Identification of the Marginal Treatment Effect with Multivalued Treatments

Toshiki Tsuda
Graduate School of Economics, University of Tokyo, Tokyo

Abstract

Heckman et al. (2008) examine the identification of the marginal treatment effect (MTE) with multivalued treatments by extending the approach of Heckman and Vytlacil (1999). Lee and Salanié (2018) study the identification of conditional expectations given unobserved heterogeneity; in Section 5.2 of their paper, they analyze the identification of MTE under the same selection mechanism as Heckman et al. (2008).

We note that the construction of their model in Section 5.2 in Lee and Salanié (2018) is incomplete, and we establish sufficient conditions for the identification of MTE with an improved model. While we reduce the unordered multiple-choice model to the binary treatment setting as in Heckman et al. (2008), we can identify the MTE defined as a natural extension of the MTE using the binary treatment in Heckman and Vytlacil (2005). Further, our results can identify other parameters such as the marginal distribution of potential outcomes.

Keywords: Identification, treatment effect, unobserved heterogeneity, multivalued treatments, endogeneity, instrumental variable

JEL classification: C14, C31

*I am grateful to my advisor, Katsumi Shimotsu, for his continuous guidance and support. I am also thankful to Mark Henry, Sokbae Lee, Ryo Okui, Bernard Salanié and Takahide Yanagi for their insightful comments. Further, I benefited from comments of the seminar participants at The University of Tokyo and Tohoku University.

†Toshiki Tsuda, Graduate School of Economics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan; World-leading Innovative Graduate Study of Advanced Economics, The University of Tokyo; Email: tsuda-toshiki@g.ecc.u-tokyo.ac.jp
1 Introduction

Assessing heterogeneity in treatment effects is important for precise treatment evaluation. The marginal treatment effect (MTE) provides rich information on heterogeneity across economic agents in terms of their observed and unobserved characteristics. Further, once the MTE is estimated, researchers can obtain other treatment effects, such as the average treatment effect (ATE), average treatment effect on the treated, and local ATE (LATE).

For the binary treatment models, Heckman and Vytlacil (1999) establish the local instrumental variable (LIV) framework to identify MTE. They assume individuals decide their choice based on the generalized Roy model that is separable in terms of observed and unobserved variables. Vytlacil (2002) shows that the separable threshold-crossing model in the LIV approach plays the same role as the monotonicity assumption for identifying LATE (Imbens and Angrist 1994).

In many applications, selection problems cannot be adequately described using single-crossing models. For example, vocational programs provide various types of training to participants, and college choice involves numerous dimensions to respond to varied incentives. The literature has developed treatment effects with multivalued treatments, such as LATE (Angrist and Imbens 1995), MTE (Heckman et al. 2006, 2008; Heckman and Vytlacil 2007b; Heckman and Pinto 2018; Lee and Salanie 2018) and instrumental variable quantile regression (Fusejima 2022). Heckman and Pinto (2018) introduce an “unordered monotonicity” condition that is weaker than a monotonicity condition under models with multiple choices, and show the identification of several treatment effects.

Heckman et al. (2006, 2008) and Heckman and Vytlacil (2007b) examine multivalued treatment effects based on the discrete choice model that is additively separable in instruments and errors. This model is a generalization of the multiple logit model, and has been extensively studied in economics since the seminal work of McFadden (1974). In theoretical research, Matzkin (1993) establishes sufficient conditions for the nonparametric identification of the discrete choice model. In applied research, Dahl (2002) employs this model to study the effects of self-selected migration on returns to college. Kline and Walters (2016) use the discrete multiple-choice model as a self-selection model to analyze the Head Start program’s cost-effectiveness.

Heckman and Vytlacil (2007b) and Heckman et al. (2008) expand the LIV ap-
approach to a model with multivalued treatments generated by a general unordered choice model. They study the identification conditions of several types of treatment effects including a treatment effect of one specific choice versus the next best alternative, and a treatment effect of one specific group of choices versus the other group. Moreover, they show the MTE of one specified choice versus another choice can be identified if we assume utilities of all the other options are quite small. For example, when the number of treatments is three, [Heckman and Vytlacil (2007b) and Heckman et al. (2008)] assume the utility of one option is sufficiently small, and reduce the model to the binary treatment model in effect. Then, they identify the MTE using identification results in the binary model.

Lee and Salanié (2018) investigate the identification of conditional expectations given unobserved variables based on multinomial choice models characterized by a combination of separable threshold-crossing rules. They assume the existence of continuous instruments and identify several causal effects with identified thresholds. In Section 4, they introduce several commonly used models that their framework covers and show the identification of MTE for each model with sufficient conditions for the identification of thresholds. In Section 5.2, Lee and Salanié (2018) analyze the identification of the MTE studied by [Heckman and Vytlacil (2007b) and Heckman et al. (2008)]. They argue that if we know the thresholds and apply the result of Lee and Salanié (2018), it is possible to identify MTE without reducing the model to the binary treatment case.

We notice that the construction of threshold-crossing rules in Section 5.2 of Lee and Salanié (2018) is incomplete. By construction, two unobserved variables uniquely determine the other unobserved variable, which violates an assumption they require for the identification of MTE. We establish sufficient conditions for identifying MTE under an improved model. Under the assumptions for the threshold variation, we reduce the unordered multiple-choice model to a binary treatment setting, as in [Heckman and Vytlacil (2007b) and Heckman et al. (2008)]. A major distinction between Heckman et al. (2008) and our study is the object of the identification. By extending the LIV approach, [Heckman and Vytlacil (2007b) and Heckman et al. (2008)] identify MTE. We study the identification of conditional expectations on unobserved heterogeneity. While the MTE that [Heckman and Vytlacil (2007b) and Heckman et al. (2008)] identify depends on an unknown cumulative function of unobserved error terms, we identify the MTE conditional on the unobserved variable normalized
to have a uniform distribution on \([0, 1]\). In the binary treatment case, \cite{heckman2005} identify the MTE conditional on unobserved heterogeneity that uniformly ranges from 0 to 1. Therefore, the MTE we identify is a natural extension of the MTE defined in \cite{heckman2005} to the multiple-choice model. Moreover, our results help identify not only MTE but also other measurements, such as quantile treatment effect.

We further establish a sufficient condition for identifying the thresholds. \cite{heckman2008} identify the MTE given known utility functions and refer to \cite{heckman2007} for the identification of utility functions. The identification results of \cite{lee2018} are conditional on the identified thresholds. While they rely on Matzkin’s (1993) result for the identification of thresholds, she does not provide sufficient conditions for nonparametric identification of thresholds. Our result for identifying thresholds does not depend on Matzkin’s (1993) results and enables MTE estimation under multivalued treatments.

The remainder of this paper is organized as follows: In Section 2, we propose the basic settings and notation used in this study. We demonstrate that the model construction in Section 5.2 of \cite{lee2018} is incomplete. In Section 3, we suggest an improved model, explain our identification strategy, and show identification of the MTE. Section 4 establishes sufficient conditions for nonparametric identification of the thresholds. In Sections 2–4, we consider the case when the number of treatments is three. Section 5 discusses the general case and identifies the MTE. Proofs of the main results and some auxiliary results are collected in Appendix A.

\section{Multiple Discrete Choice Model and \cite{lee2018}}

\subsection{Basic Setup}

Let := denote “equals by definition,” and let a.s. denote “almost surely.” Let \( \mathbb{1}\{\cdot\} \) denote the indicator function. For random variables \( X \) and \( Z \), \( f_X(\cdot) \) denotes the probability density function of \( X \). \( F_{X|Z}(\cdot) \) and \( Q_{X|Z}(\cdot) \) denotes the distribution and quantile functions of \( X \) given \( Z \), respectively. Let \( \|\cdot\|_d \) denote the Euclidean norm.
in $\mathbb{R}^d$, and let $\tilde{N}_d(a, \delta)$ be a $\delta$-neighborhood of $a$ in $(0, 1)^d$, that is,

$$\tilde{N}_d(a, \delta) := \{x \in (0, 1)^d : \|a - x\|_d < \delta\} = \{x \in \mathbb{R}^d : \|a - x\|_d < \delta\} \cap (0, 1)^d.$$ 

For any set $A$ and $B$, $\bar{A}$ denotes the closure of $A$, and $A \setminus B$ denotes the set difference, that is, the set of elements in $A$ but not in $B$. Let $|A|$ denote the cardinality of $A$.

As in Heckman et al. (2006, 2008) and Heckman and Vytlacil (2007b), we consider the following discrete choice model. Let $\mathcal{K}$ be the set of treatments comprising $K (= |\mathcal{K}|)$ elements, and let $Z$ be the vector of the observed random variables. For each $k$, we define $R_k(Z)$ as an unknown function that maps from $\mathbb{R}^\text{dim}(Z)$ to $\mathbb{R}$ and define $U_k$ as an unobserved continuous random variable whose support is $\mathbb{R}$. Let $\{Y_k : k \in \mathcal{K}\}$ be a potential outcome. $D_k$ takes the value one if the agent receives treatment $k$. By extending the definition of the treatment variable in the binary treatment model, we formulate the treatment decision as follows:

$$D_k := 1\{R_k(Z) - U_k > \max_{j \neq k}(R_j(Z) - U_j)\}, \quad (1)$$

where $\Pr((R_k(Z) - U_k) = (R_j(Z) - U_j)) = 0$ for $j \neq k$. The observed outcome and treatment are expressed as $D = \sum_{k=0}^{K-1} kD_k$ and $Y = \sum_{k=0}^{K-1} D_kY_k$, respectively. The data contains covariates $X$ and instruments $Z$. Throughout this article, we condition on the value of $X$ and suppress it from the notation. Let the support of $Y$ and $Z$ be $Y \subset \mathbb{R}$ and $Z \subset \mathbb{R}^\text{dim}(Z)$, respectively.

Intuitively, by interpreting $R_k(Z)$ and $U_k$ as utility and cost from choice $k$, respectively, this discrete multiple-choice model states that an agent makes a choice that gives the largest benefit. From this intuition, we regard model (1) as a straightforward extension of the generalized Roy model.

Model (1) has been studied extensively in economics since the seminal work of McFadden (1974). Matzkin (1993) establishes sufficient conditions for the nonparametric identification of the utility functions and the cumulative joint density of unobserved random terms. The multinomial choice model has also been used in applied research. Dahl (2002) uses this model to study the effects of self-selected migration on the return to college. Kline and Walters (2016) adopts the discrete multiple-choice model as a self-selection model and analyzes the Head Start program’s cost-effectiveness in the presence of the substitute preschools. Kirkeboen et al. (2016) examine the
effect of types of postsecondary education on the gains in earnings. They find that
the estimated payoffs are consistent with agents choosing fields based on a discrete
multiple-choice model.

