

---

# “PATH” ANALISIS, MODELOS DE ECUACIONES ESTRUCTURALES Y VARIABLES NO OBSERVADAS

Francisco Alvira Martín y Marina Peña Fernández

---

Las variables utilizadas en un diseño de investigación pueden clasificarse en cuatro grupos (L. Kisch, 1975):

1. *Variables explicativas*: Aquellas variables cuyas relaciones se quieren investigar. Abarcan tanto a las variables propiamente explicativas como a las variables explicadas.
2. *Variables controladas*: Variables que están auténticamente controladas y variables que no interfieren en las interrelaciones entre las variables explicativas del primer grupo.
3. *Variables no controladas* con relaciones funcionales con la(s) variable(s) independiente(s), pero no con la(s) variable(s) dependiente (s).
4. *Variables no controladas* que guardan relaciones funcionales tanto con la(s) variable(s) independiente(s) como con la(s) variable(s) dependiente(s) (1).

---

Uno de los objetivos esenciales de todo diseño de investigación es lograr que el grupo 4 sea un conjunto vacío y, en cualquier caso, que sea máximo

<sup>1</sup> El término “variable” se utiliza como contraposición a “constante” y no como contraposición a “atributo”; por otra parte, lo que se llaman variable(s) independiente (s) abarca tanto a variables exógenas como a variables predeterminadas.

---

el número de variables del grupo 2. En los diseños experimentales se logra que desaparezcan las variables del grupo 4, a la vez que las variables del grupo 3, mediante la aleatorización de los sujetos sobre los tratamientos. En los diseños no experimentales, la aleatorización no es posible por definición. Es necesario aceptar supuestos simplificadores para rechazar las explicaciones alternativas. La plausibilidad teórico-empírica de estos supuestos dependerá de cada diseño en concreto.

Esta visión clásica de la dicotomía experimentación/no experimentación ha sido puesta en entredicho, tanto desde la experimentación como desde la no-experimentación.

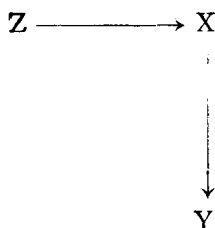
*Desde la experimentación*, las investigaciones de Orne (1962) y Rosenthal (1963) sobre los efectos del experimentador y la reactividad de los sujetos experimentales ponen en cuestión la no necesidad de supuestos simplificativos en experimentación.

*Desde la no experimentación*, A. D. Miller (1971) ha puesto de relieve la similitud de la lógica del experimento y de la lógica de los diseños no experimentales.

El experimento más simple puede representarse utilizando un esquema de tres variables en la forma en que se presenta en el gráfico 1.

### GRAFICO 1

*Representación esquemática de un experimento con una variable independiente y una dependiente*



Z es el tratamiento manipulado por el experimentador; X es el constructo teórico que se supone causa de Y, e Y, la variable dependiente<sup>2</sup>.

<sup>2</sup> Esta es la interpretación que presenta Miller; sin embargo, parece más lógico pensar que "Z" fuera el constructo teórico (variable independiente) y "X" el indicador de "Z" (en este caso el tratamiento operativo empleado por el experimentador). En cualquier caso la lógica general y los problemas que se plantean no se ven alterados por esta diferencia. Seguiré, por tanto, la presentación de Miller.

Tal y como están representadas las relaciones en el esquema, se destacan muy claramente los dos supuestos básicos de la lógica experimental:

- a) que Z no es causa de Y, sino sólo a través de X (Z no es causa directa de Y);
- b) que Z no guarda relación alguna con ninguna de las causas no medidas de Y que no hagan sentir su efecto sobre ésta exclusivamente a través de X.

La validez de los dos supuestos es independiente de la existencia o no de otras variables relacionadas con (causa de) X e Y, siempre que éstas no estén relacionadas con Z.

El supuesto *b)* tiene su basamento empírico en la asignación aleatoria de los sujetos experimentales a los tratamientos. A su vez, esta asignación aleatoria es sólo posible porque el experimentador manipula los tratamientos; es decir, controla qué, cuándo, cómo, dónde y sobre quién recaerán. Cada sujeto experimental recibe, por tanto, un valor de Z no predeterminado con anterioridad, dada la aleatorización. De aquí se sigue que estos valores de Z no dependerán de otras causas no medidas de Y excepto por azar.

El *a)*, sin embargo, es un supuesto que carece de contrastación empírica directa. Se da por supuesto a no ser que haya dos tratamientos,  $Z_1$  y  $Z_2$ . Caso que los haya, es posible intentar refutar que bien  $Z_1$ , bien  $Z_2$ , influyen directamente en Y. (Véase A. D. Miller, 1971, págs. 280 y sigs.)<sup>3</sup>.

En la investigación no experimental, la ausencia de aleatorización y manipulación lleva consigo que:

1. La hipótesis de que X causa Y se debilita, puesto que no hay un orden temporal entre las variables.
2. La no relación entre Z y otras causas no medidas de Y no puede darse por desecontada sin más. Deben apuntarse datos que corroboren la plausibilidad de este supuesto.

¿Significa esto la imposibilidad de investigar relaciones causales por métodos no experimentales? La respuesta, obviamente, es no. Lo que significa

<sup>3</sup> Miller sugiere para ello la comprobación de que estimando la ecuación

$$(y - bx) = f + gZ_1 + hZ_2$$

por el método de mínimos cuadrados, se obtenga  $g = h = 0$ . El coeficiente b es el correspondiente coeficiente de regresión de Y en X pero utilizando no los valores empíricos reales de esta última variable, sino los valores predichos partiendo de  $Z_1$  y  $Z_2$ , o sea X. Tal como está estimado b, g y h, tienen que ser iguales a 0 si realmente no hay influencia directa de  $Z_1$  y  $Z_2$  en Y.

es que la investigación no experimental controlará en mucha menor medida posibles explicaciones alternativas y estará más dependiente de la teoría sustantiva correspondiente. Pero la plausibilidad de los supuestos simplificativos es una cuestión de grado y no absoluta.

