Capítulo 13

Modelación jerárquica en las finanzas públicas

Mario Ojeda & Fernando Velasco

M.Ojeda & F.Velasco
Benemérita Universidad Autónoma de Puebla. 3 Oriente 4 sur no. 104, Col Centro., Centro Histórico, 72000, Puebla.Universidad Veracruzana. Mueseo 133, Unidad Magisterial, 91010 Xalapa, Veracruz.
mojeda@uv.mx

M.Ramos, F.Miranda (eds.) *Optimización-Estocástica-Recursiva-Coherente-Sistémica y sus variantes (probabilidad, econometría y estadística aplicada),* Temas Selectos de Optimización-©ECORFAN-Santiago de Compostela, España, 2012.

Abstract

Hierarchical linear models are a general class of models that allow modeling in a variety of situations in which there are data that have a hierarchical structure. The hierarchical data structure occur frequently in studies of public finances, which are commonly analyzed variables measured on entities (states or provinces), which in turn are composed of sub-entities (towns or municipalities). It also presents the case of entities that are studied over a period of several years, so we have a sample of years nested entities. This paper provides an introduction to hierarchical linear models and illustrates its application to problems of public finances. The document is divided into three sections: the first provides a characterization of the data with hierarchical structure, in the second hierarchical linear models are developed in particular-that experience of the authors are the most application-, and finally in the third section it is presented two examples of application to public finances.

13 Introducción

Una de las tareas fundamentales de un científico de la economía y las finanzas públicas es la de encontrar patrones de asociación entre variables que le permitan probar hipótesis, desarrollar descripciones de fenómenos económicos en el espacio y en el tiempo. Por esta razón la modelación estadística ha encontrado en esta área del conocimiento tanto la inspiración para el desarrollo de teorías y metodologías como el espacio para la aplicación de las mismas. La retroalimentación es una constante en el desarrollo de tanto de la estadística teórica como de la economía y las finanzas. En este contexto, la econometría –en particular los modelos de regresión aplicados a la economía y las finanzas- ha tenido un desarrollo vertiginoso después de la década de los noventa del siglo pasado. Se abrió una importante vertiente de desarrollo con la aparición de los métodos de estimación –basados en principios bayesianos y algoritmos computacionales que aplican mínimos cuadrados iterativos (Raudenbush y Bryk, 2002)- para modelos lineales generales (que permiten la postulación de modelos más realistas a las situaciones de estudios espaciales, temporales, espaciotemporales). Es así que, las contribuciones que ofrece hoy en día la teoría y herramientas computacionales asociadas a los modelos lineales generales permiten que se cuente con una amplia gama de metodologías de modelación que hacen que los estudios en estas áreas de aplicación puedan abordar con mayor realismo y eficiencia las tareas de modelación estadística. Aunque se cuenta con una variedad de libros de texto y monografías sobre la modelación lineal multinivel –que también es denominada modelación lineal jerárquica- se requiere aún de la promoción de estas metodologías entre los profesionales, técnicos y científicos de las disciplinas particulares, lo que es posible lograr a través de trabajos suscintos, de nivel introductorio, autocontenidos, y que a la vez muestren con ilustraciones en problemas concretos el potencial de estas herramientas para la investigación. En este escrito se encontrará una introducción a la modelación lineal jerárquica enfatizando su aplicación en problemas relacionados a las finanzas públicas. Se presentará brevemente una caracterización de los datos con estructura jerárquica, se describirán algunos de los modelos lineales jerárquicos particulares de mayor uso y se presentará la ilustración de su aplicación con dos ejemplos de finanzas públicas.

El contenido se estructura en tres secciones, además de una introducción y las conclusiones. En la primera sección 13.1 se sientan las bases de la modelación jerárquica; el objetivo es introducir al lector en los conceptos básicos de la modelación jerárquica. En la sección 13.2, se presentan dos casos particulares de los modelos lineales jerárquicos: el modelo intercepto aleatorio y el modelo de pendientes aleatorias.

Finalmente en la sección 13.3, se presentan dos aplicaciones de esta metodología a problemas de finanzas públicas: en el primero de ellos se analiza la evolución del Gasto Público en Salud (GPS) y del Producto Interno Bruto (PIB), por entidad federativa en la República Mexicana, en un periodo de tiempo. En el segundo ejemplo se analiza la relación que existe entre los ingresos del sector eléctrico y del sector petrolero en relación con el sector primario desde 2003 hasta 2008, en las 32 entidades federativas que conforman a México.

En las finanzas públicas modernas están involucradas varias disciplinas, que van desde la sociología del estado pasando por la metodología de la investigación, hasta llegar a las tecnologías de la informática y las comunicaciones. Una de estas disciplinas es la estadística, que es una herramienta fundamental para la realización de procesos de investigación en las ciencias factuales que utilizan la investigación cuantitativa. Se debe de considerar que la metodología estadística comprende tres grandes pasos en el desarrollo de una investigación; 1) el diseño adecuado para la obtención de datos; 2) el análisis de éstos; y 3) la interpretación y presentación de los resultados en forma apropiada. Respecto al análisis de datos es de gran utilidad conocer además de las técnicas exploratorias univariadas y multivariadas, las técnicas de modelación estadística. La más conocida de éstas es la regresión lineal, que cuenta con una amplia promoción entre los economistas y especialistas en finanzas públicas (Gujarati, 2003; Montgomery, Peck y Vining. 2004). Cabe hacer notar que los modelos de regresión simple, múltiple, multivariante, para datos en series de tiempo, etc. pueden encontrar diversas situaciones en las que los supuestos se violan por la estructura de los datos (anidados o de clasificación cruzada), lo que hace que los resultados de la modelación no se correspondan con lo que dicta la teoría y hasta a veces el sentido común (Wooldridge, 2009). Por tal razón podemos decir que en diversas ocasiones las muestras o poblaciones de estudio presentan una estructura jerárquica o de datos estructurados en varios niveles de anidamiento. Los datos con estructura jerárquica son bastante comunes en diferentes áreas de las ciencias sociales, como en educación (los estudiantes aparecen agrupados en escuelas, escuelas en zonas, etc.), en salud (pacientes, hospitales, regiones, etc.) y en economía (estudios longitudinales, de grupos anidados de empresas, economía comparada de países, etc. Esta situación se presenta particularmente en los estudios que abordan las finanzas públicas, donde se analizan comúnmente variables que se miden sobre las entidades federativas, las cuales a su vez están formadas (y los datos se desagregan) por los municipalidades, y a veces es necesario llegar hasta el nivel de áreas geoestadísticas básicas (AGEB´s). Cuando el caso es el de las entidades federativas (estados o provincias) que se estudian en un periodo de varios años, se tiene un conjunto de series de tiempo (una para cada entidad), lo cual constituye una muestra anidada (años en entidades). En fin, que las estructuras de datos y poblaciones de referencia ordenadas jerárquicamente son muy frecuentes, con lo que los problemas –llamados multinivel- plantean la necesidad del uso de metodologías de modelación estadística adecuadas. Para tratar este tipo de problemas la metodología estadística cuenta con una serie de técnicas, métodos y modelos que en la actualidad están bien definidos y se encuentran disponibles junto con el software que permite su adecuada aplicación para plantear y resolver problemas de este tipo, a través de la postulación, ajuste y utilización de modelos para la interpretación de los fenómenos bajo estudio.

