

INIE
Instituto Nacional de Investigaciones Económicas
Ciudad de la Habana, Cuba

**Influencia de variables
socio-económicas en el proceso
educativo.**

Autora: Susset Rosales

Julio 2006

Introducción

La educación constituye un factor central y estratégico para el desarrollo socioeconómico de cualquier país, siendo uno de los componentes básicos de la inversión en capital humano. Es una herramienta poderosa para impulsar el desarrollo humano sostenible, aumentar la productividad y la competitividad de los países en el mundo moderno y constituye uno de los elementos más importantes en la formación y calificación de la fuerza de trabajo. La educación es un derecho humano fundamental que adquiere hoy día cada vez más valor, al devenir en un factor de indiscutible peso para el mejoramiento de la calidad de vida, el logro de una mayor equidad en la distribución de los recursos, la reducción de la pobreza, etc.

El objetivo principal de la educación en nuestro país, es elevar el desarrollo humano de todos los cubanos, al propiciar una cultura general e integral desde la infancia y promover la asimilación de conocimientos sólidos y profundos, resaltándose la educación para la vida a partir de valores acordes con el modelo social cubano, prestando atención a la diversidad de los alumnos e integrando a los propios estudiantes, al personal de las escuelas, a la familia, al personal de salud, a las instituciones y organizaciones y a los trabajadores sociales (Álvarez, 2004).

El sistema educativo cubano tiene como misión proporcionar con equidad una educación de calidad para todos, basado en la formación de profesionales integrales preparados para la vida y para las exigencias planteadas por el desarrollo económico y social del país. Lograr y mantener una máxima cobertura educativa, elevar la retención escolar y la calidad de las prestaciones, no sólo constituyen las tradicionales funciones del Estado, sino que hoy también reviste características estratégicas, debido a la importancia de lograr una mayor formación integral de los niños y jóvenes, insertado al amplio espectro de programas educativos que se llevan a cabo.

La educación es un proceso de decisiones secuenciales. Si se parte de que todos los individuos asisten a la escuela primaria, los próximos niveles de decisión serían: completar la primaria, continuar de la primaria a la secundaria básica, completar la secundaria, continuar al nivel medio superior, completar la media superior y asistir y completar la educación terciaria. Cada uno de estos niveles de decisión puede analizarse en forma separada.

El objetivo principal de este estudio es identificar y evaluar la importancia de las variables socio-económicas que inciden y explican el comportamiento de tres variables educativas fundamentales: el acceso de los niños a la primaria los 6 años de edad, la continuidad de estudios a los 15 años y el completamiento en tiempo de los estudios a los 18 años. Se pretende identificar los principales determinantes¹ y la relación funcional existente entre las variables del sistema, así como cuantificar e interpretar los efectos de cada variable sobre la probabilidad de satisfacer cada una de las decisiones. Los resultados de este estudio pueden

¹ Variables socio-económicas que influyen en las diferentes decisiones que se van a analizar a lo largo de este trabajo.

resultar particularmente útiles a la hora de trazar las estrategias que contribuyan a perfeccionar la política educacional en el país.

El estudio realizado hace especial énfasis en la educación primaria y secundaria, lo que está estrechamente relacionado con el objetivo de la meta de desarrollo del milenio de lograr una enseñanza primaria universal. Como este objetivo está cerca de ser alcanzado en la mayoría de los países latinoamericanos -en el caso de nuestro país la matrícula en esta enseñanza representa alrededor del 100% de la población en las edades establecidas para este nivel de 6 a 11 años- el análisis también cubrirá niveles más altos de educación. Estos niveles educativos pueden jugar un rol decisivo en el crecimiento, pues son los que tienen el mayor impacto en la calidad de vida, en la salud de la población, en la capacitación del trabajo, en la distribución del ingreso, etc.

El trabajo se ha estructurado de la siguiente manera. En la sección 1 se presenta una breve reseña teórica de los estudios realizados sobre los determinantes del funcionamiento educativo en general. En la sección 2 se presentan algunos de los métodos empíricos propuestos para la determinación de estos factores. En la sección 3 se describe la fuente de información y las variables utilizadas. La sección 4 contiene las estimaciones de los determinantes asociados a las distintas variables dependientes empleadas y los principales resultados que se derivan del análisis. El estudio concluye en la sección 5 con los comentarios finales y recomendaciones.

Sección 1: Marco Teórico y Evidencia empírica del análisis de los determinantes.

La búsqueda de mejores métodos para evaluar las políticas educativas y dentro de ellas los gastos que requiere este sector, ha generado un extenso desarrollo teórico sobre la educación y sus determinantes. La literatura económica sobre el tema ha venido enriqueciéndose con el paso de los años y en la actualidad, se pueden encontrar investigaciones que analizan los determinantes para muchas de las variables educativas importantes, como por ejemplo: la matrícula, la retención, la deserción, el acceso, entre otras.

En esta sección se pretende describir brevemente los diferentes estudios internacionales o regionales que han explorado los determinantes de la educación, destacándose las metodologías empleadas y los principales resultados obtenidos. Estos estudios constituyen una fuente de obligada referencia para un mejor entendimiento del tema.

1.1 Evidencia internacional

La mayoría de los estudios que analizan los determinantes de la educación, parten de un modelo convencional de capital humano, donde la educación es vista como una inversión que depende de los costos y beneficios asociados con la matriculación Ben-Porath (1967) y Heckman (1976). Esta formulación teórica inicial ha sido extendida utilizando modelos dinámicos de ingresos y la importancia de la inversión en capital humano y se han formulado modelos extendidos de capital humano para entender la inversión del gobierno sobre la calidad de la educación y su impacto en los ingresos futuros.

Un primer estudio que aborda esta temática se basa en la metodología de estimación de los determinantes en la enseñanza a través de un modelo costo-efectividad. El modelo básico propuesto por Gertler y Glewwe (1990) es un análisis microeconómico que utiliza la encuesta de hogares y datos de entrada educacional para cuantificar el impacto de cambios en algunas variables sobre los resultados escolares. Este tipo de modelo analiza los factores principales que determinan la matrícula escolar, haciendo distinción entre variables por el lado de la demanda (ingresos y costos asociados), variables por el lado de la oferta (calidad de las entradas escolares) y variables institucionales tales como: la autonomía de las escuelas, las citas descentralizadas de los profesores, etc. Esta última dimensión para ellos resulta importante con el propósito de lograr un plan de políticas de intervención más rentable en la educación.

Dos años más tarde, estos mismos autores Gertler y Glewwe (1992) analizaron estadísticamente a través de un modelo logit multinomial anidado, la opción de asistir a la escuela contra la opción de no hacerlo. Entre los resultados más destacables de este nuevo modelo se encuentran: que los padres en las zonas rurales de Perú asignan mayor valor a la asistencia a la escuela de los niños hombres que la de las niñas y que la demanda por educación aumenta con el nivel de gasto del hogar (elasticidad ingreso positiva). También encontraron que los padres demandan menos educación para las hijas; que la educación de

estos tiene un efecto menor en la demanda por educación (en particular comparado con el de la madre) y que las variables que reflejan calidad (tiempo de viaje y falta de profesores) tienen un efecto negativo sobre la asistencia.

Otro estudio interesante es el realizado Skyt (1998), donde se analiza la asistencia a la escuela y la participación en la fuerza de trabajo de menores de entre 7 y 14 años en Zambia. Para ello emplea un modelo probit bivariado que permite analizar el efecto de las variables consideradas en las decisiones: solamente estudiar, solamente trabajar, trabajar y estudiar y no trabajar ni estudiar. Es esta variedad de análisis de las diferentes alternativas de decisión posibles, lo que proporciona mayor interés a este modelo que ha servido de guía para muchos otros autores. El modelo de Skyt es estimado en dos etapas por la existencia de regresores endógenos: primero se estima por Mínimos Cuadrados Ordinarios cada uno de los regresores endógenos con variables instrumentales y luego, se realiza un probit, incluyendo las variables endógenas y los residuos de mínimos cuadrados de la etapa anterior.

Entre los resultados más destacables se encuentran: las variables que miden calidad en general tienen el signo opuesto al esperado; el efecto del ingreso es muy pequeño y finalmente, que las variables más importantes son la tradición de la familia y el acceso al colegio.

Por su parte, Beyer (1998) con datos de Chile estima dos logits, uno para explicar la probabilidad de estar desempleado y un segundo para explicar la probabilidad que un joven de 15 a 24 esté estudiando. En este último logit incluye las variables edad, sexo, educación de los padres, situación familiar (un padre o dos), localización (urbano o rural), situación civil (casado o no) y el ingreso per cápita del hogar. El centro de atención del trabajo es la desocupación juvenil, y este logit cumple el propósito de proveer evidencia a una posible solución de la misma, vías subsidios a la educación (reteniendo a los jóvenes en la escuela). El trabajo no cuenta con un marco teórico formal, y atribuye el retorno de los egresados a sus diferencias con los no egresados (no considera la posibilidad de que haya un sesgo de selección).

Los resultados obtenidos proveen evidencia preliminar de que mientras más bajo sea el ingreso de la familia, más alta es la probabilidad de que un joven deje de estudiar. Por esto concluye que no deben descartarse políticas sociales que reduzcan la deserción en dichos sectores (como el subsidio educacional).

Eckstein y Wolpin (1999) estiman un modelo de decisión secuencial de asistencia a la enseñanza media y trabajo. Se supone que los jóvenes eligen entre combinaciones de trabajo-escuela en orden a maximizar el valor presente de la utilidad esperada en cada período. Los resultados obtenidos indican que los que desertan tienen rasgos distintos a los que se gradúan: menor habilidad para el colegio y/o menor motivación, menores expectativas de los beneficios de graduarse, ventajas comparativas en trabajos que realizan los “no graduados”, mayor valor del ocio, menor valor de consumo de asistir al colegio. Se constata además que las legislaciones que prohíben el trabajo tienen poco impacto en las tasas de graduación; y que para disminuir la deserción en educación media hay que afectar

la habilidad y la motivación, y también los rasgos iniciales, como el valor esperado de graduarse.

Para el caso argentino, son conocidos varios trabajos que analizan los determinantes en la educación. El primero de ellos realizado por Corbacho (1999) para la provincia de Mendoza, analiza los factores que afectan la asistencia a establecimientos públicos en los cuatro niveles educativos, estimando para ello los efectos de las características familiares sobre la educación mediante un modelo secuencial y, posteriormente, en el 2001 aplica un modelo ordinal generalizado. Por su parte, Cossa (2000) para la región metropolitana, determina cuáles son las variables que tienen mayor incidencia sobre distintos niveles de educación con un modelo de transiciones escolares.

Un tercer trabajo mucho más completo y reciente es el realizado por Bertranou (2001). La autora investiga sobre la base de encuestas de desarrollo social 1997, los factores que determinan la educación adquirida por una persona como un proceso secuencial de decisiones, cuyas categorías son: primaria incompleta, primaria completa, secundaria incompleta, secundaria completa y terciaria. Para ello construye un modelo logit secuencial para estimar las probabilidades de completar la primaria y secundaria y asistir a la secundaria y a la terciaria. Los resultados indican que las características socioeconómicas determinan significativamente la educación de una persona y tienen importantes efectos sobre las probabilidades de satisfacer los distintos niveles educativos.

Específicamente, obtiene que la edad tiene un efecto significativo positivo y negativo sobre la probabilidad de completar el nivel primario y en la decisión de asistir al nivel secundario respectivamente. Por otro lado, la variable sexo e ingreso total de la familia son significativas y positivas en los cuatro modelos estimados, aunque el ingreso va perdiendo un poco de significatividad en los últimos dos niveles.

La educación del padre y de la madre es muy relevante en la determinación de la educación de sus hijos, mientras que la cantidad de hermanos que habitan en el hogar arroja valores significativos pero negativos. Con respecto a las dummies provinciales, estas sólo ejercen su influencia sobre el nivel primario y secundario, no siendo de la misma manera sobre la asistencia al nivel terciario.

