf_{\alpha \in \mathbb{R}} [G(s - \varepsilon + \alpha) - G(\alpha)]$$

$$> \sup_{\xi \in \mathbb{R}} [G(t + \varepsilon + \xi) - G(\xi)] \geq \sup_{\xi \in \mathbb{R}} [G((z', 1)b + \xi) - G(\xi)] > 0.$$  

By Lemma 12, it follows that for any $z \in \mathcal{Z}_\tau$, $A((z', 1)\beta) \cap A((z', 1)b) = \emptyset$.

**Case 2:** suppose $(z'_*, 1)\beta < (z'_*, 1)b$. The argument is analogous to Case 1 except that we swap the roles of $\beta$ and $b$. We omit it for simplicity.

Therefore, we have proved that for any $b \in \mathcal{B}_+$ with $b \neq \beta$, there exists an open set $\mathcal{Z}_b \subset \mathcal{Z}$ such that for any $z \in \mathcal{Z}_b$, $A(w'\beta) \cap A(w'b) = \emptyset$ with $w = (z', 1)'$. Clearly $\mathcal{Z}_b \cap \mathcal{Z} \neq \emptyset$. By Lemma 3, $P(Z \in \mathcal{Z}_b) > 0$.  

\[ \square \]
A.4 Proof of results in Section 3

Proof of Lemma 2. We prove the first claim since the proof of the second claim is analogous.

The “only if” part is easy. Assume that \( \phi(X) = E(Y_1 - Y_0 \mid X) \leq 0 \) almost surely. Then \( \rho(q) = E1\{W'q \geq 0\} \phi(X) \leq 0 \) for any \( q \in \mathbb{R}^K \).

We now prove the “if” part. Suppose that \( \sup_{q \in \mathbb{R}^K} \rho(q) \leq 0 \). We consider \( q = \beta \). By Lemma 1, \( \rho(\beta) = E1\{W'\beta \geq 0\}(Y_1 - Y_0) = E1\{W'\beta \geq 0\} \cdot \phi(X) = E1\{\phi(X) \geq 0\} \cdot \phi(X) \). Since \( \rho(\beta) \leq \sup_{q \in \mathbb{R}^K} \rho(q) \leq 0 \), it follows that \( P(\phi(X) > 0) = 0 \), which means that \( \phi(X) \leq 0 \) almost surely. 

References

Aliprantis, C. D. and Border, K. C. (2006). Infinite Dimensional Analysis: A Hitchhiker’s Guide. Springer Science & Business Media.

Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. Journal of the Royal Statistical Society: Series B (Methodological), 32(2):283–301.

Cattaneo, M. D., Jansson, M., and Nagasawa, K. (2020). Bootstrap-based inference for cube root asymptotics. Econometrica, 88(5):2203–2219.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. The Review of Economic Studies, 47(1):225–238.

Chamberlain, G. (2010). Binary response models for panel data: Identification and information. Econometrica, 78(1):159–168.

Davezies, L., D’Haultfoeuille, X., and Laage, L. (2021). Identification and estimation of average marginal effects in fixed effects logit models. arXiv:2105.00879.

Davezies, L., D’Haultfoeuille, X., and Mugnier, M. (2020). Fixed effects binary choice models with three or more periods.

Dudley, R. M. (2002). Real Analysis and Probability. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition.
Honoré, B. E., Muris, C., and Weidner, M. (2021). Dynamic ordered panel logit models. *arXiv:2107.03253*.

Honoré, B. E. and Weidner, M. (2020). Moment conditions for dynamic panel logit models with fixed effects. *arXiv:2005.05942*.

Kim, J. and Pollard, D. (1990). Cube root asymptotics. *The Annals of Statistics*, pages 191–219.

Manski, C. F. (1987). Semiparametric analysis of random effects linear models from binary panel data. *Econometrica*, 55:357–362.

Rasch, G. (1960). Probabilistic models for some intelligence tests and attainment tests. *Danish Institute for Educational Research, Copenhagen, Denmark*.

Rockafellar, R. T. (1970). *Convex Analysis*, volume 36. Princeton University Press.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Science & Business Media.