

VALOR AGREGADO MULTINIVEL Y FACTORES CONTEXTUALES EN EDUCACIÓN: UNA COMPARACIÓN NO PARAMÉTRICA ROBUSTA

Claudio Thieme¹

Facultad de Economía y Empresa

Universidad Diego Portales (Chile)

Emili Tortosa-Ausina

Departamento de Economía

Universitat Jaume I, Castellón (España)

Diego Prior

Departamento de Economía de la Empresa

Universitat Autònoma de Barcelona (España)

René Gempp

Facultad de Economía y Empresa

Universidad Diego Portales (Chile)

¹Correspondencia: C. Thieme, Facultad de Economía y Empresa, Universidad Diego Portales, Av. Manuel Rodríguez Sur 253, Santiago de Chile. E-mail: claudio.thieme@udp.cl. Teléfono: 56 -2-6762211

Claudio Thieme, René Gempp y Emili Tortosa-Ausina agradecen el financiamiento del Fondo Nacional de Desarrollo Científico y Tecnológico (FO NDECYT), proyecto #1121164. Diego Prior y Emili Tortosa-Ausina agradecen el apoyo financiero del Ministerio de Ciencia e Innovación (ECO2010-18967/ECON y ECO2011-27227/ECON). Emili Tortosa-Ausina también agradece el financiamiento de la Generalitat Valenciana (PROMETEO/2009/066). Esta investigación utilizó como fuente de información las bases de datos del SIMCE del Ministerio de Educación de Chile. Los autores agradecen al Ministerio de Educación el acceso a la información. Todos los resultados del estudio son de responsabilidad de los autores y en nada comprometen a dicha Institución.

VALOR AGREGADO MULTINIVEL Y FACTORES CONTEXTUALES EN EDUCACIÓN: UNA COMPARACIÓN NO PARAMÉTRICA ROBUSTA

1.- INTRODUCCIÓN

Un elemento central de los esfuerzos que los países realizan para implementar mejoras en sus sistemas educativos, es el desarrollo de indicadores para evaluar la calidad de la educación que imparten (Battauz, Bellio & Gori, 2011). En particular, en el campo de la política pública en educación se aprecia una creciente preocupación en la evaluación del aprendizaje de los estudiantes (Denvir & Brown, 1986; Ercikan, 2006).

En muchos países, lo anterior se ha traducido en la adopción de sistemas de rendición de cuentas (*accountability*), cuyo propósito es evaluar, e informar a los actores involucrados, los resultados de logro académico de los estudiantes u otros indicadores de calidad de las instituciones educativas. Esta tendencia se ha visto fortalecida con la acumulación de evidencia empírica que demuestra que mecanismos adecuados de rendición de cuentas (los que son capaces de establecer responsabilidades claras para todos los actores del sistema) permiten mejorar las prácticas organizacionales al interior de las escuelas (Rouse et al., 2007) y optimizar los resultados educacionales (Carnoy y Loeb, 2002; Hanushek y Raymond, 2005).

Todo lo anterior presupone que las metodologías de evaluación e indicadores utilizados para ello son los adecuados. Al respecto, los significativos avances metodológicos, principalmente el desarrollo de modelos multinivel (Bryk y Raudenbush, 1992; Goldstein, 1995) han mejorado la conceptualización y medición de las causas que explican el aprendizaje de los estudiantes (Aitkin y Longford, 1986). El concepto general es que el logro académico de los estudiantes depende tanto de características personales de los alumnos como de las características de la escuela y su contexto. Para analizar estas situaciones, los distintos niveles son considerados como sistemas jerárquicos de alumnos y escuelas, con individuos y grupos definidos en separados niveles jerárquicos, a través de variables que son definidas en cada nivel (Hox, 2002).

Este importante progreso permite solucionar el principal problema metodológico de los estudios pioneros en este ámbito, al descomponer los distintos efectos anidados que explican los resultados educativos de los estudiantes, pudiendo determinar el porcentaje de logro de los estudiantes que se debe a las diferentes variables en diferentes niveles organizacionales (distrito, escuela, clase, alumno).

Lo anterior es de alta relevancia en el momento de arbitrar medidas (específicas a nivel de alumno, colegio o entorno), y por tanto, permite aportar conocimientos útiles tanto para desarrollar procesos de mejora en los establecimientos educativos, desalentar las prácticas oportunistas de los gestores, señalar la correcta asignación de recursos, premios y castigos, como para la toma de decisiones de políticas públicas.

En tal sentido, una revisión bibliográfica muestra que existe un amplio acuerdo entre los especialistas internacionales respecto a que las comparaciones entre escuelas no pueden establecerse sobre los resultados brutos de logro académico, sino que deben basarse en el progreso de sus estudiantes a lo largo de los años de escolaridad, idealmente corrigiendo el efecto del rendimiento inicial y de otras condiciones de entrada relevantes (Goldstein et al., 1993; Goldstein y Thomas, 1996; Gray, Jesson, Goldstein, Hedger y Rasbash, 1996; Mortimore, Sammons y Thomas, 1994; Sammons, 1995). De hecho, el consenso entre los investigadores es que una escuela verdaderamente efectiva no es aquella cuyos estudiantes obtienen mejores resultados, sino aquella en la que los estudiantes *progresan* más allá de lo que puede esperarse, debido al efecto de la escuela. En otras palabras, que las escuelas efectivas son aquellas que **agregan valor** a los resultados de sus alumnos en comparación con otras escuelas que atienden a poblaciones estudiantiles con características equivalentes.

En el contexto del accountability, se entiende por **Valor Agregado (VA)** de una escuela a *la contribución que realiza al progreso neto de los estudiantes hacia objetivos de aprendizaje establecidos, una vez eliminada la influencia de otros factores ajenos a la escuela que pueden contribuir a dicho progreso* (Meyer, 1997). Para estimar el VA se utilizan un conjunto de procedimientos estadísticos que permiten hacer inferencias sobre la eficacia de las escuelas a través del seguimiento de la trayectoria de los estudiantes, analizando los resultados de dos o más años (Tekwe et al., 2004). Estos datos de la evolución de los estudiantes se transforman en **indicadores de VA** que pueden utilizarse para evaluar la eficacia de la escuela de manera mucho más válida y justa que el simple resultado promedio de sus estudiantes en un año cualquiera (Raudenbush, 2004). Durante los últimos años ha sido creciente la incorporación de modelos de VA más complejos (modelos contextuales), que permiten ajustar también los resultados por variables socioeconómicas y demográficas que están fuera de control por parte de la escuela, con el objeto de obtener una estimación del desempeño neto del establecimiento eliminando el efecto de las diferencias preexistentes entre los alumnos (Ballou, Sandres & Wright, 2004).

Debido a su capacidad para estimar la contribución neta de la escuela a los resultados de aprendizaje, los modelos de VA permiten comparar la efectividad de las escuelas aunque atiendan a poblaciones muy diversas de estudiantes y guiar los cambios educativos necesarios, tanto en el nivel de la escuela, como de las reformas políticas (Drury y Doran, 2004; McCaffrey, Lockwood, Koretz, y Hamilton, 2003). De este modo, los indicadores de VA aparecen como una alternativa sumamente atractiva para los gobiernos (que necesitan contar con medidas de accountability equitativas), para los políticos (que desean que la evaluación de las escuelas considere su diversidad), para los investigadores (que requieren estudiar los factores que contribuyen a la efectividad escolar mediante indicadores no contaminados por las condiciones de entrada de los alumnos), para los profesores y directivos escolares (que solicitan medidas objetivas de su desempeño, adaptadas al tipo de población estudiantil que atienden), para los apoderados (que necesitan elegir escuelas para sus alumnos en función del aporte que ésta haga al estudiante) y para la sociedad en su conjunto, en la medida que implica una evaluación más justa y equitativa de las escuelas del país.

Todas las razones anteriores han provocado que internacionalmente los Modelos de Valor Agregado sean cada vez más populares tanto en la comunidad de investigadores, como entre los políticos y gestores escolares. Un reporte del año 2008 de la OECD alienta decididamente a sus países miembros a emplear medidas de VA para evaluar el desempeño de sus escuelas y entrega lineamientos específicos y pautas claras para hacerlo correctamente.

En este ámbito, existen una gran cantidad de modelos estadísticos para su estimación pero difieren en la definición e inclusión de variables de ajuste (Tewke et al., 2004). Sin embargo, la postura mayoritaria es la de incluir variables de ajuste, especialmente cuando se establece alguna forma de rendición de cuentas o de difusión de los resultados, ya que la equidad será cuestionable si no se tienen en cuenta características contextuales de los estudiantes y de las escuelas (McCaffrey et al., 2003; McCaffrey et al., 2004).

A pesar de la gran cantidad de aportes metodológicos y empíricos, esta línea de investigación no está exenta de críticas (Bock, Wolfe y Fisher, 1996; Kuppermintz, 2003; Mccafrey, et al., 2003). Una de ellas dice relación que la naturaleza de su estimación corresponde a una comparación con la media, no suponiendo un real incentivo para el desempeño de excelencia. En efecto, la gran mayoría de los estudios de valor agregado utilizan regresión o análisis multinivel. Ejemplo de ello son los estudios de Goldstein et al. (1993), Gray et al. (1996), Cervini, 2009; Blanco, 2010, entre muchos otros.

Por ello, los modelos provenientes desde la investigación en operaciones que evalúan el desempeño de las organizaciones educativas, pero principalmente utilizando modelos frontera no paramétricos (ya sea utilizando *Data Envelopment Analysis (DEA)* o *Free Disposal Hull (FDH)*), que proporcionan ventajas en este campo al adecuarse con el concepto de comparación con un óptimo y permitir la existencia de varios outputs de manera simultánea, resultan una alternativa más adecuada. En este campo, se aprecia en los últimos años una importante evolución al desarrollarse modelos cada vez más robustos y capturar el concepto de valor agregado (Silva Portela y Thanassoulis, 2001; De Witte et al. 2010) y análisis multinivel (Thieme et. al 2012).

Sin embargo y a pesar de los avances que ello supone, las metodologías existentes no han capturado aún el concepto de valor agregado contextualizado, por lo que la sola inclusión de datos de alumno con resultados ex ante y ex post de su logro académico se estaría incorrectamente estimando el efecto escuela ya que ello supone erróneamente la existencia de inputs óptimos no controlables por los gestores, sin responder por tanto a un verdadero análisis de valor agregado que pretende medir correctamente la contribución de la escuela, eliminando la influencia de factores ajenos a esta unidad organizacional que pudieron haber contribuido a este progreso.

Por tal razón, el objeto de este artículo es doble y pretende aportar desde la perspectiva metodológica y empírica a dos brechas existentes a esta evolución. En primer lugar, el aporte es metodológico al proponer un modelo de frontera robusta contextualizado de valor agregado que recoge los aportes de los trabajos de valor agregado contextualizado provenientes de la investigación en educación; los trabajos de metafrontera (Battese, Rao y O'Donnel, 2004), y los de

análisis frontera de *order-m* (Cazals, Florens y Simar, 2002), provenientes de la investigación de operaciones. En segundo lugar, el aporte es empírico al utilizar esta propuesta en un país como Chile, que si bien tiene una larga historia de evaluación desde 1980 a nivel nacional con pruebas estandarizadas en tres distintos niveles de escolaridad, utiliza indicadores brutos de logro académico de sus estudiantes como medida de desempeño de las escuelas que los cobijan.

