UNIVERSIDAD TORCUATO DI TELLA

DEPARTAMENTO DE ECONOMIA

MAESTRIA EN ECONOMETRIA

DETERMINANTES DEL DESEMPEÑO EN MATEMATICAS PARA COLOMBIA. UN ENFOQUE ALTERNATIVO MEDIANTE MODELOS LINEALES Y LOGISTICOS MULTINIVEL

ANDRES EDUARDO RANGEL JIMENEZ

TUTOR: MARTIN GONZALEZ ROZADA

AGOSTO, 2013

Resumen

Evitando realizar un estudio más con modelos multinivel de gran complejidad aprovechando la riqueza de la información de PISA, se estimaron modelos que permitieran comprobar la hipótesis central sobre la diferenciación al interior de las escuelas públicas y privadas con base en el impacto de la orientación vocacional y el efecto “par”.

Desde el punto de vista metodológico el estudio realiza la defensa del modelo mediante test de correcta especificación, entre ellos, test sobre el término de error de nivel 1 y 2, test sobre la existencia o no de variación aleatoria en los coeficientes de nivel uno, exogeneidad de nivel 2 y sobre el número de niveles a incluir. Encontrando endogeneidad de nivel 2 se procede a estimar el modelo con la aproximación de Mundlak. Se concluye que el efecto par se diferencia por tipo de escuela y que el tipo de orientación vocacional afecta el desempeño en la prueba PISA según sea el tipo de colegio.

Palabras Clave: Modelos Multinivel, Endogeneidad

Contenidos

1. Introducción

Los estudios disponibles sobre los sistemas educativos latinoamericanos muestran que estos tienden a reproducir las desigualdades preexistentes, perpetuando las históricas desigualdades rurales-urbanas y diferencias en la calidad de la educación según niveles socioeconómicos.

La universalización produjo la llegada de la escuela secundaria a nuevos sectores sociales dentro del marco de los objetivos del milenio, el cual propone un universalismo básico de la educación. Este concepto apunta a asegurar un conjunto de servicios homogéneos con estándares de calidad para todos los ciudadanos, que en el caso educación secundaria se alcanzó mediante aumento de la cobertura de la educación pública.

En este punto la concepción de movilidad social como objetivo de la educación se ha vuelto un poco difusa, en cierto grado debido a que ha sido mediatizada por la dinámica del mercado de trabajo. Diversos estudios muestran que al final de la secundaria el proceso de inserción laboral es muy bajo, pues no hay puestos de trabajo para todos los egresados ni mucho menos empleos de calidad (Jacinto, 2009).

Durante los años noventa el consenso generalizado en América Latina giraba entonces alrededor de si el mercado laboral requería enfatizar en una educación general de calidad o si por el contrario requería de un conjunto de saberes de trabajo. Esta iniciativa y posterior adopción parece haber estado influenciada por la complejidad de los mercados de trabajo latinoamericanos, cuya segmentación permea a la educación llevándola a plantearse el interrogante: cuál debe ser el papel de la escuela en la orientación de los jóvenes egresados que enfrentan a un mercado laboral duro e incierto? Es conveniente proponer generalizadamente que la escuela secundaria no forme para nada en específico?

En algunos países las iniciativas para introducir saberes del trabajo en la escuela secundaria apenas comienzan, mientras otros como México, Colombia y Brasil han introducido una amplia y explicita gama de saberes del trabajo. La formación para el trabajo dentro del aula de clase se denomina en Colombia orientación vocacional técnica, diferenciándola de la clásica orientación académica que se enfoca en la preparación para la universidad[1]. La existencia de orientaciones vocacionales en Colombia constituye el instrumento mediante el cual la escuela cumpliría un rol diferenciador como formadora para el trabajo y para migrar a la educación universitaria.

Para ambos tipos de orientación las nuevas estrategias se enfocan en mejorar la formación que reciben los estudiantes y así disminuir los problemas de aprendizaje que impiden avanzar en el proceso escolar. Para citar una de las más importantes, se encuentra la introducción de tutores para el acompañamiento escolar, siendo esta ultima la medida más implementada en los últimos años en el país[2].Dado que el objetivo del trabajo es identificar los determinantes del desempeño en matemáticas, es pertinente identificar mediante modelos logísticos multinivel, los determinantes de la elección de estas tutorías por parte de los estudiantes, siendo de especial interés el probar la importancia del efecto contextual o "efecto par" dentro de dicha elección.

El estudio se distancia de la discusión de la diferencia de la calidad entre la educación privada y la pública, puesto que esta regularidad empírica ya está sobre diagnosticada con una ventaja clara a favor de la primera. Es pertinente cuestionar hasta donde la simple comparación de promedios en los puntajes es válida, toda vez que se tratan de estructuras educativas

[1]La escuela secundaria en su origen selectivo y con un currículo académico debió dar respuesta a una sociedad desigual respecto a las oportunidades de los jóvenes de insertarse en el mercado laboral.

[2]Complementan acciones de orientación a los estudiantes con miras al proceso universitario, aquellas centradas en la formación para el trabajo.

totalmente distintas en cuanto a la interacción social dada al interior de cada una de ellas. Se postula un impacto diferencial de los “efectos par” en el desempeño académico del estudiante, según sea el tipo de establecimiento educativo, público o privado.

En qué grado la función de formación para el trabajo en la escuela secundaria superior (grados décimo y undécimo) es compatible con el brindar saberes y competencias generales y transversales en las que todos acuerdan? Pues bien, este interrogante ha estado por fuera de los estudios de calidad de la educación secundaria en Colombia ignorando las brechas de calidad que puedan existir al interior de escuelas privadas y públicas por tipo de orientación vocacional[3].

En este sentido la investigación aborda si el desarrollo de habilidades laborales dentro de la escuela interfiere con una adecuada formación académica, más allá de que los saberes del trabajo sean o no integrados con la cotidianidad de las clases dentro del aula[4].

[3]Hasta el momento se había dado por sentado que la calidad educativa está del lado de las escuelas privadas en el tradicional análisis frente a las escuelas públicas, sin entrar a mirar de manera diferenciada el desempeño académico en las primeras según orientación vocacional.

[4]Es importante anotar que este estudio no pretende analizar en qué medida la iniciativa de saberes del trabajo dentro del aula de clase, contribuye al enorme desafío que tienen los jóvenes colombianos en su proceso de inserción laboral.

1.1 Justificación de la investigación

Con el fin de evitar el riesgo de realizar un estudio más con modelos jerárquicos lineales y la tentación de construir y estimar numerosos modelos aprovechando la riqueza de la información contenida en el Programa Internacional para la Evaluación de Estudiantes (PISA), se escogieron modelos que permitieran comprobar la hipótesis central sobre la diferenciación al interior de las escuelas públicas y privadas con base en dos elementos, el impacto de la orientación vocacional y el "efecto par" o efecto contextual.

Desde el punto de vista metodológico el estudio realiza un aporte que sin llegar a ser pretencioso al nivel de llamarlo novedoso, hace operacional la construcción y "defensa del modelo" al realizar los test de correcta especificación del modelo, entre ellos los test sobre el término de error de nivel 1 y nivel 2, homocedasticidad, normalidad y exogeneidad de nivel 2. Adicionalmente se presentan los test sobre el número de niveles a incluir y sobre la existencia de variación aleatoria o no aleatoria en los coeficientes de primer nivel. Por lo general los artículos que utilizan este tipo de modelos dejan por fuera las pruebas sobre correcta especificación, limitándose a la presentación de resultados.

