

Modelamiento Estructural: Principios y Ejemplos

Ernesto San Martín

Facultad de Matemáticas, Pontificia Universidad Católica de Chile, Chile

The Economics School of Louvain, Université catholique de Louvain, Belgium

LIES Laboratorio Interdisciplinario de Estadística Social, Pontificia Universidad Católica de Chile, Chile

Seminario de Estadística Educacional

LIES, 2 de noviembre de 2017

Problema de Especificación

- Objetivo de los métodos estadísticos: reducir los datos para poder entenderlos.
- Población infinita de observaciones \iff Distribución de muestreo indexada por parámetros.
- Parámetros: representan características de las unidades bajo estudio.
- Formalmente, un modelo estadístico corresponde a

$$\mathcal{E} = \{(S, S), P^\theta : \theta \in \Theta\}.$$

- Es importante notar que

$$\Phi : \Theta \longrightarrow \mathcal{P}(S, S), \quad \theta \longmapsto P^\theta.$$

Propiedades de Φ .

Problema de Especificación

- Las ideas anteriores están contenidas en Fisher (1922).
- “Problems of Specification. These arise in the choice of the mathematical form of the population”.

- Koopmans and Reiersol (1950):

“In many fields the objective of the investigator’s inquisitiveness is not just a “population” in the sense of a distribution of observable variables, but a physical structure projected behind this distribution, by which the latter is thought to be generated. The word “physical” is used merely to convey that the structure concept is based on the investigator’s ideas as to the “explanation” or “formation” of the phenomena studied, briefly, on his theory of these phenomena, whether they are classified as physical in the literal sense, biological, psychological, sociological, economic or otherwise”.

- Como consecuencia de lo anterior (que conllevará una reformulación del Problema de Especificación), surge un nuevo problema: el Problema de identificación.
- Veamos el ejemplo propuesto por Koopmans y Reiersol: los parámetros de interés son α y β en la relación

$$\eta_1 = \alpha + \beta\eta_1,$$

donde η_1, η_2 son observadas con error, es decir,

$$y_i = \eta_i + u_i, \quad E(u_i) = 0 \quad i = 1, 2.$$

Problemas de Identificación

- Discuten el problema para el caso

$$\eta - 1 \perp\!\!\!\perp (u_1, u_2), \quad (u_1, u_2) \sim \mathcal{N}(\cdot, \cdot).$$

- En primer lugar, notemos que si η_1 es también distribuido normalmente entonces β no es identificado pues

$$\begin{pmatrix} \text{var}(y_1) & \text{cov}(y_1, y_2) \\ \text{cov}(y_1, y_2) & \text{var}(y_2) \end{pmatrix} = \begin{pmatrix} 1 & \beta \\ \beta & \beta^2 \end{pmatrix} \text{var}(\eta_1) + \begin{pmatrix} \text{var}(u_1) & \text{cov}(u_1, u_2) \\ \text{cov}(u_1, u_2) & \text{var}(u_2) \end{pmatrix}$$

Dado los observables, la igualdad anterior es válida para β 's distintos combinados con valores de $\text{var}(\eta_1)$ suficientemente pequeños.

- Se puede demostrar que β queda determinado por la distribución conjunta de (y_1, y_2) si esta distribución no es normal.

- Estructura $S = (F, \phi)$ es un par donde F es una distribución de probabilidad de los no-observables o no-observados, y ϕ permite relacionar estructuralmente los observables y los no-observados:

$$\phi(y, u) = 0.$$

- Por lo tanto, la estructura S determina la distribución de los observables $H(y | S)$.
- ¿Dada la distribución de los observables, es posible determinar una única estructura S que generó dichos observables?
- Problema de identificación como un problema pre-estadístico (pre-inferencial, pre-estimación)

- Koopmans y Reiersol insisten en la importancia de este problema:

“Scientific honesty demands that the specification of a model be based on prior knowledge of the phenomenon studied and possibly on criteria of simplicity, but not on the desire for identifiability of characteristics in which the researcher happens to be interested.”

- Estructura: representa la explicación sustantiva de un fenómeno.
- Identificabilidad: asegura que el modelo matemático/probabilístico corresponde a una explicación del fenómeno.

¿Qué es un Problema de Identificación?

