Notes on Twin Models

Rodrigo Pinto

University of Chicago
HCEO Seminar
April 19, 2014
This draft, April 19, 2014 8:17am
Figure 1: Standard Confounding Model
Figure 2: Non-parametric IV Solution
IV Properties

1. **Independence:** $V \perp Z$ and $Y \perp Z|(X, V)$.
2. **Or:** $Y(x) \perp Z|X$ and $V \not\perp X$.
3. **Identification:** Based on Assumptions on the causal link between $Z$ and $X$. 

Rodrigo Pinto
Gene-environment Interaction and Causality, April 19, 2014 8:17am
Theorem 1

In the standard confounding model with IV, where $Z$ is the instrument that takes values on \{${z_1, \ldots, z_{N_Z}}$\}, $V$ are unobserved variables and $X$ is the categorical treatment, the following statements are equivalent:

(i) An utility function $u$ representing rational preferences over $X$, $V$, $Z$ is additive in $Z$, $V$, i.e.:

$$u(X, Z, V) \equiv u_1(X, V) + u_2(X, Z);$$

(ii) For any $z, z' \in \text{supp}(Z)$ and $v, v' \in \text{supp}(V)$ such that:

$$X|(Z = z, V = v) = t \text{ and } X|(Z = z', V = v') = t,$$

then $X|(Z = z, V = v') = t$ or $X|(Z = z', V = v) = t$.

(iii) Each Stata Matrices $A_x; x \in \text{supp}(X)$ is a Lonesum Matrix.

(iv) Each Stata Matrices $A_x; x \in \text{supp}(X)$ is equivalent to its maximal matrix $\bar{A}_t$.

(v) Treatment $X$ is separable on $V$ and $Z$, that is:

there exist functions, $f_x : \text{supp}(Z) \to [0, 1]$ and $g_x : \text{supp}(V) \to [0, 1]; \forall x \in \text{supp}(X)$
such that $P_E\left( X = \sum_{x \in \text{supp}(X)} 1[f_x(Z) \leq g_x(V)] \cdot x \right) = 1.$
Twins - Two Useful Properties

1. **Confounding Dependence:** Twins share, to some degree, the confounding variables causing outcomes.

2. **Independent Variation:** While Monozygotic twins (MZ) share the same *genetic* endowment, Dizygotic twins don’t (DZ);
Figure 3: Twin Identification – Two Properties

\[ \frac{dz}{mz} \]
Classification of Confounding Variables

Properties (1) and (2) suggest a partition of all confounding variables into:

1. **A. Genetic Endowment:** confounding variables that are equal only if twins are MZ;

2. **C. Shared Environment:** confounding variables that are equal for twins regardless of type (MZ or DZ);

3. **E. Non-shared Environment:** remaining confounding variables.

4. **Exclusion Restriction:** No confounding variables is equal for DZ and different for MZ;

<table>
<thead>
<tr>
<th>Are Confounding Variables equal across twins?</th>
<th>Genetic Type</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Genetic Endowment</td>
<td>DZ</td>
<td>✓</td>
</tr>
<tr>
<td>2. Shared Environment</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>3. Non-shared Environment</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>4. Exclusion Assumption</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Statistical Properties

Classification based on Properties (1) and (2) generate the following properties:

1 **Genetic Endowment:**
   \[
   \Pr(A_1^X = A_2^X | MZ = 1) = 1.
   \]

2 **Shared Environment:**
   \[
   \Pr(C_1^X = C_2^X | MZ = 1) = \Pr(C_1^X = C_2^X | MZ = 0) = 1.
   \]

3 **Non-shared Environment (Assumption):**
   \[
   E_1^X \perp \perp E_2^X | (MZ = 1), (E_1^X \perp \perp E_2^X | MZ = 0).
   \]

4 **Does not justify:**
   \[
   A_i^X \perp \perp C_i^X, A_i^X \perp \perp E_i^X, C_i^X \perp \perp E_i^X;
   \]
Figure 4: **Standard Univariate ACE Model**

\[
\begin{align*}
E_{1}^{X} & \rightarrow X_{1} \\
A_{1}^{X} & \rightarrow C_{1}^{X} \\
A_{2}^{X} & \rightarrow C_{2}^{X} \\
E_{2}^{X} & \rightarrow X_{2} \\
E_{1}^{X} & = e_{xx} \\
A_{1}^{X} & = a_{xx} \\
C_{1}^{X} & = c_{xx} \\
A_{2}^{X} & = a_{xx} \\
C_{2}^{X} & = c_{xx} \\
dz & = 0.5mz \\
mz & = dz = 1
\end{align*}
\]
Properties of the Standard ACE Model

1. **Variance Decomposition**: heritable \( \left( \frac{a_{xx}^2}{a_{xx}^2 + c_{xx}^2 + a_{xx}^2} \right) \), shared environment \( \left( \frac{c_{xx}^2}{a_{xx}^2 + c_{xx}^2 + a_{xx}^2} \right) \) or non-shared environment \( \left( \frac{e_{xx}^2}{a_{xx}^2 + c_{xx}^2 + a_{xx}^2} \right) \).