Otros estudios se concentran en un nivel educativo en particular. Sosa y Marchionni (1999) examinan el rol de la estructura de hogar y del género en la decisión de enviar a los hijos a la escuela secundaria y Ripani (2000) analiza los factores que afectan la decisión de asistir al nivel superior para varios aglomerados urbanos de Argentina.

Ponce, Bedi y Vos (2004) aplicaron un modelo de determinantes de enseñanza para Ecuador, identificando la importancia de factores por el lado de la demanda y por el lado de la oferta y cuantificando el impacto de cambios en tales variables sobre los resultados en la escuela. Analizaron los determinantes de la matriculación neta para la educación básica y secundaria encontrando que la importancia relativa de los determinantes de la enseñanza realmente varían sustancialmente entre la educación básica y la educación secundaria, así como la conducta decisión-realización difiere para los hogares urbanos y rurales y para los pobres y los menos pobres.

Se estima el modelo a través de una especificación logit multinomial y se ofrecen las posibles implicaciones de políticas educacionales de estos resultados. Se concluye que para la educación primaria los costos directos e indirectos de educación tienen un efecto débil pero negativo en la matriculación, por su parte el nivel educativo de la madre tiene un efecto significativo y positivo. El género para este caso, apenas influye en la decisión de la instrucción primaria y la porción de maestros especializados y la preparación de estos tiene una influencia importante en la matriculación en las áreas urbanas.

Sin embargo, para la educación secundaria los costos indirectos y directos tienen un efecto negativo pero fuerte sobre la matrícula particularmente en los hogares pobres, mientras que el nivel educativo de los padres y el género son variables explicativas muy importantes de la matrícula en las áreas urbanas y rurales. La porción de maestros por aula, el número de profesores con grado universitario, así como la cantidad de alumnos por aula tienen una significativa influencia en las decisiones de matriculación de los niños en las escuelas sobre todo en las áreas urbanas.

El Cuadro 1 presenta un resumen del efecto de las principales variables consideradas en los diferentes trabajos. Puede observarse que en general los resultados empíricos validan las suposiciones sobre las variables relacionadas con la permanencia en la escuela como ingreso, sexo, educación de los padres y otras variables demográficas de la familia. Sin embargo, su impacto no es sistemático.

Cuadro No 1

Resumen de los principales resultados sobre asistencia a la escuela

Variable	Gertler y Glewwe	Skyt	Beyer	Bertranau	Ponce, Bedi y Vos.
Ingresos * Costos* *	Positiva*	Positivo y pequeño*	Positivo e importante*	Positivo. Significativo para los dos primeros niveles y luego pierde significatividad*	Negativo. Débil en la educación primaria y fuerte en la secundaria**
Edad	-	-	Negativo	Negativo e importante	-
Sexo	Importante		No significativo	Importante	No significativo primaria y muy significativo en la secundaria
Educación de los padres	Positivo y pequeño para los dos	-	Positivo pero no se separa entre madre y padre	Positivo e importantes en todas las decisiones	Positivo e importante en la primaria y secundaria
VARIABLES demográficas familiar	-	-	Significativas	Significativas	-
VARIABLES reflejan calidad	Negativo	-	-	-	Positivo e importante (porción de maestros especializados)

Como puede constatarse de la discusión bibliográfica, surgen varios hechos aceptados y numerosas dudas respecto a los determinantes de la decisión de asistir a la escuela:

- En la mayoría de los trabajos consultados se encuentra un efecto positivo e importante del ingreso en la decisión de asistir a la escuela, mientras que en otros, este efecto es pequeño. En cualquiera de los casos, el ingreso resulta ser una de las variables principales que se tienen en cuenta cuando se realizan estudios de este tipo, relacionados con la educación y la salud.
- Respecto a la importancia del sexo y la edad del niño también se encuentran resultados muy diferentes, en particular en términos de magnitud. Las variables demográficas del hogar (estado civil, situación familiar, tamaño del núcleo, número de hermanos, etc) inciden según algunos estudios y no inciden según otros.
- Los trabajos analizados, validan en general la presunción de que la educación de los padres es una de las variables más importante en la decisión de asistir al influir de forma significativa en la mayoría de los trabajos realizados sobre este tema.

Para el caso cubano, no se tiene conocimiento de ningún trabajo precedente orientado a estimar y analizar los determinantes de la educación, por lo que el presente estudio constituye un primer intento de aproximación a lo que sería la cuantificación de estos impactos sobre variables educativas claves con su consecuente interpretación e implicación para la toma de decisiones de las autoridades.

Sección 2: Métodos empíricos propuestos para la determinación de estos factores.

2.1 Metodología internacionalmente utilizada

En esta sección se aborda la conceptualización básica de algunos aspectos teóricos fundamentales que son necesarios a la hora de obtener y cuantificar los determinantes. Al tiempo que se pretende describir la metodología que se utiliza para el cálculo de los determinantes en el caso de nuestro país.

El marco analítico que ha sido utilizado en la literatura sobre el tema, responde a modelos de elección discreta, los cuales permiten la modelación de variables cualitativas. Para la utilización de este tipo de variables (que miden cualidades) es necesario la codificación como paso previo a la modelización proceso por el cual las alternativas de las variables se transforman en códigos o valores cuánticos, susceptibles de ser modelizados mediante el empleo de técnicas econométricas.

Existe una amplia tipología de modelos de elección discreta. En concreto, según el número de alternativas incluidas en la variable endógena, se distinguen los modelos de respuesta dicotómica o binaria frente a los denominados modelos de respuesta o elección múltiple. Según la función utilizada para la estimación de la probabilidad existe el modelo de probabilidad lineal, el modelo logit y el modelo probit. Según que las alternativas de la variable endógena sean excluyentes o incorporen información ordinal se distingue entre los modelos con datos no ordenados y los modelos con datos ordenados. Dentro de los primeros, según que los regresores hagan referencia a aspectos específicos de la muestra o de las alternativas entre las que se ha de elegir, se distingue entre los modelos multinomiales y los condicionales.

En la literatura existen dos enfoques para la interpretación estructural de los modelos de elección discreta. El primero hace referencia a la modelización de una variable latente a través de una función índice, que trata de modelizar una variable inobservable o latente. El segundo de los enfoques permite interpretar los modelos de elección discreta bajo la teoría de la utilidad aleatoria, de tal manera que la alternativa seleccionada en cada caso será aquella que maximice la utilidad esperada (Medina, 2003).

Bajo este último enfoque un individuo debe adoptar una decisión que le permita elegir entre dos alternativas excluyentes, la 1 o la 0, lo que hará maximizando la utilidad esperada que le proporciona cada una de las alternativas posibles sobre las que tiene que decidir. Es decir, el individuo i -ésimo elegirá una de las dos alternativas dependiendo de que la utilidad que le proporciona dicha decisión sea superior a la que le proporciona su complementaria.

La metodología básica internacionalmente utilizada para el análisis de los determinantes se propone en el estudio de Gertler y Glewwe (1990), y más tarde en los trabajos de Ponce, Bedi y Vos (2004) se puede consultar una descripción más detallada de los resultados cuantitativos y procedimientos de estimación.

Los modelos empíricos iniciales que subyacen detrás de este tipo de estudio son los análisis costo-efectividad o eficiencia de la educación. Estos son elaborados con el propósito de hacer proyecciones presupuestarias de acuerdo con los costos y beneficios resultantes de si el niño asiste a la escuela o no. Pero el interés por estudiar esta temática ha llevado a extender este horizonte e incluso se ha logrado vincularlos con el logro de las metas de desarrollo del milenio en educación.

Actualmente hay estudios que están interesados en cuantificar cuánto el presupuesto de educación necesitaría aumentar o cómo hacer ese gasto más eficiente para alcanzar una determinada meta educativa, siendo necesario para todos ellos el análisis de los determinantes de la enseñanza.

Los estudios que se han realizado sobre los determinantes en la educación recurren al segundo enfoque para una mejor interpretación de este tipo de modelos. La especificación empírica del modelo bajo esta teoría, parte del supuesto de que los hogares maximizan su función de utilidad condicionada a la matrícula escolar, $U_1 = U(b, c_1)$ donde U: utilidad hogar condicionada a la matrícula escolar, b: vector de beneficios asociados a la asistencia a la escuela y c: consumo de los hogares.

A su vez los principales beneficios asociados a la enseñanza (b) quedarían definidos por la expresión: $b = B(h, w, z)$ donde h: es un vector de las características individuales de los niños, w: es un vector de las características de los hogares y z: es un vector de las características de la escuela (incluye calidad de las entradas a la escuela y los aspectos institucionales).

El hogar maximiza su utilidad asociado a la siguiente restricción presupuestaria $y = C_1 + p$ donde y: es el ingreso del hogar y p: representa el total de costos asociados con la matrícula.

Finalmente la función de utilidad puede ser estimada a través de un modelo probabilístico lineal porque es más fácil de interpretar y hay un efecto precio directo. En este caso la función de utilidad toma la siguiente forma: $U_1 = \beta_1 b + \beta_2 c_1 + \varepsilon_1$ donde β : son los coeficientes a ser estimados y ε_1 son los términos de error normalmente distribuidos (se asume que tiene media cero).

La oportunidad de asistir a la escuela puede ser expresada como una función del individuo socio-demográfica y características de los hogares, la calidad de las entradas a la escuela y los costos indirectos y directos de matriculación. En términos de un modelo de probabilidad lineal estas relaciones funcionales pueden ser escritas:

$$\Pr[a = 1] = \Pr[\beta_1 + \beta(h, w, z) - \beta_2 p + \varepsilon_a > 0]$$

donde a es un binomio variable que toma valor 1 si matricula y 0 si no se matricula².

La estimación de los modelos probabilísticos lineales por mínimos cuadrados plantea una serie de inconvenientes que han llevado a la búsqueda de otros modelos alternativos que permitan estimaciones más fiables de las variables dicotómicas. Entre las principales limitaciones de este tipo de modelos se encuentran:

- El modelo no garantiza que los valores estimados se encuentren en el intervalo [0-1], lo cual carece de sentido cuando lo que se pretende es estimar una probabilidad. Esto es uno de los problemas más importantes.
- El término de error no se distribuye normalmente. Dado los valores que toma el término de perturbación aleatoria, no se puede asegurar que ésta se distribuya normalmente al tratarse de una distribución binaria o dicotómica. Esta ausencia de normalidad imposibilita el uso de los estadísticos habituales utilizados para el contraste de hipótesis.
- El término de perturbación aleatoria es heterocedástico, es decir, la varianza de la perturbación aleatoria no es constante y por tanto la estimación del modelo mediante el método de mínimos cuadrados generalizados obtiene unos estimadores de los coeficientes de regresión con varianza no mínima, es decir, no eficientes.

Con el propósito de solucionar los problemas que plantean este tipo de modelos de probabilidad lineal se realiza la modelación a través de funciones no lineales que permitan acotar el rango de la estimación. Las funciones más comúnmente utilizadas han sido la función logística, que ha dado lugar a la modelación logit y la función de distribución de la normal tipificada que ha dado lugar a la modelización probit.

De manera general los estudios recientes sobre esta temática, realizan las estimaciones sobre la base de los modelos logit fundamentalmente al ser estos más fáciles de estimar. En realidad, actualmente ambos modelos son igualmente sencillos de estimar y la mayoría de los programas econométricos traen ambos ya incorporados.

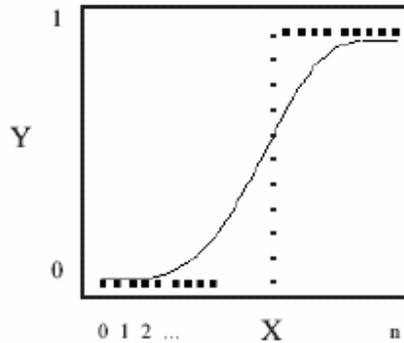
2.2 Metodología utilizada en la estimación de los determinantes de la educación para Cuba.

Para el caso de Cuba y con la finalidad de determinar los factores socio-económicos que pudieran estar incidiendo en el comportamiento de algunas variables educacionales claves, se ha utilizado un modelo de regresión logístico, de especial utilidad cuando la variable dependiente puede adoptar sólo dos valores.