● El Problema de Reflexión I:

- Supongamos que observamos los movimientos casi simultáneos de una persona y de su imagen en un espejo.
- ¿Es la imagen en el espejo la que causa los movimientos de la persona, o la imagen refleja los movimientos de la persona, o tanto la persona como la imagen se mueven conjuntamente en respuesta a un estímulo común externo?
- Las **solas observaciones empíricas** no son suficientes para responder estas preguntas, incluso si tenemos la posibilidad de observar innumerables ejemplos como el anterior.
- Para poder concluir **requerimos entender algo de óptica y del comportamiento humano**. De lo contrario, no es posible **distinguir empíricamente** entre las tres explicaciones anteriores.

¿Qué es un Problema de Identificación?

● El Problema de Reflexión I:

- Supongamos que observamos los movimientos casi simultáneos de una persona y de su imagen en un espejo.
- ¿Es la imagen en el espejo la que causa los movimientos de la persona, o la imagen refleja los movimientos de la persona, o tanto la persona como la imagen se mueven conjuntamente en respuesta a un estímulo común externo?
- Las **solas observaciones empíricas** no son suficientes para responder estas preguntas, incluso si tenemos la posibilidad de observar innumerables ejemplos como el anterior.
- Para poder concluir **requerimos entender algo de óptica y del comportamiento humano**. De lo contrario, no es posible **distinguir empíricamente** entre las tres explicaciones anteriores.

¿Qué es un Problema de Identificación?

● El Problema de Reflexión I:

- Supongamos que observamos los movimientos casi simultáneos de una persona y de su imagen en un espejo.
- ¿Es la imagen en el espejo la que causa los movimientos de la persona, o la imagen refleja los movimientos de la persona, o tanto la persona como la imagen se mueven conjuntamente en respuesta a un estímulo común externo?
- Las **solas observaciones empíricas** no son suficientes para responder estas preguntas, incluso si tenemos la posibilidad de observar innumerables ejemplos como el anterior.
- Para poder concluir **requerimos entender algo de óptica y del comportamiento humano**. De lo contrario, no es posible **distinguir empíricamente** entre las tres explicaciones anteriores.

¿Qué es un Problema de Identificación?

● El Problema de Reflexión I:

- Supongamos que observamos los movimientos casi simultáneos de una persona y de su imagen en un espejo.
- ¿Es la imagen en el espejo la que causa los movimientos de la persona, o la imagen refleja los movimientos de la persona, o tanto la persona como la imagen se mueven conjuntamente en respuesta a un estímulo común externo?
- Las **solas observaciones empíricas** no son suficientes para responder estas preguntas, incluso si tenemos la posibilidad de observar innumerables ejemplos como el anterior.
- Para poder concluir **requerimos entender algo de óptica y del comportamiento humano**. De lo contrario, no es posible **distinguir empíricamente** entre las tres explicaciones anteriores.

¿Qué es un Problema de Identificación?

● El Problema de Reflexión I:

- Supongamos que observamos los movimientos casi simultáneos de una persona y de su imagen en un espejo.
- ¿Es la imagen en el espejo la que causa los movimientos de la persona, o la imagen refleja los movimientos de la persona, o tanto la persona como la imagen se mueven conjuntamente en respuesta a un estímulo común externo?
- Las **solas observaciones empíricas** no son suficientes para responder estas preguntas, incluso si tenemos la posibilidad de observar innumerables ejemplos como el anterior.
- Para poder concluir **requerimos entender algo de óptica y del comportamiento humano**. De lo contrario, no es posible **distinguir empíricamente** entre las tres explicaciones anteriores.

¿Qué es un Problema de Identificación?

- **Galileo, Bellarmino:**

- En los *Escritos Copernicanos*, Galileo muestra lo que estaba en juego en aquella polémica:

Es verdad que no es lo mismo demostrar que con el movimiento de la Tierra y la inmovilidad del Sol se salvan mejor las apariencias (*σωζεῖν τὰ φαινόμενα*) que el demostrar que tales hipótesis son realmente verdaderas en la naturaleza; pero es igualmente correcto y más verdadero que con el otro sistema comúnmente aceptado no se puede dar razón de tales apariencias. Aquél es incuestionablemente falso, de la misma forma que es claro que éste, que se ajusta muy bien, **podría ser verdadero**.

- Hay un problema de identificación . . . y una preocupación por buscar verdades *in rerum natura*.

¿Qué es un Problema de Identificación?

- **Galileo, Bellarmino:**

- En los *Escritos Copernicanos*, Galileo muestra lo que estaba en juego en aquella polémica:

Es verdad que no es lo mismo demostrar que con el movimiento de la Tierra y la inmovilidad del Sol se salvan mejor las apariencias (*σωζεῖν τὰ φαινόμενα*) que el demostrar que tales hipótesis son realmente verdaderas en la naturaleza; pero es igualmente correcto y más verdadero que con el otro sistema comúnmente aceptado no se puede dar razón de tales apariencias. Aquél es incuestionablemente falso, de la misma forma que es claro que éste, que se ajusta muy bien, **podría ser verdadero**.