Heckman and Vytlacil (2007b) and Heckman et al. (2008) expand the LIV ap-
proach to model (1). They study the identification conditions of three types of
treatment effects; treatment effect of one specific choice versus the next best alter-
native, treatment effect of one specific group of choices versus the other group and
treatment effect of one specified choice versus another choice. They establish the
identification conditions for LATE and MTE with discrete instruments and continu-
ous instruments, respectively. For the the MTE of one specified choice versus another
choice, they impose a large support conditions. By reducing model (1) to the binary
treatment model in effect, they identify the MTE using identification results in the
binary model.

2.2 Section 5.2 of Lee and Salanié (2018)

Lee and Salanié (2018) employ the following model:

\[ D_k = d_k(V, Q(Z)) \]

where

\[ d_k(V, Q(Z)) = \sum_{l \in \mathcal{L}} c^k_l \prod_{j \in l} S_j(V, Q(Z)) = \sum_{l \in \mathcal{L}} c^k_l \prod_{m=1}^{|l|} S_{l_m}(V, Q(Z)), \]

and \( c^k_l \) is an integer. Let \( Q(Z) \) denote the vector of functions of the instruments
\( Q_i(Z) \), that is,

\[ Q(Z) := (Q_1(Z), \ldots, Q_J(Z))^\prime. \]

Let \( V = (V_1, \ldots, V_J)^\prime \) be a vector of continuous random variables whose support
is \([0, 1]^J\). For each \( j \in \{1, \ldots, J\} \), define \( S_j(V, Q(Z)) := 1(V_j < Q_j(Z)) \), which is an
element of the selection mechanism (2). Let \( \mathcal{J} \) be the set \( \{1, \ldots, J\} \), and let \( \mathcal{L} \) be
the set of all the subsets of \( \mathcal{J} \).

\( V \) is a vector of unobserved heterogeneity and \( Q_j(Z) \) serves as a threshold for
each \( S_j \) when \( Z \) is given. Hence, \( S_j \) is a separable threshold-crossing model as in the
generalized Roy model. Model (2) can express any decision model that comprises
sums, products, and differences of their indicator functions $S_j$.

In their Section 5.2, Lee and Salanić (2018) apply their main theorem to the multiple discrete choice model. Example 5.2 of Lee and Salanić (2018) analyzes three treatments, $K = \{0, 1, 2\}$. Using the definitions we set in Section 2.1, they define

\[
\tilde{R}_{0,1}(Z) := R_0(Z) - R_1(Z), \quad \tilde{R}_{0,2}(Z) := R_0(Z) - R_2(Z), \quad \tilde{R}_{1,2}(Z) := R_1(Z) - R_2(Z),
\]
\[
\tilde{U}_{0,1} := U_0 - U_1, \quad \tilde{U}_{0,2} := U_0 - U_2, \quad \tilde{U}_{1,2} := U_1 - U_2.
\]

(3)

Subsequently, they define

\[
Q_{0,1}(Z) := F_{\tilde{U}_{0,1}}(\tilde{R}_{0,1}(Z)), \quad Q_{0,2}(Z) := F_{\tilde{U}_{0,2}}(\tilde{R}_{0,2}(Z)), \quad Q_{1,2}(Z) := F_{\tilde{U}_{1,2}}(\tilde{R}_{1,2}(Z)),
\]
\[
V_{0,1} := F_{\tilde{U}_{0,1}}(\tilde{U}_{0,1}), \quad V_{0,2} := F_{\tilde{U}_{0,2}}(\tilde{U}_{0,2}), \quad V_{1,2} := F_{\tilde{U}_{1,2}}(\tilde{U}_{1,2}).
\]

(4)

Note that

\[
V_{0,1} < Q_{0,1}(Z) \iff R_1(Z) - U_1 < R_0(Z) - U_0,
\]
\[
V_{0,2} < Q_{0,2}(Z) \iff R_2(Z) - U_2 < R_0(Z) - U_0,
\]
\[
V_{1,2} < Q_{1,2}(Z) \iff R_2(Z) - U_2 < R_1(Z) - U_1.
\]

Hence, under this setting, treatments are determined as follows:

- $D = 0$ iff $V_{0,1} < Q_{0,1}(Z)$ and $V_{0,2} < Q_{0,2}(Z)$,
- $D = 1$ iff $V_{0,1} > Q_{0,1}(Z)$ and $V_{1,2} < Q_{1,2}(Z)$,
- $D = 2$ iff $V_{0,2} > Q_{0,2}(Z)$ and $V_{1,2} > Q_{1,2}(Z)$.

Evidently, this corresponds to the decision rule based on model (1).

We first collect relevant assumptions and definitions in Lee and Salanić (2018).

Assumption 2.1 (Lee and Salanić, 2018) Any of the following three equivalent statements holds:

(i) the treatment variable $D$ is measurable with respect to the $\sigma$-field generated by the events

\[
\{V_{0,1} < Q_{0,1}(Z)\}, \quad \{V_{0,2} < Q_{0,2}(Z)\}, \quad \{V_{1,2} < Q_{1,2}(Z)\};
\]
(ii) each event \( \{ D = k \} = \{ D_k = 1 \} \) is a member of this \( \sigma \)-field;

(iii) for each \( k \), there exists a function \( d_k \) that is measurable with respect to this \( \sigma \)-field, such that \( D_k = d_k(V, Q(Z)) \), where \( Q(Z) = (Q_{0,1}(Z), Q_{0,2}(Z), Q_{1,2}(Z))' \).

Moreover, every treatment value \( k \) has positive probability.

Assumption 2.2 (Lee and Salanié, 2018) \( Y_0, Y_1, Y_2 \) and \( (V_{0,1}, V_{0,2}, V_{1,2})' \) are jointly independent of \( Z \).

Assumption 3.2 (Lee and Salanié, 2018) The joint distribution of \( (V_{0,1}, V_{0,2}, V_{1,2})' \) is absolutely continuous with respect to the Lebesgue measure on \( \mathbb{R}^3 \) and its support is \([0,1]^3\).

Assumption 3.4 (Lee and Salanié, 2018) The point \( q \in (0,1)^3 \) belongs to the interior of the range of variation of the thresholds \( Q \).

\( q \) is a point where we can identify the MTE. \( Q \) is defined as follows:

**Definition 3.1** (Lee and Salanié, 2018) Let \( Z \) denote the support of \( Z \), and let \( Q = (Q_{0,1}(Z), Q_{0,2}(Z), Q_{1,2}(Z))' \) denote the range of variation of \( (Q_{0,1}(Z), Q_{0,2}(Z), Q_{1,2}(Z))' \).

Lee and Salanié (2018) argue that their main theorems (Theorem 3.1 and Theorem A.1) can identify the MTE if \( Q_{0,1}(Z), Q_{0,2}(Z), Q_{1,2}(Z) \) are identified and their Assumptions 2.1–2.2 and 3.2–3.4 hold. From this result, they state that we can identify MTE without monotonicity. Moreover, because they identify the MTE via multidimensional cross derivatives, they do not rely on the identification-at-infinity strategy.

The following discussion shows that the model in Section 5.2 of Lee and Salanié (2018) is not sufficient to identify the MTE. When we set \( V := (V_{0,1}, V_{0,2}, V_{1,2})' \), by construction we have

\[
F_{\tilde{U}_{0,1}}^{-1}(V_{0,1}) = U_0 - U_1,  \\
= (U_0 - U_2) - (U_1 - U_2),  \\
= F_{\tilde{U}_{0,2}}^{-1}(V_{0,2}) - F_{\tilde{U}_{1,2}}^{-1}(V_{1,2}).
\]

Hence, even if \( V \) is absolutely continuous with respect to the Lebesgue measure on \( \mathbb{R}^3 \), its support is not equal to \([0,1]^3\) as in Figure 1, which violates the Assumption 3.2 in Lee and Salanié (2018). Consequently, we cannot achieve the identification of the MTE with multi-valued treatments.
Figure 1: This figure shows the support of $V$ when $U_0 \sim N(0, 0.5)$, $U_1 \sim N(1, 1)$, $U_2 \sim N(-1, 1)$ and $(U_0, U_1, U_2)'$ are mutually independent. The right bottom axis, the left bottom axis and the vertical axis correspond to the value of $V_{0,1}, V_{0,2}$ and $V_{1,2}$, respectively.

3 Model and Identification

This section establishes the identification of MTE in model \[1\] in Section 2.1. We construct a model of three alternatives with two thresholds instead of three thresholds in \cite{Lee and Salanić (2018)}. In the discussion below, we regard treatment 1 as the baseline.

Define

\[
\tilde{R}_{1,0}(Z) := R_1(Z) - R_0(Z), \quad \tilde{R}_{1,2}(Z) := R_1(Z) - R_2(Z),
\]
\[
\tilde{U}_{1,0} := U_1 - U_0, \quad \tilde{U}_{1,2} := U_1 - U_2.
\]

Assume $\tilde{U}_{1,0}$ and $\tilde{U}_{1,2}$ are continuously distributed. Let

\[
Q_1(Z) := F_{\tilde{U}_{1,0}}(\tilde{R}_{1,0}(Z)), \quad Q_2(Z) := F_{\tilde{U}_{1,2}}(\tilde{R}_{1,2}(Z)),
\]
\[
V_1 := F_{\tilde{U}_{1,0}}(\tilde{U}_{1,0}), \quad V_2 := F_{\tilde{U}_{1,2}}(\tilde{U}_{1,2}).
\]
Set

\[ S_1 = \mathbb{1}\{V_1 < Q_1(Z)\}, \quad S_2 = \mathbb{1}\{V_2 < Q_2(Z)\}. \]

Note that

\[ V_1 < Q_1(Z) \iff F_{\tilde{U}_{1,0}}(U_1 - U_0) < F_{\tilde{U}_{1,0}}(R_1(Z) - R_0(Z)), \]
\[ \iff R_0(Z) - U_0 < R_1(Z) - U_1, \]

and a similar argument gives

\[ V_2 < Q_2(Z) \iff R_2(Z) - U_2 < R_1(Z) - U_1. \]

Therefore, \( D_1 = S_1S_2 \) by definition.

