6. Expresión matemática

Un modelo teórico, una explicación en definitiva, puede encontrar diferentes formas de expresión; ya sea en la apariencia de un diagrama, adoptando una enunciación verbal o escrita, en todos los casos se trata del mismo modelo. Una forma nueva de representar el mismo modelo es mediante un sistema de ecuaciones. Para ello, deberemos adoptar una serie de convenciones para poder formular el modelo ecuacionalmente. No existe una notación universalmente aceptada, (evidentemente, no existe una notación natural) y la que empleamos no deja de ser una más de las existentes.

6.1. Notación de sistemas estructurales

Las variables endógenas (dependientes) las notaremos mediante una Y con subíndice que expresa un número que la diferencia. Para el caso de las variables exógenas (independientes) emplearemos una X con subíndice.

Variable endógena	$\mathbf{y_i}$
Variable exógena	$\mathbf{x_i}$

En lo que se refiere a las relaciones o efectos, aquel que se postula entre variables endógenas lo notaremos β con dos subíndices (ij) donde se identifican las variables que intervienen en dicha relación. El subíndice (i) para la variable que recibe el efecto (y por tanto que es explicada) y el subíndice (j) para la variable que explica.

$$\mathbf{y_j} \xrightarrow{\beta_{ij}} \mathbf{y_i}$$

Para la relación de una variable exógena sobre una endógena emplearemos una γ con la misma intencionalidad en los subíndices.

$$\mathbf{x_j} \xrightarrow{\gamma_{ij}} \mathbf{y_i}$$

Como ya se advirtió al hablar del contenido de las perturbaciones, estas serán notadas ζ con el subíndice de la variable correspondiente. Evidentemente, suponemos

una perturbación por cada ecuación. Ya nos es posible especificar un sistema de ecuaciones lineales, donde habitualmente los efectos son aditivos. El sistema tendrá tantas ecuaciones como variables endógenas contenga, dado que cada variable endógena posee alguna previa que explica su variabilidad. En ese sentido, recordemos que la variación y covariación entre variables endógenas están, de algún modo, determinadas por la variación y covariación entre variables exógenas, lo que nos lleva a reconocer que la varianza y covarianza de las variables exógenas son fundamentales en todo modelo.

Necesitamos, por lo tanto, una forma de notación para las varianzas de cada variable y las covarianzas entre ellas. La cuantía de la $\mathbf{x_i}$ (la variación de la variable exógena) se nota $\mathbf{\Phi_{ii}}$. Cuando se trate de la covarianza entre dos variables exógenas $\mathbf{x_i}$ e $\mathbf{x_j}$ serán los subíndices los encargados de indicarlo.

$$\mathbf{\Phi_{ij}}$$

La varianza de las perturbaciones ζ_i se nota como

$$\mathbf{\Psi_{ii}}$$

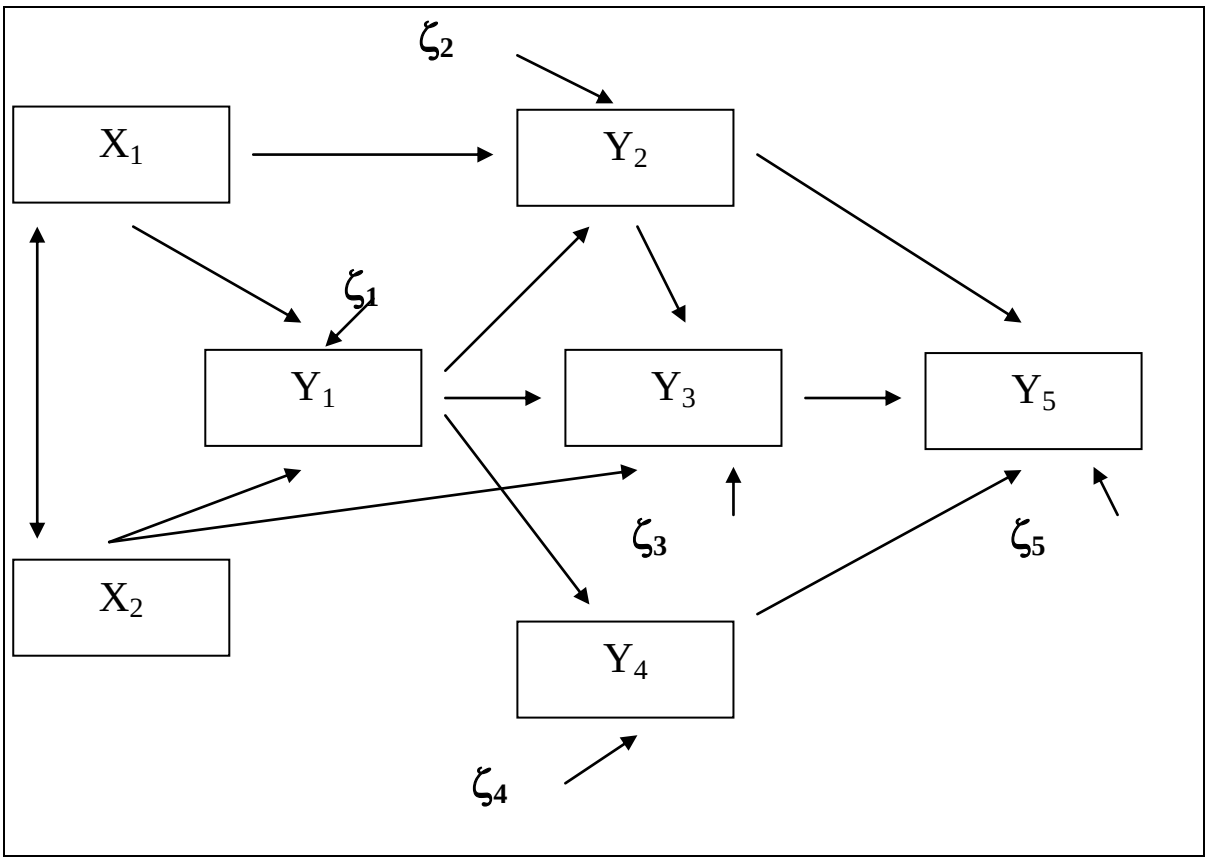
y nuevamente cuando nos refiramos a la covarianza entre dos perturbaciones ζ_i e ζ_j

$$\mathbf{\Psi_{ij}}$$

Una vez acordadas las convenciones de notación, podemos utilizarlas para construir ecuaciones.

6.2. Sistemas de ecuaciones

Antes de comenzar, debemos recordar que un modelo estructural no es simplemente un sistema de ecuaciones. Lo esencial es que dicho sistema represente el mecanismo causal que ha producido los valores observados en las variables endógenas. En ese sentido, el diagrama causal siguiente expresaría una secuencia explicativa.



Sobre la base del sistema de notación que se ha introducido, las relaciones entre variables endógenas se expresaran mediante una Beta β con los subíndices correspondientes a las variables que esta relacionando. Recordemos que primero se posiciona el subíndice de la variable efecto (la que recibe el grafo) y seguidamente el subíndice correspondiente a la variable que se propone como causa de ella.

$$y_1 = 0y_1 + 0y_2 + 0y_3 + 0y_4 + 0y_5 + \gamma_{11}x_1 + \gamma_{12}x_2 + \alpha_1 + \zeta_1$$
$$y_2 = \beta_{21}y_1 + 0y_2 + 0y_3 + 0y_4 + 0y_5 + \gamma_{21}x_1 + 0x_2 + \alpha_2 + \zeta_2$$
$$y_3 = \beta_{31y1} + \beta_{32}y_2 + 0y_3 + 0y_4 + 0y_5 + 0x_1 + \gamma_{32}x_2 + \alpha_3 + \zeta_3$$
$$y_4 = \beta_{41}y_1 + 0y_2 + 0y_3 + 0y_4 + 0y_5 + 0x_1 + 0x_2 + \alpha_4 + \zeta_4$$
$$y_5 = 0y_1 + \beta_{52}y_2 + \beta_{53}y_3 + \beta_{54}y_4 + 0y_5 + 0x_1 + 0x_2 + \alpha_5 + \zeta_5$$

6.2.1. Presunciones

En el planteamiento de modelos causales son habitualmente necesarias un conjunto de presunciones que definan el marco de la especificación del sistema que se propone. Estas presunciones son testadas durante la fase de ajuste empírico del sistema de ecuaciones sobre los datos. Para un modelo expresado con las variables no trasformadas, es decir tal y como se han registrado, encontraremos normalmente cuatro presunciones básicas.