2. **Concept**: Decompose the confounding variables \( V_1, V_2 \) into:
   - Variables that are independent between twins \( (E_1^X, E_1^X) \):
     \[ (E2, E1) \perp \perp (A1, A2, C1, C2, Y2) \text{ and } E1 \perp \perp E2 \]
   - Variables that are equal for both twins types \( (C_1^X, C_1^X) \):
     \[ \Pr(C_1^X = C_2^X|MZ) = \Pr(C_1^X = C_2^X|DZ) = 1 \]
   - Variables that are not equal for both twins types \( (A_1^X, A_1^X) \):
     \[ \Pr(A_1^X = A_2^X|MZ) = 1, \Pr(A_1^X = A_2^X|DZ) \neq 1 \]

3. **Critics**: No Gene\( \times \)Environment Interaction \( (A_1^X, A_2^X) \perp \perp (C_1^X, C_2^X) \).
Figure 5: What about the Bivariate Case?

\[
e_{yx} e_{yx} = m z = d z
\]
Figure 6: Bivariate ACE Diagram

dz = 0.5 mz
mz = dz
dz = 0.5 mz
mz = dz
Basic Goal of Economics

- **Main Goal:** Evaluating the impact of $X$ on $Y$ instead of Variance Decomposition.
- **Question:** How Twin Data can help on identifying the causal impact of $X$ on $Y$?

Analysis based on the Bivariate ACE Diagram

- **Approach:** Add a causal link $X \rightarrow Y$ in the Bivariate ACE Model.
- **Problem:** Model not identified.
- **Solution:** Assume that all confounding effects are equally shared by MZ Twins (ACE-$\beta$ Model).
- **Estimation:** Linear Effect of $X$ on $Y$ can be estimated by Twin Fixed Effects using MZ Data.
Figure 7: Bivariate ACE Diagram
Figure 8: Bivariate ACE Diagram With Causal Link $X \rightarrow Y$
Figure 9: ACE-β Diagram

dz = 0.5 mz
mz = dz
dz = 0.5 mz
mz = dz

\[ e_{xx} \rightarrow X_1 \]
\[ a_{xx} \rightarrow A_1^X \]
\[ c_{xx} \rightarrow C_1^X \]
\[ a_{yx} \rightarrow A_1^Y \]
\[ c_{yx} \rightarrow C_1^Y \]
\[ e_{yy} \rightarrow E_1^Y \]
\[ e_{xx} \rightarrow X_2 \]
\[ a_{xx} \rightarrow A_2^X \]
\[ c_{xx} \rightarrow C_2^X \]
\[ a_{yx} \rightarrow A_2^Y \]
\[ c_{yx} \rightarrow C_2^Y \]
\[ e_{yy} \rightarrow E_2^Y \]
\[ \beta \}

\[ \text{dx} = 0.5 \text{mz} \]
\[ \text{mz} = \text{dz} \]
\[ \text{dz} = 0.5 \text{mz} \]
\[ \text{mz} = \text{dz} \]
Interpreting the ACE-β Model

- **Identification**: Strong assumption of no confounding effects under outcome difference for linear models.
- **Neglects**: the potential use of Property (2) of Twins data as an identification tool, only needs MZ to estimate β.
- **Generate an IV**: $(\Delta X) \perp\!\!\!\!\!\perp (A, C)$ but $\Delta X \not\perp X$.
- **Estimation**: Standard Twin Fixed Effects.
- **Pending Critics**: No Gene × Environment Interaction;
Figure 10: Non-parametric Monozygotic and Dizygotic ACE
Characteristics of the Nonparametric ACE Model

Setting the Problem

- **No IV**: MZ does not have IV properties.
- **Conditional Independence Relations**: Model generates 341 total conditional independence relations and 141 unique ones.
- **Additional Restrictions**: Model cannot be characterized solely as a DAG.
- **Four Additional Independence Relations**:
  \[ A_i^X \perp \perp \{ T \setminus \{ A_i^X, A_i^X \} \} \mid (A_j^X, MZ = 1); i, j \in \{1, 2\} \]
- **Equality in Distributions**:
  \[ X_1 =^d X_2, \quad Y_1 =^d Y_2, \quad A_1^X =^d A_2^X, \quad A_1^Y =^d A_2^Y, \quad E_1 =^d E_2. \]
Figure 11: Non-parametric Monozygotic and Dizygotic ACE
Characteristics of the Nonparametric MZ ACE Model
Setting the Problem

- **Conditional Independence Relations:** Model generates 32 unique conditional independence relations.

- **Four Additional Independence Relations:**

  \[ A_i^X \perp \{ T \setminus \{ A_j^X, A_i^X \} \} \mid (A_j^X, MZ = 1); i, j \in \{1, 2\} \]

- **Additional Distribution Equalities:**

  \[ X_1 =^d X_2, \quad Y_1 =^d Y_2, \quad E_1 =^d E_2, \]
Identification of the Intergenerational Elasticity Equation (IGE) Under Gene-Environment Interactions

- **IGE**: Often modeled as an AR(1):

\[
X = \beta Y + \epsilon
\]  

- **Problems:**
  - AR(1) is a simplistic approach of a complex process;
  - $\beta$ is not a causal effect of $X$ in $Y$,
  - but rather a summary of confounding factors that operate on the income of both parents and children.

- **Question**: How Twin Data can help identify the IGE process?
Figure 12: IGE AR(1) Diagram

\[ X_1 \rightarrow \beta \rightarrow Y_1 \leftarrow E_1^\gamma \]
Figure 13: Underlying IGE Process Diagram
Figure 14: Underlying IGE Process Diagram (Parental Twins)
Figure 15: ACE/IGE – Partitioning \( \lor \) (Twin Parents)
Figure 16: ACE/IGE Diagram (Twin Parents)
Problems with the ACE/IGE Model

• Equal Variances by Twin Type: $\text{Var}(X|MZ) = \text{Var}(X|DZ)$.
• Even Worse: The Model is not Identified!
• Solution: Generalize.
  • Twin Siblings can affect each other income.
  • Allow for cross fertilization due to twin’s genetic endowment.
Figure 17: ACE/IGE Diagram (Twin Parents)
Figure 18: Modified ACE/IGE Diagram (Twin Parents)
Figure 19: Bivariate ACE/IGE Diagram With GxE Interaction