

# **Potencial de logro educativo, entorno socioeconómico y familiar: una aplicación empírica con variables latentes para Colombia<sup>1</sup>**

*Individual potential of educational achievement, socioeconomic and family background: an empiric application with latent variables for Colombia*

*Potencial de desempenho educativo, entorno socioeconômico e familiar: uma aplicação empírica com variáveis latentes para Colômbia*

**HARVY VIVAS PACHECO**

*Docente del Departamento de Economía de la Universidad del Valle, Cali  
harvivas@univalle.edu.co*

**JUAN BYRON CORREA FONNEGRA**

*Docente del Departamento de Economía de la Universidad del Valle, Cali  
jbcorrea@univalle.edu.co*

**JORGE ANDRÉS DOMÍNGUEZ MORENO**

*Economista Universidad del Valle, miembro del Grupo de Investigación en Economía Laboral,  
Cali  
andresdomin@hotmail.com*

Recibido: 01.08.11  
Aprobado: 12.09.11

---

1 Especial agradecimiento a Oscar Fabián Riomaña Trigueros y a Víctor Hugo Zuluaga González por su colaboración y apoyo en la elaboración de este artículo. Igualmente, al estadístico Héctor Fabio Ramírez, coordinador del laboratorio de estadística social de la Facultad de Ciencias Sociales y Económicas de Univalle por sus contribuciones en las discusiones del modelo. Agradecemos los valiosos comentarios y sugerencias de los dos evaluadores anónimos que tuvieron a su cargo la revisión del artículo.

## Resumen:

Esta investigación propone una estrategia empírica de estimación del potencial individual de logro educativo para Colombia (PCE) a través de modelos de ecuaciones estructurales (SEM). El potencial de logro educativo es una variable latente que proviene de la combinación multidimensional de los efectos de vecindad y del background familiar. Un aporte de la investigación radica en la utilización de métodos bayesianos de estimación de las relaciones estructurales sugeridas por la teoría y compara los resultados con los obtenidos mediante técnicas de máxima verosimilitud. Estos resultados aportan evidencia de que la calidad de los entornos locales (efectos de vecindad) presentan los efectos más grandes sobre el PCE. Las estimaciones fueron consistentes por los dos métodos utilizados (estimación bayesiana y máxima verosimilitud, M.L.).

**Palabras clave:** Logro Educativo, Modelos de Ecuaciones Estructurales, Estimación Bayesiana, Efectos de Vecindad, Background Familiar.

## Abstract:

This article proposes a method to estimate the individual potential of educational achievement (PCE) for Colombia using Structural Equation Model approach (SEM). The latent variable PCE is derived from a multidimensional array of the neighborhood effects and the family background. The main contribution of this study was to advance in Bayesian estimates for the structural relations suggested by the theory and to compare these outcomes with those derived from Maximum Likelihood Statistical techniques (M.L). The results show evidence that the quality of local linkages effects (specifically the neighborhood environment) makes the largest contribution to variation in PCE. This conclusion was consistent when estimated by the two methodologies.

**Key Words:** Educational Achievement, Structural Equation Models, Bayesian Estimation, Neighborhood Effects, Family Background.

**JEL classification:** D1, I21, I32.

## Resumo:

Esta pesquisa propõe uma estratégia empírica para estimar o potencial individual de desempenho educacional para a Colômbia (PCE) através da modelagem de equações estruturais (SEM). O potencial do desempenho educativo é uma variável latente que vem da combinação multidimensional de efeitos de vizinhança e do background familiar. A contribuição da pesquisa reside na utilização de métodos bayesianos de estimação das relações estruturais sugeridas pela teoria, comparando os resultados com aqueles obtidos pela utilização de técnicas de máxima verossimilhança. Estes resultados fornecem evidência de que a qualidade dos ambientes locais (os efeitos de vizinhança) tem o maior efeito sobre o PCE. As estimações foram consistentes para ambos os métodos (estimação bayesiana e máxima verossimilhança, ML).

**Palavras-chave:** Desempenho Educativo, Modelos de Equações Estruturais, Estimación Bayesiana, Efeitos de Vizinhança, Background familiar.

## 1. Introducción

Este artículo propone una estrategia empírica de estimación del potencial de logro educativo mediante la utilización de métodos de análisis estructural. El potencial de logro mide la capacidad de producción esperada de las realizaciones educativas que se genera en el interior de los hogares, de acuerdo con los perfiles socioeconómicos y sociodemográficos de sus integrantes, en particular, los asociados a los jefes de los hogares. En vista de que el potencial de logro no se puede observar directamente debido a su naturaleza multivariada, se configura entonces como una variable latente que proviene de la combinación multidimensional de un conjunto de factores, entre los que se destacan los efectos de vecindad y del background familiar que, a su vez, son también constructos que definen una relación no mensurable con el potencial de calidad de vida de los hogares.

Tal y como lo afirman los postulados básicos de la nueva economía del ciclo vital, Cunha y Heckman (2007), las diferencias iniciales en las estructuras cognitivas de los educandos en sus entornos inmediatos repercuten en la edad adulta, condicionando así las trayectorias y los logros efectivos en los ámbitos educativos y laborales. Variables como el promedio educativo del hogar, la educación de los jefes del hogar, el grupo étnico, la proporción de capital humano de alta calidad que configura el entorno sociodemográfico del hogar, entre otras, juegan un papel fundamental a la hora de explicar el potencial de logros educativos de cada uno de los miembros de los hogares.

Aquellos niños que se forjan en ambientes hostiles con cabezas de familia femenina, localizados en entornos microlocales con prevalencia de minorías étnicas de origen rural, con medios económicos precarios y entornos familiares y de vecindad desfavorables en los niveles educativos, cuentan con desventajas acumuladas. Estas desventajas se replican a lo largo de su ciclo vital, frente a otros sujetos que se forman en ambientes propicios con hogares completos y con vínculos laborales, ubicados en entornos socioeconómicos de elevada educación promedio de los familiares y vecinos (Vivas, 2007). En este último caso la probabilidad de impulsar trayectorias virtuosas de acumulación de capital humano y buenas oportunidades de inserción futura a los mercados laborales es alta.

Tal y como lo postula las nuevas teorías del ciclo vital, el hogar y el entorno microlocal inmediato (localidad, barrio, segmento, manzana, cuadra) juegan, desde esta perspectiva un papel preponderante y, de esta manera, los factores de riesgo de los dos grupos son disímiles y se replican en los diferentes estadios de desarrollo de los individuos y, en particular, en las conquistas o logros educativos finalmente alcanzados.

¿Cómo inciden el hogar y el entorno local en el potencial de logro educativo? Los enfoques que han seguido la larga tradición inaugurada por Becker se apoyan corrientemente en la existencia de una función de producción del hogar en la que la utilidad

se deriva de las “*commodities*” producidas en su interior mediante la combinación de bienes y servicios adquiridos en el mercado con servicios y atributos no mercantiles (tiempo personal de los miembros del hogar dedicado al cuidado y educación de los niños).

Desde esta perspectiva, a partir de la definición de una función de producción que incluye la cantidad de tiempo disponible en el cuidado y formación de los niños, la renta disponible y el background familiar, se supone que los hogares escogen la calidad educativa potencial que luego se reflejará en los logros escolares, en los máximos niveles educativos alcanzados por los individuos y, en efecto, en los niveles salariales logrados en su vida laboral.

De manera general, este potencial se puede expresar como:  $PCE_j = f((1-t), Y, \mathbf{F}, \mathbf{N}; u)$ ; donde  $t$  es el tiempo disponible del hogar para el trabajo y el ocio,  $Y$  la renta disponible,  $\mathbf{F}$  un vector de atributos asociados a los antecedentes familiares (*family background*), entre los que se incluye el nivel educativo de los padres, el nivel educativo prevaleciente en el entorno familiar y otros atributos del hogar;  $\mathbf{N}$  es un vector de características asociadas al entorno inmediato en el que viven los individuos (*neighborhood effects*).

Recientemente, siguiendo algunos de los trazos iniciales señalados por el trabajo pionero de Ben-Porath (1967), los contrastes empíricos realizados por Cunha y Heckman (2007), muestran que las brechas en la dotación de habilidades observadas en la edad adulta emergen antes de la iniciación del ciclo escolar. La disponibilidad de ingresos familiares y los antecedentes socioeconómicos del hogar operan como factores decisivos en la determinación de las trayectorias de vida, de tal modo que las “*habilidades engendran habilidades*” a través de un proceso de múltiples estados en el que los atributos del hogar y las inversiones iniciales tienen implicaciones en las inversiones posteriores a través de un multiplicador de habilidades.