«... Por muy elaborado que sea un diseño, siempre hay que llevar a cabo ciertos supuestos simplificativos. En especial, y en algún momento, hay que presuponer que los efectos de las variables no controladas (tipo 4) son pequeños. La aleatorización ayuda a dejar fuera parte de estas variables, pero la plausibilidad de este tipo específico de supuesto simplificador es siempre cuestión de grado. Quiero subrayar este hecho para hacer hincapié en la semejanza subyacente entre la lógica para realizar inferencias causales desde diseños experimentales y desde diseños no experimentales» (Blalock, 1964, pág. 26).

#### «PATH» ANALISIS Y MODELOS RECURSIVOS

La investigación no experimental puede aproximarse a la experimental partiendo de modelos causales en los que se introduzcan las variables necesarias para refutar aquellas relaciones entre variables no derivadas de la teoría. La teoría señala también el orden temporal de las variables, aunque en este caso el empleo de diseños longitudinales, así como el empleo de variables desfasadas en el tiempo (*lagged*), ayuda a controlar alternativas explicativas de tipo temporal.

La publicación del libro de Blalock *Causal Inferences in non-experimental Research* marca un hito en el camino de mejorar la utilización de los diseños no experimentales. Las ideas de Blalock, junto con el artículo de Duncan en 1964 sobre el *path* análisis, replantean el tema de la causalidad con datos no experimentales en sociología, retomando la polémica desde la biología, la bioestadística y la econometría.

El modelo recursivo causal incluye todas las posibles relaciones entre

\* Como pone de relieve A. S. Goldberger (1971) en el modelo clásico, un conjunto de indicadores ( $Y$ ) están determinados linealmente por un conjunto de variables no observadas ( $X$ ) que son los factores comunes y un conjunto de términos de error no observables ( $V$ ). De forma matricial,

$$Y = BX + V$$

y la matriz de varianzas/covarianza

$$\Sigma = E(Y Y') = \beta \Phi \beta' + \theta$$

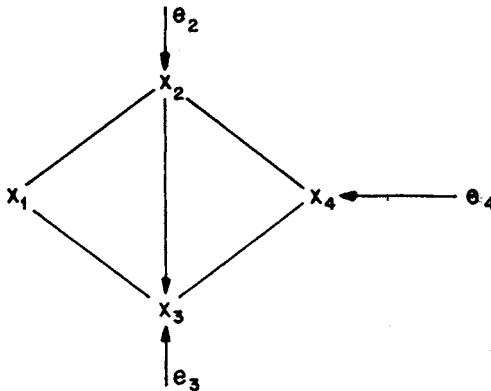
donde  $E(X X') = \Phi$        $E(X V') = 0$

$$E(V V') = \theta \text{ diagonal.}$$

las variables del modelo, de forma que la relación entre cada par de variables sea unidireccional, es decir, causal.

Si nos centramos en un modelo de cuatro variables,  $X_1$ ,  $X_2$ ,  $X_3$  y  $X_4$ , de las cuales  $X_1$  es una variable exógena (no  $X$  determinada por el modelo) y el resto variables endógenas (determinadas por el modelo), el sistema de relaciones causales existentes entre las variables se expresa gráfica y analíticamente.

*Gráficamente:* El sistema de relaciones causales se presenta a través de un diagrama en el cual la influencia causal entre las variables intervinientes en el modelo se expresa mediante flechas unidireccionales. La causalidad no explicada por el modelo (variable error o residuos) debe también estar representada por una flecha unidireccional.



*Analíticamente:* El sistema de relaciones causales viene expresado por un conjunto de ecuaciones estructurales de las cuales se deducen los coeficientes *path*.

$$x_2 = p_{21}x_1 + e_2$$

$$x_3 = p_{32}x_2 + p_{31}x_1 + e_3$$

$$x_4 = p_{43}x_3 + p_{42}x_2 + p_{41}x_1 + e_4$$

Los coeficientes *path* ( $p_{ij}$ ) expresan de forma matemática la relación causal entre las variables integrantes del modelo midiendo el grado de variación producida en la variable dependiente por cada una de las variables independientes, permaneciendo las otras constantes.

Miden, por tanto, el poder explicativo que tienen las variables independientes sobre la variable dependiente y la evaluación matemática del grado de error medirá el porcentaje de influencia causal que queda sin explicar por las variables independientes.

Es necesario también determinar los residuos (medidas de error implícitos en las variables endógenas). Este error expresa la variación total producida en la variable dependiente, pero no causada por el resto de las variables.

## METODOS DE ESTIMACION

### *El método de las variables instrumentales*

Este método parte de tres supuestos en el proceso de estimación:

- a) Las variables contenidas en el modelo aparecen en forma estandarizada

$$\begin{aligned} E(x_i) &= 0 \\ E(x_i^2) &= 1 \end{aligned} \quad i = 1, 2, 3, 4$$

- b) La variable exógena  $X_1$  no está correlacionada con ninguno de los residuos (son independientes)

$$E(x_1e_2) = E(x_1e_3) = E(x_1e_4) = 0$$

- c) No existe correlación entre los residuos y las variables endógenas pre-determinados anteriormente en el modelo

$$E(x_2e_3) = E(x_2e_4) = E(x_3e_4) = 0$$

En base a estos tres supuestos, el método consiste en multiplicar cada ecuación del modelo por diferentes variables del mismo y hallar los valores esperados de las ecuaciones resultantes y expresar este último resultado en forma de correlaciones y de coeficientes *path*.