² Ver Anexo No. 1 en el cual se describe con más detalle el modelo de Gertler and Glewwe (1990).

El modelo logístico tiene la forma concreta de una curva logística y para estimar un modelo se ajusta esta curva a los datos de las variables dependientes. Así, el gráfico muestra un buen ajuste de los datos a la curva logística. Los valores de Y representan los valores de la variable dependiente para cada valor de la variable independiente X .

Gráfico No 1
Ejemplo de una relación bien definida entre las observaciones de (X, Y) y una curva logística.



El objetivo de este tipo de modelos reside en explicar el comportamiento de una variable dicotómica dependiente (Y) en función de un grupo de variables explicativas (*el vector X*) con el objetivo de hallar la probabilidad de ocurrencia de algún evento o no.

En este trabajo y siguiendo la metodología teórica básica, el modelo logit propuesto descansa sobre el análisis costo-beneficio, en el cual el nivel de inversión de cada individuo proviene de un proceso de optimización en el que se igualan los beneficios y costos marginales de educarse. Son tres las variables dependientes que serán analizadas en este estudio, razón por la cual se estimarán tres modelos logit diferentes, condicionados al conjunto de factores que determinan la decisión. La especificación de la función de distribución logística en cada caso es la siguiente:

$$P_i = F(z_i) = \frac{1}{1 + e^{-z_i}}$$

donde P : es la probabilidad de que el acontecimiento objeto de estudio ocurra y z : es la combinación lineal: $Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$ donde los β : son los coeficientes a estimar, X : son las variables independientes y e : es la base del logaritmo natural, aproximadamente 2,718.

Si la probabilidad de que el suceso ocurra es P , entonces la probabilidad de que no ocurra es $(1-P)$, en consecuencia, la razón entre ambas será igual a:

$$\frac{P}{1-P} = \frac{1 + e^z}{1 + e^{-z}} = e^z = \exp(Z)$$

Al cociente entre ambas probabilidades se le denomina *ratio odds*. Su interpretación es la ventaja o preferencia de la opción 1 frente a la 0, es decir, el número de veces que es más probable que ocurra el fenómeno frente a que no ocurra.

Desde el punto de vista analítico y tomando logaritmo neperiano del ratio odds, se linealiza la ecuación del modelo Logit, respetando el objetivo de que los valores estimados se encuentren dentro del rango [0-1], obteniéndose la expresión:

$$\text{Ln} \left(\frac{P}{1-P} \right) = \text{Ln} (e^z) = \text{Ln} (e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

La estimación de modelos de determinantes de la enseñanza presenta un problema de complejo tratamiento relacionado con la posibilidad de distinguir con precisión el efecto de las distintas variables causales. En este sentido y a partir de esa función, se pueden obtener diferentes interpretaciones de los coeficientes y cocientes, estos son:

- Los coeficientes estimados (β) son las medidas de los cambios en $\ln[P/(1-P)]$ causado por el cambio unitario de X. La interpretación del signo del coeficiente es intuitiva e indica la dirección en que se mueve la probabilidad: un coeficiente positivo aumenta la probabilidad de éxito y un coeficiente negativo la disminuye.
- $\text{Exp}(\beta)$ indica el cambio en $(P/(1-P))$, causado por un cambio unitario de X
- $\beta \cdot P(1-P)$ indica el cambio en P (probabilidad de ocurrencia) causado por el cambio unitario de X (elasticidad).

Los coeficientes de un modelo logit no son fáciles de interpretar a simple vista más allá de su signo, pues están reflejando el efecto de cambio en la variable $\ln[P/(1-P)]$. En este tipo de modelos, los signos de los coeficientes permiten observar la relación entre las variables explicativas y la variable dependiente, no siendo sin embargo, un reflejo directo de las elasticidades o efectos marginales asociadas a dicha relación.

Para tener una medida de los efectos de cada variable sobre la probabilidad de satisfacer el evento son reportados los efectos marginales. Estos se definen como la variación de la probabilidad de que la variable dependiente presente un determinado suceso $P(Y=1)$, debido a un incremento del valor de un determinado regresor (X_{ki}), bajo el supuesto de que los valores del resto de los regresores del modelo se mantienen constantes. En otras palabras, es la probabilidad de éxito de que el evento ocurra ante un cambio marginal en la variable independiente.

En la estimación de los coeficientes del modelo ha sido utilizado el método de máxima verosimilitud ya que al tratarse de un modelo no lineal no se puede utilizar el método de mínimos cuadrados ordinarios. Este método se basa, como su nombre lo indica en la función de verosimilitud de la muestra, definiéndose dicha función como la probabilidad de que se den las observaciones muestrales, dependiendo lógicamente de los parámetros poblacionales.

El método elige como estimadores de los coeficientes β a aquellos maximizan la función de verosimilitud, es decir, los estimadores máximo-verosímiles son aquellos valores para los cuales la función de densidad conjunta alcanza un máximo. Los estimadores obtenidos con la utilización de este modelo son consistentes (insesgados y asintóticamente eficientes).

Debido a que se trabaja con un sistema de ecuaciones no lineales, es necesario la aplicación de un método iterativo o algoritmo de optimización que permita la convergencia en los estimadores³.

Para comprobar si los modelos estimados por máxima verosimilitud son los adecuados se llevan a cabo la contrastación y validación de los mismos mediante la aplicación de una serie de pruebas de hipótesis que son comentadas a continuación:

1. Significatividad de los coeficientes estimados

1.1 Test de Wald

Se emplea con el objetivo de comprobar la significatividad estadística de los parámetros estimados, se testa la hipótesis nula de que los parámetros sean igual a 0, contra la alternativa de que no lo sean⁴. Con el propósito de testar la hipótesis nula se calcula la razón entre los valores estimados de los parámetros⁵ y su error típico.

$$Wald = \frac{b}{SE_b}$$

Esta razón resultante debería seguir una distribución asintóticamente normal, por lo que el valor obtenido se compara con una distribución normal estandarizada. El contraste se realiza a través del nivel de significación asociado a dicho estadístico de tal manera que si el nivel de significación es mayor a 0,05 la variable no resulta significativa estadísticamente.

2. Medidas de bondad de ajuste del modelo

2.1 El Test de Ratio de Probabilidad o Test global de parámetros (Chi Cuadrado)

El estadístico chi-cuadrado es una medida de bondad de ajuste de los valores observados y los valores esperados. La más grande diferencia o desviación de los valores observados de los valores esperados, conllevará a un más pobre ajuste del modelo.

³ Medina, Eva (dic 2003).

⁴ La hipótesis nula $H_0 = \beta_1 = \beta_2 = \beta_k = 0$ contra la alternativa $H_0 \neq \beta_1 \neq \beta_2 \neq \beta_k \neq 0$.

⁵ Generalizando se utiliza como regla común que parámetros con valores superiores a 1,96, en valores absolutos, pueden considerarse significativos a un nivel de significación de 0,05.

Las hipótesis a contrastar son las siguientes: $H_0: B=0$ el modelo no es significativo contra $H_1: B \neq 0$ el modelo es significativo. Para el caso de la regresión logística y en lugar de usar la desviación (-2LL) para juzgar el ajuste global del modelo, se utiliza este estadístico Chi-Cuadrado:

$$G = \chi^2 = -2 \ln \left(\frac{L_{null}}{L_k} \right)$$

Si la probabilidad asociada al estadístico chi-cuadrado es menor que 0.05, se rechaza la hipótesis nula siendo de esta manera el modelo global estadísticamente significativo.

Para medir la bondad del ajuste también se utilizan otros estadísticos, los cuales tienen cierta semejanza con el coeficiente de determinación calculado en la estimación lineal. Son considerados tres tipos de estadísticos:

2.2 Mc-fadden o Ratio de verosimilitud o pseudo R^2

Este estadístico compara el valor de la función de verosimilitud de dos modelos: uno corresponde al modelo estimado que incluye todas las variables explicativas (modelo completo L) y el otro sería el modelo cuya única variable es la constante (modelo restringido $L(0)$).

$$ICV = 1 - \frac{\log L}{\log L(0)}$$

El ratio calculado tendrá valores comprendidos entre 0 y 1 de forma que:

- Valores próximos a 0 se obtendrán cuando $L(0)$ sea muy parecido a L , situación en la que nos encontraremos cuando las variables incluidas en el modelo sean poco significativas, por lo que en este caso la capacidad explicativa del modelo será muy reducida.
- Cuanto mayor sea la capacidad explicativa del modelo, mayor será el valor de L sobre el valor de $L(0)$, y más se aproximará el ratio de verosimilitud calculado al valor 1.

Este valor tiende a ser más pequeño que el R^2 y los valores de 0.2 a 0.4 son considerados muy satisfactorios.

2.3 Cox and Snell

Este estadístico también está basado en el log likelihood, pero tiene en cuenta el tamaño de la muestra. Pero al igual que el anterior no puede alcanzar un máximo de 1.

$$R_{CS}^2 = 1 - \exp\left[-\frac{2}{n}[LL(B) - LL(0)]\right]$$

2.4 Nagelkerke

Esta medida ajusta la medida de Cox and Snell para que el valor máximo de 1 sea alcanzado. El Cox & Snell puede interpretarse como R^2 en una regresión múltiple, pero no puede alcanzar un valor máximo de 1. El Nagelkerke R^2 puede alcanzar un máximo de 1.

$$R_{max}^2 = 1 - \exp[2(n^{-1})LL(0)]$$

2.5 Prueba de Hosmer-Lemeshow

Esta es otra medida global de la exactitud predictiva del modelo, pero a diferencia de las anteriores no esta basada en el valor de la función de verosimilitud sino en la predicción real de la variable dependiente. Dicho contraste consiste en realizar comparaciones entre el valor estimado y el observado por grupos. Para ello las observaciones se dividen en J grupos (generalmente 10) aproximadamente iguales, dividiendo el recorrido de la probabilidad en deciles de riesgo.

A partir de esta información se construye una tabla de contingencias donde se compara tanto la distribución de ocurrencia, como la de no ocurrencia y los valores realmente observados. El contraste se realiza siguiendo las siguientes hipótesis: H_0 : el modelo ajusta los datos (no hay mucha diferencia entre los valores esperados y los valores observados), contra la alternativa H_1 : el modelo no lo ajusta.

Hosmer y Lemeshow demuestran que cuando el modelo es correcto (se obtiene un buen ajuste) el estadístico HL sigue una distribución chi-cuadrado con J-2 grados de libertad, y los valores de probabilidad asociados al estadístico son superiores al 5% del nivel de significación.

2.6 Porcentaje de aciertos

Otra de las vías utilizadas para determinar la bondad del ajuste de un modelo logit es predecir con el modelo los valores de la variable endógena Y_i de tal manera que $Y_i = 1$ si $p_i \geq 0,5$ o $Y_i = 0$ si $p_i < 0,5$. Dado que los valores reales de Y_i son conocidos, basta con contabilizar el porcentaje de aciertos para decir si la bondad del ajuste es elevada o no.

3 Prueba Ómnibus sobre los coeficientes del modelo

El contraste se realiza a través del nivel de significación asociado al estadístico Chi-Cuadrado. Contrasta la hipótesis nula de un mal ajuste del modelo (adicionando variables explicativas al modelo no se ha incrementado significativamente la capacidad de éste para predecir las decisiones hechas por los sujetos), contra la alternativa de que éste alcanza un buen ajuste y en consecuencia incrementa su capacidad explicativa.

El contraste se realiza a través del nivel de significación asociado al estadístico, de tal manera que si el nivel de significación es mayor a 0,05 el ajuste del modelo es malo.

Sección 3: Descripción de los Datos: fuentes de información, bases de datos y variables utilizadas en el análisis

En esta sección y con el objetivo de poder realizar con posterioridad el análisis empírico que se propone en este estudio, se hace necesaria una descripción sintetizada de la fuente de información empleada, los datos utilizados, así como las variables dependientes y las variables socio-económicas que inciden en las mismas.