- Hay un problema de identificación . . . y una preocupación por buscar verdades *in rerum natura*.

¿Qué es un Problema de Identificación?

- **Galileo, Bellarmino:**

- En los *Escritos Copernicanos*, Galileo muestra lo que estaba en juego en aquella polémica:

Es verdad que no es lo mismo demostrar que con el movimiento de la Tierra y la inmovilidad del Sol se salvan mejor las apariencias (*σωζεῖν τὰ φαινόμενα*) que el demostrar que tales hipótesis son realmente verdaderas en la naturaleza; pero es igualmente correcto y más verdadero que con el otro sistema comúnmente aceptado no se puede dar razón de tales apariencias. Aquél es incuestionablemente falso, de la misma forma que es claro que éste, que se ajusta muy bien, **podría ser verdadero**.

- Hay un problema de identificación . . . y una preocupación por buscar verdades *in rerum natura*.

¿Qué es un Problema de Identificación?

- **El Problema de Reflexión II (Manski, 1993):**
 - ¿Cómo interpretar la observación común que los individuos que pertenecen a un mismo grupo tienden a comportarse de manera similar?
- Dos hipótesis explicativas se han propuesto:
 - **Efectos endógenos:** la tendencia de un individuo a comportarse de una cierta manera varía con el comportamiento predominante del grupo.
 - **Efectos correlacionados:** los individuos de un mismo grupo tienden a comportarse de manera similar porque enfrentan ambientes similares y tienen características similares.

¿Qué es un Problema de Identificación?

- **El Problema de Reflexión II** (Manski, 1993):
 - ¿Cómo interpretar la observación común que los individuos que pertenecen a un mismo grupo tienden a comportarse de manera similar?
- Dos hipótesis explicativas se han propuesto:
 - **Efectos endógenos:** la tendencia de un individuo a comportarse de una cierta manera varía con el comportamiento predominante del grupo.
 - **Efectos correlacionados:** los individuos de un mismo grupo tienden a comportarse de manera similar porque enfrentan ambientes similares y tienen características similares.

¿Qué es un Problema de Identificación?

- **El Problema de Reflexión II** (Manski, 1993):
 - ¿Cómo interpretar la observación común que los individuos que pertenecen a un mismo grupo tienden a comportarse de manera similar?
- Dos hipótesis explicativas se han propuesto:
 - **Efectos endógenos:** la tendencia de un individuo a comportarse de una cierta manera varía con el comportamiento predominante del grupo.
 - **Efectos correlacionados:** los individuos de un mismo grupo tienden a comportarse de manera similar porque enfrentan ambientes similares y tienen características similares.

¿Qué es un Problema de Identificación?

- **El Problema de Reflexión II** (Manski, 1993):
 - ¿Cómo interpretar la observación común que los individuos que pertenecen a un mismo grupo tienden a comportarse de manera similar?
- Dos hipótesis explicativas se han propuesto:
 - **Efectos endógenos:** la tendencia de un individuo a comportarse de una cierta manera varía con el comportamiento predominante del grupo.
 - **Efectos correlacionados:** los individuos de un mismo grupo tienden a comportarse de manera similar porque enfrentan ambientes similares y tienen características similares.

¿Qué es un Problema de Identificación?

- **El Problema de Reflexión II** (Manski, 1993):
 - ¿Cómo interpretar la observación común que los individuos que pertenecen a un mismo grupo tienden a comportarse de manera similar?
- Dos hipótesis explicativas se han propuesto:
 - **Efectos endógenos:** la tendencia de un individuo a comportarse de una cierta manera varía con el comportamiento predominante del grupo.
 - **Efectos correlacionados:** los individuos de un mismo grupo tienden a comportarse de manera similar porque enfrentan ambientes similares y tienen características similares.

¿Qué es un Problema de Identificación?

- Observaciones empíricas (que pueden incluir mediciones por medio de cuestionarios) del comportamiento de individuos en grupos **no pueden *per se* distinguir entre estas hipótesis.**
- Para extraer conclusiones se requiere que la evidencia empírica se combine con **fuertes hipótesis mantenidas** acerca de la naturaleza del comportamiento individual y de las interacciones sociales.
- Sin embargo, dichas **hipótesis mantenidas** pueden resultar implausibles . . . en cuyo caso **no es posible distinguir entre las dos hipótesis anteriores.**

¿Qué es un Problema de Identificación?