Ahora bien, tanto el modelo tradicional de capital humano como los nuevos enfoques del ciclo vital enfrentan dificultades de contrastación empírica debido al carácter no observable de las variables clave. Modelos como el de producción del hogar que incorporan de manera exitosa la variable tiempo como factor de producción y que teóricamente permiten abordar su valor económico, enfrentan la dificultad de aproximarse empíricamente a las “*commodities*” debido a que no son observables directamente, tal y como suele suceder con la utilidad o el bienestar. Algo similar ocurre con los modelos desarrollados recientemente por Cunha y Heckman (2007) que en muchos de los casos deben recurrir a ejercicios de micro-simulación dinámica. Lo mismo sucede con los efectos de vecindad que –tal y como ya lo advertía Durlauf (2004)–, al no poderse medir directamente genera escepticismo sobre la verdadera incidencia que tienen sobre la persistencia de la pobreza, su naturaleza y la magnitud de tales efectos.

El estudio aquí propuesto presenta diferencias con otras investigaciones que abordan estos mismos problemas. El aporte reside en operar de manera explícita con variables latentes que se modelan mediante los *métodos de ecuaciones estructurales* de tipo confirma-

torio. Esta estrategia metodológica permite contrastar los nexos entre los constructos<sup>2</sup> efectos de vecindad, background familiar y el potencial de logro educativo, especificando simultáneamente los modelos de medida adecuados (que utilizarán un conjunto de indicadores relevantes) y un modelo de regresión estructural (que contiene, de manera subyacente, un modelo factorial confirmatorio) mediante el cual es posible corroborar las relaciones causales entre las variables latentes. En tal sentido, los ejercicios empíricos propuestos intentarán contrastar los postulados teóricos que se desprenden de los diferentes modelos aportados por la teoría tradicional y las nuevas teorías del ciclo vital.

Otro aporte de esta investigación radica en la utilización de técnicas bayesianas, las cuales son propicias para superar la dificultad que se deriva cuando se viola el supuesto de normalidad o cuando se utilizan variables e indicadores cualitativos que invalidan la utilización directa de correlaciones de Pearson. Comúnmente, los modelos de ecuaciones estructurales se estiman mediante procesos iterativos de máxima verosimilitud que exigen el cumplimiento del supuesto de normalidad para poder probar la congruencia o discrepancia entre la matriz de varianzas y covarianzas poblacional y la definida por el modelo teórico.

No obstante, cuando el modelo considera variables categoricas y se viola el supuesto de normalidad, las matrices de correlaciones relevantes son policóricas, poliseriales o tetracóricas<sup>3</sup> y exigen la presencia de muestras grandes y supuesto de distribución asintóticamente libres. La alternativa de estimación propuesta aquí utiliza técnicas bayesianas y contrasta los resultados (de acuerdo con el tamaño de la muestra utilizada) con los obtenidos por el método de máxima verosimilitud.

La estructura del artículo es la siguiente: el apartado 2 presenta de manera sucinta la literatura relacionada, el apartado 3 describe los datos utilizados y las características de la muestra. El 4 formaliza el modelo de contrastación empírica, definiendo los modelos de medida y el modelo estructural finalmente adoptado después de varias contrastaciones. El apartado 5 analiza el conjunto de resultados obtenidos y, finalmente, en el 6 aparece la síntesis de los principales hallazgos del estudio.

---

2 Es común encontrar en las ciencias sociales y de la conducta variables que se pueden medir y modelar directamente y de manera explícita (edad, sexo, salarios, etc.); pero también se identifican variables que no se pueden medir ni observar directamente (actitudes, percepciones, inteligencia, efectos de vecindad, background familiar, etc.) y que suelen denominarse como constructos, variables latentes, factoriales, entre otros nombres. Algunas de estas variables -que corresponden a conceptos teóricos- únicamente son cuantificables a partir de algunas variables observables que operan como indicadores y sus relaciones se abordan a través de modelos factoriales y de regresión estructural, tal y como es el caso del presente estudio.

3 La utilización de variables en escalas intervalares, dicotómicas o a nivel de ítem, implica la utilización de correlaciones biserials (al menos una de ellas es dicotómica o en categorías) o en el caso más amplio poliserials. Cuando se tiene dos variables dicotómicas (por ejemplo, sexo y respuesta) se utilizan correlaciones phi que se desprenden del análisis de contingencia. Los coeficientes de correlación tetracóricas aparecen cuando las variables dicotomizadas provienen de transformaciones artificiales de otras variables (por ejemplo, una variable originalmente continua que luego se dicotomiza o expresa en varios niveles). En todos estos casos la utilización directa de la matriz de correlaciones de Pearson no es conveniente.

## 2. Literatura relacionada

Las investigaciones sobre las trayectorias de acumulación de capital humano muestran que: la educación de los padres, los tamaños medios de los hogares, la calidad de los entornos locales (efectos de vecindad), la riqueza inicial, la pertenencia a redes sociales, étnicas, religiosas y la fortaleza de las instituciones, entre otros factores; juegan de manera conjunta con las condiciones iniciales y tienen repercusiones de mediano y largo plazo en estas trayectorias, a través de las pérdidas y ganancias que generan en el proceso de interacción social.

La influencia positiva o negativa que puedan ejercer las vecindades o los lugares en los que habitan los individuos, se conjuga con los efectos de los compañeros de clase en las escuelas y el de las familias para determinar los logros y realizaciones futuros de los hijos. Por tal razón, si la trayectoria dinámica que describe el proceso de acumulación lleva a mayores grados de segregación entre ricos y pobres, evidentemente la distribución de talentos, habilidades y capacidades productivas reflejará este comportamiento en los espacios efectivos de realizaciones de las diferentes dinastías, en las condiciones objetivas y en los sentimientos de privación relativa entre los individuos de una misma generación, tal y como ya lo habían planteado Merton y Kitt (1950), así como Sen (1982, 1992).

De este modo, los logros y realizaciones individuales provienen del proceso de interacción social y de la posibilidad de los hogares para generar capacidades.

Desde el año 1967 Ben-Porath había formulado el carácter especial de las funciones de producción de capital humano –al incorporar el proceso de acumulación y formación de capacidades en el interior de los hogares– y había resaltado el papel que juegan las variables sociodemográficas en las trayectorias de acumulación bajo diferentes escenarios (perfectamente competitivo, restricciones de liquidez). Mostraba que el resultado repercute finalmente en la determinación de los perfiles salariales a lo largo del ciclo vital de los individuos.

Aunque los trabajos de Schultz (1961) y Becker (1964) ya habían resaltado que la vinculación de los individuos al sistema educativo corresponde a una decisión de inversión productiva en función de mayores ingresos futuros, los desarrollos teóricos posteriores trascendieron estos planteamientos iniciales y algunos de ellos lograron incorporar las formulaciones de Ben-Porath por mucho tiempo desconocidas.

Con el trabajo de Becker y Tomes (1986), la literatura de las funciones de producción de capital humano se abre a una nueva perspectiva de análisis que contempla al *background* familiar como un insumo vital dentro del proceso de acumulación de capacidades de los individuos.

A través de un modelo teórico de dinastías con horizonte temporal finito, donde los padres transmiten sus patrones de ingresos, capacidades y consumo a sus descendientes, postulan factores como las inversiones que realizan los padres en sus hijos, el acervo de habilidades como elementos centrales que determinan tanto la calidad como la cantidad de las capacidades acumuladas a través de la educación.

Lazear (1980) involucra la influencia del *background* familiar dentro del proceso de elección del nivel educativo óptimo de un individuo. Para ello plantea un modelo de maximización de la riqueza, donde postula que el nivel de ingreso del grupo familiar determina no exclusivamente el logro del agente dentro del sistema escolar, sino que ejerce una gran influencia sobre sus ingresos salariales y su posicionamiento en el mercado laboral. Este autor corrobora empíricamente sus planteamientos haciendo uso de los datos provenientes del National Longitudinal Survey (1975).

Ermisch y Francesconi (2001) presentan un modelo donde los padres definen un patrón de inversión en la formación educativa de sus hijos, de acuerdo con las restricciones de liquidez presentes en la economía y a su dotación de capital humano, estado socioeconómico y estructura familiar. Partiendo de la información estadística proporcionada por British Household Panel Study (1991-1997), los autores encuentran que los antecedentes familiares asociados al ingreso y la dotación de educación de los padres juegan un papel potencial en el logro educativo de sus hijos, lo cual se traduce también en los diferenciales salariales que los mismos obtienen dentro del mercado laboral.

Siguiendo la tradición iniciada por Ben-Porath (1967), Cunha y Heckman (2007) proponen una tecnología de formación de capacidades (*skills*) en el ciclo vital. Dicha propuesta intenta formalizar recientes hallazgos en la literatura empírica sobre el desarrollo de los niños. La producción de capacidades es, entonces, un proceso con varias etapas, donde las producciones de períodos pasados tienen un efecto directo (*self-productivity*) e indirecto (*dynamic complementarity*) sobre la formación de habilidades en el futuro.

Una característica importante del planteamiento de estos autores es la compatibilidad de la tecnología propuesta con la existencia de múltiples capacidades (cognitivas y no-cognitivas). En este sentido, la formación de una capacidad cognitiva se ve potenciada por la posesión de una o varias no-cognitivas (*cross-complementarity*). Igualmente, la existencia de múltiples habilidades da lugar a períodos sensibles y críticos en la formación de las mismas.