El modelo propuesto se aplica a una muestra de 47.076 alumnos, pertenecientes a 948 centros educativos de educación básica de Chile de todos los tipos que rindieron las pruebas de Lenguaje y Matemáticas del Sistema de Medición de la Calidad Educativa (SIMCE) de octavo básico (13 años) el año 2009 y de cuarto básico el año 2005 (a la edad de 9 años) y para los cuales se contaba con información socioeconómica individualizada de sus familias. Con el objeto de contar con una muestra confiable y homogénea de la estimación agregada a nivel de establecimientos, se consideraron sólo los alumnos de los establecimientos que habiendo rendido ambas pruebas y se contaba con información socioeconómica a nivel individual, sus escuelas cumplieran el requisito de tener 30 o más alumnos en estas condiciones y que dicho valor fuese al menos del 60% del total de alumnos del establecimiento que rindieron la prueba el año 2009.

El texto que sigue es organizado de la siguiente manera: en la segunda sección, se describe el marco teórico relevante. En la tercera sección se entregan detalles de la metodología propuesta. Los antecedentes de la aplicación empírica y la descripción de la base de datos utilizada se presentan en la cuarta sección. Los resultados comparativos entre modelos son analizados en la quinta parte, para finalizar con las principales conclusiones del estudio.

2.- Fundamentos teóricos

Nuestro trabajo difiere principalmente en dos aspectos de las propuestas existentes: en el modelo de datos y en la metodología de evaluación utilizada. Por ende, resulta conveniente realizar un análisis de los fundamentos al respecto.

2.1.- El modelo de datos

A pesar de que los diversos sistemas nacionales o estatales de *accountability* tienen como meta evaluar el aprendizaje que los estudiantes obtienen en una determinada escuela, difieren de manera significativa entre sí en la operacionalización del concepto de calidad y consecuente con ello, en el modelo de evaluación utilizado. Estas diferencias conducen a diferencias en la definición de qué escuelas son recompensadas y de cuáles son sujetas a recibir atención por parte del sistema de *accountability*, entre otras muchas otras repercusiones. Estos distintos modelos de evaluación han sido clasificados de diversa forma (Carlson, 2006; XXXX).

Uniando estas clasificaciones con la literatura de la medición del desempeño escolar en el área de investigación de operaciones podemos relevar dos aspectos esenciales: (i) existe una clara evolución metodológica y empírica de evaluación a nivel de escuela hacia una evaluación del

desempeño de la escuela con datos a nivel de alumno; (ii) A su vez, en estos estudios más recientes de frontera no paramétrica a nivel de alumno podemos distinguir dos grandes enfoques para monitorear el desempeño de las escuelas: los modelos de status que utilizan un único año de evaluación de los resultados académicos de los alumnos, y los modelos de valor agregado (o ganancia) que utilizan dos o más años de resultados de los estudiantes. Del mismo modo, en estos enfoques podemos distinguir los modelos que utilizan variables contextuales (a nivel de alumno y/o escuela) para evaluar el desempeño escolar y los que no las utilizan. Esta situación y las preguntas fundamentales que intentan responder pueden ser representadas en una matriz 2*2.

Tabla 1: Tipos de modelos de evaluación

| | Sin variables contextuales | Con variables contextuales |
|---|--|---|
| Estatus (1 evaluación del alumno) | Modelo 0 ¿Cuál es el nivel de logro absoluto de los estudiantes en esta escuela? | Modelo 1 ¿Cuál es el logro de los estudiantes en esta escuela, acorde a factores contextuales de sus alumnos y/o de la escuela? |
| valor agregado (ganancia) (2 o más evaluaciones del alumno) | Modelo 2 ¿Es una escuela efectiva? Dado el nivel de logro de sus estudiantes cuando ellos entran, ¿cuánto ellos aprenden o desarrollan mientras ellos están en la escuela? | Modelo 3 ¿Es una escuela efectiva? Dado el nivel de logro de sus estudiantes cuando ellos entran, ¿cuánto ellos aprenden o desarrollan mientras ellos están en la escuela, acorde a factores contextuales de sus alumnos y/o de la escuela? |

El modelo 0 implica considerar en el modelo de evaluación sólo outputs de logro académico de los alumnos y un valor unitario, similar para todos los alumnos, como input. Si bien esta situación es factible de modelar, el considerar evaluar a todos los estudiantes y escuelas considerando óptimas y similares condiciones de entrada y de perfil socioeconómico, pareciera no corresponder a una evaluación adecuada y por ende la literatura de investigación de operaciones no da cuenta de una evaluación de estas características. En efecto, estos modelos atribuyen todo el logro académico de los estudiantes a los establecimientos educativos, desconociendo una abundante y sólida literatura que indica que gran parte de la explicación de logro académico de los estudiantes proviene de factores contextuales, no controlables ni atribuibles a la escuela (Teddlie y Reynolds, 2000). Las principales críticas apuntan a que estas mediciones podrían generar incentivos perversos para la consecución del objetivo deseado, asignando menos recursos a aquellos alumnos con peores resultados que no contribuyen a acercar al establecimiento al objetivo deseado. Así también, esto podría generar una selección de alumnos al interior de los establecimientos o forzar a una autoselección de éstos (Wilson, 2004). Debido a la alta correlación entre el NSE (nivel socioeconómico) y el desempeño de los estudiantes, se considera que las medidas en niveles no logran transmitir información acerca de la eficacia de los establecimientos pues establecimientos con mayor NSE tendrán mayor facilidad, dada su composición, para

alcanzar los objetivos (Wilson and Piebalga, 2008). Al respecto Goldschmidt y Choi (2007) concuerdan con los riesgos señalados por Wilson (2004) y consideran que la tarea por alcanzar un porcentaje de alumnos en un nivel de logro dado termina por enfatizar en las características de los alumnos. Sin embargo, y a pesar de lo anterior, hemos querido rescatar este modelo ya que Chile ha adoptado un modelo de estatus de logro académico sin variables contextuales para la evaluación de sus escuelas, y por tanto bien vale la pena comparar estos resultados con los que se obtendrían en el modelo que proponemos en este estudio.

El modelo 1 contempla evaluar el desempeño de las escuelas considerando el valor observado de logro académico de los estudiantes como output, y como inputs los factores del entorno no atribuibles al centro, ya sea a nivel de alumno como a nivel de escuela. Al respecto, Thieme et al. (2012) plantean un modelo multinivel que incorpora tanto factores del proceso educativo en educación atingentes a la escuela como factores contextuales a nivel de alumno y de escuela. A pesar del considerable avance que ello representa, el no considerar como inputs variables de logro académico de entrada de los alumnos es similar que considerar que ellas son iguales y óptimas para todos los estudiantes, situación que obviamente dista de la realidad y podría conducir a un error de interpretación de los resultados.

El modelo 2 corresponde a un modelo de valor agregado puro donde se consideran exclusivamente como inputs y outputs, resultados de logro académico de los estudiantes, tanto al inicio como al final del proceso educativo bajo evaluación. Desde la investigación en educación se reconoce que las medidas de valor agregado (de ganancia) son una medida más informativa acerca de la eficacia de los establecimientos, aislando el efecto de la escuela del progreso de los alumnos (Wilson and Piebalga, 2008) y permiten reducir los incentivos a comportamientos deshonestos. Tres recientes estudios concuerdan con este modelo. Sin embargo, al igual que en el modelo anterior, adolece de no controlar por elementos ajenos a la escuela que influyen en dicho proceso.

El modelo 3 resuelve las deficiencias de los modelos 1 y 2. Sin embargo, no ha sido estudiado aún desde la investigación en operaciones. Por el contrario, desde la investigación en educación, este tipo de modelos han surgido con fuerza como un refinamiento de las medidas de ganancia, y se le ha denominado CVA (*contextual value added*). El CVA se utilizó por primera vez el año 2006 en las escuelas Británicas, con esta medida se pretendía aislar el impacto real de la escuela sobre el progreso del alumno. En este sentido existe evidencia de la relación que existe entre las características del establecimiento y las mejoras que éstos experimentan. Levacic y Woods (2002) señalan que elementos como el rendimiento del establecimiento en el año base de la medición o la cantidad de niños en desventaja social y la variación en el presupuesto, tienen una incidencia significativa en el desempeño que alcanzan los establecimientos. El cálculo del CVA incluye una estimación de los resultados según una serie de factores como lo son el sexo, la etnia, la lengua de origen, entre otros. La diferencia entre la estimación del modelo y el resultado que realmente obtiene el alumno se denominó CVA (Wilson and Piebalga, 2008).

Como hemos visto, a pesar de los grandes avances que suponen los modelos 1 y 2 desde la investigación de operaciones, es el modelo 3 el que permite aislar de mejor forma el impacto real de la escuela sobre el progreso de los estudiantes como lo indican los avances desde el área tradicional de evaluación escolar que utiliza preferentemente análisis multinivel paramétrico. Este modelo, que aún no ha sido analizado, presenta diferencias sustanciales con respecto a sus predecesores.

2.2. La metodología de evaluación

Como ya hemos señalado, en los últimos años ha existido un considerable avance metodológico en la evaluación del desempeño escolar, especialmente producto del desarrollo de modelos multinivel (Bryk y Raudenbush, 1992; Goldstein, 1995). El concepto general es que el logro académico de los estudiantes depende tanto de características personales de los alumnos como de las características de la escuela y su contexto. Para analizar estas situaciones, los distintos niveles son considerados como sistemas jerárquicos de alumnos y escuelas, con individuos y grupos definidos en separados niveles jerárquicos, a través de variables que son definidas en cada nivel (Hox, 2002).

Este importante progreso permite solucionar el principal problema metodológico de los estudios pioneros en este ámbito, al descomponer los distintos efectos anidados que explican los resultados educativos de los estudiantes, pudiendo determinar el porcentaje de logro de los estudiantes que se debe a las diferentes variables en diferentes niveles organizacionales (distrito, escuela, clase, alumno).

En este ámbito, existen una gran cantidad de modelos estadísticos para su estimación pero difieren en la definición e inclusión de variables de ajuste (Tewke et al., 2004). Sin embargo, la postura mayoritaria es la de incluir variables de ajuste, especialmente cuando se establece alguna forma de rendición de cuentas o de difusión de los resultados, ya que la equidad será cuestionable si no se tienen en cuenta características contextuales de los estudiantes y de las escuelas (McCaffrey et al., 2003; McCaffrey et al., 2004).

A pesar de la gran cantidad de aportes metodológicos y empíricos, esta línea de investigación no está exenta de críticas (Bock, Wolfe y Fisher, 1996; Kuppermintz, 2003; Mccafrey, et al., 2003). Una de ellas dice relación que la naturaleza de su estimación corresponde a una comparación con la media, no suponiendo un real incentivo para el desempeño de excelencia. En efecto, la gran mayoría de los estudios de valor agregado utilizan regresión o análisis multinivel. Ejemplo de ello son los estudios de Goldstein et al. (1993), Gray et al. (1996), Cervini, 2009; Blanco, 2010.