- Observaciones empíricas (que pueden incluir mediciones por medio de cuestionarios) del comportamiento de individuos en grupos **no pueden *per se* distinguir entre estas hipótesis.**
- Para extraer conclusiones se requiere que la evidencia empírica se combine con **fuertes hipótesis mantenidas** acerca de la naturaleza del comportamiento individual y de las interacciones sociales.
- Sin embargo, dichas **hipótesis mantenidas** pueden resultar implausibles . . . en cuyo caso **no es posible distinguir entre las dos hipótesis anteriores.**

¿Qué es un Problema de Identificación?

- Observaciones empíricas (que pueden incluir mediciones por medio de cuestionarios) del comportamiento de individuos en grupos **no pueden *per se* distinguir entre estas hipótesis.**
- Para extraer conclusiones se requiere que la evidencia empírica se combine con **fuertes hipótesis mantenidas** acerca de la naturaleza del comportamiento individual y de las interacciones sociales.
- Sin embargo, dichas **hipótesis mantenidas** pueden resultar implausibles . . . en cuyo caso **no es posible distinguir entre las dos hipótesis anteriores.**

¿Qué es un Problema de Identificación?

- Observaciones empíricas (que pueden incluir mediciones por medio de cuestionarios) del comportamiento de individuos en grupos **no pueden *per se* distinguir entre estas hipótesis.**
- Para extraer conclusiones se requiere que la evidencia empírica se combine con **fuertes hipótesis mantenidas** acerca de la naturaleza del comportamiento individual y de las interacciones sociales.
- Sin embargo, dichas **hipótesis mantenidas** pueden resultar implausibles . . . en cuyo caso **no es posible distinguir entre las dos hipótesis anteriores.**

¿Qué es un Problema de Identificación?

- **Distinguir** entre las dos hipótesis anteriores es relevante pues tienen diferentes implicancias para la política pública.
- Por ejemplo, entender cómo los estudiantes interactúan en una sala de clase es crítico para la evaluación de muchos aspectos de la política educacional tales como *classroom effect size*, programas de interacción racial o social, etc.

¿Qué es un Problema de Identificación?

- **Distinguir** entre las dos hipótesis anteriores es relevante pues tienen diferentes implicancias para la política pública.
- Por ejemplo, entender cómo los estudiantes interactúan en una sala de clase es crítico para la evaluación de muchos aspectos de la política educacional tales como *classroom effect size*, programas de interacción racial o social, etc.

¿Qué es un Problema de Identificación?

- Si no somos incapaces de interpretar los patrones observados, podemos buscar la *opinión experta* de dos científicas sociales:
 - Uno, tal vez un sociólogo, asegura que la presión por conformarse a las normas del grupo hace que los individuos en un grupo tiendan a comportarse de manera similar.
 - Otro, un economista, asevera que las personas con características similares escogen asociarse con otros similares.

¿Qué es un Problema de Identificación?

- Si no somos incapaces de interpretar los patrones observados, podemos buscar la *opinión experta* de dos científicas sociales:
 - Uno, tal vez un sociólogo, asegura que la presión por conformarse a las normas del grupo hace que los individuos en un grupo tiendan a comportarse de manera similar.
 - Otro, un economista, asevera que las personas con características similares escogen asociarse con otros similares.

¿Qué es un Problema de Identificación?

- Si no somos incapaces de interpretar los patrones observados, podemos buscar la *opinión experta* de dos científicas sociales:
 - Uno, tal vez un sociólogo, asegura que la presión por conformarse a las normas del grupo hace que los individuos en un grupo tiendan a comportarse de manera similar.
 - Otro, un economista, asevera que las personas con características similares escogen asociarse con otros similares.

¿Qué es un Problema de Identificación?

- **Ambas** afirmaciones son consistentes con la evidencia empírica.
- Los datos por sí solos **no pueden decirnos cuál de las dos aseveraciones anteriores es correcta.**
- Quizás ambas lo son ... **este es un problema de identificación.**

¿Qué es un Problema de Identificación?

- **Ambas** afirmaciones son consistentes con la evidencia empírica.
- Los datos por sí solos **no pueden decirnos cuál de las dos aseveraciones anteriores es correcta.**
- Quizás ambas lo son ... **este es un problema de identificación.**

¿Qué es un Problema de Identificación?

