



ASOCIACION ARGENTINA
DE ECONOMIA POLITICA

ANALES | ASOCIACION ARGENTINA DE ECONOMIA POLITICA

XLVII Reunión Anual

Noviembre de 2012

ISSN 1852-0022

ISBN 978-987-28590-0-8

CAPITAL HUMANO COMO UNA VARIABLE
LATENTE MULTIDIMENSIONAL

Laverde Henry

CAPITAL HUMANO COMO UNA VARIABLE LATENTE MULTIDIMENSIONAL[†]

Henry Laverde[‡]

Facultad de Ciencias Económicas
Universidad Nacional de La Plata, UNLP

Esta versión: Agosto de 2012

Resumen

El objetivo de este documento es estimar el valor monetario del capital humano nacional para Colombia usando el método sugerido por Dagum y Slottje (2000). El método combina una técnica de variables latentes multidimensionales con un enfoque actuarial que permite obtener el valor agregado del capital humano. Las estimaciones están basadas en la Encuesta Continua de Hogares del año 2001. Aunque, los resultados indican que la educación es un componente importante en la formación del capital humano también permiten dilucidar que factores relacionados a los entornos laborales y sociodemográficos juegan un papel significativo y, por lo tanto, que los indicadores que se han venido utilizado tradicionalmente como proxy, los cuales se basan en conceptos unidimensionales, podrían estar sesgados por cuanto no alcanzan a capturar muchos de los componentes que enmarcan al concepto.

Clasificación JEL: J24, O47, E24.

Palabras claves: Capital humano, PLS-PM, Variables latentes.

[†] Este trabajo hace parte de la Tesis para optar al título de doctorado en Economía de UNLP, dirigido por el profesor Adolfo Sturzenegger.

[‡] Esta es una versión preliminar, para comentarios hlaverde@gmail.com.

1 Introducción

En las últimas décadas la búsqueda de una medida apropiada de capital humano ha recobrado una especial importancia, particularmente por su uso en las investigaciones empíricas de crecimiento económico. Trabajos como Kyriacou (1991), Wolff (2000), Bils y Klenow (2000) han señalado el sobredimensionamiento que se le ha dado al capital humano. Estos resultados, sorprendentes a la luz de los modelos de crecimiento endógeno, señalan que la utilización de una mala proxy del capital humano es la causante de esta aparente baja relación [Cohen, D. y Soto, M., 2006]. Los enfoques tradicionales para estimar este stock están plagados de limitaciones y críticas. Estos métodos tienen una larga trayectoria en la literatura y se han generalizado en tres perspectivas diferentes: retrospectivo, prospectivo y educativo¹.

El primer enfoque aproxima el capital humano a sus costos de producción, fundamentalmente en las inversiones realizadas en un individuo hasta alcanzar una edad determinada. Este enfoque ha sido muy criticado por cuanto la clasificación de lo que se considera inversión o gasto en capital humano es un tema sumamente subjetivo. Por otro lado, sus precios no son fácilmente identificables y los métodos de depreciación son cuestionables. Además, como señala Le et al. (2003) al igual que el capital físico no existe necesariamente una relación entre inversiones y la calidad del producto, el valor del capital es determinado por su demanda y no por los costos de producción.

Por otro lado, el enfoque prospectivo confina al capital humano a la perspectiva de su retorno, es decir, la capacidad que tiene una persona para generar ingresos a lo largo de su vida. Este método calcula el stock bajo el supuesto de que los ingresos de los trabajadores deben reflejar la productividad de estos, por cuanto es el mercado el mejor instrumento para identificar y valorar los componentes del capital humano. No obstante, este enfoque también cuenta con algunas limitaciones: *i*) es sensible a las tasas de descuento utilizadas para descontar los ingresos esperados, *ii*) no necesariamente las fluctuaciones de los ingresos estarán aparejadas con las de productividad de los trabajadores, *iii*) y quizás el inconveniente fundamental, especialmente para la utilización en regresiones de crecimiento económico, es el problema de circularidad que se puede presentar al intentar introducir esta medida en una función de producción.

El enfoque más utilizado en la literatura de crecimiento económico es el basado en educación. Esta rama de la literatura considera que la educación formal es el principal componente del capital humano y que, por lo tanto, estará bien caracterizado por una variable educativa, generalmente los años promedio de educación. Sin embargo, la educación formal es apenas un fragmento de un concepto mucho más complejo y, por ende, la utilización de esta variable como único elemento de consideración aleja la posibilidad de evaluar otros componentes de relevancia. Además, asumiendo homogeneidad entre los diferentes individuos, este enfoque no tiene en cuenta la calidad del capital humano.

El concepto de capital humano no puede ser concebido como unidimensional, sino por el contrario, como el resultado compuesto de la confluencia de varias variables entre las que se encuentra la educación formal y no formal, el aprendizaje en el trabajo, las habilidades innatas, las condiciones sociodemográficas (como el género, la edad, la estratificación socioeconómica, etc.), entre otras. Al mismo tiempo, el capital humano influye sobre la productividad de los trabajadores y sobre los retornos esperados. Como consecuencia se confiere a este stock una doble connotación al

¹ Le et al. (2003) muestra en profundidad las ventajas y desventajas de cada uno de estos enfoques.

ser tanto resultado como reflejo. Por otra parte, este stock es también una variable no observable con lo cual su estimación debe realizarse de manera indirecta. Así pues, un problema adicional de los métodos tradicionales es estimar al capital humano como si tratara de una variable unidimensional y observable.

Para minimizar los problemas mencionados algunos autores han optado por realizar combinaciones de diferentes enfoques. Por ejemplo, Tao y Stinson (1997) realizan una estimación del stock del capital humano mediante la integración del enfoque prospectivo y retrospectivo. Sin embargo, es la propuesta de Dagum y Slottje (2000) la que reconoce el carácter multidimensional y latente del capital humano a través de un modelo de ecuaciones estructurales (SEM, por sus siglas en inglés), el cual es luego fusionado con la metodología prospectiva para obtener una medida en términos monetarios.

Dadas las características del capital humano los modelos SEM encajan muy bien en su estimación. Estos modelos incluyen una variedad de metodologías estadísticas que permiten estimar las relaciones causales, de acuerdo a un modelo teórico, conectando dos o más variables latentes (no observables) medidas a través de indicadores observables. Dentro de este conjunto de metodologías se destaca el Modelo de Ruta de Mínimos Cuadrados Parciales (PLS Path Modeling, PLS-PM) que permite calcular relaciones entre las variables latentes por medio de modelos formativos y reflexivos. A pesar de lo apropiado de esta metodología para estimar el capital humano, por cuanto ella tiene en cuenta el carácter multidimensional y latente del concepto, es sorprendente la casi inexistente literatura empírica a su alrededor.

De esta manera el objetivo principal de este documento es estimar el capital humano a través de la utilización de PLS-PM siguiendo la propuesta de Dagum et al. (2000) para conseguir los valores monetarios de este stock para el caso colombiano. La medición de esta proxy a través de esta metodología permitirá no solo obtener un valor agregado sino también analizar la importancia relativa de cada uno de los componentes que afectan la formación del capital humano y su distribución.

Este documento se divide, además de esta sección, como sigue. En la sección 2 se discute la metodología alrededor de PLS Path Modeling. En la sección 3 se presenta la especificación de los modelos utilizados y se describe la construcción de los datos necesarios para la implementación del método. En la sección 4 se presentan y discuten los resultados. Finalmente, en la sección 5 se concluye.

2 Algoritmo del Modelo de Ruta de Mínimos Cuadrados Parciales (PLS-PM)

Los modelos de ruta (*path models*) son la lógica extensión de los modelos de regresión dado que ellos involucran el análisis de ecuaciones de regresión múltiple simultáneas, tomando en cuenta los efectos directos e indirectos entre las variables. Sin embargo, los modelos de regresión multivariados múltiples únicamente toman en cuenta las relaciones entre las variables independientes y dependientes observables. Cuando las variables internas del modelo de ruta son latentes, cuya media es inferida por un conjunto de indicadores observados, el modelo de análisis de ruta se denomina modelo de ecuaciones estructurales.