Define

\[ S_3 := \mathbb{1}\{V_1 < F_{\tilde{U}_{1,0}}^{-1}(Q_2(Z)) - F_{\tilde{U}_{1,1,2}}^{-1}(Q_1(Z))\}. \]

We show \( D_2 = (1 - S_2)S_3 \). By construction, we obtain

\[ V_1 < F_{\tilde{U}_{1,0}}^{-1}(V_2) - F_{\tilde{U}_{1,1,2}}^{-1}(Q_2(Z)) + F_{\tilde{U}_{1,0}}^{-1}(Q_1(Z)) \]
\[ \iff U_2 - U_0 < R_2(Z) - R_0(Z) \]
\[ \iff R_0(Z) - U_0 < R_2(Z) - U_2. \]

Hence, \( D_2 = (1 - S_2)S_3 \) by definition. A similar argument reveals \( D_0 = (1 - S_1)(1 - S_3) \).

In this setting, treatment 1 has the form of a double hurdle model, namely \( D_1 = 1 \) if and only if \( V_1 < Q_1(Z) \) and \( V_2 < Q_2(Z) \). Hence, we can apply Theorem 3.1 in Lee and Salanié (2018) to the identification of the conditional expectation, \( E[Y_1|V_1, V_2] \).

The result of Lee and Salanié (2018) for the double hurdle model is also useful for the identification of thresholds \( Q_1(Z) \) and \( Q_2(Z) \). For more details, see Section 4.

In a double hurdle model with three choices, we cannot define some treatments through two thresholds. For example, when treatment 1 has the form of a double hurdle model, the information in \( 1\{V_1 < Q_1(Z)\} \) and \( 1\{V_2 < Q_2(Z)\} \) is not sufficient
to determine whether the agent receives either treatment 0 or treatment 2. As a result, we cannot identify the MTE of one specific choice versus another. With the introduction of $S_3$, we successfully specify $D_0$ and $D_2$ in our model. Our model cannot be expressed in the form of model 2, however, because $S_3$ includes the nonlinear transformation of $(V_1, V_2)$ and $(Q_1(Z), Q_2(Z))$. Consequently, Theorem 3.1 in Lee and Salanié (2018) does not identify the corresponding conditional expectations of $Y_0$ or $Y_2$.

**Assumption 3.1.** \(\{V_1 < Q_1(Z)\}, \{V_2 < Q_2(Z)\}, \{V_1 < F_{\tilde{V},1,0}(F_{\tilde{U},1,2}^{-1}(V_2) - F_{\tilde{U},1,2}^{-1}(Q_2(Z)) + F_{\tilde{U},1,0}^{-1}(Q_1(Z)))\}\) are measurable sets.

**Assumption 3.2** (Conditional Independence of Instruments). $Y_0$, $Y_1$, $Y_2$ and $V = (V_1, V_2)'$ are jointly independent of $Z$.

**Assumption 3.3** (Continuously Distributed Unobserved Heterogeneity in the Selection Mechanism). The joint distribution of $V$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^2$, and its support is $[0, 1]^2$.

**Assumption 3.4** (The Existence of the Moments). For $k = 0, 1, 2$, $E[|G(Y_k)|] < \infty$, where $G$ is a measurable function defined on the support $\mathcal{Y}$ of $Y$, which can be discrete, continuous, or multidimensional.

Assumption 3.1 ensures the existence of probability of each treatment, that is, $\Pr(D_k = 1)$. Assumption 3.2 corresponds to the exogeneity of instruments. We guarantee the existence of the probability density function of $V$ by Assumption 3.3. Assumption 3.4 ensures the existence of the moments for each alternative. Assumptions 3.1, 3.2, and 3.3 correspond to Assumptions 2.1, 2.2, and 3.2 of Lee and Salanié (2018), respectively. Assumption 3.4 generalizes Assumption (A-3) of Heckman et al. (2008).

The main idea of our identification strategy is similar to that in Heckman and Vytlacil (2007b) and Heckman et al. (2008). They assume $(U_0, U_1, U_2)$ are continuously distributed whose support is equal to $\mathbb{R}^3$; that $(Y_0, Y_1, Y_2)$ and $V$ are jointly

\[1\]

A distinction between our identification strategy and that of Heckman et al. (2008) is the order of differentiation and reduction to a binary case. At first, we reduce to a binary case by extending two conditional expectations $E[G(Y)D_1|Q_1(Z) = q_1, Q_2(Z) = q_2]$, and $E[G(Y)D_2|Q_1(Z) = q_1, Q_2(Z) = q_2]$. Thereafter, we differentiate those extended conditional expectations with respect to $Q_2(Z)$. However, Heckman and Vytlacil (2007b) and Heckman et al. (2008) reduce to a binary case after taking differentiation.
independent of $Z$; and that, for $k = 0, 1, 2$, $E[|Y_k|] < \infty$. To identify MTE of one specified choice versus another specified choice, they impose a large support assumption in Theorem 8 (Heckman and Vytlacil, 2007b) and Theorem 3 (Heckman et al., 2008). This large support assumption implies that one utility function, $R_k(Z)$, can take a sufficiently negative value, which effectively reduces the model with three alternatives to a binary case. In the following subsection, we impose assumptions that play the same role as the large support assumption.

In the binary treatment case, Heckman and Vytlacil (2005) define $D^* = 1$ as the receipt of the treatment and characterize the decision rule as the generalized Roy model, that is,

$$D^* = \mathbb{1}\{\mu_D(Z) - U_D \geq 0\},$$

where $\mu_D(Z)$ is an unknown function, which maps from $\mathbb{R}^{\dim(Z)}$ to $\mathbb{R}$, and $U_D$ is an unobserved continuous random variable. As a normalization, they innocuously assume that $U_D \sim U[0, 1]$. Thereafter, Heckman and Vytlacil (2005) define MTE with binary treatment as

$$\Delta^{MTE}(u_D) \equiv E[Y_1 - Y_0 | U_D = u_D],$$

where $u_D \in (0, 1)$.

A major distinction between Heckman et al. (2008) and our study is the object of the identification. By extending the LIV approach, Heckman and Vytlacil (2007b) and Heckman et al. (2008) study the identification of MTE, that is, for any $\ell \in \mathbb{R}$ and $j, k \in K$, such that $j \neq k$,

$$E[Y_k - Y_j | R_k(Z) - R_j(Z) = \ell, R_k(Z) - U_k = R_j(Z) - U_j].$$

We study the identification of the following conditional expectations

$$E[G(Y_k) | V = q^*], \ E[G(Y_j) | V = q^*],$$

where $V = F_{U_k - U_j}(U_k - U_j)$. When we set $G(Y) = Y$ and take the difference between

---

2 In the binary treatment case, by using the identification results of the conditional expectations, Carneiro and Lee (2009) propose a semiparametric estimator of the MTE. Brinch et al. (2017) also employ the results to identify MTE with discrete instruments.
two conditional expectations, we identify the following MTE:

\[ E[Y_k - Y_j | V = q^*]. \] (7)

We identify MTE (7) conditional on \( V \). By the definition of \( V \), \( V \) uniformly ranges from 0 to 1, and \( q^* \) is the quantile of the distribution of \( U_k - U_j \), which corresponds to the definition of MTE (5) with binary treatment in Heckman and Vytlacil (2005). However, Heckman et al. (2008) and Heckman and Vytlacil (2007b) identify MTE (6) conditional on \( U_k - U_j \), which implies the shape of the distribution function of \( U_k - U_j \) affects the interpretation of MTE and the distribution function of \( U_k - U_j \) is unknown in many cases. Therefore, the MTE (7) is a natural extension of MTE with binary treatment to the multivalued treatment case.

Mountjoy (2022) introduces a new MTE with three choices. He identifies the MTE of treatment 1 versus treatment 2 conditional on a certain level of the utility of treatment 0. Because Mountjoy (2022) fixes the utility of choice 0, he does not require the large support assumption. We assume the large support condition to obtain MTE corresponding to the one with binary treatment. We view our identification result and Mountjoy’s identification result as complementing each other.

3.1 Identification of MTE of Treatment 1 versus Treatment 2

As in Heckman and Vytlacil (2007b) and Heckman et al. (2008), we impose the large support assumption for the identification of MTE of treatment 1 versus treatment 2.

Definition 3.1. Let \( Z \) denote the support of \( Z \) and \( \tilde{Q} = Q(Z) \) denote the range of variation of \( Q(Z) = (Q_1(Z), Q_2(Z))' \).

Assumption 3.5. Let \( q_2^* \) be in \((0, 1)\).

(i) (Continuity at \( q_2^* \)): \( E[G(Y_1)|V_2 = v_2] \) and \( E[G(Y_2)|V_2 = v_2] \) are continuous at \( q_2^* \).

(ii) (Open Range at \((1, q_2^*)')\): There exists \( \delta > 0 \) such that \( \tilde{N}_2((1, q_2^*'), \delta) \) is in \( \tilde{Q} \).

Assumption 3.5 ensures partial differentiability, which is important for the MTE identification. Assumption 3.5 (ii) ensures the existence of the case wherein the
agents re treatment 1 or treatment 2, which plays the same role as the large support assumption in [Heckman and Vytlacil (2007b)] and [Heckman et al. (2008)]. Assumptions 3.1 to 3.5 impose no explicit restriction on \( Z \), but Assumption 3.5 (ii) implicitly requires that at least two variables in \( Z \) must be continuous.

Assumption 3.5 (ii) implies that two conditional expectations, \( E[G(Y)D_1|Q_1(Z) = q_1, Q_2(Z) = q_2] \) and \( E[G(Y)D_2|Q_1(Z) = q_1, Q_2(Z) = q_2] \) are defined on \( \tilde{N}_2((1, q_2^*)', \delta) \). Because \((1, q_2^*)' \) is not in \( \tilde{Q} \), the range of variation of \( Q(Z) \), these two conditional expectations are not defined at \((q_1, q_2) = (1, q_2^*)' \). We identify MTE of treatment 1 and treatment 2 by first extending these two conditional expectations to \( \tilde{N}_2((1, q_2^*)', \delta) \), such that they are defined at \((1, q_2^*)' \) and, thereafter, by partially differentiating the extended functions with respect to \( Q_2(Z) \) at \((1, q_2^*)' \). To this end, we first introduce the definition of Cauchy continuity.

**Definition 3.2 (Cauchy continuity).** A function \( f : A \to \mathbb{R} \) is Cauchy continuous if \( f \) preserves Cauchy sequences, that is, if \( \{x_n\}_{n=1}^\infty \) is a Cauchy sequence in \( A \), \( \{f(x_n)\}_{n=1}^\infty \) is a Cauchy sequence in \( \mathbb{R} \).