Al respecto, Ferron, Hogarty, Dedrick, Hess, Niles & Kromrey (2008) señalan que estas cuestiones debiendo ser abordadas en este tipo de análisis regularmente son ignoradas en la presentación de estos artículos[5]. En general los resultados de diversos estudios indican que no existe suficiente información que justifique la elección del modelo estimado y mucho menos que validen en conjunto los supuestos del modelo. La presente investigación intenta suplir

[5]Ferrón analiza 98 artículos proveniente de 19 revistas especializadas.

metodológicamente ese faltante mediante una construcción que permita en cada etapa justificar la elección del modelo final junto con sus resultados.

Adicionalmente y dado que la investigación se centra en estudiar los determinantes del desempeño académico en matemáticas, se cierra la investigación con un modelo logit multinivel que estime la probabilidad de que un estudiante dentro de una estructura educativa, tome la decisión de tomar clases o tutorías adicionales con el fin suplir las deficiencias en matemáticas. Se está interesado en analizar si el efecto contextual ejerce gran influencia en la elección de este tipo de ayudas por parte de los estudiantes. Si realmente existe una estructura jerárquica dentro de la elección del estudiante, parecería natural incorporarla en el modelo mediante la estimación modelos logísticos multinivel.

2. Dos aspectos que diferencian la dinámica interna entre escuelas públicas y privadas

2.1 El "Efecto par" en educación secundaria

El supuesto del cual se parte para incluir en el análisis este efecto consiste en que los estudiantes no solo aprenden de sus profesores sino de la clase en general, esto es, de sus compañeros. El "efecto par" se presenta cuando el comportamiento de una persona es afectado por su interacción con sus compañeros de clase; la manera como se agrupen los estudiantes afectará el aprendizaje dados unos recursos. Si los "estudiantes débiles" ganan más de la interacción con los "estudiantes fuertes" en comparación a la perdida de los estudiantes fuertes con la cercanía a "estudiantes débiles", luego el aprendizaje podría incrementarse reduciendo la segregación, esto es aumentando la varianza de habilidades dentro del salón de clases. La anterior situación apoyaría una política pública de asignación aleatoria de estudiantes a colegios bajo el esquema de "vouchers".

Siguiendo a Winston y Zimmerman (2004) se toma por simplicidad la interacción entre dos estudiantes:

$$\mathrm{B_1 = f(B_2, C_2, X)} \qquad (2.1.1)$$

Donde B_1 es el comportamiento del estudiante 1, C_2 es la característica del estudiante 2 y X es un vector de otras características relevantes para el comportamiento del estudiante 1. El efecto par existe si las derivadas parciales de la ecuación (2.1.1) son diferentes de cero y asimétricas o no lineales, esto es las derivadas parciales cambian dependiendo de los diferentes niveles de B y C.

La pregunta que surge es como identificar el efecto par; pues bien dado que individuos que pertenece a un mismo grupo tienden a comportarse de manera similar, siguiendo a Manski (2000) se plantean dos hipótesis que explicarían las interacciones sociales del individuo al interior de un grupo.

- Efectos endógenos. En este primer caso, la probabilidad de que un individuo se comporte en la misma manera se incrementa con la presencia del comportamiento del grupo.
- Efectos contextuales. La probabilidad de que un individuo se comporte de una u otra manera dependerá de la distribución de variables exógenas referidas a los antecedentes del grupo. En el presente contexto, el desempeño del estudiante dependerá de la composición socioeconómica dentro del grupo.

Infortunadamente PISA no contiene información a nivel de las clases, luego el "par del grupo" está definido como estudiantes que se encuentran matriculados en el mismo colegio y grado. Esto es así pues en la mayoría de países no es común agrupar estudiantes a través de las clases, de manera que la composición de estudiantes dentro de un grado en una escuela particular no debería resultar una buena proxy de la composición de los salones de clase (Scheeweis & Winter-Ebmer, 2005).

Aunque no constituye una solución para la medición del efecto par de la clase, si constituye una mejora el diferenciar el efecto par entre escuelas privadas y públicas. Se postula entonces la existencia de un "efecto par" para ambos tipos de escuela, con una mayor magnitud en las escuelas privadas respecto a las escuelas públicas.

Estimar el efecto par no obstante es difícil pues se tiene que tratar con el problema del sesgo de selección. Se espera que la regla de selección este distorsionada producto no del diseño muestral sino más bien de una decisión económica de autoselección[6]. En un escenario no experimental es difícil identificar si se trata de un “efecto par” o simplemente se está observando estudiantes similares comportándose similarmente (Winston & Zimmerman, 2004).

PISA no aleatoriza el “efecto par”, luego el evitar el problema del sesgo de selección esta fuera del alcance de los objetivos planteados[7]. La auto-selección en Colombia como en la mayoría de países se produce a través de la segregación de los estudiantes en diferentes tipos de escuela, en este caso de acuerdo a su financiación, en públicas y privadas. Respecto a estudiantes que asisten al mismo tipo de escuela, puede asumirse que sus padres comparten características, unas observables, otras no. Con el fin de disminuir el problema de auto-selección, una estrategia a seguir es la de incrementar información sobre los antecedentes familiares de los estudiantes y así reducir el sesgo en las estimaciones por variables omitidas[8].

Incluso una asignación aleatoria del efecto par mediante sistema de “vouchers” aún dista del escenario ideal, por lo menos en lo que respecta a los cuasi experimentos hechos en Colombia. Al respecto Bettinger, Kremer y Saavedra (2010) utilizan el Programa de Ampliación de Cobertura en Educación Secundaria (PACES, 2004) para acercarse al escenario ideal. Este programa que inicia en 1998 financio alrededor 125000 becas para estudiantes de bajos ingresos. Con estas becas el beneficiario podía acceder a cualquier escuela privada que aceptara la beca, no obstante muchos colegios sobre todo los de elite las rechazaron. Las escuelas que participaron en su mayoría tenían estudiantes con bajos ingresos, una ratio alta de profesor-estudiante por lo que los

[6] En presencia de mecanismos de autoselección, los efectos par quedan sobreestimados.
[7] Asimismo el presente estudio no pretende identificar si el efecto par es simétrico o asimétrico, esto es si afecta o no por igual a los estudiantes en todos los niveles respecto a sus antecedentes socioeconómicos, lo cual puede alcanzarse mediante la metodología de regresión por cuantiles.
[8] El sesgo en las estimaciones puede reducirse sustancialmente si se incluyen variables que afectan ambos, el desempeño académico y la formación del efecto par.

precios de su matrícula era considerablemente más bajas que aquellas que no participaron, manteniendo parte del sesgo de autoselección que inicialmente el cuasi experimento quería aislar.

Las hipótesis planteadas por Manski (2000) que explicarían las interacciones sociales del individuo al interior de un grupo son los efectos endógenos y efectos contextuales. Para modelar el primer efecto se utilizara la aproximación de Mundlak de datos de panel para modelos multinivel, entretanto para el segundo, se propone como variable contextual la media del índice socioeconómico y cultural construido por PISA. La adaptación de la solución Mundlak de datos de panel para tratar el problema de endogeneidad de nivel 2 en modelos multinivel, en si misma justifica la inclusión del efecto contextual o "efecto par", cuyo indicador es la composición socioeconómica del grupo[9]. Un segundo argumento que apoya la hipótesis de realizar un efecto diferenciado por tipo de escuela lo constituye la hipótesis de que el efecto par opera de manera diferente por tipo de escuela.