Por ello, los modelos provenientes desde la investigación en operaciones que evalúan el desempeño de las organizaciones educativas, pero principalmente utilizando modelos frontera no paramétricos (ya sea utilizando *Data Envelopment Analysis (DEA)* o *Free Disposal Hull (FDH)*), que proporcionan ventajas en este campo al adecuarse con el concepto de comparación con un

óptimo y permitir la existencia de varios outputs de manera simultánea, resultan una alternativa más adecuada. En el área de educación son innumerables los estudios que se han realizado utilizando estas técnicas, como por ejemplo Bessent et al. (1982), Ruggiero et al. (1995), Mancebón y Mar Molinero (2000), Bifulco y Bretschneider (2001), Mizala et al. (2002), Ouellette y Vierstraete (2005). Sin embargo, estas metodologías tampoco están exentas de problemas, en particular la característica determinista y no probabilística de estos modelos, la dimensionalidad y fuerte dependencia a la inexistencia de *outliers*, y la incorporación exclusiva de datos a nivel de alumno, con lo cual se estaría incorrectamente estimando el efecto escuela ya que ello supone erróneamente la existencia de inputs óptimos (controlables o no controlables), sin establecer por tanto un análisis multinivel.

Para solucionar aquello, nuestro enfoque sigue esta línea de trabajo, haciéndose cargo de los problemas descritos anteriormente, a través de la integración de los trabajos de análisis multinivel contextualizado provenientes de la investigación en educación (McCaffrey et al., 2003; McCaffrey et al., 2004), con los trabajos provenientes de la investigación operativa, en particular, los de metafrontera (Battese, Rao y O'Donnell, 2004) y los de análisis frontera de orden-*m* (Cazals, Florens y Simar, 2002).

En tal sentido, este trabajo es coincidente con la evolución que han experimentado los estudios en esta área. En efecto, el estudio pionero corresponde a Silva Portela y Thanassoulis (2001) quienes (siguiendo un modelo 2) descomponen la eficiencia global en efecto estudiante y efecto escuela, utilizando DEA (*Data Envelopment Analysis*) como metodología de análisis. Posteriormente, usando el mismo modelo de datos, De Witte et al., (2010) avanzan en el plano metodológico planteando una metodología de frontera robusta, en particular de orden-*m*, para esta evaluación.

Sin embargo, ambos estudios adolecen de considerar factores contextuales, tanto a nivel de alumno como de escuela, en la evaluación del desempeño escolar.

Por su parte, Thieme et al. (2012), siguiendo un modelo de datos 2 realizan una descomposición multinivel de frontera robusta con datos de corte transaccional (un modelo de status contextual); pero se diferencia de ellos al utilizar adicionalmente resultados de logro académico de los estudiantes en dos momentos del tiempo. Estos autores utilizan igualmente orden-*m* en sus estimaciones, logrando una mayor descomposición, y por ende explicación, de los distintos efectos que componen la ineficiencia global.

En resumen, la propuesta de una evaluación de valor agregado multinivel contextualizado de frontera robusta que mide y separa los efectos netos del alumno y de la escuela, controlando por nivel socioeconómico, tanto a nivel de alumno como a nivel de escuela; utilizando un modelo de análisis de frontera robusta, eliminado (o al menos disminuyendo drásticamente) los potenciales problemas causados por la existencia de *outliers* y de dimensionalidad de las evaluaciones, resulta la evolución normal de esta área de estudio.

3.- Metodología

Nuestra propuesta se inspira en la descomposición planteada inicialmente por Silva Portela y Thanassoulis (2001). En este trabajo, los autores consideran dos fronteras: la frontera local, específica para cada escuela orientada a estimar la eficiencia de cada estudiante dentro de su propia escuela (student-within-school efficiency), y la frontera global, usada para estimar la eficiencia del estudiante dentro de todas las escuelas (student within-all-schools efficiency). La distancia a la frontera local depende de la eficiencia del estudiante (la que denominamos efecto estudiante, de aquí en adelante STE), mientras que la distancia que separa la frontera local de la global se refiere a la eficiencia de la escuela (la que denominamos efecto escuela, o SCE). El modelo tipo 2 de la figura 1 da cuenta de esta descomposición.

El estudiante (c) bajo análisis obtiene los resultados representados por y_c , con un nivel de recursos x_c (en este caso, el puntaje de logro en una prueba estandarizada al inicio del segundo ciclo de educación básica, bajo análisis).

Cuando comparamos el desempeño académico de este estudiante con respecto a su frontera local (que corresponde a la escuela en la cual está matriculado), es obvio que el estudiante c es ineficiente, dado que en la frontera se ubican estudiantes más eficientes quienes se atienden en la misma escuela y obtienen mejores resultados (y'_2) con el mismo nivel de recursos o inputs (en este caso, conocimiento previo, x_c). Acorde a ello, el efecto estudiante (el student-within-school efficiency, en términos de Silva Portela y Thanassoulis, 2001) puede ser determinado como el ratio: el potencial output dividido por el output observado ($STE_2 = y'_2/y_c$). El efecto estudiante es mayor a la unidad cuando el estudiante es ineficiente (como el caso presentado en la Figura 1), e igual a la unidad en caso contrario. Cuando se compara la frontera global (metafrontera, o el student within-all-schools efficiency en términos de Silva Portela y Thanassoulis, 2001), el coeficiente de eficiencia para el estudiante bajo análisis es $OE = OE_2 = y''_2/y_c$.

Teniendo estas dos fronteras de referencia, el efecto escuela (una suerte de gap tecnológico que resulta del ratio que separa la frontera de una escuela específica de la frontera global) es determinada a través de la comparación entre la frontera global y las fronteras locales respectivas ($SCE_2 = y''_2/y_2 = OE_2/STE_2$).

En resumen, la propuesta de Silva Portela y Thanassoulis (2001) descompone la eficiencia global en dos efectos:

$$\text{Eficiencia Global (OE2)} = \text{Efecto Estudiante (STE2)} * \text{Efecto Escuela (SCE2)} \quad (1)$$

Acorde a la taxonomía propuesta en la tabla 1, la descomposición corresponde a un modelo de evaluación de tipo 2, o de valor agregado puro (cuando la dotación de inputs corresponde al conocimiento de entrada del estudiante). Como hemos mencionado, nosotros seguimos parcialmente esta propuesta al considerar al estudiante como unidad de análisis. Sin embargo, contrario a Silva Portela y Thanassoulis (2001), nuestro interés es desarrollar una descomposición multinivel de valor agregado contextualizado (CVA). Esto implica considerar no sólo resultados académicos previos del estudiante, sino también factores contextuales relacionados con la escuela y el propio estudiante. Para tal fin, nosotros seguimos una propuesta previa clasificada como tipo 1 (que considera sólo variables de status contextual pero no de valor agregado), por Thieme et. al

(2012) que introduce sucesivas descomposiciones de variables específicas. El efecto final es la modificación del efecto escuela, después de introducir variables contextuales que dan cuenta del nivel socioeconómico promedio de las familias que acuden al mismo establecimiento educativo.

En la figura 1 ilustramos las diferencias entre la propuesta de Silva Portela y Thanassoulis (2001), precisamente un ejemplo del tipo 2 consiste en la evaluación de valor agregado puro, y nuestra propuesta, que da cuenta del modelo de evaluación de tipo 3, consiste en una evaluación de valor agregado contextual.

Con respecto al efecto estudiante, vemos que $y'_2 > y'_3$ porque consideramos como inputs no sólo los resultados de logro académico previo sino también el nivel socioeconómico y cultural de la familia del estudiante. Esto significa en orden a estimar y'_3 se debe considerar el contexto familiar de cada estudiante mientras que y'_2 implícitamente asume que su contexto es óptimo y no interfiere sobre los resultados de logro académico de los estudiantes. En otras palabras, el *benchmark* y'_3 proviene de un estudiante que tiene similar nivel socioeconómico y de resultados académicos previos pero que obtiene mejores resultados al finalizar el ciclo educativo. Por contraste, el *benchmark* y'_2 proviene de otro estudiante que tiene mejor situación socioeconómica.

El impacto de las variables contextuales también ejerce impacto sobre el efecto escuela. En efecto, esto es también observable tal que $y''_2 > y''_3$ porque el entorno socioeconómico debiera afectar el logro académico de los estudiantes en y''_3 , pero precisamente, el Modelo de tipo 2 asume que este factor no tiene ningún efecto.

Resumiendo, cuando comparamos el Modelo de evaluación de tipo 2 y el Modelo de tipo 3 vemos que una parte de lo que es considerado como ineficiencia del estudiante (STE_2), el Modelo tipo 3 lo descompone como efecto de variables contextuales ($STCE_3$) y el efecto neto del estudiante (STE_3). Similarmente, el efecto escuela del Modelo 2 (SCE_2) es descompuesto en orden a considerar el impacto del contexto debido a socioeconómicos factores ($SCCE_3$) y el efecto neto de la escuela (SCE_3). Esto significa que un potencial gap tecnológico (representado por $STCE_3$ y $SCCE_3$) aparece cuando el contexto tiene un significativo impacto sobre los puntajes de logro académico que el estudiante puede alcanzar. En el modelo 3 este gap puede o no ser significativo, mientras que en el Modelo tipo 2, por definición, este impacto no existe.

En resumen, la descomposición correspondiente al Modelo de tipo 3 es:

$$\begin{aligned} \text{Eficiencia Global} &= \text{Efecto Contextual sobre la Eficiencia} * \text{Efecto Neto Global} \\ &= \text{Efecto Contextual sobre la Eficiencia} * \text{Efecto Estudiante} * \text{Efecto Escuela} \end{aligned} \quad (2)$$

O:

$$OE = (y''_2 / y_c) = (y''_2 / y''_3) * (y'_2 / y'_3) * (y'_3 / y_c) * (y''_3 / y'_2) \quad (3)$$

3.1. La estimación de orden-m de los coeficientes de eficiencia frontera

Una importante decisión a tomar antes de comenzar a estimar niveles de ineficiencia y *benchmarks* frontera se relaciona a la especificación de la tecnología prevalente utilizada en el proceso educativo. Esta especificación no es trivial, y tiene directa implicación vis-à-vis con el nivel de eficiencia de la escuela. De esta manera, cuando asumimos una tecnología convexa, los modelos DEA operan como puntos virtuales, estableciendo combinaciones lineales entre las observaciones reales. De manera alternativa, una tecnología no convexa define observaciones reales como frontera de referencia. Como consecuencia, cada estudiante ineficiente será comparado con un real estudiante más eficiente y no uno virtual construido a partir de combinaciones lineales. Esta es la gran ventaja del proceso de evaluación FDH.