De la propuesta de Cunha y Heckman (2007) se siguen dos conclusiones importantes: en primer lugar, las inversiones tempranas tienen un rendimiento mayor que las inversiones tardías, esto se debe a la auto-productividad y a la complementariedad dinámica de las inversiones –en un caso extremo este planteamiento lleva a la conclusión de que remediar bajas inversiones tempranas es excesivamente costoso (sino imposible). En segundo lugar, el *background* familiar afecta la formación de capacidades, ya que determina la inversión que podrían realizar los padres en sus hijos, además de acentuar el efecto de las restricciones de liquidez.

Pfeiffer y Reuß (2008), con base en el trabajo de Heckman y Cunha (2007), modelan una tecnología de formación de capacidades, adicionando un término de depreciación que depende de la edad. Este modelo, además, intenta capturar razones tanto sociales como biológicas para la heterogeneidad en la formación de capacidades, para esto se introducen multiplicadores que intentan capturar el talento de los individuos



a la hora de adquirir cierto tipo de habilidades. Estos autores, igualmente, realizan una estimación de su modelo calibrando los resultados con datos para Alemania. Como principales conclusiones encuentran que las diferencias en las habilidades innatas tienen un efecto mayor sobre la desigualdad que aquellas provenientes del entorno. Adicionalmente, muestran que una sociedad que pretenda maximizar el acervo de capital humano debe invertir en aquellas personas más hábiles y con un entorno mucho más favorable; por el contrario, si el objetivo es maximizar el retorno relativo de las inversiones en educación, los recursos deberían estar dirigidos hacia aquellas personas con mayores desventajas.

El modelo planteado en este artículo muestra la consistencia empírica de las relaciones teóricas planteadas en la mayoría de los trabajos señalados previamente y corrobora que, efectivamente, existe un estrecho nexo entre la calidad de los entornos inmediatos de los individuos y los logros educativos potenciales que pudiesen alcanzar. La proporción de capital humano de alta calidad o su densidad en las vecindades, así como la segregación por color de piel y el estatus de pobreza condensan la calidad de los entornos microlocales (efectos de vecindad) y afectan de manera directa la formación de capacidades y los logros potenciales de los individuos, a la vez que presentan una estrecha correlación con el *background* familiar debido a los efectos del *sorting* socioeconómico.

### 3. Datos, variables y métodos

#### 3.1. Datos y variables

El ejercicio empírico aquí propuesto se realiza para Colombia con información proveniente de las Encuestas de Calidad de Vida. En particular, la información estadística con la cual se trabaja en este artículo proviene de la Encuesta de Calidad de Vida de 2003 del Departamento Administrativo Nacional de Estadística (DANE) y utiliza la información correspondiente a los individuos que en el momento de la Encuesta se encontraban estudiando y que, además, tenían edades entre los 5 y los 30 años.

Es preciso anotar que el DANE realizó una nueva versión de la ECV para el año 2008, pero en el momento en el que se escribió este artículo no se contaba con la información completa de los microdatos. Una de las principales limitaciones de este instrumento es que, aunque se tiene información acerca de los hogares en los cuales habitan los individuos, la encuesta no indaga sobre los logros educativos de los encuestados.

No obstante, tal y como se planteó en la introducción y se formalizará en el apartado 4, en este estudio el potencial de logro educativo, los efectos de vecindad y el *background* familiar son variables latentes o constructos teóricos que no se pueden observar y medir directamente. Su operacionalización se hace a través de indicadores o variables observables que se configuran como medidas falibles de los constructos



que están afectados por errores de medición y que provienen de la imperfección de los datos y de los instrumentos de medida.

El conjunto de variables, parámetros y errores se describen a continuación (la notación *Lisrel*<sup>4</sup> aparecen entre paréntesis):

N: constructo exógeno o variable latente que sintetiza los efectos de vecindad ( $\xi_1$ ).

F: constructo exógeno o variable latente que sintetiza el background familiar ( $\xi_2$ ).

PCE: constructo endógeno que condensa el potencial de logro educativo ( $\eta_1$ ).

PROSPCO1: coeficiente de densidad de capital humano de alta y baja calificación ( $X_1$ ).

ETNIA: grupo étnico que toma valor de 1 si no es afrodescendiente y 0 si lo es ( $X_2$ ).

ESTATUS: estatus socioeconómico que toma valor de 1 si no es pobre y 0 si lo es ( $X_3$ ).

GENGEF: género del jefe del hogar que toma valor de 1 si es hombre y 0 mujer ( $X_4$ ).

TAMHOG: tamaño del hogar ( $X_5$ ).

VIVPADRE: procedencia de los padres 1 urbano y 0 rural ( $X_6$ ).

CLIMEDU: clima educativo del hogar medidos como el promedio educativo de la población económicamente activa ( $Y_1$ ).

EDUJEF: educación del jefe del hogar ( $Y_2$ ).

ESTABLEC: tipo de establecimiento educativo 1 privado y 0 oficial ( $Y_3$ ).

dni: errores de medida asociados a los efectos de vecindad ( $\delta_{Ni}$ ).

dfi: errores de medida asociados al background familiar ( $\delta_{Fi}$ ).

dpi: errores de medida asociados al potencial de logro educativo ( $\delta_{Pi}$ ).

Las estimaciones se realizan con una muestra de 23629 individuos a la que se aplicó técnicas de análisis estructural.

### 3.2. Métodos

Tal y como se anotó en la introducción, una de las innovaciones de este artículo reside en la utilización de modelos estructurales<sup>5</sup> que intentan dilucidar cuáles son las variables de mayor relevancia en la explicación del potencial de logro educativo de los hogares, y que viene a ser un constructo que no se puede observar ni medir directamente.

Los métodos de ecuaciones estructurales (SEM) han venido ganando importancia en el campo de las ciencias económicas en los últimos años, debido a su capacidad para “confirmar” explicaciones causales sobre variables latentes y para evaluar la con-

4 Linear Structural Relation planteado desde la década de los 70's por Karl Jöreskog y Sörbom para el tratamiento de modelos de ecuaciones estructurales (Structural Equation Modeling, SEM).

5 Este método integra los enfoques tradicionales de la econometría y la psicometría. La primera proporciona las bases de la inferencia estructural mediante la introducción de sistemas de ecuaciones simultáneas y la psicometría aporta los métodos de aproximación a la medición de constructos o variables latentes que no son observables directamente como la inteligencia, por ejemplo.

sistencia de las hipótesis planteadas<sup>6</sup> a la luz de la teoría adoptada. Las hipótesis o nexos causales que se derivan de los planteamientos teóricos relacionan las variables latentes o constructos mediante modelos de medida y un modelo estructural.

El primero –el modelo de medida– formula el conjunto de relaciones entre las variables observables y las variables latentes comprometidas en el análisis. El segundo –el modelo estructural– generaliza el modelo convencional de regresión, incluyendo las relaciones causales hipotetizadas entre las variables latentes.

Los procedimientos estándar de modelización inician con la fase de especificación en la que se formulan de manera conjunta los submodelos de medida y el estructural. Luego se aborda el problema de identificación en el que se evalúa la correspondencia entre la cantidad de información contenida en la matriz de varianzas y covarianzas muestral y el número de parámetros a estimar. De acuerdo con este análisis, el modelo puede ser identificado, subidentificado o sobreidentificado. La fase de estimación de los parámetros se lleva a cabo bajo la restricción del modelo teórico y la matriz de datos. Entre los métodos más usuales del proceso de estimación resalta el de Máxima Verosimilitud, cuando el supuesto de normalidad no es cuestionable, o mediante métodos de distribución asintóticamente libre o Bayesianos, cuando no se garantiza la condición de normalidad o se introducen variables categóricas.

La nomenclatura utilizada en este artículo sigue la tradición planteada en los modelos Lisrel (Lisrel-SEM) que se fundamenta en la definición de un conjunto de vectores y matrices, denotadas por  $\mathbf{X}$ ,  $\mathbf{Y}$ , correspondientes a las variables exógenas observadas y endógenas observadas, (Gómez, 1995).

En principio se definen  $\Lambda_x$ ,  $\Lambda_y$ ,  $\mathbf{B}$ ,  $\mathbf{\Gamma}$ : matrices de coeficientes y  $\Phi$ ,  $\Psi$ ,  $\Theta$ : las matrices de covarianza. Las variables latentes endógenas se representan con  $\eta$  y las exógenas con  $\xi$ , las variables observadas endógenas con  $\mathbf{Y}$ ; las exógenas con el vector  $\mathbf{X}$ . Los errores de medida de las variables observadas endógenas se denotan con  $\varepsilon$  y los errores de medida de las variables observadas exógenas con  $\delta$ . Los términos de perturbación se representan mediante  $\zeta$ , los cuales incluyen los efectos de las variables omitidas, los errores de medida y la aleatoriedad del proceso especificado.  $\psi$  denota la variación en el término de perturbación. La covariación entre los términos de perturbación  $i$ -ésimo y  $j$ -ésimo se denota por  $\psi_{ij}$ . Los coeficientes de regresión se representan mediante  $\lambda$  (relaciones entre las variables latentes con los indicadores). Los coeficientes de regresión se pueden representar mediante  $\gamma$ ,  $\beta$ ,  $\phi$  que captan las relaciones entre las variables latentes, así como entre las variables observadas.