13.1 Modelación jerárquica

Los modelos lineales jerárquicos forman una clase general de modelos que permiten la modelación en una gran variedad de situaciones en las cuales se tienen datos que presentan una estructura jerárquica.

Estos modelos tienen una gran variedad de aplicaciones en diversas áreas, tales como: investigación educativa (efectividad de escuelas, logro escolar), biología (curvas de crecimiento, estudios genéticos), investigación social (análisis de encuestas, estudios de mercado), psicología (análisis de conducta), medicina (medidas repetidas), entre otras. Los modelos lineales jerárquicos tienen una gran historia, pero han recibido especial atención desde finales de la década de los ochenta, aunque sus orígenes se remontan varios años atrás. Recientes desarrollos en cómputo han permitido que se incremente el uso de modelos lineales jerárquicos en el análisis de datos con estructura jerárquica. Los Modelos lineales jerárquicos, son también conocidos bajo una gran variedad de nombres: Modelos de componentes de la varianza (Searle et al., 1992), Modelos de coeficientes aleatorios (Longford 1995), Modelos multinivel (Goldstein, 1995) o como Modelos de efectos mixtos (Laird y Ware, 1982; Littell, Milliken, Stroup y Wolfinger, 1996).

Los datos con estructura jerárquica surgen en una gran variedad de situaciones. Recordemos que la parte fundamental de un análisis de datos son las unidades de estudio. Éstas se definen como el conjunto de observaciones de las cuales obtenemos información y a través de las cuales los valores medidos variarán. Las unidades pueden ser de varios tipos de acuerdo al contexto del problema. Sin embargo, en el caso de la modelación jerárquica tienen una característica fundamental y ésta es que se encuentran anidadas, estructuradas o agrupadas en un cierto número de niveles o clasificaciones. Por ejemplo: investigaciones educativas frecuentemente están relacionadas con problemas de investigación de relaciones existentes entre alumnos y el grupo de clase en el que éstos se desenvuelven. El concepto general es que el alumno interactúa con el grupo de clase al cual éste pertenece. Generalmente los alumnos y el grupo de clase se conceptualizan como un sistema con estructura jerárquica, donde los alumnos y los grupos de clase son definidos en niveles separados de esta estructura jerárquica. En general supóngase que se tienen datos con estructura jerárquica; es decir, se tienen J grupos con n_j unidades en el j-ésimo grupo, $j = 1,...,J$. A cada grupo se le denomina unidad de nivel 2; así se tienen J unidades de nivel 2, y a cada unidad las n_j unidades en cada grupo se le denomina unidad de nivel 1; con lo que se tienen n_j unidades de nivel 1 en la j-ésima unidad de nivel 2. El número n_j de unidades de nivel 1 no tiene que ser necesariamente igual en cada unidad de nivel 2. Así también en investigaciones sociales se tratan problemas relacionados con la interacción entre los individuos en su contexto social, significando que las personas son influidas por los grupos sociales a los cuales pertenecen; generalmente los individuos y los grupos se conceptualizan como un sistema con estructura jerárquica, donde los individuos son las unidades de nivel 1 y los grupos sociales las unidades de nivel 2. En estudios sociales los miembros de una familia dentro de las familias son las unidades de nivel 1 y las familias las unidades de nivel 2; en estudios empresariales los empleados dentro de compañías. Los datos longitudinales o de curvas de crecimiento pueden representarse mediante un sistema con estructura jerárquica en la que las observaciones de medidas repetidas se conceptualizan anidadas dentro de sujetos.

También se presentan sistemas que tienen una estructura jerárquica pero de tres niveles; por ejemplo, en investigaciones educativas además de las relaciones existentes entre alumnos y el grupo de clase en el que éstos se desenvuelven, los grupos de clase están anidados dentro de escuelas, las cuales serian las unidades de nivel 3. En general en un sistema con estructura jerárquica se pueden presentar varios niveles.

En un sistema de estructura jerárquica puede ser de interés estudiar la relación existente entre una variable respuesta, la cual se mide en las unidades de nivel 1, y variables explicatorias las cuales se pueden medir en cada uno de los niveles de la estructura jerárquica.

Se tiene que para cada una de las n_j unidades de nivel 1 en la j-ésima unidad de nivel 2 se registraron mediciones sobre una variable respuesta y_{ij}, y sobre m variables explicatorias $x_1,...,x_m$; éstas se denominan variables explicatorias a nivel 1. Además se puede medir otro conjunto de variables explicatorias $w_1,...,w_q$ en cada una de las unidades de nivel 2, las que se denominan variables explicatorias a nivel 2. Por ejemplo, en estudios de educación se tienen estudiantes, que están anidados en grupos de clase, y puede ser de interés la calificación en determinada asignatura, obtenida por el estudiante al finalizar el curso; ésta sería la variable respuesta. Una posible variable explicatoria a nivel estudiante podría ser las horas de estudio dedicadas a tal asignatura por el estudiante; una posible variable explicatoria a nivel grupo de clase podría pensarse como los años de experiencia que tienen el profesor que imparte la asignatura en cada gripo de clase.

La importancia de los modelos lineales jerárquicos radica en que se puede tener una mejor comprensión de la variabilidad de los datos, pues permite conocer la varianza entre las unidades de nivel 1 y entre las unidades de nivel 2. En el modelo lineal jerárquico la varianza de la variable respuesta puede ser descompuesta como la suma de las varianzas nivel 1, σ_e^2 y nivel 2, σ_{u0}^2. Esta línea de investigación es muy potente, pues otras técnicas de análisis estadístico no permiten obtener esta información. Retomando el ejemplo que se ha presentado, si se utiliza un modelo de dos niveles, es posible llegar a conocer la variación que existe entre los grupos de clase y entre los estudiantes en cada grupo. Por otro lado al ajustarse un modelo de un solo nivel (Regresión ordinaria), se ignorarían los efectos de agrupamiento y por lo tanto, se obtendrían estimadores sesgados que conducirían a inferencias erróneas. Las técnicas usuales no están diseñadas para dividir la variación de esta manera y sólo estiman un término para explicar esta diferencia. En la modelación jerárquica esta variación presenta una estructura relevante susceptible de ser analizada y que aporta mucha información al problema.