La literatura existente da cuenta de algunas limitaciones con respecto a los métodos de estimación con frontera no paramétrica: la dimensionalidad, la carencia de propiedades estadísticas –dado que por su naturaleza son deterministas, y el potencial impacto de *outliers*. La propuesta consiste en la identificación y eliminación de los casos extremos súper eficientes. Sin embargo, esto es controversial, dado que la eliminación de una unidad súper eficiente podría llevar a perder importante información si su nivel de eficiencia no es causada por un error. Su eliminación podría incrementar el valor de la eficiencia promedio, magnificar la mediocridad y reducir el potencial de ganancia de eficiencia que podría ser alcanzado.

Para hacer frente a las limitaciones anteriores, algunas propuestas han establecido propiedades estadísticas al estimador FDH (Kneip et al., 1998; Simar y Wilson, 2000), al igual que otros indicadores no paramétricos. De estos estudios podemos deducir que los modelos FDH experimentan problemas de dimensionalidad debido a su lenta tasa de convergencia. Todo lo anterior compromete seriamente los indicadores FDH. Para solucionar estos problemas y establecer indicadores más robustos una serie de enfoques han sido propuestos en la literatura (Wilson, 1993, 1995). Más recientemente, Simar (2003), introdujo el concepto de fronteras de orden-m, siendo una excelente herramienta para mitigar los problemas de dimensionalidad, la presencia de observaciones extremas y *outliers*; además de posibilitar la inferencia estadística mientras mantiene su naturaleza no paramétrica. Una breve descripción del enfoque se realiza a continuación.

Asumamos que contamos con información de los vectores de inputs y outputs $(x_{c1}, x_{c2}, \dots, x_{ci}, \dots, x_{cn})$ e $y_c = (y_{c1}, y_{c2}, \dots, y_{cj}, \dots, y_{cn})$, respectivamente) para cada estudiante en la muestra $(1, 2, \dots, C)$. Caracterizando los elementos del vector entero de actividad como $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_C)$ y el coeficiente de eficiencia como α_c^{FDH} , el coeficiente FDH orientado al output se obtiene resolviendo el siguiente programa lineal:

$$\begin{aligned}
& \max_{\{\alpha_c^{FDH}, \lambda_1, \lambda_2, \dots, \lambda_c\}} \alpha_c^{FDH}, \\
& \text{s.t.} \\
& \sum_{s=1}^C \lambda_s x_{s,i} - x_{c,i} \geq 0, \quad i = 1, \dots, I \\
& - \sum_{s=1}^C \lambda_s y_{s,j} + \alpha_c^{FDH} y_{c,j} \geq 0, \quad j = 1, \dots, J \\
& \sum_{s=1}^C \lambda_s = 1, \\
& \lambda_s \in \{0, 1\}, \quad s = 1, \dots, S
\end{aligned} \tag{4}$$

Para cada estudiante c encontrado ser FDH-ineficiente, el programa (4) identifica otro estudiante en la muestra que presenta un rendimiento superior (precisamente el estudiante tiene un coeficiente de $\lambda_s^* = 1$) y estima el incremento requerido de los outputs para alcanzar la frontera no convexa ($\alpha_c^{FDH} > 1$), siendo $(1 - \alpha_c^{FDH})$ el nivel de aumento proporcional requerido en el output. Para los estudiantes declarados FDH-eficientes, el programa (4) ofrece un vector de actividad $\lambda_c = 1$, y un coeficiente de eficiencia igual a la unidad ($\alpha_c^{FDH} = 1$).

Parte de los problemas relacionados con las estimaciones FDH (por ejemplo, la falta de robustez debido a la problemática de la dimensionalidad y al efecto de unidades súper-eficientes unidades) puede ser resuelto con recientes extensiones en el marco de la eficiencia no convexa. Por ejemplo, Cazals et al. (2002) y Simar (2003) introducen la estimación de orden m , siendo una excelente herramienta para mitigar los problemas de dimensionalidad, reducir el impacto de las observaciones extremas y, adicionalmente, para hacer posible la inferencia estadística, manteniendo su naturaleza no convexa y no paramétrico. Una breve descripción de la evaluación de la orden m se detalla en los siguientes párrafos.

Considere un entero positivo m fijo. Para un nivel dado de entrada $(x_{c,i})$ y de salida $(y_{c,j})$, la estimación define el valor esperado del máximo de m variables aleatorias $(y_{1,j}, \dots, y_{m,j})$, elaborado a partir de la distribución condicional de la matriz de salida Y observando la condición $y_{m,j} > y_{c,j}$.

Formalmente, el algoritmo propuesto para calcular el estimador de orden m -tiene cuatro pasos:

Para un nivel dado de $y_{c,j}$, extraer una muestra aleatoria de tamaño m con sustitución entre los $y_{m,j}$, tal que $y_{m,j} \geq y_{c,j}$.

1. Calculamos el programa (4) y estimamos $\tilde{\alpha}_c$.
2. Repetimos pasos 1 y 2 B veces y obtenemos B coeficientes de eficiencia $\tilde{\alpha}_c^b$ ($b=1,2,\dots,B$). Valores altos de B mejoran la calidad de la estimación, elegimos $B=2000$.
3. Calculamos la media de las estimaciones de las B muestras como:

$$\alpha_c^m = \frac{1}{B} \sum_{b=1}^B \tilde{\alpha}_c^b$$

A medida que aumenta m , el número de observaciones considerados en la estimación se aproxima a la unidades observadas que cumplen la condición $y_{m,j} > y_{c,j}$ y el estimador esperado de orden- m en cada una de las iteraciones b tiende al coeficiente FDH de eficiencia $\tilde{\alpha}_c^{FDH}$. De esta forma, m es un valor entero positivo arbitrario, pero siempre es conveniente observar las fluctuaciones de los coeficientes $\tilde{\alpha}_c^b$ en función del nivel de m . Para valores aceptables de m , normalmente $\tilde{\alpha}_c^b$ presentará valores mayores que la unidad (esto indica que estas unidades son ineficientes, debido a que pueden aumentar su nivel de output sin modificar su nivel de inputs. Cuando $\tilde{\alpha}_c^b < 1$, la unidad c puede ser etiquetada como súper-eficiente, producto que la frontera orden- m presenta unos niveles más bajos de los outputs que la unidad de análisis.

Como se ha mencionado, la estimación de orden m es una excelente herramienta para mitigar los problemas de dimensionalidad, la presencia de observaciones extremas y de valores atípicos. Sin embargo, esta evaluación puede ser de poca utilidad si parte de la ineficiencia encontrada depende de los factores contextuales y no introducimos estas variables en la evaluación. Para ajustar el proceso de evaluación a esta situación, como ya se ha discutido en el Modelo de tipo 3, definimos un proceso de evaluación frontera multinivel que podría servir para estimar el impacto de los factores socioeconómicos en el nivel de eficiencia de los estudiantes. Esta estimación multinivel es posible adaptando lo que Battese et al. (2004), Battese y Rao (2002), y O'Donnell et al. (2008) definieron como función de producción metafrontera. Para el modelo de tipo 3, antes mencionado, este proceso contiene los siguientes pasos:

(a) Clasificar los estudiantes (1, 2, ..., C), dependiendo de la escuela en la que están matriculados (1, 2, ..., D).

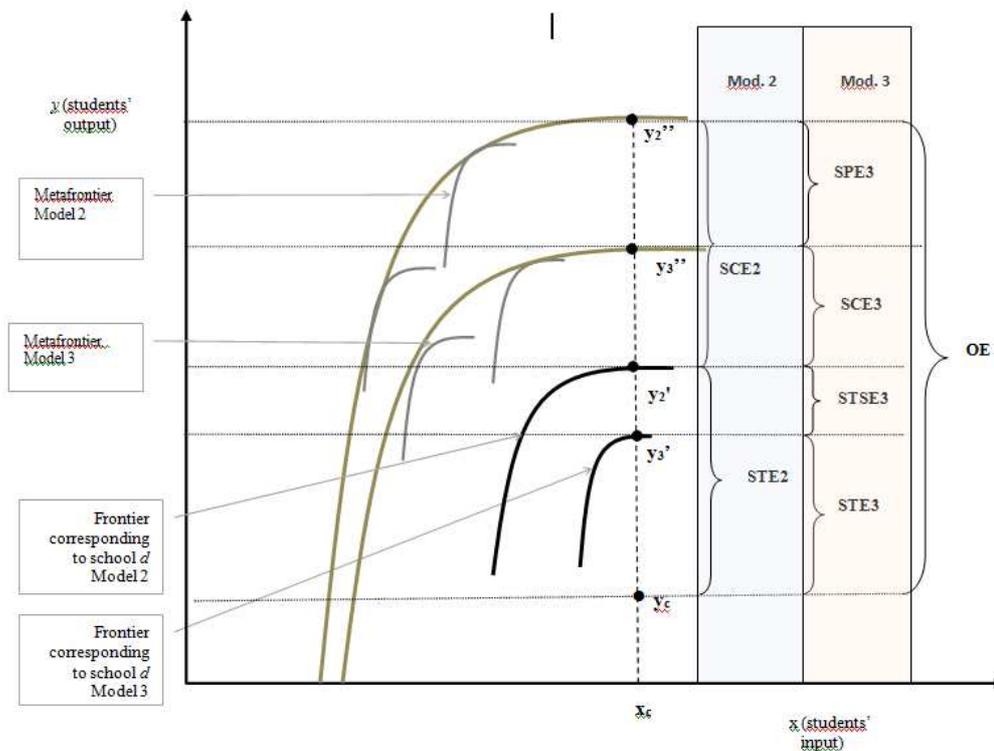
(b) Realizar los pasos 1 a 4 para calcular los coeficientes de eficiencia correspondientes a cada estudiante en la escuela respectiva en que está matriculado ($\tilde{\alpha}_c^m$) (significa, considerar el punto de la frontera de la escuela representado por y'_3 en la figura X en orden a estimar STE). A fin de facilitar la comparación cruzada de los resultados, independientemente del número de estudiantes clasificados en cada una de las escuelas, el mismo valor de m se le asignará en todas las estimaciones. Haciendo esto, los problemas de dimensionalidad y el impacto potencial de los valores extremos es neutralizado.

(c) Después de completar las fronteras condicionales, se añaden nuevas variables de entrada (el nivel socio-económico y cultural que corresponde a la familia del estudiante y el promedio de la misma variable para la escuela) y se aplican nuevamente los pasos 1 a 4 de la estimación de orden m de la muestra completa y se calculan los coeficientes de eficiencia con respecto a la metafrontera $\tilde{\alpha}_{c,1}^m$. Estos nuevos coeficientes proporcionan una evaluación de la eficiencia del alumno con respecto a la metafrontera global que tendrá en cuenta sólo a las escuelas que operan

sin mejores factores ambientales que la escuela donde se ubica el estudiante (lo cual es representada por el punto y''_3 en la figura 1).

Para cada estudiante encontrado ser FDH-ineficiente, el programa (4) identifica otro estudiante en la muestra con un rendimiento superior con el fin de estimar el aumento de la producción necesaria para alcanzar la frontera no convexa ($\alpha'_3 > 1$), siendo el aumento requerido proporcional en el nivel de salida. Para los estudiantes estimados FDH-eficientes, el programa (4) ofrece un coeficiente de eficiencia igual a la unidad ($\alpha'_3 = 1$).

