



Munich Personal RePEc Archive

**Are the subsidized particular schools
better than the public schools?
Estimation of the equation of scholastic
profit for Chile**

Idrovo Aguirre, Byron

Centro de Economía Aplicada - Universidad de Chile

3 December 2007

Online at <https://mpra.ub.uni-muenchen.de/10665/>
MPRA Paper No. 10665, posted 29 Sep 2008 15:24 UTC

¿Son las escuelas particulares subvencionadas mejores que las municipales? Estimación de la ecuación de logro escolar para Chile*

Byron Idrovo A.^ψ

Resumen

Este trabajo estima el desempeño relativo de colegios municipales (MUN) y particulares subvencionados (PS) con distintas estrategias de identificación, ordenadas desde la más básica –suponiendo la ausencia de sesgo de selección y características multiniveles en las observaciones– hasta aquella que relaja tales supuestos. La fuente de información utilizada es la prueba SIMCE de matemáticas rendida por los alumnos de 4° año básico en el 2002, las características de los estudiantes y de sus familiares que recoge el SIMCE en la encuesta de padres, y la información sobre los establecimientos educacionales que entrega el Ministerio de Educación.

Dependiendo de la metodología utilizada, las estimaciones revelaron que un estudiante que asiste a un colegio particular subvencionado, obtiene una ganancia en la prueba SIMCE de matemáticas que fluctúa entre 11,6 puntos (estimado por Mínimos Cuadrados Ordinarios) y 17,3 puntos (estimado por Mínimos Cuadrados en dos etapas con errores robustos) versus si el estudiante hubiese asistido a un colegio municipal. Por su parte, las estimaciones que utilizan variables instrumentales muestran que, dependiendo del método usado, un estudiante que asiste a un colegio particular subvencionado obtiene entre 14,1 puntos y 17,3 puntos adicionales en la prueba SIMCE de matemáticas que si hubiese asistido a un colegio municipal. Con relación a la variabilidad de las estimaciones, las estrategias de identificación más básicas muestran una menor diferencia en el desempeño escolar entre colegios particulares subvencionados y municipales. Por otra parte, entre las técnicas que corrigen por las características multiniveles y sesgo de selección en las observaciones, el método de Heckman –asumiendo distintos retornos a las características de los colegios MUN y PS, así como también distintas especificaciones anidadas de variables instrumentales– reportó una menor variabilidad en la estimación del desempeño relativo de ambos tipos de colegios. En tanto, para una determinada especificación de variables instrumentales, todas las técnicas revelaron marcadas diferencias en sus resultados, y aquellas que controlaron por el hecho que los estudiantes están anidados en establecimientos (clusters) exhibieron mayor varianza en la estimación.

Finalmente, si bien la presente tesis se focaliza en las diferencias de puntaje entre los dos tipos de colegios, los coeficientes estimados para las restantes variable (características de los alumnos y establecimientos) coinciden con lo esperado y con lo encontrado en otros estudios. No obstante, cabe mencionar que una limitación del presente estudio es que no controla por efecto pares (*peer effects*).

Palabras Clave: Desempeño escolar, estimación paramétrica, sesgo de selección.

Clasificación JEL: C13, C31, I21

* Este trabajo fue presentado como tesis de grado para la obtención del título de Magister en Economía Aplicada de la Universidad de Chile. Biblioteca Universidad de Chile www.cybertesis.cl. Se agradecen los comentarios de los académicos Alejandra Mizala, Andrea Repetto, Pablo García, y Mauricio Farías.

^ψ Economista. Gerencia de Estudios de la Cámara Chilena de la Construcción. bidrovo@cchc.cl.

I. Introducción

El desempeño relativo de colegios municipales y particulares subvencionados ha sido objeto de recientes debates en el ámbito académico a nivel nacional e internacional. En su gran mayoría, los debates giran en torno a la efectividad de la política de elección de colegios (school choice). La hipótesis detrás de la elección de colegios es que los colegios particulares subvencionados poseen un mejor desempeño que los colegios municipales (Sapelli, 2003). Esta hipótesis supone que la elección racional de los padres –basada en el desempeño de los establecimientos– induce a la competencia entre colegios municipales y particulares, y a su vez la competencia conduce a mejorar la calidad de la educación entre ambos colegios (Aedo y Sapelli, 2001). En América Latina, Chile ha sido pionero en instaurar un sistema universal de elección escolar a través de un esquema de subvenciones similar a lo que se conoce en la literatura como sistema de “voucher”¹. Este sistema entrega a los estudiantes la opción de ser atendidos en colegios particulares subsidiados por el gobierno² (Mizala y Romaguera, 2000, González, 2000).

Los estudios realizados en Chile pueden clasificarse en dos grupos, los primeros que se realizaron cuando sólo existían datos agregados a nivel de colegios³ y los posteriores que utilizan datos a nivel de alumnos. Con datos a nivel de alumnos, Mizala y Romaguera (2001), y Sapelli y Vial (2002) coinciden en que los colegios privados subvencionados presentan un mejor desempeño que los colegios municipales. Recientemente, Anand, Mizala y Repetto (2007) demuestran que estudiantes becados en escuelas privadas subvencionadas con financiamiento compartido obtienen un desempeño levemente superior al que obtendrían si fuesen atendidos por una escuela pública.

¹ La actual configuración del sistema educativo se fundamenta en las reformas implementadas en 1981 que incluyó la descentralización del sistema de educación público. A esta reforma también se suma la creación de pruebas estandarizadas para medir el rendimiento de los estudiantes como indicador de la calidad entregada por cada establecimiento (e.g. el Sistema de Medición de la Calidad de la Educación, SIMCE).

² El sistema de elección escolar en Chile, entrega a los familiares la opción de elegir una escuela privada subvencionada. Por otro lado, los colegios privadas y las municipales con exceso de demanda son libres de ser selectivas en sus admisiones.

³ Con datos agregados por colegios, Mizala y Romaguera (2000) mostraron que luego de considerar a todos los colegios y variables socioeconómicas del alumno, las diferencias en el desempeño de colegios municipales y particulares subvencionados no difería significativamente. Tokman (2002) mostró que los colegios municipales tienen un mejor desempeño que los privados subvencionados en el caso de alumnos de bajo nivel socioeconómico. Por otra parte, McEwan (2001) encontró que los colegios particulares pagados y los colegios católicos subvencionados mostraban un mejor desempeño que los colegios municipales. Sin embargo, no encontraron diferencias entre colegios particulares subvencionados no religiosas y colegios municipales.

El siguiente cuadro resume las diferencias que se observan en el desempeño escolar con distintas técnicas de estimación, base de datos y variables utilizadas por un grupo de trabajos de investigación en esta materia.

| Cuadro 1 | | | | |
|--|---|---|--|-----------|
| Cuadro resumen de las principales características y resultados de algunos estudios recientes | | | | |
| Autor (año) | Prueba SIMCE utilizada | Técnica de estimación | Diferencia estimada del rendimiento de escuelas PS vs MUN (error estándar) | |
| | | | ATE PS-MUN=9.6 (0.1) | |
| Sapelli y Vial (2002) | - Prueba estandarizada de matemáticas y lenguaje del Sistema de Medición de Calidad de la Educación (SIMCE), aplicado a la población de estudiantes del 4to grado básico y escuelas en 1999. | - Efecto promedio de tratamiento (Average Treatment Effect, ATE) - Efecto de tratamiento sobre los tratados (Treatment on the treated, TT) por quintiles de ingreso | ATE | |
| | | | PS-MUN=9.6 (0.1) | |
| | | | TT | |
| | | | Quintiles | |
| | | | 1er | 5.7 (3.0) |
| | | | 2do | 6.1 (1.4) |
| 3er | 7.4 (1.5) | | | |
| 4to | -3.6 (1.7) | | | |
| 5to | -75.2 (2.9) | | | |
| Sapelli (2003) | - Prueba estandarizada de matemáticas del Sistema de Medición de Calidad de la Educación (SIMCE), aplicado a la población de estudiantes del 4to grado básico y escuelas en 1999. - Encuesta de padres y apoderados, que por ser contemporánea al momento de la evaluación SIMCE, permite asociar las características socioeconómicas del estudiante con su rendimiento escolar. | - Mínimos Cuadrados Ordinarios corrigiendo por sesgo de selección. | PS-MUN=39.46 (3.54) | |
| Mizala, Romaguera, y Ostoić (2004) | - Prueba estandarizada de matemáticas del Sistema de Medición de Calidad de la Educación (SIMCE), aplicado a la población de estudiantes del 4to grado básico y escuelas en 1999. - Encuesta de padres y apoderados, que por ser contemporánea al momento de la evaluación SIMCE, permite asociar las características socioeconómicas del estudiante con su rendimiento escolar. | Modelo Lineal Jerárquico (Hierarchical Linear Model, HLM) | PS-MUN= 3.986 (0.755). | |
| Sapelli y Vial (2005) | - Prueba estandarizada de matemáticas y lenguaje del Sistema de Medición de Calidad de la Educación (SIMCE), aplicado a la población de estudiantes del 4to grado básico y escuelas en 2002. | - Mínimos Cuadrados en dos etapas (a la Heckman) - Efecto promedio de tratamiento (Average Treatment Effect, ATE) - Efecto de tratamiento sobre los tratados (Treatment on the treated, TT) | ATE, controlando por características pares: - Efecto de matricular a un estudiante en una escuela municipal PS-MUN=21.4 (8.1) - Efecto de matricular a un estudiante en una escuela municipal, junto con sus pares PS-MUN=17.1 (7.7) TT, controlando por características pares: - Efecto de trasladar a un estudiante de una escuela privada subvencionada a | |

| Cuadro 1 | | | |
|--|---|---|---|
| Cuadro resumen de las principales características y resultados de algunos estudios recientes | | | |
| Autor (año) | Prueba SIMCE utilizada | Técnica de estimación | Diferencia estimada del rendimiento de escuelas PS vs MUN (error estándar) |
| | | | una escuela municipal PS-MUN=31.1 (5.9) |
| | | | - Efecto de trasladar a un estudiante junto con sus compañeros a una escuela municipal PS-MUN=28.4 (5.9) |
| Anand, Mizala y Repetto (2007) | - Prueba estandarizada de matemáticas y lenguaje del Sistema de Medición de Calidad de la Educación (SIMCE), aplicado a la población de estudiantes del 4to grado básico y escuelas en 2002. - Encuesta de padres y apoderados, que por ser contemporánea al momento de la evaluación del estudiante, permite recolectar las características socioeconómicas del estudiante. | - Mínimos Cuadrados Ordinarios corrigiendo por sesgo de selección (controlando por la probabilidad que el alumno reciba una beca) - Match propensity scores. | PS-MUN=entre 8.8 y10.8 |

Fuente: Elaboración propia.

El estudio del desempeño relativo de los colegios municipales y particulares con subvención del Estado presenta importantes desafíos metodológicos, el más relevante es la dificultad para resolver el sesgo de selección presente al comparar a los estudiantes de estos dos tipos de colegios (Mizala y Romaguera 2000; Contreras, 2001; Tokman, 2002; Anand, Mizala y Repetto, 2006). Este sesgo de selección se debe a que hay elección de colegios por parte de los padres y hay elección de estudiantes por parte de los colegios (Stephen y Whitney, 2001).

Este trabajo estima el desempeño relativo de colegios municipales (MUN) y particulares subvencionados (PS) con distintas especificaciones anidadas de variables instrumentales y estrategias de identificación, ordenadas desde la estrategia más básica –es decir, aquella que no corrige por sesgo de selección y características multiniveles en las observaciones– hasta aquella que exige más de los datos: Mínimos Cuadrados Ordinarios (Ordinary Least Squares, OLS); Modelo Lineal Jerárquico⁴ (Hierarchical Linear Model, HLM); Método de corrección de Heckman –asumiendo distintos retornos entre los tipos de establecimientos; Estimación del Treatment Effect con Variables Instrumentales (Instrumental Variables, IV) –con y sin controlar por la presencia de clusters de

⁴ Wong y Entwistle (1983); Goldstein (1991); Bryk, A. and S. Raudenbush. (1992); Yang et al. (2000); Moerbeek y Wong (2002); Mizala Romaguera y Ostoic (2005).

escuelas⁵, y aplicando Mínimos Cuadrados en dos etapas con errores robustos (Two Stage Least Squares, TSLS). Por otra parte, para analizar las diferencias de desempeño escolar, una vez corregido por sesgo de selección y asumiendo distintos retornos entre los colegios MUN y PS, se estima el efecto promedio de tratamiento (average treatment effect, ATE) y el tratamiento sobre los tratados (treatment on the treated, TT)⁶. La muestra utilizada en este estudio corresponde a los puntajes de la prueba SIMCE de matemáticas, y la información sobre características de los estudiantes, sus familiares y los establecimientos educativos del nivel 4° básico del año 2002.

Esta investigación muestra que la diferencia del rendimiento académico de colegios MUN y PS es sensible a los instrumentos y al método utilizado. Los datos corresponden a los puntajes de la prueba SIMCE de matemáticas, y la información sobre características de los estudiantes, sus familiares y los establecimientos educativos del nivel 4° básico del año 2002.