Respecto al número de unidades que deben ser incluidas en cada nivel del modelo, es una de las preguntas más frecuentes cuando se utiliza este tipo de metodología. La respuesta a esta interrogante estará en función principalmente de los intereses del investigador y de las unidades de estudio. Por ejemplo, si el objetivo es estudiar la variación entre las universidades del país respecto al tiempo que tardan sus estudiantes de doctorado en obtener el grado, se necesitará información de varias universidades con el objetivo de obtener estimadores confiables. Esto significa que no se podría utilizar información sólo de dos universidades aunque se tuvieran datos de 500 estudiantes titulados en esa universidad. Goldstein (1995) recomienda que dada la magnitud de los efectos que es común encontrar entre las diferencias de las escuelas, se requiere información de al menos 25 centros escolares para proporcionar un estimador preciso de la varianza entre las escuelas. Por su parte, Snijders y Bosker (1993) señala que la robusticidad de las pruebas estadísticas usualmente depende del tamaño de la muestra y ha diseñado un software especializado, llamado PinT, de las siglas de Power Analysis in Two Level Designs para la determinación del tamaño de muestra óptimo en diseños multinivel (Véase Snijders, 2005).

13.2 Algunos modelos lineales jerárquicos

En esta sección se presentan dos casos particulares de los modelos lineales jerárquicos, el modelo intercepto aleatorio y el modelo de pendientes aleatorias.

Para analizar datos con estructura jerárquica se tienen que emplear técnicas estadísticas que tomen en cuenta dicha estructura.

En esta situación, es razonable postular un modelo de regresión que considere una posible diferencia entre las unidades de nivel 2, es decir, plantear un modelo de regresión tal que, para cada unidad de nivel 2, se tengan diferentes coeficientes de regresión. Bajo esta situación el modelo lineal jerárquico de dos niveles permite simultáneamente hacer un estudio de unidades de nivel 1 y un estudio de unidades de nivel 2, tomando en cuenta variables explicatorias para las unidades de nivel 1 y variables explicatorias para las unidades de nivel 2. En los modelos lineales jerárquicos cada uno de los niveles de la estructura jerárquica es representado formalmente con su propio submodelo. Un tratamiento y abundantes referencias acerca de estos modelos se puede encontrar en Goldstein (1987, 1995), Longford, (1993, 1995), Kreft y De Leeuw (1998), Snijders y Bosker, (1999), Raudenbush y Bryk, (2002), Hox, J. (2002).

En la actualidad existe software estadístico el cual permite analizar datos con estructura jerárquica de acuerdo al modelo apropiado, MLwiN, (Rasbash et al., 2009), S-PLUS (Pinheiro y Bates, 2000), SAS (Little, et al., 2002, Singer, 1998). Una revisión exhaustiva puede encontrarse en Kreft y De Leeuw (1998).

A continuación se describen algunos de los modelos lineales jerárquicos.

13.2.1 Modelo intercepto aleatorio

El caso más simple de un modelo lineal jerárquico es el denominado modelo intercepto aleatorio, el cual no contiene ni variables explicatorias a nivel 1, ni variables explicatorias a nivel 2. En este modelo solamente se tiene variabilidad entre las unidades de nivel 2 y dentro de las unidades de nivel 2. Este modelo puede ser expresado como un modelo donde la variable respuesta, y_{ij}, es la suma de una media general dada por β_{00}, un efecto aleatorio a nivel 2 dado por u_{0j}, y un efecto aleatorio a nivel 1 dado por e_{ij}; El modelo para la i-ésima unidad de nivel 1, la cual se encuentra en la j-ésima unidad de nivel 2, tiene la forma:

$$
\begin{aligned}
&y_{ij} = \beta_{00} + u_{0j} + e_{ij}, \\
&\mathrm{E}(e_{ij}) = 0, \quad \mathrm{Var}(e_{ij}) = \sigma_e^2, \qquad e_{ij} \sim \mathrm{N}(0, \sigma_e^2), \\
&\mathrm{E}(u_{0j}) = 0, \quad \mathrm{Var}(u_{0j}) = \sigma_{u0}^2.
\end{aligned}
\tag{13.1}
$$

Donde $\mathrm{N}(0, \sigma_e^2)$ denota la distribución normas con media 0 y varianza σ_e^2. Los parámetros en el modelo (1) son tres: El coeficiente β_{00} y los componentes de la varianza σ_e^2 y σ_{u0}^2. En el modelo intercepto aleatorio la varianza de la variable respuesta es descompuesta como la suma de las varianzas nivel 1, σ_e^2 y nivel 2, σ_{u0}^2,

$$
\mathrm{Var}(y_{ij}) = \sigma_{u0}^2 + \sigma_e^2. \tag{13.2}
$$

El modelo para el nivel 1 tiene la forma:

$$
y_{ij} = \beta_{0j} + e_{ij},
$$

Y el modelo para el nivel 2 tiene la forma:

$$\beta_{0j} = \beta_{00} + u_{0j}.$$

El coeficiente de correlación intraclase es unamedida del porcentaje de la variabilidad de la variable respuesta que es atribuida a las unidades de nivel 2, éste está dado por medio de:

$$\rho = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2}. \tag{13.3}$$

13. 2.2 Modelo intercepto aleatorio con una explicatoria a nivel 1

En el modelo intercepto aleatorio el valor esperado de la variable respuesta puede ser explicado en términos de variables explicatorias a nivel 1. Así la siguiente etapa es la inclusión de variables explicatorias a nivel 1, esto con el objetivo de tratar de explicar el comportamiento de la variable respuesta. Con una variable explicatoria a nivel 1 el modelo intercepto aleatorio tiene la forma:

$$\begin{aligned} &y_{ij} = \beta_{00} + \beta_1 x_{ij} + u_{0j} + e_{ij}, \\ &\mathrm{E}(e_{ij}) = 0, \quad \mathrm{Var}(e_{ij}) = \sigma_e^2, \qquad e_{ij} \sim \mathrm{N}(0, \sigma_e^2), \\ &\mathrm{E}(u_{0j}) = 0, \quad \mathrm{Var}(u_{0j}) = \sigma_{u0}^2. \end{aligned} \tag{13.4}$$

El modelo (4) se denomina modelo intercepto aleatorio con una variable explicatoria a nivel 1. Los parámetros en el modelo (4) son cuatro: los coeficientes de regresión β_{00} y β_1, y las varianzas σ_e^2 y σ_{u0}^2.