Figura 1: Descomposición modelos de Valor agregado puro (Modelo 2) y Valor agregado Contextual (modelo 3)



4.- Muestra, modelos de evaluación, inputs y outputs

La base de datos utilizada en este estudio se construyó a partir de las bases de datos del Sistema para la evaluación de la calidad de la educación en Chile (SIMCE), el cual desde mediados de los años 1990 ha evaluado de manera censal el aprendizaje de los estudiantes de cuarto y octavo año básico, y segundo año medio, a través de pruebas estandarizadas. Las pruebas SIMCE evalúan el logro de los Objetivos Fundamentales y Contenidos Mínimos Obligatorios del Marco Curricular vigente en diferentes subsectores de aprendizaje, a través de una medición que se aplica a nivel nacional, una vez al año, a los estudiantes que cursan un determinado nivel educacional. A partir del año 2006, se evalúa todos los años a 4° Básico y se alternan 8° Básico y 2° Medio. Además de las pruebas asociadas al currículum, el SIMCE también recoge información sobre docentes, estudiantes y padres y apoderados a través de cuestionarios de contexto. Esta información se utiliza para contextualizar y analizar los resultados de los estudiantes en las pruebas SIMCE.

En particular, para este estudio se utilizó información de los resultados de los alumnos de cuarto año básico del año 2005 y de octavo año básico del año 2009. De igual forma, se extrajo adicionalmente la información sobre nivel socioeconómico y cultura de las familias de los alumnos a través del cuestionario aplicado a padres en la evaluación del año 2005.

De esta forma, de los 142.109 alumnos que dieron ambas pruebas y se mantuvieron en el mismo colegio, sólo 57.000 cumplieron adicionalmente con pertenecer a una escuela con más de 29 estudiantes que dieron la prueba y se contase con información completa de más del 60% de los alumnos que el año 2009 dieron la prueba SIMCE en ese colegio. De este universo se extrajo una muestra de 47.076 alumnos pertenecientes a 948 establecimientos, de los cuales 395 (41.67%) tienen dependencia Municipal (públicos), 460 (48.52%) son particulares con subvención estatal (particulares subvencionados) y 93 (9.81%) son particulares pagados.

La tabla 1 hace una descripción completa de la muestra tanto a nivel de alumnos como de centro educativo.

Tabla 2: Descripción de la muestra

| Type of School | Pupils | | Schools | |
|------------------------------|--------|--------|---------|--------|
| | Number | % | Number | % |
| Public | 18.021 | 38,28% | 395 | 41,67% |
| Privately - owned subsidized | 23.987 | 50,95% | 460 | 48,52% |
| Privately - owned fee paying | 5.068 | 10,77% | 93 | 9,81% |
| Total | 47.076 | 100% | 948 | 100% |

Los distintos modelos a evaluar se basan en la disponibilidad de información recogida a nivel de alumno (5 variables) y de centro (1 variable). Como variables de output se consideró los puntajes obtenidos por los alumnos de 8º básico en el año 2009 en las pruebas de matemáticas y lenguaje del SIMCE. Como variables de inputs a nivel de alumno se consideraron dos variables de logro

académico previo como son el puntaje obtenido por los mismos alumnos en las pruebas de matemáticas y lenguaje del SIMCE cuando cursaban 4º básico en el año 2005; y un índice de nivel socioeconómico y cultural de las familias construido a través de Componentes Principales utilizando las variables de nivel educacional del padre, nivel educacional de la madre e ingresos mensuales del hogar, obtenidas del cuestionario aplicado a las familias junto con la prueba SIMCE durante el año 2005. Igualmente se utilizó como proxie del nivel socioeconómico del centro el promedio del índice de nivel socioeconómico y cultural de las familias.

Tabla 3: Estadísticos descriptivos de las variables

| Level | Variable | Description | Minimun | Maximun | Mean | Std. Dev. | Percentil 25 | Percentil 75 |
|---------------|----------|--|----------|----------|----------|-----------|--------------|--------------|
| Student level | y1 | Mathematics Score, 8º grade, 2009 | 138,1600 | 401,7100 | 274,1052 | 51,4150 | 237,0400 | 311,1000 |
| | y2 | Language Score, 8º grade, 2009 | 97,2000 | 376,8700 | 263,9506 | 50,1096 | 229,3100 | 300,1200 |
| | x1 | Mathematics Score, 4º grade, 2005 | 90,8300 | 363,6200 | 267,0472 | 50,5391 | 233,4200 | 304,3200 |
| | x2 | Language Score, 4º grade, 2005 | 103,7400 | 364,7700 | 273,7696 | 48,4829 | 244,0550 | 308,0450 |
| School level | x3 | Socioeconomic and cultural level, student's family | -2,6437 | 3,3798 | 0,2923 | 1,0111 | -0,3841 | 0,8137 |
| | x4 | Socioeconomic and cultural level, school average | -1,1227 | 2,6654 | 0,2917 | 0,8372 | -0,2842 | 0,5766 |

Model type 0: Overall Effect: $(y_1, y_2)/2$

Model type 1: Contextual Status: y_1, y_2, x_3, x_4

Model type 2: Value added: y_1, y_2, x_1, x_2 .

Model type 3: Contextual value added: $y_1, y_2, x_1, x_2, x_3, x_4$

Siguiendo la clasificación señalada previamente, se configuraron 4 modelos de evaluación (uno de ellos, el modelo 0, correspondiente al promedio SIMCE), de los que se da cuenta en la tabla 2.

5.- Resultados

La tabla 4 presenta los resultados de los cuatro tipos de evaluaciones realizadas: tipo 0 (status sin variables contextuales), tipo 1 (status con variables contextuales), tipo 2 (valor agregado puro), tipo 3 (valor agregado contextual). Los resultados del tipo 0 corresponden a los resultados brutos de los estudiantes promedio de cada colegio, similares a los que se reportan año a año en Chile; los restantes tres tipos de modelos de evaluación corresponden a los resultados de orden-m obtenidos para cada uno de ellos.

De los resultados de tipo 0 podemos concluir que existen importantes diferencias de desempeño entre tipos de escuelas dependiendo de su dependencia. Los establecimientos particulares pagados muestran mejor desempeño que los particulares subvencionados, y estos a su vez, mejores desempeños que los establecimientos municipales. Sin embargo, cabe recordar que en este tipo de evaluación se asigna erróneamente todas las diferencias (overall effect) a un efecto escuela. Ello obedece a que este tipo de evaluación supone que todos los alumnos tienen similares y óptimas capacidades cognitivas de entrada y factores del entorno (nivel socioeconómico y cultural de las familias del estudiante).

Table 4: Results for the overall effect (OE) and its components, arithmetic means

| | Type of Effect | Type 0 (Status) | Type 1 (CS) | Type 2 (VA) | Type 3 (CVA) |
|------------------------------|-------------------|-----------------|-------------|-------------|--------------|
| Total | Overall Effect | 265,5373 | 1,1737 | 1,0545 | 1,0545 |
| | Student's Effect | | 1,1515 | 1,0283 | 1,0179 |
| | School Effect | | 1,0192 | 1,0255 | 0,9976 |
| | Contextual Effect | | | | 1,0385 |
| Public | Overall Effect | 246,2879 | 1,2335 | 1,1019 | 1,1019 |
| | Student's Effect | | 1,1678 | 1,0289 | 1,0177 |
| | School Effect | | 1,0563 | 1,0710 | 1,0294 |
| | Contextual Effect | | | | 1,0518 |
| Privately - owned subsidized | Overall Effect | 273,5235 | 1,1500 | 1,0347 | 1,0347 |
| | Student's Effect | | 1,1479 | 1,0274 | 1,0176 |
| | School Effect | | 1,0019 | 1,0071 | 0,9845 |
| | Contextual Effect | | | | 1,0329 |
| Privately - owned fee paying | Overall Effect | 315,7072 | 1,0508 | 0,9610 | 0,9610 |
| | Student's Effect | | 1,1021 | 1,0304 | 1,0202 |
| | School Effect | | 0,9534 | 0,9326 | 0,9325 |
| | Contextual Effect | | | | 1,0101 |

Sin embargo, un análisis sobre estos resultados dan cuenta de una correlación intraclass de 34%. Este resultado entrega una indicación de la dependencia de los puntajes de logro de los estudiantes sobre la escuela. Alrededor de 34% de la varianza final de logro de los estudiantes puede ser adscrita a la escuela. Esto significa que la mayor cantidad de varianza ocurre a nivel de estudiante. Situación que no se visualiza en este tipo de reporte de resultados.

Los resultados de orden –m para una evaluación de tipo 1 (contextual status), que descompone el efecto total entre efecto alumno y efecto escuela, y que considera los valores observables (y por tanto no óptimos) de factores no controlables a nivel e alumno y de escuela. Para el total de la muestra la ineficiencia global es de 1.1737; compuesta principalmente por el efecto estudiante que alcanza un valor de 1.1515, y en bastante menor medida por el efecto escuela que toma un valor de 1.0192. Por tanto, en promedio, la contribución del efecto estudiante es mucho mayor que el atribuible a la escuela. La descomposición de estos resultados por tipo de dependencia indica que en los tres casos el efecto estudiante tiene mucha mayor relevancia que el efecto escuela sobre la ineficiencia global. De igual manera se aprecia que persisten, aunque en menor medida, las diferencias de desempeño entre tipos de establecimientos.

Los resultados para la evaluación de tipo 2 (valor agregado puro), considera valores observados de logro académico de los estudiantes tanto al inicio como al final del segundo ciclo de enseñanza básica. Sin embargo, supone que todos los alumnos tienen una dotación óptima de factores del entorno. Para el total de la muestra, la ineficiencia global es de 1.0545, siendo el efecto estudiante (1.0283) sólo levemente superior al efecto escuela (1.0255). Sin embargo, esta situación cambia al realizar un análisis por tipo de dependencia. En efecto, si bien el efecto estudiante se mantiene relativamente estable (entre 1.0274 para el caso de los establecimientos públicos y 1.0304 para

