CAPITAL HUMANO Y PLS PATH MODEL

Gallese, Elda
Instituto de Investigaciones Económicas, Escuela de Economía

INTRODUCCIÓN

Según el Profesor Guarini (1996), entre otros, el impulso de las transformaciones tecnológicas de los últimos decenios, sin duda incrementó la imperiosa necesidad de invertir, especialmente en capital humano (instrucción, formación, capacitación), para activar el desarrollo, asegurar una adecuada formación profesional y crear así nuevos puestos de trabajo.

El profesor Dagum (1994) realizó un estudio donde explica, en un modelo, el ingreso de las familias en función de la riqueza neta y el capital humano que ellas poseen. En dicho trabajo el profesor Dagum propone que el valor monetario de la "**variable latente**" capital humano sea estimada por medio del modelo de Herman Wold con mínimos cuadrados parciales ("**PLS PATH model**").

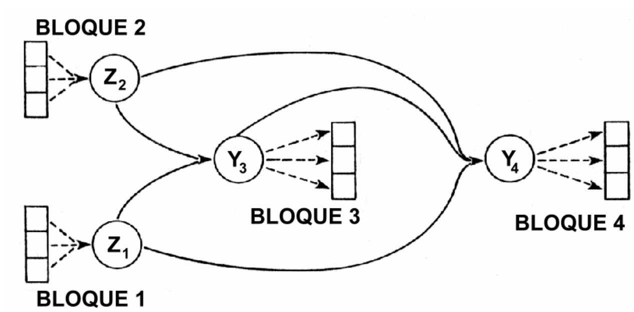
En este trabajo se desarrollará el marco teórico de los PLS PATH model de Herman Wold en el contexto de los modelos multiecuacionales.

Este modelo fue diseñado por H. Wold (1980) principalmente para el análisis causal-predictivo, de problemas con alta complejidad y poca información.

La **alta complejidad** está dada por la intervención de un gran número de indicadores (variables observables), relacionados a variables no observables o variables latentes.

La falta de relevamiento de datos a través del tiempo, y por consiguiente el desconocimiento de las propiedades de las distribuciones estadísticas de las variables relevantes, reflejan el hecho de **poca información y escaso conocimiento**.

El esquema de flechas del cuadro ilustra el modelo teórico conceptual. En él se observan:



- Cuáles son las variables lantes, **LV** que entran en el modelo.

- Cuáles son los indicadores (variables observables), que intervienen en cada bloque para medir indirectamente a las variables latentes.

- Qué relaciones internas se supone que existen entre las variables latentes.

Una variable latente **LV**, combina los items de un bloque (indicadores), para ser usado luego en el modelo causal – predictivo.

1. **Definición Teórica del Modelo**.

El modelo teórico conceptual, está ilustrado por el diagrama de flechas del cuadro de flechas.

Notación:

$j = 4$: número de bloques o variables latentes
k_j = número de indicadores en el bloque j-ésimo
z_1 y z_2 = variables latentes predeterminadas
y_3 e y_4 = variable latente endógena
x_{jk} = k-ésima, variable observable del j-ésimo bloque

- La relación de cada indicador con su bloque, **LV**, configura la **ESTRUCTURA DE BLOQUE**, ilustrada en el diagrama por las líneas de puntos.

Formalmente:

$$x_{jk} = f(z_j) + v_{jk} \qquad j = 1, 2 \qquad (1)$$
$$x_{sk} = f(y_s) + v_{sk} \qquad s = 3, 4$$

con

$$\mathrm{var}(z_j) = \mathrm{var}(y_s) = 1$$

$$r(z_j\, z_i) = \rho_{ij} \neq 0$$
$$r(y_s\, y_r) = \rho_{sr} \neq 0$$

$$r(v_{ih}, v_{ik}) = r(v_{ih}\, z_i) = r(v_{ih}\, y_j) = 0$$

$$i \neq j \qquad i, j = 1, \ldots, j \qquad h = 1, \ldots, k_i \qquad k = 1, \ldots, k_j$$

- Las relaciones de las variables latentes entre sí, son las **RELACIONES INTERNAS** (sistema de ecuaciones interdependientes)[1] , ilustradas en el diagrama por flechas de líneas llenas.
- Las flechas, indican los canales de información del modelo, mostrando cuáles variables son exógenas (z_1 y z_2), y cuáles son endógenas (y_3 e y_4).