Entre los principales hallazgos, se tiene que las estrategias de identificación que obvian las características multiniveles y sesgo de selección en las observaciones, tienden a estimar una menor diferencia en el desempeño escolar entre ambos tipos de establecimientos. Por otra parte, entre las técnicas que exigen más de los datos, el método de Heckman –asumiendo distintos retornos a las características de los estudiantes de colegios MUN y PS– mostró una mayor estabilidad en las estimaciones de las distintas especificaciones anidadas de variables instrumentales. En tanto, para una determinada especificación de variables instrumentales, todas las técnicas revelaron diferencias en sus resultados, y aquellas que controlaron por clusters de establecimientos exhibieron una mayor varianza en la estimación. En efecto, dependiendo del método utilizado, las estimaciones revelaron que un estudiante atendido por un colegio particular subvencionado, obtiene una ganancia entre 12,8 (estimado por OLS) y 17,3 (estimado por TSLS con errores robustos) puntos adicionales en la prueba SIMCE de matemáticas (versus si el estudiante hubiese asistido a un colegio municipal). No obstante, las estimaciones obtenidas de métodos que corrigen por sesgo de selección mostraron menores diferencias. En este caso las estimaciones revelaron que un estudiante atendido por un colegio particular subvencionado, obtiene una ganancia entre 14,1 (estimado por IV) y 17,3 (estimado por TSLS con errores robustos) puntos adicionales en la prueba SIMCE de matemáticas (versus si el estudiante hubiese asistido a un colegio municipal). Finalmente, cabe señalar que si bien el presente

⁵ Permite corregir los problemas de heteroscedasticidad del desempeño entre los colegios y la correlación del rendimiento escolar al interior de los colegios que típicamente están presentes en datos multiniveles de corte transversal. Manski (1993).

⁶ Angrist, 2003.

estudio se focaliza en las diferencias de puntaje entre los dos tipos de colegios, los coeficientes estimados para las restantes variables (características de los alumnos y establecimientos) coinciden con lo esperado y con lo encontrado en otros estudios.

El presente estudio se divide en cinco capítulos: además de esta introducción del documento; el segundo capítulo comprende el marco teórico; el tercer capítulo describe las variables utilizadas y el análisis descriptivo de los datos; en el cuarto capítulo se presenta el rendimiento escolar estimado con las distintas estrategias de identificación; finalmente, el capítulo cinco concluye.

II. Metodología

A continuación se discuten las metodologías que permiten identificar los determinantes del desempeño escolar y testear la diferencia en la calidad educativa entre colegios municipales y particulares subvencionados.

a. Modelo Lineal Jerárquico

Uno de los modelos que usaremos para testear las diferencias en el rendimiento escolar entre colegios MUN y PS, es el Modelo Lineal Jerárquico (Hierarchical Linear Model, HLM) en dos niveles⁷. El HLM permite responder acerca de cómo las organizaciones afectan a los individuos dentro de ellas. En el presente estudio, las organizaciones son los establecimientos educativos y los individuos son los estudiantes. En este sentido, el HLM permite separar el efecto que tienen las características de los estudiantes (estimación de primer nivel) del efecto que tienen las características de los colegios (estimación de segundo nivel) sobre el rendimiento escolar.

En la estimación de primer nivel, el desempeño escolar es afectado por las características propias del estudiante (género, características familiares, educación de los padres, ingreso familiar, etc), y se presume que la relación entre el rendimiento escolar y dichas características varían entre las escuelas. Por ello, la técnica de primer nivel consiste en estimar una ecuación de logro para cada escuela, y capturar el efecto-estudiante sobre el rendimiento escolar al interior de las escuelas. En un segundo nivel, los coeficientes de regresión del primer nivel son considerados como variables explicadas (o variables dependientes). Los determinantes de las variables dependientes del segundo nivel son hipotéticamente las características del establecimiento (colegios de jornada escolar completa, PS,

⁷ Seitzer 1995; Bryk y Raudenbush 1992; Raudenbush and Bryk 1992; Goldstein 1987; Raudenbush and Bryk 1986

MUN, mixtos, religiosos, etc). En este caso, la estimación de segundo nivel permite capturar el efecto total que tienen las características de los colegios (efecto-colegio) sobre el rendimiento escolar entre los establecimientos. En resumen, el presente estudio utiliza HLM para analizar, luego de controlar por las características de los estudiantes, el efecto total que tienen los colegios PS (versus los colegios MUN) sobre el rendimiento escolar.

Formalmente, el resultado del primer nivel es obtenido por la estimación de la siguiente ecuación de logro estocástica (Nivel 1).

$$SIMCE_{ij}^{mat} = \beta_{oj} + \sum_{k=1}^K x_{ijk} \beta_{kj} + \varepsilon_{ij} \begin{cases} i = 1, 2, \dots, I_j \\ j = 1, 2, \dots, J \\ k = 2, 3, \dots, K \end{cases} \quad [1]$$

Donde, $SIMCE_{ij}^{mat}$ es el resultado del test en matemáticas del estudiante i , atendido en la escuela j ; x_{ijk} son variables explicativas en la especificación del HLM (ver Anexo 1). Por su parte, ε_{ij} es el componente de error aleatorio de la ecuación de primer nivel. El componente de error captura todas las características no observables de los alumnos que afectan el rendimiento escolar (e.g. motivación individual). Los parámetros $(\beta_{1j}, \beta_{2j}, \dots, \beta_{Kj})$ que preceden a las variables del lado derecho de la ecuación 1, constituyen la tasa de contribución de cada variable explicativa sobre el rendimiento escolar del alumno. HLM asume que estos parámetros difieren entre las escuelas. Luego de controlar por las características de los estudiantes, el parámetro β_{oj} es el promedio condicional del rendimiento escolar de la escuela j , tal como se muestra a continuación:

$$SIMCE_{ij}^{mat} - \sum_{k=1}^K x_{ijk} \beta_{kj} = \beta_{oj} + \varepsilon_{ij} \quad [1.1]$$

$$\text{Tomando esperanza matemática: } E\left(SIMCE_{ij}^{mat} - \sum_{k=1}^K Ex_{ijk} \beta_{kj}\right) = \beta_{oj} \quad [1.2]$$

Formalmente, los parámetros β 's se determinan tal que minimicen la suma de los errores al cuadrado,

$$\forall j \in N, \left\{ \beta_{0j}, \sum_{k=2}^K \beta_{kj} \right\} = \arg \min \sum_{i=1}^I \sum_{j=1}^J \varepsilon_{ij}^2 \quad [2]$$

Los parámetros del primer nivel son explicados por las características de los establecimientos en el modelo HLM del segundo nivel, i.e., en la estimación de segundo nivel, los parámetros del primer nivel son las variables dependientes y las características de los colegios son las variables independientes, tal como se muestra a continuación:

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + \gamma_{01}PS_j + \sum_{k=1}^K \gamma_{0k}E_{jk} + \mu_{0j} \\ \beta_{kj} &= \gamma_{k0} \forall k \in [1, K] \\ \mu_{0j} &\sim iid(0, \sigma_{\mu_0}^2)\end{aligned}\quad [3]$$

Donde, $\{\gamma_{00}, \gamma_{10}, \dots, \gamma_{K0}\} \in \Theta$ son los coeficientes que miden el efecto total que tienen las características de las escuelas sobre el rendimiento escolar; y Θ es el espacio paramétrico. El parámetro γ_{01} mide el efecto total que tienen los colegios PS_j sobre el rendimiento escolar promedio de los diferentes establecimientos (versus el efecto de los colegios MUN). PS_j es una variable dicotómica que toma el valor 1 si la escuela j es privada subvencionada y valor 0 si es municipal. Por ello, si la variable PS_j es significativa, entonces el coeficiente γ_{01} mide, luego de controlar por el efecto-estudiante, la ganancia que obtendría un estudiante atendido por una escuela PS en la prueba SIMCE de matemáticas (versus si hubiese asistido a una escuela MUN). Agrupando el primer y segundo nivel de la estimación HLM, la siguiente ecuación permite mostrar formalmente la diferencia en el desempeño escolar (parámetro γ_{01}) de colegios PS y MUN.

$$SIMCE_{ij}^{mat} = \gamma_{00} + \gamma_{01}PS_j + \sum_{k=1}^K \gamma_{0k}E_{jk} + \sum_{p=1}^P \gamma_{0p}x_{ijk} + \mu_{0j} + \varepsilon_{ij} \quad [4]$$

Una vez estimado el HLM en dos niveles, las diferencias en el rendimiento escolar entre colegios PS y MUN se testean formalmente mediante la siguiente hipótesis:

$$\begin{aligned}H_0 &: \gamma_{01} = 0 \\ H_1 &: \gamma_{01} \neq 0\end{aligned}$$

Si $\gamma_{01} = 0$, entonces no hay suficiente evidencia para concluir que los colegios PS son mejores que los colegios MUN. Por otra parte, si $\gamma_{01} \neq 0$, entonces se concluye que el desempeño de los colegios PS es superior en γ_{01} puntos que los colegios MUN.

Una característica importante de la muestra es que los estudiantes no están distribuidos aleatoriamente en los distintos colegios, i.e., existen factores no observables para el investigador (e.g. habilidad cognitiva y motivación del estudiante), que explican la elección de los familiares por un determinado colegio y la elección de estudiantes por parte de los establecimientos. Esto es lo que se conoce en la literatura como sesgo de selección, el cual impide identificar la verdadera diferencia en el desempeño escolar ya que las estimaciones de la ecuación de logro son sesgadas e inconsistentes.

Existen varios estudios que tratan de resolver el problema de sesgo de selección⁸. Para ello, emplean una regresión auxiliar (ecuación de selección) con variables instrumentales que aproximan los factores no observables de la selección escolar. A continuación dos técnicas estadísticas serán aplicadas para resolver el problema de sesgo de selección: Variables Instrumentales (IV) y Corrección del sesgo de selección propuesto por Heckman (1979).

b. Variables instrumentales

Las variables explicativas de un modelo de regresión podrían no ser completamente exógenas (problema de endogeneidad). Por ejemplo, el que un alumno asista a una escuela PS o MUN no es una decisión del todo exógena para los familiares, i.e., existen factores socioeconómicos y atributos de la escuela que influyen en la decisión de los padres al momento de matricular a sus hijos en un determinado establecimiento. Por otra parte, las escuelas también son selectivas en sus postulaciones. Esto quiere decir que los alumnos no están asignados aleatoriamente en los diferentes colegios.

Las estimaciones con problemas de endogeneidad son inconsistentes. No obstante, el problema de endogeneidad puede ser resuelto si se dispone de instrumentos válidos⁹. Para ello, los instrumentos elegidos no deben ser las variables que ya estén contenidas en la ecuación de logro escolar; las variables instrumentales deben estar relacionadas con la variable candidata a instrumentalizar (variable con problema de endogeneidad); finalmente, los instrumentos no deben estar fuertemente correlacionado con el término de error de la ecuación de logro escolar.

Un importante problema en la estimación con variables instrumentales (IV) es la identificación de instrumentos válidos, pues los estimadores IV son altamente sensibles a la selección de los

⁸ Dan D. Goldhaber y Eric R. Eide (2003); Gallego (2002); Fligio D. y Stone J. (1999); Jimenes E. Lockheed M., y Wattanawaha N. (1988); Goldberger and Cain (1982); Anand, Mizala y Repetto (2007).

⁹ Amemiya T. (1985); Stephen G., Whitney K. (2001).

instrumentos¹⁰. Por ello, el presente estudio utiliza el test de Sargan (1988) como un estadístico de prueba para contrastar la validez de los instrumentos. La definición del estadístico de prueba del test de Sargan se presenta en el Anexo al final del documento.

La variable a instrumentalizar es PS_{ij} que toma el valor 1 si el estudiante es atendido por una escuela PS y 0 en caso de ser atendido por una escuela MUN. La significancia estadística de esta variable latente marcará la diferencia en el rendimiento escolar de un estudiante atendido por una escuela PS (versus si hubiese asistido a una escuela MUN).

El modelo a estimar es:

$$SIMCE_{ij}^{mat} = \beta_0 + \beta_1 PS_{ij} + \sum_{k=4}^K q_{ijk} \beta_k + \tau_{ij} \quad [5]$$

$$PS_{ij} = \Phi(z'_{ij} \theta) + \eta_{ij} \quad [6]$$

Propiedad I: las variables instrumentales z no están incluidas en la ecuación de logro escolar (ecuación 5).

Propiedad II: $\rho_{z_{ij}, PS_{ij}}$ (coef. de correlación entre z_{ij} y PS_{ij}) $\cong 1$

Propiedad III: $z_{ij} \perp \tau_{ij}$

Supuestos: $z_{ij} \perp \eta_{ij}$; $\tau_{ij} \perp \eta_{ij}$

$(\beta_0, \beta_1, \dots, \beta_K)$ son parámetros que acompañan a las variables de la ecuación de logro (ecuación 5). Cada parámetro mide el grado de influencia que tienen las variables explicativas sobre el resultado de la prueba SIMCE en matemáticas ($SIMCE_{ij}^{mat}$) de un alumno i atendido en el establecimiento j . PS_{ij} es un variable latente que indica si un alumno pertenece a un establecimiento PS ($PS_{ij} = 1$) o MUN ($PS_{ij} = 0$); $\{q_{ijk}\}_{k=4}^K$ es el conjunto de variables de la ecuación de logro escolar (ver cuadro 2 y 3); y τ_{ij} es el término de error que captura los componentes aleatorios de la ecuación de logro escolar. La ecuación 6, es la regresión con IV cuya variable endógena es PS_{ij} (variable a

¹⁰ Stephen G. Donald K. Whitney K, Econometría. (2001)

instrumentalizar) y el conjunto de variables $\{z_{ijk}\}_{k=1}^K$ son los instrumentos¹¹. Debido a que PS_{ij} es una variable latente, la ecuación de IV será estimada como un modelo no lineal de probabilidad (*modelos Probit*).