La primera a considerar afirma que la media de los errores es cero para todas las ecuaciones. Lo que se afirma mediante esta presunción es que la ecuación estructural explica correctamente la variable endógena, en la medida que el efecto de las variables que no están en el modelo (y que son representadas por el error) tienden a cancelarse entre si.

$$\mu_{\zeta i} = 0 \text{ para todo } i \qquad (1)$$

Una segunda presunción importante afirma que los errores de las diferentes ecuaciones no covarian con las variables exógenas. La razón principal por la que el error y las variables exógenas pueden covariar es que ambas tengan alguna causa previa que sea común. La presunción indica que no existen causas comunes omitidas a variables endógenas y exógenas.

$$\text{Cov}(\zeta_i , x_j) = 0 \text{ para todo } i , j \qquad (2)$$

La tercera presunción afirma que los errores no covarian. La interpretación de dicha covariación, en el caso de producirse, es esencialmente que se han olvidado variables que son causa común a las endógenas en la fase de especificación. No debe pensarse que habitualmente la varianza de un error sea cero, dado que esto implicaría que el error es cero o una constante, cosas bastante improbable. La media de un error si que puede ser cero, pero no su variación alrededor de la media.

$$\Psi_{ij} = 0 \text{ para todo } i \neq j \qquad (3)$$

Por último, una cuarta presunción plantea la posibilidad de que las variables exógenas, es decir, que no son explicadas dentro del modelo, puedan presentar covariación entre ellas.

$$\Phi_{ij} \neq 0 \text{ para todo } i,j \qquad (4)$$

Estas cuatro presunciones vienen a plantear las condiciones de funcionamiento del modelo, orientando a su vez sobre los posibles problemas que este muestra en su ajuste a los datos. Sin embargo, no es habitual que el sistema se formule para las variables expresadas en términos "brutos" sino que estas sufren una serie de trasformaciones. Como veremos la finalidad de estas trasformaciones es conseguir una mayor facilidad de estimación de parámetros así como mejorar la comparabilidad entre los coeficientes. A su vez, dichas trasformaciones dejaran su huella sobre las presunciones.

6.2.2. Transformaciones

La primera de las trasformaciones produce efectos interesantes en el sistema de ecuaciones. En primer lugar suprime el coeficiente constante (α) de la ecuación. Debemos considerar que el coeficiente constante es un parámetro a estimar y sin embargo, con frecuencia, es un mero apoyo matemático para ajustar la solución. De hecho, al expresar el valor de la dependiente para determinadas combinaciones de valores de las que la explican, puede estar asociada a una situación sin significado. Eliminarla no supone ningún problema porque puede recuperarse en caso de que se necesite. Otro efecto interesante es que las medias de las variables se hacen igual a 0. ($\mu y^d_i = 0$ y $\mu x^d_i = 0$). No obstante, no produce ningún efecto sobre los coeficientes, que permanecen expresados al igual que en la ecuación original. Para trasformar las variables calculamos su desviación a la media.

$$y^d_i = y_i - \mu y_i \text{ para todos los i}$$

$$x^d_i = x_i - \mu x_i \text{ para todos los i}$$

El impacto sobre la notación es una d como superíndice sobre las variables.

$$\mathbf{y^d_1 = \gamma_{11} x^d_1 + \gamma_{12} x^d_2 + \zeta_1}$$
$$\mathbf{y^d_2 = \beta_{21} y^d_1 + \gamma_{21} x^d_1 + \zeta_2}$$
$$\mathbf{y^d_3 = \beta_{31} y^d_1 + \beta_{32} y^d_2 + \gamma_{32} x^d_2 + \zeta_3}$$
$$\mathbf{y^d_4 = \beta_{41} y^d_1 + \zeta_4}$$

$$\mathbf{y^d_5 = \beta_{52} y^d_2 + \beta_{53} y^d_3 + \beta_{54} y^d_4 + \zeta_5}$$

Y una modificación en la primera presunción donde se indica que la media de todas las variables en la ecuación es igual a 0

$$\mathbf{\mu y^d_i = \mu x^d_i = \mu_{\zeta i} = 0 \text{ para todo } i} \quad (1)$$
$$\mathbf{Cov(\zeta_i , x^d_j) = 0 \text{ para todo } i , j} \quad (2)$$
$$\mathbf{\Psi_{ij} = 0 \text{ para todo } i \neq j} \quad (3)$$
$$\mathbf{\Phi_{ij} \neq 0 \text{ para todo } i,j} \quad (4)$$

Otra transformación muy frecuente consiste en normalizar las variables mediante la división de éstas, expresadas en desviación a la media por la desviación típica de la variable.

$$y^s_i = y^d_i / \sigma_{yi}$$

$$X^s_i = X^d_i / \sigma_{xi}$$

Es importante notar que la transformación mediante la división de las variables por la desviación típica afecta a los parámetros y a su interpretación. Así, para un coeficiente normalizado la interpretación de β^s_{ij} es que y^s_i cambiara β^s_{ij} desviaciones típicas cuando y^s_j cambie una desviación típica, para todas las demás variables permaneciendo sin cambios. Una interpretación equivalente para el caso de los coeficientes normalizados que expresan los efectos directos de las variables exógenas $\mathbf{y^s_{ij}}$. Así, una variable y^s_i cambiara $\mathbf{y^s_{ij}}$ desviaciones típicas cuando x^s_j cambie una desviación típica, para todas las demás variables permaneciendo constantes.

El sistema de ecuaciones se expresa, con notación simplificada para las variables normalizadas introduciendo un superíndice s en todas las variables y parámetros así como un apostrofe en el error.

$$\mathbf{y^s_1 = \gamma^s_{11} x^s_1 + \gamma^s_{12} x^s_2 + \zeta'_1}$$
$$\mathbf{y^s_2 = \beta^s_{21} y^s_1 + \gamma^s_{21} x^s_1 + \zeta'_2}$$
$$\mathbf{y^s_3 = \beta^s_{31} y^s_1 + \beta^s_{32} y^s_2 + \gamma^s_{32} x^s_2 + \zeta'_3}$$
$$\mathbf{y^s_4 = \beta^s_{41} y^s_1 + \zeta'_4}$$
$$\mathbf{y^s_5 = \beta^s_{52} y^s_2 + \beta^s_{53} y^s_3 + \beta^s_{54} y^s_4 + \zeta'_5}$$

En lo referido a las presunciones, es necesario añadir una quinta. Esta presunción indica que la variabilidad de las variables en el modelo (normalizadas) es igual a 1.

$\mu y^{s}_{i} = \mu x^{s}_{i} = \mu_{\zeta',i} = 0$ **para todo** i (1)

$Cov(\zeta'_{i}, x^{s}_{j}) = 0$ **para todo** i, j (2)

$\Psi'_{ij} = 0$ **para todo** $i \neq j$ (3)

$\Phi^{s}_{ij} \neq 0$ **para todo** i,j (4)

$\sigma_{x}{}^{s}{}_{i} = \sigma_{y}{}^{s}{}_{i} = 1$ **para todo** i (5)

Consideremos las ventajas y desventajas de las diferentes formas de expresar los coeficientes, normalizados o no. Una de las ventajas de emplear coeficientes no normalizados es que en el caso de aplicar el modelo a diferentes poblaciones, estos tenderán a ser los mismos, aún cuando la variabilidad interna de las variables no lo sea. Los coeficientes normalizados pueden cambiar más fácilmente (menos robustos al cambio) cuando se trata con poblaciones diferentes. Esto es debido a que los coeficientes normalizados son función de la desviación típica. Si varía la distribución típica de una variable al comparar poblaciones distintas, provocará cambios en los coeficientes inducidos por dichas diferencias en la variabilidad. En ese sentido, cabe recomendar el uso de los coeficientes sin normalizar para ajustar modelos sobre diferentes poblaciones.

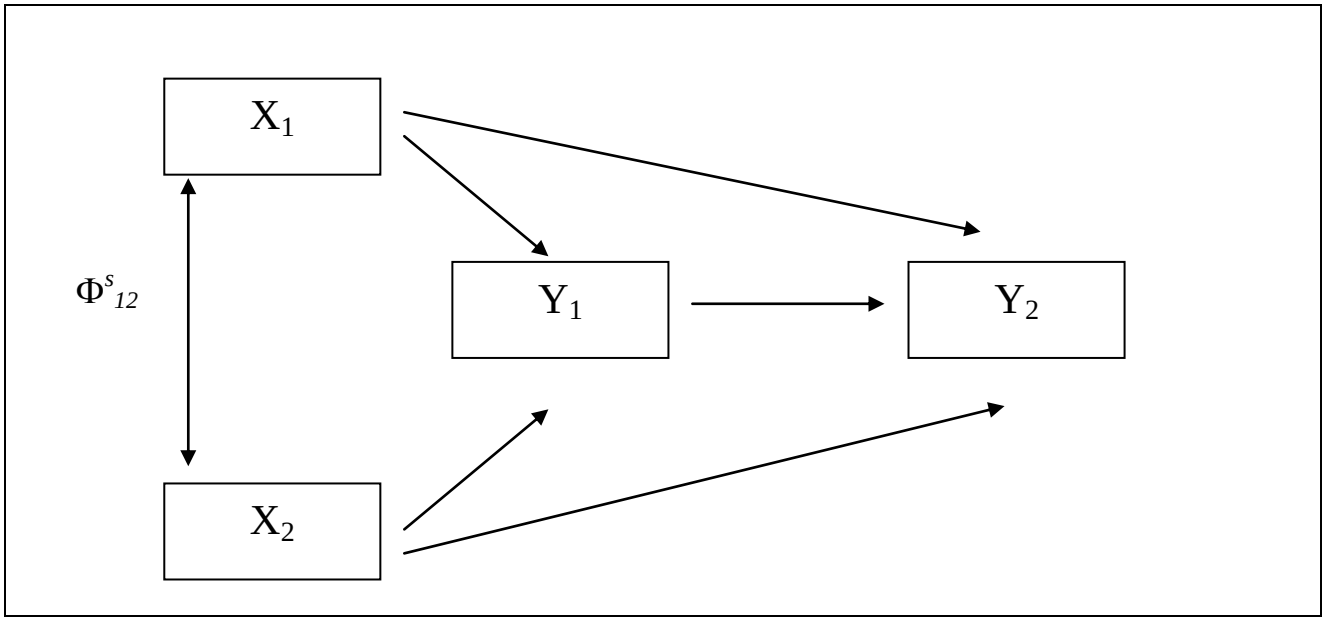
Por otra parte, si en el modelo se mezclan diferentes tipos de variables con rangos muy dispares y distintas escalas (tanto en el mismo modelo o para comparar entre modelos que provienen de diferentes investigaciones) será conveniente el empleo de coeficientes normalizados, dado que ello facilita la comparación entre modelos.