Formalmente:

$$y_3 = \alpha_3 + \gamma_{31} z_1 + \gamma_{32} z_2 + \delta_3 \qquad (2)$$

$$y_4 = \alpha_4 + \beta_{43} y_3 + \gamma_{41} z_1 + \gamma_{42} z_2 + \delta_4$$

- La especificación de las **RELACIONES INTERNAS** y la **ESTRUCTURA DE BLOQUE**, constituye la "**DEFINICIÓN TEÓRICA DEL MODELO**".
- Una característica del modelo que no está ilustrada explícitamente por el esquema de flechas, es la especificación de las **RELACIONES CAUSAL-PREDICITIVA**.
- La **ESTRUCTURA DE BLOQUE**, las **RELACIONES INTERNAS** y las **RELACIONES CAUSAL – PREDICTIVA**, se denominan "**RELACIONES ESTRUCTURALES**" del modelo.

Recordemos que las relaciones causales son siempre predictivas, pero las relaciones predictivas no son necesariamente causales. La decisión respecto de la causalidad de una variable exógena es del especialista del tema en estudio. No es responsabilidad de la estadística.

2. Ajuste del modelo y estimación de los parámetros con PLS.

- Las variables observadas (indicadores), se someten al procedimiento de estandarización, razón por la cual, todas las regresiones van a aparecer estimadas sin el intercepto u ordenada al origen.
- Cada variable latente se estima como un agregado ponderado de sus indicadores.
- Cada variable latente se somete a la estandarización.

2.1 – Estimación de las Variables Latentes y las Ponderaciones:

1er paso: Se realiza la primera estimación de todas las variables latentes como un agregado de sus indicadores con **ponderación unitaria**.

[1] Una introducción al tratamiento de los sistemas de ecuaciones interdependientes o modelos multiecuacionales se presenta en el anexo.

2do paso: Se estiman las **ponderaciones** como un promedio del grado de asociación de cada indicador con los bloques (variables latentes estimadas en el 1er paso) relacionados con él, por el diagrama de flechas.

3er paso: Se realiza la segunda estimación de todas las variables latentes como un agregado de sus indicadores con las **ponderaciones obtenidas en el segundo paso**.

4to paso: Se estima las ponderaciones como en el 2do paso **sobre las variables latentes estimadas en el tercer paso**.

Este proceso iterativo sigue hasta que las diferencias entre las estimaciones de las distintas iteraciones sean despreciables.

De este modo, se obtienen las estimaciones de las **variables latentes**, y las estimaciones de las **ponderaciones**.

2.2 Estimación de las Relaciones Internas.

La estimación de las relaciones internas se obtiene "**estimando**" los coeficientes del siguiente modelo biecuacional interdependiente:

$$y_3 = \alpha_3 + \gamma_{31} z_1 + \gamma_{32} z_2 + \delta_3$$

$$y_4 = \alpha_4 + \beta_{43} y_3 + \gamma_{41} z_1 + \gamma_{42} z_2 + \delta_4$$

2.3 Estimación de la Estructura de Bloque.

La estimación de la estructura de bloque se obtiene "**estimando**" los coeficientes de las regresiones de cada **indicador**, con su correspondiente **variable latente** (bloque).

Estos coeficientes se denominan "**carga**" (load).

2.4 Estimación de las Relaciones Causal – Predictiva, en términos de variables latentes.

La estimación de las relaciones causal – predictiva en términos de **variables latentes**, se obtiene reemplazando, en la estructura de bloque (de cada indicador endógeno), **cada variable latente por su relación interna**.

Pone de manifiesto la relación entre cada indicador (endógeno) y las variables latentes relaciones con él por el diagrama de flechas.

2.5 Estimación de las Relaciones Causal – Predictiva, en términos de variables manifiestas.

La estimación de las relaciones causal – predicitiva en términos de **variables manifiestas**, se obtiene reemplazando las **variables latentes** en 2.4, por el **agregado ponderado de sus indicadores** obtenidos en 2.1.