Esta notación permite representar cualquier modelo de estructuras de covarianzas y apreciar de manera gráfica y matricial el sistema completo de interrelaciones y cada uno de los submodelos que lo compone.

De acuerdo con la especificación restringida que se deriva del modelo teórico y la naturaleza de las variables consideradas, la literatura estadística identifica diversos tipos de modelos de ecuaciones estructurales que van desde los modelos de trayectoria,

6 Ver Gómez en Arnau, Jaume (editor) (1995) para una excelente referencia al respecto.

los factoriales confirmatorios, los modelos de regresión estructural, los *mimic* (multiple indicators and multiple causes of a single latent variable), entre otros.

#### 4. El modelo

El artefacto analítico planteado en este estudio corresponde a un modelo de regresión estructural que establece asociación entre las variables latentes (F, N y PCE) y define los modelos de medida entre estas variables no observables y un conjunto de indicadores o variables observadas.

El submodelo de medida del constructo *background* familiar, F, utiliza el género de los jefes del hogar, el tamaño del hogar y la procedencia rural o urbana como indicadores o variables observables.

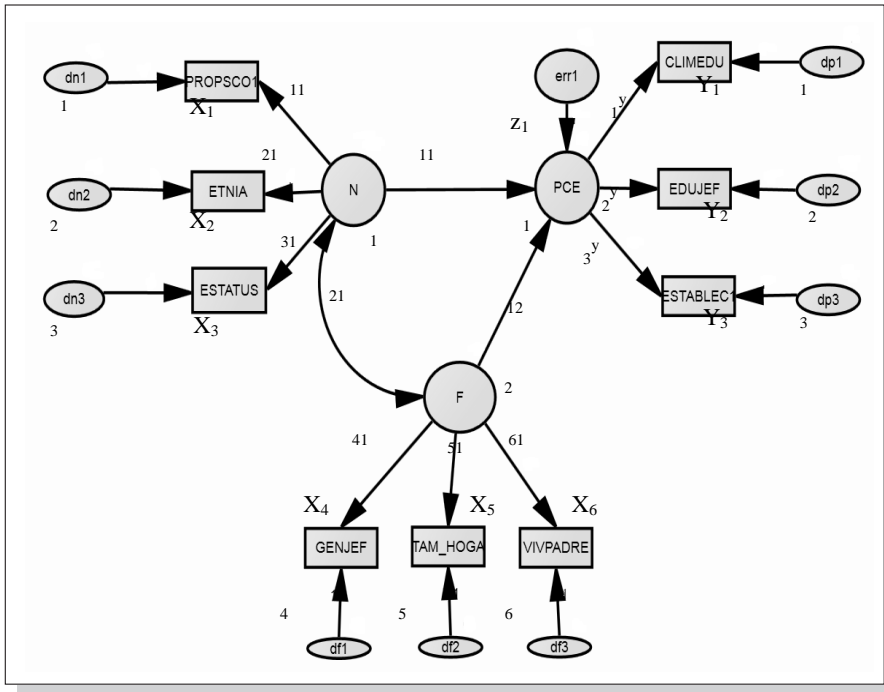
El submodelo de medida del constructo de vecindades, N, utiliza como indicadores la relación de densidad de capital humano de alta y baja calificación, el grupo étnico y el estatus de pobreza (de acuerdo con la línea de pobreza del DANE).

El submodelo de medida del potencial de logro educativo incorpora como indicadores el clima educativo del hogar, el nivel educativo de los jefes del hogar y el tipo de establecimiento (oficial o privado).

El modelo estructural planteado establece la relación causal de F y N hacia el potencial de logro educativo, PCE, así como la covariación entre N y F.

El diagrama No. 1 presenta el sistema general de interrelaciones y las equivalencias con la notación Lisrel utilizada frecuentemente en las especificaciones SEM.

Diagrama No. 1: Representación<sup>7</sup> del sistema general de interrelaciones



Variables latentes endógenas  $\eta$ , latentes exógenas  $\xi$

Variables observadas endógenas  $Y$ , observadas exógenas  $X$ .

Errores de medida de las variables observadas endógenas  $\xi$ , y de las variables observadas exógenas  $\delta$ .

Término aleatorio de error  $\zeta$  (incluye efectos de las variables omitidas, los errores de medida y la aleatoriedad inmanente).

La variación en el término de perturbación se simboliza por  $\Psi$  y la covariación entre los términos de perturbación  $i$ -ésimo y  $j$ -ésimo se denota por  $\Psi_{ij}$ .

Coefficiente de regresión entre las variables latentes y los indicadores:  $\lambda$ .

Coefficientes de regresión  $\gamma$ ,  $\beta$ ,  $\phi$  que relacionan las variables latentes entre sí, y las variables observadas entre sí.

Fuente: Elaboración propia.

7 Es preciso tener en cuenta que las relaciones causales aparecen representadas con flechas unidireccionales entre las variables exógenas ( $\xi$ ) y las variables endógenas ( $\eta$ ), mientras que las correlaciones se representan a través de líneas bidireccionales o trazos curvilíneos en dirección recíproca. Los rectángulos contienen las variables endógenas y exógenas observables, mientras que los trazos circulares representan las variables latentes (endógenas y exógenas). Los óvalos contienen los errores de medida y las flechas representan el sentido de los nexos, condensando en valores paramétricos los coeficientes de regresión correspondientes a la estructura de estas relaciones. Esta representación recibe el nombre de path analysis y es frecuente en la modelación de ecuaciones estructurales. Software especializado como el Lisrel y el Spss/Amos facilitan una interfaz de modelamiento orientada a objetos que sigue esta misma nomenclatura.

La estructura matricial del modelo completo representa los submodelos de medida que especifican las ecuaciones que relacionan las variables latentes con los indicadores o variables observadas (exógenas y endógenas X, Y, respectivamente). Este submodelo introduce un error de medición aleatorio ( $\delta$  y  $\xi$ ) en los modelos causales entre las variables observables, las cuales operan como indicadores de las variables latentes ( $\xi$  = variable latente exógena y  $\eta$  = variable latente endógena).

Tal y como se anotó con anterioridad, las variables latentes corresponden a constructos hipotéticos no observables directamente y que deben cumplir con la condición de unidimensionalidad. La representación matricial del submodelo de medida se resume en la siguiente expresión:

$$\begin{aligned} \text{[Expr. 1]} \quad X &= \Lambda_x \xi + \delta \\ Y &= \Lambda_y \eta + \varepsilon \end{aligned}$$

X: vector (6x1), en donde 6 es el número de variables observables exógenas.

Y: vector (3x1), en donde 3 es el número de variables observables endógenas.

$\Lambda_x$ : matriz de coeficientes de los indicadores de las variables exógenas (Lambdas).

$\Lambda_y$ : matriz de coeficientes de los indicadores de las variables endógenas (Lambdas).

$\xi$ : vector (2x1) de constructos exógenos (F y N).

$\eta$ : constructo endógeno (PCE).

$\delta$ : vector de errores de medición de las variables observadas exógenas.

$\varepsilon$ : vector de errores de medición de las variables observadas endógenas.

Valores altos de los coeficientes lambda sugieren que los indicadores utilizados representan bien a los constructos y, en consecuencia, se corroboraría la pertinencia de los indicadores finalmente seleccionados en la medición de estas variables hipotéticas. Los  $\lambda^2$  aproximan la cantidad de varianza explicada por cada una de las ecuaciones (fiabilidad del indicador).

Es importante observar que debido a que el submodelo de medida obtiene la relación entre las variables exógenas latentes y las observables, los vínculos son direccionales desde cada constructo a sus indicadores correspondientes.

Ahora bien, el submodelo estructural especifica las ecuaciones causales lineales entre las variables latentes del modelo. La estructura matricial generalizada de este submodelo toma la siguiente forma:

$$\text{[Expr. 2]} \quad \eta = B\eta + \Gamma\xi + u$$

En donde:

$\eta$ : vector (mx1) de variables latentes endógenas.

$B$ : matriz de coeficientes de las relaciones entre variables endógenas. Su rango es (mxm) y contiene ceros en la diagonal principal.

$\Gamma$ : matriz de coeficientes de las relaciones entre las variable endógenas y las exógenas. Rango (mxn).

$\xi$ : vector (nx1) de constructos exógenos.

U: perturbaciones aleatorias de rango (mx1). Indica que las variables latentes endógenas no están perfectamente determinadas por las ecuaciones estructurales<sup>8</sup>.