Si utilizamos las variables  $X_3$  como variable instrumental en la ecuación  $X_4$ , obtenemos

$$x_4x_3 = p_{43} \cdot x_3x_3 + p_{42}x_2x_3 + p_{41}x_1x_3 + e_4x_3$$

y hallando la esperanza matemática tendríamos:

$$E(x_4x_3) = p_{43}E(x_3)^2 + p_{42}E(x_2x_3) + p_{41}E(x_1x_3) + E(e_4x_3)$$

al estar las variables estandarizadas:

$$E(x_ix_j) = r_{ij}$$

$$E(x_i)^2 = 1$$

$$E(x_ie_j) = 0$$

Utilizando las diferentes variables del modelo, el sistema de ecuaciones resultante quedará de la forma siguiente:

$$\begin{aligned} r_{12} &= p_{12} \\ r_{13} &= p_{31} + p_{32}r_{12} \\ r_{23} &= p_{31}r_{12} + p_{32} \\ r_{14} &= p_{41} + p_{42}r_{12} + p_{43}r_{13} \\ r_{24} &= p_{41}r_{12} + p_{42} + p_{43}r_{23} \\ r_{34} &= p_{41}r_{13} + p_{42}r_{23} + p_{43} \end{aligned}$$

Resolviendo los coeficientes *path*  $p_{ij}$  en términos de coeficientes de correlación  $r_{ij}$ , tendremos:

— para la primera ecuación,

$$p_{21} = r_{12}$$

— para la segunda ecuación,

$$p_{31} = \frac{(r_{13} - r_{12}r_{23})}{(1 - r_{12}^2)}$$

$$p_{32} = \frac{(r_{23} - r_{12}r_{13})}{(1 - r_{12}^2)}$$

— para la tercera ecuación

$$\text{siendo } P = \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{vmatrix}$$

$$P_{41} = \frac{\begin{vmatrix} r_{14} & r_{12} & r_{13} \\ r_{24} & 1 & r_{23} \\ r_{34} & r_{23} & 1 \end{vmatrix}}{P}$$

$$P_{42} = \frac{\begin{vmatrix} 1 & r_{14} & r_{13} \\ r_{12} & r_{24} & r_{23} \\ r_{13} & r_{34} & 1 \end{vmatrix}}{P}$$

$$P_{43} = \frac{\begin{vmatrix} 1 & r_{12} & r_{14} \\ r_{12} & 1 & r_{24} \\ r_{13} & r_{23} & r_{34} \end{vmatrix}}{P}$$

La regla de la multiplicación de Wright señala que la correlación entre  $X_i$  y  $X_j$ , donde  $X_j$  aparece más tarde que  $X_i$  en el modelo, es igual al producto de los coeficientes *path* directos e indirectos a lo largo del *path* desde  $X_j$  para atrás y hacia adelante en un solo paso. El logaritmo que facilita las ecuaciones que relacionan las correlaciones y los coeficientes *path* es:

$$r_{ij} = \sum_q P_{iq} r_{jq}$$

en donde «i» y «j» son dos variables del modelo específico y «q» implica a todas las variables de las que salen *paths* directos hacia  $X_i$ .

El método de los mínimos cuadrados simple (OLS) es lo suficientemente conocido como para no necesitar una explicación. Simplemente señalar que es necesario llevar a cabo tantas regresiones múltiples como ecuaciones haya en el modelo objeto de investigación.

## MODELOS NO RECURSIVOS

En primer lugar cabe interpretar los modelos no recursivos simplemente como modelos relacionales que definen distribuciones conjuntas de posibilidades de ocurrencia de variables dependientes condicionadas a las variables predeterminadas. Puede, asimismo, hablarse simplemente de que todas las variables predeterminadas conjuntamente causan las variaciones de las variables dependientes del modelo. Este es el sentido que tendría, según Strotz y Wold, la idea cotidiana de que las disponibilidades alimenticias determinan la población de peces existentes. Pero en este tipo de causalidad no sabríamos qué tipos de peces dependen de otros peces, ni las interacciones entre ellos.



También cabe hablar de dos posibilidades más, siempre según los autores anteriores:

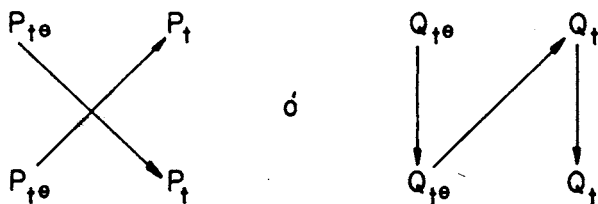
- círculos causales o causación recíproca;
- bicausalidad.

En el primer caso se trata realmente de variables operando en el tiempo. Si tenemos el siguiente sistema:

$$p(t) = a_1 + b_1q(t) + c_1z_1(t) + v_1(t)$$

$$q(t) = a_2 + b_2p(t) + c_2z_2(t) + v_2(t)$$

entonces realmente estaremos ante una de estas dos posibilidades, u otras parecidas



con lo cual realmente sólo existe causación recíproca porque hemos observado el sistema en condiciones de equilibrio (estudio sectorial y no longitudinal).

En el segundo caso —bicausalidad—, tal como señalan los autores, tampoco se da causación recíproca. Lo que debe deducirse de todo esto es que la bicausalidad o causación recíproca es más aparente que real.

«La interpretación causal de un coeficiente en cualquiera de estos dos modelos en equilibrio debe de buscarse en el modelo dinámico subyacente, que será una noción de causalidad similar a la usada en el laboratorio» (R. H. Strotz y H. O. Wold, 1960: 426).

*Problemas de estimación.*—En los modelos y sistemas no recursivos se plantea, de una manera más tajante que en los recursivos, el problema de la estimación. Este problema está condicionado por el de la identificación del sistema. Desde el punto de vista de la identificación sólo existen tres posibilidades:

1. El modelo está subidentificado.
2. El modelo está exactamente identificado.
3. El modelo está sobreidentificado.

El primer caso implica que el sistema de ecuaciones que determina el modelo no ofrece suficiente información para ser resuelto y para estimarse por tanto los diferentes coeficientes. En los otros dos casos se tiene suficiente información para hallar las soluciones de las ecuaciones. La única diferencia entre el 2 y el 3 estriba en que en este último caso existe un superávit de información a favor de dicha solución.

De una manera más estricta se dice que un sistema de ecuaciones está identificado si cumple dos condiciones:

- *La condición de orden*: Si el número de variables predeterminadas excluidas de una ecuación es igual o mayor al número de variables endógenas en la ecuación menos uno, decimos que dicha ecuación está identificada.
- *La condición de rango*: Que cada ecuación del modelo sea distinta de todas las demás y de las posibles combinaciones lineales de éstas.

