

ECUACIONES ESTRUCTURALES, UNA TÉCNICA ESTADÍSTICA PARA FORMULAR Y CONTRASTAR MODELOS DE RELACIONES CAUSA-EFECTO

ANTONIO HUMBERTO CLOSAS

UNIVERSIDAD NACIONAL DEL NORDESTE
UNIVERSIDAD TECNOLÓGICA NACIONAL

*“El progreso y el desarrollo son imposibles si uno
sigue haciendo las cosas tal como siempre las ha hecho”.*

Wayne W. Dyer

A&P

ACTUALIDAD&PROSPECTIVA

RESUMEN

Al pretender explicar un determinado fenómeno, con frecuencia algunas de las variables consideradas como posibles predictoras de la respuesta de interés pueden ser reflejo de un constructo o factor subyacente. Esta complejidad puede ser tratada mediante modelos de ecuaciones estructurales (MEE), ya que los mismos permiten ordenar una gran cantidad de variables, observadas y latentes, cuya combinación –en bloques vinculados según un esquema teórico previo– contribuye a esclarecer las relaciones de causa-efecto que pueden existir. Por ello, el propósito de este trabajo es dar a conocer diferentes aspectos inherentes a los MEE, una secuencia de modelos matemático-estadísticos que posibilitan estudiar determinadas situaciones, teniendo en cuenta la estructura de covarianzas derivada de las asociaciones entre diversas variables. No obstante, como en cualquier proceso de modelización estadística es esencial dar al conocimiento de la teoría del área de aplicación la máxima importancia, ya que del mismo surgirán el planteamiento y la validación de las hipótesis sustantivas. Desde hace algunos años, se encuentran disponibles en el mercado distintos programas estadísticos (EQS, AMOS, LISREL, MPLUS, R, SAS, etc.) los cuales se requieren para comprobar los MEE; esto es, estimar los parámetros y contrastar la validez de las medidas de bondad de ajuste del modelo teórico a los datos muestrales. En vista de lo expuesto, se puede sostener que la metodología basada en los MEE es uno de los procedimientos que más se aconseja utilizar en la actualidad de cara a la explicación de muchos fenómenos presentes en la compleja realidad educativa, económica y social.

Palabras clave: análisis multivariante, modelización estadística, ecuaciones estructurales, programas informáticos.

STRUCTURAL EQUATIONS: A STATISTICAL TECHNIQUE FOR FORMULATING AND TESTING MODELS OF CAUSE-EFFECT RELATIONS

ABSTRACT

When trying to explain a certain phenomenon, often some of the variables identified as possible predictors for the interest answer could actually be based on an underlying factor. This complexity can be treated by Structural Equation Modeling (SEM), as they allow to organize a great deal of variables (observed and latent). Their combination in blocks (according to a previous theoretical scheme) contribute to pinpoint the causeeffect relationships that could exist. The purpose of the present paper is to introduce different aspects inherent to the SEM, a sequence of mathematical-statistical models allowing to model situations through the covariance structure derived from di-

fferent association of the variables. Nonetheless, in any statistical modeling process, it is vital to focus in the application area theory, as the hypothesis and their validations will come from it. From some time now, are available in the market some statistical tools (EQS, AMOS, LISREL, MPLUS, R, SAS, etc.) required to validate of the SEM, i.e., assess the parameters and check how the theoretical model fits the sample data. In summary, SEM is one of the most advisable procedures to explain complex phenomenon in the educational, economical and social areas.

Keywords: multivariate analysis, statistical modeling, structural equations, computer programs.

1. INTRODUCCIÓN

En diversas áreas de conocimiento cuando por alguna razón es necesario proporcionar una explicación cuantitativa de la relación que existe entre una variable dependiente (criterio, explicada o endógena) y más de una variable independiente (predictora, explicativa o exógena), que ciertamente pensamos que influyen en la respuesta de interés, es habitual utilizar para ello un modelo de regresión lineal múltiple. Sin embargo, en aquellos casos que la situación se torna más compleja, como por ejemplo, cuando se trata de estudiar de qué manera se relacionan dos o más variables latentes (no medidas en forma directa), evidentemente, es necesario contar con un sistema de ecuaciones funcionales que permita describir en su totalidad la estructura que se ha generado como consecuencia de las diferentes relaciones planteadas.

Este tipo de problemáticas, que con frecuencia se presentan en Educación, Biometría, Marketing, Psicometría, Ciencias Sociales, entre otras disciplinas, pueden ser abordadas mediante modelos de ecuaciones estructurales (*SEM, Structural Equation Modeling*). Se trata de un método práctico que permite formular y comprobar modelos matemáticos que describen relaciones entre variables (observadas y latentes). Este método extiende las ideas de correlación y regresión en el sentido de admitir en el estudio redes complejas de variables interconectadas, en lugar de efectuar una separación artificial de éstas en dependientes e independientes.