**Lemma 3.1 (Theorem 7, Snipes (1977)).** Assume \( A \) is dense in \( X \), where \( X \subset \mathbb{R}^d \). If a function \( f : A \to \mathbb{R} \) is Cauchy continuous, a unique continuous function \( \tilde{f} : X \to \mathbb{R} \) that extends \( f \) exists.

The following lemma extends \( E[G(Y)D_1|Q_1(Z) = q_1, Q_2(Z) = q_2] \) and \( E[G(Y)D_2|Q_1(Z) = q_1, Q_2(Z) = q_2] \) defined on \( \tilde{N}_2((1, q_2^*)', \delta) \) to \( \tilde{N}_2((1, q_2^*)', \delta) \), which includes \((1, q_2^*)' \), by applying Lemma 3.1 to them.

**Lemma 3.2.** Under Assumptions 3.1 to 3.5 there exist continuous functions \( \tilde{E}[G(Y)D_1|Q_1(Z) = q_1, Q_2(Z) = q_2] \) and \( \tilde{E}[G(Y)D_2|Q_1(Z) = q_1, Q_2(Z) = q_2] \) on \( \tilde{N}_2((1, q_2^*)', \delta) \), which uniquely extend \( E[G(Y)D_1|Q_1(Z) = q_1, Q_2(Z) = q_2] \) and \( E[G(Y)D_2|Q_1(Z) = q_1, Q_2(Z) = q_2] \), respectively.

Conditional on the assumption that \( Q_1(Z) \) and \( Q_2(Z) \) are identified, we identify the conditional expectations \( E[G(Y_1)|V_2 = q_2^*] \) and \( E[G(Y_2)|V_2 = q_2^*] \) by partially differentiating the extended conditional expectations \( \tilde{E}[G(Y)D_1|Q_1(Z) = q_1, Q_2(Z) = q_2] \) and \( \tilde{E}[G(Y)D_2|Q_1(Z) = q_1, Q_2(Z) = q_2] \) with respect to \( Q_2(Z) \) at \((1, q_2^*)' \) as in [Lee and Salanié (2018)]. In Section 4, we give a sufficient condition of the identification.

---

3The Cauchy continuity slightly differs from continuity and uniform continuity. See Appendix B for further details.
of thresholds $Q(Z)$ that Heckman and Vytlacil (2007b) and Heckman et al. (2008) also require.

**Theorem 3.3.** Assume Assumptions 3.1 to 3.5 hold. Then, the conditional expectations of $G(Y_1), G(Y_2)$ are given by

$$E[G(Y_1)|V = q_2^*] = \frac{\partial E[G(Y)D_1|Q_1(Z), Q_2(Z)]}{\partial Q_2(Z)} \bigg|_{(Q_1(Z), Q_2(Z))=(1, q_2^*)},$$

and

$$E[G(Y_2)|V = q_2^*] = -\frac{\partial E[G(Y)D_2|Q_1(Z), Q_2(Z)]}{\partial Q_2(Z)} \bigg|_{(Q_1(Z), Q_2(Z))=(1, q_2^*)}.$$ 

The identification result of the conditional expectations enables us to identify measures of treatment effects. For example, if we set $G(Y) = Y$ as we did previously, we obtain

$$E[Y_1 - Y_2|V = q^*].$$

If we let $G(Y) = 1(Y \leq y)$ for $y \in \mathbb{R}$, we can identify

$$F_{Y_1|V=q^*}(y), \quad F_{Y_2|V=q^*}(y).$$

If $F_{Y_1|V}$ and $F_{Y_2|V}$ are invertible, we identify the quantile treatment effect by taking the difference between the two, that is,

$$Q_{Y_1|V=q^*}(\tau) - Q_{Y_2|V=q^*}(\tau),$$

where $\tau \in (0, 1)$.

### 3.2 Identification of MTE of Treatment 0 versus Treatment 1

In this subsection, similar to Section 3.1, we state sufficient conditions for the identification of MTE of treatment 0 versus treatment 1. We impose assumptions similar to Assumptions 3.5 in Section 3.1.

**Assumption 3.6.** Let $q_1^*$ be in $(0, 1)$. 


(i) (Continuity at $q_1^*$). $E[G(Y_0)|V_1 = v_1]$ and $E[G(Y_1)|V_1 = v_1]$ are continuous at $q_1^*$.

(ii) (Open Range at $(q_1^*, 1)'$) There exists $\delta > 0$ such that $\tilde{N}_2((q_1^*, 1)', \delta)$ is in $\tilde{Q}$.

Similar to Section 3.1, we identify MTE of treatment 0 versus treatment 1 by first extending two conditional expectations $E[G(Y)D_0|Q_1(Z) = q_1, Q_2(Z) = q_2]$ and $E[G(Y)D_1|Q_1(Z) = q_1, Q_2(Z) = q_2]$, defined on $\tilde{N}_2((q_1^*, 1)', \delta)$, such that they are defined at $(q_1^*, 1)'$, and, thereafter, by partially differentiating the extended functions with respect to $Q_1(Z)$ at $(q_1^*, 1)'$. By applying Lemma 3.4 to them, we extend $E[G(Y)D_0|Q_1(Z) = q_1, Q_2(Z) = q_2]$ and $E[G(Y)D_1|Q_1(Z) = q_1, Q_2(Z) = q_2]$ to $\tilde{N}_2((q_1^*, 1)', \delta)$, which includes $(q_1^*, 1)'$.

**Lemma 3.4.** Under Assumptions 3.1 to 3.4 and 3.6 there exist continuous functions $E[G(Y)D_0|Q_1(Z) = q_1, Q_2(Z) = q_2]$ and $E[G(Y)D_1|Q_1(Z) = q_1, Q_2(Z) = q_2]$ on $\tilde{N}_2((q_1^*, 1)', \delta)$, which uniquely extend $E[G(Y)D_0|Q_1(Z) = q_1, Q_2(Z) = q_2]$ and $E[G(Y)D_1|Q_1(Z) = q_1, Q_2(Z) = q_2]$, respectively.

A similar argument to Theorem 3.3 provides the identification results for the conditional expectations $E[G(Y_1)|V_1 = q_1^*]$ and $E[G(Y_0)|V_1 = q_1^*]$.

**Theorem 3.5.** Assume Assumptions 3.1 to 3.4 and 3.6 hold. Then, the conditional expectations of $G(Y_0), G(Y_1)$ are given by

$$E[G(Y_1)|V_1 = q_1^*] = \left. \frac{\partial E[G(Y)D_1|Q_1(Z), Q_2(Z)]}{\partial Q_1(Z)} \right|_{(Q_1(Z), Q_2(Z)) = (q_1^*, 1)}$$

and

$$E[G(Y_0)|V_1 = q_1^*] = - \left. \frac{\partial E[G(Y)D_2|Q_1(Z), Q_2(Z)]}{\partial Q_1(Z)} \right|_{(Q_1(Z), Q_2(Z)) = (q_1^*, 1)}.$$

**4 Identification of thresholds $Q(Z)$**

In this section, we provide a sufficient condition for the nonparametric identification of $Q_1(Z)$ and $Q_2(Z)$. Heckman et al. (2008) identify the MTE given known utility functions; they refer to Heckman and Vytlacil (2007a) for the identification of utility...
functions. Lee and Salanié (2018) establish the sufficient conditions for the identification of the MTE conditional on the assumption that $Q(Z)$ is already identified. They rely on Matzkin’s (1993) result, which establishes sufficient conditions for the nonparametric identification of the utility functions and the cumulative joint density of unobserved random terms. However, because Matzkin (1993) does not intend to identify thresholds, her result is not sufficient for the nonparametric identification of thresholds. We establish a sufficient condition for the identification of thresholds, thereby enabling the estimation of MTE with the results in Theorems 3.3 and 3.5.

A sufficient condition requires the existence of at least one instrument that significantly and negatively affects only one utility. Let $Z[\ell]$ denote the $\ell$-th component of $Z$. Let $Z[\ell]^\ell$ be all the instruments except for the $\ell$-th component.

**Assumption 4.1.** For $k = 0, 2$, there exists at least one element of $Z$, say $Z[\ell_k]$, and at least one value $a[\ell_k]$ such that, given any $z[\ell_k]^\ell \in Z[\ell_k]^\ell$,

$$\lim_{z[\ell_k] \to a[\ell_k]} R_k(z[\ell_k], z[\ell_k]^\ell) = -\infty$$

and $R_j(Z)$ is constant for $j \neq k$.

Assumption 4.1 imposes a type of exclusion restriction. Conditional on all the regressors except $Z[\ell_k]$, one can vary $R_k(Z)$ independently. Moreover, we assume the existence of one value $a[\ell_k]$ such that, as $z[\ell_k] \to a[\ell_k]$, the value of the utility function $R_k(Z)$ becomes sufficiently small given any $z[\ell_k]^\ell$.

**Theorem 4.1.** Assume Assumption 4.1 holds. Then, $Q_1(Z), Q_2(Z)$ are identified as

$$\lim_{z[\ell_2] \to a[\ell_2]} H(Z) = Q_1(Z)$$

$$\lim_{z[\ell_0] \to a[\ell_0]} H(Z) = Q_2(Z),$$

where

$$H(Z) := \Pr(D_1 = 1|Z) = F_V(Q_1(Z), Q_2(Z)).$$

Theorem 4.1 corresponds to Theorem 4.2 in Lee and Salanié (2018). Because treatment 1 has the form of a double hurdle model, Theorem 4.1 provides the identification result of thresholds in a double hurdle model, as in Theorem 4.2 in Lee

---

4Assumption 4.1 implies Assumption 3.5 (ii) and Assumption 3.6 (ii) because $Q_1(Z), Q_2(Z)$ approach one as $R_0(Z), R_2(Z)$ approach minus infinity, respectively.
and Slanić (2018). While Lee and Slanić (2018) restrict the range of instrumental variables and the functional form of \((Q_1(Z), Q_2(Z))\)', Assumption 4.1 requires the utility of treatment 0 or treatment 2 can be arbitrarily small. Heckman and Vytlacil (2007b) and Heckman et al. (2008) also require Assumption 4.1 to identify the MTE of treatment 1 versus treatment 2 and the MTE of treatment 1 versus treatment 0.

5 Generalization

This section generalizes the framework in Section 3 to identify the MTE for the discrete choice model with more than two treatments.

5.1 Model and Assumptions

In this subsection, we construct a model regarding treatment \(k\) as the baseline and consider the identification of the MTE of treatment \(k\) versus treatment \(j\) for any \(j \neq k\), where \(j, k \in \mathcal{K}\) and \(|\mathcal{K}| \geq 3\).

