1 Extending to Multiple Treatments (unordered)

In this section, I extend the previous analysis to allow for multiple treatments. Consider the following switching-regression model with multiple outcome states. Let \( J \) denote the agent’s choice set. I assume that \( J \) contains a finite number of elements. I specify that the value to the agent of choosing option \( j \in J \) is

\[
V_j(Z_j) = \vartheta_j(Z_j) - U_j,
\]

where \( Z_j \) are the agent’s observed characteristics that affect the utility from choosing choice \( j \), and \( U_j \) is the unobserved shock to the agent’s utility from choice \( j \). I will often write \( V_j \) for \( V_j(Z_j) \). Let \( D_{J,j} \) be an indicator variable for whether the agent would choose option \( j \) if confronted with choice set \( J \):

\[
D_{J,j} = 1 \text{ if } V_j \geq V_k \quad \forall k \in J = 0 \text{ otherwise.}
\]

Let \( I_J \) denote the choice that would be made by the agent if confronted with choice set \( J \):

\[
I_J = j \iff D_{J,j} = 1.
\]

Let \( Y_J \) be the outcome variable that would be observed if the agent faced choice set \( J \), determined by

\[
Y_J = \sum_{j \in J} D_{J,j} Y_j,
\]

where \( Y_j \) is the potential outcome, observed only if option \( j \) is chosen. I assume that \( Y_j \) is determined by

\[
Y_j = \nu_j(X_j, \varepsilon_j),
\]

where \( X_j \) is a vector of the agent’s observed characteristics and \( \varepsilon_j \) is an unobserved scalar. I assume that \( (Z, X, I_J, Y_J) \) is observed.\(^1\) I define \( V_J \) as:

\[
V_J = \max_{j \in J} (V_j) = \sum_{j \in J} D_{I,J,j} V_j
\]

We thus have the traditional representation of the decision process that choice \( j \) being optimal implies that choice \( j \) is better than the “next best” option:

\[
I_J = j \iff V_j \geq V_{J \setminus j},
\]

or more generally,

\[
I_J \in K \iff V_K \geq V_{J \setminus K}.
\]

As I will show, this simple, well-known representation is the key intuition for understanding how nonparametric instrumental variables estimates the effect of a given choice versus the “next best” alternative.

I will make the following assumptions, which generalize assumptions (A-1) to (A-5) for the multiple treatment case:

\[\text{(B-1) } \vartheta_j(Z_j) \text{ is a nondegenerate random variable conditional on } X = x \text{ for each } j \in J.\]

\(^1\)One possible extension is to the case where one does not observe which choice was made, but only whether one particular choice was made, i.e., one observes \( D_{J,0} \) but not \( I_J \). The analysis of Thompson (1989) suggests that that extension should be possible.
The distribution of \( \{U_j, \varepsilon_j\}_{j \in \mathcal{J}} \) is nondegenerate and absolutely continuous with respect to Lebesgue measure on \( \prod_j \mathbb{R} \).

\( \{(U_j, \varepsilon_j)\}_{j \in \mathcal{J}} \) is independent of \( \{(X_j, Z_j)\}_{j \in \mathcal{J}} \).

\( Y_j \) is integrable for all \( j \in \mathcal{J} \).

\( \Pr(I_{\mathcal{J}} = j) > 0 \) for all \( j \in \mathcal{J} \).

Assumption (B-2) implies the \( V_j \neq V_k \) w.p.1 for \( j \neq k \), so that \( \arg \max \{V_j\} \) is unique w.p.1. Assumption (B-3) is required for the mean treatment parameters to be well defined. It will also allow me to integrate to the limit, which will be a crucial step for all identification analysis. Assumption (B-5) requires that at least some individuals participate in the program. I thus do not study the more difficult problem of evaluating programs that have not yet been implemented.

Given these assumptions, we have that

\[
\Pr(I_{\mathcal{J}} = j | Z = z) = \frac{\Pr(V_k(z) \leq V_j(z) \forall k \in \mathcal{J})}{\Pr(V_{\mathcal{J}\setminus j}(z) \leq V_j(z))}
\]

and that

\[
E(Y | X = x, Z = z) = E(Y_{\mathcal{J}} | X = x, Z = z) = \sum_{j \in \mathcal{J}} \Pr(I_{\mathcal{J}} = j | Z = z) \times E(Y_j | X = x, Z = z, I_{\mathcal{J}} = j)
\]

From assumption (v), we have that \( E(Y | X = x, Z = z) \) is finite a.e. \( F_{X,Z} \).

I will sometimes impose the following assumption for identification purposes:

For \( j \in \mathcal{J} \), the distribution of \( \vartheta_j(Z_j) \) is nondegenerate conditional on \( \{\vartheta_k(Z_k) = \vartheta_k(z_k)\}_{k \in \mathcal{J}\setminus j} \).