En realidad, se trata de un conjunto de técnicas basadas en el análisis de las covarianzas o de las correlaciones entre las variables, que incluyen diversos desarrollos: análisis factorial confirmatorio, relaciones estructurales lineales, análisis de estructuras de covarianza, etc., que han surgido en respuesta a la necesidad de establecer inferencias de tipo explicativo de cierta clase de fenómenos. En definitiva, siguiendo a Biddle y Marlin (1987), puede decirse que es un instrumento que combina las ventajas del análisis factorial con las de la regresión múltiple. De igual forma que en el análisis factorial, el método puede trabajar con factores basados en la covarianza de las variables y, además, como se hace en la regresión múltiple, es posible describir la fuerza de las relaciones estructurales que se dan entre los factores subyacentes (Lunnenborg y Abbott, 1983).

La evaluación de SEM se realiza a través de dos instancias: a) el estudio analítico, a efectos de determinar y contrastar las relaciones entre las variables postuladas en el modelo; b) el

análisis del grado de ajuste global del mismo, con el fin de comprobar en qué medida el modelo teórico reproduce correctamente las relaciones existentes en la matriz de correlaciones de datos empíricos (véase Jackson, 2007; Kaplan, 2009).

El estudio analítico, a su vez, contempla dos pasos importantes: el análisis de medición y el análisis estructural. El primero es un análisis factorial confirmatorio, mientras que el segundo estima las relaciones entre los factores obtenidos en el análisis de medición. La ventaja de este estudio es que permite medir al mismo tiempo los efectos directos e indirectos que tiene una variable latente u observada sobre otra(s) variable(s) (Asparouhov y Muthén, 2009; Corral-Verdugo, 2002).

El análisis de ecuaciones estructurales, cuyo origen se encuentra en el *path analysis* de Wright (1934), estuvo disponible a partir de los años setenta (Goldberger y Duncan, 1973), aunque fue en la década del ochenta que comenzaron a elaborarse investigaciones en el campo educativo utilizando esta metodología. Jöreskog (1978), Muthén (1984) y Bentler (1990), han sido los pioneros, a quienes muchos han seguido luego, en trabajar con esta serie de técnicas desarrolladas con el objetivo de evaluar el grado de ajuste entre un modelo teórico y los datos empíricos, utilizando para ello la información contenida en las covarianzas o correlaciones de los datos.

Mediante los modelos de ecuaciones estructurales, al principio conocidos como modelos causales, es posible ordenar la gran cantidad de variables que intervienen, con el fin de intentar explicar algún fenómeno, cuya combinación –en bloques relacionados según un esquema teórico previo– contribuye a esclarecer las relaciones de dependencia causal que pueden existir; aunque esta es una cuestión que debe ser tratada con prudencia puesto que excede al ámbito estadístico. En rigor, el concepto del método se refiere a un conjunto de estrategias cuya finalidad es elaborar modelos que sean explicaciones plausibles de los fenómenos, con el objeto de validarlos empíricamente (Bollen, 1989). No es por tanto un método para descubrir causas, sino para validar estructuras elaboradas por el investigador basándose en una teoría (Bisquerra, 1989).

Actualmente, se encuentran disponibles diversos programas informáticos, algunos de los cuales aparecieron en los años ochenta, que permiten estimar y contrastar la validez de las medidas de bondad de ajuste del modelo a los datos. Realmente, la etapa de informatización comienza en el año 1972 con la aparición del programa LISREL (Linear Structural Relationship) escrito por Karl G. Jöreskog y M. Van Thillo, claro que años después fueron apareciendo sucesivas versiones mejoradas del mismo. Este programa ha sido el primero que se comercializó para aplicar a modelos de ecuaciones estructurales y en principio lo había incorporado SPSS (Statistical Package for the Social Sciences), en las versiones LISREL 7 y 8 (Jöreskog y Sörbom, 1989), pero a la fecha es AMOS el programa que sustenta SPSS, aunque en módulo aparte.

Por cierto, existen además otros paquetes también muy conocidos, surgidos con posterioridad a LISREL, entre los que se encuentran EQS, MPLUS, R, SAS CALIS, LISCOMP; gracias a los cuales es posible ahorrar los cálculos engorrosos que antes era necesario realizar en forma manual para ajustar los datos. Las últimas versiones de todos estos programas presentan inter-

faces con el usuario de uso bastante coloquial, por lo que el trabajo con modelos de ecuaciones estructurales ha experimentado una simplificación notable que continuamente está generando una mayor popularidad de éstos. El problema que de inmediato se presenta es que el uso indiscriminado de los paquetes informáticos ocasiona, muchas veces, que las técnicas multivariantes (ecuaciones estructurales se encuentra entre las denominadas técnicas de dependencia) no se empleen adecuadamente, dado que las mismas son relativamente complejas y requieren para su utilización un conocimiento profundo de sus fundamentos y condiciones de aplicabilidad.