Define, for each \(i \neq k\) in \(\mathcal{K}\),

\[
\tilde{R}_{k,i}(Z) := R_k(Z) - R_i(Z), \quad \tilde{U}_{k,i} := U_k - U_i.
\]

Assume \(\tilde{U}_{k,i}\) s are continuously distributed. Let

\[
Q_i(Z) := F_{\tilde{U}_{k,i}}(\tilde{R}_{k,i}(Z)), \quad V_i := F_{\tilde{U}_{k,i}}(\tilde{U}_{k,i}), \quad S_i = 1\{V_i < Q_i(Z)\}.
\]

By construction, \(Q_i(Z), V_i\) and \(S_i\) are defined for each \(i\) in \(\mathcal{K}\) except \(k\). Note that

\[
V_i < Q_i(Z) \iff F_{\tilde{U}_{k,i}}(U_k - U_i) < F_{\tilde{U}_{k,i}}(R_k(Z) - R_i(Z)),
\]

\[
\iff R_i(Z) - U_i < R_k(Z) - U_k,
\]

Therefore, \(D_k = \prod_{i \in \mathcal{K} \setminus \{k\}} S_i\) by definition.

Define, for each \(i \neq j, k\) in \(\mathcal{K}\),

\[
S_i^* := 1\{V_i < F_{\tilde{U}_{k,i}}^{-1}(F_{\tilde{U}_{k,j}}^{-1}(V_j) - F_{\tilde{U}_{k,j}}^{-1}(Q_j(Z)) + F_{\tilde{U}_{j,k}}^{-1}(Q_j(Z))))\}.
\]
By construction, we obtain

\[ V_i < F_{\tilde{U}_{k,i}}^{-1}(F_{\tilde{U}_{k,j}}^{-1}(V_j) - F_{\tilde{U}_{k,j}}^{-1}(Q_j(Z))) + F_{\tilde{U}_{k,i}}^{-1}(Q_i(Z))) \]

\[ \Leftrightarrow F_{\tilde{U}_{k,i}}^{-1}(V_i) - F_{\tilde{U}_{k,j}}^{-1}(V_j) < F_{\tilde{U}_{k,i}}^{-1}(Q_i(Z)) - F_{\tilde{U}_{k,j}}^{-1}(Q_j(Z)) \]

\[ \Leftrightarrow U_j - U_i < R_j(Z) - R_i(Z) \]

\[ \Leftrightarrow R_i(Z) - U_i < R_j(Z) - U_j. \]

Hence, \( D_j = \prod_{i\in\mathcal{K}\setminus\{k,j\}} S_i^*(1 - S_j) \) by definition.

We use the same strategy to identify MTE of treatment \( j \) versus treatment \( k \) \( Y_j \) and \( Y_k \) as in Section 3. Assumptions 5.1 to 5.5 in the following correspond to Assumptions 3.1 to 3.5, respectively.

**Assumption 5.1.** For each \( i \neq k \) and \( \ell \neq k,j \) in \( \mathcal{K} \), \{\( V_i < Q_i(Z) \)\} and \{\( V_\ell < F_{\tilde{U}_{k,i}}^{-1}(F_{\tilde{U}_{k,j}}^{-1}(V_j) - F_{\tilde{U}_{k,j}}^{-1}(Q_j(Z))) + F_{\tilde{U}_{k,\ell}}^{-1}(Q_\ell(Z))) \)\} are measurable sets.

**Assumption 5.2** (Conditional Independence of Instruments). \( Y_j, Y_k \) and \( V = (V_0, \ldots, V_{k-1}, V_{k+1} \cdots, V_K) \) are jointly independent of \( Z \).

**Assumption 5.3** (Continuously Distributed Unobserved Heterogeneity in the Selection Mechanism). The joint distribution of \( V \) is absolutely continuous with respect to the Lebesgue measure on \( \mathbb{R}^{K-1} \), and its support is \([0,1]^{K-1}\).

**Assumption 5.4** (The Existence of the Moments). \( E[|G(Y_j)|] < \infty \) and \( E[|G(Y_k)|] < \infty \), where \( G \) is a measurable function defined on the support \( \mathcal{Y} \) of \( Y \), which can be discrete, continuous, or multidimensional.

**Definition 5.1.** Let \( \mathcal{Z} \) denote the support of \( Z \) and \( \tilde{Q} = Q(Z) \) denote the range of variation of \( Q(Z) = (Q_0(Z), \ldots, Q_{k-1}(Z), Q_{k+1}(Z) \cdots, Q_{K-1}(Z))' \).

**Assumption 5.5.** Let \( q_j^* \) be in (0,1).

(i) (Continuity at \( q_j^* \)). \( E[G(Y_k)|V_j = v_j] \) and \( E[G(Y_j)|V_j = v_j] \) are continuous at \( q_j^* \).

(ii) (Open Range at \( q_j^* \)) There exists \( \delta > 0 \) such that \( \tilde{N}_{k-1}(q_j^*, \delta) \) is in \( \tilde{Q} \), where \( q_j^* = (1, \cdots, q_j^*, \cdots, 1)' \).
To reduce the model with multivalued treatments to a binary treatment model, we need to impose Assumption 5.5 which enables us to define two conditional expectations, \( E[G(Y)D_k|Q(Z) = q] \), and \( E[G(Y)D_j|Q(Z) = q] \) at \( q_j^* \) by extension, and to partially differentiate two extended conditional expectations with respect to \( Q_j(Z) \) at \( q_j^* \).

**Lemma 5.1.** Under Assumptions 5.1 to 5.5, there exist continuous functions \( E[G(Y)D_k|Q(Z) = q] \) and \( E[G(Y)D_j|Q(Z) = q] \) on \( \tilde{N}_{K-1}(q_j^*, \delta) \), which extend \( E[G(Y)D_k|Q(Z) = q] \) and \( E[G(Y)D_j|Q(Z) = q] \) defined on \( \tilde{N}_{K-1}(q_j^*, \delta) \), respectively. This extension is unique.

### 5.2 Identification Result of MTE

Using \( E[G(Y)D_k|Q(Z) = q] \), \( E[G(Y)D_j|Q(Z) = q] \) from Lemma 5.1, we identify the MTE at \( V_j = q_j^* \) in a similar way to Theorems 3.3 and 3.5.

**Theorem 5.2.** Let Assumptions 5.1 to 5.5 hold. Then, the conditional expectations of \( G(Y_k), G(Y_j) \) are given by

\[
E[G(Y_k)|V_j = q_j^*] = \frac{\partial E[G(Y)D_k|Q(Z)]}{\partial Q_j(Z)} \bigg|_{Q(Z) = q_j^*},
\]

and

\[
E[G(Y_j)|V_j = q_j^*] = -\frac{\partial E[G(Y)D_j|Q(Z)]}{\partial Q_j(Z)} \bigg|_{Q(Z) = q_j^*}.
\]

### 6 Conclusion

We study the identification of MTE with multivalued treatments. We note that the model in Section 5.2 of [Lee and Salanié (2018)] is incomplete for identifying MTE and propose an improved model. By reducing the multiple discrete choice model to the case with binary treatment, we achieve the identification of MTE that is a natural extension of MTE with binary treatment defined in [Heckman and Vytlacil (2005)]

Further, we establish a sufficient condition for the identification of thresholds.
References

Angrist, J. D. and Imbens, G. W. (1995), “Two-stage least squares estimation of average causal effects in models with variable treatment intensity,” Journal of the American statistical Association, 90, 431–442.

Brinch, C., Mogstad, M., and Wiswall, M. (2017), “Beyond LATE with a discrete instrument,” Journal of Political Economy, 125, 985–1039.

Carneiro, P. and Lee, S. (2009), “Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality,” Journal of Econometrics, 149, 191–208.

Dahl, G. B. (2002), “Mobility and the return to education: Testing a Roy model with multiple markets,” Econometrica, 70, 2367–2420.

Fusejima, K. (2022), “Identification of multi-valued treatment effects with unobserved heterogeneity,” arXiv preprint, arXiv:2009.00553.

Heckman, J. J. and Pinto, R. (2018), “Unordered Monotonicity,” Econometrica, 86, 1–35.

Heckman, J. J., Urzua, S., and Vytlacil, E. (2006), “Understanding Instrumental Variables in Models with Essential Heterogeneity,” The Review of Economics and Statistics, 88, 389–432.

— (2008), “Instrumental Variables in Models with Multiple Outcomes: the General Unordered Case,” Annales d’Économie et de Statistique, 151–174.

Heckman, J. J. and Vytlacil, E. (2005), “Structural Equations, Treatment Effects, and Econometric Policy Evaluation1,” Econometrica, 73, 669–738.

Heckman, J. J. and Vytlacil, E. J. (1999), “Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects,” Proceedings of the National Academy of Sciences of the United States of America, 96, 4730–4734.

— (2007a), “Chapter 70 Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation,” Elsevier, vol. 6 of Handbook of Econometrics, pp. 4779–4874.

— (2007b), “Chapter 71 Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments,” Elsevier, vol. 6 of Handbook of Econometrics, pp. 4875–5143.

Imbens, G. W. and Angrist, J. D. (1994), “Identification and Estimation of Local Average Treatment Effects,” Econometrica, 62, 467–475.
Kirkeboen, L. J., Leuven, E., and Mogstad, M. (2016), “Field of study, earnings, and self-selection,” *The Quarterly Journal of Economics*, 131, 1057–1111.

Kline, P. and Walters, C. R. (2016), “Evaluating public programs with close substitutes: The case of Head Start,” *The Quarterly Journal of Economics*, 131, 1795–1848.

Lee, S. and Salanié, B. (2018), “Identifying Effects of Multivalued Treatments,” *Econometrica*, 86, 1939–1963.

Matzkin, R. (1993), “Nonparametric identification and estimation of polychotomous choice models,” *Journal of Econometrics*, 58, 137–168.

McFadden, D. (1974), “Conditional logit analysis of qualitative choice behavior,” *Frontiers in Econometrics*, 105–142.

Mountjoy, J. (2022), “Community colleges and upward mobility,” *American Economic Review*, 112, 2580–2630.

Snipes, R. (1977), “Functions that preserve Cauchy sequences,” *Nieuw archief voor Wiskunde*, 25, 409–422.