Assumption (B-6) requires that one be able to independently vary the index for the given value function. This will clearly be satisfied if one has an exclusion restriction – in particular, if, for \( j \in \mathcal{J} \), there are elements of \( Z_j \) that are not elements of \( Z_k \) (for \( j \neq k \)) and are not elements of \( X_k \) for any \( k \). Assumption (B-6) will also be satisfied if there are no exclusion restrictions but if instead \( Z \) contains a sufficient number of continuous variables and there is sufficient variation in the \( \vartheta_k \) function across \( k \).

In the binary treatment case, we had that conditioning on the propensity score was equivalent to conditioning on the underlying index. The following lemma shows the analogous result for the multiple treatment case. Let \( P_j(Z) = \Pr(I_{\mathcal{J}} = j | Z = z) \). Let 0 denote an element of \( \mathcal{J} \), and set \( V_{\mathcal{J}}(Z_j) = 0 \).

**Lemma 1.1** Assume that \( \text{Supp}(\{U_j\}) = \prod_j \mathbb{R} \). For \( j \in \mathcal{J} \), \( p_j(v_j) = \Pr(I_{\mathcal{J}} = j | \{V_j(Z) = v_j\}) \). Then:

\( \{P_j = p_j(v_j) : j \in \mathcal{J}\} \leftrightarrow \{V_j = v_j\} \)

**Proof.** See Appendix ??.

---

\(^2\)See the analysis of Cameron and Heckman (1998a) and Chen et al. (1999a,b).
1.1 Definition of Treatment:

I will define treatment effects as the difference in the counterfactual outcomes that would have been observed if the agent faced different choice sets. For any two choice sets, $\mathcal{K}, \mathcal{L} \subset \mathcal{J}$, I will define

$$\Delta_{\mathcal{K}, \mathcal{L}} = Y_{\mathcal{K}} - Y_{\mathcal{L}}.$$  

This is the effect of the individual being forced to choose from choice set $\mathcal{K}$ versus choice set $\mathcal{L}$. The conventional treatment effect is defined as the difference in potential outcomes between two specified states,

$$\Delta_{k, l} = Y_k - Y_l,$$

which is nested within this framework by taking $\mathcal{K} = \{k\}$, $\mathcal{L} = \{l\}$. It is the effect for the individual of having no choice except to choose state $k$ versus having no choice except to choose state $l$.

$\Delta_{\mathcal{K}, \mathcal{L}}$ will be zero for agents who make the same choice when confronted with choice set $\mathcal{K}$ and choice set $\mathcal{L}$. Thus, $I_{\mathcal{K}} = I_{\mathcal{L}}$ implies $\Delta_{\mathcal{K}, \mathcal{L}} = 0$, and we have

$$\Delta_{\mathcal{K}, \mathcal{L}} = 1(I_{\mathcal{L}} \neq I_{\mathcal{K}})\Delta_{\mathcal{K} = \mathcal{L}, \mathcal{L}}$$  

$$(1)$$

$$= 1(I_{\mathcal{L}} \neq I_{\mathcal{K}}) \left( \sum_{j \in \mathcal{K} = \mathcal{L}} D_{\mathcal{K}, j} \Delta_{j, \mathcal{L}} \right).$$  

$$(2)$$

Two examples will be of particular importance for this analysis. First, consider choice set $\mathcal{K} = \{k\}$ versus choice set $\mathcal{L} = \mathcal{J} \setminus \{k\}$. In this case, $\Delta_{k, \mathcal{J} \setminus k}$ is the difference between the agent’s potential outcome in state $k$ versus the outcome that would have been observed if he or she had not been allowed to choose state $k$. If $I_{\mathcal{J}} = k$, then $\Delta_{k, \mathcal{J} \setminus k}$ is the difference between the outcome in the agent’s preferred state and the outcome in the agent’s “next-best” state. Second, consider the set $\mathcal{K} = \mathcal{J}$ versus choice set $\mathcal{L} = \mathcal{I} \setminus \{k\}$. In this case, $\Delta_{\mathcal{J}, \mathcal{J} \setminus k}$ is the difference between the agent’s observed outcome and what his or her outcome would have been if state $k$ had not been available. Note that

$$\Delta_{\mathcal{J}, \mathcal{J} \setminus k} = D_{\mathcal{J}, k} \Delta_{k, \mathcal{J} \setminus k}.$$  

$$(3)$$

**Example: GED Certification (continued)** Consider the case where $\mathcal{J} = \{ \{\text{GED}\}, \{\text{HS Degree}\}, \{\text{Permanent Dropout}\}\}$. Let $j = \{\text{GED}\}$, $k = \{\text{HS Degree}\}$, and $l = \{\text{Permanent Dropout}\}$. Suppose one wishes to study the effect of the GED. Then possible definitions of the effect of the GED include:

- $\Delta_{j, k}$, is the individual’s outcome if he or she received the GED versus if he or she had graduated from High School;
- $\Delta_{j, l}$, is the individual’s outcome if he or she received the GED versus if he or she had been a permanent dropout;
- $\Delta_{j, \mathcal{J} \setminus j}$, is the individual’s outcome if he or she had received the GED versus what the outcome would have been if he or she had not had the option of receiving the GED;
- $\Delta_{\mathcal{J}, \mathcal{J} \setminus j}$, is the individual’s outcome if he or she had the option of receiving the GED versus the outcome if he or she did not have the option of receiving the GED.

1.2 Treatment Parameters

The conventional definition of the average treatment effect (ATE) is

$$\Delta_{k, l}^{ATE}(x) = E(\Delta_{k, l}|X = x),$$

3
which immediately generalizes to the class of parameters discussed in this section as:

$$\Delta_{K,L}^{ATE}(x) = E(\Delta_{K,L}|X = x).$$

The conventional definition of the treatment on the treated (TT) parameter is

$$\Delta_{k,l}^{TT}(x) = E(\Delta_{k,l}|X = x, I_\mathcal{J} = k),$$

which I will generalize to

$$\Delta_{K,L}^{TT}(x) = E(\Delta_{K,L}|X = x, I_\mathcal{J} \in K - L).$$

I can also define marginal versions of the parameters. Two examples will be of particular importance for my study of nonparametric instrumental variables. The first is the average effect conditional on being indifferent between the best option among choice set $K$ versus the best option among choice set $L$ at some given fixed value of the instruments, $Z = z$:

$$E(\Delta_{K,L}|X = x, V_K(z) = V_L(z)),$$

(4)

The second example is average effect for someone for whom the optimal choice in choice set $K$ is preferred to the optimal choice in choice set $L$ at $Z = z$, but who prefers the optimal choice in choice set $L$ to the optimal choice in choice set $K$ at $Z = \tilde{z}$:

$$E(\Delta_{K,L}|X = x, V_K(z) \geq V_L(z), V_L(\tilde{z}) \geq V_K(\tilde{z}))$$

(5)

If the treatment effect of interest is $\Delta_{\mathcal{J},\mathcal{L}}$, then without loss of generality we can instead work with $K = \mathcal{J} \setminus \mathcal{L}$. To see this, note that from equation (1), we have that:

$$\Delta_{j,\mathcal{L}}^{TE}(x) = Pr(I_\mathcal{J} \notin \mathcal{L})E(\Delta_{\mathcal{J},\mathcal{L}}|X = x).$$

with similar expressions for the other mean treatment parameters. Since $I_\mathcal{J}$ is identified (it is the observed choice), we have that $Pr(I_\mathcal{J} \notin \mathcal{L})$ is identified and thus that by a known rescaling of $E(\Delta_{\mathcal{J},\mathcal{L}}|X = x)$ we attain $E(\Delta_{\mathcal{J},\mathcal{L}}|X = x)$ For example, if we identify $E(\Delta_{j,\mathcal{J}\setminus j}|X = x)$, we can multiply the parameter by $Pr(I_\mathcal{J})$ to attain the $E(\Delta_{\mathcal{J},\mathcal{J}\setminus j}|X = x)$ parameter.

### 1.3 Heterogeneity in Treatment Effects

Consider heterogeneity in $\Delta_{j,k}$. We have that

$$\Delta_{j,k} = Y_j - Y_k = \nu_j(X_j, \varepsilon_j) - \nu_k(X_k, \varepsilon_k),$$

which in general will vary with both observables ($X$) and unobservables ($\varepsilon_j, \varepsilon_k$). Since I have not assumed that the error terms are additively separable, the treatment effect will in general vary with unobservables even if $\varepsilon_j = \varepsilon_k$.

The mean treatment parameters for $\Delta_{j,k}$ will differ if the effect of treatment is heterogeneous and agents base participation decisions, in part, on their idiosyncratic treatment effect. In general, the ATE, TT, and the marginal treatment parameters for $\Delta_{j,k}$ will differ as long as there is dependence between $(\varepsilon_j, \varepsilon_k)$ and the decision rule, i.e., if there is dependence between $(\varepsilon_j, \varepsilon_k)$ and $\{U_1\}$. If we impose that $\{U_1\}$ is independent of $(\varepsilon_j, \varepsilon_k)$, then the treatment effect will still be heterogeneous, but the average treatment effect, average effect of treatment on the treated, and the marginal average treatment effects will be the same.\(^3\)

\(^3\)Heckman (1997) emphasizes the importance for social program evaluation of whether agents base participation decisions on their idiosyncratic gains from participating.