Más allá de este hecho, la posibilidad de contar en la actualidad con métodos estructurales, unido a la notable ventaja que brindan el empleo de computadoras personales y de eficientes programas estadísticos para el proceso de análisis de datos, han reimpulsado el estudio en distintas disciplinas desde el paradigma correlacional. Las investigaciones, desde esta perspectiva, se habían debilitado en años anteriores a los setenta, probablemente a causa de la carencia de modelos y métodos estadísticos apropiados para llevarlas adelante.

2. ASPECTOS OPERATIVOS

En relación con la estructura de un modelo teórico, cabe indicar que éste ha de contener todas las variables que se consideran importantes, disponiéndolas tal y como funcionan en la realidad. El modelo debe marcar las directrices de la recolección de datos, es decir, las variables que deben ser medidas. En cualquier caso, esta interpretación de la entidad modelovariables que se propone, si bien permite, en principio, dar respuesta al problema, no está exenta de reformas. Por el contrario, lo más probable es que sea modificada (se denomina reespecificación del modelo), una y otra vez, con el objeto de acercarse a un modelo lo más concreto y perfilado posible. A su vez, la teoría al respecto establece la forma que han de tener las ecuaciones del modelo y luego de determinado el tipo de datos a recolectar, el método por el que éstos serán analizados.

Las variables que intervienen en un modelo pueden ser de dos tipos: a) *observadas* (medidas o manifiestas), b) *latentes* (no medidas o subyacentes). Una flecha recta de una variable latente a una variable observada indica una relación de causalidad. La relación causal o path entre las variables puede ser *unidireccional* (recursivo o jerárquico) o *bidireccional* (no recursivo). El parámetro que mide la intensidad de esta relación, es el término que en el análisis factorial exploratorio se denomina saturación o carga factorial, o el coeficiente estandarizado asociado a una variable independiente en una regresión múltiple. Además, a cada variable dependiente (observada o latente) se le debe añadir su *residuo* o *error residual*; este término en el caso de una variable latente se denomina *perturbación* (*disturbance*).

Para la estimación de parámetros se recomienda en muestras¹ pequeñas, siempre que

¹ Habitualmente se requieren tamaños muestrales superiores a los 100 sujetos (entre 200 y 1000 observaciones son una buena garantía). Como regla general, deben obtenerse cinco veces más observaciones que el número de variables que van a ser analizadas y el ratio óptimo (sujetos-variables) es de diez a uno.

sea plausible la asunción de normalidad e independencia entre los residuos y las variables latentes, los dos procedimientos siguientes: a) método de máxima verosimilitud y b) método de mínimos cuadrados generalizados. No obstante, en caso que ambos supuestos no parezcan razonables, se sugiere recurrir a la estimación de máxima verosimilitud denominada “robusta” (Satorra y Bentler, 2001), la cual se encuentra disponible en varios programas (e.g., opción ROBUST en EQS, opción MLM en MPLUS).

Ahora bien, si uno o más de los indicadores es categórico (o si la ausencia de normalidad es extrema), deberá optarse por otros métodos de estimación (e.g., mínimos cuadrados ponderados, ponderados diagonalizados, ponderados robustos o no ponderados), debido a que en estos casos el uso del método de máxima verosimilitud puede producir estimaciones atenuadas de las relaciones entre los indicadores (en especial cuando existen efectos “suelo” o “techo”), conducir a la formación de pseudofactores (originados por la dificultad o el carácter extremo de los ítems), producir errores estándar incorrectos o estimaciones erróneas de los parámetros, llegar a resultados imprecisos si se parte de una matriz de correlaciones en lugar de una de covarianzas (Brown, 2006).

Básicamente, existen dos formas equivalentes de representar los modelos causales, la gráfica y la algebraica. La representación gráfica se realiza mediante diagramas de caminos y la forma algebraica a través de ecuaciones (en la mayoría de los casos lineales) que relacionan las variables entre sí.

Emplearemos, a fin de ejemplificar, la notación de Bentler y Weeks (1980), dado que es en la que se basa el sistema EQS, una de las aplicaciones más usadas en la actualidad y la que habitualmente utilizamos en nuestras investigaciones (véase Closas, 2009; Closas, Sanz de Acedo y Ugarte, 2011). En esta notación, para cada variable del modelo, latente u observada, sólo es relevante si es una variable dependiente o independiente.