Vytlacil, E. (2002), “Independence, Monotonicity, and Latent Index Models: An Equivalence Result,” *Econometrica*, 70, 331–341.
Appendix

A Proofs and Auxiliary Results

A.1 Proof of Lemma 3.2

In view of Lemma 3.1, we can prove the state result by showing that $E[G(Y)D_1|Q_1(Z) = q_1, Q_2(Z) = q_2]$ and $E[G(Y)D_2|Q_1(Z) = q_1, Q_2(Z) = q_2]$ are Cauchy continuous on $\mathcal{N}_2((1, q_2^*), \delta)$. We first show $E[G(Y)D_1|Q_1(Z) = q_1, Q_2(Z) = q_2]$ is Cauchy continuous on $\mathcal{N}_2((1, q_2^*), \delta)$. Under the assumptions of the theorem, for any $q$ in the range of $Q(Z)$, we obtain

$$E[G(Y)D_1|Q_1(Z) = q_1, Q_2(Z) = q_2] = \int E[G(Y_1)|D_1 = 1, Q_1(Z) = q_1, Q_2(Z) = q_2] \Pr[D_1 = 1|Q_1(Z) = q_1, Q_2(Z) = q_2]$$

where $Q(Z) = (Q_1(Z), Q_2(Z))^*$, $q = (q_1, q_2)^*$, $V = (V_1, V_2)^*$, $d_1(V, Q(Z)) = \mathbb{1}\{V_1 < Q_1(Z)\} \times \mathbb{1}\{V_2 < Q_2(Z)\}$. The third equality holds by Assumption 3.2. Hence, for any $(q_1, q_2)^* \in \mathcal{N}_2((1, q_2^*), \delta)$, we have

$$E[G(Y)D_1|Q(Z) = q] = \int E[G(Y_1)|V = v]g(v_1, v_2; q_1, q_2) f_v(v) dv_1 \times v_2. \quad (A.2)$$

where $v = (v_1, v_2)^*$ and $g(v_1, v_2; q_1, q_2) := \mathbb{1}\{v_1 < q_1\} \mathbb{1}\{v_2 < q_2\}$.

Assume $\{a_n\}_{n=1}^{\infty} = \{(a_{1n}, a_{2n})\}_{n=1}^{\infty}$ is a Cauchy sequence in $\mathcal{N}_2((1, q_2^*), \delta)$. Take any $a_m$ and $a_\ell$ in $\{a_n\}_{n=1}^{\infty}$. Then, we obtain

$$|E[G(Y)D_1|Q(Z) = a_m] - E[G(Y)D_1|Q(Z) = a_\ell]|$$

$$= \left| \int E[G(Y_1)|V = v](g(v_1, v_2; a_{1m}, a_{2m}) - g(v_1, v_2; a_{1\ell}, a_{2\ell})) f_v(v) dv_1 \times v_2 \right|.$$ 

Because $E[|G(Y_1)|]$ is finite by Assumption 3.4, we have $E[|G(Y_1)||V = v]f_v(v) < \infty$ a.s. Further, $g(v_1, v_2; a_{1m}, a_{2m}) - g(v_1, v_2; a_{1\ell}, a_{2\ell}) \to 0$ a.s. as $m, \ell \to \infty$. Therefore, as $m, \ell \to \infty$,

$$E[G(Y_1)|V = v](g(v_1, v_2; a_{1m}, a_{2m}) - g(v_1, v_2; a_{1\ell}, a_{2\ell})) f_v(v) \to 0, \quad a.s.$$
Hence, it follows from the dominated convergence theorem (DCT) that as \( m, \ell \to \infty \),

\[
|E[G(Y)D_1|Q(Z) = a_m] - E[G(Y)D_1|Q(Z) = a_{\ell}]| \to 0. \tag{A.3}
\]

This implies \( E[G(Y)D_1|Q(Z) = q] \) is Cauchy continuous on \( \tilde{N}_2((1, q_2^*), \delta) \).

We complete the proof by showing that \( E[G(Y)D_2|Q_1(Z) = q_1, Q_2(Z) = q_2] \) is Cauchy continuous on \( \tilde{N}_2((1, q_2^*), \delta) \). From a similar argument to (A.1), under the assumptions of the theorem, we obtain

\[
E[G(Y)D_2|Q_1(Z) = q_1, Q_2(Z) = q_2] = E[E[G(Y_2)|V = v]h(v_1, v_2; q_1, q_2) f_V(v)d(v_1 \times v_2), \tag{A.4}
\]

where \( v = (v_1, v_2)' \) and \( h(v_1, v_2; q_1, q_2) := 1 \{ v_1 < F_{\tilde{U}_1,0}^{-1}(F_{\tilde{U}_2,2}^{-1}(Q_2(Z)) + F_{\tilde{U}_1,2}^{-1}(Q_1(Z))) \} \times 1 \{ v_2 \geq Q_2(Z) \} \). Hence, for any \( (q_1, q_2) \in \tilde{N}_2((1, q_2^*), \delta) \), we have

\[
E[G(Y)D_2|Q(Z) = q] = \int E[G(Y_2)|V = v]h(v_1, v_2; q_1, q_2) f_V(v)d(v_1 \times v_2).
\]

Assume \( \{b_n\}_{n=1}^{\infty} = \{(b_{1n}, b_{2n})\}_{n=1}^{\infty} \) is a Cauchy sequence in \( \tilde{N}_2((1, q_2^*), \delta) \). We take any \( b_m \) and \( b_{\ell} \) in \( \{b_n\}_{n=1}^{\infty} \). Then, we obtain

\[
|E[G(Y)D_2|Q(Z) = b_m] - E[G(Y)D_2|Q(Z) = b_{\ell}]| = \left| \int E[G(Y_2)|V = v](h(v_1, v_2; b_{1m}, b_{2m}) - h(v_1, v_2; b_{1\ell}, b_{2\ell})) f_V(v)(dv_1 \times v_2) \right|.
\]

Because \( E[|G(Y_2)|] \) is finite by Assumption 3.4, we have \( E[|G(Y_2)||V = v]f_V(v) < \infty \) a.s. Moreover, \( h(v_1, v_2; a_{1m}, a_{2m}) - h(v_1, v_2; a_{1\ell}, a_{2\ell}) \to 0 \) a.s. as \( m, \ell \to \infty \). Therefore, a similar argument to (A.3) gives

\[
|E[G(Y)D_2|Q(Z) = b_m] - E[G(Y)D_2|Q(Z) = b_{\ell}]| \to 0.
\]

This implies \( E[G(Y)D_2|Q(Z) = q] \) is Cauchy continuous on \( \tilde{N}_2((1, q_2^*), \delta) \).

\[\square\]

\section*{A.2 Proof of Theorem 3.3}

We first show that

\[
E[G(Y_1)|V = q_2'] = \frac{\partial E[G(Y)D_1|Q_1(Z), Q_2(Z)]}{\partial Q_2(Z)} \bigg|_{(Q_1(Z), Q_2(Z)) = (1, q_2^*)}. \tag{A.5}
\]
Because \((1, q_2) \in \mathcal{N}_2((1, q_2^*)', \delta)\) for any \(q_2 \in (q_2^* - \delta, q_2^* + \delta)\) by Lemma \[A.1\] and \(E[G(Y)D_1|Q_1(Z) = q_1, Q_2(Z) = q_2]\) is continuous on \(\mathcal{N}_2((1, q_2^*)', \delta)\), it follows from \(\[A.2\]\) that

\[
E[G(Y)D_1|Q_1(Z) = 1, Q_2(Z) = q_2] = \lim_{q_1 \uparrow 1} \int E[G(Y_1)|V = v]g(v_1, v_2; q_1, q_2)f_{V}(v)\,dv_1\,dv_2,
\]

where \(g(v_1, v_2; q_1, q_2) = \mathbb{1}\{v_1 < q_1\}\mathbb{1}\{v_2 < q_2\}\). Because \(E[|G(Y_1)|]\) is finite by Assumption \[3.4\] we have \(E[|G(Y_1)||V = v]f_{V}(v) < \infty\) a.s. Hence, as \(q_1 \uparrow 1\),

\[
E[G(Y_1)|V = v]g(v_1, v_2; q_1, q_2)f_{V}(v) \rightarrow E[G(Y_1)|V = v]\mathbb{1}\{v_2 < q_2\}f_{V}(v), \quad a.s.
\]

Therefore, it follows from the DCT and Fubini’s theorem that:

\[
E[G(Y)D_1|Q_1(Z) = 1, Q_2(Z) = q_2] = \lim_{q_1 \uparrow 1} \int E[G(Y_1)|V = v]g(v_1, v_2; q_1, q_2)f_{V}(v)\,dv_1\,dv_2
\]

\[
= \int E[G(Y_1)|V = v]1[v_2 < q_2]f_{V}(v)\,dv_1\,dv_2
\]

\[
= \int \int E[G(Y_1)|V_1 = v_1, V_2 = v_2]1[v_2 < q_2]f_{V}(v_1, v_2)\,dv_1\,dv_2
\]

\[
= \int_0^{q_2} E[G(Y_1)|V_2 = v_2]\,dv_2,
\]

where the last equality follows from \(V_2 \sim U[0, 1]\).

From the Leibniz integral rule, we have

\[
\frac{\partial}{\partial q_2} \int_0^{q_2} E[G(Y_1)|V_2 = v_2]\,dv_2 \bigg|_{q_2 = q_2^*} = E[G(Y_1)|V_2 = q_2^*], \quad (A.6)
\]

and the requirement result \((A.5)\) follows.

We complete the proof by showing that

\[
E[G(Y_2)|V_2 = q_2^*] = -\frac{\partial E[G(Y)D_2|Q_1(Z), Q_2(Z)]}{\partial Q_2(Z)}\bigg|_{(Q_1(Z), Q_2(Z))=(1,q_2^*)} \quad (A.7)
\]

Because \((1, q_2)^* \in \mathcal{N}_2((1, q_2^*)', \delta)\) for any \(q_2 \in (q_2^* - \delta, q_2^* + \delta)\) by Lemma \[A.1\] and \(E[G(Y)D_1|Q_1(Z) = q_1, Q_2(Z) = q_2]\) is continuous on \(\mathcal{N}_2((1, q_2^*)', \delta)\), it follows from \(\[A.4\]\) that

\[
E[G(Y)D_2|Q_1(Z) = 1, Q_2(Z) = q_2] = \lim_{q_1 \uparrow 1} \int E[G(Y_2)|V = v]h(v_1, v_2; q_1, q_2)f_{V}(v)\,dv_1\,dv_2,
\]
where \( h(v_1, v_2; q_1, q_2) = \mathbb{1} \left\{ v_1 < F_{\tilde{v}_{1,0}}^{-1}(F_{\tilde{v}_{1,2}}^{-1}(v_2) - F_{\tilde{v}_{1,2}}^{-1}(q_2) + F_{\tilde{v}_{1,0}}^{-1}(q_1)) \right\} \mathbb{1} \{ q_2 \leq v_2 \} \).