Si la diferencia en el rendimiento escolar de los alumnos atendidos por colegios PS es representativamente mayor al rendimiento escolar de los alumnos atendidos en colegios MUN, entonces el coeficiente estimado que acompaña a la variable PS_{ij} es significativo y mayor que cero ($\hat{\theta}_{IV} > 0$). Formalmente, y asumiendo iguales retornos a las características de estudiantes de colegios PS y MUN se tiene:

$$SIMCE_{ij}^{mat} = \begin{cases} \theta_0 + \theta_{(IV)} E(PS_{ij}/z_{ijk}) + \sum_{k=2}^K q_{ijk} \theta_k + \tau_{ij} & \text{si } PS_{ij} = 1 \\ \theta_0 + \sum_{k=4}^K q_{ijk} \theta_k + \tau_{ij} & \text{si } PS_{ij} = 0 \end{cases}$$

Por lo tanto:

$$E(SIMCE_{ij}^{mat} / E(PS_{ij} = 1/z_{ij}), q_{ij1}, \dots, q_{ijK}) - E(SIMCE_{ij}^{mat} / E(PS_{ij} = 0/z_{ij}), q_{ij1}, \dots, q_{ijK}) = \theta_{(IV)}$$

$$\hat{SIMCE}_{ij}^{mat} \Big|_{\hat{PS}_{ij}=1} - \hat{SIMCE}_{ij}^{mat} \Big|_{\hat{PS}_{ij}=0} = \hat{\theta}_{(IV)} > 0$$

Sabemos que la estimación por variables instrumentales corrige el problema de la asignación no aleatoria de los estudiantes en las diferentes escuelas (problema de endogeneidad de la variable PS_{ij}).

c. Método de Corrección de Heckman

El método de Corrección de Heckman (1979), consiste en estimar simultáneamente la ecuación de logro (ecuación 5) y una ecuación de selección. La ecuación de selección captura los determinantes de la probabilidad de que un estudiante pertenezca a una escuela MUN o PS. La ecuación de selección incluye, además de los instrumentos, a las variables de la ecuación de logro. El Método de Corrección de Heckman se plantea tal que todas las variables de la ecuación de logro están contenidas en la ecuación de selección, pero no todas las variables de la ecuación de selección están contenidas en la

¹¹ El set de IV son: número de colegios PS y M en la comuna de residencia del alumno (Gallego 2005) y las razones que motivaron a los padres a matricular a sus hijos en el establecimiento en que son atendidos (Mizala, Romaguera y Ostoic 2005).

ecuación de logro. Esto se explica porque no todas las variables de la ecuación de selección son relevantes para explicar el rendimiento de los estudiantes.

La ecuación de logro a estimar es similar a la descrita en el capítulo anterior (ecuación 5). El Método de Corrección de Heckman supone que los residuos de la ecuación de logro escolar se distribuyen normalmente. El presente estudio utiliza esta metodología para la estimación por separado de dos ecuaciones de logro (una para colegios MUN y otra para colegios PS) con su respectiva ecuación de selección. Goldhaber et. al (2003), sostiene que estimar las regresiones por separado permite modelar diferentes retornos a las características de los estudiantes y colegios. En otras palabras, al tener una ecuación para cada dependencia, permite que los retornos a sus características puedan ser diferentes entre colegios PS y MUN.

Formalmente, las ecuaciones 7 y 8 corresponden a las ecuaciones de logro de los colegios PS y MUN respectivamente.

$$SIMCE_{ij}^{matPS} = \beta_{0j}^{PS} + \sum_{k=3}^K q_{ijk} \beta_k^{PS} + \tau_{ij}^{PS} \quad [7]$$

$$SIMCE_{ij}^{matM} = \beta_{0j}^M + \sum_{k=3}^K q_{ijk} \beta_k^M + \tau_{ij}^M \quad [8]$$

Las ecuaciones de selección capturan los factores que explican la pertenencia de un estudiante a una escuela PS o MUN. Formalmente se tiene:

$$PS_{ij} = \alpha_0^{PS} + \sum_{k=1}^K q_{ijk} \alpha_k^{PS} + \sum_{z=1}^Z z_{ijz} \delta_z^{PS} + \mu_{ij}^{PS} \quad [8.1]$$

$$MUN_{ij} = \alpha_0^M + \sum_{k=1}^K q_{ijk} \alpha_k^M + \sum_{z=1}^Z z_{ijz} \delta_z^M + \mu_{ij}^M \quad [8.2]$$

PS_{ij} es una variable latente que toma el valor 1 si el alumno i de la escuela j es atendido por una escuela PS y 0 en otro caso. MUN_{ij} es una variable latente que toma el valor 1 si el alumno i de la

escuela j es atendido por una escuela MUN y 0 en otro caso. Las variables explicativas en ambas ecuaciones de selección son las variables de la ecuación de logro (q_{ijk}) y los instrumentos z_{ijk} ¹².

Heckman supone que los términos de error en las ecuaciones de logro ($\tau_{ij}^{PS}, \tau_{ij}^M$) y de selección ($\mu_{ij}^{PS}, \mu_{ij}^M$) se distribuyen conjuntamente como una normal bivariada.

$$\begin{bmatrix} \tau_{ij}^T \\ \mu_{ij}^T \end{bmatrix} \sim N \left(\mathbf{0}_{2 \times 1}, \begin{bmatrix} \sigma_\tau^T & \rho_T \sigma_\tau^T \sigma_\mu^T \\ \rho_T \sigma_\tau^T \sigma_\mu^T & \sigma_\mu^T \end{bmatrix} \right) \text{ con } T = PS, MUN$$

Donde ρ_T es el coeficiente de correlación simple entre ambos términos de error y en consecuencia, $\rho_T \sigma_\tau^T \sigma_\mu^T$ es la covarianza entre ambos términos de error.

$$\rho_T = \frac{\text{Cov}(\tau_{ij}^T, \mu_{ij}^T)}{\sigma_\tau^T \sigma_\mu^T}$$

Una de las tantas ventajas de la distribución normal es que no se ve afectada por transformaciones lineales, por lo que la distribución condicional también es una normal.

$$f\left(\tau^T / \mu^T\right) \sim N\left(\rho_T \frac{\sigma_\tau^T \sigma_\mu^T}{\sigma_{\mu\mu}^T} \mu^T, \sigma_{\tau\tau}^T (1 - \rho_T^2)\right)$$

o bien estandarizando queda,

$$f\left(\tau^T / \mu^T\right) \sim \phi\left(\frac{\tau^T - \rho_T \frac{\sigma_\tau^T \sigma_\mu^T}{\sigma_{\mu\mu}^T} \mu^T}{\sigma_\tau^T \sqrt{1 - \rho_T^2}}\right)$$

A partir de la expresión anterior es posible demostrar que la esperanza condicional de τ^T tiene la siguiente forma funcional:

¹² Para evitar problemas de multicolinealidad en la estimación de la ecuación de selección, las variables z_{ijk} se escogieron de tal forma que no estén correlacionadas con las variables de la ecuación de logro. Para ello, se eligió una matriz de información cuyo número de condición sea menor a 20 (umbral de multicolinealidad severa).

$$E\left(\frac{\tau^T}{\mu}\right) = \rho_T \sigma_\tau^T \lambda\left(\frac{\mathbf{z}'\xi}{\sigma_\mu^T}\right) \equiv \rho_T \sigma_\tau^T \lambda_T(\mathbf{z}'\xi)$$

$\lambda_T(\mathbf{z}'\xi)$ es el inverso del ratio de Mills (IRM) que mide la verosimilitud de que un estudiante pertenezca a una escuela PS o MUN según sea el caso (e.g. $\lambda_{PS}(\mathbf{z}'\xi)$ es la probabilidad que un estudiante asista a una escuela PS condicional a que el estudiante pertenece a la muestra que se está analizando). Heckman demostró que la inclusión del IRM como una de las variables explicativa en la ecuación de logro, resuelve el problema de sesgo de selección en las estimaciones¹³. Formalmente se tiene:

$$E\left(\frac{SIMCE_{ij}^{matPS}}{SP_{ij}=1}\right) = \beta_0^{PS} + \sum_{k=1}^K q_{ijk} \beta_k^{PS} + \rho_{PS} \sigma_\tau^{PS} \lambda_{PS}(\mathbf{z}'\xi) \quad [9]$$

$$E\left(\frac{SIMCE_{ij}^{matM}}{MUN_{ij}=1}\right) = \beta_0^M + \sum_{k=1}^K q_{ijk} \beta_k^M + \rho_M \sigma_\tau^M \lambda_M(\mathbf{z}'\xi) \quad [10]$$

Donde $\lambda_{PS}(\mathbf{z}'\xi)$ y $\lambda_M(\mathbf{z}'\xi)$ son el IRM de las ecuaciones de logro PS y MUN respectivamente,

$$\lambda_{PS}(\mathbf{z}'\xi) = \frac{\phi(\mathbf{z}'\xi)}{\Phi(\mathbf{z}'\xi)}$$

$$\lambda_{MUN}(\mathbf{z}'\xi) = \frac{-\phi(\mathbf{z}'\xi)}{1 - \Phi(\mathbf{z}'\xi)}$$

$\phi(\mathbf{z}'\xi)$ es la función de densidad de una distribución normal estándar y $\Phi(\mathbf{z}'\xi)$ es la función de distribución acumulada para una distribución normal estándar; los parámetros $\rho_{PS} \sigma_\tau^{PS}$ y $\rho_M \sigma_\tau^M$ son los coeficientes del IRM de las ecuaciones de logro PS y MUN respectivamente. La significancia estadística del coeficiente que acompaña al IRM determinará la presencia del sesgo de selección¹⁴. Intuitivamente, en presencia de sesgo de selección, el IRM es una variable relevante de la ecuación de logro. Por ello, ignorar el IRM en la especificación de la ecuación de logro escolar, genera sesgo en las estimaciones (similar al sesgo por omisión de variables relevantes).

Una vez que la ecuación de logro es corregida por la presencia de sesgo de selección (inclusión de la variable explicativa IRM en la ecuación de logro), se analiza las diferencias en el rendimiento escolar

¹³ Otro método para resolver el sesgo de selección es la estimación por máxima verosimilitud propuesto por Amemiya (1985)

¹⁴ Wooldrige (2000)

de colegios PS y MUN mediante la estimación del efecto promedio de tratamiento (average treatment effect, ATE) y la estimación del tratamiento sobre los tratados (treatment on the treated, TT). Ambos efectos corresponden a la ganancia que obtendría un estudiante por asistir a un establecimiento PS versus un establecimiento MUN. El ATE es la ganancia esperada en el puntaje de la prueba SIMCE de matemáticas que un alumno habría obtenido con la asistencia a un establecimiento particular subvencionado (versus uno municipal). El TT es la ganancia esperada en puntaje de la prueba SIMCE de matemáticas para aquellos estudiantes que efectivamente asisten a un establecimiento particular subvencionado (versus su puntaje esperado si asistieran a un establecimiento municipal)¹⁵. Formalmente, dada las siguientes ecuaciones de logro estimadas:

$$SIMCE_{ij}^{matPS} = \hat{\beta}_0^{PS} + \sum_{k=1}^K q_{ijk} \hat{\beta}_k^{PS} + \hat{\rho}_{PS} \hat{\sigma}_\tau^{PS} \hat{\lambda}_{PS}(\mathbf{z}'\hat{\xi}) \quad (\text{Particulares Subvencionadas, PS})$$

$$SIMCE_{ij}^{matMUN} = \hat{\beta}_0^{MUN} + \sum_{k=1}^K q_{ijk} \hat{\beta}_k^{MUN} + \hat{\rho}_M \hat{\sigma}_\tau^{MUN} \hat{\lambda}_M(\mathbf{z}'\hat{\xi}) \quad (\text{Municipales, MUN})$$

Se define una variable dummy PS cuyo valor es 1 si el estudiante escoge un establecimiento PS y 0 si elige un establecimiento MUN.

$$PS = \begin{cases} 1 & \text{si } \mathbf{z}'\hat{\xi} > 0 \\ 0 & \text{en otro caso} \end{cases}$$

A partir de las ecuaciones de logro estimadas (PS y MUN), el ATE se calcula de la siguiente forma:

- ATE condicional (ATE_C)

$$ATE_C(q_{ij1}, \dots, q_{ijK}, \hat{\lambda}) = (\hat{\beta}_0^{PS} - \hat{\beta}_0^{MUN}) + \sum_{k=1}^K q_{ijk} (\hat{\beta}_k^{PS} - \hat{\beta}_k^{MUN}) + (\hat{\rho}_{PS} \hat{\sigma}_\tau^{PS} \hat{\lambda}_{PS}(\mathbf{z}'\hat{\xi}) - \hat{\rho}_{PS} \hat{\sigma}_\tau^{MUN} \hat{\lambda}_{MUN}(\mathbf{z}'\hat{\xi}))$$

- ATE no condicional (ATE_{NC})

$$ATE_{NC}(q_{ij1}, \dots, q_{ijK}, \hat{\lambda}) = \int_1^N ATE_C(q_{ij1}, \dots, q_{ijK}, \lambda) dF(q_{ij1}, \dots, q_{ijK}, \lambda) \cong \frac{1}{N} \sum_{i=1}^N ATE_C(q_{ij1}, \dots, q_{ijK}, \lambda)$$

Donde N es el tamaño de muestra.