2.1.1 Esquema Empírico: Función de producción educativa

El análisis empírico es llevado a cabo utilizando la base de datos para el Programa de Evaluación Internacional del estudiante (PISA) conducido por la Organización para la Cooperación y el Desarrollo (OECD) para el año 2009. Con estudiantes de 15 años el estudio mide el desempeño en lenguaje, matemáticas y ciencias, y adicionalmente recoge información a nivel del estudiante y de la escuela. Colombia participa en 2006, 2009 y en la más reciente 2012

[9] Posteriormente cuando se aborde este apartado se explicara que el problema de endogeneidad de nivel 2 se presenta por la correlación entre el termino de error de nivel 2 con variables de nivel 1. En este caso se sospecha que el índice socioeconómico y cultural (ESCS) esa correlacionada con la media del índice, esto es con el indicador del "efecto par". Es de anotar que se descarta el desempeño en PISA del grupo como indicador de calidad de los pares debido al problema de simultaneidad.

cuya base de datos al momento no es publica para los investigadores. Puesto que uno de los objetivos es medir el impacto de la orientación vocacional sobre el desempeño en matemáticas, se trabajara solo con estudiantes de grados decimo y once según la clasificación internacional para programas educativos (ISCED) adoptada por la UNESCO en la cual el nivel 3 corresponde a los niveles mencionados. La razón para escoger los dos últimos años de educación se debe a que en estos grados se imparten los diferentes énfasis vocacionales.

El impacto de la orientación vocacional según tipo de escuela, la existencia de endogeneidad y el efecto contextual serán estimados mediante una función de producción educativa cuya variable dependiente es el puntaje PISA en este caso en el área de matemáticas. Dado que el estudio PISA es realizado mediante diseño muestral por clúster, se obtiene una estructura de datos jerárquica, que requiere el uso de modelos lineales multinivel.

2.2 Orientación Vocacional en Colombia: Una mirada

La orientación vocacional en Colombia tiene su origen en la participación del Estado en la formación para el trabajo (FpT) justificada por fallos del mercado de trabajo y consideraciones redistributivas (Saavedra & Medina, 2012). Como es bien sabido el Estado interviene ante este tipo de fallos, como lo es la escasa formación en competencias cuando se deja en manos de la iniciativa privada, produciendo un resultado socialmente ineficiente[10].

El Estado Colombiano interviene de tres formas en la FpT. La primera es mediante escuelas públicas con orientación técnica a nivel medio (grados decimo y once) mientras que las

[10] Toda vez que una empresa no puede garantizar la permanencia de sus empleados después de capacitarlos y puesto que la capacitación trae consigo un aumento en la productividad laboral que se considera la misma en todas las empresas, no es rentable ante esta incertidumbre que las empresas tengan incentivos para ofrecer capacitación en FpT (Becker, 1994)

restantes son las instituciones de formación técnica y tecnológica públicas, siendo el Servicio Nacional de Aprendizaje (SENA) el referente[11].

Para la definición de FpT se utiliza la consagrada en el decreto 2020 de 2006 el cual define la FpT como "el proceso educativo, formativo, organizado y sistemático, mediante el cual las personas adquieren y desarrollan a lo largo de su vida competencias laborales, especificas o transversales, relacionadas con uno o varios campos ocupacionales referidos a la Clasificación Nacional de Ocupaciones, que le permiten ejercer una actividad productiva como empleado o emprendedor de forma individual o colectiva".

Ahora, no solo las escuelas públicas pueden ofrecer formación para el trabajo dentro de sus currículos, pues también se reconocen los programas que desarrollan las empresas privadas. Dependiendo del enfoque, los estudiantes pueden adquirir saberes para el trabajo en las áreas industrial, comercial o agrícola en escuelas públicas y privadas. La pregunta que surge ahora es si ambos propósitos son compatibles, es decir si la existencia de la orientación vocacional técnica no interfiere con el buen desempeño académico de los estudiantes que les permitan su posterior inclusión en la educación superior.

Saavedra & Medina (2012) encuentran que en términos de desempeño académico medido por los puntajes de las pruebas de Estado para el ingreso a la educación superior, hay poca diferencia en el promedio entre los graduados de escuelas media con orientación técnica respecto a aquellos con orientación académica. Los autores llegan a esta conclusión sin discriminar por tipo de escuelas, luego, ¿Debe colocarse en duda esta regularidad empírica tanto para escuelas privadas como públicas?

[11] La educación media superior (grados decimo y once) con orientación vocacional técnica tiene énfasis en formación comercial, industrial, agropecuaria o pedagógica.

Sorprende que en general sin discriminar por tipo de escuela, se afirme que no existe diferencia entre los desempeños académicos. Si la educación media con orientación técnica tiene asignado el menos el 25% del tiempo total de clases a actividades de orientación vocacional, luego no deberían esperarse rendimientos diferentes según el tipo de orientación vocacional? La respuesta intuitiva es que si, puesto que el estudiante con orientación vocacional técnica tiene menos contenido en horas de clase de cursos como ciencias, matemática y lenguaje, esto sin contar con el tipo de pasantías por tipo de escuela y su interrelación con las actividades académicas.

Se constituye entonces en una de las hipótesis a probar mediante la estimación de modelos multinivel, el que la orientación vocacional es uno de los determinantes del desempeño académico medido por PISA. La hipótesis de que la orientación vocacional tiene un efecto diferenciado en escuelas públicas y privadas, se refuerza sobre el hecho de que el 25% de las escuelas públicas con orientación vocacional técnica tienen enfoque industrial y el 64% currículo comercial[12]. Entretanto, solo el 4% de las escuelas privadas con orientación vocacional técnica tienen un enfoque industrial y el 92% tienen enfoque comercial. La diferencia en el enfoque constituiría una respuesta a la recomposición del empleo que favorece al sector de servicios más que al sector manufacturero (Saavedra & Medina, 2010).

La anterior regularidad empírica constituye la primera razón para postular como hipótesis principal de trabajo, el que las escuelas públicas y privadas exhiben dinámicas educativas totalmente diferentes, lo cual puede relacionarse con el tipo de pasantías asignadas a la orientación técnica en uno u otro tipo de establecimiento.

[12] El enfoque de la oferta de los colegios privados con orientación vocacional técnica está en las áreas de contabilidad, administración de negocios, comunicaciones y tecnologías informáticas entre tanto el enfoque de la publica esta en el área industrial.

Según el estudio adelantado por Bettinger, Kremer & Saavedra (2010), el tipo de pasantías que ofrecen las escuelas públicas con orientación técnica difiere de las privadas, pues mientras en las primeras las pasantías suelen adelantarse en el SENA, en las segundas las pasantías se adelantan en el sector privado ya sea en oficinas o en empresas pequeñas.

El análisis del desempeño educativo debe ir más allá de la tradicional comparación de desempeños académicos entre los estudiantes de ambos tipos de escuela; el comparar puntajes promedio de pruebas de colegios públicos y privados resulta algo muy simplista. En este orden de ideas es pertinente estudiar el efecto diferenciado de la orientación vocacional sobre el desempeño académico de los estudiantes de colegios privados y públicos. La incorporación de las variables, tipo de escuela y orientación académica mediante la utilización de variables falsas permite diferenciar el efecto de la orientación por tipo de escuela[13].

[13] Dado que los diferentes tipos de currículo pueden atraer a diferentes tipos de estudiantes, la diferencia en el desempeño académico observado puede ser el resultado de las características de los estudiantes y no del tipo de orientación vocacional elegida. Es por ello que en los modelos multinivel en la ecuación de nivel 1 se aislara el efecto de la orientación vocacional después de tener en cuenta las características del estudiante.

3. Modelos Multinivel

3.1 Una introducción a los modelos multinivel

En investigación social a menudo los problemas de investigación están inmersos en estructuras jerárquicas en las cuales las unidades se agrupan en diferentes niveles o etapas. Tal como lo anota Goldstein (2002) la existencia de estructuras jerárquicas en los datos no es accidental y por lo tanto no puede ignorarse. El término de datos multinivel es típicamente usado para describir este tipo de estructura en la cual las unidades de análisis de nivel 1 son consideradas un subconjunto de unidades de nivel 2[14].

El objetivo del análisis multinivel mediante la estimación de modelos jerárquicos lineales (HLM) es modelar la variación de la variable dependiente que es medida en el nivel más bajo considerando la información de todos los niveles de análisis (Steenbergen & Jones, 2002).