Sólo si se cumplen ambas condiciones para todas las ecuaciones se puede hablar de identificación del sistema.

La estimación de los coeficientes del sistema de ecuaciones sólo es posible cuando el modelo esté identificado (exactamente o sobreidentificado). En el caso de subidentificación no existe suficiente información y, por tanto, o se incorpora nueva información al modelo o se imponen nuevas restricciones, que es exactamente igual.

*Los métodos de estimación* existentes, tal y como son enumerados en la literatura, son:

1. Mínimos cuadrados indirectos.
2. Variables instrumentales.
3. Máxima verosimilitud información limitada.
4. Máxima verosimilitud información completa.
5. Mínimos cuadrados bietápicos.
6. Mínimos cuadrados trietápicos.
7. Punto fijo.

(Véanse J. Kmenta, 1971; R. Karns, 1974; C. Dagun y E. M. Bee de Dagun, 1971, entre otros.)

Los métodos de mínimos cuadrados indirectos y variables instrumentales sólo pueden ser utilizados cuando el sistema está exactamente identificado, mientras que los demás métodos pueden utilizarse tanto en el caso de sobreidentificación como de exacta identificación. Utilizando este criterio (identi-

---

ficación exacta o en exceso) junto con el de si el método de estimación tiene en cuenta toda la información del sistema o no, se llega al cuadro siguiente:

	<i>Exactamente identificado</i>	<i>Sobreidentificado</i>
Estimación de ecuaciones una a una	— Variables instrumentales.	— Mínimos cuadrados bietápicos. — Máxima verosimilitud con información parcial.
Estimación de todo el sistema	— Mínimos cuadrados indirectos.	— Mínimos cuadrados trietápicos. — Máxima verosimilitud con información completa. — Punto fijo.

Claramente, ciertos métodos de estimación no son adecuados en determinadas ocasiones (sobreidentificación); pero de los que sí lo son, ¿cómo elegir el más adecuado? Karns (1974) y Kmenta (1971) ofrecen razones teóricas y empíricas al respecto. Ambos señalan que para sistemas sobreidentificados que no utilizan la información de todo el modelo, el método de mínimos cuadrados bietápicos parece ligeramente superior al de máxima verosimilitud.

En todo lo que respecta a los diferentes métodos de modelos sobreidentificados e información completa, los datos existentes no permiten deducir nada sobre cada uno de ellos respecto a los demás, pero sí como bloque respecto a los métodos que no usan toda la información. Contrariamente a lo que cabría pensar resultan más eficientes los métodos de información parcial. Esto se debe ante todo a la vulnerabilidad de los otros métodos ante errores de especificación del modelo, algo relativamente frecuente en ciencias sociales. Así, Karns puede concluir diciendo:

«Como conclusión, el actual tratamiento de los estimadores de los sistemas no recursivos y la presentación de los resultados de la investigación por el método Montecarlo, sugieren que el método de mínimos cuadrados bietápicos es el más útil para los científicos sociales» (R. Karns, 1974: 31).

## EL MODELO DE KARL JÖRESKOG

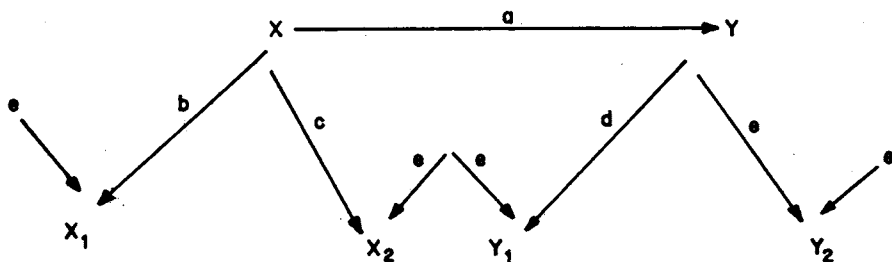
La no recursividad y el empleo de coeficientes estructurales acercan el análisis multivariable desarrollado alrededor del «path» análisis a las necesidades realmente sentidas por los investigadores al hacer más realistas los modelos.

El progresivo afianzamiento del enfoque constructivista en el estudio de la relación teoría/observaciones, así como la escasa fuerza explicativa de la mayoría de los modelos multivariables desarrollados dentro de las teorías apuntadas en este artículo, han llevado a un progresivo interés por el problema del error de medición, error tanto aleatorio como sistemático.

En el gráfico,  $X_1$ ,  $X_2$ ,  $Y_1$  e  $Y_2$  son las variables efectivamente observadas. Lo que interesa al investigador es, sin embargo, la relación entre  $X$  e  $Y$ , que son constructos teóricos no observados directamente (véanse Costner, 1969; N. K. Namboodiri, L. F. Carter y H. M. Blalock, 1975, y A. S. Goldberger y O. D. Duncan, 1973). Cada uno de estos constructos teóricos se mide a través de dos indicadores distintos y el modelo permite error en la medición ( $e$ ), pero este error es un error aleatorio, no sistemático. Se asume que los indicadores son efectos y no causas de los constructos, pues son las variaciones de estos últimos lo que origina variaciones en los indicadores, y no al revés.

## GRAFICO 2

*Un modelo de dos indicadores con variables no observadas*



El modelo del gráfico 2 es un modelo sobreidentificado, pues hay cinco coeficientes que estimar (a, b, c, d, e) y se dispone de seis correlaciones para su estimación (cada variable observada con las cuatro restantes). En base a esta sobre identificación, Costner establece un «criterio de asistencia» del modelo como

$$\gamma_{X_1 Y_1} \gamma_{X_2 Y_2} = \gamma_{X_1 Y_2} \gamma_{X_2 Y_1}$$

lo que equivale a

$$(bad)(cae) = (bae)(cad)$$

$$ba^2cde = ba^2cde$$

Si el criterio de consistencia es aplicable, entonces los coeficientes serán estimables unívocamente. Este enfoque de indicadores múltiples ha sido extendido a más de dos constructos teóricos, a más de dos indicadores y a modelos longitudinales. Asimismo se han estudiado modelos en los que se da cabida al error sistemático no aleatorio (véanse C. E. Werts, R. L. Linn y K. G. Jöreskog, 1971; D. R. Heise, 1969; H. M. Blalock, 1971; D. E. Wiley, 1973; Karl G. Jöreskog, 1973, y H. L. Costner y R. Schoenberg, 1973).