2.6 Consideraciones generales.

En las relaciones entre las variables latentes y sus indicadores, el modelo da dos medidas de la relevancia de un indicador, su **ponderación** y su **carga**.

- La **carga** de un indicador, mide la relación **individual** entre el indicador y la variable latente de su propio bloque.
- La ponderación de un indicador, por su ubicación en la construcción de la variable latente, mide la relación entre el indicador y su propio bloque, **removiendo la influencia de los otros indicadores** del mismo bloque.

Por el modo de obtención, la ponderación de un indicador también mide el grado de relación de éste con las variables latentes de los bloques que están vinculados con él por medio del diagrama de flechas.

Las dos medidas son de importancia si se las interpreta correctamente, por consiguiente en los resultados, se presentan ambas.

El procesamiento de datos en lo referente al PLS PATH model se hace mediante el paquete estadístico de Lohmöler (1984).

INTRODUCCIÓN A LOS MODELOS MULTIECUACIONALES Y EL PROBLEMA DE LA IDENTIFICACIÓN

MODELOS MULTIECUACIONALES

Al enfrentarse con un problema económico el primer paso, fundamental, es la formulación del modelo. Ello se efectúa teniendo en cuenta ciertos supuestos "**a priori**", que parten de la observación de la realidad, de la experiencia anterior, de los elementos que proporciona la teoría económica, etc. Todo depende en última instancia de la sagacidad y buen sentido de aquel que está realizando la investigación y se ocupará de la formulación de los modelos tratando de interpretar el fenómeno de la forma más fidedigna posible, con la mayor aproximación, tratando de minimizar los errores y aumentar la precisión de su modelo.

Se tratará ahora el problema de la resolución de los mismos; aunque previamente se clarificarán algunos conceptos a los efectos de su correcta interpretación.

Se denominan "**variables**", al conjunto de características que varían en un determinado período de tiempo. Estas se clasifican en:

$$
\begin{array}{l}
\text{Endógenas}\begin{cases}\text{coetáneas (a explicar)} \quad (y_{it}) \\ \text{de retardo (o desfazadas}\end{cases} \\
\text{Exógenas}\begin{cases}\text{coetáneas}\\ \text{de retardo}\end{cases} \Bigg\} \text{pre-determinadas (explicativas)} \\
\text{Aleatorias}\begin{cases}\text{con error en las ecuaciones}\begin{cases}\text{con función de prob. conocida}\\ \text{con función de prob. desconocida}\end{cases}\\ \text{con error en las variables}\end{cases}
\end{array}
$$

Al conjunto de características que permanece constante, se lo denomina "**estructura del modelo**".

Los valores numéricos que asumen estas estructuras van a determinar los **"**parámetros estructurales**".**

Es importante distinguir bien los roles de parámetro y variable, es decir, si una característica entra en el modelo como variable o como parámetro. Luego habrá que especificar cuáles serán las variables predeterminadas y las endógenas: esto ya es una especificación del modelo; dependerá del problema que se trate.

El conjunto de variables y de parámetros estructurales adecuadamente combinados forman las ecuaciones estructurales.

El conjunto de ecuaciones estructurales constituyen el modelo.

Las variables endógenas influyen sobre las ecuaciones estructurales, pero a la vez están influenciadas por ellas; las variables exógenas influyen sobre el conjunto de ecuaciones estructurales pero no están influidas por ellas; y las variables aleatorias son las que están sometidas al cálculo de probabilidades.

Existen restricciones que hacen que no pueda incluir en el modelo todas las variables.

Modelo Teórico

$$\beta_{11} y_{1t} + \beta_{12} y_{2t} + \dots + \beta_{1n} y_{nt} + \gamma_{11} z_{1t} + \gamma_{12} z_{2t} + \dots + \gamma_{1m} z_{mt} = \mu_1$$

$$\beta_{21} y_{1t} + \beta_{22} y_{2t} + \dots + \beta_{2n} y_{nt} + \gamma_{21} z_{1t} + \gamma_{22} z_{2t} + \dots + \gamma_{2m} z_{mt} = \mu_2$$

.
.
.

$$\beta_{n1} y_{1t} + \beta_{n2} y_{2t} + \dots + \beta_{nn} y_{nt} + \gamma_{n1} z_{1t} + \gamma_{n2} z_{2t} + \dots + \gamma_{nm} z_{mt} = \mu_n$$

n: número de variables endógenas

m: número de variables predeterminadas
t = 1, 2, ..., T
Se trata de variables centradas pues en el modelo no aparece término independiente.