A las variables observadas se les asigna la etiqueta V y a los términos de error de éstas la E, mientras que las variables latentes (factores comunes) se indican con la letra F y a los errores de las estimaciones de éstas con la D (de *disturbance*). En la Figura 1, se muestra un ejemplo simple de representación gráfica de un modelo de ecuaciones estructurales o, como también se denomina, modelo de estructuras de covarianza.

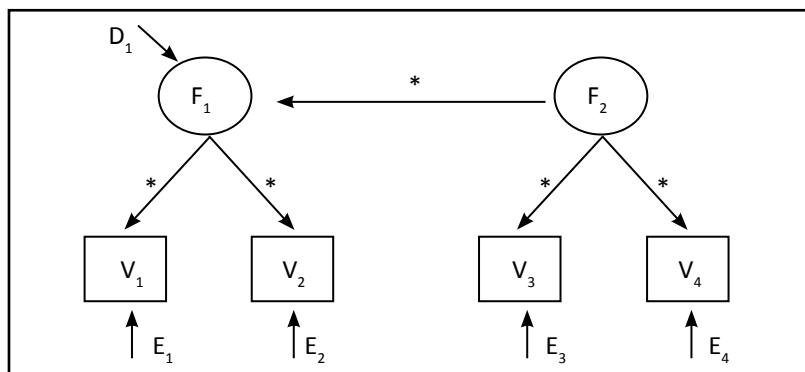


Figura 1. Modelo mediante la notación Bentler-Weeks

Si denotamos con un asterisco los parámetros que serán estimados, el modelo del ejemplo estaría representado por las ecuaciones que se muestran en la Tabla 1 (observar que los coeficientes de regresión de las variables observadas respecto a los términos de error o de los factores dependientes respecto a los términos de error fueron fijados a 1).

Componente de medida	$V_1 = *F_1 + E_1$ $V_2 = *F_1 + E_2$
	$V_3 = *F_2 + E_3$ $V_4 = *F_2 + E_4$
Componente estructural	$F_1 = *F_2 + D_1$

Tabla 1. Formulación algebraica del modelo de la Figura 1

Desde luego, los modelos de ecuaciones estructurales utilizados en situaciones reales emplean mayor número de variables, tanto observadas como latentes, e interrelaciones entre las mismas que dan origen a representaciones gráficas y matemáticas bastantes más complejas que las observadas en el ejemplo presentado.

3. APRECIACIONES COMPLEMENTARIAS

Si bien, en el apartado anterior ya hicimos referencia a las ecuaciones estructurales como metodología de análisis de datos, brindamos a continuación, en forma breve, algunas cuestiones que creemos ameritan ser mencionadas con el objeto de completar la información presentada hasta el momento.

Comenzamos por señalar que el cumplimiento de los supuestos paramétricos en que se basa la mayoría de las pruebas de significación en análisis multivariante (normalidad, linealidad, independencia, homoscedasticidad), es evidentemente extensible a las ecuaciones estructurales, las cuales se encuentran, como ha sido señalado, entre las denominadas técnicas explicativas o de dependencia.

Al trabajar con datos reales, en muchas ocasiones resulta que las hipótesis de normalidad y homoscedasticidad multivariante no se verifican estrictamente. Sin embargo, en la mayoría de estas situaciones dichos supuestos “casi” se cumplen; de todas formas, lo que realmente interesa en estos casos es conocer los efectos que pueden tener sobre los resultados las violaciones de estas hipótesis. Existe clara evidencia tanto matemática como empírica de que las pruebas con una sola variable dependiente (análisis de la varianza, de regresión, de la covarianza, etc.) son altamente robustas bajo la violación de los principios de normalidad y homoscedasticidad, excepto cuando las muestras son muy pequeñas y desiguales. En el análisis multivariante, sostiene Harris (1985), la tendencia está en considerar que en muestras grandes

($n > 30$) sus métodos son lo suficientemente robustos como para ser insensibles a ligeras desviaciones de los supuestos paramétricos.

En lo atinente a los estudios estadísticos de carácter confirmatorio, corresponde puntualizar que una manera de formalizar los mismos es, precisamente, por medio de un conjunto de técnicas denominadas sistemas de ecuaciones estructurales (análisis factorial confirmatorio y modelos de estructuras de covarianza). De hecho, en la actualidad, el análisis factorial confirmatorio es una de las técnicas más utilizada en investigaciones sobre temas sociales (Kline, 2011); se emplea con frecuencia, por ejemplo, en el desarrollo de instrumentos estadísticos de medición (test de personalidad, de inteligencia, de calidad educativa, entre muchas otras pruebas).