En el modelo intercepto aleatorio con una variable explicatoria a nivel 1 la varianza de la variable respuesta puede ser descompuesta como la suma de las varianzas nivel 1, σ_e^2 y nivel 2, σ_{u0}^2, de la siguiente manera:

$$\mathrm{Var}(y_{ij}) = \sigma_{u0}^2 + \sigma_e^2. \tag{13.5}$$

A los términos σ_{u0}^2, σ_e^2 se les denomina componentes de la varianza.

El modelo para el nivel 1 tiene la forma:

$$y_{ij} = \beta_{0j} + \beta_1 x_{ij} + e_{ij},$$

Y el modelo para el nivel 2 tiene la forma:

$$\beta_{0j} = \beta_{00} + u_{0j}.$$

13. 2.3 Modelo intercepto aleatorio con variables explicatorias a nivel 1

Al igual que en el modelo de regresión múltiple, más de una variable explicatoria a nivel 1 puede ser usada en el modelo intercepto aleatorio. La generalización del modelo (4) para incluir más variables explicatorias a nivel 1; es decir, el modelo intercepto aleatorio con varias variables explicatorias a nivel 1 tiene la forma:

$$
\begin{aligned}
&y_{ij} = \beta_{00} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \cdots + \beta_m x_{mij} + u_{0j} + e_{ij}, \\
&\mathrm{E}(e_{ij}) = 0, \quad \mathrm{Var}(e_{ij}) = \sigma_e^2, \qquad e_{ij} \sim \mathrm{N}(0, \sigma_e^2), \\
&\mathrm{E}(u_{0j}) = 0, \quad \mathrm{Var}(u_{0j}) = \sigma_{u0}^2.
\end{aligned}
\tag{13.6}
$$

El modelo (6) se denomina modelo intercepto aleatorio con varias variables explicatorias a nivel 1. Los parámetros del modelo intercepto aleatorio con varias variables explicatorias a nivel 1 son $m+3$; los $m+1$ coeficientes de regresión $\beta_{00}, \beta_1, \beta_2, ..., \beta_m$ y los componentes de la varianza σ_e^2 y σ_{u0}^2.

En el modelo intercepto aleatorio con varias variables explicatorias a nivel 1 la varianza de la variable respuesta puede ser descompuesta como la suma de las varianzas nivel 1, σ_e^2 y nivel 2, σ_{u0}^2,

$$
\mathrm{Var}(y_{ij}) = \sigma_{u0}^2 + \sigma_e^2. \tag{13.7}
$$

El modelo para el nivel 1 tiene la forma:

$$
y_{ij} = \beta_{0j} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \cdots + \beta_m x_{mij} + e_{ij},
$$

y el modelo para el nivel 2 tiene la forma:

$$
\beta_{0j} = \beta_{00} + u_{0j}.
$$

13. 2.4 Modelo de pendientes aleatorias

En el modelo lineal jerárquico intercepto aleatorio con variables explicatorias a nivel 1, solo el intercepto se supone aleatorio, mientras que los demás coeficientes de regresión se suponen fijos para todas las unidades de nivel 2. En ocasiones la relación entre las variables explicatorias y la variable respuesta puede ser diferente en las unidades de nivel 2. Lo anterior da surgimiento al modelo de pendientes aleatorias. En este modelo los coeficientes de algunas o de todas las variables explicatorias están variando entre las unidades de nivel 2, es decir, la relación existente entre cada una de las variables explicatorias y la variable respuesta no es la misma en todas las unidades de nivel 2. Como los coeficientes varían entre las unidades de nivel 2 se les denomina coeficientes aleatorios. Para el caso de una variable explicatoria a nivel 1 lo anterior se expresa en el siguiente modelo:

$$
\begin{aligned}
&y_{ij} = \beta_{00} + \beta_{10} x_{ij} + u_{0j} + u_{1j} x_{ij} + e_{ij}, \\
&\mathrm{E}(e_{ij}) = 0, \quad \mathrm{Var}(e_{ij}) = \sigma_e^2, \\
&\mathrm{E}(u_{0j}) = 0, \quad \mathrm{Var}(u_{0j}) = \sigma_{u0}^2, \quad \mathrm{E}(u_{1j}) = 0, \quad \mathrm{Var}(u_{1j}) = \sigma_{u1}^2, \\
&\mathrm{Cov}(u_{0j}, u_{1j}) = \sigma_{u01}, \quad \mathrm{Cov}(u_{kj}, e_{ij}) = 0.
\end{aligned}
\tag{13.8}
$$

El cual se denomina modelo de pendientes aleatorias con una variable explicatoria a nivel 1.

Los parámetros del modelo de pendientes aleatorias con una variable explicatoria a nivel 1 son seis: los coeficiente de regresión β_{00} y β_{10}, y los componentes de la varianza σ_e^2, σ_{u0}^2, σ_{u1}^2 y σ_{u01}.

En el modelo de pendientes aleatorias con una variable explicatoria a nivel 1 la varianza de la variable respuesta se descompone de la siguiente forma:

$$
\mathrm{Var}(y_{ij}) = \sigma_{u0}^2 + \sigma_{u1}^2 x_{ij}^2 + 2\sigma_{u01} x_{ij} + \sigma_e^2. \tag{13.9}
$$

De la ecuación (9) se tiene que en el modelo de pendientes aleatorias con una variable explicatoria a nivel 1 la varianza de la variable respuesta depende de la variable explicatoria a nivel 1, x_{ij}.

El modelo para el nivel 1 tiene la forma:

$$
y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + e_{ij},
$$

y el modelo para el nivel 2 tiene la forma:

$$
\beta_{0j} = \beta_{00} + u_{0j}, \qquad \beta_{1j} = \beta_{10} + u_{1j},
$$

Aquí se observa que los coeficientes de regresión β_{0j}, así como β_{1j} son aleatorios, es decir cambian de unidad de nivel 2 a unidad de nivel 2.

Existen muchas variantes de los modelos anteriores, ya sea añadiendo más variables explicatorias a nivel 1 o variables explicatorias a nivel 2, o más aún añadiendo más niveles.

13.3 Prototipos

En esta sección se muestran dos problemas de finanzas públicas en los que se aplicó la modelación jerárquica, el primero de ellos se analiza la evolución del Gasto Público en Salud (GPS) y del Producto Interno Bruto (PIB), por entidad federativa de la República Mexicana. Paralelamente, a través de una modelación jerárquica se determinó la relación entre GPS y el PIB, para analizar si hay variabilidad entre esta relación durante el periodo 2000-2008 por entidad federativa, teniendo como resultado, que sí existe una relación directa entre el gasto en salud y el crecimiento económico, y que sí hay variabilidad significativa entre los años bajo estudio y entre las entidades federativas.