Vamos a presentar y a tratar aquí, un modelo multiecuacional simple y deduciremos así, las condiciones que deben darse para que se puedan estimar los parámetros intervinientes en cada ecuación estructural. Supondremos aquí que las variables erráticas inciden en las ecuaciones, es decir que en este estudio dejaremos de lado el caso en el que las componentes erráticas se ciñen a cada variable, así como el caso compuesto en el que hay erraticidad en las variables y en las ecuaciones.

Sea el modelo:

$$\left.\begin{aligned} \beta_{11} y_{1t} + \beta_{12} y_{2t} + \gamma_{11} z_{1t} + \gamma_{12} z_{2t} &= \mu_1 \\ \beta_{21} y_{1t} + \beta_{22} y_{2t} + \gamma_{21} z_{1t} + \gamma_{22} z_{2t} &= \mu_2 \end{aligned}\right\} \begin{array}{l}\text{Ecuaciones}\\ \text{Estructurales}\end{array}$$

El tiempo t tomará los valores 1, 2, 3, ..., T, y para mayor sencillez usaremos solamente los subíndices que se refieren a cada variable. Se entenderá que previo a los cálculos pertinentes, las variables deberán ser deflactadas por sus respectivos índices.

Las “z” son las variables exógenas o predeterminadas que actúan como datos. Las “y” son las variables endógenas cuyas leyes se tratan de estimar. Las “μ” son las componentes erráticas de las ecuaciones estructurales. El econometrista es el que debe diagnosticar la naturaleza aleatoria diciendo, por ejemplo, si son normales o no y si la serie de las μ_{it} tienen un carácter independiente o no, así como si puede existir una vinculación entre las μ_1 y las μ_2.

Los supuestos que se hagan respecto a la naturaleza de las variables erráticas forman parte integrante del modelo econométrico. En el caso que estudiamos, como los errores actúan en cada ecuación, la razón de los mismos debe ser imputable a que se han omitido otras variables. Estas variables omitidas son, en general, en gran número y entonces sí se puede admitir que cada variable omitida se adiciona a las otras, invocando el Teorema Central del Límite se puede conjeturar que las “μ” que acusan los resultados finales de esas variables, se distribuirán normalmente.

Como el modelo que hemos escrito no tiene términos constantes se supone que todas las variables están medidas a partir de sus respectivas medias aritméticas, o sea son desvíos.

El modelo quedará normalizado en la forma:

$$\begin{aligned} y_1 + \beta_{12} y_2 + \gamma_{11} z_1 + \gamma_{12} z_2 &= \mu_1 \\ \beta_{21} y_1 + y_2 + \gamma_{21} z_1 + \gamma_{22} z_2 &= \mu_2 \end{aligned} \qquad \text{(A)}$$

Ya que se habrá podido dividir la primera y segunda ecuación por β_{11} y β_{22}, ya que dichas variables correspondientes deben estar en la primera y segunda ecuación. Para mayor sencillez se han mantenido los mismos coeficientes a pesar de que su contenido haya variado por la normalización.

Aplicando la regla de Cramer se tiene:

$$y_1 = \frac{1}{\Delta} \begin{vmatrix} -\gamma_{11} z_1 - \gamma_{12} z_2 + \mu_1 & \beta_{12} \\ -\gamma_{21} z_1 - \gamma_{22} z_2 + \mu_2 & 1 \end{vmatrix}$$

con:

$$\Delta = \begin{vmatrix} 1 & \beta_{12} \\ \beta_{21} & 1 \end{vmatrix} = 1 - \beta_{21}\ \beta_{12}$$

deberá entonces verificarse que:
$\beta_{21} \cdot \beta_{12} \neq 1$

Se tiene también:

$$y_2 = \frac{1}{\Delta} \begin{vmatrix} 1 & -\gamma_{11} z_1 - \gamma_{12} z_2 + \mu_1 \\ \beta_{21} & -\gamma_{21} z_1 - \gamma_{22} z_2 + \mu_2 \end{vmatrix}$$