6.3. Parámetros teóricos y estimados empíricos

Se han definido los sistemas de notación de los modelos, así como los parámetros que los constituyen. Sin embargo, sobre la base de los datos solo nos es posible obtener coeficientes de covarianza o de correlación, varianzas, etc. y desde ellos debemos definir los diferentes parámetros y efectos. Para ello se definen dos reglas de descomposición que nos vinculan teóricamente dichos coeficientes y los parámetros del modelo.

Observemos el siguiente ejemplo. El modelo causal tiene asociada un diagrama causal, la matriz de correlaciones, un sistema de ecuaciones y sus presunciones estandarizadas.

diagrama causal



matriz de correlaciones

x_1	ρ_{x1x1}			
x_2	ρ_{x2x1}	ρ_{x2x2}		
y_1	ρ_{y1x1}	ρ_{y1x2}	ρ_{y1y1}	
y_2	ρ_{y2x1}	ρ_{y2x2}	ρ_{y2y1}	ρ_{y2y2}
	x_1	x_2	y_1	y_2

sistema de ecuaciones

$$y^S_1 = \gamma^S_{11}x^S_1 + \gamma^S_{12}x^S_2 + \zeta'_1$$

$$y^S_2 = \beta^S_{21}y^S_1 + \gamma^S_{21}x^S_2 + \gamma^S_{22}x^S_2 + \zeta'_2$$

presunciones estandarizadas

$\mu y^s_i = \mu x^s_i = \mu_{\zeta' i} = 0$ para todo *i* (1)
$Cov(\zeta'_i , x^s_j) = 0$ para todo *i , j* (2)
$\Psi'_{ij} = 0$ para todo $i \neq j$ (3)
$\Phi^s_{ij} \neq 0$ para todo *i,j* (4)
$\sigma_{x^s_i} = \sigma_{y^s_i} = 1$ para todo *i* (5)

6.3.1. Primera regla de descomposición

Es denominada así, dado que descompone la correlación observada entre variables en cuatro componentes de variación. Definición: el coeficiente de correlación entre dos variables es igual a la suma de los efectos directos, los efectos indirectos, las relaciones espurias y los efectos conjuntos.

La diagonal de la matriz de correlaciones no se ve afectada por esta primera regla. La correlación observada entre las variables predeterminadas es igual al parámetro que expresa su covariación.

$$\rho x^s_1 x^s_2 = \phi^s_{12}$$

Correlación entre x^s_1 e y^s_1 $(\rho y^s_1 x^s_1)$ es igual a un efecto directo (γ^s_{11}) más un efecto conjunto entre x^s_1 e x^s_2 $(\gamma^s_{12}\ \phi^s_{21})$ luego

$$\rho y^s_1 x^s_1 = \gamma^s_{11} + \gamma^s_{12}\,\phi^s_{21}$$

El último es un efecto conjunto dado que no sabemos si es un efecto indirecto a través de x^s_2 o espurio debido a x^s_2

$$\rho\ y^s_1 x^s_2 = \gamma^s_{12} + \gamma^s_{11}\,\phi^s_{21}$$

$$\rho\ y^s_2 x^s_1 = \gamma^s_{21} + \beta^s_{21}\,\gamma^s_{11} + \gamma^s_{22}\,\phi^s_{21} + \beta^s_{21}\,\gamma^s_{12}\,\phi^s_{21}$$

$$\rho\ y^s_2 x^s_2 = \gamma^s_{22} + \beta^s_{21}\,\gamma^s_{12} + \gamma^s_{21}\,\phi^s_{21} + \beta^s_{21}\gamma^s_{11}\,\phi s_{21}$$

$$\rho\ y^s_2 y^s_1 = \beta^s_{21} + \gamma^s_{21}\,\gamma^s_{11} + \gamma^s_{22}\,\gamma^s_{12} + \gamma^s_{22}\,\phi^s_{21}\,\gamma^s_{11} + \gamma^s_{21}\,\phi^s_{21}\,\gamma^s_{12}$$

6.3.2. Segunda regla de descomposición

La segunda regla de descomposición responde de la variabilidad apreciada en la diagonal de la matriz de correlación. Definición: la varianza total de una variable endógena es igual a la cantidad de varianza explicada por las variables causantes de dicha variable endógena, más una cantidad de varianza no explicada. En definitiva lo que se viene a afirmar es que la varianza de la variable endógena estandarizada es igual a la varianza explicada por las variables causales y a la varianza no explicada por estas. Dado que las variables están estandarizadas su varianza es igual a 1. Debemos recordar que la varianza de las variables predeterminadas no se explica desde otras variables contenidas en el modelo, Por lo tanto, la varianza observada en las variables predeterminadas es igual a la varianza del modelo. Así,

$$\rho\, x^s_1 x^s_1 = \phi^s_{11}$$
$$\rho\, x^s_2 x^s_2 = \phi^s_{22}$$

Por otra parte tenemos la varianza de las variables endógenas. La proporción de varianza explicada mediante un grupo de variables se denota

$$R^2 y_1 \bullet x_1, x_2$$

siendo R^2 el coeficiente de determinación, la primera variable aquella endógena que se desea explicar y separada de las demás por un punto.

$$\rho\, y_1 y_1 = 1 = \quad R^2_{y1\bullet x1,x2} + \psi'_{11}$$
$$\rho\, y_2 y_2 = 1 = \quad R^2_{y2\bullet y1,x1,x2} + \psi'_{22}$$

El coeficiente de determinación es una función de los parámetros del modelo estructural y no un nuevo parámetro del modelo. Se puede demostrar que "para cualquier variable endógena, la proporción de varianza explicada puede obtenerse sumando los productos de los efectos directos y los coeficientes de correlación entre la variable endógena y cada una de las variables causales que les afecta directamente".

$$R^2_{y1\bullet x1,x2} = \gamma^s_{11} \quad \rho_{y1x1} + \gamma^s_{12}\, \rho_{y1x2}$$
$$R^2_{y2\bullet y1,x1,x2} = \beta^s_{21} \rho_{y2y1} + \gamma^s_{21}\, \rho_{y2x1} + \gamma^s_{22}\, \rho_{y2x2}$$

La proporción de varianza no explicada es igual a la varianza de las perturbaciones estandarizada ψ'_{ii}

7. Tipologías de sistemas causales

Las diferentes tipologías de sistemas causales se establecen sobre la base de diferentes criterios que dan pie a conjuntos específicos de terminologías. No obstante tal y como advirtiera Bentler (1994), todas las tipologías se apoyan sobre la noción básica de un conjunto de ecuaciones estructurales lineales. Las variantes simplemente expresan las diferentes formas que este conjunto de ecuaciones adquiere en función a la finalidad de su utilización. Así, se diferenciaran entre sistemas recursivos o no recursivos en función a la direccionalidad del sistema según este totalmente ordenado o no. Uno de los aspectos principales de esta diferencia es el problemas de la identificación. Es decir, de la complejidad que puede suponer la resolución matemática del sistema.

El análisis estructural, puede emplearse combinando variables latentes, (del mismo modo que el análisis factorial), junto con otras variables dentro del modelo explicativo; así mismo, puede referirse a datos en un solo momento del tiempo o en varios (como en el análisis de panel) o en simulaciones mediante ecuaciones simultaneas, etc. En cualquiera de éstas formas de utilización, el elemento básico es la idea de estructura.