3.1 Fuente de Información

La principal fuente de información en este trabajo es la Encuesta sobre la Situación Económica de los Hogares (ESEH) de 2004. La encuesta de los hogares tiene por objetivo generar indicadores que permitan conocer los niveles de bienestar y las condiciones de vida de los hogares. Más específicamente, la ESEH contiene información sobre las características demográficas y socioeconómicas de los individuos, la asistencia de éstos a establecimientos educativos y laborales, así como los ingresos (en moneda nacional y en divisas) de los ocupados según sectores de propiedad y su relación con el empleo. Este tipo de encuesta no brinda información sobre el uso de servicios de salud.

La encuesta es procesada por la Oficina Nacional de estadística (ONE) y es captada para la zona urbana y rural y se ha mantenido con el mismo alcance desde 1997 hasta el 2004. El universo y alcance de la encuesta abarca a toda la población residente en viviendas particulares y locales de trabajo (con viviendas particulares) de la parte urbana y rural del país. Se excluyen los residentes en colectividades, ingresados permanentemente en hospitales y sanatorios. El tamaño de la muestra fue de 40 000 viviendas y ésta fue seleccionada del Marco Muestral Maestro elaborado por la ONE para la realización de encuestas de hogares en 1995. El Marco Muestral Maestro está constituido por Áreas Geográficas Muestrales⁶.

El muestreo fue por Conglomerados Trietápico Estratificado, con estratificación de las unidades primarias. Las provincias y dentro de ellas los municipios constituyeron los estratos.

Para cada vivienda seleccionada se obtuvieron datos de los hogares que la ocupaban y de sus respectivos miembros. La encuesta está compuesta por 4 módulos fundamentales. El primer bloque de preguntas es un módulo fijo que está referido a las características generales y educacionales para todas las edades (color de la piel, sexo, edad, parentesco, educación, etc). El segundo bloque es un módulo específico de ocupación para personas de 15 años y más de edad. El tercer módulo corresponde a la captación de los ingresos que obtienen esos individuos, divididos estos en una variedad de tipologías diferentes, mientras que el último recoge los gastos en moneda nacional y en divisas.

⁶ Son territorios que tienen en promedio 180 viviendas y que se clasifican en 8 categorías: urbano llano, urbano montañoso, rural llano (concentrado, mixto y disperso) y rural montañoso (concentrado, mixto y disperso).

Esta base de datos incluye dos preguntas relevantes para el análisis de los determinantes en el sector educativo. La primera es cuál es el nivel educativo máximo aprobado a la fecha de la entrevista y en el módulo para personas de 15 años y más, en cuál situación se encontraba usted la semana anterior, a la que habría que responder estudiante.

3.2 Las Bases de Datos

Para el análisis de los datos se ha realizado un muestreo estratificado de la población, seleccionando los diferentes estratos de la muestra de acuerdo a las diferentes variables dependientes seleccionadas. La heterogeneidad de los estratos implica que no todos tengan la misma representatividad, pero para cada uno de ellos es aplicable un factor de elevación de la muestra ponderado por el tamaño de la misma.

La base de población cuenta con aproximadamente 24 115 observaciones. Para poder obtener y evaluar las probabilidades de ocurrencia o no de un conjunto de sucesos en función de variables socioeconómicas de demanda y oferta educacionales, es necesario primeramente definir los diferentes eventos objeto de estudio. En cada uno de los eventos bajo análisis se utiliza una submuestra, es decir, un grupo de observaciones consideradas relevantes para la decisión en juego.

El primer nivel de decisión corresponde al acceso de los niños a la enseñanza primaria. Esta submuestra está compuesta por 297 observaciones y comprende los niños que entran en la escuela a la edad de 6 años y responden a la pregunta *que entraron en la enseñanza primaria*. El segundo nivel de decisión comprende la continuidad de estudios a la edad de 15 años representado por 280 observaciones. Esta es una etapa de transición del nivel secundario al nivel preuniversitario o a la enseñanza politécnica.

El tercer nivel de decisión que se estudia se relaciona con el completamiento de estudios en tiempo a la edad de 18 años. La población de interés para el análisis sería aquella formada por todos los individuos que tuvieran 18 años y que respondieran la pregunta correspondiente a los estudios medios terminados. En este caso el número de observaciones con que se realizaron las estimaciones es 428.

3.3 Descripción de las variables

La esencia de este estudio consiste en relacionar las diferentes decisiones objeto de análisis con variables socio-económicas del individuo o del hogar. Las variables utilizadas han sido seleccionadas tomando como base de datos la Encuesta Permanente de los hogares para el año 2004, así como información ofrecida por la Dirección de Planeación y Estadística del Ministerio de Educación.

El modelo teórico especificado con anterioridad se ajustará teniendo en cuenta tres variables dependientes cualitativas diferentes: el acceso a la enseñanza primaria a los 6 años, la continuidad de estudios a los 15 años y el completamiento de los estudios en

tiempo a los 18 años. En cada una de las submuestras antes mencionadas, la variable dependiente es un indicador binario que toma el valor de 1 si el resultado de la decisión es afirmativo y 0 en caso contrario.

Al mismo tiempo, los modelos logísticos que serán estimados tienen en cuenta un conjunto de variables independientes, las cuales son generalmente aceptadas y nombradas en la literatura como los *determinantes*. La selección de estas variables explicativas se ha visto condicionada por tres aspectos fundamentales: la literatura existente al respecto, la disponibilidad de los datos y las especificidades de la política social cubana, y en particular de la política educacional.

La literatura económica sugiere que si bien el proceso de identificar factores específicos que afectan a la educación es complejo, existen algunos elementos fundamentales cuya evidencia empírica es notable. A continuación se enumeran una serie de cuestiones que pueden ser relevantes en el análisis:

- *Rol de la educación de los padres*: Una mayor educación de los padres debiera incidir directamente en una mayor educación de sus hijos, pues le dan una mayor valorización a la educación.
- *Rol del ingreso familiar*: el ingreso mide la cantidad de recursos disponibles para financiar la educación. Esta variable también puede captar otros efectos directamente relacionados: como la educación de los padres y la situación de empleo de los miembros del hogar.
- *Efecto de escala*: Los efectos del tamaño de la familia afectan la posibilidad de asistir a la escuela de cada uno de los hijos, así como la manera de reasignar los recursos para estudiar dentro de la familia.
- *Efecto género*: Es muy interesante distinguir si cada uno de los factores mencionados afecta de una forma diferente a la educación de los hijos y de las hijas.
- *Efecto de cambios en la estructura familiar*: resulta interesante estudiar cómo afectan a la decisión de estudiar las diferentes conformaciones familiares (padres viudos, divorciados, casados, solteros).

Los determinantes de estas tres variables dependientes -enmarcadas todas dentro del proceso educativo- pueden encontrarse en la accesibilidad física a la escuela y la calidad de las entradas escolares, es decir, pueden clasificarse en variables por el lado de la demanda y oferta educativas, lo que está en correspondencia con la metodología propuesta en la mayoría de los estudios sobre este tema.

En el cuadro que se muestra a continuación se realiza una descripción de las variables explicativas internacionalmente empleadas de acuerdo a la anterior clasificación, muchas de las cuales serán utilizadas en este trabajo.

Cuadro No. 2
Variables Explicativas

Determinantes	
Por el lado de la demanda educativa	Por el lado de la oferta educativa
<ul style="list-style-type: none"> ➤ Accesibilidad física a la escuela: <ul style="list-style-type: none"> • Nivel de escolaridad de los padres • Situación económica del hogar (ingreso familiar disponible per cápita) • Costos de oportunidad • Sexo • Edad del niño • Localización • Color de la piel • Total de personas en el núcleo • Tasa de mortalidad infantil • Estado civil de los padres • Cantidad de hermanos 	<ul style="list-style-type: none"> ➤ Calidad de las entradas escolares: <ul style="list-style-type: none"> • Número de alumnos por profesor • Cantidad de escuelas • Gasto en educación por alumno • Porcentaje de profesores calificados • Infraestructura creada en este sector • Disponibilidad de libros de texto • Disponibilidad de uniformes escolares • Capacitación de los profesores • Porcentaje de matriculados en seminternado e internados • Porcentaje de matriculados en doble sesión

Después de haberse realizado una evaluación de las variables provistas por la encuesta y otras fuentes, se presenta un resumen de las variables seleccionadas para realizar las estimaciones en el país:

Definición de las variables dependientes:

- (Y= ACCESO6) $Y_i \left\{ \begin{array}{l} 1 \text{ si accede a la primaria} \\ 0 \text{ si no accede} \end{array} \right\}$

- (Y= CON15) $Y_i \left\{ \begin{array}{l} 1 \text{ si continua los estudios} \\ 0 \text{ si no los continua} \end{array} \right\}$

- de nivel medio a nivel medio superior

- (Y= COMPLETE18) $Y_i \left\{ \begin{array}{l} 1 \text{ si completa los estudios} \\ 0 \text{ si no los completa} \end{array} \right\}$

- del nivel medio superior

Definición de las variables independientes a incluir indistintamente en los tres modelos:

Variables	Descripción
TPERNUC	Total de personas por núcleo
ESCONUC	Escolaridad del núcleo
CPIEL	Color de la piel
SEX	Sexo
TMORMEN5	Tasa de mortalidad infantil menores de 5 años
INGRESOS	Total de ingresos per cápita del hogar
ALPROMES	Alumnos por profesor en la media superior
ALPROMEB	Alumno por profesor en la media básica
ALPROPRI	Alumno por profesor en la primaria
ESCUEMEB	Escuelas en la media básica
ESCUEMES	Escuelas en la media superior
ESCUEPRI	Escuelas en primaria
GASPRIAL	Gasto en primaria por alumno
GASSECAL	Gasto en secundaria por alumno
PREMSUP	Premio en la enseñanza superior
PREMMED	Premio en la media
SEMIMEDS	Seminternado en la media superior
SEMIPRIM	Seminternado en primaria
SEMIMEDB	Seminternado en la media básica
INTEMEDB	Internados en la media básica
INTERMEDS	Internados en la media superior
H2ODOMI	Cobertura de agua domiciliaria

Las primeras cuatro variables que aparecen en el cuadro, representan características individuales y familiares. El sexo y la edad de una persona se suponen potencialmente determinantes en las diferentes decisiones relacionadas con el proceso educativo. También se incluyen el nivel educativo de los padres⁷ y el número de personas que conviven en el núcleo, al entenderse que cada decisión no depende exclusivamente de la persona, sino del entorno que lo rodea. Si los padres son más educados preferirán una mejor educación para sus hijos, es decir, el nivel educativo de los padres captura esencialmente las preferencias de éstos por la educación de sus hijos.

La tasa de mortalidad infantil sólo es incluida en el primer modelo de acceso a la enseñanza primaria, pues en dependencia del número de niños menores de 5 años que mueren, así se comportará la entrada a este primer nivel de enseñanza. La inclusión de este indicador también responde al objetivo de relacionar las diferentes metas del milenio que se han propuesto alcanzar los países para el 2025. En este caso, si el país muestra bajas tasas de

⁷ Esta variable probablemente haya sido analizada en todos los estudios sobre determinantes de la educación.

mortalidad infantil, mayor número de niños podrán acceder a la enseñanza primaria universal.

El ingreso total familiar per cápita resulta un determinante muy típico e importante en la literatura sobre el tema, al mostrar en qué medida la menor disponibilidad de recursos económicos restringe el proceso educativo en su conjunto. En nuestro país, y teniendo en cuenta las peculiaridades de la política social cubana, este factor no parece ser un determinante relevante en el estudio, al ser la educación completamente gratuita para todos los niveles educativos, rechazándose a los ingresos individuales como el motor de la política social.

El otro grupo de variables restantes, corresponden con las variables por el lado de la oferta educacional dadas por el número de alumnos por profesor, el número de escuelas, la cantidad de niños seminternados e internados en los diferentes niveles, los gastos corrientes por alumnos en primaria y secundaria y la infraestructura pública.

Entre las prioridades más importantes del gasto público en nuestro país se encuentra la educación, a la cual se le dedican crecientes recursos desde 1998. Se entiende que en la medida en que se destine más recursos a la educación, mejores serán las condiciones materiales para lograr el acceso, continuidad y completamiento de estudios de los estudiantes. Los gastos de este sector representaron el 12% del PIB en el 2004. De igual manera, el indicador que recoge los gastos en primaria y secundaria por alumnos para las diferentes provincias, también muestra incrementos significativos con algunas diferenciaciones regionales.