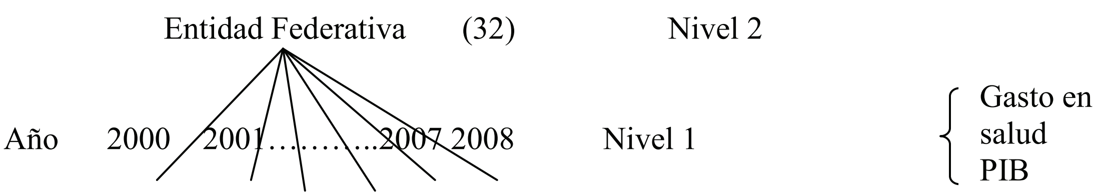
En el segundo ejemplo se analiza la relación que existe entre los ingresos del sector eléctrico y del sector petrolero en relación con el sector primario; para la obtención de la información se recurre a una base de datos del Instituto Nacional de estadística Geografía e Informática (INEGI) de México, respecto a los ingresos de petróleo y de energía eléctrica desde 2003 hasta 2008, en las 32 entidades federativas de México. Se aplicaron una serie de modelos multinivel para analizar la influencia del tiempo, de los ingresos del sector eléctrico y de los ingresos del sector petrolero en el sector primario y determinar si existe variabilidad entre las entidades federativas y los 6 años del periodo de estudio. Para mayor información sobre estos ejemplos los autores ponen a disposición los trabajos más extensos donde se presentan estas aplicaciones (Ojeda, et al., 2011)

Ejemplo 1 Análisis del gasto en salud y su relación con el crecimiento económico de México en el periodo 2000-2008

En este ejemplo se analiza la relación que existe entre el GPS y el PIB. Los datos utilizados para la realización del análisis fueron tomados de la página electrónica del Sistema Nacional de Información en Salud (SINAIS), entidad dependiente de la Secretaría de Salud (SSA) de México, así como del Instituto Nacional de INEGI, contando con una muestra de 288 observaciones, correspondientes a la información del PIB y del GPS, en el periodo 2000-2008, de cada entidad federativa.

Dado que la información que se obtuvo presenta una estructura de anidamiento, y se desea modelar la relación existente entre el PIB de cada entidad por año con el GPS de cada entidad por año, se hizo uso de la modelación jerárquica, considerando un modelo de dos niveles (Goldstein, 1995; Raudenbush y Bryk, 2002); como unidades de nivel 1 se tomaron los 9 años que comprende este estudio y como unidades de nivel 2 las 32 entidades federativas (Figura 1).

Figura 13.1 Diagramas de unidad para la estructura jerárquica de los datos bajo estudio.



A través de la modelación jerárquica, se pretende tener una mejor comprensión de la variabilidad del PIB, pues permite conocer la varianza entre los años y las entidades federativas respecto al PIB, tomando en consideración la posible relación con el tiempo y el GPS. El modelo para ajustar la relación está dado por medio de la ecuación:

$$PIB_{ij} = \beta_{0j} + \beta_1 TIEMPO + \beta_2 GPS_{ij} + e_{ij}$$

$$\beta_{0j} = \beta_0 + u_{0j} \qquad i = 1,\ldots,9$$

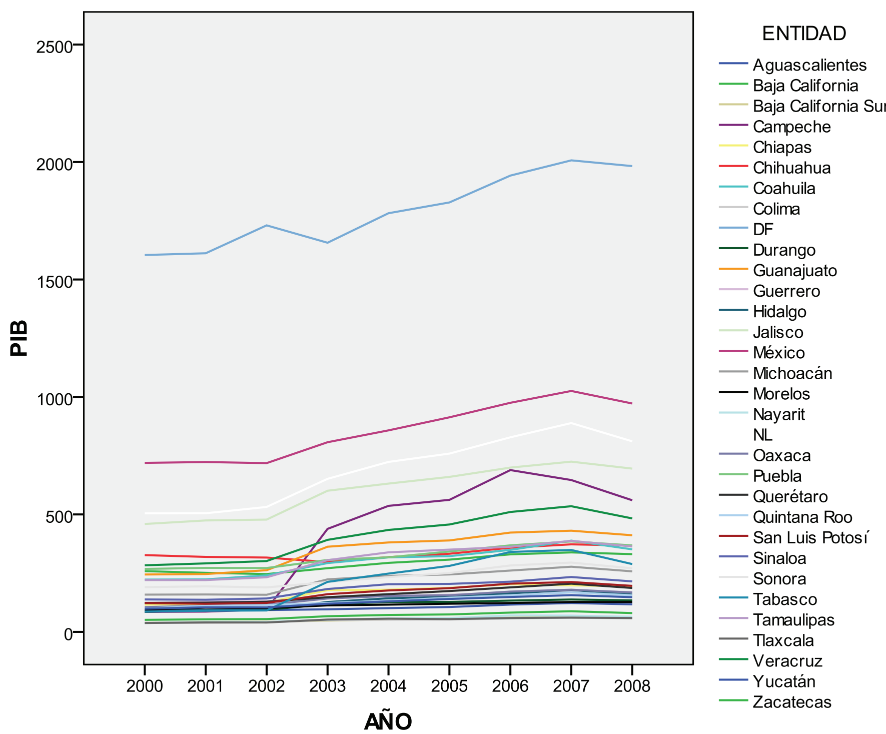
$$e_{ij} \sim N\left(0, \sigma_e^2\right) \qquad j = 1,\ldots,32$$

$$u_{0j} \sim N\left(0, \sigma_{u0}^2\right)$$

Donde β_0 denota el intercepto o la media global del PIB para todas las entidades federativas en todos los años; β_1 y β_2 constituyen la pendiente o el cambio en la media del PIB, cuando hay un cambio unitario en cada variable explicatoria $TIEMPO$, y GPS, respectivamente, manteniendo las otras variables constantes, e_{ij} denota el error aleatorio correspondiente a la i-ésima unidad de nivel 1 (año) en la j-ésima unidad de nivel 2 (entidad federativa) y u_{0j} denota el j-ésimo error aleatorio a nivel 2. Con este modelo, lo que interesa es conocer si alguna variable como el $TIEMPO$ o el GPS influyen en el comportamiento del PIB.

Al relacionar el PIB con los años del periodo de estudio para cada entidad federativa, se aprecia en la Figura 2 una primera aproximación de la relación lineal existente entre ambas variables. De la Figura 2, se observa una tendencia de crecimiento a través de los años del PIB, además se observa que hay una variabilidad entre las entidades respecto al PIB la cual se mantiene durante el periodo de estudio 2000-2008.

Figura 13.2 Relación entre el PIB y el tiempo para cada entidad federativa.



Para corroborar los factores que contribuyen a explicar por qué hay variación entre los años y entre las entidades federativas respecto al PIB, se ajustaron 3 modelos multinivel.