Cada modelo SEM involucra dos niveles de relaciones: la primera toma en cuenta la relación entre las variables observadas (usualmente llamados indicadores o variables manifiestas

(VM)) y la correspondiente variable latente (VL), a este tipo de modelo se le conoce como externo o de medida. El segundo modelo considera la relación causal entre las variables latentes, el cual recibe el nombre de modelo interno o estructural.

Algunos métodos han sido desarrollados para estimar los parámetros tanto del modelo de medida como del estructural, entre ellos PLS-PM [Wold, 1975 y 1982]. El objetivo de esta metodología es el de dar una estimación de las variables latentes de tal manera que ellas estén lo más correlacionadas posible (de acuerdo a la estructura de un diagrama de ruta) y lo más representativas a cada bloque compuesto por las variables manifiestas.

En concreto, PLS-PM intenta estimar, a través de un sistema de ecuaciones interdependientes bajo regresiones simples y múltiples, la cadena de relaciones entre las variables manifiestas y sus propias variables latentes y entre las variables latentes del modelo.

Formalmente, asumamos P variables observables sobre N unidades ($i = 1, \dots, N$). Los datos resultantes x_{npq} son agrupados en una matriz particionada de datos estandarizados X :

$$X = [X_1, \dots, X_q, \dots, X_Q],$$

donde X_q es el q -ésimo bloque genérico. En el modelo PLS-PM para cada variable latente endógena, el modelo estructural puede ser escrito como:

$$\xi_j = \beta_{0j} + \sum_{q: \xi_q \rightarrow \xi_j} \beta_{qj} \xi_q + \zeta_j \quad (1)$$

donde ξ_j ($j = 1, \dots, J$) es la variable latente endógena genérica, β_{qj} es el coeficiente de ruta genérico interrelacionando a la q -ésima variable latente exógena con la j -ésima endógena y ζ_j es el error en la relación interna. Este modelo está sujeto a hipótesis de especificación, la cual implica cero correlaciones entre los residuales y las variables latentes:

$$E(\zeta_j / \xi_j, \xi_i) = 0 \quad (2)$$

Por otra parte, la formulación del modelo de medida depende de la naturaleza de las relaciones entre las variables latentes y las correspondientes variables manifiestas. Diferentes tipos de modelos de medidas están disponibles: el modelo reflexivo, el modelo formativo y el modelo MIMIC.

En un modelo reflexivo un bloque de variables manifiestas se relaciona a una variable latente, y por lo tanto, se asume medir un único concepto subyacente. Cada variable manifiesta refleja (es un efecto de) la correspondiente variable latente y juega un rol de variable endógena en el bloque que especifica el modelo de medida. En el modelo de medida reflexivo, los indicadores que conectan a la misma variable latente deberían covariar: cambios en un indicador implica cambios en los otros. Por lo tanto, la consistencia interna tiene que ser comprobada, es decir, cada bloque es asumido ser homogéneo y unidimensional². En términos más formales, en un modelo reflexivo cada

² La unidimensionalidad implica que un solo rasgo latente o constructo se encuentra en la base de un conjunto de indicadores. En otras palabras, un instrumento será unidimensional si las respuestas dadas a él son producidas en base a un único atributo. Con esto lo que se quiere lograr es que la mayor cantidad de la varianza observada en las repuestas a los indicadores sea explicada por un solo atributo latente.

variable manifiesta es relacionada a la correspondiente variable latente por un modelo de regresión simple:

$$x_{pq} = \lambda_{p0} + \lambda_{pq}\xi_q + \epsilon_{pq} \quad (3)$$

donde λ_{pq} es la carga asociada a la p -ésima variable manifiesta en el q -ésimo bloque y el termino de error ϵ_{pq} representa la imprecisión en el proceso de medida. Un supuesto detrás de este modelo es que el error ϵ_{pq} tiene media cero y no esta correlacionado con la variable latente del mismo bloque:

$$E(x_{pq}/\xi_q) = \lambda_{p0} + \lambda_{pq}\xi_q. \quad (4)$$

Por su parte, en un modelo formativo, cada variable manifiesta o cada sub-bloque de variables manifiestas representa una diferente dimensión del concepto subyacente. Por lo tanto, a diferencia del modelo reflexivo, el modelo formativo no asume homogeneidad ni unidimensionalidad del bloque. La variable latente es definida como una combinación lineal de las correspondientes variables manifiestas, así cada variable manifiesta es una variable exógena en el modelo de medida. Estos indicadores no necesitan covariar: cambios en un indicador no implican cambios en los otros y la consistencia interna ya no es más un problema. De esta manera el modelo de medida puede ser expresado como:

$$\xi_q = \sum_{p=1} \omega_{pq}x_{pq} + \delta_q \quad (5)$$

donde ω_{pq} es el coeficiente que conecta cada variable manifiesta a la correspondiente variable latente y el término de error δ_q representa la fracción de la correspondiente variable latente no contabilizada por el bloque de variables manifiestas. Este modelo también está sujeto al supuesto de especificación del predictor, es decir:

$$E(\xi_q/x_{pq}) = \sum_{p=1} \omega_{pq}x_{pq} \quad (6)$$

Finalmente, el modelo MIMIC es una combinación de los modelos reflexivo y formativo dentro del mismo bloque de variables manifiestas.

Independientemente del tipo de modelo de medida, bajo convergencia del algoritmo, los valores de la variable latente estandarizada ($\hat{\xi}_q$) asociada a la q -ésima variable latente (ξ_q) es calculada como una combinación lineal de su propio bloque de variable manifiestas por medio de las llamadas *relaciones de peso* (*weight relations*) definida como:

$$\hat{\xi}_q = \sum_{p=1} w_{pq}x_{pq} \quad (7)$$

donde las variables x_{pq} son estandarizadas y los w_{pq} son los pesos externos (*outer weights*). Esos pesos son logrados bajo convergencia del algoritmo y luego transformados para producir los valores de la variable latente estandarizada. En un modelo PLS-PM un proceso iterativo permite la estimación de los parámetros del modelo, es decir los pesos externos (w_{pq}) y los valores de las variables latentes (ξ_q). El proceso de estimación es llamado parcial porque resuelve los bloques uno a la vez por medio de alternar regresiones simples y múltiples. Los coeficientes de ruta (*path coefficients*) se consiguen después de una regresión regular entre los valores estimados de las

variables latentes. La tabla No. 1 resume el proceso iterativo³. En términos generales el algoritmo PLS-PM se realiza bajo tres pasos. En el primer paso se estima las variables latentes. Como ni las cargas ni las variables latentes son conocidas una estandarización es necesaria. Generalmente, todas las variables latentes son estandarizadas, con media cero y varianza unitaria.

Tabla No. 1. Algoritmo: PLS P-M basado sobre el algoritmo de Lohmoller con las siguientes opciones: esquema centroide, valores de las variables latentes estandarizadas, regresiones por medio de OLS.

Insumo: $X = [X_1, \dots, X_q, \dots, X_Q]$, es decir Q bloques de variables manifiestas estandarizadas;

Producto: w_q, ξ_q, β_j ;

1: para todo $q = 1, \dots, Q$ hacer

2: valor inicial w_q

3: $v_q \propto \pm \sum_{p=1}^{P_q} w_{pq} x_{pq} = \pm X_q w_q$;

4: $e_{qq'} = \text{signo}[\text{cor}(v_q, v_{q'})]$ siguiendo el esquema centroide

5: $\vartheta_q \propto \sum_{q'=1}^{Q'} e_{qq'} v_{q'}$

6: actualizar w_q ;

(a) $w_{pq} = \text{cov}(x_{pq}, \vartheta_q)$ para el modo A

(b) $w_q = \left(\frac{X_q' X_q}{N} \right)^{-1} \left(\frac{X_q' \vartheta_q}{N} \right)$ para el modo B

7: fin

8: los pasos 1-7 son repetidos hasta converger, cuando los pesos externos son logrados, es decir hasta:

$$\max\{w_{pq} \text{ de la iteracion actual} - w_{pq} \text{ de la iteracion anterior}\} < \Delta$$

donde Δ es una tolerancia de convergencia usualmente ajustada a 0.0001 o menos

9: Bajo convergencia:

- (1) Para cada bloque los valores de las variables latentes estandarizadas son calculadas como los agregados ponderados de las variables manifiestas:

$$\hat{\xi}_q \propto X_q w_q,$$

- (2) Para cada variable endógena ξ_j ($j = 1, \dots, J$), el vector de coeficientes de ruta es estimado por medio de una regresión OLS como:

$$\beta_j = (\hat{\Xi}' \hat{\Xi})^{-1} \hat{\Xi}' \hat{\xi}_j,$$

Donde $\hat{\Xi}$ incluye los valores de las variables latentes que explican la j -ésima variable latente endógena ξ_j , y $\hat{\xi}_j$ es el valor de la variable latente de la j -ésima variable latente endógena.