De acuerdo con la estructura aquí propuesta, la variable latente endógena (potencial de logro educativo) toma una forma simplificada que viene expresada por:

[Expr. 3]

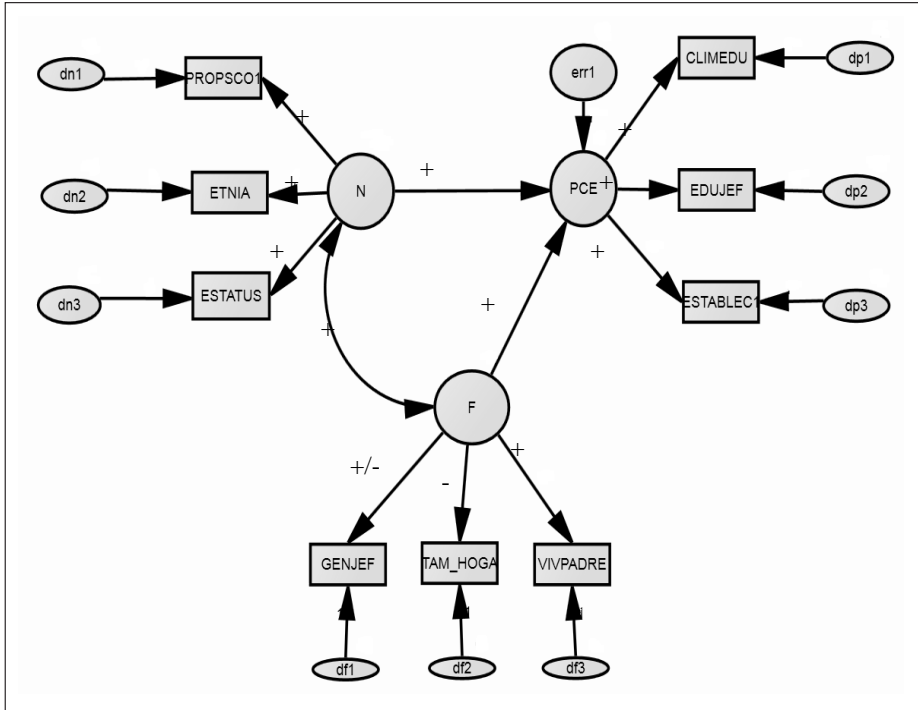
$$\eta_1 = \gamma_{11}\xi_1 + \gamma_{12}\xi_2 + z_1$$

El diagrama No. 2 presenta el resumen de las hipótesis de comportamiento, teniendo en cuenta la definición de las variables en la base de datos utilizada.

---

8 Si las variables están representadas respecto a la media, el modelo no presenta interceptos. El modelo supone errores homocedásticos y con correlaciones nulas entre sí y con las demás variables. El sistema de modelación admite estructuras recursivas (relaciones recíprocas), aunque en estos casos es probable que se reporten problemas de identificación debido a que los términos de perturbación no serán independientes. Véase Gómez, 1995:264 en Arnau, 1995.

Diagrama No. 2: Hipótesis de comportamiento



Fuente: Elaboración propia.

Por ejemplo, la alta proporción de capital humano en los ámbitos microlocales se espera que afecte positivamente al constructo de calidad del entorno local (vecindades, N). El hecho de que los individuos estén rodeados de vecinos con niveles educativos relativamente altos eleva la calidad de los entornos microlocales.

Asimismo, se espera que la no pertenencia al grupo étnico de afrodescendientes afecte positivamente a este constructo, de tal manera que cuando este indicador toma el valor de cero (pertenencia al grupo étnico de afrodescendientes), el efecto esperado sobre N se asocia a una menor calidad de los entornos locales, ya que el constructo contiene información subyacente de la etnicidad.

Es preciso recalcar aquí que la variable latente establece una relación con los indicadores de medida que debe interpretarse como la presencia de un factor común subyacente. En este caso, el constructo de vecindad opera como un factor común que subyace en el conjunto de indicadores.

De manera análoga se debe interpretar para el caso del *background* familiar y el potencial de logro educativo: el tamaño del hogar, el origen urbano o rural de los padres y su género contienen información subyacente del constructo *background* familiar.



En cuanto a los valores  $\lambda_{ji}$  –que miden la magnitud de las relaciones entre las variables latentes y sus indicadores– se espera que sean estadísticamente significativos y que en su forma estandarizada corroboren las relaciones teóricas previamente definidas.

Tal y como se anotó en los aparatados anteriores, el desarrollo de habilidades cognitivas, lingüísticas y de capacidades para relacionarse satisfactoriamente en ambientes competitivos, se forjan desde las edades tempranas en el hogar y en el vecindario inmediato, de tal manera que más adelante serán decisivas para la inserción al sistema de educación formal y al mercado laboral. Las hipótesis de comportamiento formuladas en el modelo planteado marchan en esta dirección e intentan, a través de los instrumentos planteados para el manejo de variables latentes, corroborar y medir la magnitud de los nexos entre los efectos de vecindad, los antecedentes familiares y el potencial de logro educativo.

## 5. Estimaciones, hallazgos y discusión

A continuación se exponen los métodos utilizados en las estimaciones del modelo y se discuten los principales resultados.

### 5.1. Método de estimación

Los métodos de estimación de los modelos estructurales utilizados frecuentemente recurren a las técnicas de máxima verosimilitud, siempre y cuando no exista duda sobre la condición de normalidad. En estos métodos no bayesianos los verdaderos valores de los parámetros del modelo se consideran estocásticamente fijos pero desconocidos (Byrne, 2006 y Arbuckle, 2007).

La estimación bayesiana, por el contrario, considera que los verdaderos parámetros del modelo son desconocidos y aleatorios. De esta manera, los parámetros se asignan a una distribución conjunta *a priori* y otra a *posteriori* que luego se evalúan en su discrepancia. La distribución conjunta se fundamenta en el teorema de Bayes y devela la combinación de las creencias iniciales acerca de los valores de los parámetros y la evidencia empírica.

Ahora bien, cuando los modelos utilizan variables categóricas es preciso definir de manera exacta la matriz de correlaciones relevante. Cuando las variables tienen escalas ordinales, por ejemplo, las correlaciones convencionales de Pearson no son adecuadas y es necesario recurrir a otro tipo de estructuras de correlación. Adicionalmente, la presencia de este tipo de variables debe suponer, bajo los métodos no bayesianos, que cada una de las variables categóricas se distribuye normalmente, lo cual resulta difícil de sostener en este caso. Por tal razón, entre las alternativas propuestas se identifica la utilización de distribuciones asintóticamente libres (métodos de mínimos cuadrados ponderados<sup>9</sup>) o el tratamiento explícito de las matrices de correlaciones policóricas.

9 Los mínimos cuadrados generalizados, MCG, también operan bajo el supuesto de normalidad, aunque tienen la característica de operar con una matriz de pesos. El método de mínimos cuadrados ponderados (método de

Una alternativa a los métodos anteriores consiste en la utilización de inferencia bayesiana, tal y como se hace en este artículo en el que se comparan los resultados con las salidas convencionales de máxima verosimilitud, ML, bajo la condición de una muestra suficientemente grande.

## 5.2. Hallazgos y discusión

La tabla No. 1 presenta los resultados de las estimaciones bayesianas una vez se logra la convergencia que garantiza la estabilidad de los valores de los parámetros<sup>10</sup>.

La primera columna muestra los valores medios<sup>11</sup> de la distribución posterior de los parámetros de regresión, los interceptos, las varianzas y las covarianzas. Estos resultados pueden interpretarse como las estimaciones puntuales bayesianas basadas en los datos y en la distribución *a priori*. Cada una de las filas muestra los valores asociados a la distribución posterior de cada uno de los parámetros del modelo.

La comparación de estos resultados con los de máxima verosimilitud –ver tabla No. 2– permite observar pequeñas discrepancias en los valores puntuales, lo cual proporciona evidencia favorable a la robustez de los procedimientos utilizados.

Por ejemplo, el parámetro que relaciona el clima educativo del hogar con el potencial de logro educativo presenta un valor centrado de 13.674 en la estimación bayesiana y bajo el método de máxima verosimilitud 13.672 (parámetro no estandarizado) –comparar con los resultados presentados en el diagrama path del anexo A1–. Obsérvese que el valor estandarizado de este mismo parámetro (bajo máxima verosimilitud, ML) es de 0.96, tal y como se puede ver en el A2. La tabla 2 muestra el detalle de los resultados de ML<sup>12</sup>, comparados con el método bayesiano.

Ahora bien, las columnas 2 y 3 de la tabla No. 1 muestran valores pequeños de S.E y S.D que indican un valor medio estimado en la distribución posterior que difiere muy poco del verdadero valor del parámetro. S.E corresponde a un error estándar estimado y S.D se aproxima al valor de la desviación estándar bajo máxima verosimilitud.

---

distribución asintóticamente libre) puede resultar adecuado cuando se viole el supuesto de normalidad de los datos. Es imprescindible si el modelo contiene una o más variables categóricas y por lo tanto se trabaja con matrices policóricas, poliserials y tetracóricas. Un requisito es que la muestra sea considerablemente grande.

