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**

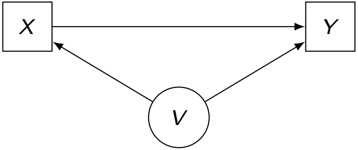
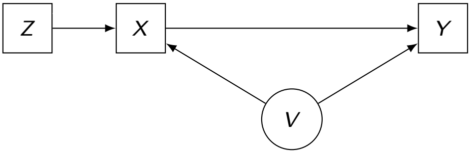


Figure 2: **Non-parametric IV Solution**



IV Properties

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

In Pinto(2013):

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:*

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

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

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

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

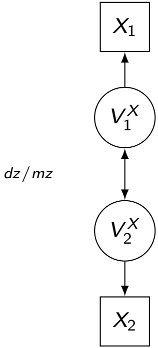
$$\text{there exist functions, } f_x : \text{supp}(Z) \to [0,1] \text{ and } g_x : \text{supp}(V) \to [0,1];\ \forall\, x \in \text{supp}(X)$$
$$\text{such that } \mathbf{P}_E\left(X = \sum_{x \in \text{supp}(X)} \mathbf{1}[f_x(Z) \le g_x(V)] \cdot x\right) = 1.$$

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);

THE UNIVERSITY OF CHICAGO

Figure 3: **Twin Identification – Two Properties**



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;

	Genetic Type		
Are Confounding Variables equal across twins?	DZ	MZ	Variables
1. Genetic Endowment	✗	✓	(A_1^X, A_2^X)
2. Shared Environment	✓	✓	(C_1^X, C_2^X)
3. Non-shared Environment	✗	✗	(E_1^X, E_2^X)
4. Exclusion Assumption	✓	✗	–

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:**

$$\Pr(C_1^X = C_2^X | MZ = 1) = \Pr(C_1^X = C_2^X | MZ = 0) = 1.$$

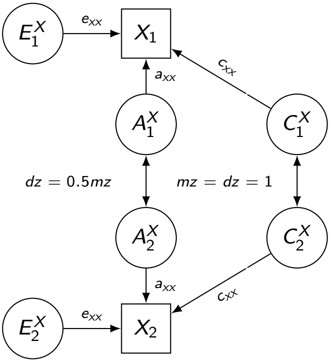
3. **Non-shared Environment (Assumption):**

$$E_1^X \perp\!\!\!\perp E_2^X | (MZ = 1), (E_1^X \perp\!\!\!\perp E_2^X | MZ = 0).$$

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**



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×Environment Interaction $(A_1^X, A_2^X) \perp\!\!\!\perp (C_1^X, C_2^X)$.

THE UNIVERSITY OF CHICAGO

Figure 5: **What about the Bivariate Case?**

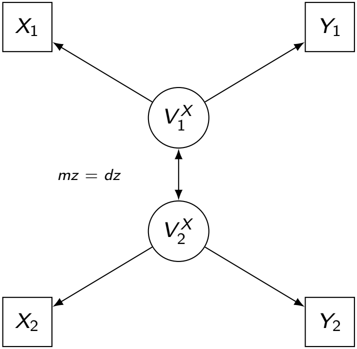
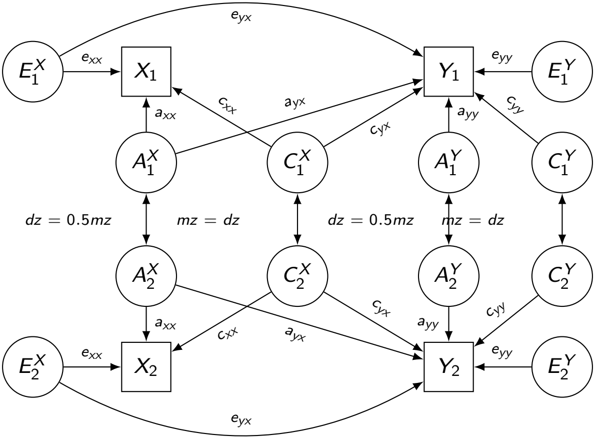


Figure 6: **Bivariate ACE Diagram**



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 $X \to Y$ in the Bivariate ACE Model.
- **Problem:** Model not identified.
- **Solution:** Assume that all confounding effects are equally shared by MZ Twins (ACE-β Model).
- **Estimation:** Linear Effect of X on Y can be estimated by Twin Fixed Effects using MZ Data.