Fuente: tomado de V. Esposito Vinzi et al. (2010)

Cada variable latente es calculada como una combinación lineal de sus variables manifiestas de tal manera que no únicamente logra la máxima correlación con sus variables manifiestas sino también se relaciona con otras variables latentes de acuerdo con el modelo interno. Las variables latentes calculadas son luego conectadas a las ecuaciones internas y externas para derivar sus parámetros. Una vez logrado la primera estimación de las variables latentes el algoritmo actualiza

³ Para detalles del algoritmo ver V. Esposito Vinzi et al. (eds), Handbook of Partial Least Squares, Springer Handbook of Computational Statistics (2010).

los pesos externos w_{pq} . Para ello, existen dos modos diferentes de actualizar dichos pesos usualmente relacionados a las dos clases de modelos de medida (es decir, el esquema formativo o el reflexivo):

- Modo A: cada peso externo w_{pq} es el coeficiente de regresión en una regresión simple de la p -ésima variable manifiesta del q -ésimo bloque (x_{pq}) en la estimación interna de la q -ésima variable latente ϑ_q . Es importante notar que como los valores de la variable latente están estandarizados, el peso externo genérico (w_{pq}) es obtenido como en 6(a) del algoritmo (tabla No. 1), es decir como la covarianza entre cada variable manifiesta y la correspondiente estimación interna de la variable latente.
- Modo B: el vector w_q de los pesos w_{pq} asociados a las variables manifiestas del q -ésimo bloque es el vector de coeficientes de la regresión múltiple de la estimación interna de la q -ésima variable latente ϑ_q sobre sus variables manifiestas estandarizadas X_q , 6(b) del algoritmo.

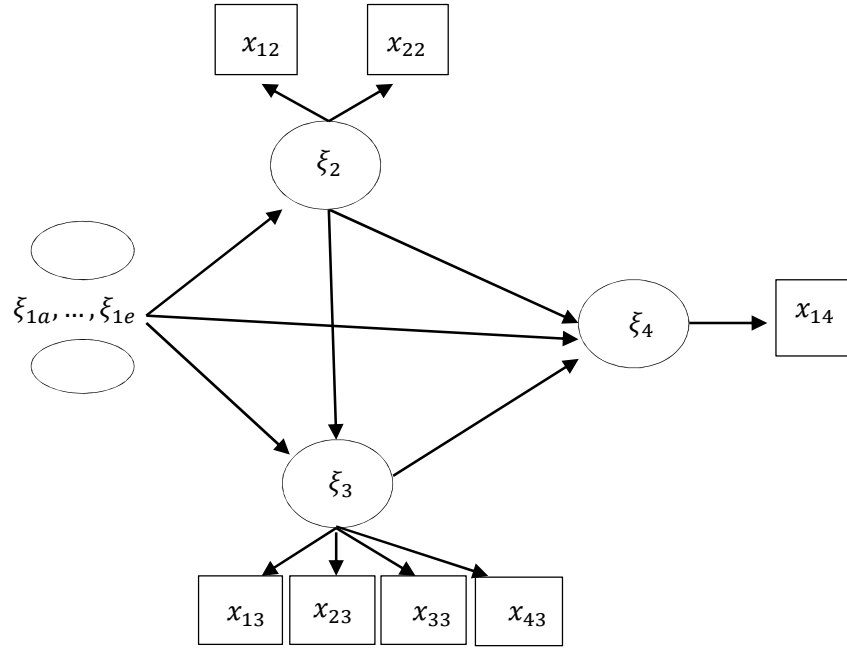
La elección del modo de estimación de los pesos está estrictamente relacionada a la naturaleza del modelo. Para un modelo reflexivo el modo A es más apropiado, mientras el modo B es mejor para los modelos formativos. Además, el modo A es sugerido para variables latentes endógenas, mientras el modo B para las exógenas.

El algoritmo es iterado hasta converger. Después de la convergencia los coeficientes estructurales son estimados a través de una regresión múltiple entre los valores de las variables latentes.

3 Datos y especificación del modelo

El modelo utilizado en este documento es ilustrado en la figura 1, mediante un diagrama de ruta en el cual las elipses o círculos representan las variables latentes y los rectángulos o cuadrados se refieren a las variables manifiestas. Las flechas representan la relación de causalidad entre las variables (latentes o manifiestas). Al mismo tiempo la dirección de estas flechas, en los modelos de medida, determinan si estos son reflexivos o formativos. En este modelo se ha querido caracterizar al capital humano con la ayuda de otras tres variables latentes. En primer lugar, el concepto latente que abarca los antecedentes familiares (condiciones sociales, demográficas y económicas) (ξ_1) se considera exógeno por cuanto las diferencias en las estructuras de los entornos inmediatos repercuten y condicional las trayectorias y los logros efectivos en los ámbitos educativos y laborales. Aquellos individuos que se forjan en ambientes hostiles con cabezas de familia femeninos, localizados en entornos de menor desarrollo, con medios económicos precarios y entornos familiares desfavorables, cuentan con desventajas acumuladas que se replican a lo largo de su ciclo vital y que les impide generar trayectorias virtuosas de acumulación de capital humano y buenas oportunidades de inserción futura a los mercados laborales.

Figura 1: Diagrama de ruta del modelo de media y estructural para el CH multidimensional



Ahora, las habilidades innatas y cognitivas, el entorno educativo de la familia y la escuela, entre otros, determinarán los logros educativos (ξ_2), los cuales a su vez influyen sobre los resultados laborales y los niveles de capital humano.

Por otro lado, el tipo de trabajo, el estatus y el sector donde laboran los trabajadores reflejan los logros laborales (ξ_3), aquellos trabajadores que realizan actividades de servidumbre, que ocupan posiciones de bajo perfil y se ubican en sectores de bajo valor agregado disminuyen su capacidad para negociar las condiciones contractuales y salariales.

Finalmente, el capital humano es el resultado de cada uno de estos conceptos latentes y reflejo de las trayectorias salariales de los trabajadores. De esta manera el modelo interno está compuesto por una variable exógena (antecedentes familiares, ξ_1) y tres endógenas (logro educativo (ξ_2), logro laboral (ξ_3) y capital humano (ξ_4)).

En el modelo externo o de medida, el bloque de antecedentes familiares es subdividido en varias características siendo, para cada una de ellas, las variables latentes sus propias variables manifiestas y estimadas por el modo B. Los logros educativos, laborales y el capital humano son estimados por el modo A. La tabla No. 2 presenta las variables latentes y manifiestas junto con las ecuaciones de los modelos internos y externos. Como algunas variables manifiestas son de carácter categóricas y, aunque, PLS-PM es estructurado para manejar variables cuantitativas esto no es un problema porque puede ser transformada en variables dummy. De esta manera, variables como la localización geográfica es desagregada en seis variables dummy representando seis ciudades del

país, separadas de tal manera que las tres primeras ciudades (Bogotá, Medellín, Barranquilla) se espera que sean las que más aporten al desarrollo del capital humano por su importancia relativa en el país.