Se puede escribir así que:

$$y_1 = \frac{1}{\Delta} \left[(-\gamma_{11} + \beta_{12} \gamma_{21}) z_1 + (-\gamma_{12} + \beta_{12} \gamma_{22}) z_2 + \mu_1 - \beta_{12} \mu_2 \right]$$

$$y_2 = \frac{1}{\Delta} \left[(-\gamma_{21} + \beta_{21} \gamma_{11}) z_1 + (-\gamma_{22} + \beta_{21} \gamma_{12}) z_2 + \mu_2 - \beta_{21} \mu_1 \right]$$

Observemos que los parámetros "β" y "γ" aún no están determinados y que deberán ser estimados en base a la muestra de las observaciones sobre los tiempos t = 1, 2, 3, ..., T.
Las dos últimas expresiones pueden escribirse sintéticamente como:

$$\left.\begin{aligned} y_1 &= \Pi_{11} z_1 + \Pi_{12} z_2 + \varepsilon_1 \\ y_2 &= \Pi_{21} z_1 + \Pi_{22} z_2 + \varepsilon_2 \end{aligned}\right. \quad \text{(B)}$$

que no son más que las ecuaciones reducidas.
Se tendrán entonces que verificar las siguientes relaciones:

$$\Pi_{11} = \frac{1}{\Delta} (-\gamma_{11} + \beta_{12} \gamma_{21})$$

$$\Pi_{12} = \frac{1}{\Delta} (-\gamma_{12} + \beta_{12} \gamma_{22})$$

$$\Pi_{21} = \frac{1}{\Delta} (-\gamma_{21} + \beta_{21} \gamma_{11})$$

$$\Pi_{22} = \frac{1}{\Delta} (-\gamma_{22} + \beta_{21} \gamma_{12})$$

$$\varepsilon_1 = (\mu_1 - \beta_{12} \mu_2) / \Delta$$

$$\varepsilon_2 = (\mu_2 - \beta_{21} \mu_1) / \Delta$$

De las dos últimas relaciones se deduce que si las "μ" son normales, también serán normales las "ε", por ser combinaciones lineales de aquellas. Las primeras cuatro relaciones podrán escribirse:

$$\begin{aligned} (1 - \beta_{12} \beta_{21}) \Pi_{11} &= -\gamma_{11} + \beta_{12} \gamma_{21} \\ (1 - \beta_{12} \beta_{21}) \Pi_{12} &= -\gamma_{12} + \beta_{12} \gamma_{22} \\ (1 - \beta_{12} \beta_{21}) \Pi_{21} &= -\gamma_{21} + \beta_{21} \gamma_{11} \\ (1 - \beta_{12} \beta_{21}) \Pi_{22} &= -\gamma_{22} + \beta_{21} \gamma_{12} \end{aligned}$$

Sistema de parámetros

Los valores de Π se estimarán por mínimos cuadrados sobre las expresiones reducidas (B). Luego los Π son cuatro valores que se conocerán por dicha estimación y se tienen dos "β" a estimar y cuatro "γ" a estimar también, o sea seis incógnitas.

Si por ejemplo hacemos $\beta_{12} = 0$ se tendrá:

$$\left.\begin{aligned} \hat{\Pi}_{11} &= -\hat{\gamma}_{11} \\ \hat{\Pi}_{12} &= -\hat{\gamma}_{12} \end{aligned}\right\} \quad (C)$$

$$\left.\begin{aligned} \hat{\Pi}_{21} &= -\gamma_{21} + \beta_{21}\,\gamma_{11} \qquad \text{o sea: } \hat{\Pi}_{21} = -\gamma_{21} - \beta_{21}\,\hat{\Pi}_{11} \\ \hat{\Pi}_{22} &= -\gamma_{22} + \beta_{21}\,\gamma_{12} \qquad \text{o sea: } \hat{\Pi}_{22} = -\gamma_{22} - \beta_{21}\,\hat{\Pi}_{12} \end{aligned}\right\} \quad (D)$$

Las (C) se obtuvieron haciendo $\beta_{12} = 0$, o sea sacando un parámetro en la primera ecuación.