The literature almost always imposes additive separability: \( Y_j = \nu_j(X_j) + \varepsilon_j \) and \( Y_k = \nu_k(X_k) + \varepsilon_k \). In that case, we have the standard result that a common treatment effect is equivalent to the error term not varying with the treatment state: \( \varepsilon_j = \varepsilon_k \). Thus, in the special case of additive separability, the treatment parameters for \( \Delta_{j,k} \) will be the same even if there is dependence between \((\{U_l\})\) and \((\varepsilon_j, \varepsilon_k)\) as long as \( \varepsilon_j \) and \( \varepsilon_k \) are perfectly dependent.\(^4\)

There is an additional source of treatment heterogeneity in the more general case of \( \Delta_{k,l} \), heterogeneity in which states are being compared. Consider, for example, \( \Delta_{j,J\setminus j} \). We have that

\[
\Delta_{j,J\setminus j} = \sum_{k\in J\setminus j} D_{j,k}\Delta_{j,k},
\]

which will vary over individuals even if each individual has the same \( \Delta_{j,k} \) treatment effect. Consider the corresponding ATE and TT parameters:

\[
\Delta_{j,J\setminus j}^{ATE}(x) = E(\Delta_{j,J\setminus j}|X = x)
= \sum_{k\in J\setminus j} Pr(I_{J\setminus j} = k)E(\Delta_{j,k}|X = x, I_{J\setminus j} = k)
\]

and

\[
\Delta_{j,J\setminus j}^{TT}(x) = E(\Delta_{j,J\setminus j}|X = x, I_{J} = j)
= \sum_{k\in J\setminus j} Pr(I_{J\setminus j} = k|I_{J} = j)E(\Delta_{j,k}|X = x, I_{J} = j, I_{J\setminus j} = k)
\]

Even in the case where \( U \) is independent of \( \varepsilon \), so that \( E(\Delta_{j,k}|X = x, I_{J\setminus j} = k) = E(\Delta_{j,k}|X = x, I_{J} = j, I_{J\setminus j} = k) \), it will still in general be the case that \( \Delta_{j,J\setminus j}^{ATE}(x) \neq \Delta_{j,J\setminus j}^{TT}(x) \) since in general \( Pr(I_{J\setminus j} = k) \neq Pr(I_{J\setminus j} = k|I_{J} = j) \) unless \( U_j \) is independent of the remaining \( U_k \) shocks.

In summary, \( \Delta_{j,k} \) will be heterogeneous depending on the functional form of the \( \nu \) equation and on the pairwise dependence between \( \varepsilon_j \) terms. The \( \Delta_{j,k} \) mean treatment parameters will also vary depending on the dependence between \( \varepsilon \) and \( U \). For \( \Delta_{j,J\setminus j} \), there is an additional source of heterogeneity – what option is optimal in the set \( J \setminus j \). Even if there is no heterogeneity in the pairwise \( \Delta_{j,k} \) terms, there will still be heterogeneity in \( \Delta_{j,J\setminus j} \), and heterogeneity in the corresponding mean treatment parameters if there is dependence in the \( U \) terms.

### 1.4 Identification for Multiple Treatments

I examine the extension of the techniques of the previous sections to the above latent variable framework with multiple treatments. For notational purposes, for any \( j,k,\in J \), define \( U_{j,k} = U_j - U_k \), and let \( \vartheta_{j,k}(Z) = \vartheta_j(Z_j) - \vartheta_k(Z_k) \). I first examine a version of the local average treatment effect (LATE) parameter of Imbens and Angrist (1994) where I use the \( \vartheta_j \) index as the instrument.\(^5\) First note that

\[
E(Y|X = x, Z = z) = E(Y|X = x, \{\vartheta_j(Z_j) = \vartheta_j(z)\}_{j\in J})
\]

and that

\[
Pr(I_{J} = j|X = x, Z = z) = Pr(I_{J} = j|\{\vartheta_j(Z_j) = \vartheta_j(z)\}_{j\in J}).
\]

---

\(^4\)Because the literature almost always assumes additive separability, questions of a common treatment effect becomes a question of whether the additively separable error terms differ by treatment state. If the errors terms differ by treatment state, there will be differences in the treatment parameters according to whether the differences in the error terms are stochastically dependent on the participation decision. See, e.g., Heckman and Robb (1985), Heckman, Smith and Clements (1997), and Heckman (1997). As an exception, see Aakvik et al. (1999), who examine the case where the outcome variable is binary so that an additive separability assumption is not appropriate.