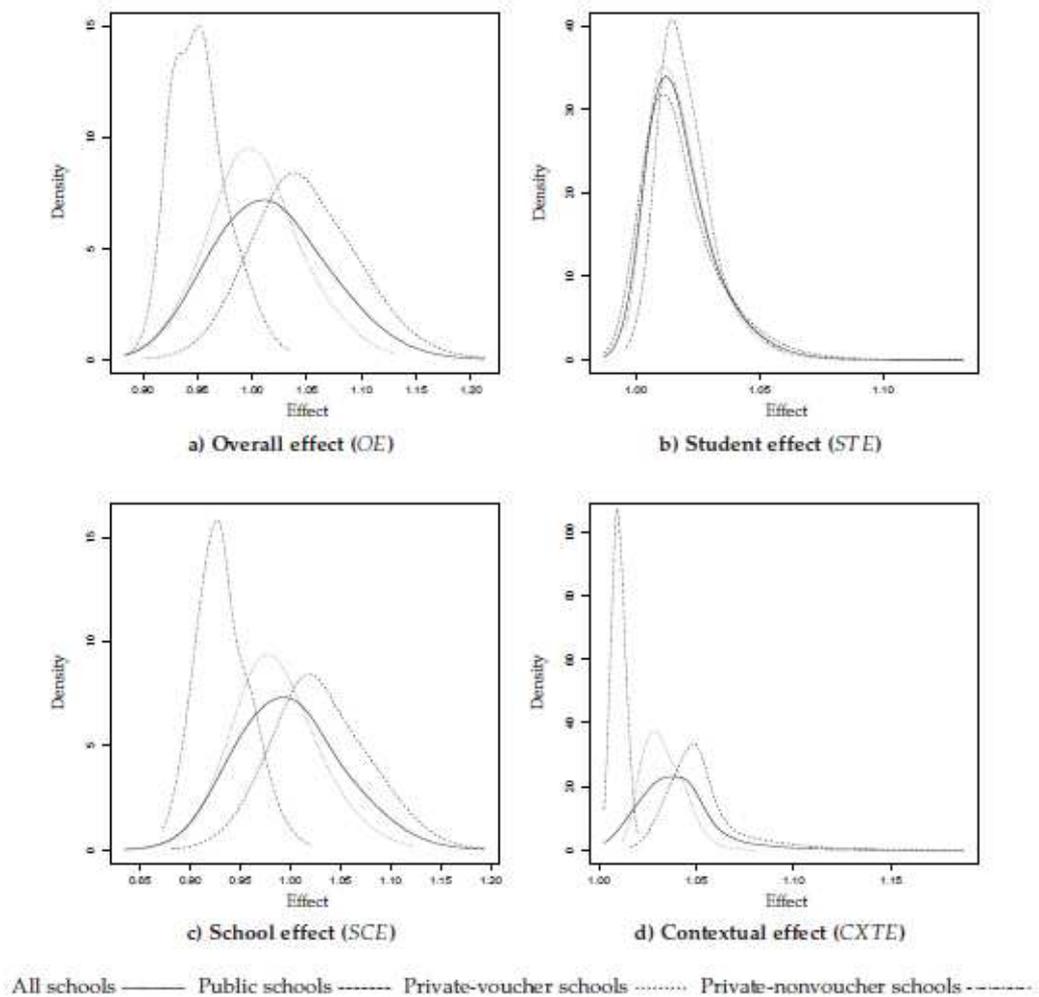
los establecimientos particulares pagados), el efecto escuela cambia drásticamente. En promedio, los establecimientos particulares pagados muestran ser eficientes (0.9326); los particulares pagados exhiben bajos niveles de ineficiencia. Sólo los establecimientos públicos muestran altos niveles de ineficiencia (1.0710).

Sin embargo, esta situación cambia al incorporar al modelo de valor agregado precedente variables de contexto a nivel de alumno y de escuela. Los resultados de la evaluación de tipo 3 (valor agregado contextual) dan cuenta de ello. Para el total de la muestra la ineficiencia global se mantiene inalterable, variando los efectos que la componen. En efecto, la mayor parte de esta ineficiencia global obedece a variables de contexto (1.0385), seguida de efecto de estudiante (1.0179), mientras que, en promedio, el efecto escuela resulta eficiente (0.9976).

Al descomponer por tipo de dependencia, los resultados son más aclaradores de la situación. Tal como podría esperarse, el efecto estudiante muestra no sólo igual media para todos los tipos de dependencia, sino también, como puede apreciarse en figura 2, similar distribución. Similar situación se aprecia con respecto a los efectos contextuales que afectan negativamente principalmente a los establecimientos públicos (1.0518), en menor medida a los establecimientos particulares subvencionados (1.0329) y muy poco a los establecimientos particulares pagados (1.101). Por último, los resultados del efecto escuela, muestran sólo promedios ineficientes para el caso de los establecimientos municipales. Tanto los establecimientos particulares subvencionados como los particulares pagados muestran, que en promedio, son eficientes.

Una visión más global de la situación analizando la función de distribución puede verse en la figura 2. En ella se aprecia una curva normal de distribución del efecto escuela con media cercana a 1.000, pero con diferentes curvas de distribución por tipo de dependencia. En efecto, casi la totalidad de los establecimientos particulares pagados son eficientes, al igual que más de la mitad de los establecimientos particulares subvencionados y sólo cerca de 1/3 de los establecimientos públicos.

Figure 2: Kernel density plots by type of school, model 3



Comparación entre diferentes tipos de modelos de evaluación

La primera metodología para comparar los diferentes tipos de modelos de evaluación fue estimar la correlación del efecto escuela entre los cuatro tipos de modelos de evaluación. La tabla 6 presenta los resultados de las correlaciones. Estos indicadores pueden dar una indicación de qué tanto las estimaciones de los distintos tipos de modelo de evaluación dan cuenta de similar constructo. Las altas correlaciones de las estimaciones del efecto escuela entre los distintos tipos de modelos de evaluación sugieren que subyace el mismo constructo para los distintos indicadores de efectos de escuela. La correlación entre los resultados de valor agregado contextual (de tipo 3) y las restantes estimaciones de efecto escuela muestran altas correlaciones que varían entre 0.8469 y 0.9445; incluso la correlación entre las estimaciones brutas de resultados Simce promedio (tipo 0) con las de valor agregado (tipos 2y 3) son muy altas (0.9039 y

0.8469, respectivamente). Estos resultados no difieren significativamente de los reportados en estudios que comparan distintos modelos de valor agregado utilizando metodologías tradicionales jerárquico lineales (Timmermans et al., 2011; Gorard, 2006; Gorard, 2008).

Tabla 5: Correlations for the overall effect (OE) across models, all schools

| | Type 0 (Status) | Type 1 (CS) | Type 2 (VA) |
|--------------|-----------------|-------------|-------------|
| Type 1 (CS) | 0,7931 | | |
| Type 2 (VA) | 0,9039 | 0,78 | |
| Type 3 (CVA) | 0,8469 | 0,8025 | 0,9445 |

Todas los valores $p < 0.001$ (two-tailed); $N = 948$

La segunda metodología para comparar las estimaciones de los distintos tipos de modelos de evaluación corresponde a tablas de contingencia. Las comparaciones entre las clasificaciones del efecto escuela por quintiles, entre distintos tipos de evaluación se muestran en la tabla 6.

Tabla 6: Tablas de contingencia entre tipos de modelos de evaluación

| | | Upper quintile, model 3: | | | | | (Number) |
|--------------------------|-----|--------------------------|------|------|------|------|----------|
| | | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 | |
| Upper quintile, model 0: | 0.2 | 0.69 | 0.26 | 0.04 | 0.01 | 0.00 | (190) |
| | 0.4 | 0.28 | 0.39 | 0.29 | 0.05 | 0.00 | (189) |
| | 0.6 | 0.02 | 0.25 | 0.40 | 0.28 | 0.05 | (190) |
| | 0.8 | 0.01 | 0.10 | 0.19 | 0.45 | 0.25 | (189) |
| | 0.9 | 0.00 | 0.01 | 0.08 | 0.22 | 0.69 | (190) |

a) Model 0 to 3

| | | Upper quintile, model 3: | | | | | (Number) |
|--------------------------|-----|--------------------------|------|------|------|------|----------|
| | | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 | |
| Upper quintile, model 2: | 0.2 | 0.83 | 0.17 | 0.01 | 0.00 | 0.00 | (190) |
| | 0.4 | 0.17 | 0.65 | 0.18 | 0.00 | 0.00 | (189) |
| | 0.6 | 0.00 | 0.16 | 0.63 | 0.20 | 0.01 | (190) |
| | 0.8 | 0.00 | 0.02 | 0.15 | 0.64 | 0.19 | (189) |
| | 1.0 | 0.00 | 0.00 | 0.03 | 0.16 | 0.81 | (190) |

b) Model 2 to 3

La comparación entre los tipos de evaluación 0 y 3 indica que del 100% de los establecimientos que se clasifican en el quintil superior (q_5) acorde al promedio Simce (tipo 0), el 69% se mantienen en ese quintil acorde a un análisis de valor agregado, el 22% se traslada al quintil 4, el 8% al quintil 3 y un 1% al quintil 2. Por otra parte, del 100% de los establecimientos que se clasifican en el quintil inferior (q_1 , el 20% con peor desempeño acorde al promedio Simce), el 69% se encontraría en esa misma clasificación acorde a una evaluación de valor agregado contextual, un 26% debiera ser clasificado como del quintil 2, un 4% como del quintil 3 y un 1% como del quintil 4. Como se aprecia en las tablas 7a y 7b, las mayores similitudes de ordenamiento se producen en los quintiles extremos, existiendo mayor variabilidad en los quintiles intermedios (q_2 , q_3 y q_4).

En general, un ordenamiento de escuelas con resultados brutos de logro académico (tipo 0) sólo es capaz de reproducir un 52,4% del ordenamiento resultante con un modelo de valor agregado contextual (tipo 3). Por su parte, un ordenamiento de escuelas acorde a su indicador de valor agregado puro (tipo 2) en quintiles es en el 71,2% casos, similar al que se obtendría en un ordenamiento con estimaciones de valor agregado contextual (tipo 3).