10 De acuerdo con el algoritmo MCMC –Simulación Montecarlo de Cadenas Markovianas– se generaron 81 muestras adicionales hasta alcanzar valores de convergencia del estadístico C.S cercanos a 1 (ver columna 4 de la tabla No. 1). Los valores de los parámetros corresponden a los más cercanos posibles a los verdaderos valores.

11 Los cuales corresponden a la estimación final del parámetro. De acuerdo con Arbuckle (2007) y Byrne (2006), cuando las muestras son suficientemente grandes estos valores medios deben coincidir con las estimaciones de máxima verosimilitud.

12 Observar que todos los parámetros resultaron estadísticamente significativos.

Tabla No. 1: Estimaciones Bayesianas del modelo de ecuaciones estructurales

	Mean	S.E.	S.D.	C.S.	Median	50% Lower bound	50% Upper bound	Skewness	Kurtosis	Min	Max
<b>Regression weights</b>											
CLIMEDU<--PCE	13.67	0	0.15	1	13.673	13.57	13.776	0.035	-0.077	13.08	14.23
EDUJEF<--PCE	17.54	0	0.19	1	17.541	17.407	17.677	0.055	-0.175	16.81	18.23
PROPSCO1<--N	37.8	0.09	2.11	1	37.734	36.353	39.159	0.192	-0.006	30.77	45.63
PCE<--N	5.398	0.02	0.44	1	5.38	5.097	5.672	0.334	0.355	3.929	7.423
TAM_HOGA<--F	-	0	0.08	1	-3.758	-3.815	-3.703	-0.09	0.001	-	-
GENJEF<--F	3.759	-	0	0.02	1	-0.145	-0.156	-0.133	0.059	0.102	4.147
PCE<--F	0.145	-	0	0.05	1	0.179	0.143	0.214	-0.084	0.125	0.384
ESTATUS<--N	0.178	0	0.05	1	0.179	0.143	0.214	-0.084	0.125	0.033	-
ESTATUS<--N	8.55	0.02	0.48	1	8.536	8.224	8.856	0.197	0.059	6.98	10.19
<b>Intercepts</b>											
CLIMEDU	8.093	0	0.02	1	8.093	8.077	8.109	-0.008	-0.058	7.993	8.179
EDUJEF	7.785	0	0.03	1	7.785	7.763	7.808	-0.041	-0.129	7.65	7.9
ESTABLEC1	0.311	0	0	1	0.311	0.309	0.313	0.012	-0.136	0.299	0.321
PROPSCO1	1.692	0	0.02	1	1.692	1.681	1.702	0.029	0.03	1.634	1.752
ETNIA	0.917	0	0	1	0.918	0.916	0.919	-0.078	-0.151	0.91	0.924
VIVPADRE	0.785	0	0	1	0.785	0.783	0.786	0.013	0.014	0.773	0.795
TAM_HOGA	5.078	0	0.01	1	5.078	5.069	5.087	0.019	-0.101	5.027	5.129
GENJEF	0.703	0	0	1	0.703	0.701	0.705	0	-0.052	0.692	0.714
ESTATUS	0.473	0	0	1	0.473	0.471	0.476	-0.078	0.031	0.462	0.49
<b>Covariances</b>											
F<-->N	0.007	0	0	1	0.007	0.007	0.007	0.192	0.104	0.006	0.009
<b>Variances</b>											
N	0.001	0	0	1	0.001	0.001	0.001	0.298	0.11	0.001	0.002
F	0.047	0	0	1	0.047	0.046	0.048	0.097	-0.05	0.041	0.055
err1	0.01	0	0	1	0.01	0.009	0.01	-0.114	0.063	0.007	0.012
dp1	0.904	0	0.04	1	0.904	0.878	0.93	0	-0.103	0.746	1.052
dp2	5.867	0	0.08	1	5.866	5.812	5.922	0.024	-0.166	5.564	6.155
dp3	0.152	0	0	1	0.152	0.151	0.153	0.007	-0.008	0.146	0.157
dn1	3.835	0	0.04	1	3.834	3.804	3.864	0.103	0.031	3.644	4.031
dn2	0.074	0	0	1	0.074	0.074	0.075	0.04	0.006	0.072	0.077
df3	0.122	0	0	1	0.122	0.121	0.123	0.001	-0.119	0.116	0.129
df2	3.276	0	0.04	1	3.276	3.253	3.3	0.05	-0.042	3.144	3.411
df1	0.208	0	0	1	0.208	0.206	0.209	0.06	0.058	0.201	0.215
dn3	0.153	0	0	1	0.153	0.152	0.154	0.005	-0.024	0.146	0.161

Fuente: Estimaciones propias con el módulo bayesiano de Amos.

La tabla No. 1 presenta de manera adicional los demás parámetros de la distribución (skewness, kurtosis, min, max).

De acuerdo con las hipótesis inicialmente planteadas, la primera columna de la tabla No. 1 permite apreciar que el sentido de las relaciones postuladas en la base teórica (signos esperados en el valor de los parámetros) se corrobora en el ejercicio empírico con variables latentes.

Los coeficientes de regresión de los submodelos de medida y los que relacionan las variables latentes presentan los signos esperados en las corridas bayesianas y resultaron estadísticamente significativos cuando se utilizó el método de máxima verosimilitud.

La alta densidad de capital humano en los entornos locales, la baja vulnerabilidad

de los hogares a los que pertenecen los individuos (medido a través de la línea de pobreza del DANE) y la baja composición de población afrodescendiente, contienen información subyacente del constructo “calidad de los entornos locales” (efectos de vecindad) que explica las relaciones positivas que presenta el submodelo de medida.

El origen de los padres (urbano o rural), el género del jefe y el tamaño del hogar resultaron ser indicadores adecuados del constructo background familiar. El tamaño del hogar y el género (0= femenino, 1=hombre) arrojaron efectos negativos sobre la variable latente y el origen urbano un efecto positivo, tal y como se esperaba *a priori*.

En relación con el submodelo de medida del constructo “potencial de logro educativo”, los resultados corroboran que la capacidad de producción esperada de las realizaciones educativas de los individuos se asocia fuertemente con el clima educativo del hogar y el nivel educativo de los jefes.

Los perfiles educativos de los integrantes del hogar, en particular los de los jefes del hogar, juegan un papel crucial a la hora de explicar las posibles trayectorias de acumulación de capital humano de los demás miembros de la dinastía. Este potencial (que desde una perspectiva más amplia definen una relación no mensurable directamente con el potencial de calidad de vida de los hogares) presenta un estrecho nexo con los constructos N y F, tal y como se puede corroborar en los valores de los coeficientes de regresión estructural de la tabla No.1 y en los resultados de máxima verosimilitud presentados en la tabla No. 2 y en los anexos.

Los coeficientes de regresión estructural  $\gamma$  (que relacionan las variables latentes entre sí) permiten apreciar que el mayor efecto proviene del constructo “calidad de los entornos locales”.

Las estimaciones bayesianas y por máxima verosimilitud (ML) muestran resultados muy próximos:  $N \rightarrow PCE = 5.398$  en los pesos de regresión por el método bayesiano y 5.368 por ML. El valor estandarizado del mismo coeficiente toma el valor de 0.78, mientras que el valor que relaciona el background familiar con este potencial de logro,  $F \rightarrow PCE$ , alcanza el valor de 0.15 en la versión estandarizada (ver tabla No. 2 y A2).