\(^5\)All results will hold varying the \( Z \) vector instead of the \( \vartheta \) indices. In practice, the \( \vartheta \) indices are not known so that one must vary the \( Z \) vector instead.
For \( \vartheta_j > \tilde{\vartheta}_j \), let \( z \) and \( \tilde{z} \) be s.t. \( \vartheta_k(z_k) = \vartheta_k(\tilde{z}_k) = \vartheta_k \) for \( k \neq j \), and \( \vartheta_j(\tilde{z}) \equiv \tilde{\vartheta}_j \leq \vartheta_j \equiv \vartheta_j(z_j) \). I define the LATE parameter as follows,

\[
\Delta_j^{\text{LATE}}(x, \vartheta_j, \tilde{\vartheta}_j, \{ \vartheta_k(z_k) \}_{k \in \mathcal{J} \setminus j}) \equiv \frac{E(Y|X=x, Z=z) - E(Y|X=x, Z=\tilde{z})}{\text{Pr}(U_{j,j}=1|X=x, Z=z) - \text{Pr}(U_{j,j}=1|X=x, Z=\tilde{z})}
\]

I also examine the form of the local instrumental variables parameter (LIV) introduced in Heckman (1997),

\[
\Delta_j^{\text{LIV}}(x, \{ \vartheta_k(z_k) \}_{k \in \mathcal{J}}) \equiv \frac{\partial E(Y|X=x, (\vartheta_k(z_k))_{k \in \mathcal{J}})}{\partial \vartheta_j} \frac{\partial \text{Pr}(U_{j,j}=1|\vartheta_k(z_k))_{k \in \mathcal{J}}}{\partial \vartheta_j}
\]

LIV is thus the limit form of the LATE parameter. Given my previous assumptions, this limit exists w.p.1.6

I now have the following Lemma,

**Lemma 1.2**

\[
\Delta_j^{\text{LIV}}(x, \{ \vartheta_k(z_k) \}_{k \in \mathcal{J}}) = E(\Delta_{j,J \setminus j}|X = x, V_j(z) = V_{J \setminus j}(z))
\]

\[
\Delta_j^{\text{LATE}}(x, \vartheta_j, \tilde{\vartheta}_j, \{ \vartheta_k(z_k) \}_{k \in \mathcal{J} \setminus j}) = E(\Delta_{j,J \setminus j}|X = x, V_j(z) \geq V_{J \setminus j}(z) \geq V_j(\tilde{z}))
\]

**Proof.** See Appendix ??.

\( \Delta_j^{\text{LIV}}(x, \{ \vartheta_k(z_k) \}_{k \in \mathcal{J}}) \) is the average effect of switching to state \( j \) from state \( I_{J \setminus j} \) (the best option besides state \( j \)) for individuals who are indifferent between state \( j \) and \( I_{J \setminus j} \) at the given values of the selection indices (at \( \{ \vartheta_k(Z_k) = \vartheta_k(z_k) \}_{k \in \mathcal{J}} \)). \( \Delta_j^{\text{LATE}}(x, \vartheta_j, \tilde{\vartheta}_j, \{ \vartheta_k(z_k) \}_{k \in \mathcal{J} \setminus j}) \) is the average effect of switching to state \( j \) from state \( I_{J \setminus j} \) for individuals who would choose \( I_{J \setminus j} \) at \( \vartheta_j(Z_j) = \vartheta_j \) but would choose \( j \) at \( \vartheta_j(Z_j) = \tilde{\vartheta}_j \), holding the values of the other indices fixed at a given level.

The average effect of state \( j \) versus state \( I_{J \setminus j} \) (the next best option) is a weighted average over \( k \in \mathcal{J} \setminus j \) of the effect of state \( j \) versus state \( k \), conditional on \( k \) being the next best option, weighted by the probability that \( k \) is the next best option. For example, for the LATE parameter,

\[
E(\Delta_{j,J \setminus j}|X = x, V_j(z) \geq V_{J \setminus j}(z) \geq V_j(\tilde{z})) = \sum_{k \in \mathcal{J} \setminus j} \text{Pr}(I_{J \setminus j} = k|Z = z, V_j(z) \geq V_{J \setminus j}(z) \geq V_j(\tilde{z})) \times E(\Delta_{j,k}|X = x, V_j(z) \geq V_k(z) \geq V_j(\tilde{z}), I_{J \setminus j} = k)
\]

How heavily each option is weighted in this average depends on \( \text{Pr}(I_{J \setminus j} = k|Z = z, V_j(z) \geq V_k(z) \geq V_j(\tilde{z})) \), which in turn depends on the \( \{ \vartheta_k(z_k) \}_{k \in \mathcal{J} \setminus j} \) evaluation points. The higher \( \vartheta_k \), holding the other evaluation points constant, the larger the weight given to state \( k \) as the base state.