Figure 3: Transitions from model 0 to 3, contour plots

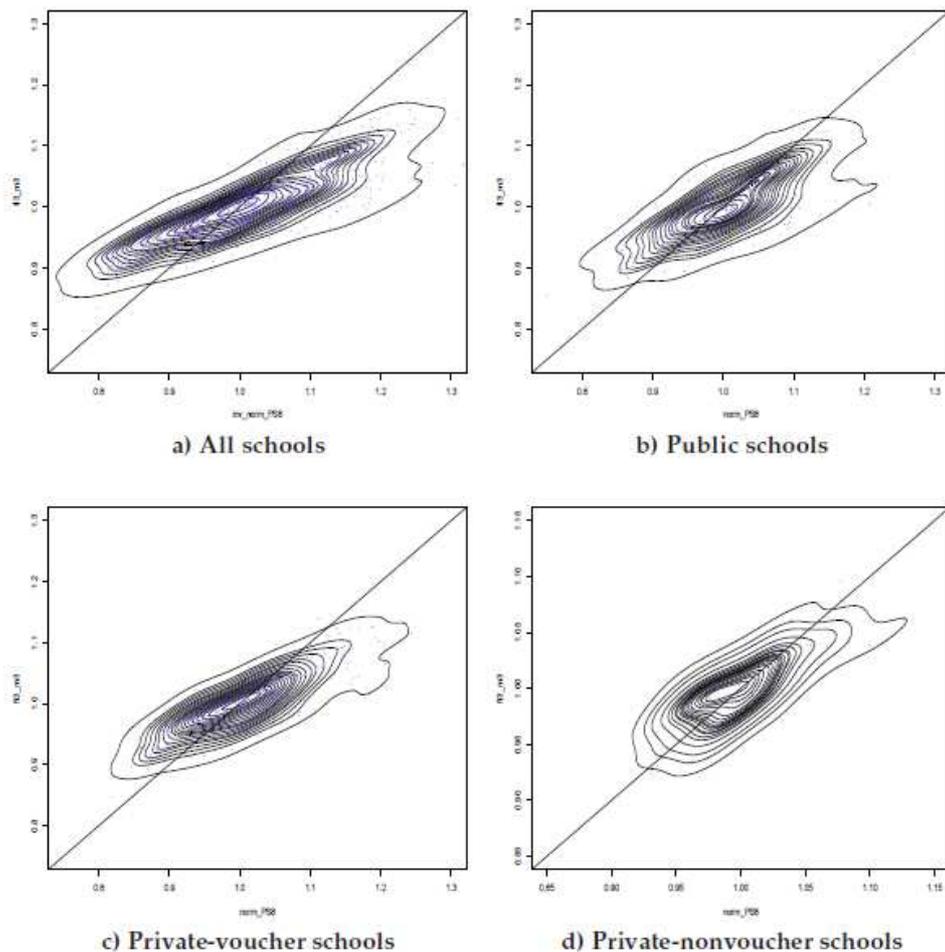
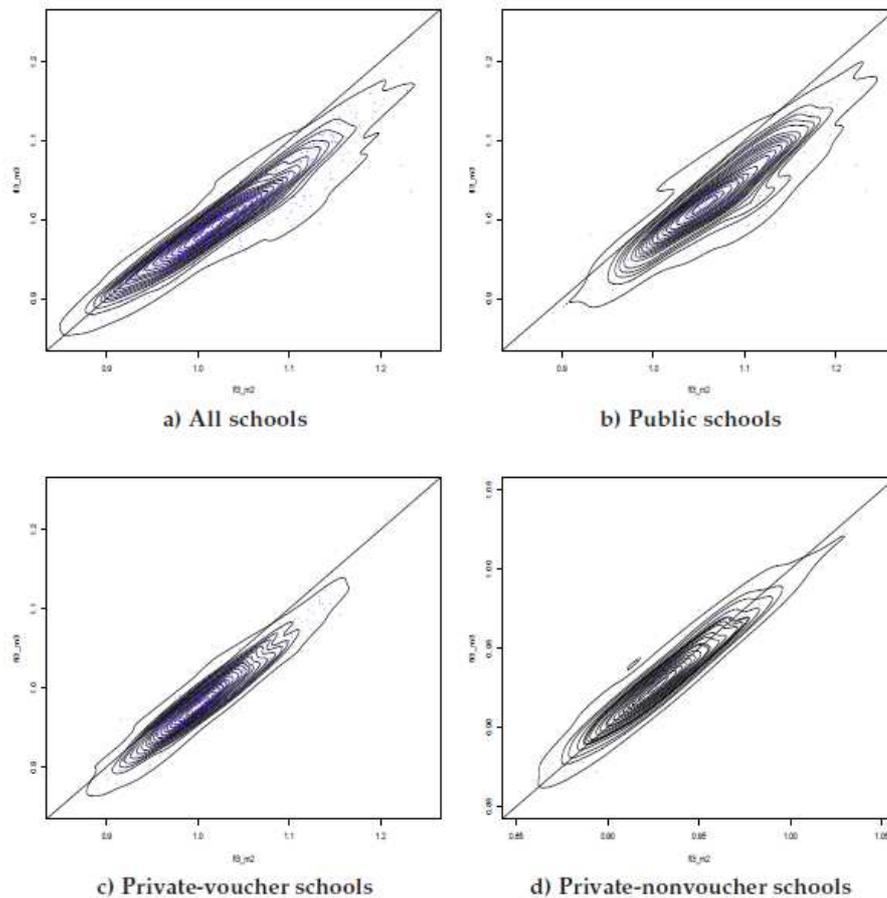


Figure 4: Transitions from model 2 to 3, contour plots



Comparación con metodología convencional para estimar Valor Agregado

Un aspecto importante a considerar en el análisis es comparar los resultados obtenidos mediante metodologías fronteras con las metodologías tradicionales para estimar Valor Agregado en investigación educativa. En general, la metodología más frecuente y aceptada son los Modelos de Regresión Multinivel (McCaffrey, 2004), también conocidos como Modelos Jerárquicos Lineales o modelos de regresión con efectos aleatorios (Goldstein, 2003; Raudenbush & Bryk, 2002). De manera análoga a la metodología que utilizamos en este trabajo, en los modelos de regresión multinivel se intenta explicar el rendimiento individual de los estudiantes a partir de una serie de predictores individuales (variables de Nivel 1) y de características contextuales de la escuela (variables de Nivel 2). Algebraicamente, un modelo de 2 niveles puede ser conceptualizado como un modelo de regresión lineal en el cual el intercepto β_0 y las k pendientes β_k asociadas a los x predictores de Nivel 1 son estimadas de manera tal que puedan variar para cada uno de los j grupos definidos en el Nivel 2 (en este caso escuelas), mediante la especificación de términos de error aleatorios u_{0j} y u_{kj} , respectivamente, que capturan la contribución específica de cada grupo a la explicación de la varianza total. De este modo, por ejemplo, un modelo multinivel con un predictor X en el Nivel 1 y un predictor W en el Nivel 2, es descrito como:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1j} + r_{ij}$$

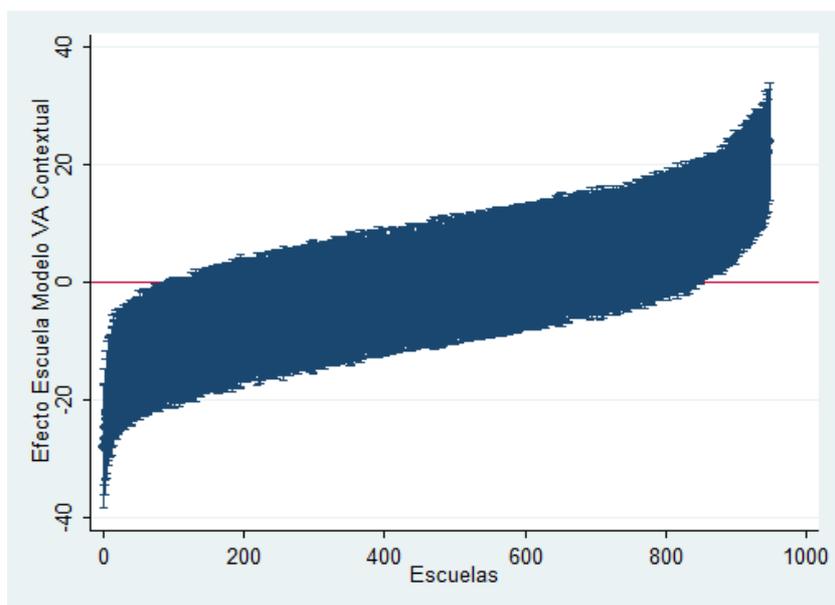
$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j}$$

En este sistema de ecuaciones, el predictor W , en el Nivel 2, explica la variación tanto del intercepto como de la pendiente del predictor X de Nivel 1. De este modo, el puntaje de un estudiante en la variable dependiente Y (i.e. SIMCE) puede descomponerse en tres elementos: la media global para toda la muestra (γ_{00}), la desviación del puntaje del estudiante respecto del promedio de su escuela (r_{ij}) y la desviación del promedio de la escuela respecto de la media global (u_{0j}). Este último término es definido como el *efecto aleatorio* para el intercepto, porque puede variar entre las escuelas. Cuando las variables incluidas en el modelo corresponden a características individuales de los estudiantes (e.g. X = nivel socioeconómico de la familia) o a características de la escuela (e.g. W =composición socioeconómica del estudiantado) que explican el rendimiento de los estudiantes, u_{0j} es un estimador de Status Contextual de eficacia de la escuela. Cuando las variables incluidas en el modelo son corresponden al desempeño pasado del estudiante (e.g. X = rendimiento en SIMCE en 4° básico), u_{0j} se puede interpretar como un estimador de Valor Agregado de la escuela. Finalmente, si las variables incluidas en el modelo consideran tanto el desempeño pasado como características individuales y/o de la escuela que podrían explicar el desempeño de los estudiantes, u_{0j} es utilizado como indicador de Valor Agregado Contextual de la escuela.

En línea con los resultados presentados en las secciones previas del artículo, en esta sección utilizamos modelos multinivel para estimar los tres modelos de efectos escolares previamente calculados con modelos frontera. Previamente se estimó un Modelo 0 (modelo sin predictores) que arrojó un coeficiente de correlación intraclase ICC=0.34. Este resultado indica que el 34% de la varianza total de los resultados en 8° básico puede explicarse únicamente por diferencias entre las Escuelas.

El Modelo 1, fue un modelo de Status Contextual, utilizando como predictores el NSE individual y el NSE promedio de la escuela. El Modelo 2 fue un modelo de VA simple en el que se utilizó como predictor el rendimiento SIMCE promedio de 4° Básico. Finalmente, el Modelo 3 fue un modelo de VA contextual, en que se incluyeron los tres predictores. Es importante indicar que el Modelo 3 arroja un ICC=0.12, lo que significa que, una vez corrigiendo el efecto de rendimiento previo y nivel socioeconómico, la varianza de los resultados explicada sólo por las diferencias entre escuelas, se reduce a un 12%. La siguiente figura muestra los residuales estandarizados para este modelo.



En la siguiente tabla se presentan las correlaciones entre los estimados de efectos escolares para cada metodología multinivel y las correlaciones entre las estimaciones para metodologías multinivel versus metodología frontera.

| | SIMCE8_mean | Modelo 1 | Modelo 2 | Modelo 3 |
|----------|-------------|----------|----------|----------|
| SIMCE8 | 1 | ,576** | ,789** | ,440** |
| Modelo 1 | ,576** | 1 | ,597** | ,768** |
| Modelo 2 | ,789** | ,597** | 1 | ,823** |
| Modelo 3 | ,440** | ,768** | ,823** | 1 |

Hay cuando menos dos resultados interesantes de observar en esta tabla. El primero es que Modelo 2 (VA simple) es el que mejor correlaciona con el promedio SIMCE no corregido. Aunque esto resulta en principio contra intuitivo, probablemente obedece al fuerte efecto que ejerce el NSE sobre los resultados. En efecto, dados que los estudiantes cursaron el menos 5 años consecutivos en el mismo establecimiento, es razonable suponer que el NSE ejerció un efecto inicial sobre sus resultados. Esta presunción es plausible si se considera que los Modelos que incorporan efectos contextuales (Modelos 1 y 3) tienen una correlación sustancialmente más baja con SIMCE. En ese sentido, un segundo resultado interesante de apreciar, es que la correlación del Modelo 3 con SIMCE es más baja que la obtenida para el mismo análisis en el caso de los modelos frontera.

En línea con lo anterior, en esta tabla se presenta la correlación entre los efectos escolares estimados mediante metodología multinivel versus metodología frontera.

| | fi3_m1 | SCE2 | SCE3 |
|-------------------------|---------|---------|---------|
| Efecto Escuela Modelo 1 | -,707** | -,620** | -,751** |
| Efecto Escuela Modelo 2 | -,657** | -,894** | -,897** |
| Efecto Escuela Modelo 3 | -,531** | -,645** | -,753** |

Puede observarse que, tal como ha sido reportado en estudios anteriores en que se comparan ambos tipos de metodología (De Witte et al., 2010), los efectos estimados como con modelos frontera y modelos multinivel tienden a converger, especialmente en el caso que utilicen las mismas variables predictoras, como queda en evidencia al analizar los valores en la diagonal de la tabla. No obstante, es preciso notar que correlaciones en torno a 0.75, como las que se observan para ambos tipos de modelo de VA contextual, representan sólo un 56% de varianza común entre ambas estimaciones de efecto, lo que sugiere que si se utilizaran ambas para, por ejemplo, hacer un ranking de escuelas, los resultados no serían del todo concordantes.