- **Ambas** afirmaciones son consistentes con la evidencia empírica.
- Los datos por sí solos **no pueden decirnos cuál de las dos aseveraciones anteriores es correcta.**
- Quizás ambas lo son ... **este es un problema de identificación.**

¿Qué es un Problema de Identificación?

- **Ambas** afirmaciones son consistentes con la evidencia empírica.
- Los datos por sí solos **no pueden decirnos cuál de las dos aseveraciones anteriores es correcta.**
- Quizás ambas lo son . . . **este es un problema de identificación.**

¿Qué es un Problema de Identificación?

- Consideremos las siguientes variables:
 - Y : logro de un joven estudiante al fin de segundo año medio.
 - X : atributos que caracterizan el grupo de referencia (por ejemplo, el grupo-curso de un joven estudiante).
 - Z : atributos que afectan directamente Y (por ejemplo, nivel socio-económico y habilidad).

¿Qué es un Problema de Identificación?

- Consideremos las siguientes variables:
 - Y : logro de un joven estudiante al fin de segundo año medio.
 - X : atributos que caracterizan el grupo de referencia (por ejemplo, el grupo-curso de un joven estudiante).
 - Z : atributos que afectan directamente Y (por ejemplo, nivel socio-económico y habilidad).

¿Qué es un Problema de Identificación?

- Un investigador observa (Y, X, Z) y asume que

$$E(Y | X, Z) = \alpha + \beta E(Y | X) + E(Z | X)' \gamma + X' \delta + Z' \eta.$$

- Los siguientes efectos están incluidos en este modelo:
 - Si $\beta \neq 0$, entonces se incluye un **efecto endógeno**.
 - Si $\gamma \neq 0$, entonces se incluye un **efecto exógeno**.
 - Si $\delta \neq 0$, entonces se incluye un **efecto correlacional**.
 - Si $\eta \neq 0$, entonces se incluye un **efecto directo**.

¿Qué es un Problema de Identificación?

- Un investigador observa (Y, X, Z) y asume que

$$E(Y | X, Z) = \alpha + \beta E(Y | X) + E(Z | X)' \gamma + X' \delta + Z' \eta.$$

- Los siguientes efectos están incluidos en este modelo:
 - Si $\beta \neq 0$, entonces se incluye un **efecto endógeno**.
 - Si $\gamma \neq 0$, entonces se incluye un **efecto exógeno**.
 - Si $\delta \neq 0$, entonces se incluye un **efecto correlacional**.
 - Si $\eta \neq 0$, entonces se incluye un **efecto directo**.

¿Qué es un Problema de Identificación?

- Un investigador observa (Y, X, Z) y asume que

$$E(Y | X, Z) = \alpha + \beta E(Y | X) + E(Z | X)' \gamma + X' \delta + Z' \eta.$$

- Los siguientes efectos están incluidos en este modelo:
 - Si $\beta \neq 0$, entonces se incluye un **efecto endógeno**.
 - Si $\gamma \neq 0$, entonces se incluye un **efecto exógeno**.
 - Si $\delta \neq 0$, entonces se incluye un **efecto correlacional**.
 - Si $\eta \neq 0$, entonces se incluye un **efecto directo**.

¿Qué es un Problema de Identificación?

- Un investigador observa (Y, X, Z) y asume que

$$E(Y | X, Z) = \alpha + \beta E(Y | X) + E(Z | X)' \gamma + X' \delta + Z' \eta.$$

- Los siguientes efectos están incluidos en este modelo:
 - Si $\beta \neq 0$, entonces se incluye un **efecto endógeno**.
 - Si $\gamma \neq 0$, entonces se incluye un **efecto exógeno**.
 - Si $\delta \neq 0$, entonces se incluye un **efecto correlacional**.
 - Si $\eta \neq 0$, entonces se incluye un **efecto directo**.

¿Qué es un Problema de Identificación?

- Un investigador observa (Y, X, Z) y asume que

$$E(Y | X, Z) = \alpha + \beta E(Y | X) + E(Z | X)' \gamma + X' \delta + Z' \eta.$$

- Los siguientes efectos están incluidos en este modelo:
 - Si $\beta \neq 0$, entonces se incluye un **efecto endógeno**.
 - Si $\gamma \neq 0$, entonces se incluye un **efecto exógeno**.
 - Si $\delta \neq 0$, entonces se incluye un **efecto correlacional**.
 - Si $\eta \neq 0$, entonces se incluye un **efecto directo**.

¿Qué es un Problema de Identificación?