Figure 1 graphically illustrates the case where \( J = \{1, 2, 3\} \). The rectangle \( A \) is the set of \( (U_{31}, U_{32}) = (U_3 - U_1, U_3 - U_2) \) points such that the individual would choose state 3 if the indices were set at \( (\vartheta_{31}, \vartheta_{32}) = (\tilde{\vartheta}_3 - \vartheta_1, \tilde{\vartheta}_3 - \vartheta_2) \). The line \( V_1 = V_2 \) is the set of points where the individual is indifferent between state 1 and state 2. To the northwest of this line are \( (U_{31}, U_{32}) \) points such that \( I_{J \setminus 3} = 1 \), and to the southeast of the line are points \( (U_{31}, U_{32}) \) such that \( I_{J \setminus 3} = 2 \).

As \( \vartheta_3 \) is increased, holding \( (\vartheta_1, \vartheta_2) \) constant, the rectangle of points corresponding to \( I_j = 3 \) expands to the northeast, with the upper-right vertex staying along the \( V_1 = V_2 \) ray. \( B \) corresponds to the \( (U_{31}, U_{32}) \) points s.t. the individual would choose state 1 at \( \vartheta_3 \) but would choose state 3 at \( \vartheta_3 \). These points correspond to individuals who would switch from state 1 to state 3 because of the change in \( \vartheta_3 \). Likewise, the points in \( C \) correspond to individuals who would switch from state 2 to state 3 because of the change in \( \vartheta_3 \). The \( V_1 = V_2 \) line will have a slope of 45 degrees regardless of the values of the \( \{ \vartheta_k \} \) indices, but its intercept is determined by \( \vartheta_1 - \vartheta_2 \). The LATE parameter is the average effect of switching to state 3

---

6 The proof is trivial, but is included in the appendix for completeness.
for individuals with \((U_{31}, U_{32}) = (U_3 - U_1)\) values in \(B\) and \(C\). The LIV parameter is the average effect of switching to state 3 for individuals with \((U_{31}, U_{32}) = (U_3 - U_1)\) values at the top and right edges of the \(A\) rectangle.

Alternatively, one can define averaged versions of the LIV and LATE parameters, that take \(\vartheta_j\) and \((\vartheta_j, \vartheta_j)\) as arguments, respectively, but average over \(\{\vartheta_k(Z_k)\}_{k \in \mathcal{J} \cup \mathcal{J}}\):

\[
\Delta_j^{LIV}(x, \vartheta_j) = \int \Delta_j^{LIV}(x, \vartheta_j, \{\vartheta_k(Z_k)\}_{k \in \mathcal{J} \cup \mathcal{J}}) dF_{Z_k}
\]

\[
= E(\Delta_j, \mathcal{J}, \vartheta_j \mid X = x, \vartheta_j - U_j = V_{\mathcal{J} \cup \mathcal{J}}(Z))
\]

\[
\Delta_j^{LATE}(x, \vartheta_j, \vartheta_j) = \int \Delta_j^{LATE}(x, \vartheta_j, \vartheta_j, \{\vartheta_k(Z_k)\}_{k \in \mathcal{J} \cup \mathcal{J}}) dF_{Z_k}
\]

\[
= E(\Delta_j, \mathcal{J}, \vartheta_j \mid X = x, \vartheta_j + U_j \geq V_{\mathcal{J} \cup \mathcal{J}}(Z) \geq \vartheta_j + U_j)
\]

We can follow the same analysis to examine \(\Delta_{K, \mathcal{J}, \mathcal{K}}\). Define:

\[
\Delta_K^{LATE}(x, h, \{\vartheta_k\}_{k \in \mathcal{J}}) = A
\]

where

\[
A = E(Y \mid X = x, \{\vartheta_k(Z_k) = \vartheta_k + h\}_{k \in \mathcal{K}}, \{\vartheta_l(Z_l) = \vartheta_l\}_{l \in \mathcal{J} \cap \mathcal{K}}) - E(Y \mid X = x, \{\vartheta_l(Z_l) = \vartheta_l\}_{l \in \mathcal{J}})
\]

\[
B = Pr(I_{\mathcal{J}} \in \mathcal{K} \mid X = x, \{\vartheta_k(Z_k) = \vartheta_k + h\}_{k \in \mathcal{K}}, \{\vartheta_l(Z_l) = \vartheta_l\}_{l \in \mathcal{J} \cap \mathcal{K}}) - Pr(I_{\mathcal{J}} \in \mathcal{K} \mid X = x, \{\vartheta_l(Z_l) = \vartheta_l\}_{l \in \mathcal{J}})
\]

And define

\[
\Delta_K^{LIV}(x, \{\vartheta_k\}_{k \in \mathcal{J}}) = \lim_{h \to 0} \Delta_K^{LATE}(x, h, \{\vartheta_k\}_{k \in \mathcal{J}})
\]