Si en el modelo del gráfico 2 nos olvidamos de la relación entre X e Y, es decir, del coeficiente a, para cada constructo teórico X e Y la estimación de b, c, d y e se transforma simplemente en un caso especial del modelo clásico del análisis factorial<sup>1</sup>.

A diferencia de éste, sin embargo, la especificación de ciertas relaciones funcionales entre observables y factores comunes, así como de las interrelaciones entre los factores, llevan a lo que se denomina análisis factorial confirmatorio. De aquí a la formulación de un modelo estructural para los propios factores no hay más que un paso, y esto es precisamente lo que lleva a cabo el modelo general y estructural de covarianza de Jöreskog.

Este modelo general se transformó posteriormente en el programa LISREL (Jöreskog y Van Thillo, 1972). El método y modelo de LISREL permite el uso de constructos hipotéticos o variables latentes, así como el empleo de indicadores múltiples observados para cada variable no observada; permite asimismo el manejo tanto de errores en las ecuaciones (residuos...) como en las variables observadas (medición u observación).

Se parte de dos vectores de variables aleatorias  $\eta'$  y  $\xi'$ , que representan, respectivamente las variables dependientes e independientes verdaderas, y del siguiente sistema de relaciones estructurales

$$\beta \eta = \Gamma \xi + \varepsilon$$

donde  $\beta$  y  $\Gamma$  son matrices de coeficientes y  $\zeta$  es un vector de términos de error. Se asume que tanto  $\eta$  como  $\xi$  y  $\zeta$  están estandarizadas y asimismo que  $\zeta$  y  $\xi$  son ortogonales, siendo  $\beta$  una matriz no singular.

Los vectores observados serán  $Y'$  y  $X'$  y no  $\eta$  e  $\xi$ , de modo que  $Y$  y  $X$  son las variables observadas, y  $\eta$  y  $\xi$ , las verdaderas variables o constructos teóricos. Sean  $\Phi$  y  $\Psi$  las matrices de varianzas/covarianzas de  $\xi$  y  $\zeta$ , respectivamente, y  $\Theta_\epsilon^2$  y  $\Theta_\delta^2$  las matrices diagonales de errores de  $Y$  y  $X$ , respectivamente; entonces, dado que

$$Y = \mu \Lambda_y \eta + \epsilon \quad \text{siendo } \mu = E(Y)$$

y

$$X = \gamma + \Lambda_x \xi + \delta \quad \gamma = E(X)$$

y  $\Lambda_y$  e  $\Lambda_x$ , las matrices de regresión de  $Y$  en  $\eta$ , y  $X$  en  $\xi$ , la matriz de varianzas/covarianzas de  $Z$ .

Donde  $Z = (Y', X)'$  será:

$$\Sigma = \begin{pmatrix} \Lambda_y [B^{-1} \Gamma \Phi \Gamma' B^{-1} + B^{-1} \Psi B^{-1}] \Lambda_y' + \Theta_\epsilon^2 & \Lambda_y B^{-1} \Gamma \Phi \Lambda_x' \\ \Lambda_x \Phi \Gamma' B^{-1} \Lambda_y' & \Lambda_x \Phi \Lambda_x' + \Theta_\delta^2 \end{pmatrix}$$

Como se ve, los elementos de  $\Sigma$  son función de las matrices

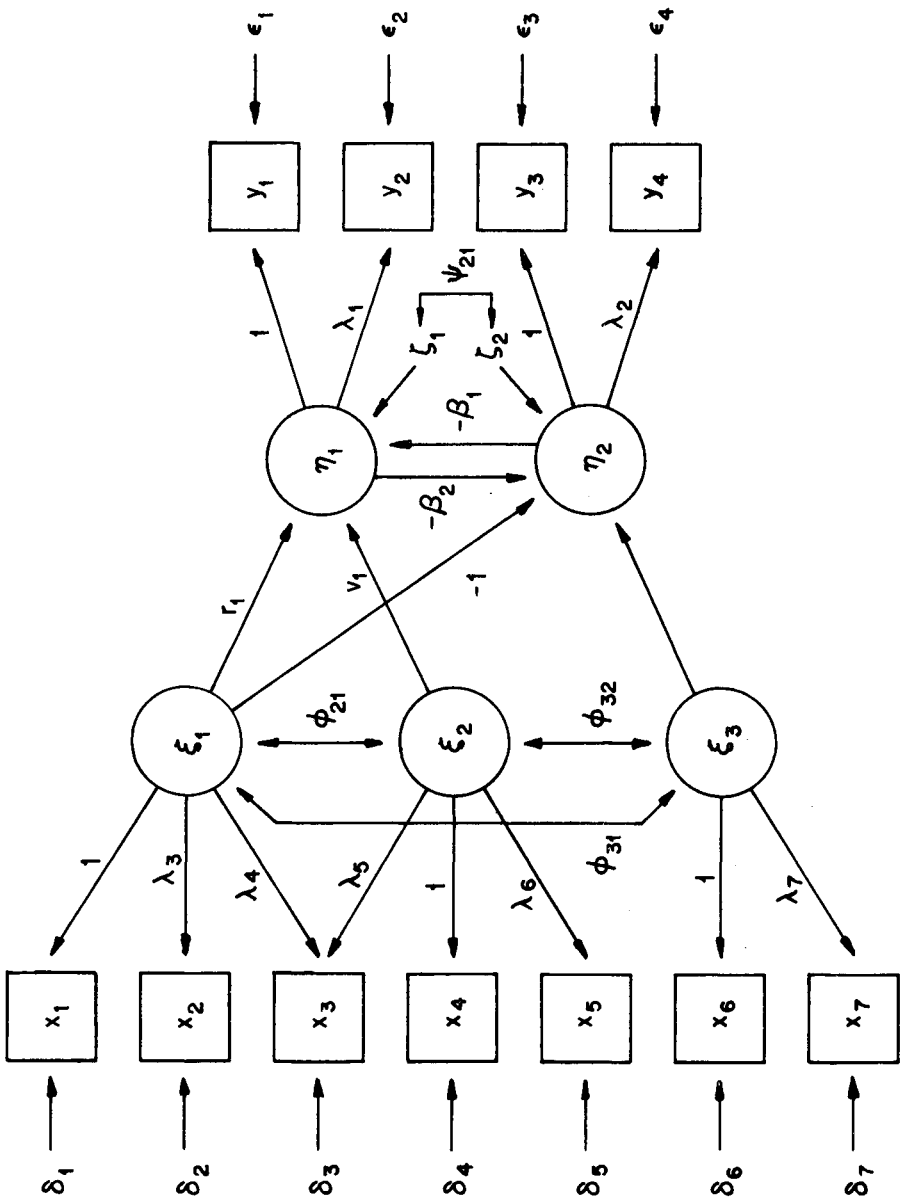
$$\Lambda_y, \Lambda_x, B, \Gamma, \Phi, \Psi, \Theta_\delta \text{ y } \Theta_\epsilon$$