¹⁵ Las definiciones conceptuales del ATE y TT fueron adaptadas de Mizala A, Romaguera P (2000). Para mayores detalles ver Heckman et al (2001), Angrist J (2004).

El TT se calcula de la siguiente forma:

- TT condicional (TT_C)

$$TT_C(q_{ij1}, \dots, q_{ijK}, \hat{\lambda}) = ATE_C + (\hat{\sigma}_{qu}^{PS} \hat{\lambda}_{PS}(-\mathbf{z}'\hat{\xi}) - \hat{\sigma}_{qu}^{MUN} \hat{\lambda}_{MUN}(-\mathbf{z}'\hat{\xi}))$$

Donde $\hat{\sigma}_{qu}$ es la estimación de la covarianza entre el residuo de la ecuación de logro escolar y el residuo de la ecuación de selección para cada establecimiento (PS y MUN).

- TT no condicional (TT_{NC})

$$TT_{NC}(q_{ij1}, \dots, q_{ijK}, \hat{\lambda}, \mathbf{z}, PS = 1) = \int_1^N TT_C(q_{ij1}, \dots, q_{ijK}, \hat{\lambda}, \mathbf{z}, PS = 1) dF(q_{ij1}, \dots, q_{ijK}, \hat{\lambda}, \mathbf{z} / PS = 1) \\ \cong \frac{1}{N_{PS}} \sum_{i=1}^N (PS) TT_C(q_{ij1}, \dots, q_{ijK}, \hat{\lambda}, \mathbf{z}, PS = 1)$$

Donde N_{PS} es el número de observaciones (alumnos) que asisten a establecimientos PS. Finalmente, si ATE y TT son significativamente mayores que cero, entonces los colegios PS poseen mejor calidad educativa que los colegios MUN. La significancia del valor ATE y TT no condicionales pueden ser testeados a partir de un estadístico t.

d. Mínimos Cuadrados en dos etapas (TSLS) con errores robustos

El método de estimación lineal TSLS con errores robustos (o Método Generalizado de Momentos en el caso de estimaciones no lineales), es ampliamente utilizado en estudios sobre macroeconomía y finanzas. Este método, por ser un caso general de todos los métodos de estimación anteriormente analizados, será aplicado al presente estudio. En otras palabras, los métodos HLM, IV y TSLS a la Heckman, son métodos particulares de momentos. TSLS con errores robustos es muy útil cuando se trabaja con numerosas variables instrumentales, i.e., cuando el número de variables instrumentales supera al número de variables de la ecuación de logro. En este caso, se dice que la ecuación de logro está sobre identificada. En otras palabras, existen más restricciones (variables instrumentales) que parámetros a estimar (Domínguez y Lobato, 2004; Han y Phillips, 2006).

Suponga que la teoría especifica la siguiente relación:

$$SIMCE_i = h(\mathbf{x}_i, \boldsymbol{\beta}) + \varepsilon_i, \quad [11]$$

donde $h(\bullet)$ es una función lineal; $\boldsymbol{\beta}$ es un vector $K \times 1$ de parámetros que se quiere estimar y \mathbf{x}_i es un vector de características del alumno y su escuela. Podrá no tratarse de un relación de regresión estándar, dado que es posible que

$$Cov[\varepsilon_i, \mathbf{x}_j] \neq 0 \quad \text{para todo } i \text{ y } j.$$

Se asume que:

$$\begin{aligned} E(\boldsymbol{\varepsilon}) &= \mathbf{0} \\ E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') &= \boldsymbol{\Sigma} \end{aligned}$$

Donde $\boldsymbol{\Sigma}$ no está restringido. Los residuos pueden ser heteroscedásticos entre las escuelas y correlacionados al interior de las escuelas. Excepto por la posibilidad de correlación entre regresores y residuos, éste modelo podría tratarse de un modelo de regresión generalizado. Seguidamente, se supone que para cada observación i se observan un vector de L variables instrumentales \mathbf{z}_i , tal que estas variables instrumentales no están correlacionadas con los residuos $\boldsymbol{\varepsilon}_i$. Los supuestos implican establecer un conjunto de condiciones de ortogonalidad

$$E(\mathbf{z}_i \boldsymbol{\varepsilon}_i) = \mathbf{0},$$

que pueden ser suficientes para sobre identificar (si $L > K$) los parámetros del modelo.

Por conveniencia, se define

$$\mathbf{e}(\mathbf{X}, \hat{\boldsymbol{\beta}}) = y_i - h(\mathbf{x}_i, \hat{\boldsymbol{\beta}}), \quad i=1, \dots, n,$$

y

$$\mathbf{Z} = n \times \text{matriz } L \text{ cuya } i\text{-ésima fila es } \mathbf{z}_i'$$

El estimador consistente de TLSL, se define estableciendo las siguientes condiciones de momentos muestrales para $\boldsymbol{\beta}$,

$$\bar{\mathbf{m}}(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i e(\mathbf{x}_i, \boldsymbol{\beta}) = \frac{1}{n} \mathbf{Z}' \mathbf{e}(\mathbf{X}, \boldsymbol{\beta}) \quad [12]$$

El estimador TSLS con errores robustos o de mínima distancia, será aquel que minimice la siguiente función convexa

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \arg \min(\mathbf{m}(\boldsymbol{\beta})' \mathbf{W}^{-1} \mathbf{m}(\boldsymbol{\beta})) \\ q &= \left(\frac{1}{n^2}\right) [\mathbf{e}(\mathbf{X}, \hat{\boldsymbol{\beta}})' \mathbf{Z}] \mathbf{W}^{-1} [\mathbf{Z}' \mathbf{e}(\mathbf{X}, \hat{\boldsymbol{\beta}})]\end{aligned}\quad [13]$$

Donde \mathbf{W} es la matriz de varianza asintótica de $\mathbf{m}(\boldsymbol{\beta})$. Hansen (1982) demostró que la elección óptima de \mathbf{W} para el estimador de $\boldsymbol{\beta}$ que alivie el problema de matriz varianza-covarianza no escalar es:

$$W_{GMM} = \text{Varianza asintótica} \left[\frac{1}{n} \mathbf{Z}' \mathbf{e}(\mathbf{X}, \boldsymbol{\beta}) \right] = \frac{1}{n^2} \mathbf{Z}' \boldsymbol{\Sigma} \mathbf{Z}$$

Introduciendo esta expresión en la ecuación 13, se tiene

$$q = \left[\left(\frac{1}{n}\right) \mathbf{e}(\mathbf{X}, \hat{\boldsymbol{\beta}})' \mathbf{Z} \right] \left(\frac{1}{n^2} \mathbf{Z}' \boldsymbol{\Sigma} \mathbf{Z} \right)^{-1} \left[\frac{1}{n} \mathbf{Z}' \mathbf{e}(\mathbf{X}, \hat{\boldsymbol{\beta}}) \right] \quad [14]$$

La intuición detrás de la función cuadrática 14, es estimar elasticidades finitas y consistentes de los factores explicativos del desempeño escolar. Estas elasticidades son aquellas que minimizan el producto ponderado de las condiciones de momentos (q). En el centro de la expresión 16, se encuentra la matriz de ponderaciones (matriz de varianza covarianza generalizada de los residuos).

Test de validación de restricciones: Si los parámetros están sobre identificados, las ecuaciones (restricciones) de momentos que se planteen influirán significativamente en las estimaciones. Si la hipótesis del modelo (estructura de la ecuación de logro) que lleva a las ecuaciones de momentos, es inicialmente incorrecta, al menos algunas de las restricciones muestrales de momentos se verán sistemáticamente violadas. Por lo que, surge la necesidad de contrastar la validez de las restricciones. La expresión 14, se asocia al estadístico de Wald y sigue una distribución asintótica chi-cuadrado con (J-K) grados de libertad. J es el número de condiciones de momentos y K los parámetros a estimar,

$$q \sim \chi^2 [J - K]$$

Para el caso de exacta identificación, hay cero grados de libertad y $q = 0$.

III. Base de datos y definición de variables

El presente estudio utiliza una base de datos proveniente de tres fuentes de información estadística. La primera, consiste en una prueba estandarizada llamada Sistema de Medición de la Calidad de la Educación (SIMCE). El SIMCE evalúa periódicamente el logro de los estudiantes del 4to básico, 8vo básico y 2do medio; a partir del año 2005, las mediciones de 4to básico se ejecutarán anualmente, alternando el curso a medir entre 8vo básico y 2do medio. De esta primera fuente se utilizó, como variable del rendimiento escolar, los resultados de la prueba SIMCE en matemáticas (SIMCEmat) rendida por los alumnos del 4to básico 2002. La segunda fuente de información estadística, corresponde al Cuestionario de Padres y Apoderados (CPA) de los alumnos del nivel 4to básico 2002. El CPA recoge información de las familias y características del alumno. Por ejemplo, género, edad del alumno, edad de los padres, composición del hogar, gastos en educación, nivel de educación alcanzado por los padres, ingreso familiar, etc. La tercera fuente de información, corresponde a las características de los establecimientos educativos (privado subvencionado, municipal, mixto, sólo mujeres, sólo hombres, religioso, etc) publicada por el Ministerio de Educación de Chile (MINEDUC).

La ventaja de utilizar el CPA, es que provee una serie de información cualitativa que señala las razones por la que los padres decidieron matricular a sus hijos en un determinado establecimiento educativo. Por ejemplo, en la pregunta 17 del cuestionario se les pide a los padres que elijan 3 de las 9 razones (ordenadas de menor a mayor importancia), por la cual decidieron matricular a sus hijos en la escuela a la que actualmente asisten. Esta pregunta será potencialmente útil a la hora de definir los determinantes que aproximan las razones por la que un estudiante pertenece a una escuela PS o MUN. Asimismo, el cuestionario también permite identificar el número de colegios ubicados en la comuna de residencia del estudiante, lo que permite construir una variable Proxy de la competencia de colegios en dicha comuna¹⁶.

Por otro lado, se eliminaron de la muestra a los estudiantes cuya edad fue menor a 6 años y mayor a 14 años; esto se resolvió considerando que la edad promedio de un estudiante de 4to básico es 10 años. Se definió la variable educación de los padres como el promedio simple de los años de educación obtenidos por el padre y de la madre del alumno. Esta variable permite medir el efecto conjunto de la educación del padre y madre sobre el rendimiento escolar del estudiante. Los cuadros

¹⁶ Gallego 2005.

A1-1 y A1-2 del Anexo1, presentan la lista de variables utilizadas a nivel de estudiantes y a nivel de colegios, respectivamente.

Los cuadros 2 y 3, muestra las estadísticas descriptivas para algunas variables de interés (resultados de la prueba SIMCEmat, educación de los padres, ingreso familiar, entre otras).

| Cuadro 2 | | | | | |
|---|--------|-----------|------------|-------|---------|
| Estadísticas descriptivas a nivel de estudiantes | | | | | |
| Variable | Obs | Media | Desv. Std. | Min | Max |
| SIMCEmat | 232935 | 243.4508 | 52.75609 | 94 | 379 |
| Años de educación padres | 199519 | 10.41 | 3.094889 | 0 | 21 |
| Ingreso familiar | 227917 | 194627.9 | 211896.5 | 50000 | 2000000 |
| Num. Personas del hogar | 231656 | 5.139224 | 1.840775 | 2 | 17 |
| Dummy sexo del alumno | 254495 | 0.5124305 | 0.4998464 | 0 | 1 |
| Recursos econ. | 235410 | 403.1568 | 128.8362 | 0.6 | 3591.1 |
| Dummy zona rural | 255085 | 0.1279064 | 0.3339862 | 0 | 1 |
| Num. Libros en el hogar (*) | 230838 | 37.30931 | 53.93053 | 2 | 250 |
| Elección por proximidad (*) | 235819 | 0.4029998 | 0.4905018 | 0 | 1 |
| Elección porque enseña valores (*) | 235819 | 0.2698129 | 0.4438634 | 0 | 1 |
| Elección por su bajo costo económico (*) | 235819 | 0.3604629 | 0.4801358 | 0 | 1 |
| Elección porque otros familiares estudian en la escuela (*) | 235818 | .21013660 | .40740630 | 0 | 1 |
| Elección porque no tuvo otra opción (*) | 235819 | .10307910 | .3040628 | 0 | 1 |
| Número de colegios particulares subvencionados (PS) en la comuna de residencia del estudiante (*) | 222877 | 17.5189 | 20.47859 | 0 | 86 |
| Número total de colegios en la comuna de residencia del estudiante (*) | 211767 | 47.07086 | 28.53857 | 1 | 134 |

(*) Variables utilizadas como instrumentos.

Fuente: Estimación propia

| Cuadro 3 | | | | | |
|---|------|------|-----------|------|--------|
| Estadísticas descriptivas a nivel de establecimientos | | | | | |
| Variable | Obs | Mean | Std. Dev. | Min | Max |
| Dummy particular subvencionado (PS) | 5549 | 0.38 | 0.49 | 0 | 1 |
| Establecimiento de sólo hombres | 5549 | 0.02 | 0.13 | 0 | 1 |
| Establecimiento de sólo mujeres | 5549 | 0.03 | 0.16 | 0 | 1 |
| Experiencia docentes | 5382 | 6 | 1 | 0 | 10 |
| Cuociente alumno-profesor | 5376 | 2.70 | 1.92 | 2 | 62 |
| Ln matriculados | 5390 | 5.63 | 1.13 | 2.40 | 8.83 |
| Establecimiento religioso (*) | 5549 | 0.06 | 0.23 | 0 | 1 |
| Pago por escolaridad (*) | 5492 | 4929 | 8015 | 0 | 118885 |

(*) Variables utilizadas como instrumentos.