Tabla No. 2: Variables, parámetros y relaciones del modelo

Variables latentes y ecuaciones del modelo interno	Variables manifiestas	Ecuaciones del modelo externo
$\xi_{1a}, \dots, \xi_{1e}$ Antecedentes familiares (Exógena)	x_{11} : Genero x_{21} : Edad x_{31} : Edad al cuadrado x_{41} : localización geográfica x_{51} : Estratificación económica	Las variables latentes son las mismas variables manifiestas. Modo B
ξ_2 Logros educativos (Endógena) $\xi_2 = \beta_{12}\xi_1 + \zeta_2$	x_{12} : nivel educativo alcanzado x_{22} : alfabetismo	$x_{p2} = \lambda_{p2}\xi_2 + \epsilon_{p2}$ Modo A
ξ_3 Logros laborales (Endógena) $\xi_3 = \beta_{13}\xi_1 + \beta_{23}\xi_2 + \zeta_3$	x_{13} : Experiencia x_{23} : Experiencia cuadrado x_{33} : Estatus laboral x_{43} : Rama de actividad laboral	$x_{p3} = \lambda_{p3}\xi_3 + \epsilon_{p3}$ Modo A
ξ_4 Capital Humano (Endógena) $\xi_4 = \beta_{14}\xi_1 + \beta_{24}\xi_2 + \beta_{34}\xi_3 + \zeta_4$	x_{14} : Ingresos laborales	$x_{p4} = \lambda_{p4}\xi_4 + \epsilon_{p4}$ Modo A

Los datos para estimar el valor monetario del stock de capital humano provienen de la Encuesta Continua de Hogares (ECH) colombiana del año 2001 recolectados por el Departamento Administrativo Nacional de Estadísticas (DANE) de periodicidad trimestral. En este estudio solo se tiene en cuenta a las personas que se encuentran en la fuerza laboral, esto incluye a los que están trabajando o buscando trabajando. Además, se acota las edades de los individuos al rango comprendido entre 16 y 62 años. El límite inferior de este rango fue elegido por dos razones: en primer lugar, puesto que este estudio se centraliza en las actividades del mercado laboral, y a pesar de que en Colombia se considera a una persona capaz de trabajar a partir de los 12 años, en promedio un individuo con 16 años se halla en la culminación de sus estudios secundarios, edad en la cual una porción importante de la población colombiana ingresa al mercado laboral. En segundo lugar, se trata de seguir un rango similar al hecho en otros estudios para posibilitar las comparaciones entre países. Además, se eliminan aquellos individuos que no reportan su nivel de educación.

Con esto, el tamaño de la muestra final es 20339 individuos para las siete principales áreas metropolitanas del país. Las encuestas cuentan con un factor de expansión que permite aplicarlo a los datos muestrales, con lo que se consigue para cada individuo de la muestra el peso o representación que le corresponde en el universo investigado.

Un problema que surge con este tipo de encuestas lo constituyen los ingresos laborales de los trabajadores. Una parte importante de la encuesta no contiene información sobre los ingresos laborales. Por lo tanto, para corregir los posibles sesgos que puede generar este problema se implementa un modelo regresión multivariado al estilo Mincer (1974), para que de esta manera se

pueda imputar los ingresos no declarados de los trabajadores. A través de este modelo se consigue asignarle a un trabajador el ingreso de otro que cuenta con las mismas características⁴.

Los indicadores o variables manifiestas tomadas en cuenta para la estimación del capital humano son distribuidos en cuatro bloques que determinan las variables latentes: antecedentes familiares, logros educativos, logros laborales y formación en capital humano.

El enfoque tradicional de medición del capital humano sugiere que la región de residencia, la edad, el género y las características de los hogares pueden cambiar los efectos sobre las cantidades invertidas en capital humano, (Wobmann, 2003). Por esta razón, en el primer bloque se incluyen variables como género, edad, edad al cuadrado, localización geográfica y estratificación socioeconómica.

Para estimar los logros educativos se suele utilizar indicadores, recomendados en la medición del CH, tales como el porcentaje de la población que ha completado exitosamente varios niveles de educación formal, relación número de personas-años de escolaridad incorporadas en el mercado de trabajo, entre otros (Hanushek, 1996; OECD, 1998; Wobmann, 2003). Sin embargo, esos indicadores no son apropiados para medir los logros educativos por cuanto o no se cuenta con los datos o no están suficiente desagregados. Por lo tanto, en este estudio se utiliza dos indicadores: nivel educativo alcanzado y un indicador de alfabetismo⁵.

Los efectos del tiempo de inversión en el trabajo sobre el stock de capital humano están influenciados por los niveles de entrenamiento (OECD, 1998), experiencia ganada a través de “aprendiendo haciendo” (Wobmann, 2003), estatus del trabajo, ocupación y el sector de actividad (Jorgenson y Fraumeni, 1989). De esta manera, la experiencia, experiencia al cuadrado, estatus de empleo y rama de actividad han sido seleccionados entre los indicadores del bloque para los logros laborales.

Finalmente, los retornos de toda la vida del capital humano son descritos por los ingresos laborales de los trabajadores.

4 Resultados

Para la estimación del modelo de la figura 1 se usa el software R mediante el paquete *plspm* (<http://cran.r-project.org/web/packages/plspm/index.html>) [Sánchez et al., 2009] y se utiliza la opción *centroide* para los pesos internos.

El modelo PLS-PM carece de un buen criterio de optimización global, así que no hay una función de ajuste global. De esta manera, cada parte del modelo necesita ser validada: el modelo de

⁴ Siguiendo la convención tradicional en este tipo de literatura este modelo se puede estimar por MCO, para luego obtener los coeficientes que servirán para imputar los ingresos de los trabajadores que no lo declararon a partir de sus características. Sin embargo, puede presentarse un sesgo de selección, ya que los trabajadores que no reportan ingresos pueden tener incentivos para hacerlo dado que sus ingresos pudieran ser superiores a los que declararon. En este caso, el término de error podría estar afectado por el censuramiento de los ingresos, conllevando a resultados y conclusiones que no describen como se esperaría las características de la población en estudio. Para corregir este problema se utilizó la metodología propuesta por Heckman (1979).

⁵ Los altos índices de analfabetismo que persisten en Colombia son aún muy altos. Estos índices oscilan entre un 6% en las cabeceras municipales y un 21% en las zonas rurales. La situación educativa se complica todavía más si abrimos el concepto de analfabetismo a otros campos como el tecnológico e informativo o simplemente a la lectura de los mensajes que circulan en nuestra sociedad a través de las imágenes, mediante los llamados mensajes o textos icónicos.

medida, el modelo estructural y el modelo general. En la literatura de PLS-PM se han desarrollado tres diferentes índices para evaluar cada uno de estos componentes, a saber: el índice de comunalidad, el índice de redundancia y el índice de bondad de ajuste (*GoF*). Por lo tanto, antes de analizar los coeficientes alrededor del capital humano se debe validar el modelo completo para examinar el grado en el cual se producen resultados confiables.

4.1 Calidad del modelo externo

La manera de evaluar el modelo externo depende del tipo de medida especificada por los indicadores (es decir, si son reflexivos o formativos). Cuando los indicadores son reflexivos se asume que son causados por una variable latente, entonces es lógico suponer que ellos están cercanamente relacionados, de tal manera, que ellos están en el mismo espacio dimensional. Para evaluar la unidimensionalidad de un bloque reflexivo existen los siguientes índices:

- Alfa de Cronbach: Este índice evalúa que tan bien un bloque de indicadores mide su correspondiente variable latente⁶. Como una regla general un bloque es considerado homogéneo si este índice es más grande que 0.7.
- Rho de Dillon-Goldstein⁷: Un bloque es considerado homogéneo si este índice es más grande que 0.7.
- Análisis de componentes principales de un bloque: un bloque puede ser considerado unidimensional si el primer valor propio de su matriz de correlaciones es más grande que uno, mientras que el segundo debe ser menor a uno.

De acuerdo a Chin (1998), el Rho de Dillon-Goldstein es considerado ser un mejor indicador que el de Alfa de Cronbach dado que el primero se basa en los resultados del modelo (es decir, las cargas) en lugar de las correlaciones observadas entre las variables manifiestas en la base de datos. En ese orden de ideas en este documento se evalúa la homogeneidad y unidimensionalidad de un bloque por medio de los índices de Rho de Dillon-Goldstein y análisis de componentes principales. La tabla 3 muestra que para todos los bloques se satisfacen los requerimientos antes señalados.

Por otro lado, en un modelo formativo, cada variable manifiesta o cada sub-bloque de variables manifiestas representa una diferente dimensión del concepto subyacente. Por lo tanto, los modelos formativos no asumen ni homogeneidad ni unidimensionalidad del bloque. Como el bloque de antecedentes familiares es estimado por medio de un modelo formativo y de manera independiente no es necesario este tipo de análisis. En el caso del bloque del capital humano los resultados no se muestran por cuanto este bloque cuenta con una sola variable manifiesta. Para el bloque de los logros laborales, las variables manifiestas experiencia y experiencia al cuadrado fueron excluidas del modelo por no ser significativas y de signos no esperados.