7.1. Modelos con variables latentes.

En el ámbito del análisis de senderos se determinó la posibilidad de análisis causales empleando variables latentes[1]. Veamos seguidamente un modelo simple se efecto entre variables latentes. El siguiente fué el primer modelo discutido en 1969 y donde se puede apreciar claramente su relación con los modelos causales construidos exclusivamente con variables manifiestas. En este modelo podemos apreciar dos variables latentes F1 y F2, donde F1 es la variable latente causa y F2 la variable latente efecto. El coeficiente que liga ambas variables causalmente es $\gamma_{21.}$

Podemos apreciar como la variable latente F_1 influencia las variables indicadoras x_1, x_2, x_3, mientras que la variable latente F_2 influye en las variables indicadoras y_1, y_2, y_3. Se han notado los errores ecuacionales como δ_1, δ_2, δ_3 el error para las variables x y ε_1, ε_2, ε_3 para las variables y. En este tipo de modelos la notación de las variables y de los

[1] Las variables latentes son denominadas factores en el ámbito de la psicología.

errores varia en la medida que se introduce el concepto de variable latente y variable indicador. En este modelo en particular, el interés esta centrado en la relación entre variables latentes y no entre los indicadores. Sin embargo, la correlación entre indicadores es la única información que poseemos empíricamente desde los datos. La relación entre variables latentes es algo que debemos estimar desde la correlación entre indicadores. Precisamente, este fué el descubrimiento importante desde el análisis de senderos: que es posible estimar la relación entre las variables latentes mediante la correlación conocida entre las variables indicadoras. El modo como se deriva el efecto γ_{21} es exactamente igual a como hemos considerado en las reglas de descomposición. En primer lugar, las correlaciones apreciadas entre las variables indicadoras se expresan en términos de los parámetros del modelo y en segundo lugar se estiman los parámetros empleando la información que les asocia a las correlaciones.

$$\rho_{y2y1} = \lambda^{s}_{y22}\lambda^{s}_{y12} \quad (3.1.1.)$$
$$\rho_{y3y1} = \lambda^{s}_{y32}\lambda^{s}_{y12} \quad (3.1.2.)$$
$$\rho_{x1y1} = \lambda^{s}_{x11}\gamma^{s}_{21}\lambda^{s}_{y12} \quad (3.1.3.)$$
$$\rho_{x2y1} = \lambda^{s}_{x21}\gamma^{s}_{21}\lambda^{s}_{y12} \quad (3.1.4.)$$
$$\rho_{x3y1} = \lambda^{s}_{x31}\gamma^{s}_{21}\lambda^{s}_{y12} \quad (3.1.5.)$$
$$\rho_{y3y2} = \lambda^{s}_{y32}\lambda^{s}_{y22} \quad (3.1.6.)$$
$$\rho_{x1y2} = \lambda^{s}_{x11}\gamma^{s}_{21}\lambda^{s}_{y22} \quad (3.1.7.)$$
$$\rho_{x2y2} = \lambda^{s}_{x21}\gamma^{s}_{21}\lambda^{s}_{y22} \quad (3.1.8.)$$
$$\rho_{x3y2} = \lambda^{s}_{x31}\gamma^{s}_{21}\lambda^{s}_{y22} \quad (3.1.9.)$$
$$\rho_{x1y3} = \lambda^{s}_{x11}\gamma^{s}_{21}\lambda^{s}_{y32} \quad (3.1.10)$$
$$\rho_{x2y3} = \lambda^{s}_{x21}\gamma^{s}_{21}\lambda^{s}_{y32} \quad (3.1.11)$$
$$\rho_{x3y3} = \lambda^{s}_{x31}\gamma^{s}_{21}\lambda^{s}_{y32} \quad (3.1.12)$$
$$\rho_{x2x1} = \lambda^{s}_{x21}\lambda^{s}_{x11} \quad (3.1.13)$$
$$\rho_{x3x1} = \lambda^{s}_{x31}\lambda^{s}_{x11} \quad (3.1.14)$$
$$\rho_{x3x2} = \lambda^{s}_{x31}\lambda^{s}_{x21} \quad (3.1.15)$$

Además de las ecuaciones anteriores es posible especificar seis más correspondientes a las varianzas de las variables indicadoras, descomponiéndolas en varianza explicada y varianza no explicada. No nos ocuparemos de dichas ecuaciones en la medida en que el efecto que nos ocupa teóricamente es el que liga las dos variables latentes F_1 y F_2 (γ_{21}). En resumen, tenemos 15 ecuaciones especificadas y solo 7 parámetros desconocidos. En estos términos podemos concluir que la condición necesaria para la identificación del sistema se cumple. También se cumplen la condición necesaria. Es evidente que con la información facilitada por los datos (en forma de correlaciones) y por las ecuaciones (3.1.1) (3.1.2.) (3.1.6) es posible estimar los coeficientes $\lambda^{s}_{y12,}\lambda^{s}_{y22,}\lambda^{s}_{y32.}$ Si consideramos la razón entre ρ_{y2y1} y ρ_{y3y1}

$\rho_{y2y1} / \rho_{y3y1} = \lambda^{s}_{y22} / \lambda^{s}_{y32}$

luego

$\lambda^{s}_{y22} = \lambda^{s}_{y32} (\rho_{y2y1} / \rho_{y3y1})$

sustituyendo en la ecuación (3.1.6) obtenemos la solución para λ^{s}_{y32}

$\lambda^{s}_{y32} =$ raíz $(\rho_{y3y1} \rho_{y3y1}) / \rho_{y2y1}$

habiendo determinado λ^{s}_{y32} los otros coeficiente pueden hallarse por sustitución en las ecuaciones anteriores. Aplicando el mismo procedimiento pueden estimarse los coeficientes $\lambda^{s}_{x11,} \lambda^{s}_{x21,} \lambda^{s}_{y31.}$

Tras determinar los coeficientes que ligan las variables indicadoras a las variables latentes, aún nos quedan 9 ecuaciones para determinar el valor de $\gamma_{21.}$

De este modo, el parámetro puede resolverse desde 9 ecuaciones distintas. Podemos apreciar como una vez establecido el mecanismo causal que da forma a las variables latentes es posible determinar la influencia de unas sobre otras. En este caso es incluso posible testar el modelo teórico propuesto dado que quedan 8 grados de libertad. Evidentemente, el modelado con variables latentes se ha desarrollado implicando modelos más complejos, donde se combinan variables latentes, indicadoras y manifiestas.

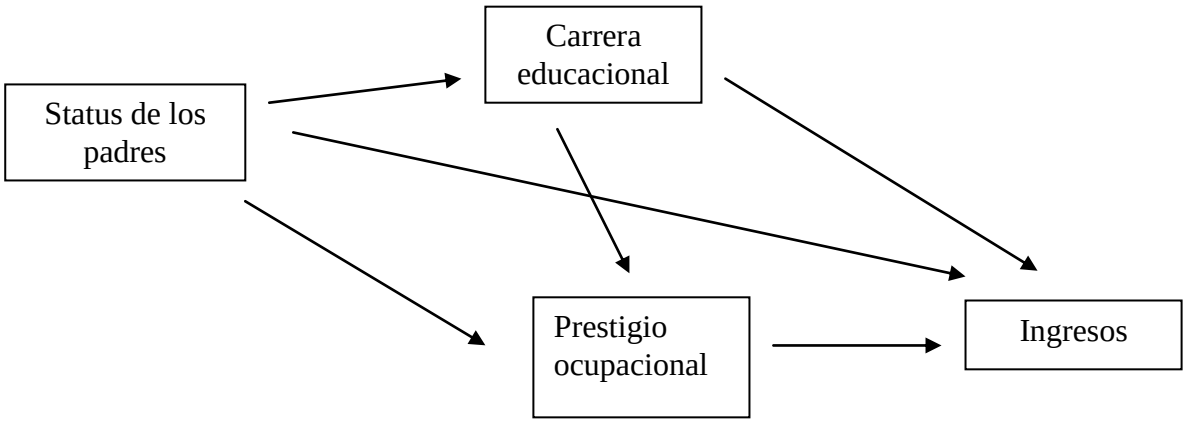
7.2. Los modelos recursivos y no recursivos

En general, podemos considerar una distinción importante entre dos tipos de sistemas, los sistemas recursivos y los no recursivos. Los modelos recursivos son aquellos modelos causales en los que todos los efectos causales se establecen en una sola dirección; es decir, se determinan relaciones asimétricas unidireccionales (y donde el error o perturbaciones está incorrelacionado entre las diferentes ecuaciones). Es decir, un modelo recursivo será a) jerárquico, donde todas las variables en el modelo pueden ser ordenadas y etiquetadas en una secuencia y_1, y_2, y_3, y_4…, y_n de tal modo que para todo y_i e y_j donde i<j y_j no se presenta como causa de yi. Por lo tanto β_{ij} será igual a cero. Según esto, la primera variable endógena solo podrá ser influida por una variable exógena. La segunda endógena solo podrá ser influida por una exógena o la endógena anterior y así sucesivamente. Según este criterio de jerarquía, en un modelo recursivo no pueden aparecer relaciones reciprocas entre dos variables ni puede pasar que una variable endógena pueda influir mediante un efecto indirecto sobre otra anterior. B) los errores deben de estar incorrelacionados entre si y con las variables exógenas. Esta

característica permite el estimar los coeficientes mediante Mínimos Cuadrados Ordinarios de forma insesgada y consistente[2]. En ese sentido, los modelos causales recursivos son fáciles de estimar. No obstante, en muchas ocasiones ambas presunciones son poco realistas. Con frecuencia, en muchos análisis es dudoso que las presunciones sean apropiadas. Por ello, no debe optarse por un modelo recursivo a la ligera, por comodidad o por conveniencia. A menos que se este perfectamente convencido de que las relaciones son estrictamente unidireccionales (jerárquicas) y que los factores (o variables no incluidas en el análisis) que están contribuyendo al error de cada ecuación son distintos para cada ecuación (no hay factores que influyan en común sobre ambas ecuaciones) no debe optarse por un modelo recursivo. El problema no debe ser de comodidad sino de acierto en la descripción completa y realista de un fenómeno social. Consideremos que si las presunciones no son ciertas (jerarquía e independencia de los errores) los estimados de los coeficientes (mediante Mínimos Cuadrados Ordinarios, OLS) serán inconsistentes y sesgados, con lo cual no solo no habremos esclarecido nada, sino que lo habremos oscurecido.