Actualmente en el sector educacional se han venido desarrollando una serie de programas, encaminados a lograr una mayor formación integral de niños y jóvenes creando alternativas que garanticen la continuidad y el completamiento de los estudios, una mayor calidad de la enseñanza, una mejor preparación de los jóvenes, el mejoramiento de las condiciones materiales, entre otros aspectos.

Entre los principales programas se encuentran: programa para la atención integral de los alumnos en la educación primaria, el programa de perfeccionamiento del nivel de secundaria básica, la universalización de la enseñanza, el programa audiovisual, el programa editorial libertad, etc. Muchos de ellos han contribuido a la reducción del número de alumnos por grupo, a la reparación de una gran cantidad de escuelas y construcción de otras nuevas, a la creación de mejores condiciones para que los niños permanezcan en sus escuelas bajo la doble sesión, entre otras.

Las variables que se corresponden a la brecha de salario (en nuestro caso, *premed* y *premsup*) se encuentran vinculadas indirectamente con los costos de oportunidad asociados a las diferentes variables dependientes. En muchos países, se puede observar la situación donde la contribución del niño a los ingresos de la familia resulta más significativa que los beneficios que le reportaría a ésta y en especial al niño si asistiera a la escuela, en consecuencia, el costo de oportunidad vinculado con la asistencia a la escuela podría ser sustancial.

En nuestro país, al ser la enseñanza gratuita y universal y la edad laboral estar establecida a partir de los 17 años, este costo de oportunidad no existe. Sin embargo, se podría incluir en el análisis los beneficios futuros (ingresos) o el premio que obtendría el niño si alcanzara el nivel secundario en contraposición con ningún nivel de educación (premed) y si terminara los estudios medios y comenzara los estudios terciarios (premsup), ambas variables medidas como un proxy del nivel de salario medio alcanzado por los trabajadores calificados y no calificados.

Por otro lado, cuando se analizan los costos directos e indirectos de la decisión de mantener al joven en la escuela –se incluye gasto en uniformes, el material educativo, los costos de transporte, entre otros- se observa que éstos son muy significativos fundamentalmente en las economías de mercado, hasta el punto de impedir la entrada o continuación del niño a la escuela si la familia no tiene los recursos necesarios. En nuestro caso, la inclusión de esta variable no parece tener ningún significado importante debido a las propias características del sistema educativo cubano.

Sin lugar a dudas, sería de suma importancia establecer si los factores de demanda son los determinantes que juegan un papel más relevante, o si lo son los de oferta. En el primer caso, si ésto sucediera no bastará con que se construyan nuevas escuelas, se mejoren los indicadores de calidad, etc para incrementar los niveles educativos. En el segundo caso, el modelo estaría demostrando, la enorme importancia que tienen y seguirán teniendo la gran mayoría de los programas sociales que se están poniendo en práctica en el país desde hace algunos años.

Sección 4: Resultados de las estimaciones econométricas para los tres modelos del proceso educativo.

Como se ha mencionado en la sección correspondiente a la metodología, en esta parte del trabajo, con base en regresiones logit, se investigan y evalúan las variables socio-económicas que potencialmente afectan las decisiones de acceder, continuar y completar los estudios. Una vez definidas las variables incluidas en el análisis, el próximo paso consiste en la estimación e interpretación de los resultados. Los resultados obtenidos permitirán formular algunas recomendaciones de políticas válidas para continuar con el perfeccionamiento de la política educacional cubana.

En primer término, se realiza una caracterización de los tres modelos empleados y se presentan los resultados de la estimación para cada uno de ellos. En adición, también son resumidas las pruebas estadísticas fundamentales que demuestran la validez y fortaleza de los mismos, así como la significatividad de cada variable independiente.

En segundo término, y sobre la base de los coeficientes estimados para cada variable explicativa se analiza el impacto cualitativo de las variables a través de los signos de sus coeficientes y luego con el objetivo de facilitar la interpretación, se estiman numéricamente las elasticidades o efectos marginales para disponer de una medida cuantitativa del efecto específico de cada una de ellas sobre la variable dependiente correspondiente.

Las estimaciones de los modelos son realizadas a través del software estadístico SPSS10, mientras que los cálculos correspondientes a las elasticidades se efectuaron en el STATA9. Los métodos de estimación -basados en el ratio de máxima verosimilitud- que reportaron modelos más relevantes y significativos fueron: el Backward LR para los primeros dos modelos y el Forward LR para el tercero. El modelo Enter, al ser el más sencillo de los tres, recogía en ocasiones, variables independientes no significativas.

En los cuadros No. 3, 4 y 5 se presentan las salidas obtenidas para los diferentes modelos logísticos. Para cada uno de ellos, de acuerdo con el valor del estadístico Chi-Cuadrado, las variables incluidas son conjuntamente significativas, rechazándose la hipótesis nula de que en conjunto no explican el modelo.

Cuadro No. 3
Salidas del modelo de acceso a la primaria a los 6 años

Modelo 1. Acceso a los 6 años				
<i>Modelo Backward</i>				
VARIABLES	Coefic.	Wald (Sig.)	Exp(B)	Elasticidad
TMORMEN5	-0.211	0.036	0.80977	-0.2187
ESCUEPRI	0.001	0.066	1.01005	0.1038
ESCONUC	0.436	0.000	1.54650	0.7855
INGRESOS	0.000	0.028	0.99970	0.0735
TPERNUC	0.434	0.000	1.54341	0.2636
ALPROMEB	-0.427	0.000	0.65246	-0.6118
Chi-Cuadrado (10) = 70.28				
Probabilidad $\chi^2 = 0.000$				
Log likelihood = -118.3355				
Pseudo R ² = 0.2290				
ML (-2LL) = 236.669				
Y = 0.86760165				

El primer modelo fue estimado involucrando a un gran conjunto de variables explicativas. Estas fueron: SEXO, CIEL, TPERNUC, ESCONUC, TMORMEN5, GASPRIAL, INGRESOS, ESCUEPRI, ESCUEMEB, PREMMEB, SEMIPRIM, SEMIMEDB, ALPROPRI, ALPROMEB, y H2ODOMI. De todas ellas, sólo resultaron relevantes para el análisis las que aparecen en la salida que se muestra a continuación. Las restantes fueron excluidas del mismo en los pasos previos del modelo Backward al presentar valores del estadístico Wald inferiores a 1.96 en valores absolutos.

Cuadro No. 4
Salida del modelo de continuidad de estudios a los 15 años

Modelo 2. Continuidad de estudios a los 15 años				
<i>MODELO BACKWARD</i>				
VARIABLES	COEFIC.	WALD(SIG.)	EXP(B)	Elasticidad
SEX	1.261	0.010	3.527	0.0103
ESCONUC	0.338	0.000	1.402	0.0525
INGRESO	0.007	0.043	1.007	0.0213
ALPROMES	-0.089	0.043	0.915	-0.0190
Chi-Cuadrado (3) = 33.8				
Probabilidad $\chi^2 = 0.000$				
Log likelihood = -89.712393				
Pseudo R ² = 0.159				
ML (-2LL) = 179.425				
Y = 0.984481				

El segundo modelo de continuidad de estudios a los 15 años ha sido estimado con las siguientes variables: SEXO, CPIEL, TPERNUC, ESCONUC, PREMMED, GASSECAL, PREMSUP, ESCUEMEB, ESCUEMES, ALPROMEB, ALPROMES, INTEMEDS, SEMIMEDS, INGRESOS y H2ODOMI. Sólo resultaron estadísticamente significativas la escolaridad del núcleo, el sexo, los ingresos y los alumnos por profesor en la media superior. El resto de las variables no fueron incluidas en el modelo estimado final al presentar cada una de ellas valores de probabilidad superiores al nivel de significación del 5%, lo que correspondería a una muy poca relevancia en el modelo.

Cuadro No. 5

Salida del modelo de completamiento de los estudios en tiempo a los 18 años

Modelo 3. Completamiento a los 18 años				
MODELO BACKWARD				
VARIABLES	COEFIC.	WALD(SIG.)	EXP(B)	Elasticidad
ESCONUC	0.445	0.000	1.561	2.1767
INGRESOS	0.002	0.055	1.002	0.1956
SEXN	0.310	0.006	1.364	0.2759
CONSTANTE	-5.694	0.000	0.003	
Chi-Cuadrado (3) = 112.19				
Probabilidad χ^2 = 0.000				
Log likelihood = -239.05394				
Pseudo R ² = 0.1901				
ML (-2LL) = 478.104				
Y = 0.55105211				

El tercer y último modelo de completamiento de los estudios en tiempo a los 18 años también ha presentado características similares al anterior en cuanto a la selección de sólo un pequeño grupo de variables significativas. En este caso en particular fueron incluidas las siguientes variables: SEXO, CPIEL, TPERNUC, INGRESOS, ESCONUC, GASSECAL, ALPROMES, ESCUEMES, PREMSUP, INTEMEDS y H2ODOMI, sólo formando parte del modelo final: el sexo, la escolaridad del núcleo y los ingresos.

Cuando se analiza cada una de las variables independientes por separado, se observa cómo las que han sido seleccionadas son también individualmente significativas, mostrando en cada caso un valor de probabilidad asociado al estadístico Wald inferior al 5% del nivel de significación. La variable ESCUEPRI en el primer modelo de acceso resulta significativa para un 10 % de significación (0.66), esta podría ser la única excepción.

Por otro lado, los Pseudo R² son elevados para este tipo de modelos, mostrando valores entre 0.18 y 0.27, situación que reafirma que las variables incluidas en los mismos son significativas y que el modelo cuya única variable es la constante no es similar al modelo completo (incluye a todas las variables explicativas).

A los efectos de explorar una especificación robusta y estable de los modelos, se realizan los contrastes de hipótesis para los diferentes estadísticos de bondad de ajuste. En los 3 casos, las tablas resumidas a continuación muestran los estadísticos que han sido explicados en la sección que aborda la metodología. Estos son: Nagelkerke, Hosmer-Lemeshow, el Test de Ómnibus y el porcentaje de predicciones correctas.

Modelo 1

NAGELKERKE	HOS-LEMESHOW	OMNIBUS (SIG.)	PREDICC.CORREC (%)
0.371	0.257	0.000	85.9

Modelo 2

NAGELKERKE	HOS-LEMESHOW	OMNIBUS (SIG.)	PREDICC.CORREC (%)
0.903	0.319	0.000	96.9

Modelo 3

NAGELKERKE	HOS-LEMESHOW	OMNIBUS (SIG.)	PREDICC.CORREC (%)
0.307	0.952	0.000	71

En todos los casos, los contrastes de hipótesis son aceptados y rechazados favorablemente. El análisis realizado se muestra a continuación:

1. El Nagelkerke, aunque no alcanza el valor máximo de 1 en ninguno de los modelos, resulta significativo de manera individual. En particular, el alto valor del estadístico presentado en el segundo modelo conjuntamente con la tabla de aciertos estaría representando una capacidad explicativa del modelo muy elevada.
2. En el caso de Hosmer-Lemeshow, el contraste se realiza teniendo en cuenta la hipótesis nula que plantea que el modelo ajusta los datos, contra la alternativa de que el modelo no lo ajusta. Los valores de probabilidad asociados al estadístico para los tres modelos son superiores al 5 % del nivel de significación, razón por la cual la hipótesis nula es aceptada y en consecuencia se plantea que no existe mucha diferencia entre los valores esperados y los valores observados en el modelo.
3. El test de Omnibus está arrojando que todas las probabilidades asociadas al estadístico Chi-Cuadrado son inferiores al nivel de significación de 0.05. Si esto sucede, se rechaza la hipótesis nula que brinda un mal ajuste del modelo, obteniéndose como resultado un modelo que si se ajusta correctamente al ir adicionando variables explicativas e ir incrementando simultáneamente la capacidad explicativa de este para predecir las decisiones hechas por los sujetos.
4. También fueron obtenidas las tablas de predicciones utilizadas para medir la bondad de ajuste de los mismos. Estas permiten observar como la capacidad explicativa del mismo es elevada, obteniéndose predicciones correctas en un 85.9, 96.9 y 71% de los casos cuando la cota de análisis se fija en 0.5.