Los resultados de las estimaciones se muestran en la tabla 13.1. En el modelo (1), modelo intercepto aleatorio, en el que no se consideran las variables explicatorias, los resultados del ajuste muestran que se tiene un PIB en promedio de 292.745 millones de pesos en cada entidad federativa por año, además de que existe variación tanto entre los años del periodo de estudio, así como se presenta variación entre las 32 entidades federativas.

De acuerdo al coeficiente de correlación intraclase, el porcentaje de la variabilidad del PIB atribuida a las entidades federativas es de aproximadamente el 95% y solo un 5% a los años, es decir, la variabilidad del PIB es atribuida en gran parte a las entidades federativas.En el modelo (2) se introdujo la variable años como variable explicatoria, se mantuvo fija la pendiente y el intercepto aleatorio, los resultados del ajuste muestran que la variable TIEMPO sí resulta significativa, es decir, que cada año el PIB de las entidades federativas se incrementa en promedio en 18.69 millones de pesos. También se observa que la variación entre los años y entre las entidades es significativa. Sin embargo, la varianza entre las entidades se mantiene alta (57.495), mientras la varianza del PIB a nivel de los años disminuyo de 5461.39 a 3257.7 respecto al modelo intercepto aleatorio.

Tabla 13.1 Resultados de las estimaciones.

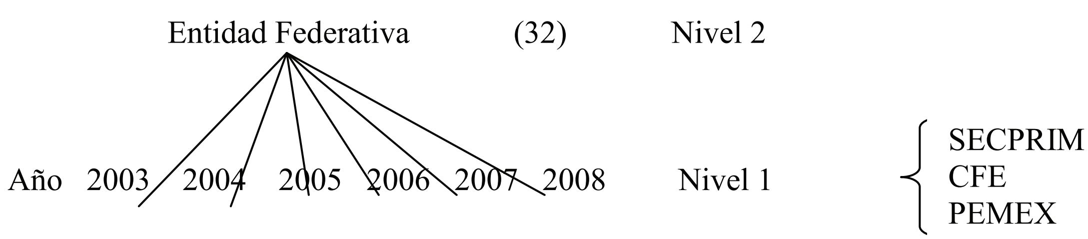
	Modelo intercepto aleatorio (1)	Modelo intercepto aleatorio con el TIEMPO (2)	Modelo intercepto aleatorio con el TIEMPO y GPS (3)
Parámetros fijos			
β_0 (Intercepto)	292.745 (58.00)	198.05 (58.2)	259.9 (22.3)
β_1 (TIEMPO)		18.69 (1.303)	6.32 (1.401)
β_2 (GPS)			29.97 (1.49)
Componentes de la varianza			
Nivel 2			
σ^2_{u0}	107054.023	106952.922	14933.551
Nivel 1			
σ^2_e	5461.385	3257.703	1136.818
Desviación			
	4167.062	3329.02	2903.018

En el modelo (3), se introdujo adicionalmente el *GPS* y se modela como fija. En los resultados mostrados en la Tabla 13.1, se observa que el *GPS* es significativo, esto quiere decir, que ante un cambio unitario en el GPS de cada entidad, el PIB se incrementan en 29.97 millones de pesos, manteniendo la variable *TIEMPO* fija, cabe destacar que al introducir la variable *GPS* al modelo, el *TIEMPO* influye de manera distinta, ahora cada año el PIB de cada entidad federativa se incrementa en promedio en 6.32 millones de pesos. Al comparar los modelos (2) y (3), se aprecia que la varianza a nivel entidad disminuyó de 106952.9 a 14933.6, y la varianza a nivel año también presenta una reducción de 3257.7 a 1136.8. También hay una disminución en el valor de la deviance de 3329 a 2903 es decir una reducción de 426, que al compararlo con una distribución χ^2 con 1 grado de libertad, resulta significativa. Lo que indica que el modelo (3) es más adecuado para el ajuste de los datos.

Ejemplo 2 Influencia del sector eléctrico y petrolero en la producción primaria 2003-2008

En este ejemplo se analiza la relación que existe entre los ingresos del sector eléctrico y del sector petrolero en relación con el sector primario; para la obtención de la información se recurre a una base de datos del INEGI, respecto a los ingresos de petróleo y de energía eléctrica desde 2003 hasta 2008, en las 32 entidades federativas.Dentro de la planeación económica y política de los ingresos públicos se puede hablar de recursos provenientes de la tributación que aporta la población con el fin de cumplir sus funciones públicas, o bien, por el producto de los ingresos que le otorgan los entes estatales que aprovechan recursos de la nación, como es el caso en México de Petróleos Mexicanos (PEMEX) y Comisión Federal de Electricidad (CFE). La teoría que justifica este tipo de intervención del Estado a través de una empresa pública es la del Estado de Bienestar (Keynes, 1981), para poder estabilizar los impactos negativos que ha traído consigo el capitalismo mismo, no tanto de una manera "prudente" sino necesaria a las necesidades de cada país, haciendo alusión a un análisis multi e interdisciplinario de todos los factores que inciden en las finanzas públicas de dicho país, con lo que se alinea a la visión de las Finanzas Públicas Modernas. Hoy en día, uno de los sectores que más atención requieren por parte del Estado Mexicano es el sector primario (SECPRIM), dado que la desigualdad regional, la rápida liberalización comercial y la creciente generación e incorporación de innovaciones tecnológicas, han arrasado con la producción agropecuaria nacional, ocasionándose un abultado déficit agropecuario externo de acuerdo a la FAO (2011) de aproximadamente 3.5 millones de dólares anuales durante 2000 a 2003, lo cual refleja el estado de dependencia alimentaria en que se ha desenvuelto la economía mexicana por más de dos décadas.Dado que la información que se obtuvo presenta una estructura de anidamiento, y se desea modelar la relación existente entre los ingresos del SECPRIM con los ingresos del sector eléctrico y del sector energético, se hizo uso de la modelación jerárquica, haciendo uso de un modelo de dos niveles. Como unidades de nivel 1 se tomaron los 6 años que comprende este estudio y como unidades de nivel 2 se tomaron a las 32 entidades federativas (Figura 3).

Figura 13.3 Diagramas de unidad para la estructura jerárquica de los datos bajo estudio.