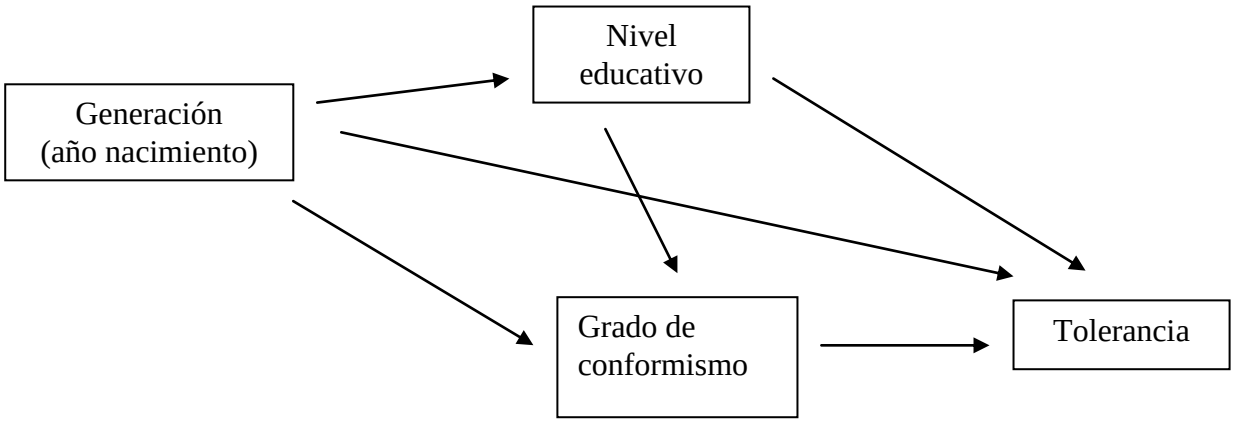
Modelos recursivos

a) estatus socioeconómico



[2] El término *insesgado* se refiere a aquel estimado que, como media, es igual al valor real del parámetro. Por otra parte, el término consistente se refiere a aquel estimado que, cuando la muestra se aproxima a infinito, la distribución del estimado se aproxima a una distribución con la mayor probabilidad de estar centrada sobre el parámetro.

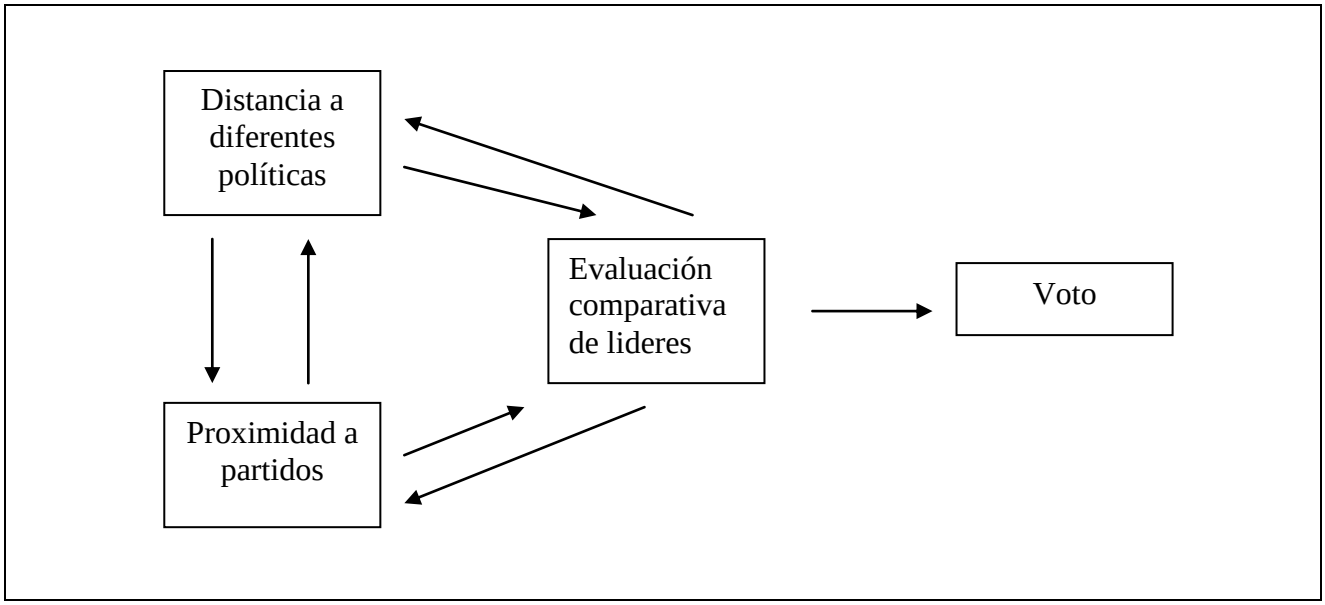
b) Tolerancia a lo distinto



Los modelos no recursivos, por el contrario, postulan la posibilidad de efectos recíprocos, o con carácter más general, que se produzcan efectos en ambas direcciones dentro del sistema. Un caso limite de no-recursividad lo plantea los modelos completamente no recursivos. En un modelo completamente no recursivo, todas las variables endógenas se ven afectadas por todas las demás variables endógenas y exógenas presentes en el modelo. No obstante, independientemente de su utilidad para la investigación no es conveniente definir modelos causales completamente no recursivos dado que dichos modelos son siempre subidentificados. Por el contrario, alguno de los parámetros del modelo no recursivo se supone que es igual a cero. Recordemos que un parámetro fijado a cero implica que hemos postulado que no existe un efecto entre dos variables. Como tendremos ocasión de comprobar cuando se considere el problema de la identificación, las presunciones que se adopten en el modelo recursivo serán de gran importancia para sus posibilidades de identificación. En general, las presunciones que empleemos serán que la media de las variables y los errores serán igual a cero (transformación mediante desviación a la media) y que los errores están incorrelacionados de las variables independientes. En un modelo no recursivo no tiene mucho sentido plantear que todos los errores están incorrelacionados de todas las variables endógenas. Siempre hay algún error que estará relacionado con alguna variable endógena, por el mismo planteamiento del modelo. Por el contrario, la presunción útil y que puede tener sentido teórico en un modelo no recursivo es que los errores están incorrelacionados entre si. Veamos los ejemplos siguientes, reflejados mediante diagramas basados en grafos orientados.

Modelos no recursivos

a) Comportamiento electoral (Page y Jones)



Podemos apreciar que mientras en los modelos recursivos la explicación esta ordenada de forma asimétrica en una sola dirección, en los modelos no recursivos, aparecen relaciones que invierten el orden de la causalidad, estableciendo relaciones reciprocas. Esta distinción es especialmente eficaz en términos de identificación del sistema, es decir, esencialmente técnicos en tanto permite o no tener soluciones. Desde el punto de vista de la explicación es evidente que los modelos causales no recursivos son bastante más realistas que los modelos recursivos. No obstante, los problemas que plantean en términos de identificación los hace bastante poco frecuentes

7.2.1. Identificación en modelos recursivos y no recursivos

El concepto de identificación, está ligado a las operaciones matemáticas que se realizan para efectuar el ajuste del modelo sobre los datos de que se disponen. En función al estado de identificación del modelo podrá o no tener un conjunto de soluciones que sean operativas para el investigador. De este modo, podremos afirmar que una ecuación esta identificada (y un modelo causal en general) cuando sus parámetros se pueden determinar de modo único a partir del conocimiento que se puede extraer de un conjunto de observaciones completas y adecuadas. Lo primero que debe destacarse es que el problema

de la identificación del sistema no es un problema de inferencia estadística. Un modelo no tendrá problemas de identificación por más inestable que sea la muestra que facilita la información para ajustar el modelo. El problema de la identificación se refiere a la relación entre información y parámetros a estimar. Se trata en definitiva de poseer mas hipótesis que información para testarlas. En resumen, la identificación del sistema no es un concepto que este relacionado con la calidad de los datos o la medición. Incluso con los mejores datos, es decir, con indicadores válidos y fiables procedentes de una gran muestra puede surgir el problema de la identificación. La identificación esta directamente relacionada con la especificación del sistema, es decir, con las relaciones que planteamos que existen a efectos de explicar un fenómeno social.