Una vez analizada la bondad de ajuste para cada uno de los modelos y la significatividad individual y conjunta de los coeficientes de los parámetros estimados, se analiza la influencia que los diferentes factores poseen sobre la variable dependiente.

En el análisis logit -como ya se había comentado con anterioridad- el valor del signo estimado para el coeficiente asociado indica directamente la dirección de la relación con respecto a la variable dependiente. Los signos de los coeficientes estimados son resumidos en los cuadros siguientes.

Cuadro No. 6

Determinantes del acceso a la enseñanza primaria

Modelo 1	
Factores de demanda	TPERNUC (+)
	ESCONUC (+)
	TMORMEN5 (-)
	INGRESOS (+)
Factores de oferta	ALPROMEB (-)
	ESCUEPRI (+)

Cuadro No. 7

Determinantes de la continuidad de estudios a los 15 años

Modelo 2	
Factores de demanda	SEXO (+)
	INGRESOS (+)
	ESCONUC (+)
Factores de oferta	ALPROMES (-)

Cuadro No. 8

Determinantes del completamiento de los estudios a los 18 años

Modelo 3	
Factores de demanda	SEXO (+)
	ESCONUC (+)
	INGRESOS (+)

Los cuadros permiten examinar que:

- En el primer modelo, se observa un efecto positivo sobre la probabilidad de acceso a la primaria de variables como la escolaridad del núcleo, del número de personas que viven en él, el número de escuelas en primaria y los ingresos per cápita. Por su parte, y tal y como se esperaba, el acceso está relacionado negativamente con la tasa de mortalidad infantil y los alumnos por profesor en la enseñanza media básica.

- En el segundo modelo, la probabilidad de continuar los estudios está positivamente asociada por el lado de la demanda, con la escolaridad del núcleo, el sexo y los ingresos de los hogares, no siendo de la misma manera el efecto (inverso) de la cantidad de alumnos por profesor en la enseñanza media superior sobre la probabilidad, lo que se refleja en el signo negativo del coeficiente.
- El último modelo, considera sólo tres variables que inciden de manera significativa sobre la variable dependiente. Cada una de ellas muestra un efecto directo positivo sobre la probabilidad de completar los estudios en tiempo a los 18 años.

La escolaridad del núcleo, es una de las variables más relevantes en la determinación de la educación de los hijos en el país. El efecto positivo y significativo observado en los tres modelos muestra que mientras más alto sea el nivel educativo familiar, mayor será la probabilidad de que el niño matricule, continúe y complete los estudios satisfactoriamente.

Relacionada con esta, la variable total de personas que conviven en el núcleo, resulta ser positiva e importante en el modelo de acceso a la primaria, lo que estaría significando que mientras más personas habiten en el hogar, mayores serán las probabilidades de que el niño acceda al primer nivel de estudio a los 6 años.

El ingreso por su parte, resulta ser una variable débilmente significativa y positiva en los dos primeros modelos. Este resultado refuerza la idea anteriormente planteada de que el ingreso de las familias no es una variable determinante para que el niño acceda o continúe sus estudios, por las características peculiares de la política educacional cubana. En el tercer modelo, ésta resulta ser un poco más significativa que en los anteriores.

Es importante tener presente que el modelo social cubano brinda un menor papel a los ingresos monetarios y a las relaciones mercantiles como condición para acceder al consumo y amplía la magnitud e importancia de bienes y servicios que se brindan a través de su política social gratuita y subsidiada, lo que aumenta respecto a otros países la equidad de acceso a servicios tan importantes como lo son la educación y la salud⁸.

La tasa de mortalidad infantil, tal y como se había comentado con anterioridad, muestra claramente a través del coeficiente estimado su signo negativo y su magnitud relevante. La relación inversa existente entre la tasa de mortalidad infantil y el acceso a la primaria puede ser explicada intuitivamente: mientras mayores sean las tasas de mortalidad de los menores de 5 años, menores será el número de niños que puedan entrar en la enseñanza primaria.

La cantidad de escuelas en primaria es otro factor que está incidiendo en la probabilidad de acceso a la primaria. El número de escuelas en el primer nivel de enseñanza incide favorable y significativamente en el modelo de acceso, es decir, mientras mayor sea el número de escuelas construidas, mayor será la probabilidad y viceversa.

⁸ Ver INIE, Segundo Informe sobre los Objetivos del Milenio para Cuba realizado en julio de 2005.

El sexo es otra variable que influye positiva y significativamente en la probabilidad de que el individuo continúe sus estudios y que los complete en tiempo. En este caso, el hecho de que el niño sea varón o hembra pudiera conllevar a diferentes situaciones que afectarían a las variables dependientes involucradas.

La cantidad de alumnos por profesor en la enseñanza media básica y en la media superior arrojan un signo negativo esperado, lo que estaría indicando que mientras mayor sea el número de alumnos por profesor, menores serán las probabilidades de acceso a la primaria y la continuidad de los estudios medios básicos a los medios superiores.

Para tener una medida de los efectos de cada variable sobre la probabilidad de satisfacer el evento son reportados los efectos marginales. Estos efectos marginales indican la variación porcentual de la variable dependiente debido a un incremento de un 1 % de la variable independiente. Las elasticidades son mostradas a continuación:

Cuadro No. 9
Elasticidades relevantes para el primer modelo

Acceso a los 6 años Y=0.86760165		
VARIABLES	EXP(B)	Elasticidad
TMORMEN5	0.80977	-0.2187
ESCUEPRI	1.01005	0.1038
ESCONUC	1.54650	0.7855
INGRESOS	0.99970	0.0735
TPERNUC	1.54341	0.2636
ALPROMEB	0.65246	-0.6118

Para este modelo, las variables que provocan un mayor cambio en la probabilidad de acceder a la enseñanza primaria son aquellas que por un lado, recogen las características socio-económicas del individuo (escolaridad del núcleo y la cantidad de personas que viven en él) y por otro, se incluyen la tasa de mortalidad infantil y la cantidad de alumnos por profesor.

La probabilidad de acceder a la enseñanza primaria aumenta en 0.78 % cuando los padres presentan un mayor nivel educativo, en 0.26 % cuando aumenta el número de personas en el hogar, y disminuye en un 0.21 % cuando aumenta la tasa de mortalidad de niños menores de 5 años y en 0.61 % cuando aumenta el número de alumnos por profesor en la media básica.

El ingreso per cápita, por su parte, resulta ser una variable estadísticamente significativa a pesar de mostrar un impacto poco relevante sobre la probabilidad de acceso. Tiene un efecto pequeño y marginal equivalente a 0.7%, es decir, ante un cambio de 1 % en el nivel de ingresos, la probabilidad aumenta 0.07%. Este resultado refuerza la idea anteriormente planteada de que el ingreso no debe ser una variable relevante en los diferentes análisis.

El número de escuelas primarias existentes resulta ser también significativa, al esperarse una subida moderada en la probabilidad de acceso en aproximadamente 0.10 % para cada incremento en un punto porcentual en la cantidad de escuelas.

La probabilidad de acceder a la educación primaria en el año en estudio ($Y = 0,86760165$) es explicada por el conjunto de elasticidades que resultaron significativas y por el adecuado ajuste del modelo. No obstante y con vistas a elevar esta probabilidad podrán ser realizadas otras estimaciones econométricas donde se incluyan nuevas variables que no han podido ser incorporadas en esta ocasión, por las características de la información recogida en la encuesta.

Cuadro No. 10
Elasticidades relevantes para el segundo modelo

Continuidad a los 15 años $Y=0.984481$		
VARIABLES	EXP(B)	Elasticidad
SEX	3.527	0.0103
ESCONUC	1.402	0.0525
INGRESO	1.007	0.0213
ALPROMES	0.915	-0.0190

El segundo modelo estimado se corresponde a la continuidad de estudios a los 15 años, siendo la probabilidad obtenida de 0,984481. La escasa variabilidad de la variable dependiente disminuye la posibilidad de encontrar nuevos determinantes que tengan un efecto significativo sobre la continuidad. Esto puede ser explicado si se analizan las estrategias educativas puestas en marcha en el país entre las que destacan: las posibilidades de continuidad de estudios que se le brinda a los adolescentes al culminar la enseñanza obligatoria de 9no grado, conjuntamente con el intenso trabajo que se desarrolla por disminuir la deserción mediante la inserción de estos, en los diferentes programas sociales en ejecución.

Los bajos valores de las elasticidades confirman las pocas posibilidades de encontrar efectos sistemáticos sobre la probabilidad. El sexo, los alumnos por profesor en la media superior, los ingresos y la escolaridad del núcleo son determinantes que tienen un efecto discreto marginal sobre la probabilidad de continuar estudios medios superiores (0.01, - 0.01, 0.02 y 0.05 %).

La variable que muestra la mayor elasticidad es la escolaridad del núcleo, mientras que el género y el ingreso influyen pero no suelen ser significativamente relevantes en este modelo. No obstante y con el objetivo de desagregar el efecto del género y su impacto sobre la probabilidad, se procedió al cálculo de la predicción del odds, lo cual arrojó que la probabilidad de que los hombres continúen los estudios medio superiores es 2.52 veces superior que para las mujeres, lo que pudiera estar explicado entre otras causas por el embarazo en edades comprendidas entre los 15 y 18 años.

Cuadro No. 11
Elasticidades relevantes para el tercer modelo

Completamiento a los 18 años $Y=0.55105211$		
VARIABLES	EXP(B)	Elasticidad
ESCONUC	1.561	2.1767
INGRESOS	1.002	0.1956
SEXN	1.364	0.2759

El tercer modelo estimado sobre la probabilidad de completar los estudios en tiempo a los 18 años (graduarse de la enseñanza media superior) muestra elasticidades significativas para los tres determinantes que se presentan en el cuadro No. 11 (la escolaridad del núcleo, el sexo y los ingresos).

La escolaridad del núcleo continúa presentando los efectos más significativos sobre la probabilidad no sólo de acceder y continuar los estudios, sino también sobre la probabilidad de completar los estudios medios superiores. Un incremento de un 1 % en la escolaridad del núcleo, provoca un aumento de 2.17 % puntos en la probabilidad de completar los estudios medios.

El hecho de que un núcleo pueda contar con un alto porcentaje de personas que hayan alcanzado un nivel educativo alto, está incidiendo favorablemente y de forma importante en la decisión de que sus hijos se gradúen satisfactoriamente en los diferentes niveles. Estas decisiones pueden verse también influenciada por la política educacional más reciente que potencia acciones específicas para que los profesores atiendan diferenciadamente a los alumnos y sus familias particularizando sus dificultades docentes y sociales, lo que permite inferir la gran importancia que se le viene confiriendo a este factor.

El sexo para este modelo sí resulta ser un determinante relevante y altamente significativo, describiendo una elasticidad de 0.27 %. El análisis diferenciado por género nos muestra que las mujeres una vez que han logrado continuar los estudiar medios tienen mayores probabilidades de graduarse que los hombres, es decir, la probabilidad de completar los estudios a los 18 años es 1.17 veces mayor para las mujeres que para los hombres. Las estadísticas de graduados de la enseñanza media superior (preuniversitario, técnico medio, magisterio) confirman estos resultados, observándose un 48% de hombres graduados del nivel medio superior contra un 51% de mujeres, porcentajes muy similares al patrón demográfico.

El ingreso per cápita también muestra una elasticidad significativa aunque no llega a alcanzar una elevada magnitud (0.10%) si se compara con la escolaridad del núcleo y con el sexo. Los ingresos, además de medir la cantidad de recursos disponibles para financiar la educación, -que para el caso de nuestro país este argumento no es aplicable-, también pueden captar otros efectos directamente relacionados como la educación de los padres y la situación económica de los miembros del hogar.