Según se ha indicado, la aplicación de los métodos estructurales permite identificar y estimar las variables relevantes como causas de otras variables, así como contrastar los datos de la muestra con el modelo teórico diseñado en base a los conocimientos sobre el tema, lo cual sitúa a esta clase de estudios en el dominio de la estadística inferencial.

En verdad, se trata de un procedimiento elemental en el sentido que trabaja, sobre todo, con modelos recursivos (relaciones de causación unidireccionales) y únicamente con variables observadas (manifiestas o medidas). Su propósito, es el estudio de las influencias directas e indirectas de unas variables tomadas como causas de otras consideradas como efectos. Esta caracterización supone un enriquecimiento de lo que es posible realizar con el análisis de regresión múltiple, ya que permite obtener el efecto total (suma de efectos directos e indirectos) y concluir de una forma más ajustada sobre la contribución de las diferentes variables en la explicación de un fenómeno de interés.

En sintonía con lo expresado se encuentra lo que sostiene J. Pearl, quien en un reciente informe técnico (Pearl, 2011) respecto a la utilidad de los SEM señala que la descomposición de los efectos en sus componentes directos e indirectos tiene verdadera importancia, ya que nos indica “cómo funciona la realidad” y, por lo tanto, nos permite predecir comportamientos bajo una rica variedad de condiciones e intervenciones.

Así pues, la utilidad del análisis de ecuaciones estructurales, en razón de su ascendencia y de las características que le son propias, reside en probar estructuras conformadas por relaciones de causa-efecto que se han planteado en función de la teoría correspondiente y ver si son congruentes o no con los datos. Si el modelo es congruente con ellos, es posible decir que apoya la teoría que lo ha generado, pero nunca se puede interpretar como prueba de la misma, puesto que con esos mismos datos correlacionales pueden ser congruentes otros modelos.

Por tanto, asumiendo en sentido estricto su incumbencia, podemos decir que el método estructural es un procedimiento que permite validar modelos causales elaborados por el investigador basándose en una teoría. Su objetivo consiste en pasar de una teoría expresada en forma coloquial a un modelo planteado de modo algebraico, puesto que, una teoría causal puede expresarse: a) verbalmente, mediante un informe; b) esquemáticamente, a través de un diagrama de camino; c) matemáticamente, por medio de un sistema de ecuaciones estructurales.

Sin duda, la elaboración de modelos causales y su comprobación empírica constituyen una metodología que en las últimas décadas está siendo profusamente utilizada por científicos en distintas áreas y en diferentes problemáticas, aunque de manera particular es aplicada en aquellas investigaciones interesadas en explicar la influencia de ciertas variables en algún determinado fenómeno. Seguramente, este protagonismo destacado que se observa en la investigación causal sea debido, en buena parte, a las posibilidades que ofrecen para su abordaje los modelos de ecuaciones estructurales y al avance, tanto de la tecnología, como de los programas estadísticos disponibles actualmente para el procesamiento y el análisis de los datos.

Por último, en vista de que mediante la técnica objeto de estudio es posible analizar relaciones de causa-efecto no sólo entre las variables predictoras y aquellas que se pretenden explicar, sino también entre las propias variables predictoras, puntualizaremos a continuación, aunque de manera resumida, ciertas cuestiones inherentes al concepto de causalidad.

4. ALGUNAS REFLEXIONES ACERCA DEL CONCEPTO DE CAUSALIDAD

En la expresión algebraica de un modelo, como ha sido citado en apartados anteriores, las funciones probabilísticas deben contener todas las variables relevantes a efectos de representar las relaciones de causa-efecto tal y como se presentan en la realidad. El conocimiento de la teoría sustantiva acerca del tema de estudio debe proporcionar la información necesaria para aislar dichas variables, con las cuales elaborar el sistema de ecuaciones funcionales.

Creemos importante enfatizar que el análisis de ecuaciones estructurales no es un sistema para descubrir las causas que dan origen a un problema, sino un método que se aplica a un modelo formulado por el investigador en base a consideraciones teóricas; su utilidad reside en evaluar una teoría, no así en generarla. Como señala Chou (1990), se trata de una técnica que permite la inclusión de variables latentes y observadas, lo cual posibilita revelar de un modo más eficiente y flexible la cadena causal de un fenómeno complejo.

Según señala Maxim (2002), para la ciencia la causalidad no es una idea sencilla. Yace inextricablemente presa de una red de discusiones metafísicas que trascienden el tiempo y la cultura, lo cual ha suscitado una gran controversia en algunos especialistas y filósofos que han dejado de verla como una entidad viable en el marco del método científico. Es probable que esta polémica tenga origen en la dificultad intrínseca de demostración de las teorías causales, pues podría suceder que distintas teorías describieran los mismos fenómenos.