⁶ Definido como $\alpha = \frac{\sum_{p \neq p'} \text{cor}(x_{pq}, x_{p'q})}{P_q + \sum_{p \neq p'} \text{cor}(x_{pq}, x_{p'q})} \times \frac{P_q}{P_q - 1}$, donde P_q es el número de variables manifiestas en el q -ésimo bloque.

⁷ Mejor conocido como “*composite reliability*”, es calculado como $\rho = \frac{(\sum_{p=1}^{P_q} \lambda_{pq})^2}{(\sum_{p=1}^{P_q} \lambda_{pq})^2 + \sum_{p=1}^{P_q} (1 - \lambda_{pq}^2)}$.

Tabla No. 3: Análisis de homogeneidad y unidimensionalidad

Bloque	Variable manifiestas	Rho de D-G	1er valor propio	2do Valor propio
Logro educativo	2	0.855	1.49	0.505
Logro laboral	2	0.770	1.25	0.749

Para la variable manifiesta que caracteriza el estatus laboral, dado su carácter categórico (particular, gobierno, domestico, cuenta propia, patrón o jefe, familiar sin remuneración, otro), se decidió transformarla en una variable dummy de tal manera que estuviera lo más correlacionada con el bloque. Como resultado los estatus trabajar con el gobierno y ser patrón representan mejor el concepto buscado para este bloque, así la variable dummy toma el valor de uno si un individuo trabaja en el gobierno y es patrón y cero en cualquier otro caso. De la misma manera para la variable manifiesta que caracteriza la rama de actividad (Agrícola, industrial, servicios, administración pública) se realizaron varios modelos donde se incluían todas las variables, descartando aquellas que presentaban una baja carga, baja significancia estadística y no pasar los criterios de homogeneidad y unidimensionalidad. Con esto la variable dummy que mejor representa al bloque es para aquellos individuos que pertenecen a la administración pública.

Además de la homogeneidad y unidimensionalidad de un bloque reflexivo se debe evaluar si los indicadores de este bloque están bien explicados por su variable latente. Una manera de hacerlo es examinando sus cargas, Gefen et al. (2000) sugiere que únicamente las variables manifiestas con cargas más grandes que 0.4 son significativas. Por otro lado, se puede utilizar el índice de comunalidad. Este índice mide cuanto de la variabilidad de las variables manifiestas en el *q-ésimo* bloque es explicado por su propia variable latente. La idea es tener más varianza compartida entre las variables latentes y sus indicadores que ruido. Por lo general, "buenos" valores de comunalidad son mayores a 0.5 [Tenenhaus et al., 2005]. Los resultados para el modelo de la figura 1 (Tabla 4) muestran que, con excepción de la variable alfabetismo, los indicadores seleccionados cumplen con este requerimiento. Como la carga de la variable alfabetismo es mayor que 0.4 y validada por el método de remuestreo bootstrapping se mantiene en el modelo.

Tabla No. 4: Estimaciones del modelo PLS-PM del modelo externo

Variable manifiesta	Cargas	Comunalidad	Redundancia
Logro educativo			
Alfabetismo	0.634	0.402	0.0705
Nivel educativo	0.986	0.972	0.1703
Logro laboral			
Actividad laboral	0.809	0.655	0.1200
Estatus laboral	0.772	0.596	0.1093
Validación por Bootstrap			
Variable manifiesta	Cargas (Boot)	Error estándar	t estadístico
Logro educativo			
Alfabetismo	0.640	3.28e-03	195.1
Nivel educativo	0.984	3.28e-03	300.0
Logro laboral			
Actividad laboral	0.808	6.99e-03	115.6
Estatus laboral	0.773	8.53e-03	90.6

4.2 Calidad del modelo interno

El modelo interno puede ser examinado ahora que el modelo externo ha sido validado. Dicha verificación puede ser realizada por las siguientes medidas: R^2 , comunalidad promedio, redundancia promedio y varianza extraída media (AVE, por sus siglas en inglés). Aunque la calidad de cada modelo estructural es medida por una simple evaluación del índice de ajuste R^2 , no es suficiente para evaluar el total del modelo estructural. En particular, como las ecuaciones estructurales son estimadas una vez la convergencia es lograda y los valores de las variables latentes son estimadas, entonces el R^2 toma en cuenta únicamente el ajuste de cada ecuación de regresión en el modelo estructural. El índice de comunalidad promedio indica la cantidad promedio de varianza entre una variable latente y sus indicadores que es común a ambos. Las variables latentes satisfactorias deberían explicar más de la mitad de la varianza, esto es, el índice de comunalidad promedio debería exceder 0.5 [Tenenhaus et al., 2005].

Tabla No. 4: Estimaciones del modelo PLS-PM del modelo interno

Variable manifiesta	R^2	Com. Promedio	Redun. Promedio	AVE
Logro educativo	0.175	0.687	0.120	0.664
Logro laboral	0.183	0.625	0.115	0.639
Capital humano	0.607	1	0.607	1
Validación por Bootstrap				
Variable manifiesta	R^2 (Boot)	Error estándar	t estadístico	
Logro educativo	0.174	0.00608	28.6	
Logro laboral	0.183	0.00490	37.3	
Capital humano	0.605	0.00609	99.3	

El índice de redundancia promedio refleja la cantidad de varianza de los indicadores en un bloque endógeno de las variables latentes independientes asociada a las variables latentes endógenas.

Por último, el índice AVE mide la cantidad de varianza explicada por la variable latente relativa a la cantidad debido al error de medida. De acuerdo a Fornell y Larcker (1981), este índice debería exceder 0.5.

De esta manera, a la luz de estos indicadores, se puede corroborar la validez del modelo estructural. Primeramente, en este estudio el modelo explica 17.5% de la variación en los logros educativos, el 18.3% de los laborales y el 60.7% para el capital humano. Estos ratios muestran un alto poder predictivo dada la complejidad y estructura del modelo, particularmente por ser uno de sección cruzada que incluye 20339 individuos, incluso a niveles de significancia del 1%. Del mismo modo, la tabla 4 confirma que el modelo interno pasa los dos índices de calidad, es decir, comunalidad promedio y AVE.

Como se mencionó antes, no existe un índice de ajuste global en PLS-PM. No obstante, un criterio de bondad de ajuste ha sido desarrollado con el objetivo de tener en cuenta el comportamiento del modelo tanto del modelo de medida como del estructural, este índice se conoce como el *GoF*. Una versión normalizada ha sido obtenida y conocida como *GoF* relativo, el cual está acotado entre 0 y 1. Este índice es descriptivo, es decir no hay un umbral de inferencia para juzgar la significancia estadística de sus valores. Sin embargo, como regla general, un valor del *GoF* relativo igual o más alto que 0.9 habla claramente en favor del modelo [Tenenhaus et al., 2005]. Para el modelo del presente estudio el índice *GoF* es igual a 0.4595 y el *GoF* relativo a 0.9007.

4.3 Estimación del capital humano

Una vez validado el modelo, lo fundamental en el análisis son los coeficientes de ruta, los cuales ilustran las asociaciones entre las variables latentes. La tabla 5 muestra las estimaciones del modelo interno. De acuerdo a ella no parece existir diferencias en la localización geográfica, la mayoría de los coeficientes de esta variable tienen un bajo impacto y no son significativos. Las diferencias de género parecen ser sólo importantes para la formación capital humano (el coeficiente de ruta es igual a 0.1435 y estadísticamente significativo). Por otra parte, la estratificación socioeconómica tiene un impacto significativo en las tres variables latentes, particularmente en los logros educativos (coeficiente de ruta igual a 0,3628).