Because \( E[|G(Y_2)|] \) is finite by Assumption 3.4, we have \( E[|G(Y_2)||V = v|f_Y(v)] < \infty \) a.s. Therefore, as \( q_1 \uparrow 1 \), in view of \( h(v_1, v_2; 1, q_2) = \mathbb{1} \{ q_2 \leq v_2 \} \)

\[
E[G(Y_2)|V = v|h(v_1, v_2; q_1, q_2)f_Y(v) \rightarrow E[G(Y_2)|V = v]\mathbb{1} \{ q_2 \leq v_2 \}f_Y(v), \quad a.s.
\]

It follows from the DCT and Fubini’s theorem that:

\[
\overline{E}[G(Y)D_2|Q_1(Z) = 1, Q_2(Z) = q_2] = \lim_{q_1 \uparrow 1} \int E[G(Y_2)|V = v|h(v_1, v_2; q_1, q_2)f_Y(v)d(v_1 \times v_2)
\]

\[
= \int E[G(Y_2)|V = v]\mathbb{1} \{ q_2 \leq v_2 \}f_Y(v)d(v_1 \times v_2)
\]

\[
= \int \int E[G(Y_2)|V_1 = v_1, V_2 = v_2]\mathbb{1} \{ q_2 \leq v_2 \}f_Y(v_1, v_2)dv_1dv_2
\]

\[
= \int_{q_2}^{1} E[G(Y_2)|V_2 = v_2]dv_2.
\]

Then, from an argument similar to that in (A.6), we obtain (A.7).

\[\square\]

### A.3 Proof of Theorem 4.1

By definition,

\[
H(Z) = \int g(v_1, v_2; Q_1(Z), Q_2(Z))f_Y(v_1, v_2)d(v_1 \times v_2),
\]

where \( g(v_1, v_2; Q_1(Z), Q_2(Z)) = \mathbb{1} \{ v_1 < Q_1(Z) \}\mathbb{1} \{ v_2 < Q_2(Z) \} \). Because

\[
Q_2(Z) = F_{\tilde{v}_{2,1}^{-1}}(R_{1,2}(Z)) = F_{\tilde{v}_{1,2}^{-1}}(R_1(Z) - R_2(Z))
\]

by Assumption 4.1 we obtain

\[
\lim_{z^{[\varepsilon_2]} \rightarrow a^{[\varepsilon_2]}} Q_2(Z) = 1.
\]

Therefore, as \( z^{[\varepsilon_2]} \rightarrow a^{[\varepsilon_2]} \),

\[
g(v_1, v_2; Q_1(Z), Q_2(Z))f_Y(v) \rightarrow \mathbb{1} \{ v_1 < Q_1(Z) \}f_Y(v), \quad a.s.
\]
Hence, it follows from the DCT and Fubini’s theorem that

\[
\lim_{Z \to a} H(Z) = \lim_{Z \to a} \int g(v_1, v_2; Q_1(Z), Q_2(Z)) f_\mathbf{V}(\mathbf{v}) d(\mathbf{v})
\]

\[
= \int \mathbb{1}(v_1 < Q_1(Z)) f_\mathbf{V}(\mathbf{v}) d(\mathbf{v})
\]

\[
= \int_0^1 \mathbb{1}(v_1 < Q_1(Z)) dv_1 = Q_1(Z),
\]

giving the stated result for \( Q_1(Z) \).

For \( Q_2(Z) \), similar to the proof for \( Q_1(Z) \), we have

\[
Q_1(Z) = F_{\tilde{U}_{1,0}}(\tilde{R}_{1,0}(Z)) = F_{\tilde{U}_{1,0}}(R_1(Z) - R_0(Z)).
\]

From Assumption 4.1, we obtain

\[
\lim_{Z \to a} Q_1(Z) = 1.
\]

Therefore, as \( Z \to a \),

\[
g(v_1, v_2; Q_1(Z), Q_2(Z)) f_\mathbf{V}(\mathbf{v}) \to \mathbb{1}(v_2 < Q_2(Z)) f_\mathbf{V}(\mathbf{v}), \ a.s.
\]

Hence, it follows from the DCT and Fubini’s theorem that:

\[
\lim_{Z \to a} H(Z) = \lim_{Z \to a} \int g(v_1, v_2; Q_1(Z), Q_2(Z)) f_\mathbf{V}(\mathbf{v}) d(\mathbf{v})
\]

\[
= \int \mathbb{1}(v_2 < Q_2(Z)) f_\mathbf{V}(\mathbf{v}) d(\mathbf{v})
\]

\[
= \int_0^1 \mathbb{1}(v_2 < Q_2(Z)) dv_2 = Q_2(Z).
\]

\[\square\]

### A.4 Proof of Lemma 5.1

Let \( q_i^* \) and \( q_j \) denote vectors \((1, \cdots, q_i^*, \cdots, 1)\)' and \((1, \cdots, q_j, \cdots, 1)\)' respectively. In view of Lemma 3.1, we prove the state result by showing that \( E[G(\mathbf{Y})D_k | Q(\mathbf{Z}) = q] \) and \( E[G(\mathbf{Y})D_j | Q(\mathbf{Z}) = q] \) are Cauchy continuous on \( \tilde{N}_{K-1}(q_i^*, \delta) \). First, we show \( E[G(\mathbf{Y})D_k | Q(\mathbf{Z}) = q] \) is Cauchy continuous on \( \tilde{N}_{K-1}(q_j, \delta) \). Under the assumptions
of the theorem, for any \( q \in \mathcal{N}_{K-1}(q^*_j, \delta) \), we obtain
\[
E[\mathbf{G}(Y)D_k|Q(Z) = q] = \mathcal{E}[\mathbf{G}(Y_k)|D_k = 1, Q(Z) = q] \Pr[D_k = 1|Q(Z) = q]
\]
\[
= \mathcal{E}[\mathbf{G}(Y_k)|d_k(V, Q(Z)) = 1, Q(Z) = q] \Pr[d_k(V, Q(Z)) = 1|Q(Z) = q]
\]
\[
= \mathcal{E}[\mathbf{G}(Y_k)|d_k(V, q) = 1] \Pr[d_k(V, q) = 1]
\]
\[
= \mathcal{E}[\mathbf{G}(Y_k)|V] \mathbb{1}(d_k(V, q) = 1),
\]  
(A.8)

where \( Q(Z) = (Q_0(Z), \ldots, Q_{K-1}(Z))^\prime \),
\( q = (q_0, \ldots, q_{K-1})^\prime \), \( V = (V_0, \ldots, V_{K-1})^\prime \), \( d_k(V, Q(Z)) = \prod_{i \in K \setminus \{k\}} \mathbb{1}\{V_i < Q_i(Z)\} \). The third equality holds by Assumption 5.2. Hence, for any \( q \in \mathcal{N}_{K-1}(q^*_j, \delta) \), we have that
\[
E[\mathbf{G}(Y)D_k|Q(Z) = q] = \int E[\mathbf{G}(Y_k)|V = v] g_{K,k}(v; q) f_V(v) dv,
\]  
(A.9)

where \( v = (v_0, \ldots, v_{K-1})^\prime \) and \( g_{K,k}(v; q) := \prod_{i \in K \setminus \{k\}} \mathbb{1}\{v_i < q_i\} \).

Assume \( \{a_n\}_{n=1}^\infty = \{(a_{0n}, \ldots, a_{k-1n}, a_{k+1n}, \ldots, a_{K-1n})\}_{n=1}^\infty \) is a Cauchy sequence in \( \mathcal{N}_{K-1}(q^*_j, \delta) \). Take any \( a_m \) and \( a_\ell \) in \( \{a_n\}_{n=1}^\infty \). Then, we obtain
\[
|E[\mathbf{G}(Y)D_k|Q(Z) = a_m] - E[\mathbf{G}(Y)D_k|Q(Z) = a_\ell]| = \left| \int E[\mathbf{G}(Y_k)|V = v](g_{K,k}(v; a_m) - g_{K,k}(v; a_\ell)) f_V(v) dv \right|.
\]

Because \( E[|\mathbf{G}(Y_k)|] \) is finite by Assumption 5.4, we have \( E[|\mathbf{G}(Y_k)||V = v] f_V(v) < \infty \) a.s. Further, \( g_{K,k}(v; a_m) - g_{K,k}(v; a_\ell) \rightarrow 0 \) a.s. as \( m, \ell \rightarrow \infty \). Therefore, as \( m, \ell \rightarrow \infty \),
\[
E[\mathbf{G}(Y_k)|V = v](g_{K,k}(v; a_m) - g_{K,k}(v; a_\ell)) f_V(v) \rightarrow 0 \quad \text{a.s.}
\]

Hence, it follows from the DCT that, as \( m, \ell \rightarrow \infty \),
\[
|E[\mathbf{G}(Y)D_k|Q(Z) = a_m] - E[\mathbf{G}(Y)D_k|Q(Z) = a_\ell]| \rightarrow 0.
\]  
(A.10)

This implies \( E[\mathbf{G}(Y)D_k|Q(Z) = q] \) is Cauchy continuous on \( \mathcal{N}_{K-1}(q^*_j, \delta) \).