Calculando una ecuación diferente para cada nivel de agregación, estos modelos expresan las relaciones que se producen entre las variables de un mismo nivel, así como las dinámicas entre las variables de diferentes niveles[15].

En este caso esto el puntaje de la prueba PISA está en el nivel 1 (nivel más bajo) correspondiente a los estudiantes y las variables explicativas se encuentran en diferentes niveles

[14] Variables del estudiante y de la escuela, con la inmediata consecuencia de que las hipótesis sobre las relaciones entre las variables se definen en diferentes niveles de la estructura jerárquica.

[15] En este caso el estudiante constituiría un primer nivel, el cual se encuentra anidado dentro de un segundo nivel, escuelas.

de la estructura jerárquica[16]. La inmediata consecuencia es que las hipótesis de las relaciones entre las variables se definen en diferentes niveles de la estructura jerárquica.

A nivel conceptual el modelo multinivel puede ser visto como un sistema jerárquico de ecuaciones, el cual permite obtener estimaciones estadísticamente más eficientes en relación al análisis tradicional que ignora esta estructura jerárquica (Hox, 2010).

3.2 Justificación de los modelos multinivel

Empezando con un análisis tradicional que lleva a cabo una regresión estándar del puntaje de la prueba PISA en matemáticas sobre variables del estudiante:

$$Y_i = \beta_0 + \beta_1 X_i + e_i \tag{3.2.1}$$

Pudiendo extenderse este modelo a un modelo múltiple introduciendo variables como género, y nivel socioeconómico de la familia, por nombrar algunas, estos modelos no reconocen que los estudiantes están inmersos dentro una estructura jerárquica. El modelo dado en ecuación (3.2.1) ignora el hecho de que los estudiantes pueden compartir características de la escuela y que estas pueden influenciar las relaciones entre el desempeño académico y las variables de los estudiantes.

Si un investigador se restringe a usar la metodología estándar de mínimos cuadrados ordinarios (MCO) y no tiene en cuenta la variación contenida dentro de la estructura de datos multinivel, luego la correlación entre los términos de error dentro de los grupos o escuelas guían a una incorrecta estimación de los errores estándar de los efectos fijos, llevando a una inferencia

[16] Un modelo de regresión lineal que no tenga en cuenta la estructura jerárquica de datos, no diferenciara entre la varianza que es debida del estudiante y la debida a la escuela, en contraste con el modelo multinivel el cual logra diferenciar que parte del logro es explicado por el estudiante y que parte por la escuela.

estadística imprecisa[17]. Por otra parte los estimadores MCO que ignoran la estructura jerárquica están sesgados hacia arriba en comparación con las estimaciones de efectos fijos del modelo multinivel.

La oportunidad de estimar los efectos fijos con variación "al interior" o "entre" constituye la segunda ventaja de estimar con modelos multinivel. ¿Qué estimador debería ser empleado? Cuando se está principalmente interesado en las variables independientes de nivel 1 más que en las variables de nivel 2, el estimador "al interior del grupo" es a menudo de interés especial, mientras que cuando el interés reside en el efecto de las variables de nivel 2, el estimador "entre grupos" es el más adecuado[18].

3.3 Estimador de datos agrupados, una primera aproximación

Cuando se estiman los parámetros de modelos cuyas variables explicativas representan características de observaciones individuales que están anidadas dentro de grupos, una regresión con datos agrupados (pooled OLS) usa una diferente combinación de variación "dentro del grupo" (within) y "entre grupos" (between) respecto al modelo multinivel.

Pooled OLS producirá estimaciones que estarán más cerca de la metodología "entre grupos" cuando la mayor parte de la variación en las variables ocurra entre grupos, y más cercana

[17] Los errores de estándar estimados bajo OLS solo serán correctos cuando los términos de error a través de las observaciones estén incorrelacionados, lo cual es bastante es improbable en estudios que usan datos multinivel en un contexto de estudiantes anidados dentro de las escuelas. Respecto a esto último, los modelos multinivel no requieren del cumplimiento estricto de incorrelación entre los términos de error de primer nivel.

[18] El programa HLM 7.0 permite entre sus opciones la de obtener estimadores "entre" o "al interior", dependiendo de cómo se especifique la manera de centrar las variables; centrando en la media general o en la media de grupos. En el capítulo 3 cuando se aborde la solución al problema de endogeneidad se abordaran en detalle estas dos formas de estimación de los efectos fijos

a la variación "dentro del grupo" cuando la mayor parte de la variación ocurra dentro de los grupos.

A pesar de la relativa fácil implementación de un modelo por mínimos cuadrados ordinarios combinado (Pooled OLS), no deja de tener una serie de limitantes para el análisis de estructuras jerárquicas, toda vez que las variables dicotómicas son solo indicadoras de diferencia de subgrupos (tiene en cuenta la heterogeneidad para las unidades de nivel 2) sin explicar el origen de la heterogeneidad (Steenbergen & Jones, 2002). Otro inconveniente con este enfoque es que el estimador de una variable de nivel 1 podría estar sesgado para una unidad particular de nivel 2 en la que la el verdadero efecto del predictor tenga un signo diferente a la estimación combinada por MCO.

3.4 Modelo Nulo

Empezando por el modelo más básico, Kreft & de Leew (1998) definen un modelo nulo en el cual el puntaje de la prueba de matemáticas para el individuo i en la escuela j es función de un intercepto al nivel del clúster (β_{oj}). Adicional a este intercepto que varía aleatoriamente a través de los clúster o escuelas, existe un término aleatorio de error individual e_{ij}asumido con distribución normal con parámetros de media cero y varianza constante, σ_e^2.

El termino de error asociado dentro del clústere_{ij} es la desviación del puntaje de la persona *i* en el clúster *j* de la media del clúster, β_{0j}

$$Y_{ij} = \beta_{0j} + e_{ij} \tag{3.4.1}$$

Existiendo un efecto que es común a todos los estudiantes dentro de la misma escuela, es necesario especificar una ecuación separada para el intercepto[19]. De esta manera se modela β_{0j} en un nivel 2 en función de la media general del puntaje para todas las escuelas (β_{00}) y de una desviación (u_{0j}) del *j's* intercepto del intercepto general, en otras palabras, una desviación de la media del clúster j de la media general (Ma, X. ,Ma, L., & Bradley K, 2008)

$$\beta_{0j} = \beta_{00} + u_{0j} \tag{3.4.2}$$

$$\text{Con } u_{0j} \sim (0, \sigma^2_{u0})$$

Combinando las anteriores ecuaciones tenemos que el modelo incondicional puede ser expresado como:

$$Y_{ij} = \beta_{00} + u_{0j} + e_{ij} \tag{3.4.3}$$

$$Y_{ij} = \beta_{00} + \xi_{ij} \tag{3.4.4}$$

$$\xi_{ij} = e_{ij} + u_{0j} \tag{3.4.5}$$

Asumiendo que ambos términos de error son independientes, $\mathrm{cov}(e_{ij}, u_{0j}) = 0$

Aunque sin variables explicativas, la estimación del modelo dado en ecuación (3.4.3) tiene como objetivo validar el uso de los modelo multinivel al estimar los componentes de la varianza del puntaje a nivel del estudiante (varianza dentro de la escuela, σ^2_e) y de la escuela (varianza entre la escuelas, σ^2_{u0}). Si los dos niveles de análisis son relevantes para estudiar el desempeño de los estudiantes, debería encontrarse mediante el análisis de descomposición de varianza (ANOVA) que los dos componentes de la varianza son estadísticamente significativos (Steenbergen & Jones, 2002).

[19] Dado que se asume que los clúster fueron extraídos de una población de clúster.