We then have the following Lemma:

**Lemma 1.3 We have that:**

\[
\Delta_K^{LIV}(x, \{\vartheta_k\}_{k \in \mathcal{J}}) = E(\Delta_{K, \mathcal{J}, \mathcal{K}} \mid X = x, V_K(z) = V_{\mathcal{J} \cup \mathcal{K}}(z))
\]

\[
\Delta_K^{LATE}(x, h, \{\vartheta_k\}_{k \in \mathcal{J}}) = E(\Delta_{K, \mathcal{J}, \mathcal{K}} \mid X = x, V_K + h \geq V_{\mathcal{J} \cup \mathcal{K}}(z) \geq V_K)
\]

where \(V_K = \max_{k \in \mathcal{K}} \{\vartheta_k + U_k\}\).

**Proof.** Proof follows with trivial modifications from proof of Lemma (1.2). □

Can we identify \(E(\Delta_{j,k} | V_j(z) = V_k(z))\) or \(E(\Delta_{j,k} | V_j(z) \geq V_k(z) \geq V_j(\hat{z}))\) using the same strategy? \(E(Y_j - Y_k | V_j(z) = V_k(z))\) and \(E(Y_j - Y_k | V_j(z) \geq V_k(z) \geq V_j(\hat{z}))\) correspond to shifting \(\vartheta_j - \vartheta_k = \vartheta_{j,k}\) while \(\{\vartheta_{l,m}\}_{(l,m) \in \mathcal{J} \times \mathcal{J} \setminus \{j,k\}}\) is held constant.\(^7\) However, given the structure of the latent variable model, these are incompatible conditions. To see this, note that \(\vartheta_{j,k} = \vartheta_{l,k} - \vartheta_{l,j}\) for any \(l\), and thus \(\vartheta_{j,k}\) cannot be shifted while holding \(\vartheta_{l,j}\) and \(\vartheta_{l,k}\) constant.

(This suggests a nonparametric test of the latent-variable model. If there exists \((z, z')\) s.t. \(Pr(I_{\mathcal{J}} = j \mid Z = z) \neq Pr(I_{\mathcal{J}} = j \mid Z = z')\) and \(Pr(I_{\mathcal{J}} = k \mid Z = z) \neq Pr(I_{\mathcal{J}} = k \mid Z = z')\), but \(Pr(I_{\mathcal{J}} = l \mid Z = z) \neq Pr(I_{\mathcal{J}} = l \mid Z = z')\) for all \(l \in \mathcal{J} \setminus \{j, k\}\), then the latent variable model is rejected. However, such shifts

\(^7\)Alternatively, one can allow \(\vartheta_{l,m}(z) \neq \vartheta_{l,m}(z')\) if \(Pr(U_{l,m} \in [\vartheta_{l,m}(z), \vartheta_{l,m}(z')]) = 0\). Such a possibility would be ruled out except “at the limit” (see below) by the standard assumption that the support of \(U_{l,m}\) is connected. Even without such an assumption, such a possibility occurring simultaneously for all \((l, m) \in \mathcal{J} \times \mathcal{J} \setminus \{j, k\}\) for a particular \(z, z'\) seems extremely implausible, and I will therefore not consider this possibility further.
in only two indices are possible for sequential models since unexpected shocks will act to shift the current decision without effecting previous decisions. Consider the following sequential model of GED certification. In the first period, the agent chooses to graduate from high school or to dropout of high school. If the agent drops out of high school in the first period, he or she has the option in the second period of attaining GED certification or staying a permanent dropout. An unexpected shock in the second period to the relative value of GED certification versus permanent dropout status will shift the GED/permanent dropout choice without changing the probability of high school graduation.)

Again, consider Figure 1. To identify $E(Y_3 - Y_1 | V_3(z) \geq V_1(z) \geq V_3(\tilde{z}))$, I would need to shift $\vartheta_{31}$ up (shift the upper edge of the rectangle up) while simultaneously holding $\vartheta_{32}$ constant (hold the right edge of the rectangle constant), and keeping $\vartheta_{12}$ constant (the $V_1 = V_2$ line). However, $\vartheta_{31} = \vartheta_{21} - \vartheta_{32}$, and so a shift in $\vartheta_{31}$ requires a shift in $\vartheta_{32}$ or $\vartheta_{21}$ (or both).