Algunos de los elementos de estas matrices son parámetros fijos, otros adoptan valores constreñidos y, por último, hay parámetros fijos. El juego entre estos tres tipos de parámetros es lo que permite la estimación o no de los coeficientes y el plantear problemas de la identificación.

Este modelo descrito puede verse en el gráfico 3 y la derivación de la estimación de los parámetros en Jöreskog y Van Thillo (1972).

GRAFICO 3

Un modelo con variables no observadas



## APLICACIONES DEL MODELO

Un ejemplo interesante, por cuanto que ha dado lugar a amplias polémicas y a numerosas investigaciones, es el de la relación entre actitud y acto\*. El planteamiento inicial de la dicotomía es un planteamiento predictivo. ¿Hasta qué punto la expresión verbal de una actitud es un buen criterio de predicción de conducta? Las investigaciones existentes ponen de relieve la falta de relación o la escasa relación existente entre actitud y acto. Sin embargo, no todos los estudios son tan negativos. Dejando a un lado los problemas que se plantean de tipo sustantivo y epistemológico (H. Schuman y M. P. Johnson, 1976, y F. Alvira, 1977), es indudable que la medición de las actitudes normalmente utilizadas en las investigaciones no guarda una relación perfecta con la actitud subyacente. Las definiciones de actitud que usualmente se manejan implican algo interno al individuo no directamente observable ni medible; sin embargo, los diseños de investigación utilizados no tienen en cuenta esta diferencia entre actitudes verbales o expresiones verbales de una actitud y la correspondiente actitud subyacente. La actitud es un constructo teórico y como tal debe ser tratada; resulta así apropiado el enfoque de indicadores múltiples y asimismo la aplicación del modelo de Jöreskog\*\*.

Tittle y Hill (1967) exploraron con datos sobre elecciones políticas la hipotética relación entre actitudes y acto. Disponían de las siguientes variables observadas:

- Voto en la elección anterior a la última ( $C_1$ ).
- Voto en la elección última ( $C_2$ ).
- Actitud expresada verbalmente ( $V_1$ ).
- Actitud medida en una escala Likert ( $V_2$ ).
- Actitud medida en una escala Thurstone ( $V_3$ ).
- Actitud medida mediante el diferencial semántico ( $V_4$ ).
- Actitud medida con una escala Guttman ( $V_5$ ).

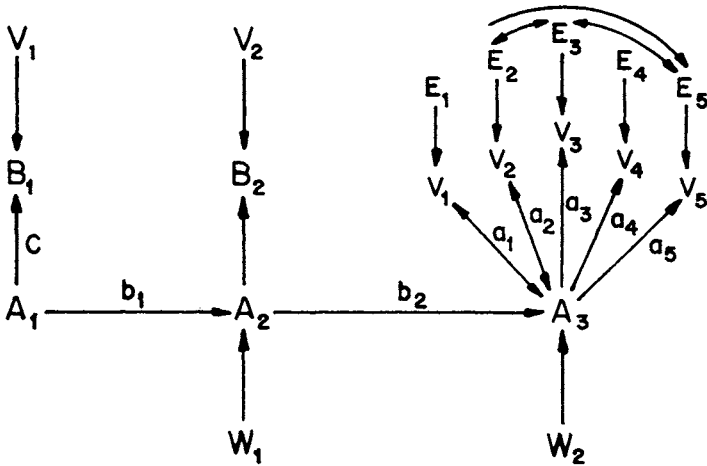
El modelo puede ponerse gráficamente como

---

\* A este respecto puede verse F. ALVIRA MARTÍN, "La relación entre actitud y acto", *REOP*, núm. 49, 1977.

\*\* Estrictamente hablando se aplica algo anterior al LISREL que es el análisis factorial confirmatorio, pero la dinámica subyacente es la misma.





En dicho modelo,  $A_1$ ,  $A_2$  y  $A_3$  son las actitudes subyacentes no observadas directamente;  $B_1$  y  $B_2$  son los votos en las elecciones;  $V_1 \dots V_5$ , las actitudes medidas verbalmente, y  $V_1, V_2, E_1 \dots E_5, W_1$  y  $W_2$ , los términos de error \*.