- Un investigador observa (Y, X, Z) y asume que

$$E(Y | X, Z) = \alpha + \beta E(Y | X) + E(Z | X)' \gamma + X' \delta + Z' \eta.$$

- Los siguientes efectos están incluidos en este modelo:
 - Si $\beta \neq 0$, entonces se incluye un **efecto endógeno**.
 - Si $\gamma \neq 0$, entonces se incluye un **efecto exógeno**.
 - Si $\delta \neq 0$, entonces se incluye un **efecto correlacional**.
 - Si $\eta \neq 0$, entonces se incluye un **efecto directo**.

¿Qué es un Problema de Identificación?

- El **problema de reflexión** surge por la presencia de $E(Y | X)$ como un regresor o factor explicativo.
- Se puede probar que, si $\beta \neq 1$ y los regresores $[1, E(Z | X), X, Z]$ son linealmente independientes, las siguientes funciones de los parámetros son identificadas:

$$\frac{\gamma + \beta\eta}{1 - \beta}, \quad \frac{\delta}{1 - \beta}, \quad \eta.$$

- Pero esta identificación no nos permite **distinguir entre los dos efectos sociales (endógeno y de correlación)**, sino sólo aseverar que **al menos uno de dichos efectos está presente**.
- Si $\frac{\gamma + \beta\eta}{1 - \beta} \neq 0$ entonces $\beta\eta \neq 0$ o $\gamma \neq 0$.

¿Qué es un Problema de Identificación?

- El **problema de reflexión** surge por la presencia de $E(Y | X)$ como un regresor o factor explicativo.
- Se puede probar que, si $\beta \neq 1$ y los regresores $[1, E(Z | X), X, Z]$ son linealmente independientes, las siguientes funciones de los parámetros son identificadas:

$$\frac{\gamma + \beta\eta}{1 - \beta}, \quad \frac{\delta}{1 - \beta}, \quad \eta.$$

- Pero esta identificación no nos permite **distinguir entre los dos efectos sociales (endógeno y de correlación)**, sino sólo aseverar que **al menos uno de dichos efectos está presente**.
- Si $\frac{\gamma + \beta\eta}{1 - \beta} \neq 0$ entonces $\beta\eta \neq 0$ o $\gamma \neq 0$.

¿Qué es un Problema de Identificación?

- El **problema de reflexión** surge por la presencia de $E(Y | X)$ como un regresor o factor explicativo.
- Se puede probar que, si $\beta \neq 1$ y los regresores $[1, E(Z | X), X, Z]$ son linealmente independientes, las siguientes funciones de los parámetros son identificadas:

$$\frac{\gamma + \beta\eta}{1 - \beta}, \quad \frac{\delta}{1 - \beta}, \quad \eta.$$

- Pero esta identificación no nos permite **distinguir entre los dos efectos sociales (endógeno y de correlación)**, sino sólo aseverar que **al menos uno de dichos efectos está presente**.
- Si $\frac{\gamma + \beta\eta}{1 - \beta} \neq 0$ entonces $\beta\eta \neq 0$ o $\gamma \neq 0$.

¿Qué es un Problema de Identificación?

- El **problema de reflexión** surge por la presencia de $E(Y | X)$ como un regresor o factor explicativo.
- Se puede probar que, si $\beta \neq 1$ y los regresores $[1, E(Z | X), X, Z]$ son linealmente independientes, las siguientes funciones de los parámetros son identificadas:

$$\frac{\gamma + \beta\eta}{1 - \beta}, \quad \frac{\delta}{1 - \beta}, \quad \eta.$$

- Pero esta identificación no nos permite **distinguir entre los dos efectos sociales (endógeno y de correlación)**, sino sólo aseverar que **al menos uno de dichos efectos está presente**.
- Si $\frac{\gamma + \beta\eta}{1 - \beta} \neq 0$ entonces $\beta\eta \neq 0$ o $\gamma \neq 0$.

¿Qué es un Problema de Identificación?

- Estudios empíricos de efectos endógenos típicamente asumen que $\gamma = \delta = 0$, de manera que **excluyen** los efectos correlacionales y los efectos exógenos!
- Por lo tanto, la regresión inicial se reduce a

$$E(Y | X, Z) = \frac{\alpha}{1 - \beta} + E(Z | X)' \left[\frac{\beta\eta}{1 - \beta} \right] + Z'\eta.$$

- Más aún, si los regresores $[1, E(Z | X), Z]$ son linealmente independiente, los parámetros $\alpha/(1 - \beta)$, $\beta\eta/(1 - \beta)$ y η son identificados.
- Además, si $\eta = 0$ o $E(Z | X)$ es linealmente dependiente de $[1, Z]$, entonces el parámetro del efecto endógeno β no es identificado.
- Esta última condición permite entender un **efecto endógeno** en contraposición a un **efecto directo**: en ausencia de este último, el primero **no es identificado**.