However, while $\{\vartheta_{l,m}\}_{(l,m) \in J \times J\setminus\{j,k\}}$ cannot be held constant, we can use a limit strategy to make the consequences of shifting them negligible. From Figure 1, we have that for a smaller $\vartheta_2$ (compared to $\vartheta_1$), the further to the right the $V_1 = V_2$ line is shifted, the larger $B$ will be relative to the $C$ rectangle, and the fraction of “compliers” who are switching from state 1 to state 3 and not from state 2 to state 3 will also be larger. I thus have that

$$\left| E(Y_3 - Y_1 | X = x, V_3(z) = V_1(z)) - E(Y_3 - Y_{J\setminus 3} | X = x, V_3(z) \geq V_{J\setminus 3}(z) \geq V_3(\tilde{z})) \right|$$

is decreasing in $\vartheta_1 - \vartheta_2$.

Under the appropriate support conditions, we can use a limit strategy to set $E(\Delta_{j,J} | X = x, V_j(z) = V_{J\setminus j})$ arbitrarily close to $E(\Delta_{j,J} | X = x, V_j(z) = V_k(z))$.

**Lemma 1.4** Assume that:

1. Support of $\{U_j\}_{j \in J}$ is $\prod_{j \in J} \mathbb{R}$.
2. For any $t \in \mathbb{R}$, $Pr\{\vartheta_j(Z_l) \leq t \ \forall \ l \in J \setminus \{j,k\}\} \vartheta_j(Z_j) = \vartheta_j(z_j), \vartheta_k(Z_k) = \vartheta_k(z_k)) \geq 0$.

Then

$$\lim_{t \to -\infty} \sup_{j \in J \setminus \{j,k\}} \left| E(\Delta_{j,k} | X = x, V_j(z) = V_k(z)) - E(\Delta_{j,J\setminus j} | X = x, V_j(z) = V_{J\setminus j}(z)) \right| = 0$$

for any $\vartheta_{J\setminus \{j,k\}}(\vartheta_j) \to -\infty$.

**Proof.** See Appendix ??.

For such $x$ values, we can thus approximate $E(\Delta_{j,k} | X = x, V_j(z) = V_k(z))$ arbitrarily well by $\Delta^{LIV}_j (x, \{\vartheta_k(z_k)\}_{k \in J})$ for arbitrarily small $\vartheta_{J\setminus \{j,k\}}$. Following Heckman and Vytlacil (1999b), we can use the LATE and LIV parameters to identify the ATE and TT parameters under the proper support conditions. Thus, if $\vartheta_j(Z_j)$ has support $\mathbb{R}$ on the values of the other indices, we can approximate $\Delta_{j,J\setminus j}^{ATE}(x, \vartheta_j, \tilde{\vartheta}_j, \{\vartheta_k(z_k)\}_{k \in J\setminus j})$ at an arbitrarily large $\vartheta_j$ value and an arbitrarily small $\tilde{\vartheta}_j$ value. Following Heckman and Vytlacil (1999b), I can show that in order to approximate $\Delta_{j,J\setminus j}^{TT}(x, I_J = j)$ arbitrarily well I need only be able to evaluate $\Delta_{j,J\setminus j}^{LATE}(x, \vartheta_j, \tilde{\vartheta}_j, \{\vartheta_k(z_k)\}_{k \in J\setminus j})$ at arbitrarily small $\tilde{\vartheta}_j$ values. The same analysis applies to $\Delta_{j,k}^{ATE}(x)$ and $\Delta_{j,k}^{TT}(x, I_J = k)$ if I additionally evaluate $\Delta_{j,J\setminus j}^{LATE}(x, \vartheta_j, \tilde{\vartheta}_j, \{\vartheta_k(z_k)\}_{k \in J\setminus j})$ at arbitrarily small values of $\vartheta_l$ for $l \in J \setminus \{j,k\}$. In summary,
\[ E(\Delta_{j,\mathcal{J}\setminus j}|X = x, V_j(z) = V_{\mathcal{J}\setminus j}(z)) \] and \[ E(\Delta_{j,\mathcal{J}\setminus j}|X = x, V_j(z) \geq V_{\mathcal{J}\setminus j}(z) \geq V_{\hat{j}}(\tilde{z})) \] can be identified without a limit argument.

\[ E(\Delta_{j,k}|X = x, V_j(z) = V_k(z)) \] and \[ E(\Delta_{j,k}|X = x, V_j(z) \geq V_k(z) \geq V_{\hat{j}}(\tilde{z})) \] can be identified with a limit argument on each index in \( \mathcal{J} \setminus \{j, k\} \).

\( \Delta_{j,\mathcal{J}\setminus j}^{ATE}(x) \) and \( \Delta_{j,\mathcal{J}\setminus j}^{ITT}(x, I_{\mathcal{J}} = j) \) can be identified with a limit argument on the \( \vartheta_j \) index.

\( \Delta_{j,k}^{ATE}(x) \) and \( \Delta_{j,k}^{ITT}(x, I_{\mathcal{J}} = j) \) can be identified with a limit argument on each index.