Tal como está especificado, el modelo está infraidentificado y es necesario bien suponer que  $b_1 = b_2$  o  $C_1 = C_2$ , o ambas a la vez. Si así lo hacemos, el modelo analítico queda como sigue

$$Z = B F + V W$$

siendo

$$Z = \begin{pmatrix} B_1 \\ B_2 \\ V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \end{pmatrix} \quad B = \begin{pmatrix} C_1 & 0 & 0 \\ 0 & C_2 & 0 \\ 0 & 0 & a_1 \\ 0 & 0 & a_2 \\ 0 & 0 & a_3 \\ 0 & 0 & a_4 \\ 0 & 0 & a_5 \end{pmatrix} \quad F = \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix}$$

$$V = \begin{pmatrix} v_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & v_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & e_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & e_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & e_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & e_4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & e_5 \end{pmatrix} \quad y \quad W = \begin{pmatrix} V_1 \\ V_2 \\ E_1 \\ E_2 \\ E_3 \\ E_4 \\ E_5 \end{pmatrix}$$

\* Véase D. F. ALWIN, "Making Inferences from Attitude-Behavior Correlations", *Sociometry*, vol. 36 (2), 1973.

La matriz de varianzas/covarianzas de las actitudes subyacentes es

$$E(F F') = \begin{vmatrix} 1 & & & \\ b_1 & 1 & & \\ b_1 b_2 & b_2 & 1 & \\ & & & \end{vmatrix}$$

Mediante este modelo y usando el AFC (análisis factorial confirmatorio) se obtiene  $C_1 = 0,657$  y  $C_2 = 0,622$ , lo que indudablemente pone de relieve que *sí que existe relación entre actitud y acto*.

Las posibilidades del modelo y técnica de Joreskög y colaboradores no se limitan a la Psicología Social, sino que abarcan prácticamente todos los aspectos investigadores de las Ciencias Sociales en general y la Sociología en particular. Joreskög y Van Thillo (1972), en la presentación que hacen de LISREL, presentan una aplicación del mismo al trabajo de Duncan, Haller y Portes sobre la influencia de los compañeros en las aspiraciones ocupacionales y educativas (1967).

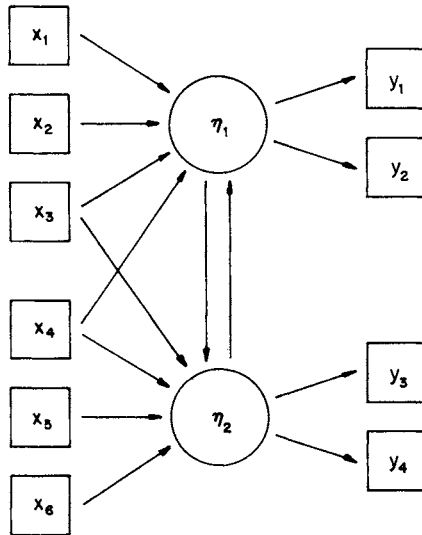
Parten estos autores de cinco variables observadas correspondientes a la persona que se analiza (el entrevistado) y cinco variables observadas análogas que corresponden al amigo del entrevistado. Estas variables observadas son:

- aspiraciones de los padres;
- inteligencia;
- *status* socio-económico de la familia;
- aspiraciones ocupaciones;
- aspiraciones educativas,

y se miden tanto para el entrevistado como para su mejor amigo.

Duncan y colaboradores introducen el constructo teórico «ambición» entre las tres primeras variables y las dos últimas. Sustantivamente se está proponiendo que la «ambición» del entrevistado es función, aparte de características del propio entrevistado (inteligencia, *status* socio-económico y aspiraciones de los padres), de la «ambición» del amigo, y que es esta «ambi-

ción» (constructo teórico) la que determina las aspiraciones ocupacionales y educativas. Gráficamente esto implica



siendo  $X_1$ ,  $X_2$  y  $X_3$  las aspiraciones de los padres, inteligencia y *status* socio-económico, respectivamente;  $\eta_1$ , la «ambición», e  $Y_1$ ,  $Y_2$ , las aspiraciones ocupacionales y educativas del entrevistado; las variables en la parte baja del diagrama corresponden al amigo.

Se parte de los siguientes supuestos:

- a) Las variables  $X_1 \dots X_6$  se miden sin error y pueden estar correlacionadas entre sí.
- b) Se permiten términos de error en la explicación de  $\eta_1$ ,  $\eta_2$ ,  $Y_1 \dots Y_4$  y que estas últimas no sean ortogonales.

Teniendo en cuenta que  $\Delta_x = I$  y  $\xi = 0$ . Si llamamos  $\gamma_1 \dots \gamma_3$  a los coeficientes estructurales que relacionan los  $X_1 \dots X_6$  con las  $\eta_i$ ;  $\xi_i$  a los errores de medición de éstos, y  $\zeta$  a los  $X$  medidos como desviaciones de las medias, se obtienen las siguientes ecuaciones estructurales que relacionan los  $X_1 \dots X_6$  y los  $\eta_1$ ,  $\eta_2$ .

$$\begin{pmatrix} 1 & \beta_1 \\ \beta_2 & 1 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} r_1 & r_2 & r_3 & r_4 & 0 & 0 \\ 0 & 0 & r_5 & r_6 & r_7 & r_8 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \\ \xi_6 \end{pmatrix} + \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix}$$

Las ecuaciones que relacionan  $Y_1, \dots, Y_4$  con  $\eta_1, \eta_2$ , llamando a los coeficientes  $\lambda$ , y al error  $V$  son las siguientes:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix} + \begin{pmatrix} \lambda_1 & 0 \\ \lambda_2 & 0 \\ 0 & \lambda_3 \\ 0 & \lambda_4 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \end{pmatrix}$$

1  
2  
donde  $\mu = E(Y)$ .

Estas ecuaciones, junto con las restricciones en los diferentes coeficientes y la matriz de intercorrelaciones, permite la resolución de los diferentes coeficientes no fijos.

## A MODO DE CONCLUSION

Se ha dado la vuelta completa: experimentación frente a correlación, *path* análisis, recursividad/no recursividad, modelos estructurales con variables no observadas. La aplicación de modelos como el de Jorëskog supone un avance importante en la búsqueda de relaciones causales entre variables, pero no es un paso definitivo ni se logra con ello desbancar la preeminencia epistemológica de la experimentación.