Tabla No.5: Coeficientes de ruta del modelo interno

	Coeficiente	Coef. (Boot)	Error estándar	T estadístico
Logro educativo				
Genero	-0.0242	-0.02379	0.00628	-3.79
Bogotá	0.0276	0.02825	0.00721	3.92
Medellín	-0.0123	-0.01212	0.00857	-1.41
Barranquilla	0.0374	0.03592	0.00896	4.0
Bucaramanga	-0.0756	-0.07549	0.00750	-10.0
Manizales	-0.0470	-0.04811	0.00769	-6.2
Pasto	-0.0231	-0.02433	0.00869	-2.8
Estrato socioeconómico	0.3628	0.36206	0.00788	45.9
Edad	0.4493	0.43734	0.04525	9.7
Edad al cuadrado	-0.6643	-0.65210	0.04710	-13.8
Logro Laboral				
Genero	0.0059	0.00653	0.00673	0.97
Bogotá	0.0111	0.01110	0.00874	1.27
Medellín	0.0080	0.00895	0.00796	1.12
Barranquilla	0.0159	0.01673	0.00924	1.81
Bucaramanga	0.0126	0.01394	0.00791	1.76
Manizales	0.0249	0.02448	0.00812	3.0
Pasto	0.0754	0.07560	0.00885	8.5
Estrato socioeconómico	0.1475	0.14918	0.00785	19.0
Edad	0.4817	0.47693	0.03432	13.9
Edad al cuadrado	-0.2941	-0.28891	0.03641	-7.9
Logro educativo	0.3134	0.31262	0.00756	41.4
Capital humano				
Genero	0.1435	0.14299	0.00446	32.0
Bogotá	0.0158	0.01629	0.00660	2.47
Medellín	0.0128	0.01348	0.00583	2.31
Barranquilla	0.0221	0.02373	0.00577	4.11
Bucaramanga	-0.0317	-0.03210	0.00580	-5.53
Manizales	-0.0244	-0.02459	0.00687	-3.6
Pasto	-0.0421	-0.04158	0.00531	-7.8
Estrato socioeconómico	0.1581	0.15915	0.00590	26.9
Edad	0.9565	0.95968	0.02827	33.9
Edad al cuadrado	-0.6829	-0.68665	0.02939	-23.4
Logro educativo	0.4984	0.49572	0.00710	69.8
Logro laboral	0.2638	0.26601	0.00712	37.4

Como se esperaba la edad es convoca y significativa. Por ejemplo, en el caso del capital humano por cada aumento de una desviación estándar esté se incrementa en 0.9565. La edad tanto en los logros laborales como en el capital humano se debe interpretar de modo particular al estar más asociada a una proxy de la experiencia de los trabajadores.

El enfoque educativo, con predominancia en la literatura de crecimiento económico, señala a la educación formal como el factor más importante en la formación del capital humano⁸, y por lo tanto, induce implícitamente a utilizar esta única variable como el elemento relevante para la comprobación empírica. Este argumento ha justificado y generalizado la utilización de variables educativas, como los años promedio de educación, para corroborar la existencia de una relación positiva entre crecimiento económico y capital humano. Sin embargo, es relevante señalar que aunque en este estudio se confirma que la educación juega un papel importante en la formación del capital humano también evidencia que factores tales como los antecedentes familiares y los laborales tienen un fuerte impacto sobre el capital humano. En concreto, mientras el capital humano se incrementa en 0.4984 desviaciones estándar por cada incremento de una desviación estándar de la educación, el impacto neto que tiene la experiencia y los logros laborales es de igual magnitud que el de la educación. Estos resultados, abonado a la importancia relativa que tienen los factores socio- demográficos, corroboran la relevancia de estimar al capital humano respetando su carácter multidimensional al señalar que la educación formal no es el único factor a tener en cuenta en dicha estimación.

Luego de obtener una versión estandarizada del capital humano (ξ_4), es posible conseguir una estimación en términos monetarios mediante la propuesta de Dagum et al. (2000). Estos autores pasan de una variable estandarizada a un valor monetario utilizando como referencia (*benchmarked*) o escala a la media de ingresos de los trabajadores que generan a lo largo de su vida, bajo el supuesto que los ingresos esperados reflejan la productividad de los trabajadores y, por lo tanto, los retornos del capital humano. En un primer paso, se realiza la siguiente transformación a las estimaciones de ξ_4 :

$$h(i) = \exp(\xi_{4i}) \quad (8)$$

donde i es la i -ésima unidad económica, luego se obtiene un valor promedio de $h(i)$ mediante la siguiente operación:

$$Av(h) = \frac{\sum_{i=1}^n h(i)f(i)}{\sum_{i=1}^n f(i)} \quad (9)$$

donde n es el tamaño de la muestra y $f(i)$ es el peso atado a la i -ésima observación de la muestra.

Para estimar los ingresos medios se utiliza un enfoque actuarial el cual supone que un trabajador espera ganar un ingreso, a la edad de x , igual al de un trabajador más viejo, con una edad de $x+t$. Por lo tanto, el promedio del capital humano de un individuo a la edad x es igual al valor presente del promedio de los ingresos por edad actualizado a una tasa de descuento y ponderado por la tasa de sobrevivencia. De esta manera, dada una tasa de descuento r y una tasa de sobrevivencia de la población p , el capital humano de un individuo, a la edad de x , es igual a:

⁸ Dado que aumenta las habilidades y productividad de los trabajadores, facilita la adsorción de nuevas tecnologías y tiene efectos esparcidos en la sociedad, como menores tasas de fertilidad, mayores niveles de educación en los niños y mejor distribución del ingreso que redundan en el crecimiento económico [Barro et al., 1993].

$$h(x) = \sum_{t=0}^{62-x} y(x+t)p(x, x+t)(1+r)^{-t} \quad (10)$$

para $x = 16, 17, \dots, 62$ y donde $y(x)$ es el ingreso promedio suavizado por un promedio móvil. Una vez estimada la ecuación (10) es posible obtener el valor promedio ponderado del capital humano de la población como:

$$AvCH(h) = \frac{\sum_{x=16}^{62} h(x)f(x)}{\sum_{x=16}^{62} f(x)} \quad (11)$$

donde $f(x)$ es el correspondiente peso que cada individuo representa en la población colombiana.

Como señala Dagum et al. (2000), para llegar al valor monetario corriente de la estimación del capital humano de las n observaciones de la muestra se realiza el cociente entre el promedio del capital humano dado por la ecuación (11) y el promedio de la transformación dado por la ecuación (9) y multiplicado este resultado por $h(i)$ dado por la ecuación (8). De esta manera, el capital humano de la i -ésima observación de la muestra es igual a:

$$CH(i) = h(i) \frac{AvCH(h)}{Av(h)}, \quad i = 1, 2, \dots, n. \quad (12)$$

El cual representa el vector de capital humano en unidades monetarias. Intuitivamente, el valor monetario del capital humano de una persona es igual a los ingresos de toda la vida de la población, ponderados por el nivel de capital humano que es relativo al capital humano promedio de la población. Finalmente, el stock de capital humano de Colombia es la sumatoria del capital humano de cada individuo.

$$CH = \sum_{i=16}^{62} CH(i) \quad (13)$$

La tabla A1 del anexo muestra los ingresos laborales de los n individuos de la muestra por edad, el tamaño de la muestra que ellos representan y el ingreso promedio. Con estos datos, y para evitar grandes fluctuaciones aleatorias, se suavizan las series de los pesos y los ingresos promedios a través de un promedio móvil, MV(3). La figura A1 muestra los valores suavizados de los ingresos promedios por edad.

A continuación, aplicando la ecuación (10) a los ingresos promedios suavizados se puede estimar el capital humano promedio de los individuos por edad para una tasa de descuento de 6,4%, el cual se muestra en la figura A2 y tabla A2 del anexo. Luego, mediante la ecuación (11), se obtiene el *benchmarked* del capital humano, es decir el promedio de sección cruzada, el cual es igual a $AvCH(h) = \$ 82565453,7$. Una vez obtenido este valor se toma como referencia y se combina con la estimación el capital humano de la ecuación (8) y (9) para obtener la distribución en pesos colombianos, con una media igual a $\$ 75705360,4$. Finalmente, mediante la ecuación (13), el stock de capital humano tiene un valor igual a 1,54 billones de pesos. Para observar que tan sensible es esta medida a las tasas de descuento se realizan los mismos cálculos para una tasa de descuento del 8,0%, lo cual arroja que el valor esperado promedio de ingresos es igual a $AvCH(h) = \$$

71557704,2 y el stock de capital humano igual a 1,33 billones de pesos. Lo cual representa una reducción del 13,3% y 13,6% respectivamente.