Podemos efectuar un planteamiento intuitivo desde el álgebra mediante el examen de un sistema de ecuaciones. Básicamente, la cuestión de la identificación se refiere a tener la suficiente información para obtener una solución única a un conjunto de incógnitas. Así, por ejemplo:

a) Identificación exacta. El siguiente sistema de ecuaciones

$$2x + 3y = 7$$
$$x - 4y = -2$$

Constituye un sistema exactamente identificado, dado que hay tantas ecuaciones linealmente independientes entre si, como incógnitas. Así, obtenemos una solución única donde

$$x = 2 \quad e \quad y = 1$$

b) Subidentificación

La subidentificación aparece cuando poseemos más incógnitas que ecuaciones linealmente independientes entre si. Por ejemplo, el sistema de ecuaciones

$$2x + 3y = 7$$
$$4x + 6y = 14$$

está subidentificado, dado que aún cuando hay dos ecuaciones con dos incógnitas, la segunda es simplemente la primera multiplicada por dos (son linealmente dependientes). Es decir, al ser la segunda ecuación simplemente la primera multiplicada por dos, no aporta ninguna información nueva que ayude a resolver de un modo único las incógnitas x e y. De

hecho, solo tendremos una ecuación con dos incógnitas, lo que lleva a un conjunto infinito de soluciones. Por ejemplo, las siguientes pueden ser soluciones al sistema anterior.

$$x = 2 \qquad y = 1$$
$$x = 3{,}5 \qquad y = 0$$
$$x = 5 \qquad y = -1$$

Todas ellas son soluciones para el sistema anterior. Como nuestra intención es obtener unos estimados con significado teórico para ese conjunto de incógnitas, la existencia de infinitas soluciones es una situación indeseable.

c) Sobreidentificación

Una situación semejante puede aparecer cuando poseemos un numero mayor de ecuaciones que de incógnitas. Por ejemplo, el sistema de ecuaciones linealmente independiente que mostramos a continuación posee dos incógnitas y tres ecuaciones.

$$2x - y = 7 \quad (1)$$
$$x + 3y = 0 \quad (2)$$
$$3x - 2y = 2 \quad (3)$$

Si se emplearan las ecuaciones (1) y (2) para resolver el sistema, obtendremos una solución única para ese sistema de dos ecuaciones de

$$x=3 \text{ e } y= -1$$

Las ecuaciones (1) y (3) dan como resultado

$$x=12 \text{ e } y=17$$

Las ecuaciones (2) y (3) ofrecen como resultado las soluciones

$x=6/11$ e $y=-2/11$

El término identificación y cada uno de los estados posibles (sub, exacta y sobre) se refieren tanto a las ecuaciones estructurales por separado como al conjunto del sistema de ecuaciones. Así, diremos que un modelo causal, esta identificado si todas y cada una de las ecuaciones que lo componen están identificadas. Por el contrario diremos que un sistema no esta identificado, cuando alguna de sus ecuaciones este subidentificada. Como podemos apreciar, la solución del sistema depende de la relación entre información e incógnitas. En el caso de la subidentificación, no tendremos solución posible (cualquier solución será indeterminada), en el caso de identificación exacta tendremos la posibilidad de estimar los parámetros mediante una solución única. Sin embargo, la situación más interesante se produce en caso de la sobre identificación. En este caso, como podremos apreciar, se presenta la posibilidad de testar la bondad del ajuste del modelo.

. Seguidamente vamos a considerar en primer lugar algunos criterios para identificar el estado del sistema de ecuaciones, para después plantear las posibles alternativas de los modelos no identificados. Evidentemente, en la medida que el problema de la identificación es un problema de especificación, solo una reelaboración de la explicación (es decir del modelo y de la teoría) puede ofrecer soluciones.

7.2.2. La determinación del estado

El problema de la identificación tiene consecuencias diferentes según se trate de sistemas recursivos o no recursivos. Como veremos, en el caso de los sistemas recursivos existe siempre la posibilidad de establecer restricciones que permitirán siempre identificar (y solucionar) el sistema. No es éste el caso de los sistemas no recursivos donde en determinadas situaciones su identificación requerirá necesariamente la modificación de este (introduciendo nuevas variables o restricciones de coeficientes o covarianzas). Podemos preguntarnos que es lo que hace a los modelos no recursivos especiales en términos de identificación. En principio, de forma intuitiva podríamos pensar que contando con suficiente datos el sistema debería tener solución. Sin embargo consideremos el ejemplo siguiente:

ζ_1 ζ_2

$$Y_1 = \beta_{12}Y_2 + \zeta_1$$
$$Y_2 = \beta_{21}Y_1 + \zeta_2$$

Este modelo no esta identificado; pero esto es evidente, en la medida que es imposible determinar en que sentido se desplaza la causalidad (cuando solo tenemos datos referidos a un solo punto en el tiempo). Así, solamente con el dato de la covariación entre ambas variables no existe ninguna forma matemática de distribuir cuanta covarianza corresponde al efecto de y_1 sobre y_2 y cuanta corresponde al efecto inverso, de y_2 sobre y_1. Esto puede pasar perfectamente en un modelo no recursivo más complejo. Además, en relación con los modelos recursivos, en un modelo no recursivo existen en general más parámetros a estimar incluso poseyendo el mismo numero de variables. Otro aspecto que influye en el problema de la identificación de modelos no recursivos es el hecho de no postular que los errores son independientes entre si. Por lo tanto, los criterios de identificación que introduciremos seguidamente son especialmente pertinentes en el caso de los modelos no recursivos.

Vamos a considerar tres criterios de evaluación del estado del modelo. El primero de ellos va considerar el sistema en conjunto y por lo tanto aportara un diagnostico global. En estas condiciones se tiene poca información para intervenir sobre el modelo, si bien es un procedimiento rápido de diagnostico. Una mayor utilidad a efectos de intervenir en el caso de no identificación del sistema son los procedimientos que evalúan el estado de cada una de las ecuaciones del sistema. De este modo, identificando las ecuaciones problemáticas es posible intervenir sobre las relaciones de forma que se posibilite la identificación del sistema. En los modelos no recursivos se emplearan dos medios para evaluar las posibles restricciones de coeficientes, las condiciones de rango y las condiciones de orden. Como se ha dicho, estos procedimientos operan evaluando cual es la situación de cada ecuación; esto viene dado porque en un modelo podrían existir ecuaciones subidentificadas, junto a otras identificadas exactamente y otras sobreidentificadas. El poder detectar cual es la situación de cada ecuación dentro del sistema ayuda claramente en el procedimiento de identificación global del sistema de ecuaciones. El procedimiento que evalúa directamente el sistema de ecuaciones en conjunto no ofrece una orientación con respecto al modo como corregir el sistema de modo que, como mínimo, se determine un conjunto finito de soluciones para las incógnitas (parámetros) a estimar. Por ultimo recordar que la identificación, en la medida que depende de la especificación, se vera afectada por las presunciones sobre el error. Aquí consideraremos, tal como se advirtió inicialmente, que las

medias de los errores y las variables es cero (desviaciones o normalización) y que los errores son independientes de las variables exógenas (es decir, no covarian). En los sistemas donde se planteen otras presunciones la identificación por los siguientes procedimientos puede verse afectada. Los dos primeros procedimientos son condiciones[3] necesarias pero no suficiente. El ultimo procedimiento es condición suficiente.

7.2.3. Identificación del sistema

Como sabemos, el problema que queremos solucionar es si los parámetros estructurales de un modo pueden ser determinados de forma única sobre la base de la información que se disponga de varianzas y covarianzas entre las variables observadas. Una regla general en álgebra es que una condición necesaria para resolver las incógnitas en un sistema de ecuaciones es que el numero de incógnitas debe ser igual o inferior que el numero de ecuaciones (linealmente independientes entre si, claro esta). Las incógnitas, en este caso, son los parámetros estructurales. Es posible determinar el numero de ecuaciones (en términos de descomposición de efectos, es decir las varianzas y las covarianzas por un lado y por el otro los parámetros). Las ecuaciones, insistimos, se refieren a las correspondientes a la relación entre parámetros y varianzas y covarianzas. En ese sentido, es fácil apreciar que tendremos tantas ecuaciones como varianzas y covarianzas. Así, si en un modelo tenemos 4 variables (tanto exógenas como endógenas) el numero de ecuaciones será igual a 1/2n(n+1), siendo n el numero de variables. 1/24(4+1) = a 6 ecuaciones. La diferencia entre el numero de ecuaciones y el numero de parámetros estructurales a estimar se denomina grados de libertad y se notan como ***df*** . Una vez definidos estos términos, el criterio para identificar el sistema puede formularse como sigue:

Una condición necesaria para la identificación de un modelo de ecuaciones estructurales es que los grados de libertad deben ser iguales o mayores que cero, es decir $df \geq 0$. Los grados de libertad resultan de comparar la información de que se dispone (varianzas y covarianzas) con los parámetros del modelo que deben estimarse.

La forma de contabilizar el número de ecuaciones (varianzas y covarianzas) es directa, contabilizando las variables exógenas y endógenas del modelo, dividiendo por dos y multiplicando por el número de variables mas uno

½ n(n+1).