Fuente: Estimación propia.

IV. Resultados

a. Mínimos Cuadrados Ordinarios

En el presente estudio la estimación por Mínimos Cuadrados Ordinarios (Ordinary Least Square, OLS) corresponde a la técnica econométrica que menos le exige a la estructura de los datos, i.e., no controla por la presencia de multiniveles y sesgo de selección en las observaciones. No obstante, pese a que OLS en su estructura más simple no corrige por las mencionadas características de los datos, una de sus variantes como lo es la estimación por clusters permite controlar por la interdependencia del rendimiento escolar al interior de las escuelas. La estimación por cluster (es decir, por establecimientos) especifica que las observaciones son independientes a través de los establecimientos, pero no necesariamente dentro de ellos. En efecto, se realizaron estimaciones de la ecuación de logro escolar por OLS con y sin cluster. A continuación se presentan ambos resultados.

| Cuadro 4 | | |
|---|------------------------|------------------------|
| Estimación de la ecuación de logro escolar con OLS. SIMCE 4to básico año 2002, matemáticas. | | |
| Variables | Técnicas de estimación | |
| SIMCEmat (dependiente) | OLS sin cluster | OLS con cluster |
| Constante | 157.66 (1.56)** | 157.66 (3.66)** |
| Dummy Particular Subvencionado | 11.63 (0.32)** | 11.63 (0.87)** |
| Años de educación padres | 4.01 (0.05)** | 4.01 (0.07)** |
| Ingreso familiar | 2.0E-05 (6.7E-07)** | 2.0E-05 (9.6E-07)** |
| Num. Personas del hogar | -1.84 (0.07)** | -1.84 (0.08)** |
| Dummy sexo del alumno (masculino) | 4.33 (0.25)** | 4.33 (0.29)** |
| Dummy zona rural | 5.51 (0.48)** | 5.51 (1.04)** |
| Establecimiento de sólo mujeres | 9.11 (0.59)** | 9.11 (1.37)** |
| Establecimiento de sólo hombres | 12.94 (0.87)** | 12.94 (2.23)** |
| Dummy JEC | -4.00 (0.34)** | -4.00 (0.50)** |
| Recursos econ. | 0.02 (9.9E-04)** | 0.02 (2.4E-03)** |
| Ln matrícula | 0.75 (0.31)** | 0.75 (0.74)* |
| Cuociente alumno-profesor | -0.02 (0.01)** | -0.02 (0.02)* |
| Experiencia docentes | 5.43 (0.30)** | 5.43 (0.70)** |
| Observaciones | 158475 | 158475 |
| Estadístico F | 1878.89 | 619.49 |
| R2-ajustado | 0.1335 | 0.1335 |
| Raíz Error Cuadrático Medio | 48.814 | 48.814 |

Nota: Error estándar entre paréntesis. Por otra parte, los grados de libertad para el estadístico F de la estimación OLS (sin cluster) es (13,158461), mientras que para la estimación OLS (con cluster) es (13,4887).

** Significativo al 95% de confianza

* Significativo al 90% de confianza

Los clusters son los establecimientos. Cabe señalar que las estimaciones empleando simultáneamente cluster y la matriz consistente de covarianza de White, no exhibió diferencia alguna respecto de la estimación por clusters.

Fuente: estimación propia.

De acuerdo a la estimación por OLS, un estudiante al ser atendido por una escuela particular subvencionada obtendría una ganancia promedio de 11,63 puntos en la prueba SIMCE de matemáticas (versus si asistiese a una escuela municipal). No obstante, la variabilidad de dicho resultados es mayor cuando se corrige por la interdependencia del rendimiento al interior de las escuelas –el error estándar de la estimación pasa de +/-0,32 a +/- 0,87.

Esta técnica fue considerada como la base de comparación para las restantes metodologías que además controlan por efectos multiniveles y sesgo de selección en los datos.

b. Modelo Lineal Jerárquico

El HLM permite responder acerca de cómo los establecimientos educativos afectan a los estudiantes dentro de ellas, i.e., el HLM permite medir el efecto que tienen los establecimientos educativos sobre el rendimiento escolar alcanzado por los estudiantes. Para el presente estudio, dicho efecto puede ser interpretado como la ganancia en la prueba SIMCE de matemáticas de un alumno que asiste a un establecimiento PS (versus si fuese atendido por un establecimiento MUN).

El Cuadro 6, muestra la estimación de la ecuación de logro escolar a través del modelo HLM de segundo nivel. El coeficiente estimado de la variable PS indica que la ganancia esperada en la prueba SIMCE de matemáticas de un estudiante atendido por un establecimiento particular subvencionado es 12.7 puntos (versus si fuese atendido por un establecimiento municipal, MUN).

| Cuadro 5 | |
|--|------------------------|
| Modelo HLM, SIMCE 4to básico año 2002, matemáticas. Efecto fijo. | |
| Variable | Estimadores |
| Para la cte β_{0j} (logro promedio del establecimiento) | |
| Constante | 188.85 (26.95)** |
| Dummy particular subvencionado (PS) | 12.77 (1.57)** |
| Establecimiento de sólo mujeres | 13.72 (1.65)** |
| Establecimiento de sólo Hombres | 17.50 (2.98)** |
| Dummy JEC | -3.78 (1.83)** |
| Recursos econ. | 0.01 (2.2E-03)** |
| Ln matrícula | 3.93 (0.95) |
| Cuociente alumno-profesor | -0.07 (0.04)** |
| Experiencia docentes | 4.13 (0.94)** |
| Para la pendiente (β_{1j}) (Años de educación padres) | |
| Constante | 5.24 (0.05)** |
| Para la pendiente (β_{2j}) (Ingreso familiar) | |
| Constante | 2.4E-05 (6.6E-07)** |
| Para la pendiente (β_{3j}) (Num. Personas del hogar) | |
| Constante | -1.92 (0.07)** |
| Para la pendiente (β_{4j}) (Dummy sexo del alumno) | |
| Constante | 3.59 (0.26)** |
| Para la pendiente (β_{5j}) (Dummy zona rural) | |
| Constante | -5.72 (0.42)** |
| Observaciones Nivel 1 (estudiantes) | 255086 |
| Observaciones Nivel 2 (establecimientos) | 4549 |
| Log-Likelihood | -9.719E+005 |

Nota: Error estándar entre paréntesis

** Significativo al 95% de confianza

Las estimaciones consideran la matriz de varianza-covarianza consistente de White (1980).

No se rechazó el supuesto de normalidad de los residuos, esto según el test de Kolmogorov - Smirnov (K S=0.013, P-value=0.17).

Fuente: Estimación propia.

El análisis de datos de sección cruzada con observaciones multiniveles, involucra potenciales problemas de matriz de varianza-covarianza no escalar de los residuos (correlación de los puntajes al interior de las escuelas y heterogeneidad del desempeño entre las escuelas). Por ello, los resultados presentados en el Cuadro 6, fueron estimados empleando métodos de estimación robusta¹⁷.

Finalmente, las pruebas de Chi-cuadrado de los parámetros de HLM indican una variabilidad significativa en la distribución del logro escolar, incluso luego de controlar por las características de los estudiantes (nivel 1) y las características de los colegios (nivel 2). En efecto, es posible concluir que el desempeño escolar de los estudiantes varía perceptiblemente entre las escuelas.

| Cuadro 6 | | | | |
|---|---------------------|---------------------------|--------------------|------------------------|
| Modelo HLM, SIMCE 4to básico año 2002, matemáticas. Efecto aleatorio. | | | | |
| Efectos aleatorios | Desviación estándar | Componente de la varianza | Grados de libertad | Chi-cuadrado (p-value) |
| Intercepto (β_{0j}) | 0.72 | 0.52 | 4917 | 377.80** (0.033) |
| Efectos del nivel 1 (ϵ_{ij}) | 49.38 | 2438.77 | | |

** Significativo al 95% de confianza.

Fuente: Estimación propia.

Por otra parte, cabe destacar que la especificación del modelo HLM no corrige por sesgo de selección. En otras palabras, el modelo HLM compara los resultados entre colegios PS y M asumiendo que los estudiantes están distribuidos aleatoriamente entre los distintos establecimientos educativos. Esto implica suponer que la variable PS es exógena cuando realmente es endógena. Si se obvia la endogeneidad de la variable PS, entonces los coeficientes estimados son sesgados e inconsistentes (problema de sesgo de selección). En las siguientes secciones, se analizarán algunas de las soluciones al problema de sesgo de selección.

c. Estimación con Variables Instrumentales

A menudo los criterios de selección de instrumentos no están exentos de críticas respecto de su rigurosidad, los cuales podrían sesgar las estimaciones con IV (Figlio D. et. al., 1999). Por ello, es importante disponer de un test estadístico que avale la selección de instrumentos. En este sentido, el test de Sargan analizado en el capítulo metodológico de estimación con IV, permite testear la validez estadística de las variables instrumentales¹⁸. El siguiente cuadro muestra el resultado del test de validación de instrumentos de Sargan:

¹⁷ Las técnicas de estimación robusta fueron: MCGF y Análisis de Cluster, las cuales no presentaron diferencias entre los coeficientes estimados.

¹⁸ Para mayor detalle sobre los instrumentos utilizados, ver cuadros del Anexo 1.

| Cuadro 7 | | | | |
|------------------------------------|----------|-----------------|---------------------------------------|---|
| Test de validación de IV de Sargan | | | | |
| Obs. | SE | σ_τ^2 | $\chi^2_{(4 \text{ gl})}$ (calculado) | $\chi^2_{(4 \text{ gl})}$ (crítico) 95% de confianza |
| 190371 | 9.10E-10 | 48.15 | 1.89E-11** | 9.49 |

** No se rechaza la hipótesis nula de validez de los instrumentos al 95% de confianza

Fuente: Estimación propia.

El valor calculado de la chi-cuadrado con cuatro grados de libertad (1.89E-11) es inferior a su valor crítico (9.49) al 95% de confianza. Por lo anterior, se concluye que no hay suficiente evidencia estadística para rechazar la hipótesis de validez de los instrumentos. En otras palabras, la selección de instrumentos contempla un argumento estadístico que la valida.

Una vez testada la validez de los instrumentos, se procedió a estimar la ecuación de logro escolar (ecuación 5). El cuadro 8, muestra la estimación robusta de la ecuación de logro escolar con Variables Instrumentales en dos etapas (IV-TSLS) y la estimación del Efecto del Tratamiento (Treatment Effect, TE). La primera etapa de la regresión con IV (regresión de la variable dicotómica PS sobre los instrumentos) se estimó mediante un Probit¹⁹. Los instrumentos seleccionados, además de ser ortogonales a los residuos de la ecuación de logro (test de Sargan), cumplieron la condición de ortogonalidad con los residuos generalizados del Probit (test de exogeneidad de los instrumentos). Los resultados del test de exogeneidad de los instrumentos se muestran en el siguiente cuadro:

| Cuadro 8 | | |
|---|--|---------------|
| Test de exogeneidad de los instrumentos | | |
| Hipótesis nula | Restricciones de momentos | Estadístico-t |
| Los instrumentos son ortogonales a los residuos generalizados del Probit | $E(\boldsymbol{\eta} \times \text{PSCORE})=0$ | 7.4E-08** |
| | $E(\boldsymbol{\eta} \times \text{MUNPSCORE})=0$ | 4.3E-09** |
| | $E(\boldsymbol{\eta} \times \text{FAMEST})=0$ | -1.7E-08** |
| | $E(\boldsymbol{\eta} \times \text{PAGOESC})=0$ | 1.4E-09** |
| $\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^N \eta_i z_i = 0$ | $E(\boldsymbol{\eta} \times \text{PROXI})=0$ | -5.9E-09** |
| | $E(\boldsymbol{\eta} \times \text{VALORES})=0$ | 1.1E-08** |
| | $E(\boldsymbol{\eta} \times \text{ECON})=0$ | 1.4E-09** |
| | $E(\boldsymbol{\eta} \times \text{NTOP})=0$ | 1.2E-08** |
| $\boldsymbol{\eta}$: Residuo generalizado del Probit. | $E(\boldsymbol{\eta} \times \text{LIBROS})=0$ | 7.6E-08** |
| | $E(\boldsymbol{\eta} \times \text{ESCREL})=0$ | 6.0E-08** |

** No se rechaza la hipótesis nula al 5% de significancia.

Fuente: estimación propia.

¹⁹ El Probit fue estimado consistentemente con la matriz de varianza-covarianza de White.

El término “Efecto del Tratamiento” se refiere al efecto causal de una variable dicotómica sobre una variable de interés científico o de política pública. Por ejemplo, el efecto que tiene sobre la prueba SIMCE de matemáticas realizada por un alumno que asiste a una escuela PS (versus si asistiera a una escuela MUN). En presencia de sesgo de selección, el TE puede ser estimado utilizando variables instrumentales²⁰. La estimación del TE de la ecuación de logro escolar se realizó con Máxima Verosimilitud (Log Likelihood, LL) y Mínimos Cuadrados en dos Etapas (Two Step Linear Square, TSLS). Los resultados de las estimaciones (IV-TSLS, TE con LL y TE con TSLS) muestran que en promedio la calidad de la educación de los colegios PS es significativamente superior a la calidad ofrecida por los colegios MUN. En otras palabras, la ganancia esperada en la prueba SIMCE de matemáticas de un alumno que asiste a una escuela PS, es entorno a 16 puntos (versus un alumno atendido por una escuela MUN).