**Tabla No. 2:** Comparación de resultados entre los métodos Bayesiano y M.L.  
(Resultados no estandarizados)

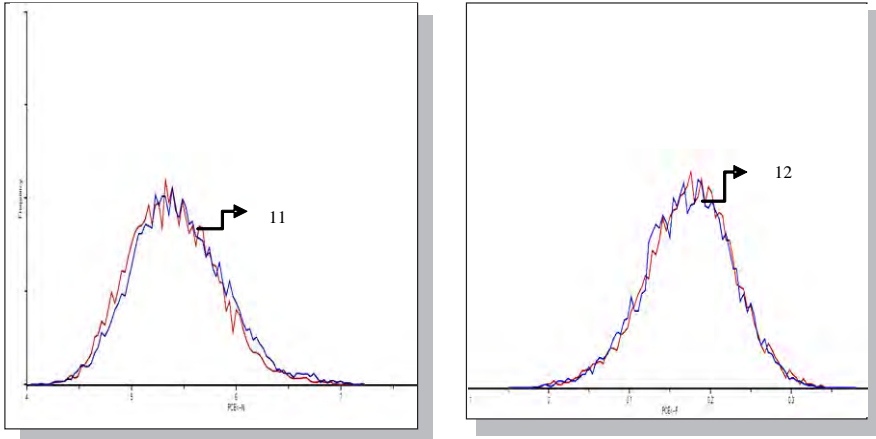
	Bayesian		M.L.			
	Estimate*	Estimate	S.E.	C.R.	P	
Regression weights						
CLIMEDU<--PCE	13.67	13.67	0.15	91.75	***	
EDUJEF<--PCE	17.54	17.54	0.20	90.14	***	
PROPSCO1<--N	37.80	37.60	2.04	18.47	***	
PCE<--N	5.40	5.37	0.42	12.78	***	
TAM_HOGA<--F	-3.76	-3.76	0.08	-46.08	***	
GENJEF<--F	-0.15	-0.15	0.02	-8.98	***	
PCE<--F	0.18	0.18	0.05	3.50	***	
ESTATUS<--N	8.55	8.50	0.46	18.54	***	
Intercepts						
CLIMEDU	8.09	8.09	0.02	350.66	***	
EDUJEF	7.79	7.79	0.03	238.94	***	
ESTABLEC1	0.31	0.31	0.00	103.24	***	
PROPSCO1	1.69	1.69	0.02	108.42	***	
ETNIA	0.92	0.92	0.00	<u>512.38</u>	***	
VIVPADRE	0.79	0.79	0.00	292.61	***	
TAM_HOGA	5.08	5.08	0.01	393.04	***	
GENJEF	0.70	0.70	0.00	236.63	***	
ESTATUS	0.47	0.47	0.00	145.74	***	
Correlations						
F<-->N	0.86	0.86				
Variances						
N	0.00	0.00	0.00	9.35	***	
F	0.05	0.05	0.00	28.97	***	
err1	0.01	0.01	0.00	16.13	***	
dp1	0.90	0.90	0.04	22.53	***	
dp2	5.87	5.86	0.09	69.22	***	
dp3	0.15	0.15	0.00	105.83	***	
dn1	3.84	3.83	0.04	88.92	***	
dn2	0.07	0.07	0.00	<u>108.31</u>	***	
df3	0.12	0.12	0.00	73.09	***	
df2	3.28	3.27	0.04	94.06	***	
df1	0.21	0.21	0.00	108.49	***	
dn3	0.15	0.15	0.00	81.26	***	

\*Estimaciones Bayesianas (valores de convergencia). \*\*Estimaciones de M.L. C.R. = estimate/S.E. \*\*\*Indica significancia estadística al 1% (probabilidad de rechazar Ho).

Fuente: Estimación y elaboración propias.

El gráfico No. 1 que se presenta a continuación muestra la distribución de los dos parámetros mencionados previamente (N→PCE y F→PCE) por el método bayesiano y compara para cada uno de ellos las funciones de densidad *a priori* y *a posteriori* (trazos claro y obscuro).

**Gráfico No. 1:** Comparación de la distribución a priori y a posteriori de  $\gamma_{11}$  y  $\gamma_{12}$

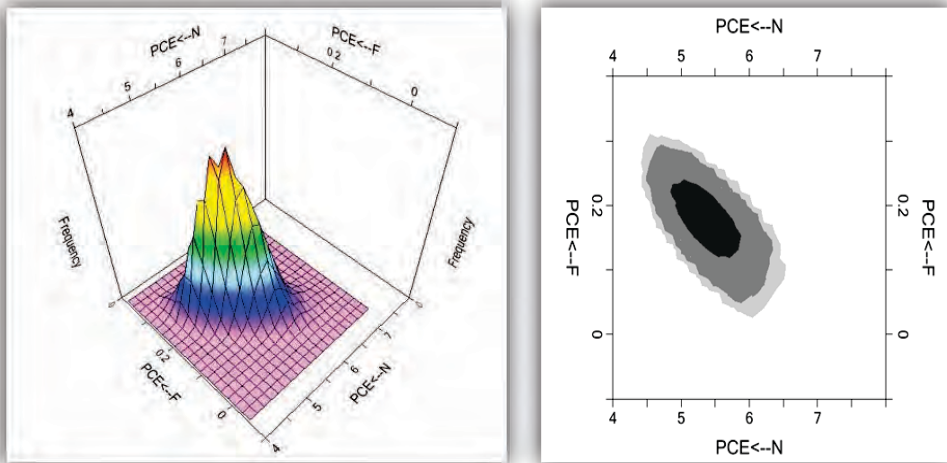


Fuente: Estimación y elaboración propias.

La distribución de la izquierda (que relaciona  $N \rightarrow PCE$ ) está centrada alrededor de 5.4 y la de la derecha (que relaciona  $F \rightarrow PCE$ ) en el entorno de 0.18. Es posible observar que las distribuciones a priori y a posteriori no presentan discrepancias significativas.

Finalmente, el gráfico No. 2 muestra la distribución bivariada de los parámetros y las curvas de contorno asociadas a estas densidades.

**Gráfico No. 2:** Distribución Bivariada de  $\gamma_{11}$  y  $\gamma_{12}$



Fuente: Estimación y elaboración propias

Los resultados del modelo de regresión estructural, en su forma estandarizada (ver A2), enseña que la variable latente endógena (potencial de logro educativo) podría expresarse como:

$$PCE = 0.78*N + 0.15*F + 0.39*z1^{13}$$

Como se puede ver en el gráfico No. 2, efectivamente la calidad de los entornos locales (vecindades, N) y del *background* familiar (F) afectan simultáneamente y de manera específica<sup>14</sup> al potencial de logro educativo, tal y como lo muestra la superficie de la izquierda (a diferencia del gráfico No. 1 el cual representa la distribución marginal posterior de cada parámetro por separado). El gráfico de contorno muestra las regiones de confianza bivariadas (regiones creíbles) al 50%, 90% y 95%.

Es preciso recalcar que los predictores de la variable explican el 84.7% de su varianza, lo que implica que el restante 15.3% corresponde a la varianza del error. El coeficiente 0.39 indica que la variable latente endógena, PCE, no está perfectamente determinada por las ecuaciones estructurales del sistema y existe un margen de error, de tal manera que si  $z1$  aumenta en una desviación estándar, PCE aumenta en 0.39 desviaciones estándar.

## 6. Conclusiones

El artículo propuso una estrategia empírica de estimación del potencial de logro educativo usando métodos de análisis estructural que permitieron trabajar con constructos teóricos o variables latentes que no se pueden medir y observar directamente, como es el caso del *background* familiar y de los efectos de vecindad. Los ejercicios empíricos aportaron evidencia de que las realizaciones educativas potenciales generadas en el interior de los hogares se ligan a los efectos propios del entorno familiar y a la calidad de las vecindades, tal y como lo sugieren los diferentes enfoques teóricos referenciados en el estudio.

Las contrastaciones se realizaron para Colombia con una muestra de 23629 individuos anidados en los hogares de la Encuesta de Calidad de Vida de 2003 a los que se aplicaron técnicas de análisis estructural.

Un aporte de esta investigación consistió en la utilización de métodos bayesianos que intentan superar los inconvenientes de los métodos factoriales convencionales cuando se introducen variables en categorías o dicotómicas. El artículo presenta la estimación de las relaciones estructurales sugeridas por la teoría y las compara con los resultados obtenidos a través de técnicas de máxima verosimilitud.

La investigación aportó evidencia de que la calidad de los entornos locales (efectos de vecindad) ejerce los efectos más grandes sobre el potencial de logro educativo, de tal modo que los individuos localizados en ambientes desfavorables cuentan con des-

13 Recordar que la especificación del modelo conduce a  $\eta_1 = \gamma_{11}\xi_1 + \gamma_{12}\xi_2 + z1$ .

14 Resultado que es evidente al incorporar consideraciones de *sorting* socioeconómico en los modelos: los individuos se localizan en lugares con perfiles sociodemográficos próximos. El coeficiente de correlación estimado entre los constructos F y N fue de 0.86, tal y como se puede ver en la tabla No. 2.



ventajas acumuladas que establecen una brecha importante, frente a otros sujetos que se forman en ambientes propicios, ubicados en entornos socioeconómicos de elevada educación promedio de los familiares y vecinos (altas densidades de capital humano).

Tal y como se anotó desde el principio, la calidad de los entornos microlocales (localidad, barrio, segmento, manzana, cuadra) juegan, desde esta perspectiva, un rol preponderante y, de esta manera, los factores de riesgo de los diferentes grupos sociales son disímiles y se replican en los diferentes estadios de desarrollo de los individuos y, en particular, en las conquistas o logros educativos que finalmente puedan alcanzar.

Las estimaciones fueron consistentes por los dos métodos utilizados y permitieron apreciar que la alta densidad de capital humano en los entornos locales, la baja vulnerabilidad de los hogares a los que pertenecen los individuos, así como el perfil de la composición étnica, presenta relaciones positivas y significativas con el constructo de vecindad. Adicionalmente, se pudo corroborar que la procedencia de los padres (urbano o rural), el género del jefe y el tamaño del hogar resultaron ser indicadores adecuados del constructo *background* familiar.

Los parámetros estimados asociados a los submodelos de medida, así como los que relacionan las variables latentes entre sí, presentaron los signos esperados por el método de Bayes y resultaron estadísticamente significativos a través de M.L. Por las dos vías los resultados permitieron apreciar que la alta densidad de capital humano en los entornos locales, la baja vulnerabilidad de los hogares a los que pertenecen los individuos (medido a través de la línea de pobreza del DANE y la baja composición de población afrodescendiente, efectivamente contienen información subyacente del constructo N (efectos de vecindad) que explica las relaciones positivas que presenta el submodelo de medida.