El primer componente se dentro de las estructuras jerárquicas se debe a la dependencia entre las respuestas observadas para las unidades que pertenecen al mismo clúster, en este caso la escuela. De hecho en la ecuación (3.4.4) es claro que la dependencia entre dos puntajes se debe únicamente al intercepto aleatorio (u_{0j}) que comparten dos observaciones dentro del mismo clúster.

En los modelos presentados en las ecuaciones (3.4.3) y (3.4.4) β_{0j} se denomina efecto aleatorio y en el nivel 2, β_{00} es denominado efecto fijo. En este caso el interés se centra en la estimación del efecto fijo y de los componentes de la varianza σ_e^2 y σ_{u0}^2.

3.5 Especificando los componentes de la varianza

Partiendo de la ecuación (3.4.4) se observa que cada puntaje individual difiere de la media del puntaje total β_{00} por un residuo total ξ_{ij}. En este caso la descomposición de la comienza con el cálculo de la varianza de la variable dependiente:

$$\mathrm{Var}(Y_{ij}) = E\left\{\left(Y_{ij} - E(Y_{ij})\right)^2\right\} \tag{3.5.1}$$

Dados los supuestos acerca de los dos términos de error se tiene:

$$E(Y_{ij}) = E(\beta_{00}) + E(\xi_{ij}) \tag{3.5.2}$$

$$E(u_{0j} + e_{ij}) = 0 \tag{3.5.3}$$

$$E(Y_{ij}) = \beta_{00} \tag{3.5.4}$$

$$\mathrm{Var}(Y_{ij}) = E\left\{(Y_{ij} - \beta_{00})^2\right\} = E\left\{(\xi_{ij})^2\right\} = E\left\{(u_{0j} + e_{ij})^2\right\} \tag{3.5.5}$$

$$\mathrm{Var}(Y_{ij}) = E(u_{0j}{}^2) + 2\mathrm{cov}(u_{0j}, e_{ij}) + E(e_{ij}{}^2) \tag{3.5.6}$$

$$\mathrm{Var}(Y_{ij}) = \sigma_{u0}^2 + \sigma_e^2 \quad (3.5.7)$$

Con el fin de estimar el peso de los diferentes niveles en el análisis se considera la participación de cada componente de la varianza en el total de la varianza $\sigma_e^2 + \sigma_{u0}^2$. El porcentaje de la variación observada en la variable dependiente atribuible a características de la escuela puede calcularse como el cociente entre σ_{u0}^2 y la varianza total:

$$\rho = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2} \quad (3.5.8)$$

Referida como correlación dentro del nivel 2, o “correlación intraclase” es siempre positiva y contenida en el intervalo 0-1. Este coeficiente ρ es similar al coeficiente de bondad de ajuste R^2 en un modelo de regresión lineal toda vez que expresa la proporción de la variabilidad total que es explicada por las variables de segundo nivel[20]. De enorme utilidad, ρ permite vislumbrar la importancia de los diferentes niveles en el análisis al considerar la participación de cada componente de la varianza en el total de la varianza $\sigma_e^2 + \sigma_{u0}^2$.

Por otra parte la ratio de σ_e^2 sobre el total de la varianza indica la importancia del nivel 1 dentro del análisis.El porcentaje de la varianza atribuible a las características del estudiantes puede calcularse como:

$$1 - \rho = 1 - \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2} = \frac{\sigma_e^2}{\sigma_{u0}^2 + \sigma_e^2} \quad (3.5.9)$$

[20] Si se encontrara una significativa varianza para u_{0j} es deseable incluir variables macro o de nivel 2 para tener en cuenta parte de esta variación.

3.6 Valores plausibles

Tomando como variable dependiente el puntaje de la prueba de matemáticas (PVMATH) debe anotarse que no está disponible el puntaje de la prueba de manera directa sino a través de valores plausibles.

Las evaluaciones internacionales a gran escala como lo es la prueba PISA y el estudio de tendencias de matemáticas y ciencias (TIMSS) suele administrar lo que se denomina una matriz de muestreo, en la cual diferentes test cortos son administrados a los estudiantes en un tiempo limitado. Esto se hace con el fin de ofrecer información comparable sobre las habilidades del estudiante y conocimiento en las áreas como matemáticas, lenguaje y ciencias.

Dado que los estudiantes no completan los diferentes test (no obstante cada uno termina una parte de su evaluación) luego el desempeño en las distintas pruebas no puede ser obtenido a través de los tradicionales pruebas si no que se basa en valores plausibles[21]. Los valores plausibles son valores imputados que se asemejan a las puntuaciones de las pruebas individuales y tienen aproximadamente la misma distribución del rasgo latente que se está midiendo.

Desarrollados como una aproximación computacional para obtener estimaciones consistentes de características de la población, estos permiten acercase de manera confiable a la verdadera distribución de los desempeños. En el caso de PISA y TIMSS, un conjunto de cinco valores plausibles debe ser usado para generar estimaciones de los estadísticos de interés.

[21] Los valores plausibles fueron desarrollados para el análisis de 1983-84 NAEP (Evaluación Nacional de Progreso Educativo) por Mislevy, Johnson Y Muraki (1990), basado en el trabajo de Rubín (1978) sobre múltiples imputaciones. Esta metodología fue utilizada en todos los estudios posteriores de NAEP, TIMSS y ahora PISA.

Utilizados para obtener una estimación precisa de la capacidad del estudiante, los valores plausibles en si no constituyen los puntajes de las pruebas de manera que deben ser combinados apropiadamente para ser utilizados en análisis multinivel[22]

[22]Utilizar como variable dependiente un promedio los valores plausibles puede producir estimaciones sesgadas y debe ser desalentado. Para la estimación de los modelos se optó por la utilización del software, HLM 7.0, toda vez que permite trabajar con valores plausibles. Aunque STATA trabaja con modelos multinivel, la opción de trabajar con valores plausibles solo esta disponibles para trabajar con modelos de regresión lineal, modelos probit y logit, logit y probit ordenado, modelos logit multinomial y regresión por cuantiles.

4. Métodos de estimación y supuestos del modelo

4.1 Máxima verosimilitud: procedimientos de estimación con datos no balanceados

Asumiendo distribuciones normales para los errores tanto del nivel y como del nivel 2, el método de máxima verosimilitud (MV) obtiene los estimadores de los efectos fijos y los componentes de la varianza que maximizan la función de verosimilitud[23]. Como es ya conocido MV tiene propiedades deseables como lo es la de estimadores consistentes y asintóticamente eficientes cuando el supuesto de normalidad se mantiene. No obstante cuando el supuesto de normalidad sea violado, el estimador de los efectos fijos es consistente (Goldstein, 2002).

Siguiendo el desarrollo de Swaminathan H. & Rogers J. (2008):

Modelo de nivel 1

$$y_j = X_j\beta + e_j \tag{4.1.1}$$

Modelo de nivel 2

$$\beta = W_j\gamma + u_j \tag{4.1.2}$$

Combinando las dos ecuaciones y obteniendo el modelo mixto:

$$y_j = X_jW_j\gamma + X_ju_j + e_j \tag{4.1.3}$$

Siendo $A_j = X_jW_j$, se tiene que:

$$y_j = X_jW_j\gamma + X_ju_j + e_j = \gamma A_j + X_ju_j + e_j \tag{4.1.4}$$

[23] Estos supuestos serán testeados posteriormente cuando se realicen los test sobre los supuestos sobre los términos de error y correcta especificación del modelo.

$$\mathrm{var}(y_j) = \mathrm{var}(X_j u_j + e_j) = X_j T X_j' + \sigma_e^2 I \equiv \Psi_j \qquad (4.1.5)$$