Cuadro 9
Estimación de la ecuación de logro escolar con IV-TSLS y Treated Effect. SIMCE 4to básico año 2002, Matemáticas

| Variables | Técnicas de estimación | | | |
|-----------------------------------|-------------------------------------|------------------------|------------------------|------------------------|
| | Treated Effect (TE) | | | |
| | Estimación por máxima verosimilitud | | | |
| | IV-TSLS (sin cluster) | TE (con clusters) | TE (sin clusters) | TSLS (sin cluster) |
| SIMCEmat (dependiente) | | | | |
| Constante | 166.73 (1.64)** | 160.59 (3.74)** | 160.59 (1.69)** | 160.92 (1.67)** |
| Dummy Particular Subvencionado | 15.55 (0.49)** | 15.85 (1.08)** | 15.85 (0.49)** | 16.28 (0.49)** |
| Años de educación padres | 3.95 (0.05)** | 3.90 (0.07)** | 3.90 (0.05)** | 3.88 (0.05)** |
| Ingreso familiar | 2.0E-05 (7.6E-07)** | 2.0E-05 (1.0E-06)** | 2.0E-05 (7.3E-07)** | 2.0E-05 (7.3E-07)** |
| Num. Personas del hogar | -1.82 (0.07)** | -1.82 (0.08)** | -1.82 (0.07)** | -1.81 (0.07)** |
| Dummy sexo del alumno (masculino) | 4.24 (0.27)** | 4.24 (0.30)** | 4.24 (0.27)** | 4.24 (0.27)** |
| Dummy zona rural | 3.63 (0.52) | 5.30 (1.05)** | 5.30 (0.52)** | 5.27 (0.52) |
| Establecimiento de sólo mujeres | 7.81 (0.58)** | 8.09 (1.40)** | 8.09 (0.62)** | 8.07 (0.62)** |
| Establecimiento de sólo hombres | 12.58 (0.86)** | 12.35 (2.29)** | 12.35 (0.90)** | 12.36 (0.90)** |
| Dummy JEC | -4.36 (0.36)** | -4.22 (0.52)** | -4.22 (0.36)** | -4.21 (0.36)** |
| Recursos econ. | 0.02 (1.1E-03)** | 0.02 (2.6E-03) | 1.1E-03 (2.6E-03) | 0.02 (1.1E-03)** |
| Ln matrícula | 2.50 (0.31) | 0.74 (0.74) | 0.74 (0.33)** | 0.76 (0.33) |
| Cuociente alumno-profesor | -0.04 (0.01)** | -0.02 (0.02) | -0.02 (7.9E-03)** | -0.02 (0.01)** |
| Experiencia docentes | 2.14 (0.27)** | 5.02 (0.70)** | 5.02 (0.31)** | 4.92 (0.31)** |
| Observaciones | 139691 | 139691 | 139691 | 139691 |
| Test de Wald $X^2_{(13g)}$ | - | 7306.21 | 21135.71 | 22271.99 |
| p-value | - | 0.0000 | 0.0000 | 0.0000 |
| Log-pseudolikelihood | - | -805321.75 | -805321.75 | -805321.75 |
| Estadístico F(13,139677) | 1640.62 | - | - | - |
| p-value | 0.0000 | - | - | - |
| R2-ajustado | 0.1324 | - | - | - |
| Raíz Error Cuadrático Medio | 48.809 | - | - | - |

Nota: Error estándar entre paréntesis. Respecto de la estimación con clusters (establecimientos educativos), el error estándar fue ajustado para 4662 clusters.

** Significativo al 95% de confianza.

Fuente: Estimación propia.

²⁰ Las variables utilizadas en la ecuación de tratamiento para la estimación del TE, fueron las mismas que se emplearon como instrumentos en la estimación de primera etapa con IV.

d. Corrección de Heckman en dos etapas

A continuación se estiman dos ecuaciones separadas para cada tipo de establecimiento (PS y MUN), las cuales son también corregidas por el potencial sesgo de selección. Estimar por separado las ecuaciones de logro para los colegios PS y MUN respectivamente, permite que los retornos a las características de los alumno o los establecimientos difieran entre los tipos de establecimiento. Cada ecuación de logro fue estimada mediante el método de corrección de Heckman en dos etapas y máxima verosimilitud.

El cuadro 11 ilustra los resultados obtenidos en la estimación de las ecuaciones de logro para establecimientos MUN y PS. La estimación del inverso del ratio de Mills (IRM) resultó significativa al 95% para la ecuación de logro en establecimientos PS, esto es, no rechazó la presencia de sesgo de selección en las observaciones. Por otro lado, la IRM no resultó significativa para el caso de la ecuación de logro de establecimientos MUN. En este caso, es posible concluir que tanto los colegios municipalizados y apoderados no presentan una conducta selectiva.

Cuadro 10
Estimación de la ecuación de logro escolar con corrección de Heckman. SIMCE 4to básico año 2002, matemáticas

| Variables | Técnicas de estimación | | | | | |
|---|-------------------------------------|-------------------------------------|-----------------------|-------------------------------------|-------------------------------------|-----------------------|
| | Ecuac. 1: Colegios PS | | | Ecuac. 2: Colegios MUN | | |
| | Máxima Verosimilitud (con clusters) | Máxima Verosimilitud (sin clusters) | TOLS (sin cluster) | Máxima Verosimilitud (con clusters) | Máxima Verosimilitud (sin clusters) | TOLS (sin cluster) |
| Constante | 164.31 (2.27)** | 164.31 (2.26)** | 164.72 (2.26)** | 170.72 (5.19)** | 170.72 (2.74)** | 171.84 (2.64)** |
| Años de educación padres | 3.82 (0.12)** | 3.82 (0.08)** | 3.78 (0.08)** | 3.85 (0.09)** | 3.85 (0.07)** | 3.86 (0.07)** |
| Ingreso familiar | 1.9E-05 (1.3E-06)* | 1.9E-05 (9.1E-07)* | 1.9E-05 (9.1E-07)* | 2.1E-05 (1.6E-06)** | 2.1E-05 (1.2E-06)** | 2.1E-05 (1.2E-06)* |
| Num. Personas del hogar | -1.46 (0.13)** | -1.46 (0.12)** | -1.45 (0.12)** | -1.97 (0.10)** | -1.97 (0.10)** | -1.97 (0.10)** |
| Dummy sexo del alumno (masculino) | 4.56 (0.43)** | 4.56 (0.41)** | 4.57 (0.41)** | 4.05 (0.41)** | 4.05 (0.36)** | 4.05 (0.36)* |
| Dummy zona rural | -0.97 (2.07)** | -0.97 (1.08)** | -0.53 (1.08) | 5.23 (1.29)** | 5.23 (1.72)** | 4.99 (0.70)* |
| Establecimiento de sólo mujeres | 7.32 (2.00)** | 7.32 (0.84)** | 7.20 (0.84)** | 9.49 (1.95)** | 9.49 (1.91)** | 9.47 (0.91)* |
| Establecimiento de sólo hombres | 14.83 (2.40)** | 14.83 (1.19)** | 14.78 (1.19)** | 9.29 (4.06)** | 9.29 (1.37)** | 9.37 (1.37)* |
| Dummy JEC | -4.38 (0.90)** | -4.38 (0.57)** | -4.33 (0.57)** | -4.26 (0.65)** | -4.26 (0.47)** | -4.29 (0.47)** |
| Recursos econ. | 0.01 (4.5E-03)* | 0.01 (1.9E-03)* | 0.01 (1.9E-03)** | 0.02 (3.4E-03)** | 0.02 (1.4E-03)** | 0.02 (1.4E-03)** |
| Ln matrícula | 0.55 (1.18) | 0.55 (0.46) | 0.40 (0.46) | -0.50 (1.26) | -0.50 (1.65) | -0.14 (0.60)** |
| Cuociente alumno-profesor | -0.03 (0.03) | -0.03 (0.01) | -0.03 (0.01)** | -0.01 (0.01) | -0.01 (0.01) | -0.01 (0.01)** |
| Experiencia docentes | 7.92 (1.04)** | 7.92 (0.45)** | 8.11 (0.44)** | 4.88 (1.40)** | 4.88 (0.76)** | 4.36 (0.68)** |
| Inverso del Ratio de Mills (IRM) | -6.00 (1.15)** | -6.00 (0.49)** | -6.75 (0.50)** | 0.20 (0.88) | 0.20 (0.63) | -0.40 (0.49) |
| Observaciones | 146496 | 146496 | 146496 | 143696 | 143696 | 143696 |
| $ \rho < 1$ | -0.13 (0.02)** | -0.13 (0.01)** | -0.14 | 0.004 (0.018) | 0.004 (0.012) | -8.0E-03 |
| Test de Wald $\chi^2_{(12 \text{ gl})}$ | 2633.81 | 8949.63 | 8789.03 | 3153.01 | 7142.04 | 7172.85 |
| P-value | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| Log-pseudolikelihood | -358976 | -358976 | - | -465678.7 | -465678.7 | - |

Nota: Error estándar entre paréntesis. Respecto de la estimación de la ecuación PS con clusters (establecimientos educativos PS), el error estándar se ajustó a 1719 clusters. Para el caso de la estimación MUN con cluster (establecimientos educativos MUN), el error estándar se ajustó a 2950 clusters.

**Significativo al 95% de confianza

*Significativo al 90% de confianza

Para todas las ecuaciones estimadas, no se rechazó la hipótesis de normalidad en los residuos al 95% de confianza, esto según el test de Jarque-Bera.

$|\rho|$: Coeficiente de correlación entre los residuos de la ecuación de logro y los residuos de la ecuación de selección.

Fuente: estimación propia.

El cuadro 10, muestra las estimaciones del ATE y TT por máxima verosimilitud y TSLS. El cálculo del ATE y TT ayudarán a comparar el rendimiento escolar de un alumno atendido por una escuela PS (tratados) versus si fuese atendido por una escuela MUN. La evaluación de la posible diferencia en el desempeño escolar entre colegios MUN y PS, se realiza mediante un enfoque de evaluación de efectividad de programas sociales. En este sentido, se considera que los beneficiarios del programa social son aquellos que recibieron un “tratamiento” y la efectividad del programa se mide por el efecto que éste causó en los tratados versus la situación en ausencia del programa²¹. Basado en dicho enfoque, asistir a un establecimiento PS es un “tratamiento” y los criterios de evaluación son el ATE y TT.

| | Máxima Verosimilitud | TSLS |
|-----|----------------------|-------------------|
| ATE | 13.87 (1.36)** | 13.91 (1.36)** |
| TT | 14.13 (1.33)** | 14.17 (1.34)** |

Nota: Error estándar entre paréntesis.

** Significativo al 95% de confianza.

Fuente: Estimación propia.

Los resultados del ATE muestran que la ganancia esperada en el puntaje de la prueba SIMCE de matemáticas que obtiene un estudiante escogido para ser atendido por una escuela PS es 14 puntos (versus si asistiera a una escuela MUN). Por otra parte, el TT señala que la ganancia esperada en el puntaje de la prueba SIMCE de matemáticas que obtiene un estudiante que efectivamente asiste a un establecimiento PS es 14 puntos (versus si fuese atendido por una escuela MUN). En conclusión, la evidencia empírica señala que el desempeño escolar medido por el SIMCE, es estadísticamente superior en los colegios PS que en los colegios MUN.

e. Mínimos Cuadrados en dos etapas (TSLS) con errores robustos

El cuadro 11, muestra los resultados de la estimación de la ecuación de logro escolar mediante TSLS con errores robustos. Los parámetros estimados de la ecuación de logro escolar se obtuvieron mediante la minimización de la distancia entre los momentos teóricos y sus análogos muestrales.

²¹ Este enfoque ha sido utilizado por Sapelli y Vial (2001) para analizar los resultados del SIMCE de 2° de Enseñanza Media de 1998.

El coeficiente estimado por TSLS con errores robustos de la variable latente PS (Cuadro 11), muestra que el desempeño de los colegios PS supera significativamente al de los colegios MUN, es decir, la ganancia esperado en el puntaje SIMCE de matemáticas de un estudiante atendido por un establecimiento PS es 17,3 puntos (versus si fuese atendido por una escuela MUN).

| Cuadro 12 | |
|--|--|
| Estimación de la ecuación de logro escolar con TSLS con errores robustos. SIMCE 4to básico año 2002, Matemáticas | |
| Variables | Técnica de estimación TSLS con errores robustos |
| SIMCEmat (dependiente) | |
| Constante | 151.08 (4.10)** |
| Particular Subvencionado (PS) | 17.29 (1.92)** |
| Años educación padres | 3.75 (0.08)** |
| Ingreso familiar | 1.9E-05 (1.0E-06)** |
| Num. Personas del hogar | -1.75 (0.08)** |
| Dummy sexo del alumno (masculino) | 4.20 (0.30)** |
| Dummy zona rural | 3.97 (0.57) |
| Establecimiento de sólo mujeres | 8.07 (1.42)** |
| Establecimiento de sólo hombres | 12.28 (2.27)** |
| Dummy JEC | -4.02 (0.54)* |
| Recursos econ. | 0.02 (2.6E-03)** |
| Ln matrícula | -1.64 (1.06)* |
| Cuociente alumno-profesor | 0.02 (0.02) |
| Experiencia docentes | 8.96 (1.18)** |
| Observaciones | 139691 |
| Raíz Error Cuadrático Medio | 48.9302 |
| Hansen (q) $\chi^2_{(9 \text{ gl})}$ | 1369.7154 |
| p-value | 0.0000 |

Nota: Error estándar entre paréntesis

** Significativo al 95% de confianza

* Significativo al 90% de confianza

El estadístico de Hansen J no rechaza la hipótesis de validez de las restricciones de momentos al 95% de confianza.