¿Qué es un Problema de Identificación?

- Estudios empíricos de efectos endógenos típicamente asumen que $\gamma = \delta = 0$, de manera que **excluyen** los efectos correlacionales y los efectos exógenos!
- Por lo tanto, la regresión inicial se reduce a

$$E(Y | X, Z) = \frac{\alpha}{1 - \beta} + E(Z | X)' \left[\frac{\beta\eta}{1 - \beta} \right] + Z'\eta.$$

- Más aún, si los regresores $[1, E(Z | X), Z]$ son linealmente independiente, los parámetros $\alpha/(1 - \beta)$, $\beta\eta/(1 - \beta)$ y η son identificados.
- Además, si $\eta = 0$ o $E(Z | X)$ es linealmente dependiente de $[1, Z]$, entonces el parámetro del efecto endógeno β no es identificado.
- Esta última condición permite entender un **efecto endógeno** en contraposición a un **efecto directo**: en ausencia de este último, el primero **no es identificado**.

¿Qué es un Problema de Identificación?

- Estudios empíricos de efectos endógenos típicamente asumen que $\gamma = \delta = 0$, de manera que **excluyen** los efectos correlacionales y los efectos exógenos!
- Por lo tanto, la regresión inicial se reduce a

$$E(Y | X, Z) = \frac{\alpha}{1 - \beta} + E(Z | X)' \left[\frac{\beta\eta}{1 - \beta} \right] + Z'\eta.$$

- Más aún, si los regresores $[1, E(Z | X), Z]$ son linealmente independiente, los parámetros $\alpha/(1 - \beta)$, $\beta\eta/(1 - \beta)$ y η son identificados.
- Además, si $\eta = 0$ o $E(Z | X)$ es linealmente dependiente de $[1, Z]$, entonces el parámetro del efecto endógeno β no es identificado.
- Esta última condición permite entender un **efecto endógeno** en contraposición a un **efecto directo**: en ausencia de este último, el primero **no es identificado**.

¿Qué es un Problema de Identificación?

- Estudios empíricos de efectos endógenos típicamente asumen que $\gamma = \delta = 0$, de manera que **excluyen** los efectos correlacionales y los efectos exógenos!
- Por lo tanto, la regresión inicial se reduce a

$$E(Y | X, Z) = \frac{\alpha}{1 - \beta} + E(Z | X)' \left[\frac{\beta\eta}{1 - \beta} \right] + Z'\eta.$$

- Más aún, si los regresores $[1, E(Z | X), Z]$ son linealmente independiente, los parámetros $\alpha/(1 - \beta)$, $\beta\eta/(1 - \beta)$ y η son identificados.
- Además, si $\eta = 0$ o $E(Z | X)$ es linealmente dependiente de $[1, Z]$, entonces el parámetro del efecto endógeno β no es identificado.
- Esta última condición permite entender un **efecto endógeno** en contraposición a un **efecto directo**: en ausencia de este último, el primero **no es identificado**.

¿Qué es un Problema de Identificación?

- **Equivalencia observacional:** cuando las observaciones no nos permiten distinguir entre diferentes parametrizaciones.
- Una parametrización no identificada **no tiene sentido empírico.**
- Estimar parámetros no-identificados no tiene, por tanto, sentido pues **no es posible interpretarlos a la luz de los mismos datos.**
- El problema de identificación es, por tanto, una parte esencial de la construcción de los modelos estadísticos y, en consecuencia, **fija los límites de la inferencia (estadística).**

¿Qué es un Problema de Identificación?

- **Equivalencia observacional:** cuando las observaciones no nos permiten distinguir entre diferentes parametrizaciones.
- Una parametrización no identificada **no tiene sentido empírico**.
- Estimar parámetros no-identificados no tiene, por tanto, sentido pues **no es posible interpretarlos a la luz de los mismos datos**.
- El problema de identificación es, por tanto, una parte esencial de la construcción de los modelos estadísticos y, en consecuencia, **fija los límites de la inferencia (estadística)**.

¿Qué es un Problema de Identificación?