Bajo el supuesto de que u_j y e_j tienen distribución normal multivariada:

$$y_j \sim N(A_j\gamma, \Psi_j) \qquad (4.1.6)$$

La función de verosimilitud es:

$$L(y:\gamma,\sigma_e^2,T) = \prod_{j=1}^{J} \frac{1}{|\Psi_j|} \exp\left\{-\frac{1}{2}(y_j - A_j\gamma)'\Psi_j^{-1}(y_j - A_j\gamma)\right\} \qquad (4.1.7)$$

Tomando el logaritmo de la función:

$$\log L(y:\gamma,\sigma_e^2,T) = \sum_{j=1}^{J} \log(|\Psi_j|) - \frac{1}{2} - \frac{1}{2}\sum_{j=1}^{J}(y_j - A_j\gamma)'\,\Psi_j^{-1}(y_j - A_j\gamma) \qquad (4.1.8)$$

Las estimaciones de máxima verosimilitud son obtenidas resolviendo las siguientes ecuaciones[24]:

$$\frac{\partial}{\partial\theta}\log L(y:\gamma,T,\sigma^2) = \begin{bmatrix} \frac{\partial}{\partial\gamma}\log L(y:\gamma,T,\sigma^2) \\ \frac{\partial}{\partial\sigma^2}\log L(y:\gamma,T,\sigma^2) \\ \frac{\partial}{\partial T}\log L(y:\gamma,T,\sigma^2) \end{bmatrix} \qquad (4.1.9)$$

Existen dos metodologías de estimación por máxima verosimilitud que son comúnmente usadas en modelos multinivel lineales. El primero es Full Máxima Verosimilitud (FML) el cual incluye los coeficientes de regresión y los componentes de la varianza en la función de verosimilitud. El segundo método de estimación es el de Máxima Verosimilitud Restringida (RML), el cual solo incluye en la función mencionada a los componentes de la varianza; ambas metodologías tienen como supuesto que los residuos en el nivel más bajo se distribuyen normal con media 0 y varianza σ_e^2.

La diferencia entre ambos métodos es que FML trata las estimaciones de los coeficientes como cantidades conocidas cuando los componentes de la varianza son estimados, mientras que

[24] Denotando a θ como el vector de parámetros

RML trata las estimaciones de los parámetros como si tuvieran cierto grado de incertidumbre. En este orden de ideas RML es más realista y debería en teoría guiar a mejores estimaciones especialmente cuando el número de grupos es pequeño (Raudenbush & Bryk, 2002).

Ahora, FML tiene dos ventajas sobre RML, además de que los cálculos son más fáciles, puesto que los coeficientes de regresión son incluidos en la función de verosimilitud, una ratio de verosimilitud puede ser usado para realizar test de diferencias entre dos modelos que difieren solo en el número de coeficientes fijos. Entretanto con RML solo se pueden comparar modelos con diferente número de componentes de la varianza, pero igual número de coeficientes fijos[25].

En una situación ideal para la estimación de un modelo multinivel como lo es un diseño balanceado de los datos (igual número de unidades de nivel 1 por unidades de nivel 2), las ecuaciones de máxima verosimilitud dadas en ecuación (3.1.9) pueden ser resueltas analíticamente en la forma cerrada.

Cuando no se tienen datos balanceados, las tres ecuaciones (componentes de la varianza y los efectos fijos) deben ser resueltas simultáneamente y dado que se trata de ecuaciones no lineales, una solución explicita no es posible. En este caso una solución iterativa es alcanzada mediante el empleo de métodos numéricos, entre cuyas alternativas se encuentran el algoritmo de Newton-Raphson, el algoritmo de puntuación de Fisher, los mínimos cuadrados generalizados factibles (IGLS) y el algoritmo de maximización EM; se expondrá brevemente este último dado que está programado en la rutina de estimación del software utilizado, HLM 7.0.

[25] Ambos RML y FML producen idénticas estimaciones de efectos fijos, no obstante la última toma en cuenta los grados de libertad de los efectos fijos y por lo tanto produce estimaciones de los componentes de la varianza menos sesgadas.

4.2 Estimación máxima verosimilitud con el algoritmo E-M.

Siguiendo el desarrollo Swaminathan H. & Rogers J. (2008) se expone este método para un modelo nulo de dos niveles (en aras de simplificar la exposición). Recordando el modelo sin covariables:

$$\mathrm{Y_{ij}} = \beta_{00} + \mathrm{u_{0j}} + \mathrm{e_{ij}} \tag{4.2.1}$$

Con:

$$\mathrm{var}(\mathrm{u_{0j}}) = \sigma^2_{\mathrm{u0}} \quad ; \quad \mathrm{var}(\mathrm{e_{ij}}) = \sigma^2_{\mathrm{e}} \tag{4.2.2}$$

Si u_{0j} es observado luego $Y_{ij}^* = (Y_{ij} - u_{0j})$ y los parámetros de interés pueden ser estimados. Puesto que los u_{0j} no son observados, el conjunto de datos Y_{ij} son denominados datos incompletos y los u_{0j} considerados "missing". En situaciones de datos faltantes, se procede a derivar un procedimiento que reemplace los valores faltantes por cantidades observables en orden a estimar los parámetros.

Son los autores Dempster, Laird & Rubin (1977) quienes desarrollan un algoritmo para estimar parámetros en modelos complejos. Demuestran que sustituir la expectativas condicionadas de los estadísticos para los valores perdidos en la función de máxima verosimilitud, maximiza la función al momento de obtener las estimaciones.

Requiriendo la esperanza condicional de u_{0j}, se condicionara sobre $\bar{Y}_j$ dado que es equivalente condicionar sobre Y_{ij} toda vez que el modelo puede ser expresado como[26]:

$$\bar{\mathrm{Y}}_{\mathrm{j}} = \beta_{00} + \mathrm{u_{0j}} + \bar{\mathrm{e}}_{\mathrm{j}} \tag{4.2.3}$$

La esperanza condicional de u_{0j} dado $\bar{Y}_j$ esta dado por la usual expresión:

---

[26] Por motivos de simplicidad se utilizara el modelo nulo para esta exposición.

$$E\left(u_{0j}\middle|\bar{Y}_j\right) = E\left(u_{0j}\right) \; + \frac{cov\left(u_{0j},\bar{Y}_j\right)}{var\left(\bar{Y}_j\right)} + \left[\bar{Y}_j - E\left(\bar{Y}_j\right)\right] \tag{4.2.4}$$

De la ecuación [3.2.3] y dado el supuesto de $E\left(u_{0j}\right) = 0$, se tiene que

$$E\left(\bar{Y}_j\right) = \beta_{00} \; ; \; var\left(\bar{Y}_j\right) = \sigma^2_{u0} + \frac{\sigma^2_{u0}}{n_j} \tag{4.2.5}$$

Calculando la covarianza:

$$cov\left(u_{0j},\bar{Y}_j\right)= cov\left(u_{0j}, u_{0j} + \bar{e}_j\right)= cov\left(u_{0j}, u_{0j}\right) + cov\left(u_{0j}, \bar{e}_j\right) = \sigma^2_{u0} \tag{4.2.6}$$

Sustituyendo (3.2.5) y (3.2.6) en (3.2.4) obtenemos:

$$E\left(u_{0j}\middle|\bar{Y}_j\right) = E\left(u_{0j}\right) + \sigma^2_{u0}\left[\sigma^2_{u0} + \frac{\sigma^2_{u0}}{n_j}\right]^{-1} + \left[\bar{Y}_j - E\left(\bar{Y}_j\right)\right] = \lambda_j\left(\bar{Y}_j - \beta_{00}\right) \tag{4.2.7}$$

Siendo:

$$\lambda_j = \frac{\sigma^2_{u0}}{\sigma^2_{u0}+\sigma^2_e/n_j} \tag{4.2.8}$$

El algoritmo E-M requiere de valores iníciales para los parámetros que en este caso se les denota con un subíndice adicional, $\beta_{000}, \sigma^2_{u00}\, y\, \sigma^2_{e0}$, los cuales son obtenidos mediante un análisis de varianza (ANOVA). Obtenidos estos valores iníciales se calcula λ_j el cual entra en el calculo de $u^*_j = \lambda_{j0}\left(\bar{Y}_j - \beta_{000}\right)$. Este valor u^*_j es usado después para calcular en la estimación de los efectos fijos.