Para hacer resolubles las (D), o un γ_{21} o bien un β_{21} debe anularse.

Luego en cada ecuación (A) se anulará un parámetro, o lo que es lo mismo se omite una variable.

Hay n = 2 variables endógenas; en cada ecuación desaparece: n – 1 = 2 – 1 = 1 variables. En esta forma el modelo es identificable.

Ahora supongamos que hemos escrito directamente las ecuaciones reducidas. Hay que estimar las Π.

Recurriendo al método de los mínimos cuadrados deberá tenerse que:

$$\sum_{t=1}^{T} (y_{1t} - \Pi_{11}\, z_{1t} - \Pi_{12}\, z_{2t})^2 \quad \text{debe ser mínimo}$$

$$\sum_{t=1}^{T} (y_{2t} - \Pi_{21}\, z_{1t} - \Pi_{22}\, z_{2t})^2 \quad \text{debe ser mínimo}$$

de aquí se deduce el siguiente sistema de ecuaciones normales:

Para el primero:

$$\hat{\Pi}_{11} \sum_{t=1}^{T} z_{1t}^2 + \hat{\Pi}_{12} \sum_{t=1}^{T} z_{1t}\, z_{2t} = \sum_{t=1}^{T} y_{1t}\, z_{1t}$$

$$\hat{\Pi}_{11} \sum_{t=1}^{T} z_{1t}\, z_{2t} + \hat{\Pi}_{12} \sum_{t=1}^{T} z_{2t}^2 = \sum_{t=1}^{T} y_{1t}\, z_{2t}$$

y para el segundo:

$$\hat{\Pi}_{21} \sum_{t=1}^{T} z_{1t}^2 + \hat{\Pi}_{22} \sum_{t=1}^{T} z_{1t}\, z_{2t} = \sum_{t=1}^{T} y_{2t}\, z_{1t}$$

$$\hat{\Pi}_{21} \sum_{t=1}^{T} z_{1t}\, z_{2t} + \hat{\Pi}_{22} \sum_{t=1}^{T} z_{2t}^2 = \sum_{t=1}^{T} y_{2t}\, z_{2t}$$

Estas ecuaciones nos dan las estimaciones por mínimo cuadrados (que serán estimaciones máximas verosímiles si las componentes erráticas son normales e independientes serialmente cada una).

Conociendo así las estimaciones de las Π veamos cómo yendo a las ecuaciones del modelo original normalizado se podrán estimar sus parámetros.

Para ello, en la primera ecuación estructural del modelo normalizado (A) sustituyendo y_1 e y_2 por sus respectivas ecuaciones reducidas, se tiene:

$$(\Pi_{11} z_1 + \Pi_{12} z_2) + \beta_{12} (\Pi_{21} z_1 + \Pi_{22} z_2) + \gamma_{11} z_1 + \gamma_{12} z_2 = \mu_1$$

o sea:

$$(\Pi_{11} + \beta_{12} \Pi_{21} + \gamma_{11}) z_1 + (\Pi_{12} + \beta_{12} \Pi_{22} + \gamma_{12}) z_2 = \mu_1$$

Aplicando esperatas a ambos miembros, siendo la esperata del segundo miembro igual a cero, y manteniéndose el primer miembro como era, por ser una constante, se deduce que:

$$\Pi_{11} + \beta_{12} \Pi_{21} + \gamma_{11} = 0$$

$$\Pi_{12} + \beta_{12} \Pi_{22} + \gamma_{12} = 0$$

Son dos ecuaciones y tres incógnitas: β_{12} , γ_{11} y γ_{12} , que son los tres parámetros de la primera ecuación estructural.

Para tener una solución, una de esas debe ser cero, ya sea β_{12} , γ_{11} o γ_{12} . Esto equivale a decir que en la primera ecuación estructural y_2 o z_1 o z_2 , debe ser omitido. Esto coincide con la exigencia (como veremos en el caso general) en que si hay n variables endógenas, entonces en cada ecuación estructural n-1 variables no deben presentarse para que haya identificación.