Sin embargo, como afirma Bisquerra (1989), la idea de causalidad está presente explícita o implícitamente en la mayoría de las investigaciones científicas y, según la opinión de muchos otros especialistas, constituye uno de los elementos más importantes de la metodología de la ciencia aplicada (Bunge, 1979; Holland, 1986; Scriven, 1971). El motivo de esta persuasión radica, preferentemente, en que el principio causal brinda una manera asequible de organizar

el terreno empírico a través de una serie de pautas razonables que se conocen.

Casi todos los análisis de las propiedades fundamentales de la causalidad proponen (Davis, 1985), quizás variando la forma pero no el espíritu, tres condiciones para identificarla: a) una secuencia temporal entre dos variables, es decir, la variable causa debe preceder en el tiempo a la variable efecto; b) el factor causa debe estar correlacionado con el factor efecto; c) la única explicación plausible del efecto observado tiene que ser el factor causa.

A su vez, la condición necesaria y suficiente del principio de causalidad debe ser expresada, según sostiene Mackie (1980), específicamente en los siguientes términos: “*X es una condición necesaria de Y*, significa que cada vez que ocurra un evento de tipo Y, también ocurrirá uno de tipo X; y *X es una condición suficiente para Y*, significa que cada vez que suceda un evento de tipo X, también lo hará uno de tipo Y” (p. 57).

A pesar de estos aspectos precisos y severos del concepto, muchos estudios correlacionales, señala Bisquerra (1989), mantienen una posición ambigua sobre términos como, por ejemplo, asociación, vinculación, influencia, dependencia, causalidad. Otros, en cambio, sustituyen el concepto de causalidad por el de dependencia entre variables (James, Mulaik y Brett, 1982), que habitualmente se expresa en términos probabilísticos mediante relaciones funcionales, a las que se añade el error aleatorio, que contienen todas las variables relevantes seleccionadas de la teoría sustantiva.

Como corolario de lo planteado, podríamos decir que en las investigaciones científicas es necesario recurrir a marcos causales, más allá que desde la visión filosófica se tengan reservas sobre el concepto de causalidad. Incluso podemos agregar que, en muchas ocasiones, si bien el término causa puede estar ausente, la idea de causalidad está presente.

De lo que precede, es que en lugar de *causación fuerte* (*A es la única causa de B*) entre variables, se hable de *causación débil* (*A es una de las causas de B*) la cual, siguiendo a Davis (1985), tiene las siguientes características: a) se refiere a tendencias, esto es, se esperan excepciones individuales a la relación entre las variables; b) puede haber más de una causa para la variable criterio; c) la correlación entre las variables no implica necesariamente causación.

Por su parte, Moore (1991) sostiene que la evidencia para afirmar que una variable causa cambios en otra se fortalece siempre que se den las tres siguientes condiciones: a) la asociación entre ambas variables se reproduce en diferentes circunstancias, lo que reduce el riesgo de que los cambios se deban a variables no controladas; b) se dispone de una explicación plausible que hace razonable el hecho de que una variable cause cambios en la otra; c) no existe un factor válido que pueda causar cambios en ambas variables conjuntamente.

Con estas limitaciones, el análisis causal, conocido en la actualidad como análisis de ecuaciones estructurales², según hemos señalado, hace referencia a un conjunto de estrategias cuyo propósito es elaborar modelos causales o de relaciones causa-efecto que sean explicaciones plausibles de los fenómenos con el objeto de validarlos o rechazarlos empíricamente. Estos

² Para otros detalles sobre SEM y causalidad, así como respecto del cambio de denominación, puede consultarse Pearl (2011), pp. 3-4.

modelos utilizan el término *causa* en el sentido citado de la causación débil que, desde luego, es también la manera que tenemos de interpretarlo en las investigaciones que realizamos en nuestro ámbito de trabajo.

5. APRECIACIONES SOBRE TEORÍAS Y MODELOS

En razón de que es fundamental la búsqueda de significados que permitan precisar de alguna manera las expresiones teorías y modelos, así como sus interrelaciones, dado que son términos frecuentemente utilizados en el contexto del análisis de ecuaciones estructurales, presentamos a continuación algunos aspectos vinculados con ambos conceptos.

Así pues, iniciamos este apartado señalando en primer lugar que las teorías pueden definirse como “grupos de proposiciones lógicamente interconectadas, de las que pueden deducirse uniformidades empíricas” (Merton, 1983, p. 56).

En tanto que, un modelo se define, siguiendo a Voght (1993), como una representación o descripción de un fenómeno o un conjunto de relaciones que ayuda a entenderlo o a estudiarlo; por ello puede decirse que el modelo es una representación formal de la teoría.