Fuente: Estimación propia.

f. Comparación del desempeño escolar PS (versus MUN) entre las distintas técnicas de estimación empleadas

Un objetivo común en la investigación sobre las diferencias en el desempeño escolar PS y MUN es realizar inferencias sobre la ganancia esperada en la prueba SIMCE de un estudiante atendido por un establecimiento PS (versus si fuese atendido por una escuela MUN). El cuadro 13, muestra las distintas estimaciones de la ganancia esperada en la prueba SIMCE de matemáticas con sus respectivas desviaciones estándar. La estimación de la desviación estándar está basada en la técnica de simulación bootstrap²². El bootstrap se diferencia del enfoque tradicional paramétrico en que permite estimar la desviación estándar de los coeficientes de interés empleando un gran número de cálculos repetitivos para estimar la forma de la distribución muestral del estadístico, en lugar de basar las estimaciones en fuertes supuestos distribucionales (por ejemplo: normalidad, independencia). Esto permite hacer inferencias en casos donde tales supuestos son insostenibles. En resumen, el bootstrap puede considerarse como un tipo especial de simulación denominada simulación basada en los datos (simulación a partir de una estimación de la población basada en los datos). Efron y Tibshirani (1993).

| IV | TE | | TT | | TOLS |
|----------|----------------|----------|----------------|----------|----------------------|
| | Log Likelihood | TOLS | Log Likelihood | TOLS | con errores robustos |
| 15.6 | 15.9 | 16.3 | 14.1 | 14.2 | 17.3 |
| (1.05)** | (0.50)** | (0.51)** | (0.31)** | (0.30)** | (0.70)** |

Nota: Error estándar entre paréntesis. Los errores estándar fueron estimados mediante la técnica de bootstrap con 10,000 réplicas de la muestra

* Significativo al 5% de significancia

Fuente: Estimación propia

La diferencia entre los estimadores se debe a las distintas ventajas y limitaciones que poseen cada una de las técnicas para resolver problemas asociados a la estructura de los datos. Se esperaría que el verdadero desempeño escolar de los establecimientos PS (versus MUN) se encuentre dentro del conjunto de estimadores presentados en el cuadro anterior. En cuanto al tamaño del error estándar de las estimaciones, el estimador TT derivado de la corrección de Heckman –asumiendo distintos retornos a las características de los estudiantes de colegios MUN y PS– mostró el menor error estándar.

²² Introducido por Efron 1979.

Por otra parte, al comparar los resultados en función de las variables instrumentales utilizadas, se observó que el método de Heckman exhibió menos variabilidad en las estimaciones para las distintas especificaciones anidadas de variables instrumentales. En tanto, para una determinada especificación de variables instrumentales, todas las técnicas revelaron diferencias en sus resultados, y aquellas que controlaron por clusters de establecimientos exhibieron una mayor varianza en la estimación. En efecto, dependiendo de la metodología utilizada, las estimaciones revelaron que un estudiante atendido por un colegio particular subvencionado, obtiene una ganancia entre 11.63 (estimado por OLS) y 17.29 (estimado por TSLS con errores robustos) puntos adicionales en la prueba SIMCE de matemáticas (versus si el estudiante hubiese asistido a un colegio municipal).

Cuadro 14
Sensibilidad de la brecha de rendimiento escolar entre escuelas privadas subvencionadas y municipales

| Variables e instrumentos | OLS | | HLM | Método con variables instrumentales | | | | | |
|---|-------------------|-------------------|-----------------|-------------------------------------|-------------------|----------------------|-------------------|---------------------------|---------------------|
| | Sin cluster | Con cluster | | Heckman (AIE) (1) | | Treated Effect | | TSLS con errores robustos | |
| | | | | M. Verosimilitud | | Máxima verosimilitud | | | |
| | | | | Sin cluster | Con cluster | Sin cluster | Con cluster | | TSLS |
| Sin instrumentos: Utilizando sólo las variables a nivel de estudiantes y a nivel de escuelas | 11.63 (0.32)** | 11.63 (0.87)** | 127 (0.98)** | - | - | - | - | - | - |
| Con instrumentos: | | | | | | | | | |
| Número de colegios particulares subvencionados (PS) en la comuna de residencia del estudiante | - | - | - | 11.56 (0.33)** | 11.56 (0.87)** | -2.85 (1.30)** | -2.85 (3.58) | -2.18 (1.31)* | -34.67 (15.28)** |
| + Número total de colegios en la comuna de residencia del estudiante | - | - | - | 11.04 (0.34)** | 11.04 (0.89)** | -6.05 (1.30)** | -6.05 (3.74)** | -5.39 (1.33)** | -24.01 (4.97)** |
| + Elección porque otros familiares estudian en la escuela | - | - | - | 11.04 (0.34)** | 11.04 (0.89)** | -4.51 (1.28)** | -4.51 (3.51)* | -3.80 (1.29)** | -2.02 (1.053) |
| + Pago por escolaridad | - | - | - | 11.00 (0.34)** | 11.00 (0.89)** | 11.65 (1.26)** | 11.65 (1.26)** | 12.69 (0.62)** | 19.29 (1.92)** |
| + Elección por proximidad | - | - | - | 10.92 (0.34)** | 10.92 (0.89)** | 12.27 (0.59)** | 12.27 (1.23)** | 13.16 (0.60)** | 19.16 (1.68)** |
| + Elección porque enseña valores | - | - | - | 11.00 (0.34)** | 11.00 (0.89)** | 15.96 (0.54)** | 15.96 (1.11)** | 16.73 (0.55)** | 16.43 (1.42)** |
| + Elección por su bajo costo económico | - | - | - | 11.00 (0.34)** | 11.00 (0.89)** | 15.92 (0.54)** | 15.92 (1.11)** | 16.66 (0.55)** | 15.29 (1.41)** |
| + No tuvo otra opción | - | - | - | 10.31 (0.37)** | 10.31 (0.90)** | 16.06 (0.54)** | 16.06 (1.11)** | 16.81 (0.55)** | 16.21 (1.40)** |
| + Num. Libros en el hogar | - | - | - | 11.03 (0.34)** | 11.03 (0.89)** | 18.22 (0.54)** | 18.22 (1.15)** | 18.81 (0.55)** | 16.62 (1.40)** |
| + Establecimiento religioso | - | - | - | 13.87 (0.35)** | 13.87 (1.36)** | 15.85 (0.49)** | 15.85 (1.08)** | 16.28 (0.49)** | 17.29 (1.92)** |

+ Sele como "la variable anterior se añade."

Error estándar entre paréntesis curvo

** significativo al 95% de confianza

* significativo al 90% de confianza

(1) AIE = Average Treatment Effect. El AIE es obtenido de estimar por separado las ecuaciones de logro para escuelas PS y MUN (supone retomos distintos entre ambas escuelas) corrigiendo por sesgo de selección a Heckman.

Fuente: Estimación propia.

V. Conclusiones

Esta investigación ha estimado la diferencia del rendimiento académico de colegios municipales (MUN) y particulares subvencionados (PS) utilizando distintos métodos. Para ello, se analizó la variación de los resultados que se obtienen al comparar el desempeño relativo de colegios MUN y PS con distintas especificaciones anidadas de variables instrumentales y estrategias de identificación, ordenadas desde la estrategia más básica —es decir, aquella que no corrige por sesgo de selección y características multinivel en las observaciones— hasta aquella que exige más de los datos. De lo anterior, cabe señalar que si bien el presente estudio se focaliza en las diferencias de puntaje entre los dos tipos de colegios, los coeficientes estimados para las restantes variable (características de los alumnos y establecimientos) coinciden con lo esperado y con lo encontrado en otros estudios.

De acuerdo a la estimación de la ecuación de logro escolar a partir de la técnica de Mínimos Cuadrados Ordinarios, las escuelas PS obtienen 11,6 puntos adicionales en la prueba SIMCE de matemáticas que las escuelas MUN. Por su parte, la estrategia HLM que controla por las características multinivel en los datos, reveló que la ganancia esperada en la prueba SIMCE de matemáticas de un estudiante atendido por un colegio PS es 12.8 puntos (versus si asistiese a un establecimiento MUN). Sin embargo, estos resultados no son consistentes debido a la presencia de sesgo de selección en los datos, puesto que los alumnos no están distribuidos aleatoriamente en los diferentes colegios. Por ello, se emplearon las técnicas de estimación de la ecuación de logro con Variables Instrumentales, las cuales revelaron que un alumno atendido por un colegio particular subvencionado obtendría 16 puntos adicionales en la prueba SIMCE de matemáticas (versus si hubiese asistido a un colegio municipal). Otra técnica de estimación utilizada para detectar y resolver el problema de sesgo de selección, es el método de corrección de Heckman, el cual estima dos ecuaciones de logro escolar por separado: una ecuación para colegios particulares subvencionados y otra para colegios municipales y por tanto permite que los retornos a las características de los alumnos y colegios difieran entre ambas dependencias. La diferencia en el desempeño se obtiene calculando el Efecto Promedio del Tratamiento (ATE) y la estimación de Tratamiento sobre los Tratados (TT). El ATE mostró que la ganancia esperada en el puntaje de la prueba SIMCE de matemáticas que obtiene un estudiante escogido para ser atendido por un colegio particular subvencionado es 14 puntos (versus si hubiese asistido a una escuela municipal). Por su parte, el TT señala que la ganancia esperada en el puntaje de la prueba SIMCE de matemáticas que obtiene un estudiante que efectivamente asiste a un establecimiento particular subvencionado es 14 puntos (versus si hubiese asistido a un colegio municipal). Finalmente, a partir de la estimación de la

ecuación de logro por TSLS con errores robustos se encontró que un estudiante atendido por un colegio particular subvencionado, obtiene en promedio 17 puntos adicionales en la prueba SIMCE de matemáticas (versus si hubiese asistido a un colegio municipal).

Por otra parte, al comparar los resultados en función de las variables instrumentales utilizadas, se tiene que las estimaciones por el método de Heckman – tuvieron menos variación en relación a las distintas especificaciones anidadas de variables instrumentales. En tanto, para una determinada especificación de variables instrumentales, todas las técnicas revelaron diferencias en sus resultados, y aquellas que controlaron por clusters de establecimientos exhibieron una mayor varianza en la estimación. En efecto, dependiendo del método utilizado, las estimaciones revelaron que un estudiante que asiste a un colegio particular subvencionado, obtiene una ganancia entre 11.63 puntos, en el caso de la estimación por Mínimos Cuadrados Ordinarios, y 17.29 puntos, en el caso de la estimación por Mínimos Cuadrados en dos etapas con errores robustos, en la prueba SIMCE de matemáticas versus si el estudiante hubiese asistido a un colegio municipal. Por su parte, las estimaciones que utilizan variables instrumentales muestran que, dependiendo de la metodología, un estudiante que asiste a un colegio particular subvencionado obtiene entre 14,1 puntos y 17,3 puntos adicionales en la prueba SIMCE de matemáticas que si hubiese asistido a un colegio municipal.

La diferencia en las estimaciones del desempeño relativo de colegios particular subvencionado y municipales se debe a las distintas ventajas y limitaciones que poseen cada una de las técnicas para resolver problemas asociados a la estructura de los datos.

Referencias

- [1] Anand P., Mizala A. y Repetto A. (2007), "Using School Scholarships to Estimate the Effect of Government Subsidized Private Education on Academia Achievement in Chile". Serie Económica No. 220. Centro de Economía Aplicada, Universidad de Chile. Santiago, Chile.
- [2] Han C. y Phillips P. (2006), "GMM with Many Moments Conditions". *Econometría*. Vol. 74, No 1, pp. 147.
- [3] Gallego F. (2005), "Inter.-School Competition, Incentives, and Outcomes: Evidence from Chile". Departamento de Economía, MIT.
- [4] Sapelli, C. & Vial, B. (2005). Private vs public voucher schools in Chile: New evidence on efficiency and peer effects. Documento de trabajo N° 289, Instituto de Economía, P. Universidad Católica de Chile.
- [5] Mizala A., Romguera P. y Ostoic C. (2004), "A Hierarchical Model for Studying Equity and Achievement in the Chilean School Choice System". Serie Economía No. 185. Centro de Economía Aplicada, Universidad de Chile. Santiago, Chile.
- [6] Dominguez M. y Lobato I. (2004), "Consistent Estimation of Models Defined by Conditional Moment Restrictions". *Econometría*. Vol. 72, No 5, pp. 1601.
- [7] Goldhaber D. y Eide E. (2003), "Methodological Thoughts on Measuring the Impact of Private Sector Competition on the Educational Marketplace". *Educational Evaluation and Policy Analysis*, pp. 217 – 232.
- [8] Angrist J. (2003), "Treatment Effect Heterogeneity in Theory and Practice". Departamento de Economía, MIT.
- [9] Sapelli C. (2003), "The Chilean Voucher System: Some New Results and Research Challenges". *Cuadernos de Economía*, vol.40, no.121, p.530-538. ISSN 0717-6821.
- [10] Sapelli, C. & Vial, B. (2003). Peer effects and relative performance of voucher schools in Chile. Documento de trabajo N° 256, Instituto de Economía, P. Universidad Católica de Chile.
- [11] Sapelli C., and Vial B. (2002), "The Performance of Private and Public Schools in the Chilean Voucher System," *Cuadernos de Economía*, vol. 39, no. 118, p. ISSN 0717-6821.
- [12] Tokman A. (2002), "Is Private Education Better? Evidence from Chile". Working paper del Banco Central de Chile. No 147.
- [13] McEwan, P.J (2001), "The Effectiveness of Public, Catholic, and Non-Religious Private Schools in Chile's Voucher System," *Education Economics*, Vol. 9, No.2, 103-128.
- [14] Aedo, Cristián y Sapelli C. (2001), "El sistema de vouchers en educación: una revisión de la teoría y evidencia empírica para Chile", *Estudios Públicos* 82, Santiago, Chile.
- [15] Mizala A. y Romaguera P. (2001), "Factores explicativos de los resultados escolares en la educación secundaria en Chile." *Trimestre Económico* 272, pp. 515-549.
- [16] Stephen G. y Whitney K. (2001), "Choosing the Number of Instruments". *Econometría*, Vol. 69, No. 5, pp. 1161.
- [17] Contreras D. (2001), *Vouchers, school choice and the access to higher education*. Universidad de Chile y Universidad de Yale. Santiago, Chile.
- [18] Mizala A. y Romaguera P. (2000), "School Performance and Choice: The Chilean Experience." *Journal of Human Resources* 35(2), pp. 392-417.