---

[3] Necesaria pero no suficiente. Quiere decir que si no se cumple esa condición la ecuación no puede ser identificada. Si la condición se cumple, la ecuación puede o no puede ser identificada, pero existe la posibilidad.

Sin embargo, el numero de incógnitas puede ocasionar dudas, dado que depende de la especificación del modelo. En principio, dado que consideramos las variables expresadas en desviación a la media o normalizadas, la constante α desaparece de la ecuación eliminando una incógnita a estimar. Sin embargo permanecen como incógnitas los parámetros $\beta, \gamma, \psi, \phi$. El numero de parámetros β, γ pueden determinarse directamente de las ecuaciones o de los diagramas. El número de parámetros (correspondientes a las varianzas y covarianzas de las variables exógenas) ϕ es igual a $\frac{1}{2}$ $q(q+1)$, siendo q el numero de variables exógenas (x). El numero de parámetros ψ es como mínimo p, siendo p el numero de varianzas de los errores. El total que resulta de sumar los parámetros anteriores expresa el numero de incógnitas a resolver. Es decir, que trabajando con variables expresadas en desviación sobre la media solo se trata de contar los parámetros $\beta, \gamma, \psi, \phi$.

7.2.4. Condiciones de orden

Técnicamente, la condición de orden es una condición necesaria, pero no suficiente, para la identificación de una ecuación. Sin embargo, en muchas de las situaciones que se producen en la práctica al analizar datos, esta condición funciona como necesaria y suficiente. La condición de orden afirma que si tenemos un modelo consistente en K ecuaciones lineales, para que cualquier ecuación en el modelo este identificada debe de excluir como mínimo un numero de variables igual (o mayor) a K-1, de entre todas las variables que aparecen en el modelo.

Por ejemplo, en el caso que un sistema posee 12 ecuaciones y 15 variables, para que una ecuación cualquiera este identificada debe de excluir 11 variables (12-1) de entre todas las que aparecen en el modelo. Es decir, las ecuaciones identificadas deben de excluir 11 variables (sus coeficientes = 0) y retener 4 variables (con coeficientes $\neq 0$).

7.2.5. Condiciones de rango

La condición de rango afirma (Christ,1966) que una ecuación en un modelo de K ecuaciones lineales esta identificada si existe un determinante de cualquier submatriz de coeficientes K-1 dentro de la matriz que resta después de omitir todas las columnas donde la ecuación a identificar posea coeficientes distintos de 0 y omitiendo la ecuación a identificar. El proceso se repetirá hasta identificar cada ecuación.

-Si no existiese ninguna submatriz de rango K-1 con determinante $\neq 0$ la ecuación esta subidentificada.

-Si existe solo una submatriz de rango k-1 con determinante $\neq 0$ la ecuación está determinada exactamente.

-Si existe más de una submatriz de rango k-1 con determinante $\neq 0$ la ecuación está sobreidentificada.

La condición de rango es una condición necesaria y suficiente para identificar una ecuación.

Una vez considerados diferentes procedimientos para comprobar el estado del sistema de ecuaciones, podemos ofrecer algunas conclusiones generales que servirán para diagnosticar en que status se encuentran en la práctica los modelos.

(1) Los modelos de una sola ecuación estructural donde el error no covaria (son independientes) con las variables exógenas están siempre identificados. Son conocidos como modelos de regresión. ($\sigma_{\zeta ixj} = 0$)

(2) Los modelos de ecuaciones sin efectos causales recíprocos y con las presunciones $\sigma_{\zeta ixj} = 0$ para todo i,j y que $\sigma_{\zeta i\zeta j} = 0$ para todos los $i \neq j$, siempre están identificados. Son denominados modelos recursivos.

(3) Los modelos donde existen efectos de xi sobre yj, existiendo covariación entre las variables exógenas y el error no, están identificados. $\sigma_{\zeta ixj} \neq 0$.

(4) Los modelos estructurales con efectos causales recíprocos (modelos no recursivos) no estarán identificados en el caso particular en que un conjunto de variables endógenas se vean afectadas todas ellas entre si.

Esencialmente, las conclusiones que pueden extraerse de las observaciones anteriores es que todos los modelos recursivos de ecuaciones estructurales (1) (2) están identificados siempre que las variables importantes se encuentren presentes en el modelo. Es decir, que la especificación sea la correcta. Es fundamental que todas las variables importantes estén en el modelo como garantia de que las presunciones se podrán cumplir; es decir, no covariaran las variables exógenas con el error y los errores estarán incorrelacionados entre si.

La conclusión tercera expresa la importancia de la presunción acerca de que las variables exógenas no deben covariar con el error. Como sabemos, la covariación entre las variables exógenas y el error indicara la existencia de causas comunes omitidas. Cuando exógena y error covarian puede ignorarse tal covariación, pero la consecuencia será que los parámetros γ estimados serán erróneos. La otra opción es introducir la covariación entre

exógena y error en el modelo, entre las presunciones, pero entonces es probable que el modelo se convierta en subidentificado. En definitiva, todo apunta al hecho de que la incorporación de causas comunes es realmente vital para la consistencia explicativa y matemática del modelo. Por ultimo, la cuarta conclusión indica en que condiciones un modelo no recursivo no puede ser identificado.

Como hemos podido apreciar, la identificación aparece como un problema especialmente en los sistemas no recursivos. En todo caso, podemos plantear algunas orientaciones para atenuar los problemas de identificación y sabiendo de antemano que restaran modelos matemáticamente no identificables. Dado que las condiciones de orden y de rango nos indican si una ecuación está subidentificada, identificada exactamente o sobreidentificada, el problema consiste en como actuar sobre las ecuaciones que plantean problemas.

7.3. Los procedimientos de restricción

Existen dos procedimientos básicos para intentar que un sistema de ecuaciones este identificado, las restricciones de coeficientes y las restricciones de covarianzas. Un tercer procedimiento consiste en la introducción de nuevas variables explicativas en el modelo. Las restricciones de coeficiente actúan imponiendo limitaciones sobre los coeficientes que unen las variables medidas. Ya sean fijándolos a cero, etc. Por su parte, las restricciones de covarianza efectúan presunciones sobre la correlación entre las variables residuales.

En los modelos recursivos la identificación es más simple dado que, por ejemplo, en este caso la mitad de los coeficientes son igual a cero (dado que no hay efectos recíprocos). Así, en un modelo recursivo afirmar que existe una relación entre Y1 e Y2 implica que el efecto inverso no se va a dar. Además, sabemos que en los modelos recursivos, se efectúan presunciones sobre el error que si bien no son realistas, sí se corresponden con la estructura teórica del modelo que se propone (asimétrico). En ese sentido, efectuando las presunciones habituales en un modelo recursivo tendremos garantía de que estará identificado (Boudon, 1968).

Por ejemplo consideremos un sistema no recursivo con tres variables con efectos recíprocos entre ellas. De acuerdo al criterio de sistema ½ 3(3+1), obtenemos 6 ecuaciones. Por otro lado tenemos 9 incógnitas, compuestas por 6 coeficientes y 3 errores. Tendríamos más incógnitas que ecuaciones. Este es, como sabemos, un caso evidente de subidentificación o falta de información. Una forma de solución es mediante restricciones de coeficientes y de covarianzas. Si lo convertimos en un modelo recursivo fijaremos tres coeficientes a 0. Contando con la presunción recursiva donde la covarianza de las tres

variables residuales están incorrelacionados, tendremos finalmente seis ecuaciones con seis incógnitas. Tendríamos con ello una identificación exacta. Como ya sabemos, el planteamiento de un modelo recursivo es correcto siempre que tengamos seguridad de que las causa comunes han sido incluidas en el. En ese sentido, la especificación del modelo es una fase especialmente ligada a la verosimilitud y fiabilidad de los coeficientes estimados finalmente. Es decir, de la fiabilidad del modelo.

En los modelos no recursivos, por el contrario, la situación se complica en la medida que la inclusión de todas las causa comunes no garantiza la identificación del modelo, y por lo tanto su resolución. Cuando estamos considerando un modelo no recursivo no es posible efectuar las restricciones de los modelos recursivos. En este tipo de modelos no son practicables las presunciones que establecíamos en los modelos recursivos, acerca de las covarianzas entre las variables residuales. Ello convierte los modelos no recursivos en modelos que se aproximan más a la realidad, dado que no presumen el que las variables residuales estén incorrelacionadas. Sin embargo, al eliminar esa restricción sobre las covarianzas de los errores se complica la tarea de la identificación. En los modelos no recursivos imponemos menos restricciones sobre los coeficientes y covarianzas, lo que conlleva un número mayor de incógnitas y a una mayor dificultad para obtener soluciones únicas. Además, el problema más frecuente se refiere a la situación donde las variables que explican en cada ecuación a las distintas endógenas tienden a repetirse en las diferentes ecuaciones.