Cualquier modelo que se utilice tiene necesariamente que llevar a cabo supuestos simplificadores. Unos supuestos serán más plausibles que otros; algunos de ellos podrán ser contrastados empíricamente, pero en cualquier caso no *habrá un solo modelo* que responda unívocamente a los datos empíricos.

En el ejemplo de la hipotética relación entre actitud/acto, una vez obtenidos los coeficientes del modelo se pueden estimar los coeficientes de correlación a partir de ellos y compararlos con los obtenidos empíricamente. La bondad del ajuste puede verse con cualquiera de las técnicas normalmente utilizadas. Pero el problema estriba en que no hay un único modelo que arroje un buen ajuste; para unos datos dados hay siempre varios modelos. ¿Cómo elegir entre ellos?

En ese ejemplo cabe la posibilidad de utilizar la predicción como base de comparación de modelos alternativos. Así, puesto que se dispone de  $B_1$ ,  $B_2$  y  $A_3$ , puede estimarse la correlación entre  $A_3$  y un hipotético  $B_3$ , medir éste y comparar la correlación predicha con la obtenida realmente. Esto no resuelve el problema tajantemente, y además en algunos modelos teóricos esta contrastación predictiva resultará imposible.

El interés de estas técnicas de análisis reside en el hecho señalado al principio de que su papel es limitar el número de explicaciones alternativas de unos mismos datos, de modo que la elección entre modelos teóricamente válidos sea más fácil. En este sentido, por tanto, llevan a cabo el mismo papel que tiene la aleatorización de sujetos sobre tratamientos en experimentación: dejar a un lado explicaciones alternativas potencialmente válidas.

- ALWIN, D. F.: "Making Inferences from Attitud-Behavior Correlations", *Sociometry*, vol. 36 (2), 1974.
- BLALOCK, H. M.: "Causal Models involving unmeasured variables in Stimules Response Situations", en H. M. BLALOCK, *Causal Models in the Social Sciences*. Aldine, 1971.
- COSTNER, H. L., y SILVENBERG: "Diagraming indicator ills in multiple Indicator models", en H. M. BLALOCK, *Causal Models in the Social Sciences*. Aldine, 1971.
- "Theory, Deduction and rule of correspondence". *A. J. S.*, 75.
- DARLINGTON, R. B.: "Multiple Regression in Psychological Research and Practice", *Psychological Bulletin*, 1968, vol. 69 (3).
- DUNCAN, O. D.: "Some Linear Models for two-wave, two variables panel analysis", *Psychological Bulletin*, 1969, vol. 72 (3).
- *Introduction to Structural Equation Models*. Aldine, 1975.
- HALLER, A. V., y PORTER, A.: "Peer influence and aspirations: a reinterpretation", *American Journal of Sociology*, 1968, vol. 74 (2).
- FINNEY, J. H.: "Indirect effects in Path Analysis", *Sociological Method and Research*, 1972, vol. 1 (2).
- FISHER, F. H.: "A correspondence principle for simultaneous equation Models", *Econometrica*, 1970, vol. 38 (1).
- FORBES, H. D.: "A note of caution in Causal Modelling", *American Political Sciences Review*, 1968.
- GOLDBERGER, A. S.: "On Boudon's Method of Linear Causal Analysis", *American Sociological Review*, 1970.
- Y DUNCAN, O. D.: *Structural Equation Models in the Social Sciences*. Seminar Press, 1973.
- MEISE, D. R.: *Causal Analysis*. John Wiley and Sons (Ed.), 1975.
- JORESOG, K. G., y THRIHO, M. V.: *Lisrel a general computer program for estimating a linear structural equation system involving multiple indicators of unmeasured variables*. Educational Testing Service, Princeton, 1972.
- JORESOG, K. G.: "A general Method for estimating a linear structural equation System", en GOLDBERGER, A. S., y DUNCAN, O. D., *Structural Equation Models in the Social Sciences*. Seminar Press, 1973.
- "Analyzing Psychological. Data by Structural Analysis of Covariance Matrices". Mimeografiado. Uppsala.
- y SORBOM, D.: "A General Computer Program for Estimation of Linear Structural Equation Systems by Maximun Likelihood Methods", 1978. Mimeografiado, Uppsala.
- KISH, L.: "Representation, randomization and Control", en BLALOCK y otros, *Quantitative International Sociology*. Academic Press, 1975.
-

- MAPES, R. E. A., y ALLAN, G. J. B.: "Path Analysis, A Cautionary Note", *The Sociological Review*, 1973, vol. 21.
- MILLER, A. D.: "Logic of Causal Analysis from Experimental to Non-experimental Designs", en BLALOCK, H. M. (Ed.), *Causal Models in the Social Sciences*. Aldine Pub., 1971.
- NAMHOORDINI, N. K.; CARTER, L. F., y BLALOCK, H. M.: *Applied Multivariate Analysis and Experimental Designs*. McGraw-Hill, 1975.
- NURNI, Hamnu: "Social causality and Empirical Data Reduction Techniques", *Quality and Quantity*, 1974 (8).
- STOKES, D. E.: "Compound Paths: An Expository Note", *American Journal of Political Science*, 1974, vol. 18.
- TITTLE, C. R., e HILL, R. J.: "Attitude Measurement and prediction of behavior: an evaluation of analiticae and measurement techniques", *Sociometry*, 1967.
- WERSTS, Ch. E.; LINN, R. L., y JORESKOG, K. G.: "Estimating the parameters of Path Models involving unmeasured variables", en H. M. BLALOCK (ed.), *Causal Models in the Social Sciences*. Aldine, 1971.
- JORESKOG, K. G., y LINN, R. L.: "Identification and Estimation in Path Analysis with Unmeasured variables", *American Journal of Sociology*, vol. 78 (6), 1973.
- WILEY, D. E.: "The Identification Problem for structural equations Models with unmeasured variables", en Goldberger, A. S., y Duncan, O. D. (eds.), *Structural Equations Models in the Social Sciences*. Seminar Press, 1973.
- WOLD, H.: "Causal Inference from observational Data. A review of Ends and Means", *The Journal of the Royal Statistical Society. Serie A*, 1956, vol. 119 (1).
-