Ahora, como menciona Dagum et al. (2000) algunos factores, tales como cambios tecnológicos o niveles más altos de educación, entre otros, pueden hacer variar los niveles de productividad del capital humano por lo cual se hace necesario introducir una variable que capture estos elementos. De esta manera asumiendo un incremento de la productividad del capital humano a una tasa anual de τ , se tendría la siguiente ecuación:

$$h^*(x, \tau) = \sum_{t=0}^{62-x} y(x+t)p(x, x+t)(1+\tau)^t(1+r)^{-t} \quad (14)$$

Mientras que el capital humano promedio sería igual a:

$$AvCH(h^*) = \frac{\sum_{x=16}^{62} h^*(x, \tau)f(x)}{\sum_{x=16}^{62} f(x)} \quad (15)$$

Aplicando la ecuación (14) a los ingresos promedios suavizados, para dos diferentes tasas de descuento 6,4% y 8,0%, se obtiene el valor esperado de los ingresos promedios los cuales se muestran en las columnas 6 y 7 de la tabla A2 y en la figura A3. Finalmente, aplicando la ecuación (15) al valor esperado de los ingresos para las dos tasas de descuento se puede obtener el capital humano promedio correspondiente al ciclo de vida (corregido por incrementos en productividad) el cual es igual a $AvCH(h^*) = \$ 113515965,9$ para una tasa de descuento de 6,4% y $AvCH(h^*) = \$ 95575715,5$ para una tasa de 8,0%. El capital humano para estas dos tasas de descuento es igual a \$2,12 billones y \$1,78 billones respectivamente.

5 Conclusiones

Este documento estima el stock de capital humano para Colombia mediante la combinación de dos técnicas con el objetivo de superar algunas limitantes bien conocidas de los métodos tradicionales. La aplicación de la primera técnica, la cual es una técnica estadística que permite estimar variables latentes multidimensionales, permite dilucidar que los métodos tradicionales podrían estar sesgados por cuanto consideran al capital humano como una variable unidimensional y observable. El enfoque educativo, por ejemplo, considera a la educación formal como único elemento de relevancia para evaluar al capital humano, dejando de lado elementos importantes en su valoración.

En este estudio los resultados demuestran que no solamente la educación es importante en la formación y retornos de este stock, sino que variables relacionadas con las condiciones laborales y los antecedentes familiares forman también un importante componente en la estimación del capital humano. La medida estimada en este documento al ser más integral que las que se vienen utilizando habitualmente en las relaciones con el crecimiento económico podría proporcionar mejores resultados.

Los resultados de este estudio están basados en la ECH de Colombia, la cual podría ser usada junto con la encuesta de hogares para formar una serie de tiempo comprendida en el periodo 1984-2006, lo que permitiría observar tendencias y posteriormente hacer comparaciones, extendiendo la medición a otros países. Sin embargo, una de las limitaciones de este tipo de encuestas es que no permite tener en cuenta otros aspectos del capital humano como las habilidades innatas de los individuos o condiciones de salud que se cree tener un impacto importante en la formación de capital humano.

Referencias

- Barro, R. J. y Lee, J.-W. (1993). International comparisons of educational attainment. *Journal of Monetary Economics*, 32(3). 363-394.
- Bils, M. y Klenow, P. (2000) Does schooling cause growth? *American Economic review* 90(5): 160-1183.
- Chin, W. W. (1998). The partial least squares approach for structural equation modeling, en G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 236-295). London: Lawrence Erlbaum Associates.
- Cohen, D y Soto, M (2006). Growth and human capital: good data, good results. *Journal of Economic Growth*. Volume 12, Number 1, 51-76.
- Dagum, C. y Slottje, D.J. (2000) A new method to estimate the level and distribution of the household human capital with application. *Structural Change and Economic Dynamics* 11: 67-94.
- Fornell, C. y Larcker, D. (1981). Evaluating structural equation models with unobservable variables and measurement error, *Journal of Marketing Research*, 18:39-50.
- Gefen, D., Straub, D. W., y Boudreau, M. C. (2000). Structural equation modeling and regression: guidelines for research practice. *Communications of the Association for Information Systems*, 4(7): 178.
- Hanushek, E. A. y Kimko, D. D. (2000). Schooling, labor force quality, and the growth of nations. *American Economic Review*, 90(5):1184-1208.
- Kyriacou, G. (1991). Level and growth effects of human capital: a cross country study of the convergence hypothesis. *Economic Research Reports* 91-26, NewYork University.
- Le, T., Gibson, J. y Oxley, L. (2003) Cost and income based measures of human capital. *Journal of Economic Surveys* 17(3): 271-305.
- Mincer, J. (1974) *Schooling, Experience, and Earnings*. New York: NBER Press. ECD (2001). *The well-being of Nations: the Role of Human and Social Capital*, Paris.
- R Development Core Team (2009), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Tao, H. y Stinson, T. (1997). An alternative measure of human capital stock. *Development Center Bulletin* 97/01, University of Minnesota.
- Tenenhaus, M., Esposito, V., Chatelin, Y., y Lauro, C., (2005). PLS path modeling. *Computational Statistics and Data Analysis* 48: 159-205.
- V. Esposito, Vinzi, Trinchera, L. y Amato, S., (eds), (2010). *Handbook of Partial Least Squares*, Springer Handbook of Computational Statistics.
- Wobmann, L. (2003). Specifying human capital. *Journal of Economic Surveys*, 17(3):239-270.
- Wold, H. (1975). Modellling in complex situations soft information. En: *Third World Congress of Econometric Society*, Toronto, Canada.
- Wold, H. (1982). Soft modeling: the basic design and some extensions. En: Joreskog, K. G, Wold, H (Eds), *Systems Under Indirect Observation: Causality, Structure, Prediction*, vol. 2. North-Holland, Amsterdam: 1-54.

Anexos

Tabla A1: Ingreso laboral de los trabajadores por edad

Edad	Ingreso de los trabajadores	Peso	Ingresos promedio
16	1,7089E+11	63695	2682946,73
17	3,531E+11	125078	2823005,9
18	7,673E+11	252918	3033795,97
19	1,0546E+12	316610	3330801,85
20	1,3672E+12	392885	3479921,44
21	1,551E+12	416815	3721151,05
22	1,9328E+12	503046	3842287,05
23	1,8806E+12	431094	4362460,31
24	2,2318E+12	486204	4590242,61
25	2,4124E+12	496822	4855653,83
26	2,0445E+12	415424	4921455,3
27	2,4917E+12	447715	5565302,32
28	2,4024E+12	439784	5462572,63
29	2,225E+12	414527	5367630,53
30	3,1072E+12	579371	5363112,47
31	2,0987E+12	331899	6323248,9
32	2,6245E+12	466905	5621153,71
33	3,0034E+12	476023	6309417,81
34	2,6694E+12	413639	6453394,95
35	2,817E+12	457168	6161921,42
36	2,7575E+12	415813	6631636,97
37	2,8589E+12	418890	6824856,38
38	2,5967E+12	407088	6378829,53
39	2,4253E+12	365600	6633836,01
40	3,8138E+12	540603	7054796,29
41	2,0451E+12	280155	7300019
42	3,3709E+12	413875	8144825,7
43	1,7851E+12	285086	6261677,18
44	1,6747E+12	240608	6960104,07
45	2,4273E+12	316178	7677004,52
46	2,5877E+12	294919	8774242,84
47	2,4245E+12	249544	9715752,27
48	2,0345E+12	261225	7788300,93
49	2,2792E+12	260023	8765420,99
50	2,2432E+12	311256	7206964,45
51	1,1735E+12	186736	6284108,61
52	1,3549E+12	212074	6388593,33
53	1,1813E+12	177249	6664793,22
54	1,0801E+12	144613	7468650,17

Edad	Ingreso de los trabajadores	Peso	Ingresos promedio
55	1,2931E+12	165714	7803009,66
56	1,0816E+12	134569	8037291,56
57	6,6228E+11	89360	7411326,87
58	7,1525E+11	114942	6222729,26
59	5,9341E+11	80992	7326811,59
60	5,8148E+11	113886	5105829,96
61	3,6746E+11	47553	7727345,35
62	5,079E+11	77095	6588033,91

Figura A1: Ingresos promedio suavizados y observados por edad.