A través de la modelación jerárquica, se pretende tener una mejor comprensión de la variabilidad de los ingresos del SECPRIM, pues permite conocer la varianza entre los años y las entidades federativas respecto a los ingresos del SECPRIM. El modelo propuesto está dado por la siguiente ecuación:

$$SECPRIM_{ij} = \beta_{0j} + \beta_1 TIEMPO_{ij} + \beta_2 CFE_{ij} + \beta_3 PEMEX_{ij} + e_{ij}$$

$$\beta_{0j} = \beta_0 + u_{0j} \qquad i = 1,...,6$$

$$e_{ij} \sim N\left(0, \sigma_e^2\right) \qquad j = 1,...,32$$

$$u_{0j} \sim N\left(0, \sigma_u^2\right)$$

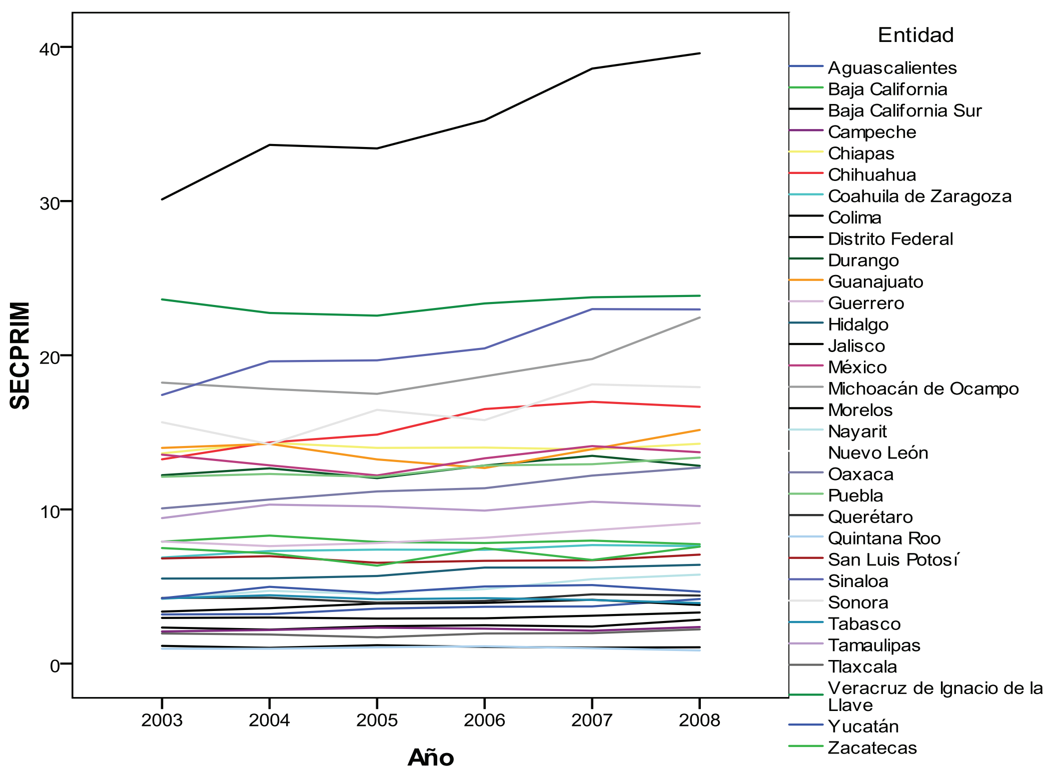
Donde β_0 denota el intercepto o la media global de los ingresos del sector primario para todas las entidades federativas en todos los años; β_1, β_2 y β_3 constituyen la pendiente o el cambio en la media de los ingresos del SECPRIM, cuando hay un cambio unitario en cada variable explicatoria tiempo ($TIEMPO$), ingresos de la CFE e ingresos de $PEMEX$, respectivamente, manteniendo las otras variables constantes, e_{ij} denota el error aleatorio correspondiente a la i-ésima unidad de nivel 1 en la j-ésima unidad de nivel 2 y u_{0j} denota el j-ésimo error aleatorio a nivel 2.

Con este modelo, lo que interesa es conocer si alguna variable como $TIEMPO$, los ingresos de la CFE de la entidad o los ingresos de $PEMEX$, influyen en el comportamiento de los ingresos de SECPRIM. Para validar los resultados del modelo, se comprobó el cumplimiento de los supuestos de normalidad de los errores en los dos niveles.

Al relacionar los ingresos del SECPRIM, con los años del periodo de estudio para cada entidad federativa, se aprecia una primera aproximación de la relación lineal existente entre SECPRIM y el tiempo.

De la Figura 4, se tiene que no se observa una tendencia a través de los años, pero se observa que hay una variabilidad entre las entidades respecto a los ingresos del SECPRIM la cual se mantiene durante el periodo de estudio.

Figura 13.4 Relación entre los ingresos del SECPRIM y los años del periodo de estudio (2003-2008).



Para corroborar los factores que contribuyen a explicar la posible variación entre los años y entre las entidades federativas respecto a los ingresos del SECPRIM, se ajustaron 4 modelos multinivel. Los resultados de las estimaciones se muestran en la Tabla 13.2.

En modelo (1), modelo intercepto aleatorio, los resultados del ajuste muestran que se tiene un ingreso promedio de 9,457 millones de pesos en cada entidad federativa por año, además de que existe variación tanto entre los años como entre las entidades federativas, siendo aproximadamente el 98% de la variación de los ingresos atribuida a las entidades federativas.

En el modelo (2) se introdujo la variable años como variable explicatoria, se mantuvo fija la pendiente y el intercepto aleatorio, los resultados del ajuste muestran que la variable TIEMPO sí resulta significativa, es decir, que cada año los ingresos del SECPRIM se incrementa en promedio en 0.254 miles de millones de pesos.

También se observa que la variación entre los años y entre las entidades es significativa. Sin embargo, la varianza entre las entidades se mantiene alta (57.495), mientras la varianza de los ingresos del SECPRIM a nivel de los años disminuyó de 0.93 a 0.71.

Tabla 13.2 Resultados de las estimaciones.

	Modelo intercepto aleatorio (1)	Modelo intercepto aleatorio con el TIEMPO (2)	Modelo intercepto aleatorio con el TIEMPO y la CFE (3)	Modelo intercepto aleatorio con el TIEMPO, la CFE y PEMEX (4)
Parámetros fijos				
β_0 (Intercepto)	9.475 (1.342)	6.811 (1.393)	8.882 (1.304)	7.503 (1.821)
β_1 ($TIEMPO$)		0.254 (0.036)	0.169 (0.053)	0.364 (0.377)
β_2 (CFE)			0.121 (0.056)	0.117 (0.056)
β_4 ($PEMEX$)				-0.130 (0.250)
Comp. de la varianza				
Nivel 2				
σ^2_{u0}	57.458 (14.634)	57.495 (14.633)	53.203 (13.297)	53.721 (13.313)
Nivel 1				
σ^2_e	0.933 (0.104)	0.712 (0.0799)	0.694 (0.078)	0.693 (0.077)
Desviación	717.4	678.0	670.994	670.723

En el modelo (3), se introdujo adicionalmente los ingresos de la CFE y se modeló como fija.