Para el cálculo de los componentes de la varianza σ^2_{u0} y σ^2_e se requiere la esperanza condicional de u^2_j, $E\left(u_{0j}\middle|\bar{Y}_j\right)^2$la cual se obtiene a continuación:

$$Var\left(u_{0j}\middle|\bar{Y}_j\right) \equiv E\left\{\left[u_{0j} - E\left(u_{0j}\right)\right]\middle|\bar{Y}_j\right\}^2 = E\left\{\left[u_{0j} - u^*_{0j}\right]\middle|\bar{Y}_j\right\}^2 \tag{4.2.9}$$

$$E\left\{\left[u_{0j} - u^*_{0j}\right]\middle|\bar{Y}_j\right\}^2 = E\left\{\left[{u_{0j}}^2 - 2u_{0j}u^*_{0j} + {u^*_{0j}}^2\right]\middle|\bar{Y}_j\right\} \tag{4.2.10}$$

Dónde:

$$\mathrm{E}\left(\mathrm{u_{0j}}^2\middle|\overline{\mathrm{Y}}_\mathrm{j}\right) = \mathrm{u}_\mathrm{j}^* + \mathrm{E}\left(\mathrm{u}_\mathrm{j}^2\right) = \mathrm{u}_\mathrm{j}^{*2} + \mathrm{v}_\mathrm{j} \qquad (4.2.11)$$

$$v_j = var(u_j)\left[1 - \frac{cov(u_{0j},\bar{y}_j)^2}{var(u_{0j})var(\bar{Y}_j)}\right] = \sigma_{u0}^2\left[1 - \frac{(\sigma_{u0}^2)^2}{\sigma_{u0}^2(\sigma_{u0}^2+\sigma_e^2/n_j)}\right] = \sigma_{u0}^2(1-\lambda_j) \qquad (4.2.12)$$

$$\hat{\beta}_{00} = \frac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}\left(\mathrm{Y_{ij}} - \mathrm{u}_\mathrm{j}^*\right)}{\mathrm{N}} \qquad (4.2.13)$$

$$\sigma_{u0}^2 = \frac{\sum_{j=1}^{J}\left(\mathrm{u}_{0j}^{*\,2} + \mathrm{v_j}\right)}{\mathrm{J}} \qquad (4.2.14)$$

$$\sigma_e^2 = \frac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}\left[\left(\mathrm{Y_{ij}} - \hat{\beta}_{00} - \mathrm{u}_\mathrm{j}^*\right) + \mathrm{v_j}\right]}{\mathrm{N}} \qquad (4.2.15)$$

Resumiendo el procedimiento que sigue le algoritmo E-M, en un primer paso se obtienen los valores iníciales para $\beta_{00}, \lambda_j, \sigma_e^2\ y\ \sigma_{u0}^2$. Una vez calculados en un segundo paso se calculan las esperanzas condicionales $E\left(u_{0j}\middle|\bar{Y}_j\right)$, $E\left(u_{0j}{}^2\middle|\bar{Y}_j\right)$ y $E\left\{\left(Y_{ij} - \hat{\beta}_{00} - u_j\right)^2\middle|\bar{Y}_j\right\}$. Por últimose calculan las estimaciones MV de los parámetros mediante la sustitución de las expectativas en las ecuaciones (4.2.13)-(4.2.15). Los pasos 2 y 3 se repiten hasta que se alcance la convergencia con un criterio preestablecido que para el software HLM es 0.00001 (Raudenbush, Brick, Cheong &. Congdon, 2004).

Aunque el algoritmo E-M es de fácil implementación, se argumenta que su convergencia es lenta frente a lo cual el software HLM 7.0 hace uso del acelerador Aitkin, mejorando la velocidad de convergencia al nivel del algoritmo de Newton Raphson[27].

[27] Nótese que la elección del modelo se realiza independiente de la elección del método y este a su vez se elige independiente de la elección del algoritmo, y la elección del algoritmo de la elección del Software (Raudenbush et al 2002)

4.3 Supuestos del modelo

La especificación del modelo multinivel está incompleto sin la especificación de los supuestos. Partiendo del siguiente modelo de dos niveles con parámetros aleatorios

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij} \tag{4.3.1}$$

$$\beta_{0j} = \beta_{00} + u_{0j} \tag{4.3.2}$$

$$\beta_{1j} = \beta_{10} + u_{1j} \tag{4.3.3}$$

Abordando los supuestos de los errores de primer nivel, e_{ij} se asumen sigue una distribución normal con media 0 y varianza σ_e^2.

$$E\left(e_{ij}\right) = 0;\ var(e_{ij}) \tag{4.3.4}$$

Entre tanto los errores a nivel de la escuela u_j son asumidos con media 0 y matriz de varianzas-covarianzasΣ. Puesto que los niveles de error del nivel escuelas son las desviaciones de las escuelas esto es equivalente a asumir que los coeficientes β_j siguen una distribución normal multivariada (Hox, 2010). Resumiendo:

$$E(u_{0j}) = E(u_{1j}) = 0 \tag{4.3.5}$$

$$var\left(u_{0j}\right) = \sigma_{u0}^2$$

$$var\left(u_{1j}\right) = \sigma_{u1}^2$$

$$Cov(u_{0j}, u_{1j}) = \sigma_{u01}$$

Este último supuesto significa que los errores de nivel 2 sobre pueden estar correlacionados (Steenbergen & Jones, 2002). Los anteriores cuatro supuestos implican que los errores de nivel 2 están distribuidos normal bivariado con media cero y matriz de varianzas y covarianzas:

$$\Sigma = \begin{bmatrix} \sigma_{u0}^2 & \sigma_{u01} \\ \sigma_{u01} & \sigma_{u1}^2 \end{bmatrix} \tag{4.3.6}$$

Retomando el modelo con pendiente e intercepto aleatorios, incluyendo covariables de nivel 2 (W_j):

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij} \tag{4.3.7}$$

$$\beta_{0j} = \beta_{00} + \beta_{01}W_j + u_{0j} \tag{4.3.8}$$

$$\beta_{1j} = \beta_{10} + \beta_{11}W_j + u_{1j} \tag{4.3.9}$$

$$Y_{ij} = [\beta_{00} + \beta_{10}X_{ij} + \beta_{01}W_j + \beta_{11}W_jX_{ij}] + \varepsilon_{ij} \tag{4.3.10}$$

$$\varepsilon_{ij} = [u_{0j}+u_{1j}X_{ij} + e_{ij}] \tag{4.3.11}$$

Se analizan ahora los supuestos del término de error compuestoε_{ij}:

$$var[\varepsilon_{ij}]=E\left[\left(u_{0j}+u_{1j}X_{ij} + e_{ij}\right)^2\right] \tag{4.3.12}$$

$$=E\left[u_{0j}^2\right] + X_{ij}^2E\left[u_{1j}^2\right] + E\left[e_{ij}^2\right] + 2X_{ij}E\left[u_{0j}u_{1j}\right] + 2E\left[u_{0j}e_{ij}\right] + 2X_{ij}E\left[u_{1j}e_{ij}\right]$$

$$=\sigma_{u0}^2 + X_{ij}^2\sigma_{u1}^2 + \sigma_e^2 + 2X_{ij}\sigma_{u01} \tag{4.3.13}$$

Es claro que aun siendo los errores de nivel 1 homocedásticos, la varianza de ε_{ij} no es constante al ser función de las variables de nivel 1. La única manera en que el termino de error ε_{ij} sea homocedastico es que el error u_{1j} sea cero lo cual significaría que la variableW_j es suficiente para modelar las diferencias en la pendiente de X_{ij} a través de las unidades de nivel 2 (Steenbergen & Jones, 2002).