Figure 7: **Bivariate ACE Diagram**

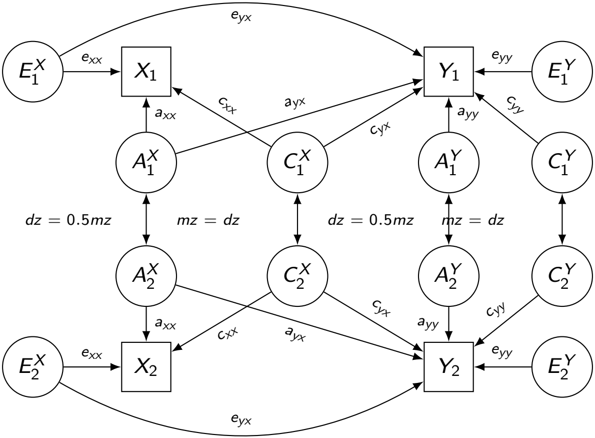


Figure 8: **Bivariate ACE Diagram With Causal Link $X \to Y$**

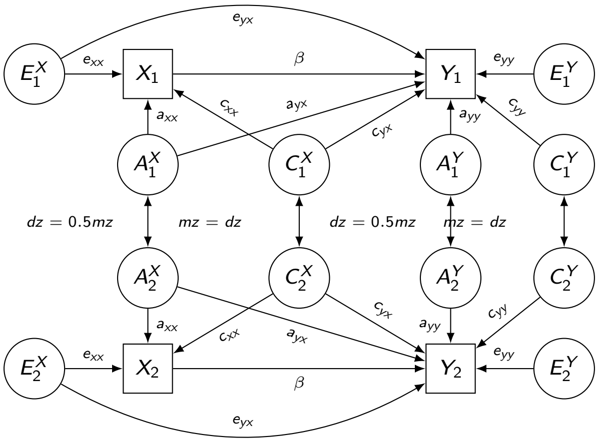
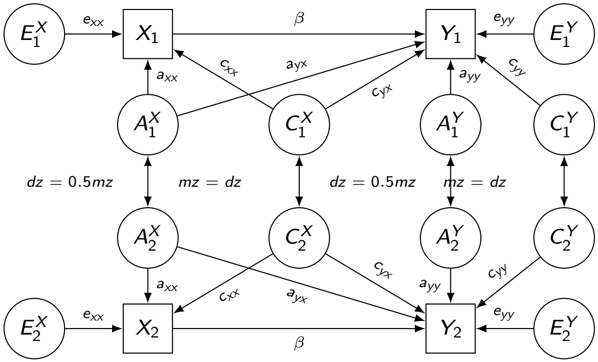


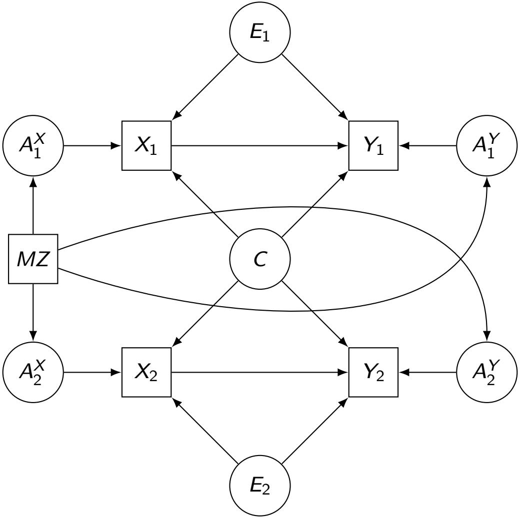
Figure 9: **ACE-β Diagram**



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 $\Delta X \not\perp\!\!\!\perp X$.
- **Estimation:** Standard Twin Fixed Effects.
- **Pending Critics:** No Gene$\times$Environment Interaction;

Figure 10: **Non-parametric Monozygotic and Dizygotic ACE**



Characteristics of the Nonparametric ACE Model
Setting the Problem

- **No IV:** MZ des 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:**

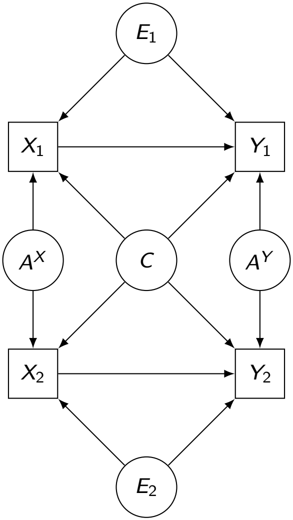
$$A_i^X \perp\!\!\!\perp \{\mathcal{T} \setminus \{A_i^X, A_i^X\}\} | (A_j^X, MZ = 1); i, j \in \{1, 2\}$$

- **Equality in Distributions:**

$$X_1 =^d X_2, \quad Y_1 =^d Y_2, \quad A_1^X =^d A_2^X, \quad A_1^Y =^d A_2^Y, \quad E_1 =^d E_2.$$

THE UNIVERSITY OF CHICAGO

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:**

$$A_i^X \perp\!\!\!\perp \{\mathcal{T} \setminus \{A_i^X, A_i^X\}\} | (A_j^X, MZ = 1); i, j \in \{1, 2\}$$

- **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-Envinonment Interactions

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

$$\underbrace{X}_{\text{Parental Income}} = \beta \underbrace{Y}_{\text{Descendants Income}} + \underbrace{\epsilon}_{\text{Exogenous Error Term}} \quad (1)$$

- **Problems:**
  - AR(1) is a simplistic approach of a complex process;
  - β 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**