Lo mismo en la segunda ecuación estructural.

IDENTIFICACIÓN

Pasemos al caso general en que hay n variables endógenas y m exógenas.

Consideremos la primera ecuación del modelo. Lo mismo se podría haber tomado cualesquiera una de las n ecuaciones que tendrá el modelo, ya que son n las endógenas que tenemos que explicar.

Sea entonces:

$$\beta_{11} y_1 + \beta_{12} y_2 + ... + \beta_{1n} y_n + \gamma_{11} z_1 + \gamma_{12} z_2 + ... + \gamma_{1m} z_m = \mu_1$$

Dejemos figurar el parámetro β_{11} , para simplificar la escritura, ya que normalizadas las ecuaciones estructurales, este coeficiente valdrá la unidad.

Podemos escribir condensadamente:

$$\sum_{j=1}^{n} \beta_{1j}\, y_j + \sum_{k=1}^{m} \gamma_{1k}\, z_k = \mu_1 \qquad \text{(S)}$$

La ecuación reducida para y_j es:

$$y_j = \Pi_{j1} z_1 + \Pi_{j2} z_2 + ... + \Pi_{jm} z_m + \varepsilon_j$$

o sea:

$$y_j = \sum_{k=1}^{m} \Pi_{jk}\, z_k + \varepsilon_j$$

Luego:

$$\sum_{j=1}^{n} \beta_{1j}\, y_j = \sum_{k=1}^{n} \beta_{1j} \sum_{k=1}^{m} \Pi_{jk}\, z_k + \sum_{j=1}^{n} \beta_{1j}\, \varepsilon_j \qquad \text{(M)}$$

Pero de (S):

$$\sum_{j=1}^{n} \beta_{1j}\, y_j = -\sum_{k=1}^{m} \gamma_{1k}\, z_k + \mu_1 \qquad \text{(N)}$$

Luego de (M) y (N) resulta que debe ser:

$$-\sum_{k=1}^{m} \gamma_{1k}\, z_k + \mu_1 = \sum_{j=1}^{n} \beta_{1j} \sum_{k=1}^{m} \Pi_{jk}\, z_k + \sum_{j=1}^{n} \beta_{1j}\, \varepsilon_j$$

Un z_k para un determinado k obliga a que:

$$-\gamma_{1k} = \sum_{j=1}^{n} \beta_{1j}\, \Pi_{jk} \qquad \text{y} \qquad \mu_1 = \sum_{j=1}^{n} \beta_{1j}\, \varepsilon_j$$

En la ecuación por tomar k los valores 1, 2, ... m, hay m ecuaciones de ese tipo y hay m "γ" y n "β", o bien n - 1 "β", pues por la normalización era $\beta_{11} = 1$. Hay m ecuaciones y hay m + n - 1 incógnitas. Para que haya solución n - 1 valores de "γ" y "β" deben anularse en la primera ecuación.

Así esa ecuación estará identificada. Habrá n - 1 variables "y" o "z" que no figuren. Lo mismo en cada ecuación restante.

BIBLIOGRAFÍA

DAGUM, C. (1994): *Human Capital, Income and Wealth Distribution Model With Applications*. S.A.E. Cuaderno Nº 11.

GREEN, P. (1976): *Mathematical Tools for Applied Multivariate Analysis*. Academic Press, Inc.

GUARINI, R. (1996): *La Sfida per lo Sviluppo*. Studi e note di economia Nº 3.

KOLMOGÓROV, A. and FORIM, S. (1975): *Elementos de la teoría de funciones y del análisis funcional*. Editorial Mir.

LOHMÖLLER, J. (1984): *Latent Variables Path Analysis With Partial Least Squares Estimation*. Program Manual. Universitatzu Koln.

MADANSKY, A. (1976): *Foundations of Econometrics*. North-Holland / American Elsevier.

RIETZ, H.L. (1924): *Handbook of Mathematical Statistics*. Houghton Mifflin. Company.

WOLD, H. (1980): *Model Construction and Evaluation When Theoretical Knowledge is Scarce*. Evaluation of Econometrics Models. Academic Press. Pag. 47-74.