Este primer elemento ha resultado ser, de hecho, el determinante más importante de la probabilidad de completamiento, el cual se encuentra directamente relacionado con el nivel de ingreso recibido por los miembros del hogar, esto es, mientras mayor sea el nivel de preparación promedio del hogar, mayores serán los ingresos percibidos. Trabajos más recientes estiman la tasa de retorno de la educación para los distintos niveles educativos terminados, la cual muestra incrementos paulatinos de los ingresos (salarios y remuneraciones) en la medida que el individuo alcanza un mayor nivel de educación.

En lo que respecta a la situación económica del hogar, se han detectado casos de comportamiento de deserción en los diferentes niveles educativos de la enseñanza media superior vinculados a los bajos ingresos per cápita que se perciben en el hogar, contribuyendo de esta forma a que un grupo pequeño de adolescentes dejen parcialmente sus estudios para comenzar a trabajar. La gran mayoría de estos estudiantes no salen del sistema educativo automáticamente, pues muchos de ellos terminan sus estudios de 12 mo grado en facultades nocturnas o se incorporaran a otros programas de la revolución que les han permitido estudiar y trabajar al mismo tiempo.

La probabilidad para el año 2004 es 0.55105211, baja si se compara con las probabilidades anteriormente estimadas, cuestión que podría estar influenciada por la exclusión de otras variables relevantes que no han sido incorporadas en el modelo. A pesar de esta pequeña acotación, los determinantes resultantes muestran elevadas elasticidades, lo que demuestra el efecto sistemático de estos regresores sobre la probabilidad que en este caso se analiza.

Por último cabe destacar que las elasticidades más altas se encuentran en la primera y la última etapa del proceso educativo que se analiza. No obstante, los resultados relativos a los efectos marginales y las relaciones causales de estas han permitido cuantificar más estrechamente la influencia que variaciones en estos regresores pueden motivar cambios significativos sobre la probabilidad de ocurrencia de las variables dependientes.

Conclusiones

El estudio pretende primeramente identificar los principales determinantes del proceso educativo visto a través de tres variables dependientes claves: el acceso a la enseñanza primaria a los 6 años, la continuidad de estudios a los 15 años y el completamiento de estudios a los 18 años, es decir, se analiza las diferentes decisiones educativas como un proceso dinámico y secuencial. Los modelos estimados apuntan a describir la relación funcional existente entre las variables del sistema, así como cuantificar e interpretar los efectos de cada variable sobre la probabilidad de satisfacer cada una de las decisiones.

Este estudio representa un avance en las siguientes direcciones:

- Se realiza una breve reseña teórica-metodológica y empírica internacional de los estudios realizados sobre los determinantes del funcionamiento educativo en general.
- Explora una compleja base de datos basada en información individual y selecciona las variables a utilizar en dependencia de los estudios realizados y la disponibilidad de información.
- Propone una metodología fácilmente reproducible para el estudio de los determinantes de la educación en Cuba.
- Implementa dicha metodología para el caso específico de Cuba con todas las particularidades y supuestos pertinentes, extrayendo conclusiones para un mejor entendimiento del tema.

Los modelos de elección binaria van más allá de las cuestiones cualitativas relacionadas con los impactos y las relaciones causales entre las variables involucradas, al proveer una medida más precisa (cuantitativa) del grado de importancia de cada uno de estos efectos sobre la probabilidad que se analiza.

En términos empíricos, se aplican tres modelos logit -utilizando datos de la encuesta de la situación económica de los hogares- construidos a partir de tres variables dependientes diferentes con los que se estima, respectivamente, la probabilidad de asistir a la escuela primaria, de continuar los estudios medio básicos y de completar los estudios medio superiores.

El análisis de una primera parte del ciclo educativo permite comparar resultados y establecer diferencias en el efecto que una misma variable podría tener sobre las probabilidades de satisfacer los distintos niveles educativos. Esto es importante porque desde el punto de vista del diseño de políticas es relevante poder determinar el peso de los distintos factores en la determinación de cada uno de los eventos.

Los resultados obtenidos en el análisis empírico desarrollado parecen confirmar que existen diversos factores socio-económicos que deben tenerse en cuenta a la hora de analizar el acceso, la continuidad y el completamiento de estudios. Los resultados más importantes de las estimaciones son los siguientes:

- La escolaridad del núcleo tiene un efecto muy importante sobre la probabilidad de satisfacer las tres decisiones analizadas, constituyendo el determinante más relevante en la determinación de la educación de los hijos.
- La cantidad de personas que viven en el núcleo también es una variable que incide en la probabilidad de que los niños accedan a la primaria. En este sentido, la presencia de personas como los abuelos, tíos y demás familiares cercanos pudiera resultar un apoyo adicional importante en la toma de decisiones relacionadas fundamentalmente con el acceso.
- La tasa de mortalidad infantil de los menores de 5 años, resulta ser esencial en el modelo de acceso a la primaria, al mostrar una elasticidad significativa y negativa.
- La separación entre sexos permite constatar que existen diferencias en la incidencia de esta variable en los últimos dos modelos. Las mujeres tienen mayores probabilidades que los varones de completar sus estudios, mientras que ellos tienen mayores probabilidades que ellas de continuar estudios medios a medios superiores, aunque esta diferencia no es marcada.
- El ingreso es un determinante débilmente significativo con un efecto marginal muy pequeño para las dos primeras decisiones estudiadas. A pesar de que para el tercer modelo, la elasticidad resulta ser un poco más significativa, su efecto sobre la probabilidad de completamiento es inferior que el resto de los determinantes estimados.
- El número de alumnos por profesor en la enseñanza media y media superior y las escuelas existentes en primaria son las variables por el lado de la oferta educativa que presentan los mayores impactos sobre la probabilidad de acceder y continuar los estudios a los 15 años respectivamente. Este resultado resulta particularmente interesante al poder verificarse estadísticamente la importancia que estas tienen en el perfeccionamiento de la política educacional en el país.

Derivado de lo anterior, los factores que cumplen un rol importante en incrementar las probabilidades analizadas, están íntimamente relacionados con las variables que están fuera del sistema educativo, es decir, aquellas vinculadas por el lado de la demanda educativa (escolaridad del núcleo, el tamaño de la familia, ingresos per cápita del hogar, sexo, tasa de mortalidad infantil). Este tipo de trabajo sugiere que cualquier acción, estrategia o política orientada a perfeccionar la política educativa cubana, deberá contemplar este tipo de variables y su relación con la política social desplegada hasta ese momento.

Por otro lado, la puesta en marcha de un conjunto de programas educativos y sociales estratégicos de corto y largo alcance, con el propósito de elevar la cultura general integral de toda la población, está teniendo resultados e impactos favorables sobre las diferentes decisiones educativas analizadas.

La cantidad de alumnos por profesor en los diferentes niveles, la cantidad de escuelas construidas y reparadas, el programa para el perfeccionamiento del nivel de secundaria básica y primaria, la creación de alternativas para garantizar la continuidad de estudios de toda la población y sobre todo de los jóvenes, entre otros, inciden directamente en la probabilidad de acceder, continuar y completar los estudios en tiempo a los 18 años.

Recomendaciones

Este estudio puede verse como una primera aproximación a la identificación y cuantificación del impacto de distintos determinantes socio-económicos sobre la probabilidad de acceso, continuidad y completamiento de los estudios. Difícilmente un solo estudio pueda recoger todos los aspectos relevantes para abordar un tema tan importante, por lo que se espera que las ideas, metodologías y resultados de este trabajo motiven investigaciones futuras que profundicen en el tema.

A tal efecto, existen varias cuestiones a considerar en trabajos futuros:

- Este estudio toma como punto de partida los niveles primario y secundario, que es donde probablemente se observen las interrelaciones más importantes entre las características familiares y la educación. En consecuencia, una primera recomendación podría ser extender los resultados de este estudio a otros niveles, como la educación terciaria y lograr vincular a través de un modelo logit secuencial los distintos niveles educativos.
- En este trabajo se utilizan técnicas econométricas simples con herramientas de fácil implementación e interpretación. En consecuencia, una línea de investigación futura pudiera ser explorar metodologías más complejas como la estimación a través de datos panel o el empleo de modelos probit que respondan a preguntas más específicas.
- Sería interesante incluir el efecto dinámico y cambiante de las variables independientes en el tiempo. Esto sería muy interesante al lograr detectar el efecto sistemático o coyuntural de una determinada variable sobre las diferentes decisiones educativas.
- La búsqueda de otras fuentes de información alternativas y el mejor diseño y procesamiento de las encuestas sobre la situación económica de los hogares, contribuirá a la incorporación de información más detallada de un conjunto de variables independientes que pudieran estar incidiendo en los diferentes modelos y que no han podido ser recogidas en esta oportunidad.
- Algunos resultados de este estudio llaman la atención sobre la estrecha interrelación que existe entre determinadas políticas que se llevan a cabo y el impacto de éstas sobre el fenómeno a estudiar. Por ejemplo, es importante entender la relación entre la reducción de la mortalidad infantil y la educación, los vínculos entre el tamaño y la preparación de la familia con los niveles de educación o cómo los diferentes programas sociales interactúan con los programas educativos. Darse cuenta de éstas interrelaciones puede contribuir notablemente al entendimiento de estos fenómenos y al diseño de las acciones y estrategias correctas.

Bibliografía

- Alvarez, Elena y Mattar, J (coord): “Política Social y reformas estructurales: Cuba a principios del siglo XXI”. Talleres de editores e Impresiones FOC, S.A de C.V, México, 2004.
- Becker, G: “Human capital: a theoretical and empirical analysis with special refence to education”. 3rd edición. University of Chicago Press, 1993.
- Ben-Porath, Y: “The production of Human Capital and the Life Cycle of Earnings”. Journal of Political Economy 75: 352-65, 1967.
- Bertranou, Evelina: “Determinantes del avance en los niveles de educación en Argentina. Análisis empírico basado en un modelo probabilístico secuencial”. Tesis de maestría en economía de la UNPL. Documento de trabajo No. 38, abril 2002.
- Beyer, Harald: “Desempleo juvenil o un problema de deserción escolar”. Serie de documentos de trabajo No. 277. Centro de estudios públicos, febrero 1998.
- Canagarajah, S and Coulombe, H: “Child labor and school attendance in Ghana”. Department for Human development and Social protection. The World Bank: Whashington DC, 1998.
- Caporale, Alejandra: “Educación-Reformas-América latina-Transformación”. Revista Nueva Sociedad Venezuela, No. 65, enero-febrero 2000.
- Carbacho, A: “The effects of family background in schooling enrollment and attainment: the case of Argentina in 1974-1997”. UTDT, 1999.
- Cortés, Fernando: “Regresión logística en la investigación social: potencialidades y limitaciones”. Revista Ciencias Sociales. Departamento Sociología, Universidad de la república Uruguay, 1997. www.rav.edu.uy/fcs/soc/revista/cortes.html.
- Cossa, R: “Determinantes de la asistencia escolar en Argentina: un análisis empírico con extensiones para la evaluación de políticas”. Universidad de Chicago, 2000.
- Di Grecia, L: “Acceso a la educación universitaria: evolución y determinantes para el caso argentino”. Mimeo, 2004.
- Eckstein, Zui and Kenneth, Wolpin: “Why youths drop out of high school: the impact of preferences, opportunities, and abilities”. Econométrica, Vol 67, november 1999.
- Ferran, M: “SPSS para WINDOWS”. McGraw-Hill, Madrid. 1996.
- Ferriol, Angela et. al: “Política Social: el mundo contemporáneo y las experiencias de Cuba y suecia”. Prontográfica, Montevideo, Uruguay. Diciembre 2004.