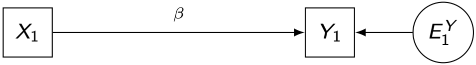


Figure 13: **Underlying IGE Process Diagram**

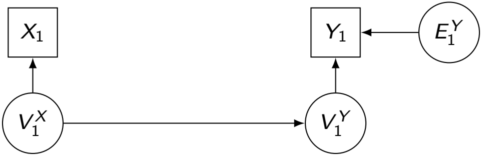


Figure 14: **Underlying IGE Process Diagram (Parental Twins)**

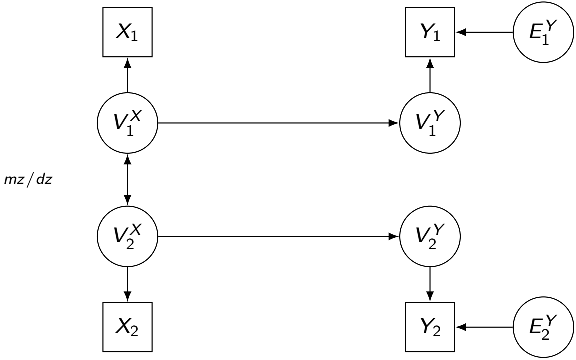


Figure 15: **ACE/IGE – Partitioning V (Twin Parents)**

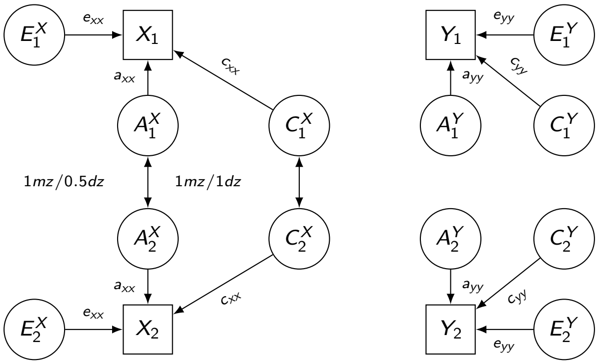
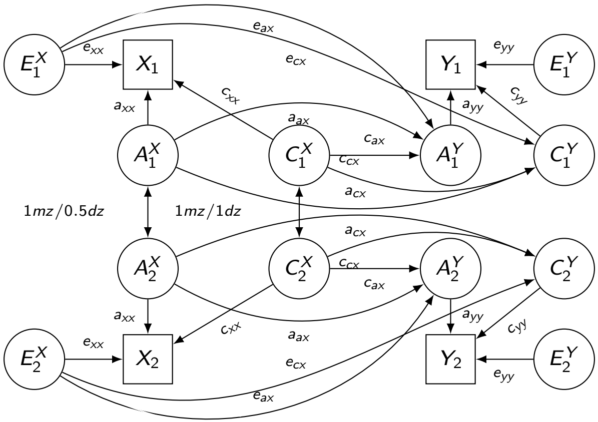


Figure 16: **ACE/IGE Diagram (Twin Parents)**



Problems with the ACE/IGE Model

- **Equal Variances by Twin Type:** $Var(X|MZ) = Var(X|DZ)$.
- **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)**

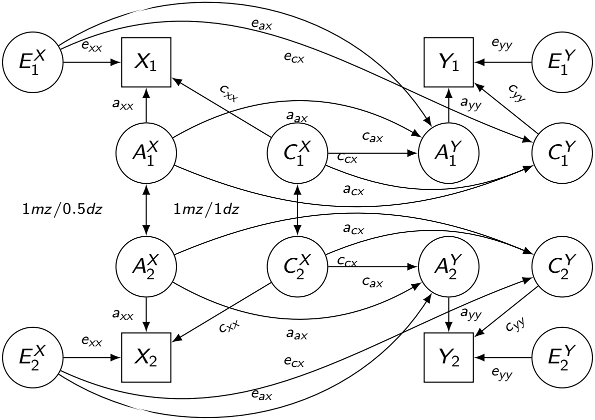


Figure 18: **Modified ACE/IGE Diagram (Twin Parents)**

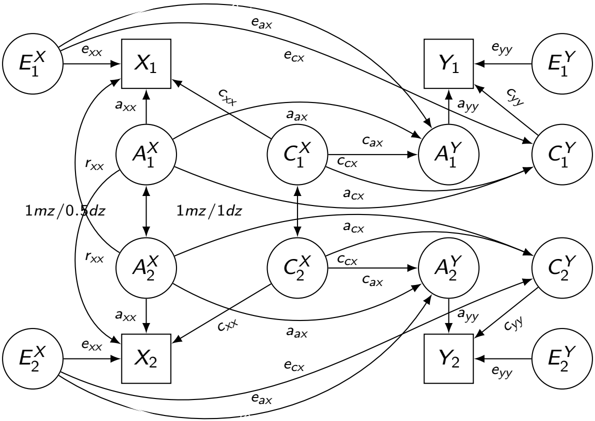


Figure 19: **Bivariate ACE/IGE Diagram With GxE Interaction**