La teoría puede ser una aproximación a la realidad y el modelo derivado de ella trata de representarla mediante convenios preestablecidos en la misma. En verdad, los modelos son sólo representaciones simplificadas y aproximadas de la realidad, por ello cuando se diseña un modelo se pretende esclarecer las relaciones que se dan en el mundo real.

Los modelos se usan en ciencia como abstracciones teóricas adecuadas para describir datos empíricos; como tales los mismos son siempre aproximaciones. En efecto, los modelos son una aproximación a la teoría sustantiva; ningún modelo contiene todos los conceptos teóricos que pasan por la cabeza del científico que lo construye. De hecho, la elaboración de un modelo puede ser una buena estrategia para un posterior desarrollo de la teoría; el intento de formular un modelo adecuado puede iluminar partes del conocimiento en los cuales la teoría es tan vaga o inconsistente que lleva al que la elabora a buscar soluciones. Este ejercicio puede ser útil incluso cuando después no se realice un análisis de los datos.

Podría decirse que los modelos son idealizaciones de la realidad inventados por la mente humana. Los modelos no son “verdaderos” y nunca dan una representación exacta de la situación que se desea mostrar. Si el modelo reflejara exactamente fenómenos reales sería tan difícil de comprender como lo es el mundo real y por lo tanto tendría poco valor práctico. Por ello se usan modelos que sean los más sencillos posibles y que aproximen la realidad lo suficientemente bien para nuestros objetivos.

6. DISCUSIÓN

Como ha sido expresado, en varias ocasiones, el análisis de ecuaciones estructurales es una de las técnicas estadísticas que más se aconseja utilizar en la actualidad para contrastar las teorías elaboradas sobre multitud de fenómenos educativos, psicológicos, económicos y sociales en general. Sin embargo, la aplicación de esta técnica se debe realizar con bastante cuidado ya que presenta ciertas limitaciones y dificultades, aunque ciertamente algunas de ellas son habituales en cualquier metodología de modelado estadístico que se utilice. Una de estas dificultades reside en elaborar modelos teóricos que sean sencillos y que a la vez reflejen la realidad que suele ser compleja. Cuando en un modelo interviene un número elevado de variables se obtiene una representación complicada, aunque el hecho de prescindir de algunas de ellas puede suponer que se ha dejado de tener en cuenta aspectos importantes del fenómeno objeto de estudio.

Este es un dilema con el que un investigador suele encontrarse frecuentemente: a) si participan muchas variables en el modelo, éste resulta difícil de comprender; b) si se simplifica el modelo, no siempre se está representando la realidad. Lo que se gana en sencillez, puede perderse en exactitud al representar la compleja realidad que se analiza. Evidentemente, definir la “simplicidad” o *parsimonia* de los modelos es una tarea que no está libre de dificultades. En los modelos estadísticos el número de parámetros que se estiman es una buena medida de la complejidad del modelo.

Por otra parte, si un investigador logra elaborar un modelo y éste resulta consistente con los datos empíricos disponibles, esto no es una prueba concluyente de las relaciones de causa-efecto propuestas. La consistencia entre los datos y el modelo no implica necesariamente la consistencia entre el modelo y la realidad. Lo único que puede afirmarse en tal caso es que los supuestos del investigador no son contradictorios y por lo tanto pueden ser válidos; aunque, ello no quiere decir que sean la única explicación del fenómeno bajo estudio, pues es posible que otros modelos también se adapten a los mismos datos.

La importancia de los modelos de estructuras de covarianza radica en que gracias a ellos se pueden rechazar los modelos no consistentes con los datos. El problema está en que el verdadero modelo es sólo uno de entre los muchos en los que pueden encajar los datos; por ello, para diseñar un modelo que se ajuste a la realidad es muy importante el conocimiento sustantivo que el investigador tenga del objeto de estudio. Sin embargo, como todo modelo, supone una elaboración teórica, una abstracción, acerca del funcionamiento y la dinámica de una parcela de la realidad, por lo que debemos ser conscientes de que no estará exento de algunas limitaciones.

Finalmente, más allá de algunas debilidades que toda elaboración científica lleva consigo, se espera que este primer acercamiento al estudio de las ecuaciones estructurales pueda aportar algo más de luz a este tema que se caracteriza por su dificultad de comprensión. Nuestra

perspectiva respecto de la aplicación de esta técnica multivariante para el análisis de datos cuantitativos, es sin duda favorable, en virtud de considerarla una metodología muy conveniente para el planteo y el contraste de modelos de relaciones causa-efecto entre diferentes variables (predictoras y criterios), con miras a explicar muchos fenómenos presentes en la compleja realidad educativa, económica y social.