- [19] González P. (2000), “Una revisión de la reforma del sistema escolar en Chile”. Serie Económica No. 92. Centro de Economía Aplicada, Universidad de Chile. Santiago, Chile.
- [20] Woldrige J. (2000), “Econometric Analysis of Cross Section and Panel Data”. Prentice Hall. MIT press, Cambridge.
- [21] Maddala G. S. (1999), “Limeted Dependent and Qualitative Variables in Economics”. Econometric Society Monographs No. 3. Universidad de Cambridge.
- [22] Efron B., y Tibshirani R. J. (1993), “An introduction to the bootstrap”, Chapman & Hall. Londres.
- [23] Bryk A. and Raudenbush S. (1992), “Hierarchical Linear Models: Applications and Data Analysis Methods”. Thousand Oaks, California: Sage Publications.
- [24] Amemiya T. (1985), “Advanced Econometrics”. Basil Blackwell, Oxford.
- [25] Efron B. (1979), “Bootstrap methods: another look at the jackknife”. The Annals of Statistics 7, 1-26.

ANEXO 1

Descripción de variables

| Cuadro A1-1. Variables incluidas en la ecuación de logro escolar: características de los estudiantes | | | |
|--|---|---|-------------------------------------|
| Definición de variables | Abreviatura | Descripción | Fuente |
| Puntaje SIMCE de matemáticas | SIMCEmat | Puntaje del test de matemáticas del SIMCE | Base de datos SIMCE |
| Años de educación de los padres | Años de educación padres | Promedio de los años de educación alcanzados entre el padre y la madre | Cuestionario de padres y apoderados |
| Ingreso familiar | Ingreso familiar | Ingreso del hogar del estudiante | Cuestionario de padres y apoderados |
| Numero de personas en el hogar | Num. Personas del hogar | Número de personas en el hogar del estudiante | Cuestionario de padres y apoderados |
| Dummy sexo del alumno | Dummy sexo del alumno | Dummy: 1 si es hombre | Cuestionario de padres y apoderados |
| Recursos económicos percápita | Recursos econ. | Recursos económicos disponibles para cada estudiante | Cuestionario de padres y apoderados |
| Zona geográfica en que vive el estudiante | Dummy zona rural | Dummy: 1 si el estudiante pertenece a una zona rural | Cuestionario de padres y apoderados |
| Número de libros en el hogar del estudiante | Num. Libros en el hogar | Promedio de libros en el hogar del estudiante | Cuestionario de padres y apoderados |
| Elección escolar por proximidad | Elección por proximidad | Dummy: 1 si a lo menos una de las tres razones por la que el estudiante está inscrito en su escuela, es por la proximidad a su hogar | Cuestionario de padres y apoderados |
| Elección escolar porque enseña valores | Elección porque enseña valores | Dummy: 1 si a lo menos una de las tres razones por la que el estudiante está inscrito en su escuela, es por el tipo de formación valórica del establecimiento | Cuestionario de padres y apoderados |
| Elección escolar: por su bajo costo económico | Elección por su bajo costo económico | Dummy: 1 si a lo menos una de las tres razones por la que el estudiante está inscrito en su escuela, es por su bajo costo | Cuestionario de padres y apoderados |
| Elección escolar porque otros familiares estudian en la escuela | Elección porque otros familiares estudian en la escuela | Dummy: 1 si a lo menos una de las tres razones por la que el estudiante está inscrito en su escuela, es porque otros familiares estudian en la escuela | Cuestionario de padres y apoderados |
| No tuvo otra opción para la elección escolar | No tuvo otra opción | Dummy: 1 si a lo menos una de las tres razones por la que el estudiante está inscrito en su escuela, es porque no tuvo otra opción | Cuestionario de padres y apoderados |

Cuadro A1-2

Variables incluidas en la ecuación de logro escolar: características de los establecimientos

| Definición de variables | Abreviatura | Descripción | Fuente |
|---|-------------------------------------|---|-------------------------------------|
| Establecimiento particular subvencionado | Dummy particular subvencionado (PS) | Dummy: 1 si la escuela es particular subvencionada (PS) y 0 si es municipal (MUN) | Base de datos SIMCE |
| Pago por escolaridad | Pago por escolaridad | Pago promedio por escolaridad | Cuestionario de padres y apoderados |
| Establecimientos de sólo hombres | Establecimiento de sólo hombres | Dummy: 1 si la escuela es sólo para niños | Base de datos SIMCE |
| Establecimientos de sólo mujeres | Establecimiento de sólo mujeres | Dummy: 1 si la escuela es sólo para niñas | Base de datos SIMCE |
| Experiencia de los profesores | Experiencia docentes | Años de experiencia del profesor | Base de datos SIMCE |
| Número de alumnos por profesor | Cuociente alumno-profesor | División del número de alumnos entre el número de profesores por escuela | Cuestionario de padres y apoderados |
| Logaritmo del número de alumnos matriculados en la escuela | Ln matriculados | Logaritmo del número de alumnos matriculados por escuela | Base de datos SIMCE |
| Establecimiento educativo religioso | Establecimiento religioso | Dummy: 1 si el establecimiento es sostenido por congregación religiosa | Base de datos SIMCE |
| Número de colegios particulares subvencionados (PS) en la comuna de residencia del estudiante | Num. Colegios PS | Número de colegios PS en la comuna de residencia del alumno | Cuestionario de padres y apoderados |
| Número total de colegios en la comuna de residencia del estudiante | Num. Colegios | Número de colegios en la comuna de residencia del alumno | Cuestionario de padres y apoderados |

ANEXO 2
Resultados de las estimaciones de la ecuación de selección

| Cuadro A2-1 Estimación de la ecuación con IV | |
|---|--|
| Variables Instrumentales | Técnica de estimación (primera etapa de TSLS) |
| Dummy particular subvencionado (dependiente) | Estimación Probit |
| Constante | -8.3E-01 (8.7E-03)** |
| Número de colegios particulares subvencionados (PS) en la comuna de residencia del estudiante | 2.1E-02 (2.9E-04)** |
| Número total de colegios en la comuna de residencia del estudiante | -6.2E-03 (2.0E-04)** |
| Elección porque otros familiares estudian en la escuela | -1.0E-01 (8.0E-03)** |
| Pago por escolaridad | 8.1E-05 (8.6E-07)** |
| Elección por proximidad | -1.1E-01 (6.9E-03)** |
| Elección porque enseña valores | 6.3E-01 (7.9E-03)** |
| Elección por su bajo costo económico | -1.9E-01 (7.0E-03)** |
| No tuvo otra opción | -1.3E-01 (1.1E-02)** |
| Num. Libros en el hogar | 2.1E-03 (6.2E-05)** |
| Establecimiento religioso | 3.0E+00 (4.8E-02)** |

Nota: Error estándar entre paréntesis

** Significativo al 95% de confianza

* Significativo al 90% de confianza

Cuadro A2-2
Estimación de la ecuación de selección (Heckman)

| Variables | Técnicas de estimación(primer etapa de TSLS) | |
|---|--|-------------------------|
| | Ecuac. 1: Colegios PS | Ecuac. 2: Colegios MUN |
| | TSLS | TSLS |
| Dummy particular subvencionado (dependiente) | | |
| Constante | 2.7E+00 (6.2E-02)** | -2.9E+00 (6.2E-02)** |
| Años de educación padres | 7.4E-02 (2.0E-03)** | -7.1E-02 (2.0E-03)** |
| Ingreso familiar | 3.4E-07 (2.9E-08)** | -3.4E-07 (3.0E-08)** |
| Num. Personas del hogar | -1.3E-02 (2.7E-03)** | 1.3E-02 (2.8E-03)** |
| Dummy sexo del alumno | 3.8E-03 (1.0E-02)** | -7.1E-03 (1.0E-02) |
| Dummy zona rural | -1.4E+00 (2.1E-02)** | 1.4E+00 (2.1E-02)** |
| Establecimiento de sólo mujeres | -2.8E-01 (3.0E-02)* | 2.8E-01 (3.0E-02)* |
| Establecimiento de sólo hombres | 2.9E-01 (3.5E-02)** | -2.8E-01 (3.5E-02)* |
| Dummy JEC | -9.9E-02 (1.4E-02)** | 1.0E-01 (1.4E-02)** |
| Recursos econ. | 1.3E-04 (4.6E-05)** | 1.1E-04 (4.5E-05)* |
| Ln matriculados | 1.3E+00 (1.2E-02) | -1.3E+00 (1.2E-02)** |
| Cuociente alumno-profesor | -1.2E-02 (3.2E-04)** | 1.2E-02 (3.2E-04)* |
| Experiencia docentes | 2.0E+00 (1.2E-02)* | 2.0E+00 (1.2E-02)* |
| Número de colegios particulares subvencionadas (PS) en la comuna de residencia del estudiante | 1.1E-02 (4.4E-04)** | -1.1E-02 (4.4E-04)** |
| Número total de colegios en la comuna de residencia del estudiante | -4.5E-04 (2.9E-04)** | 4.3E-04 (2.9E-04)** |
| Elección porque otros familiares estudian en la escuela | -1.3E-03 (1.2E-02)** | 3.9E-03 (1.2E-02)** |
| Pago por escolaridad | 6.6E-05 (5.7E-07)** | -6.9E-05 (5.9E-07)** |
| Elección por proximidad | -2.2E-01 (1.1E-02)** | 2.3E-01 (1.1E-02)** |
| Elección porque enseña valores | 5.6E-01 (1.2E-02)** | -5.6E-01 (1.2E-02)** |
| Elección por su bajo costo económico | -1.9E-01 (1.1E-02)** | 1.8E-01 (1.1E-02)** |
| No tuvo otra opción | -2.2E-01 (1.7E-02)** | 2.1E-01 (1.7E-02)** |
| Num. Libros en el hogar | 1.1E-03 (1.0E-04)** | -1.1E-03 (1.0E-04)** |
| Establecimiento religioso | 3.3E+00 (6.4E-02)** | -3.3E+00 (6.7E-02)** |

Nota: Error estándar entre paréntesis

** Significativo al 95% de confianza

* Significativo al 90% de confianza

ANEXO 3

Test de Sargan

Sargan demostró que el siguiente estadístico se distribuye, bajo la hipótesis nula de validez de los instrumentos z_{ijk} , como una chi-cuadrado con $P - K$ grados de libertad:

$$\frac{\hat{\mathbf{t}}_{IV} \mathbf{Z} (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{Z}' \hat{\mathbf{t}}_{IV}}{\hat{\sigma}_\tau^2} \equiv \frac{SE}{\hat{\sigma}_\tau^2} \overset{a}{\sim} \chi_{P-K}^2,$$

Siendo $\hat{\mathbf{t}}_{IV}$, el vector de residuos estimados de la ecuación de logro (ecuación 5), \mathbf{Z} es la matriz de variables instrumentales (z_{ijk}) y $\hat{\sigma}_\tau^2$ es la varianza estimada de los residuos de la regresión original (ecuación de logro).

El procedimiento para obtener el valor numérico del test de Sargan es:

1. Estimar el modelo original por OLS y obtener los residuos estimados \hat{t}_{ij} .
2. Regresionar la variable de residuos estimados \hat{t}_{ij} sobre las variables instrumentales z_{ijk} ; obtener la Suma Explicada (SE) de la regresión y la varianza estimada de τ_{ij} ($\hat{\sigma}_\tau^2$).
3. Plantear el siguiente estadístico:

$$\frac{SE}{\hat{\sigma}_\tau^2} \overset{a}{\sim} \chi_{P-K}^2$$

Los grados de libertad de la distribución chi-cuadrado son $P - K$, siendo P el número de instrumentos utilizados y K el número de variables explicativas del modelo original. Si el valor calculado de la distribución chi-cuadrado (con un grado de libertad) es menor que el valor de la tabla chi-cuadrado, se concluye que el conjunto de instrumentos $\{z_{ijk}\}_{k=1}^K$ es válido. En caso de que el valor calculado sea mayor al valor de la tabla chi-cuadrado, Sargan recomienda emplear un conjunto alternativo de variables instrumentales y aplicar nuevamente el test.