El modo para intentar identificar (y que por lo tanto tenga solución) los modelos no recursivos pasa por aplicar las condiciones de orden y de rango de forma que se pueda identificar las ecuaciones infraidentificadas. Cuando una ecuación esta subidentificada no existe ninguna técnica de estimación que ofrezca estimados validos. Por ello, hay que intentar transformar una ecuación subidentificada en otra identificada, generalmente introduciendo nuevas variables en el modelo. Estas variables nuevas a introducir en el modelo deberán afectar (explicar) solo a determinadas variables (con ecuación infraidentificada). En ese sentido, la identificación mediante la introducción obligatoria de nuevas variables y condicionadas a una relación concreta supone en la mayoría de los casos una cierta violencia y forzamiento teórico del modelo. Por ello, aún cuando las modificaciones del modelo vengan impuestas desde la necesidad de identificación, la introducción de nuevas variables debe de estar, en primer lugar, teóricamente orientada. Es la teoría la que debería tener la última palabra en el sentido de indicar si es posible introducir nuevas variables, cuales deban de ser estas, así como su relación con las variables endógenas del modelo.

Esto último es una cuestión importante, dado que no por el hecho de introducir nuevas variables se va a facilitar la identificación del sistema de ecuaciones, sino que esto dependerá de las pautas de asociación propuestas para las nuevas variables. Una asociación u otra facilitaran la identificación o no.[4]

En una segunda instancia, es conveniente que esas nuevas variables posean determinadas propiedades estadísticas, algunas de las cuales son consecuencia directa de la sensatez teórica. En primer lugar, es conveniente que las nuevas variables sean variables exógenas, y no correlacionadas con el error de las variables endógenas. Además, deben de estar fuertemente asociadas con aquellas variables a las que están afectando teóricamente (Fisher,1971). La búsqueda de variables exógenas (predeterminadas) con dichas características no siempre es fácil. Las alternativas son, desfigurar el modelo explicativo o abandonar cualquier esperanza de solución. En cualquier caso, la introducción de nuevas variables aparece como alternativa a la supresión de efectos (coeficientes = 0). En principio no deberían suprimirse relaciones entre variables que supongan una especificación importante del modelo. Especialmente porque la supresión de efectos importantes, si realmente los son, puede sesgar la fiabilidad de los demás parámetros dentro del modelo.

Como podemos apreciar, las condiciones que la de identificación impone sobre el modelo explicativo son bastante importantes. En el caso de los modelos recursivos, porque presume condiciones drásticas de jerarquía y de completitud de la especificación (incluyendo todas las causas comunes importantes). En el caso de los no recursivos, imponiendo la introducción de nuevas variables y además en una función relacional obligada, afectando a determinadas variables y no a otras. Por otro lado, la supresión de coeficientes (mediante la fijación de los efectos a cero) que fueron introducidos previamente en la fase de especificación del modelo implica la amenaza de sesgar los resultados estimados. Como puede verse, no son despreciables las consecuencias de la identificación (relación información e incógnitas) en la explicación que se pretende ofrecer.

El dilema es evidente, mantener una explicación que no podrá ser testada o degenerar, por imposición matemática, el modelo explicativo en función a sus posibilidades de solución. No se trata de modificaciones introducidas por el ajuste del modelo, donde las covarianzas encontradas en la estructura de los datos (entre errores y entre variables)

---

[4] Muy probablemente, el principio de parsimonia haya establecido teóricamente la conveniencia de simplificar el modelo. Puede ser conveniente a efectos de la identificación del sistema de ecuaciones recuperar variables interesantes, pero descartadas por ese criterio de simplificación.

imponen una revisión de lo que se pensaba, sino modificaciones conducidas por la mecánica interna del modelo propuesto. No es demasiado atractivo que la técnica de modelado de la realidad (explicación de esta) determine las características finales de esta explicación. Evidentemente, las trasformaciones del modelo explicativo son un aspecto crucial de la tarea de investigar. Estas deberán desarrollarse siempre que sea teóricamente aceptable en el caso de sistemas subidentificados. No seria aceptable que algo tan importante como es una explicación de los fenómenos sociales se vea sesgada por la necesidad de modificarla a efectos de ser solucionable. La prioridad debe ser siempre la mejor explicación, no la explicación que mi método de análisis de la realidad me ha permitido o me ha obligado a producir.

Este es un fenómeno que supone un riesgo evidente, en la medida que la dinámica de modelado te conduce fuera de la explicación a un terreno donde las reglas de juego las imponen las matemáticas. No debería actuarse con timidez o complacencia y una opción a plantearse seriamente, dependiendo de las condiciones teóricas que imponga la identificación del sistema, seria optar por otra estrategia de modelado que permita vías alternativas de testar la explicación. La subidentificación supone riesgos teóricos importantes, donde una de las principales ventajas es una nueva oportunidad para repensar el modelo (la explicación que se ofrece).

Cuando las ecuaciones estructurales presentan una identificación exacta o sobreidentificación no existe problema en términos de la especificación del modelo, de modo que este (la explicación que se propone) no se vera modificado por las condiciones matemáticas el sistema (relación incógnitas / información). En ambas condiciones del sistema, la sobre identificación ofrece posibilidades importantes. Un riesgo evidente de un sistema exactamente identificado es que alguno de los efectos propuestos sea cero, con lo que la identificación se ve amenazada. Por el contrario, en los modelos sobreidentificados este riesgo no se da. Por otra parte, para la solución del sistema se requiere igual numero de ecuaciones (información) que de incógnitas (coeficientes), ello hace que la información (ecuaciones) extras no utilizadas puedan emplearse para testar el modelo.

7.4. Testado de modelos causales.

Como indicábamos el exceso de información supone una oportunidad especial para testar el modelo causal. Para ello debemos considerar que el modelo se ajusta sobre un conjunto de datos y la matriz relacional que estos datos ofrecen. El modelo, realmente, aspira a reproducir dicha estructura relacional empírica pero en el contesto de una estructura

teórico explicativa. Es decir, el modelo opera aspirando a traducir una estructura relacional empírica (generada por una definición previa y una selección de lo que es importante para definir el fenómeno en estudio) en una estructura relacional de conceptos vinculados por una argumentación explicativa. En este contexto, un aspecto muy importante ligado al concepto de identificación es la posibilidad de testar el modelo; testar el modelo consiste en comparar la estructura empírica que ha sido integrada en el modelo con la estructura empírica original y sobre la que éste se apoyo.

Una situación adecuada para este test de reproducibilidad de la matriz de relaciones empíricas aparece con los sistemas sobreidentificados. Como se trató previamente, para la solución de los parámetros se hacen necesarias tantas ecuaciones como numero de parámetros. Por ello, cuando tenemos mas ecuaciones que parámetros es posible emplear el exceso de información para testar el modelo. Evidentemente, este test no es posible cuando el modelo esta exactamente identificado, dado que no quedan ecuaciones libres para testar. Veamos seguidamente el procedimiento de testado de los modelos explicativos sobreidentificados.

En primer lugar debemos distinguir, como es habitual, entre los coeficientes teóricos propuestos y los coeficientes estimados. Así, consideraremos γ como el coeficiente teórico y notaremos como $\gamma\hat{}$ el coeficiente estimado. Así, $\gamma\hat{}_{12}$ será el valor derivado del coeficiente teórico γ_{12}. Partiendo de los valores de los coeficientes estimados y mediantes la relación que el modelo propone entre coeficientes y parámetros teóricos es posible reproducir las covariaciones y varianzas. Las varianzas y covarianzas pueden derivarse de forma única desde los parámetros estructurales del modelo (como sabemos, lo contrario no es cierto y de ahí el problema de la identificación).

El grado en que las varianzas y covarianzas reproducidas desde el modelo se parezcan a las obtenidas directamente y sobre las que se apoya el modelo actuara como test del modelo propuesto. Un mismo conjunto de varianzas y covarianzas pueden ser el punto de partida para ajustar diferentes modelos; cada modelo postula un tipo distinto de relaciones. En la medida que la reconstrucción de varianzas y covarianzas depende de las relaciones propuestas, cada modelo generará una reproducción distinta de la matriz relacional empírica.

En ese sentido, cabe la noción de testar los diferentes modelos en la medida que diferentes modelos generarán diferentes varianzas y covarianzas reproducidas. De hecho, si el modelo propuesto diese cuenta completa de las covarianzas y varianzas originalmente obtenidas, empíricamente no existirían diferencias entre estas y las covarianzas y

covarianzas reproducidas. Podemos entender que un modelo es correcto en la medida que las diferencias entre varianzas y covarianzas de partida se correspondan con las generadas desde el modelo. Por el contrario, si el modelo propuesto fuese erróneo, los dos conjuntos de varianzas y covarianzas serian muy diferentes, indicando de este modo la poca bondad del ajuste del modelo sobre los datos. No tratamos aquí de la realidad, sea la que sea, sino de cómo ésta ha sido expresada mediante datos (hablada) y cómo un modelo puede ajustar mejor o peor sobre esos datos.

En todo caso, cuando se compara la matriz reproducida de varianzas y covarianzas con la matriz de datos original, es importante evaluar si las diferencias se reparten de forma igual entre todos los coeficientes. La concentración de diferencias importantes en unos coeficientes en concreto nos indicara posibles problemas de especificación del modelo en esas variables en concreto.