- **Equivalencia observacional:** cuando las observaciones no nos permiten distinguir entre diferentes parametrizaciones.
- Una parametrización no identificada **no tiene sentido empírico**.
- Estimar parámetros no-identificados no tiene, por tanto, sentido pues **no es posible interpretarlos a la luz de los mismos datos**.
- El problema de identificación es, por tanto, una parte esencial de la construcción de los modelos estadísticos y, en consecuencia, **fija los límites de la inferencia (estadística)**.

- Variación versus regularidad.
- Tradición de Hume (1748):
 - Imposibilidad de proveer fundamentos racionales para la existencia de objetos, del espacio y de las relaciones causales. Cf. Pearson, *The Grammar of the Science* (1892, capítulo V).
 - Ideas simples son combinadas para formar ideas complejas; cf. Locke y Toland.
 - Ideas complejas: semejanzas, contiguidad en el espacio y en el tiempo, causalidad.
 - Causalidad: establece una relación entre experiencias pasadas y presentes con evento que precedimos o explicamos. De ahí que todos los razonamientos relacoinados con *matter of facts* parecen fundarse sobre la relación causa-efecto.

- Conexiones causales son por tanto parte de un principio de asociación que opera en nuestra mente.
- Sucesiones regulares de de impresiones son seguidas por sucesiones regulares de ideas simples, luego la imaginación las ordena y conceptualiza sucesiones de ideas simples en ideas complejas. Así se da nacimiento a las relaciones causales.
- Pero sucesiones regulares establecidas por experiencia están faltas de la necesidad lógica que se requiere para las relaciones causales.
- Causalidad a la Hume: hábito psicológico que testimonia efectos que son regularmente seguidos por causas en el espacio y en el tiempo.

- Modelos causales modelan relaciones causales por medio del análisis de variaciones adecuadas entre variables de interés.
- Modelos causales son gobernados por el rationale de variación, no de regularidad.
- Variación: probabilidades condicionales:

$$P(E | C) \neq P(E).$$

- La variación en C causa variación en E si $P(E | C) \neq P(E)$.

- Provee un contexto causal relevante para la formulación de las hipótesis de causalidad.
- Guía la elección de las variables y de las relaciones que serán testeadas para la estabilidad estructural.
- Constituye la caja de resonancia para los resultados pues estos últimos deben ser coherentes con el conocimiento sustantivo.

- Partimos de un modelo marginal:

$$\{p_X(x | \omega) : \omega \in \Omega\}.$$

- Particionamos X como (Y, Z) :

$$p_X(x | \omega) = p_Z(z | \phi)p_{Y|Z}(y | z, \theta), \quad \omega = (\phi, \theta).$$

- El aspecto esencial es descomponer el DGP que genera x en dos componentes más simples y que permitan una secuencialidad en la generación de los datos.
- Por lo tanto, el modelo marginal y el modelo condicional deben variar libremente:

$$\{p_X(x | \omega) = p_Z(z | \phi)p_{Y|Z}(y | z, \theta), \quad \omega = (\phi, \theta) \in \Omega = \Theta \times \Phi\}.$$

- Esto define Z como exógena con respecto a ϕ y por lo tanto la descomposición marginal-condicional tiene una direccionalidad: luego Z es causa de Y .

- Descomposición recursiva \longleftrightarrow Explicación \longleftrightarrow Causalidad.
- Desafío: que la descomposición marginal-condicional represente un mecanismo causal establecido teóricamente.

- Consideremos tres variables:
 - Tabaquismo T .
 - Cáncer al sistema respiratorio C .
 - Exposición al abesto A .
 - Estatus socio-económico S .
- Nos interesa explicar el fenómeno generado por $P(T, C, A, S)$.
- ¿Qué sabemos? En ciertos lugares de Europa, se observa que:
- Aquellos de bajo S tienden a fumar más y a trabajar en medio-ambientes insalubres y pro tanto están más expuestos a A .
- Luego S es una variable causal clave que hay que incluir en un modelo de prácticas.

Ejemplo

- Es decir,

$$A \perp\!\!\!\perp T \mid S.$$

- Además, estudios clínicos muestran que A y T producen C para poblaciones diferentes. Luego

$$C \perp\!\!\!\perp SES \mid A, T.$$

- Luego, la descomposición recursiva se educa a

$$\begin{aligned} p(S, A, T, C) &= P(C \mid A, T, S) P(A, T \mid S) P(S) \\ &= P(C \mid A, T) P(A \mid S) P(T \mid S) P(S). \end{aligned}$$

- Aquí, $P(C \mid A, T)$ representa el mecanismo biológico, y $P(A \mid S)P(T \mid S)$ representa el mecanismo social.