Así mismo, el origen de los padres (urbano o rural), el género del jefe y el tamaño del hogar resultaron ser indicadores adecuados del constructo *background* familiar. El tamaño del hogar y el género de los jefes arrojaron efectos negativos sobre la variable latente y el origen urbano un efecto positivo.

En relación con el potencial de logro educativo, los resultados corroboran que la capacidad de producción esperada de las realizaciones educativas de los individuos está fuertemente asociada con el clima educativo del hogar y el nivel educativo de los jefes. Este potencial presenta un estrecho nexo con los constructos N y F, tal y como se pudo contrastar a través de los parámetros de la regresión estructural estimados que muestran las magnitudes absolutas y diferenciales de estos vínculos.

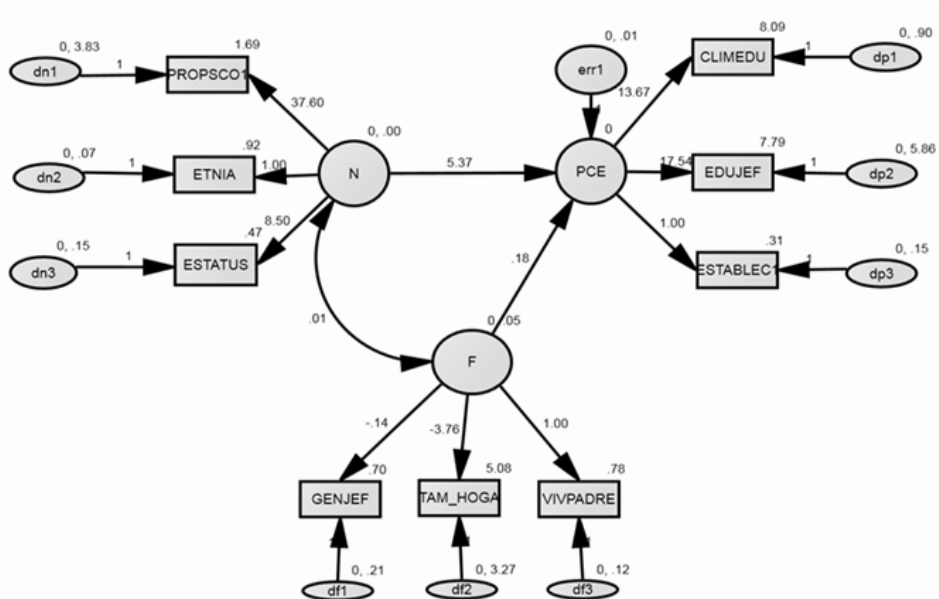
## 7. Referencias bibliográficas

- ARBUCKLE, James (2007). *AmosTM 16 user's guide*. Chicago: SPSS.
- ARNAU, Jaume (Editor) (1995). *Diseños longitudinales aplicados a las ciencias sociales y del comportamiento*, Editorial Limusa, S.A., Noriega Editores, México.
- BECKER, Gary (1964). *Human capital: a theoretical and empirical analysis, with special reference to education*. New York: National Bureau of Economics Research.

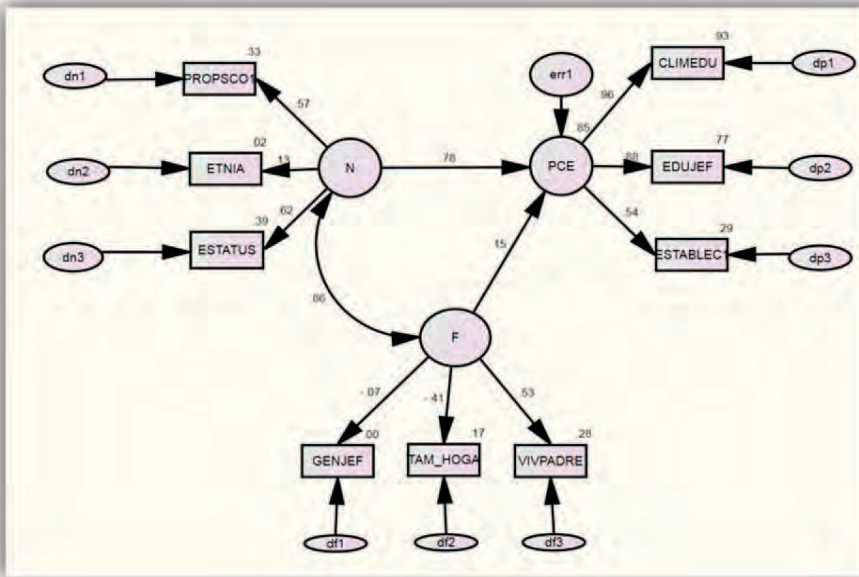
- \_\_\_\_\_ ; TOMES, Nigel (1986). "Human capital and the rise and fall of families", en *Journal of Labor Economics*, Vol. 4, No. 3, pp. 1-39.
- BEN-PORATH, Yoram (1967). "The production of human capital and the life cycle of earnings", en *Journal of Political Economy*, Vol. 75, No. 4, pp. 352-365.
- BYRNE, Barbara (2006). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Second Edition, Taylor & Francis Group.
- BROWN, Byron; SAKS, Daniel (1981) "The microeconomics of schooling", en *Review of Research in Education*, Vol. 9, pp. 217-254.
- CUNHA, Flavio; HECKMAN, James (2007). "The technology of skill formation", en *American Economics Review*, Vol. 97, No. 2, pp. 31-47.
- \_\_\_\_\_ ; \_\_\_\_\_ (2008). "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation", en *Journal of Human Resources*, Vol. 43, No. 4, pp. 738-782.
- DURLAUF, Steven (2004) "Neighborhood Effects" en, J. Vernon Henderson and Jacques-François Thisse (Eds.). *Handbook of Regional and Urban Economics*, vol. 4. Amsterdam, North Holland.
- ERMISCH, Jhon; FRANCESCONI, Marco (2001). "Family matters: impacts of family background on educational attainments", en *Economica*, Vol. 68, No. 270, pp. 137-156.
- GÓMEZ, Benito (1995). "Análisis de los diseños de medidas repetidas mediante modelos de ecuaciones estructurales", en Jume Arnau (Ed.) (1995). *Diseños longitudinales aplicados a las ciencias sociales y del comportamiento*. México: Editorial Limusa, S.A., Noriega Editores.
- HANUSHEK, Eric (1997). "Assessing the effects of school resources on student performance: an update", en *Educational Evaluation and Policy Analysis*, Vol. 19, No. 2, pp. 141-164.
- LAZEAR, Edward (1980): "Family background and optimal schooling decisions", en *Review of Economics and Statistics*, Vol. 62, No. 1, pp. 42-51.
- LUCAS, Robert (1988) "On the mechanics of economic development", en *Journal of Monetary Economics*, Vol. 22, No. 1, pp. 3-42.
- MERTON Robert; KITT, Alice (1950). "Contributions to the theory of reference group behavior", en Robert. K. Merton y Paul Lazarsfeld (Eds.), *Continuities in social research Studies in the scope and method of "The American soldier"*. Glencoe: Free Press.
- PFEIFFER, Friedhelm; REUß, Karsten (2008). "Age-dependent skills formation and returns to education", en *Labour Economics*, Vol. 15, pp. 631-646.
- SCHULTZ, Theodoro (1961). "Investment in Human Capital", en *American Economic Review*, Vol. 51, No. 1, pp. 1-17.
- SEN, Amartya (1992). *Inequality Reexamined*. Cambridge Mass., Harvard University Press.
- \_\_\_\_\_ (1982). *Poverty and Famines: An Essay on Entitlements and Deprivation*. Oxford: Clarendon Press.
- UZAWA, Hirofumi (1965). "Optimum technical change in an aggregative model of economic growth", en *International Economic Review*, Vol. 6, No. 1, pp. 18-3.
- VIVAS, Harvy (2007). *Educación, background familiar y calidad de los entornos locales en Colombia*. Tesis doctoral, Universitat Autònoma de Barcelona, España.

## Anexos

### A1. Resultados de máxima verosimilitud (no estandarizados)



## A2. Resultados de máxima verosimilitud (estandarizados)



El **path analysis** que se representa en los dos diagramas muestra la configuración final del modelo de ecuaciones estructurales utilizando la nomenclatura convencional Lisrel. La forma gráfica corresponde a la generada en la interfaz de modelamiento de Amos.

Los rectángulos contienen las variables endógenas y exógenas observables, mientras que los trazos circulares representan las variables latentes (endógenas y exógenas). Los óvalos contienen los errores de medida y las flechas representan el sentido de los nexos, condensando en valores paramétricos los coeficientes de regresión correspondientes a la estructura de estas relaciones. Los vínculos causales aparecen representados con flechas unidireccionales entre las variables exógenas y las endógenas, mientras que las correlaciones se representan a través de líneas bidireccionales o trazos curvilíneos en dirección recíproca.