REFERENCIAS BIBLIOGRÁFICAS

Asparouhov, T. y Muthén, B. (2009). Exploratory Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 397-438.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238-246.

Bentler, P. M. y Weeks, D. G. (1980). Linear structural equations with latent variables. *Psychometrika*, 45, 289-308.

Biddle, B. y Marlin, M. (1987). Causality, confirmation, credulity and structural equation-modelling. *Child Development*, 58, 4-17.

Bisquerra, R. (1989). *Introducción conceptual al análisis multivariable. Un enfoque informático con los paquetes SPSS-X, BMDP, LISREL y SPAD* (Vol. 1 y 2). Barcelona: PPU.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley & Sons.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Bunge, M. (Ed.) (1979). *Causality and Modern Science*. New York: Dover.

Chou, H. N. (1990). *Investigación of a proposed achievement model: Using LISREL Structural Modeling*. University of Illinois at Urbana-Champaign. Michigan: UMI Dissertation Services.

Closas, A. H. (2009). *Modelización estadística del rendimiento matemático con variables psicoeducativas en estudiantes universitarios*. Tesis doctoral, Universidad Pública de Navarra, España.

Closas, A. H., Sanz de Acedo, M. L. y Ugarte, M. D. (2011). Explanatory Model of the Relations Between Cognitive and Motivational Variables and Academic Goals. *Journal of Psychodidactics*, 1(16), 19-38.

Corral-Verdugo, V. (2002). Structural equation modeling. En R. B. Bechtel y A. Churchman (Eds.), *Handbook of environmental psychology* (pp. 256-270). New York: Wiley & Sons.

Davis, J. (1985). *The logic of causal order*. Beverly Hills, CA: Sage

Goldberger, A. S. y Duncan, O. D. (1973). *Structural equation models in the social sciences*. New York: Academic Press.

Harris, R. J. (1985). *A Primer of Multivariate Statistics* (2a. ed.). Orlando, FL: Academic Press.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81, 945-960.

Jackson, D. L. (2007). The effect of the number of observations per parameter in misspecified confirmatory factor analytic models. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(1), 48-76.

James, L. R., Mulaik, S. A. y Brett, J. M. (1982). *Causal Analysis. Assumptions, Models and Data*. Beverly Hills, CA: Sage.

Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, 44, 443-477.

Jöreskog, K. G. y Sörbom, D. (1989). *Lisrel 7: A guide to the Program Applications*. Chicago, IL: SPSS Inc.

Kaplan, D. (2009). *Structural Equation Modeling: Foundations and Extensions* (2a. ed.). Newbury Park, CA: Sage.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3a. ed.). New York: Guilford Press.

Lunnenborg, C. E. y Abbott, R. D. (1983). *Elementary Multivariate Analysis for the Behavioural Sciences: Applications of Basic Structure*. New York: North-Holland.

Mackie, J. L. (1980). *The Cement of the Universe: A Study in Causation*. Oxford, UK: Clarendon Press.

Maxim, P. S. (2002). *Métodos cuantitativos aplicados a las ciencias sociales*. México: Oxford University Press.

Merton, R. K. (1983). *Teoría y Estructuras Sociales*. México: Fondo de Cultura Económica.

Moore, D. S. (1991). *Statistics. Concepts and controversies*. New York: Freeman. Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical and continuous latent variable indicators. *Psychometrika*, 43, 241-250.

Pearl, J. (2011). The Causal Foundations of Structural Equation Modeling. En R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling*. New York: Guilford Press. Obtenido el 21 de septiembre de 2011 en http://ftp.cs.ucla.edu/pub/stat_ser/r370.pdf

Satorra, A. y Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66(4), 507-514.

Scriven, M. (1971). The logic of cause. *Theory and Decision*, 2, 49-66.

Voght, W. P. (1993). *Dictionary of Statistics and Methodology* (pp. 141-142). Newbury Park, CA: Sage.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, 5, 161-215.

CURRÍCULUM VITAE

Antonio Humberto Closas se desempeña como Profesor y Director de Proyectos de Investigación en la Facultad de Ciencias Económicas de la Universidad Nacional del Nordeste y en la Facultad Regional Resistencia de la Universidad Tecnológica Nacional. Es Doctor por la Universidad Pública de Navarra (Pamplona, España), en el área de Estadística e Investigación Operativa. Sus líneas de investigación están asociadas con métodos estadísticos multivariantes aplicados a estudios sociales y económicos. Es coautor de libros y de publicaciones en revistas de impacto científico nacional e internacional, reconocidas en el Social Citation Index. Participó como expositor en diversas reuniones académicas realizadas en nuestro país y en el exterior, relacionadas con su especialidad.

hclosas@eco.unne.edu.ar