- Gasparini, Leonardo: “Acceso a la educación y la salud en la provincia de Buenos Aires”. Cuadernos de Economía. No.45, Ministerio de Economía de la Provincia de Buenos Aires, noviembre 1998.
- Gertler, Paul and Glewwe, Paul: “The willingness to pay for education in developing countries: Evidence from Rural Peru”. Journal of Public Economics 42: 251-275, 1990.
- Gertler, Paul and Glewwe, Paul: “The willingness to pay for education for daughters in contrast to sons: evidence from rural Perú”. The World Bank Economic Research, Vol 6, No 1, 1992.
- Heckman, James: “A life-Cycle Model of Earnings, Learning, and Consumption”. Journal of Political Economy, 84 (4): s11-s44, 1976.
- Heckman, James; Lochner, Lance and Todd, Petra: “Fifty years of Mincer earnings regressions”. NBER WP 9732, Cambridge (MA): National Bureau for Economic Research, 2003. <http://www.nber.org/papers/w9732>.
- Instituto Nacional de Investigaciones Económicas: “Objetivos de Desarrollo del Milenio Cuba”. Segundo Informe. Mercie Group-ENPSES, CUJAE. Ciudad de la Habana, Cuba, julio 2005.
- Johnston, Jack and Dinardo, John: “Econometric Method”. Fourth Edition. The McGraw- Hill Companies, Inc.
- Maddala, G: “Limited dependent and qualitative variables in econometrics”. Cambridge University Press, 1983.
- Marchionni, Mariana y Escudero, Walter: “Los determinantes de la decisión de escolarización”. Maestría en Finanzas Públicas Provinciales y Municipales. Universidad Nacional de la Plata, Departamento de economía, 1998.
- Martín, Guillermina et al: “Introducción a la econometría”. Practice Hall Iberia, Madrid, 1997.
- Medina, Eva: “Modelos de elección discreta”, diciembre 2003. www.eva.medinaam.es.
- Orro, Alfonso: “Modelos de elección discreta con coeficientes Aleatorios”. Universidad de la Coruña, 3 de abril de 2003.
- Ponce, Juan and Vos, Rob: “Education” in: World Bank and Inter-American Development Bank (2004) Ecuador: Creating fiscal space for poverty reduction. A fiscal management and public expenditure review, Volume II, Report No. 28911-EC, Washington D.C.: The World Bank and IDB.
- Ponce, Juan; Arjun, Bedi and Vos, Rob: “¿Cómo hacer más eficiente el gasto educativo?”, in: Vos, Rob (et al.) ¿Quién se beneficia del Gasto Social en el Ecuador?, Estudios e Informes del SISESTFS. No. 4, Quito: Secretaría Técnica del Frente Social, Quito: Secretaría Técnica del Frente Social, pp. 89-118, 2003.

- Ripani, Laura: “Determinantes de la decisión de asistencia al nivel superior de educación”. La Plata, UNLP, 2000.
- Sapelli, Claudio: “Deserción Escolar y Trabajo Juvenil ¿Dos Caras de una Misma Decisión?”. Documento de trabajo No. 259. Santiago, diciembre 2003.
- Skyt, Helena: “Child labor and school attendance: two joint decisions”. Center for Labor Market and Social Research, University of Aarhus, Denmark, october, 1998.
- Stata: “Stata Statistical Software: Release 5.0”. College Statio, TX: Stata Corporation, 1997.
- Sosa, Walter y Marchionni, M: “Household Structure, gender and the economic determinants of school attendance in Argentina. Mimeo, World Bank Working Paper, 1999.
- Sosa, Walter: “Tópicos en econometría aplicada”. Notas de clase. Trabajo docente No.2. Departamento economía, universidad Nacional de la Plata, 1999.

ANEXO

Anexo No.1

Education cost-effectiveness model

A2.1 Demand for schooling

A2.2 Empirical specification

A2.3 Costs of attending school

A2.4 Benefits of attending school

A2.5 Cost analysis and budget projections

A2.1 Demand for schooling: enrolment, school attendance and student achievement

The first educational outcome that should probably be considered while analyzing the educational sector is school enrollment patterns. In addition there are several other educational outcomes that need to be considered once the enrollment decision has been made. The framework presented here focuses largely on two of these outcomes. In particular, conditional on enrollment *what are the factors that determine school attendance over the school year* and *what are the determinants of academic achievement?* The framework outlined here links both these outcomes. However, it is also possible to analyze these outcomes individually. Also note that while the focus here is on the attendance decision, it is quite straightforward to modify it to suit the enrollment decision.

The economic contribution of children to families in developing countries (especially in rural areas) and accordingly the opportunity cost associated with school attendance may be substantial. Attendance will suffer when parents perceive that the return associated with time spent in school does not justify the loss of a child's economic contribution. Parental perceptions of school inputs may also affect the attendance cost-benefit calculus as low quality teachers or limited availability of teaching materials may attenuate the expected benefits from attending school. A reduction in days attended probably exerts a negative influence on academic achievement and increases the probability of repetition and desertion.

The framework presented here, and applied in Section 6, explicitly recognizes the link between primary (and secondary) school attendance and student achievement. It is assumed that parents determine the particular pattern of school attendance for their children on the basis of expected gains and the costs of attending school. It is possible to proceed in two steps. First, estimate the effect of child, family and school characteristics on (academic achievement) test scores and obtain predicted test scores. In the second step estimate a school attendance model which includes predicted test scores as a measure of the expected gains of school attendance.

School Attendance and Student Achievement – An Analytical Framework

School attendance patterns in developing countries vary substantially across households. Some children may never enter school while others may attend only part-time. The degree of part-time schooling may vary from missing a few weeks to missing several months. The variation in attendance patterns suggests that parents evaluate differently the costs and benefits of attending school and that this evaluation for the same household may also vary according to the particular time of the year. For instance, during the harvest season the opportunity costs of attending school may far outweigh the benefits, resulting in temporary withdrawal, while at other times the benefits may outweigh the costs and result in regular school attendance. Thus, school attendance over the year may be viewed as the consequence of a daily household decision where a child attends school on a particular day if the expected benefits from attending school on that day are greater than the associated costs.

To formalize these notions and to motivate our empirical work this section presents a framework tailored to our needs. Consider that the school year consists of n days and it is day i of the school year. We assume that each household has a utility function defined over b_i and c_i , where b_i denotes the benefits associated with attending school on day i , and c_i is household consumption on day i . While attending school yields benefits it comes at a cost. Direct and opportunity costs associated with school attendance lower resources available for household consumption. Accordingly, household utility on day i conditional on school attendance (denoted by subscript 1) is given as,

$$U_{i1} = U(b_i, c_{i1}). \quad (1)$$

The associated budget constraint is,

$$y_i = c_{i1} + p_i, \quad (2)$$

where y_i is household income, and p_i represents the total cost associated with school attendance. In a similar fashion the utility associated with not attending school may be defined by,

$$U_{i0} = U(c_{i0}). \quad (3)$$

The budget constraint is $y_i = c_{i0}$. Given the utility associated with both options, households choose the option that yields the highest utility. The solution to the daily unconditional utility maximizing problem is

$$U_i^* = \max(U_{i1}, U_{i0}) \quad (4)$$

where U_i^* is the maximum utility. Alternatively, school attendance may be defined in terms of a dichotomous variable, a_i , where $a_i = 1$ if a child attends school and 0 otherwise. A child attends school i.e., $a_i = 1$ if $U_{i1} > U_{i0}$. Summing up the outcomes of these daily decisions, over the school year, leads to the observed pattern of school attendance.

A2.2 Empirical Specification

Since our purpose is to empirically explore the role of expected gains and costs on the school attendance decision we proceed by specifying linear forms of the conditional utility function. For the schooling option,

$$U_{i1} = \beta_1 b_i + \beta_2 c_{i1} + \varepsilon_{i1} \quad (5)$$

where the β 's are coefficients to be estimated and ε_{i1} is assumed to be a mean zero, normally distributed error term with positive variance. Since $c_{i1} = y_i - p_i$, we may rewrite (5) to obtain,

$$U_{i1} = \beta_1 b_i + \beta_2 (y_i - p_i) + \varepsilon_{i1} \quad (6)$$

The utility function for the non-schooling option is,

$$U_{i0} = \beta_2 y_i + \varepsilon_{i0} \quad (7)$$

Thus, an individual attends school, i.e. $a_i = 1$ if $\beta_1 b_i - \beta_2 p_i + \varepsilon_{i1} - \varepsilon_{i0} > 0$.

The chances of attending school on a particular day may be expressed in terms of a linear probability model that may be written as,

$$a_i = \beta_1 b_i - \beta_2 p_i + \varepsilon_{ia}, \quad (8)$$

where ε_{ia} is a normally distributed, mean zero, positive variance composite error term. Equation (8) depicts the probability of attending school on any particular day. Since we are interested in the yearly pattern of school attendance we may sum up the outcome of the daily attendance decision over the course of the school year,

$$\sum_{i=1}^n a_i = \sum_{i=1}^n (\beta_1 b_i - \beta_2 p_i + \varepsilon_{ia}) \quad (9)$$

to yield,

$$A = \beta_1 B + \beta_2 P + \varepsilon_A, \quad (10)$$

where yearly school attendance A , depends on B , the expected benefits associated with school attendance over the school year and P , the yearly costs of attending school.

A2.3 Costs of attending school

The total cost (P) of sending a child to school includes monetary (direct) and indirect or opportunity costs. Since education is largely subsidized the main cost incurred by households is likely to be in the form of opportunity costs. Attending school reduces a child's availability for work in and outside the home. If a child makes substantial contributions to family income, or plays an important role in supporting other working members, then the opportunity cost of attending school is likely to be high and this may curtail the attractiveness of the schooling option.

Both of these cost components are likely to differ across households. For instance, direct costs may vary due to differences in transportation costs. Opportunity costs and the value of a child's time may also differ due to personal characteristics of the child (age, sex) and the value that parents place on a child's time. Since we do not directly observe the costs of attending school we allow P to depend on a vector of child, family and other characteristics that capture the cost of attending school.

A2.4 Benefits of attending school

Parents have to ascertain the total benefits (B) associated with school attendance. We consider two types of benefits that may influence parental decision-making. The main benefit associated with attending school is likely to be the expected addition to a child's human capital. To capture this effect we need a measure of the human capital gains associated with school attendance. For this study we incorporate a measure that is widely used to indicate the benefits derived from education: test scores.

However, using actual test scores is obviously incorrect due to the potential endogeneity between test scores and attendance. In order to derive an appropriate measure of human capital gains we proceed by estimating educational production functions, one for each subject, for those students for whom test score data are available. These test score equations are specified as,

$$H = \delta Z + \varepsilon_H, \quad (11)$$

where H is a measure of human capital or in this case test scores, Z is a vector of individual, family and school characteristics that influence H , and ε_H is an error term. Estimates from the educational production functions are used to predict test scores for each individual. These predicted values (\hat{H}) are included in equation (10) in order to capture the human capital benefits associated with attending school.

In addition to school characteristics that have an impact on test scores, there may be other school characteristics that do not affect academic achievement but do signal the quality (Q) of a school and directly influence the benefits that parents attribute to school attendance.

For instance, whether a school attendance decision. Household endowments are assumed to influence the school attendance decision through opportunity costs.

However, these are easily observed signals that may be used by parents to judge the quality of a school and in turn may directly influence the benefits that parents associate with school attendance. Thus, some school inputs and facilities may directly influence parental evaluation of the benefits associated with school attendance while others may exert an influence on benefits through their impact on test scores.

To account for the different kinds of benefits that parents may associate with school attendance, equation (10) may be adjusted to accommodate both the expected human capital benefits (H) and direct benefits (Q) and may be rewritten as,

$$A = \beta_{1H}\hat{H} + \beta_{2Q}Q + \beta_3P + \varepsilon_A, \quad (12)$$

where β_{1H} is a coefficient to be estimated and β_{2Q} and β_3 are conformable coefficient vectors to be estimated. As this equation depicts, school attendance is treated as a function of expected human capital benefits, other benefits and costs.

A2.5 Cost Analysis

Recent contributions to the school effectiveness literature have emphasized the importance of incorporating cost-effectiveness analysis into school quality research designs. While there is little doubt that the identification of factors that affect learning outcomes is important, its usefulness will be considerably enhanced if accompanied by more information on the costs associated with each intervention. On the basis of the analysis suggested in the previous section it should be possible to identify the most important factors influencing educational outcomes. Combing this information with the cost of each intervention would allow an evaluation of the most efficient method of increasing achievement and attendance. This offers a method of comparing alternative policies that can be ranked in terms of achievement gains per shilling invested in each input. Examples of such static costeffectiveness analyses are available in e.g. Bedi and Marshall (1999) and Harbison and Hanushek (1992).