En los resultados mostrados en la tabla 13.1, se observa que la variable *CFE* es significativa, esto quiere decir, que ante un cambio unitario en los ingresos de la CFE de cada entidad, los ingresos del SECPRIM se incrementan en 0.121 miles de millones de pesos, manteniendo la variable *TIEMPO* constante, cabe destacar que al introducir la variable *CFE* al modelo, el *TIEMPO* influye de manera distinta, ahora cada año los ingresos del SECPRIM se incrementa en 0.169 miles de millones de pesos en promedio. Al comparar los modelos (2) y (3), se aprecia que la varianza a nivel entidad disminuyó de 57.495 a 53.203, y la varianza a nivel año también presenta una reducción de 0.712 a 0.694.

También hay una disminución en el valor de la deviance de 678 a 670.994, es decir una reducción de 7.006, que al compararlo con una distribución χ^2 con 1 grado de libertad, resulta significativa. Lo que indica que el modelo (3) está mejor ajustado a los datos.

En el modelo (4) se introdujo adicionalmente los ingresos de *PEMEX* y se modela como fija. En los resultados mostrados en la tabla 13.1, se observa que la variable *PEMEX* no es significativa.

Al comparar los modelos (3) y (4) hay una disminución en el valor de la deviance de 670.994 a 670.723 es decir una reducción de 0.271, que al compararlo con una distribución χ^2 con 1 grado de libertad, resulta no significativa. Lo que indica que los ingresos del sector petrolero no ayudan a explicar el comportamiento de los ingresos del SECPRIM, en el periodo bajo estudio.

13.4 Conclusiones

La modelación estadística permite la construcción empírica de modelos –los modelos ajustados a los datos-, con lo que es posible desarrollar descripciones y explicaciones, y probar hipótesis respecto al comportamiento de los fenómenos en muchas áreas de la ciencia y la técnica. La aplicación correcta de la modelación implica la postulación realista de ecuaciones que establecen relaciones causales para describir el fenómeno bajo estudio.

Esta circunstancia enfrenta al modelador al reto de considerar variables que se miden a diferentes niveles de agregación de las unidades de estudio además de que se debe considerar la estructura de anidamientos y entrelazamientos de la población de referencia o de la muestra de la que se obtienen los datos. Esta problemática ha trazado una línea de desarrollo para la modelación que se expresa en dos vertientes: (1) modelos cada vez más generales y más complejos, y (2) métodos de estimación y herramientas de evaluación para la selección de los modelos más parsimoniosos adecuados a cada situación.

Podemos concluir que la modelación lineal jerárquica –o modelación lineal multinivel- es un conjunto de metodologías que permiten mayor realismo y pertinencia de las investigaciones económicas y de finanzas públicas. Así mismo, que es necesario desarrollar habilidades especiales asociadas a la tarea de postulación de modelos particulares asociados a situaciones concretas. Es necesario explotar correctamente las herramientas exploratorias a fin de contar con suficientes elementos sobre la razonabilidad del modelo y de los supuestos subyacentes.

En este sentido se establece la recomendación de integrar equipos de trabajo donde se integre un especialista en estadística que maneje con solvencia las herramientas exploratorias y los elementos computacionales asociados a la estimación, diagnóstico y selección de modelos en este contexto.

Las aplicaciones concretas requieren no sólo del conocimiento de la teoría del fenómeno bajo estudio, sino así mismo de los principios y procedimientos de la modelación estadística. Cabe destacar que, en la experiencia de los autores, la colaboración inter y multidisciplinaria es estrictamente necesaria para hacer buenas aplicaciones de la modelación multinivel en finanzas públicas.

Referencias

Goldstein, H., Multilevel Models en Educational and Social Research, Griffin, London, 1987.

Goldstein, H., Multilevel Statistical Models, Second Edition, New York, Halsted Press, 1995.

Gujarati, D., Econometría. Cuarta Edición, México, McGraw-Hill Interamericana, 2003.

Hox, J.J., Multilevel Analysis, Techniques and Applications. Lawrence Erlbaum Associates, 2002.

Keynes, J. M., Teoría de la Ocupación, el Interés y el Dinero, México, Fondo de Cultura Económica, 1981.

Kreft, I. and De Leeuw, J., Introducing Multilevel Modeling, Newbury park Sage Publications, 1988.

Laird, N. M., & Ware, J. H., (1982), “Random effects models for longitudinal data”. Biometrics, 38, pp. 963-974.

Little,R. C., Stroup, W. W. and Freund, R. J., SAS for Linear Models. Cary, NC: SAS Institute, Inc., 2002.

Longford, N.T., Random Coefficient Models, New York, Oxford: University Press, 1993.

Longford, N., Random Coefficient Models. In Arminger, G., Cogg, C. C., Sobel, M. E. (eds.), Handbook of Statistiscal Modeling for the Social and Behavioral Sciences, New York, Plenum Press, 1995.

Montgomery, D., Peck, E. y Vining, G., Introducción al Análisis de Regresión Lineal. México, CECSA. Primera reimpresión, 2004.

Ojeda, M.M., Velasco, F., Cruz, C. y Tapia, P., Metodología Estadística Aplicada a las Finanzas Públicas. Xalapa, México, 2011.

Pinheiro, J. C., & Bates, D. M., Mixed-Effects Models in S and S-plus. New York: Springer-Verlag, 2000.

Raudenbush, S.W. and Bryk, A.S., Hierarchical Linear Models: Applications and Data Analysis Methods. Second edition, Newbury Park, CA: Sage, 2002.

Rasbash, J., Steele, F., Browne, W.J. y Goldstein, H., A User’s Guide to Mlwin. Version 2.10. Centre for Multilevel Modelling. University of Bristol, 2009.

Searle, S.R., Casella, G., McCulloch, C.E., Variance Components, New York, Wiley & Sons, 1992.

Singer, J. D. (1998), Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. Journal of Educational and Behavioral Statistics. 24(4), pp. 323-355.

Snijders, T.A.B., Power and Sample Size in Multilevel Linear Models; in B.S. Everitt and D.C. Howell (eds.), Encyclopedia of Statistics in Behavioral Science. Chicester. Wiley (3), 1570-1573, 2005.

Snijders, T.A.B. and Bosker, R.J. (1993), “Standard errors and simple sizes for two-level research”, Journal of Educational Statistics, 18: 237–259.

Snijders, T.A.B. and Bosker R. J., Multilevel Analysis. An Introduction to basic and Advanced Multilevel Modeling, Newbury Park/London/New Delhi: Sage Publications, 1999.

Wooldridge, J.M., Introducción a la Econometría: Un Enfoque Moderno, Cuarta Edición, Cegage, Learning Editores, 2009.