6. Conclusiones

Este artículo contribuye en dos aspectos: metodológico y empírico. La contribución metodológica consiste en la evaluación del desempeño educativo considerando una descomposición multinivel de valor agregado contextual con modelos de frontera no paramétrica robusta. A diferencia de estudios previos que consideran enfoques de regresión multinivel para medir la contribución neta del estudiante y de la escuela en el logro académico bruta del estudiante, nosotros utilizamos técnicas de frontera no paramétrica. De manera similar a recientes pero escasas contribuciones - como De Witte et al. (2010) y Thieme et al (2012)- utilizamos técnicas de orden-m para aliviar los problemas de dimensionalidad y la influencia de observaciones extremas, obteniendo resultados estadísticamente robustos. A diferencia de las propuestas anteriores de Silva Portela y Thanassoulis (2001) y De Witte et al. (2012), quienes consideran exclusivamente variables a nivel de estudiantes, nosotros consideramos, en línea con los estudios de economía de la educación, variables contextuales a nivel de alumno y de escuela.

La literatura relevante de valor agregado y de análisis multinivel en educación y los resultados obtenidos en esta aplicación revelan lo importante y necesario que resulta incluir variables contextuales en los distintos niveles bajo análisis, principalmente por dos razones. Primero, si variables contextuales a nivel de escuela no son consideradas estaríamos sub-evaluando el desempeño de las escuelas que atienden a estudiantes con desfavorables condiciones económicas y asignando responsabilidades a los gestores de estos centros por variables que no están bajo su incidencia, y peor aún, creando incentivos perversos para seleccionar estudiantes por sus características socioeconómicas. En segundo lugar, la literatura y en particular nuestro análisis

señala que las diferencias de desempeño entre los diferentes tipos de instituciones (particulares pagadas, particulares subvencionados y públicos) disminuyen al considerar estas variables.

Desde un punto de vista empírico, los resultados muestran que las grandes diferencias que se exhiben con resultados brutos de logro académico entre tipo de instituciones, disminuyen drásticamente al incorporar variables que dan cuenta de las condiciones de entrada de los estudiantes y de variables contextuales. De esta manera, una gran parte de la varianza explicada por la escuela, particularmente alta en Chile producto de su alto nivel de desagregación, desaparece al incorporar estas variables.

Referencias

- Aitkin, M. and Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society. Series A (General)*, 149(1):1–43.
- Ballou, D., Sanders, W., and Wright, P. (2004). Controlling for student background in value added assessment of teachers. *Journal of Educational and Behavioral Statistics*, 29(1):37–65.
- Battauz, M., Bellio, R., and Gori, E. (2011). Covariate measurement error adjustment for multilevel models with application to educational data. *Journal of Educational and Behavioral Statistics*, 36(3):283–306.
- Battese, G. E. and Rao, D. (2002). Technology gap, efficiency, and a stochastic metafrontier function. *International Journal of Business*, 1(2):87–93.
- Battese, G. E., Rao, D., and O'Donnell, C. J. (2004). A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. *Journal of Productivity Analysis*, 21(1):91–103.
- Bessent, A., Bessent, W., Kennington, J., and Reagan, B. (1982). An application of mathematical programming to assess productivity in the Houston independent school district. *Management Science*, 28(12):1355–1367.
- Bifulco, R. and Bretschneider, S. (2001). Estimating school efficiency: A comparison of methods using simulated data. *Economics of Education Review*, 20(5):417–429.
- Blanco, E. (2010). Eficacia escolar y desigualdad: aportes para la política educativa. *Perfiles Latinoamericanos*, 17(34).
- Bryk, A. S. and Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage, Newbury Park, CA.
- Carlson, D. (2006). "Focusing State Educational Accountability Systems: 4 Methods of Judging Quality and Progress". Working paper Center for Assessment.
- Carnoy, M. and Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, 24(4):305–331.

- Cazals, C., Florens, J.-P., and Simar, L. (2002). Nonparametric frontier estimation: a robust approach. *Journal of Econometrics*, 106:1–25.
- Cervini, R. A. (2009). Class, school, municipal, and state effects on mathematics achievement in Argentina: a multilevel analysis. *School Effectiveness and School Improvement*, 20(3):319–340.
- De Witte, K., Thanassoulis, E., Simpson, G., Battisti, G., and Charlesworth-May, A. (2010). Assessing pupil and school performance by non-parametric and parametric techniques. *Journal of the Operational Research Society*, 61(8):1224–1237.
- Denvir, B. and Brown, M. (1986). Understanding of number concepts in low attaining 7–9 year olds: Part i. development of descriptive framework and diagnostic instrument. *Educational Studies in Mathematics*, 17(1):15–36.
- Drury, D. and Doran, H. C. (2003). The value of value-added analysis. National School Boards Association Policy Research Brief, 3(1):1–4.
- Ercikan, K. (2006). Examining guidelines for developing accurate proficiency level scores. *Canadian Journal of Education*, 29(3):823–838.
- Goldschmidt, P. y Choi, K. (2007) The Practical Benefits of Growth Models for Accountability and the Limitations Under NCLB. Página de la Universidad de UCLA <http://www.cse.ucla.edu/products/policy.html>
- Goldstein, H. (1995). *Multilevel Statistical Models*. Wiley, London.
- Goldstein, H. (2003). *Statistical analysis of multilevel data* (3rd ed.). London: Edward Arnold.
- Goldstein, H. and Thomas, S. (1996). Using examination results as indicators of school and college performance. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 159(1):149–163.
- Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., Pan, H., Nuttall, D., and Thomas, S. (1993). A multilevel analysis of school examination results. *Oxford Review of Education*, 19(4):425–433.
- Gray, J., Jesson, D., Goldstein, H., Hedger, K., and Rasbash, J. (1995). A multi-level analysis of school improvement: Changes in schools' performance over time. *School Effectiveness and School Improvement*, 6(2):97–114.
- Hanushek, E. A. and Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, 24(2):297–327.
- Hox, J. J. (2002). *Multilevel Analysis: Techniques and Applications*. Lawrence Erlbaum Associates, Mahwah, New Jersey.
- Kneip, A., Park, B. U., and Simar, L. (1998). A note on the convergence of nonparametric DEA estimators for production efficiency scores. *Econometric Theory*, 14:783–793. 20

- Levacic, R. y Woods P. (2002). Raising School Performance in the League Tables (Part 1): disentangling the effects of social disadvantage. *British Educational Research Journal*, Vol. 28, No. 2
- Loader, C. R. (1996). Local likelihood density estimation. *The Annals of Statistics*, 24(4):1602–1618.
- Mancebón, M. J. y C. Mar Molinero (2000): Performance in primary schools, *Journal of the Operational Research Society*, 51, 7, 843–854.
- McCaffrey, D. F., J. R. Lockwood, D. M. Koretz and L. S. Hamilton. (2003). Evaluating Value-Added Models for Teacher Accountability. Santa Monica, CA: The RAND Corporation
- McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., and Hamilton, L. (2003). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, 29(1):67–101.
- Meyer, R. H. (1997). Value-added indicators of school performance: A primer. *Economics of Education Review*, 16(3):283–301. Education and Work, and Efficiency in Education: Essays in Memory of Charles Scott Benson.
- Mizala, A., Romaguera, P., and Farren, D. (2002). The technical efficiency of schools in Chile. *Applied Economics*, 34(12):1533–1552.
- Mortimore, P., Sammons, P., and Thomas, S. (1994). School effectiveness and value added measures. *Assessment in Education*, 1(3):315–332.
- O'Donnell, C. J., Prasada Rao, D. S., and Battese, G. E. (2008). Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. *Empirical Economics*, 34:231–255.
- Ouellette, P. and Vierstraete, V. (2005). An evaluation of the efficiency of Québec's school boards using the Data Envelopment Analysis method. *Applied Economics*, 37(14):1643–1653.
- Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods. Thousand Oaks, CA: Sage.
- Raudenbush, S.W. (2004). Schooling, Statistics, and Poverty: Can We Measure School Improvement? Educational Testing Service, Princeton, NJ.
- Rouse, C. E., Hannaway, J., Goldhaber, D., and Figlio, D. (2007). Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure. Working Paper 13681, National Bureau of Economic Research.
- Ruggiero, J., Duncombe, W., and Miner, J. (1995). On the measurement and causes of technical inefficiency in local public services: With an application to public education. *Journal of Public Administration Research and Theory*, 5(4):403–428.
- Sammons, P. (1995). Gender, ethnic and socio-economic differences in attainment and progress: a longitudinal analysis of student achievement over 9 years. *British Educational Research Journal*, 21(4):465–485.

- Silva Portela, M. C. A. and Thanassoulis, E. (2001). Decomposing school and school-type efficiency. *European Journal of Operational Research*, 132(2):357–373.
- Simar, L. (2003). Detecting outliers in frontier models: A simple approach. *Journal of Productivity Analysis*, 20(3):391–424.
- Simar, L. and Wilson, P. W. (2000). Statistical inference in nonparametric frontier models: The state of the art. *Journal of Productivity Analysis*, 13(1):49–78.
- Simar, L. and Wilson, P. W. (2008). Statistical inference in nonparametric frontier models: Recent developments and perspectives. In Fried, H., Lovell, C. A. K., and Schmidt, S. S., editors, *The Measurement of Productive Efficiency*, chapter 4, pages 421–521. Oxford University Press, Oxford, 2nd edition.
- Stevens, j. (2005). “The study of school effectiveness as a problem in research design”. En R. Lissitz (ed.), *Value-Added models in education: Theory and applications* (pp. 166-208). Maple Grove, MN: JAM Press.
- Teddlie y Reynolds (2000). *The International Handbook of School Effectiveness Research*. Falmer Press.
- Tekwe, C. D., Carter, R. L., Ma, C. X., Algina, J., Lucas, M. E., Roth, J., Ariet, M., Fisher, T., and Resnick, M. B. (2004). An empirical comparison of statistical models for value-added assessment of school performance. *Journal of Educational and Behavioral Statistics*, 29(1):11–36.
- Thanassoulis, E. and Silva Portela, M. C. A. (2002). School Outcomes: Sharing the Responsibility Between Pupil and School. *Education Economics*, 10(2):183–207.
- Thieme, C., Prior, D., and Tortosa-Ausina, E. (2012). A multilevel decomposition of school performance using robust nonparametric frontier techniques. *Economics of Education Review*. Forthcoming.
- Wilson, D. (2004). Which Ranking? The Impact of a Value-Added Measure of Secondary School Performance. *Public Money & Management*. January.
- Wilson, D. y Piebalga, A. (2008). Performance Measures, Ranking and Parental Choice: An Analysis of the English School League Tables. *International Public Management Journal*, 11:3,344 — 366.