We complete the proof by showing that \( E[\mathbf{G}(Y)D_j|Q(Z) = q] \) is Cauchy continuous on \( \mathcal{N}_{K-1}(q^*_j, \delta) \). From a similar argument to (A.8), under the assumptions of the theorem, we have
\[
E[\mathbf{G}(Y)D_j|Q(Z) = q] = E[E[\mathbf{G}(Y)|V] \mathbb{1}(d_j(V, q) = 1)],
\]

where \( d_j(V, Q(Z)) = \prod_{i \in K \setminus \{j\}} \mathbb{1}\{V_i < F_{\tilde{U}_{j,i}^{-1}}(F_{\tilde{U}_{k,j}^{-1}}(V)) - F_{\tilde{U}_{k,j}^{-1}}(Q_j(Z)) + F_{\tilde{U}_{k,i}^{-1}}(Q_j(Z))\} \times
\[ E[Q_j(Z)] = \mathbb{1}\{V_j \geq Q_j(Z)\} \]. Hence, for any \( q \in \hat{N}_{K-1}(q^*_j, \delta) \), we have

\[
E[G(Y)_j|Q(Z) = q] = \int E[G(Y)_j|V = v]h_{K,k,j}(v; q)f_V(v)dv,
\]

where \( v = (v_0, \cdots, v_{k-1}, v_{k+1}, \cdots, v_{K-1})' \) and

\[
h_{K,k,j}(v; q) := \prod_{i \in K \setminus \{k,j\}} \mathbb{1}\{q_j \leq v_j\} \times \prod_{i \in K \setminus \{k,j\}} \mathbb{1}\{v_i < F^{-1}_{\tilde{U}_{k,i}}(F^{-1}_{\tilde{U}_{k,j}}(v_j) - F^{-1}_{\tilde{U}_{k,j}}(q_j) + F^{-1}_{\tilde{U}_{k,j}}(q_i))\}.
\]

Assume \( \{b_n\}_{n=1}^\infty = \{(b_{n0}, \cdots, b_{kn-1}, b_{k+1}, \cdots, b_{K-1})\}_{n=1}^\infty \) is a Cauchy sequence in \( \hat{N}_{K-1}(q^*_j, \delta) \). Take any \( b_m \) and \( b_\ell \) in \( \{b_n\}_{n=1}^\infty \). Then, we obtain

\[
|E[G(Y)_j|Q(Z) = b_m] - E[G(Y)_j|Q(Z) = b_\ell]| = \left| \int E[G(Y)_j|V = v](h_{K,k,j}(v; b_m) - h_{K,k,j}(v; b_\ell))f_V(v)dv \right|.
\]

Because \( E[|G(Y)_j|] \) is finite by Assumption 5.4, we have \( E[|G(Y)_j||V = v|]f_V(v) < \infty \) a.s. Therefore, a similar argument to (A.10) gives

\[
|E[G(Y)_j|Q(Z) = b_m] - E[G(Y)_j|Q(Z) = b_\ell]| \to 0.
\]

This implies \( E[G(Y)_j|Q(Z) = q] \) is Cauchy continuous on \( \hat{N}_{K-1}(q^*_j, \delta) \).

**A.5 Proof of Theorem 5.2**

We first show that

\[
E[G(Y)_k|V_j = q^*_j] = \frac{\partial E[G(Y)_k|Q(Z)]}{\partial Q_j(Z)}\bigg|_{Q(Z) = q^*_j}.
\]

Because \( q^*_j \in \hat{N}_{K-1}(q^*_j, \delta) \) for any \( q^*_j \in (q^*_j - \delta, q^*_j + \delta) \) by Lemma A.1 and \( \hat{E}[G(Y)_k|Q(Z) = q] \) is continuous on \( \hat{N}_{K-1}(q^*_j, \delta) \), it follows from (A.9) that

\[
\hat{E}[G(Y)_k|Q(Z) = q_j] = \lim_{q_j \to 1_{K-2}} \int E[G(Y)_k|V = v]g_{K,k}(v; q)f_V(v)dv,
\]

where \( q^{-j} = (q_0, \cdots, q_{k-1}, q_{k+1}, \cdots, q_{j-1}, q_{j+1}, \cdots, q_{K-1})' \) and \( g_{K,k}(v; q) := \prod_{i \in K \setminus \{k\}} \mathbb{1}\{v_i < q_i\} \). \( 1_{K-2} \) is a \((K-2) \times 1\) vector of ones. Because \( E[|G(Y)_{kj}|] \) is finite by Assumption 5.4, we have \( E[|G(Y)_k||V = v|]f_V(v) < \infty \) a.s. Hence, as \( q^{-j} \to 1_{K-2} \),

\[
E[G(Y)_k|V = v]g_{K,k}(v; q)f_V(v) \to E[G(Y)_k|V = v]\mathbb{1}[v_j < q_j]f_V(v) \text{ a.s.}
\]
Therefore, it follows from the DCT and Fubini’s theorem that:

\[
\begin{align*}
\bar{E}[G(Y)D_k|Q(Z) = q_j] &= \lim_{q^{-j} \rightarrow 1_{K-2}} \int E[G(Y_k)|V = v] g_{k,k}(v; q_j) f_V(v) dv \\
&= \int E[G(Y_k)|V = v] 1[v_j < q_j] f_V(v) dv \\
&= \int_0^{q_j} E[G(Y_k)|V_j = v_j] dv_j,
\end{align*}
\]

where the last equality follows from \( V_j \sim U[0, 1] \).

From the Leibniz integral rule, we have

\[
\frac{\partial}{\partial q_j} \int_0^{q_j} E[G(Y_k)|V_j = v_j] dv_j \bigg|_{q_j = q_j^*} = E[G(Y_k)|V_j = q_j^*],
\]  

(A.13)

and the requirement result (A.12) follows.

We complete the proof by showing

\[
E[G(Y_j)|V_j = q_j^*] = -\frac{\partial \bar{E}[G(Y)D_j|Q(Z)]}{\partial Q_j(Z)} \bigg|_{Q(Z) = q_j^*}.
\]  

(A.14)

Because \( q_j \in \hat{N}_{K-1}(q_j^*, \delta) \) for any \( q_j \in (q_j^*-\delta, q_j^*+\delta) \) by Lemma A.1 and \( \bar{E}[G(Y)D_j|Q(Z) = q_j] \) is continuous on \( \hat{N}_{K-1}(q_j^*, \delta) \), it follows from (A.11) that

\[
\bar{E}[G(Y)D_j|Q(Z) = q_j] = \lim_{q^{-j} \rightarrow 1_{K-2}} \int E[G(Y_j)|V = v] h_{k,k,j}(v; q_j) f_V(v) dv,
\]

where

\[
h_{k,k,j}(v; q_j) := 1\{q_j \leq v_j\} \times \prod_{i \in \mathcal{K}\setminus \{k,j\}} 1\{v_i < F_{\tilde{U}_{k,i}}^{-1}(F_{\tilde{U}_{k,j}}^{-1}(v_j) - F_{\tilde{U}_{k,j}}^{-1}(q_j) + F_{\tilde{U}_{k,i}}^{-1}(q_i))\}.
\]

Because \( E[|G(Y_j)|] \) is finite from Assumption 5.4 we have \( E[|G(Y_j)||V = v] f_V(v) < \infty \) a.s. Hence, as \( q^{-j} \rightarrow 1_{K-2} \),

\[
E[G(Y_j)|V = v] h_{k,k,j}(v; q_j) f_V(v) \rightarrow E[G(Y_j)|V = v] 1\{q_j \leq v_j\} f_V(v) \quad a.s.
\]  

30
It follows from the DCT and Fubini’s theorem that
\[
E[G(Y_i)D_j|Q(Z) = q_j] = \lim_{q_j \rightarrow 1} \int E[G(Y_j)|V = v]h_{k,j}(v; q_j)f_V(v)dv
\]
\[
= \int E[G(Y_j)|V = v]1[q_j \leq v_j]f_V(v)dv
\]
\[
= \int_{q_j}^1 E[G(Y_j)|V_j = v_j]dv_j.
\]

Then, from an argument similar to that in (A.13), we obtain (A.14).

A.6 Auxiliary Results

Lemma A.1. Let \( d \geq 2 \) and \( 1 \leq k \leq d \). Assume there exists \( \delta > 0 \), such that \( \tilde{N}_d(q_k^*, \delta) \) is in \( \tilde{Q} \). Then, for any \( q_k \in (q_k^* - \delta, q_k^* + \delta) \), we have

\[
q_k \in \tilde{N}_d(q_k^*, \delta),
\]

where \( q_k^* := (1, \cdots, q_k^*, \cdots, 1)' \) and \( q_k := (1, \cdots, q_k, \cdots, 1)' \).

Proof

For simplicity, we assume \( [q_k - \delta, q_k + \delta] \subset (0, 1) \). We define \( \varepsilon \) as follows.

\[
\varepsilon := \sqrt{\delta^2 - (q_k^* - q_k)^2}.
\]

We proceed to show that there exists a sequence \( \{a_n\}_{n=1}^\infty \) in \( \tilde{N}_d(q_k^*, \delta) \) that converges to \( q_k \). Set \( a_n := (1 - \varepsilon/\sqrt{dn}, \cdots, 1 - \varepsilon/\sqrt{dn}, q_k, 1 - \varepsilon/\sqrt{dn}, \cdots, 1 - \varepsilon/\sqrt{dn}) \), where \( n \in N \). Apparently, \( \|a_n - q_k\|_d \rightarrow 0 \) as \( n \rightarrow \infty \). Because \( 0 < 1 - \varepsilon/\sqrt{dn} < 1 \) and \( 0 < q_k < 1 \) by definition, \( a_n \in (0, 1)^d \) for any \( n \in N \). By a trivial calculation, we obtain

\[
\|a_n - q_k^*\|_d = \sqrt{\varepsilon^2/dn \cdot (d - 1) + (q_k - q_k^*)^2 < \sqrt{\varepsilon^2 + (q_k - q_k^*)^2} = \delta}.
\]

Therefore, \( a_n \in \tilde{N}_d(q_k^*, \delta) \) for any \( n \in N \).

B Cauchy continuity

The following two examples illustrate the differences between continuity, Cauchy continuity, and uniform continuity:
Example B.1
Set $A = \mathbb{R}\{0\}$ and $X = \mathbb{R}$. Define $f(x)$ as
\[ f(x) = \sin\left(\frac{1}{x}\right). \]
Clearly, $f(x)$ is continuous on $A$. Define $\{x_n\}_{n=1}^{\infty}$ as $2/(2n+1)\pi$. Because $x_n \to 0$ as $n \to \infty$, $\{x_n\}_{n=1}^{\infty}$ is a Cauchy sequence. Without a loss of generality, we assume $n$ is even. By definition,
\[ f(x_n) = \sin\left(n\pi + \frac{\pi}{2}\right) = 1, \]
\[ f(x_{n+1}) = \sin\left((n+1)\pi + \frac{\pi}{2}\right) = \sin\left(n\pi + \frac{3\pi}{2}\right) = -1. \]
Therefore, $|f(x_n) - f(x_{n+1})| = 2$ for any $n$ and $f$ is not Cauchy continuous on $A$.

Example B.2
We set $A = \mathbb{Q}$ and $X = \mathbb{R}$. We define $g(x)$ as
\[ g(x) = x^2. \]
Apparently, $g(x)$ is not uniformly continuous on $A$.
Assume $\{a_n\}_{n=1}^{\infty}$ is a Cauchy sequence in $A$. Because any Cauchy sequence in $A$ is bounded, there exists $M$ such that $|a_n| \leq M$ for any $n \in \mathbb{N}$. Because $g(x)$ restricted to $|x| \leq M$ is uniformly continuous, $g$ is Cauchy continuous on $A$. 