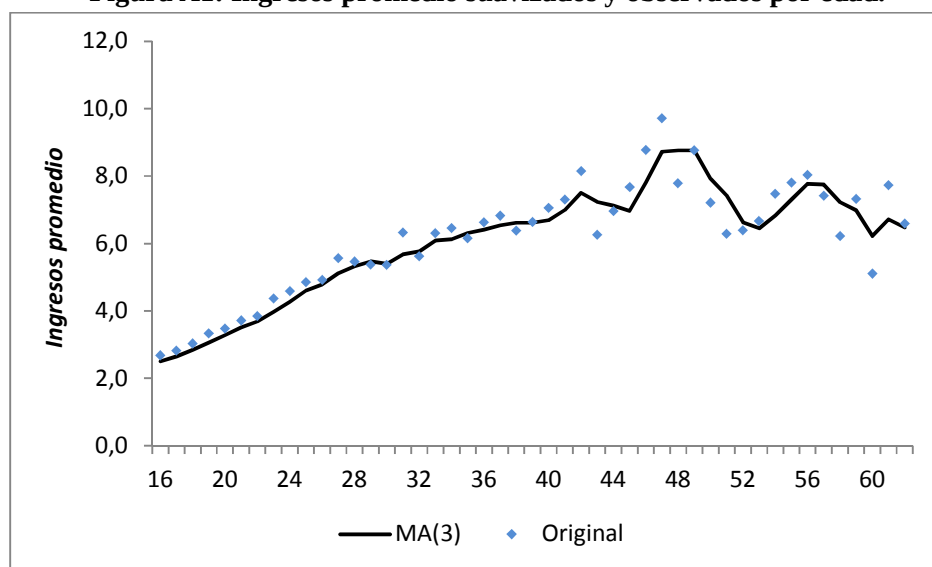


Tabla A2: Capital humano por edad (Calculados con una tasa de descuento de 6.4% y 8.0%)

Edad	Pesos	Ingresos promedios	$h(x, r = 0)$ sección cruzada		$h(x, r \neq 0)$ Ciclo de vida	
	Media móvil MA(3)	Media móvil MA(3)	6.4%	8.0%	6.4%	8.0%
x	f(x)	y(x)	6.4%	8.0%	6.4%	8.0%
16	63695,0	2498415,73	75101672,3	59579039,5	133650442,7	100684406,6
17	125078,0	2648800,49	77269164,2	61666663,1	134895416,9	102425633,9
18	147230,3	2846582,86	79416527,9	63760060,3	136045809,0	104106464,8
19	231535,3	3062534,57	81492410,3	65808874,9	137052633,9	105677422,3
20	320804,3	3281506,42	83473044,6	67790060,2	137895537,2	107117372,1
21	375436,7	3510624,78	85352025,3	69697891,2	138571893,0	108422879,6
22	437582,0	3681119,85	87109449,8	71512901,6	139066499,1	109577716,5
23	450318,3	3975299,47	88799409,4	73290466,8	139436106,4	110638467,2
24	473448,0	4264996,66	90287049,9	74895091,9	139554053,1	111476439,4
25	471373,3	4602785,58	91564153,7	76317743,6	139420861,2	112090107,4
26	466150,0	4789117,25	92567784,7	77493658,5	138985099,9	112425197,1

Edad	Pesos	Ingresos promedios	$h(x, r = 0)$ sección cruzada		$h(x, r \neq 0)$ Ciclo de vida	
	Media móvil MA(3)	Media móvil MA(3)				
x	$f(x)$	$y(x)$	6.4%	8.0%	6.4%	8.0%
27	453320,3	5114137,15	93439046,5	78564088,7	138393979,3	112628240,8
28	434307,7	5316443,42	94023135,1	79372062,5	137505218,1	112554174,3
29	434008,7	5465168,5	94431148,8	80028007,6	136438272,6	112321428,3
30	477894,0	5397771,88	94708353,1	80577146,4	135245169,3	111981815,0
31	441932,3	5684663,97	95082429,4	81250537	134098480,9	111710350,7
32	459391,7	5769171,7	97129077,7	81670975,1	135761393,7	111135411,8
33	424942,3	6084606,81	95190994,3	82034669,3	131037695,5	110452983,5
34	452189,0	6127988,82	94872289,1	82090109,1	129081217,3	109419140,9
35	448943,3	6308244,73	94487478,2	82103588,1	127027691,8	108299880,6
36	428873,3	6415651,12	93903559,5	81941041,8	124749656,4	106965007,7
37	430623,7	6539471,59	93169366,2	81650890,1	122300155,0	105465515,2
38	413930,3	6611774,29	92258026,6	81205411,2	119657503,8	103777259,1
39	397192,7	6612507,3	91212357,9	80647147,7	116868556,1	101945234,7
40	437763,7	6689153,94	90098995,8	80043440,9	114002057,3	100037291,1
41	395452,7	6996217,1	88869737,4	79346125,2	111013490,0	98007318,0
42	411544,3	7499880,33	87240676,3	78267065	107631567,4	95578297,6
43	326372,0	7235507,3	84980618,5	76567022,3	103645292,2	92530079,4
44	313189,7	7122202,32	82852401,4	75011618,2	99814688,3	89624555,5
45	280624,0	6966261,92	80706473,9	73452059,3	95991960,2	86713682,2
46	283901,7	7803783,81	78638997,5	71986771	92272116,1	83893657,6
47	286880,3	8722333,21	75569649,2	69521630,3	87607162,8	80101997,1
48	268562,7	8759432,01	71350182,7	65891257,3	81886803,7	75211603,8
49	256930,7	8756491,4	66822152,7	61931357,7	75965647,0	70072380,9
50	277501,3	7920228,79	62007381,8	57657765,1	69878319,2	64714579,8
51	252671,7	7418831,35	57863641,9	54036185,8	64566118,4	60089103,7
52	236688,7	6626555,46	53968253,7	50646153,8	59602511,6	55773088,9
53	192019,7	6445831,72	50635041,1	47808602,4	55287643,7	52075067,4
54	177978,7	6840678,91	47273607,3	44931934,8	51035348,3	48408871,3
55	162525,3	7312151,02	43292623,6	41414634,2	46277021,6	44197383,2
56	148298,7	7769650,47	38709613,3	37263484,2	40983378,4	39402008,9
57	129881,0	7750542,7	33373188,9	32313221,9	35021991,2	33877523,1
58	112957,0	7223782,56	27714449,8	26986444,1	28834418,2	28058600,1
59	95098,0	6986955,91	22223307,4	21771246,3	22910913,7	22435591,8
60	103273,3	6218456,94	16618905,5	16380587,9	16977024,0	16730020,6
61	80810,3	6719995,63	11658281,8	11576211,3	11780156,6	11696280,5
62	79511,3	6473736,41	5894304,34	5894304,34	5894304,3	5894304,3

Figura A2: Valor esperado de los ingresos promedios

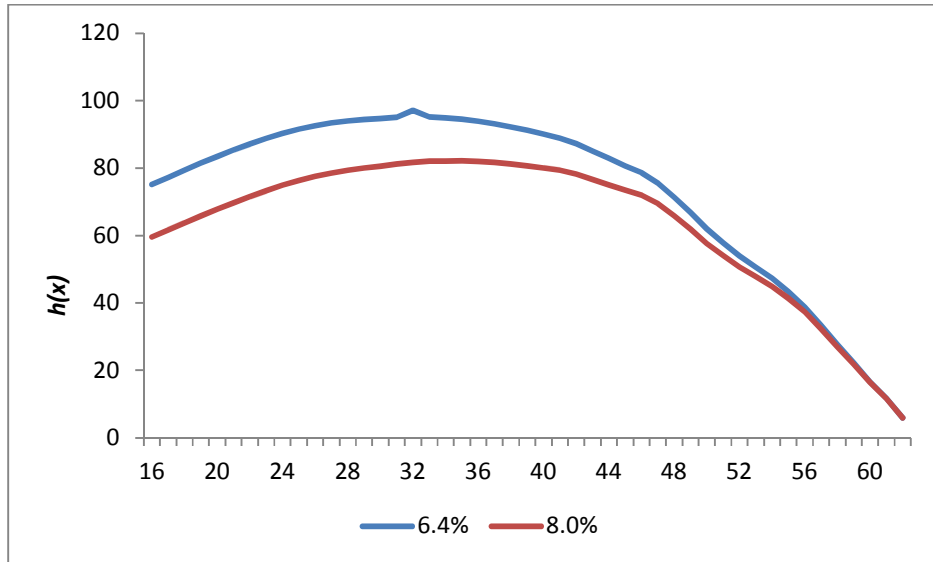


Figura A3: Valor esperado de los ingresos promedios con incrementos de productividad