El segundo supuesto que se viola en los modelos multinivel es el de no autocorrelación para los términos de perturbación de nivel 1,ε_{ij}, anidados dentro de las mismas unidades de nivel 2:

$$cov\left(\varepsilon_{ij}, \varepsilon_{kj}\right) = E\left[\left(u_{0j}+u_{1j}X_{ij} + e_{ij}\right)\left(u_{0j}+u_{1j}X_{kj} + e_{kj}\right)\right] \tag{4.3.14}$$

$$=E\left[u_{0j}^2\right] + X_{ij}E\left[u_{0j}u_{1j}\right] + E\left[e_{ij}u_{0j}\right]+ X_{kj}E\left[u_{1j}u_{0j}\right] + X_{ij}X_{kj}E\left[u_{1j}u_{1j}\right] + X_{ij}E\left[u_{1j}e_{kj}\right] + E\left[e_{ij}u_{0j}\right] +$$

$$X_{kj}E\left[e_{ij}u_{1j}\right] + E\left[e_{ij}e_{kj}\right]$$

$$=\sigma_{u0}^{2}+X_{ij}\sigma_{u01}+X_{kj}\sigma_{u01}+X_{ij}X_{kj}\sigma_{u1}^{2}$$

La covarianza será nula siempre y cuando u_{0j} y u_{1j} sean iguales a cero, lo que llevaría a la conclusion de que la variable W_j da cuenta perfecta del movimiento de los movimientos del intercepto y la pendiente de las unidades de nivel 1 a través de las unidades de nivel 2.

Es en este punto donde los modelos multinivel son necesarios dado que los datos agrupados violan el supuesto de independencia de las observaciones (Maas & Hox, 2004).Puede observarse entonces que los modelos multinivel se adaptan mejor a estructuras de datos jerárquicas en los cuales la varianza no es constante en contraste con el tradicional análisis OLS el cual asume varianza constante.

De otro lado se asume término de error de nivel 1 y las variables independientes. Es decir las características no observables de los estudiantes incluidos en el término de error no deben estar correlacionadas con las características observables de los estudiantes:

$$\mathrm{cov}(X_{ij},e_{ij})=0 \tag{4.3.15}$$

En la misma vía las variables independientes del nivel 2 no están correlacionadas con el término de error del nivel 2, esto es, las características no observables de la escuela incluidas en el término de error no deben estar correlacionadas con las características observables $\bar{X}_{.j}$ y W_j:

$$\mathrm{cov}(W_j,u_{0j})=\mathrm{cov}(W_j,u_{1j})=0 \tag{4.3.16}$$

Complementan los anteriores supuestos el que las variables independientes en cada nivel no están correlacionadas con los términos de error del otro nivel, esto es, cualquier característica observable de la escuela no debe estar correlacionada con características no observables del estudiante.

$$\mathrm{cov}(W_j,e_{ij})=0;\quad \mathrm{cov}(\bar{X}_{.j},e_{ij})=0 \tag{4.3.17}$$

Asimismo cualquier característica observable del estudiante no debe estar relacionada con las características no observables de la escuela.

$$\mathrm{cov}(X_{ij}, u_{0j}) = \mathrm{cov}(X_{ij}, u_{1j}) = 0 \qquad (4.3.18)$$

La violación de los anteriores supuestos se traduce en endogeneidad, la cual surge cuando existen covariables no observables (e incluidas en el término de error) que afectan la respuesta y están correlacionadas con las covariables incluidas en el modelo. En este caso donde se trabajan con modelos multinivel de efectos aleatorios con un término de error en cada nivel, el problema de endogeneidad puede ocurrir en cualquiera de los niveles.

En la presente investigación se prestara especial atención a la endogeneidad de nivel 2 presente cuando algunas características no observables de la escuela incluida en el término de error del nivel 2están correlacionadas con las características observables del estudiante:

$$\mathrm{cov}(X_{ij}, u_{0j}) \neq 0; \quad \mathrm{cov}(X_{ij}, u_{1j}) \neq 0 \qquad (4.3.19)$$

No controlar apropiadamente por los determinantes a nivel del grupo que están correlacionados con las variables individuales, se traducirá en un sesgo para los coeficientes de nivel 1 al contener el verdadero efecto causal a nivel individual, adicional a una parte del efecto a nivel del grupo.

Dado que la endogeneidad de nivel 2 se produce básicamente por la omisión de variables relevantes a nivel del clúster, la estrategia a seguir es eliminar el sesgo tanto como sea posible introduciendo variables contextuales de la escuela, cuya omisión se sospecha son la principal fuente de endogeneidad (Rangvid, 2008).

En particular el nivel socioeconómico y cultural de la familia se cree condiciona la elección de los padres de los “pares académicos” de los estudiantes. En este orden de ideas algunas omitir efectos contextuales de la escuela como por ejemplo el "efecto par", puede generar un sesgo por endogeneidad (Hanchane & Mostafa, 2010)

4.4 Proponiendo una solución a la endogeneidad: Una adaptación de la aproximación de Mundlak

A continuación para solucionar el problema de endogeneidad de nivel 2 se utiliza la aproximación de Mundlak (1978) utilizada en datos de panel. Partiendo del modelo con intercepto aleatorio

$$Y_{ij} = \beta_{00} + \beta_{10}X_{ij} + u_j + e_{ij} \qquad (4.4.1)$$

Este modelo análogo al de efectos aleatorios en datos de panel[28]

$$Y_{it} = \mu + \beta_1 X_{it} + \alpha_i + u_{it} \qquad (4.4.2)$$

Mundlak (1978) critica el anterior modelo dado que puede existir correlación entre la heterogeneidad no observada y las variables explicativas $\mathrm{X_{it}}$,lo cual en el modelo multinivel es análogo a la correlación entre el termino de error de nivel 2,u_j, y las variables explicativas de nivel 1, X_{ij}. Al igual que en ecuación de datos de panel (3.4.2) se asume que los efectos de nivel 2 o efectos de grupo $\left(u_j\right)$ son una función lineal de los promedio de las variables explicativas de segundo nivel. En este orden de ideas Mundlak propone aproximar $\mathrm{E}\left(\mathrm{u_j}\middle|\mathrm{X_{ij}}\right)$ mediante una regresión lineal basada en la media del clúster $\mathrm{\overline{X}_j}$

$$\mathrm{u_j} = \eta\mathrm{\overline{X}_j} + \mathrm{w_j} \qquad (4.4.3)$$

[28] En este caso el término de efectos aleatorios en datos multinivel es muy diferente al contexto de datos de panel. En el primer enfoque por ejemplo, efectos aleatorios en el intercepto se especifica como: β_{0j}=$\beta_{00} + u_j$ siendo $\beta_{00} = \mu$ y $u_j = \alpha_i$ y siendo la pendiente fijaβ_{1j}=β_{10}

Siendo $\overline{X}_j$ una proxy de la variación "entre" de X_{ij}, es necesario especificar el proceso generador de datos de la variable explicativa, esto con el fin de incorporar esta aproximación lineal de u_j dentro del modelo.

4.4.1 El proceso generador de datos

Se parte del supuesto de que las variables explicativas de nivel 1 X_{ij} difieren a través de las unidades de nivel 1 (estudiantes) dentro de una misma unidad de nivel 2 (escuela), y a través de las unidades de nivel 2 (escuelas). Llamando al primer componente "entre escuelas" y al segundo "al interior de la escuela" se postula que:

$$X_{ij} = X_j^B + X_{ij}^W \qquad (4.4.1.1)$$

Aunque la covariable X_{ij} es observable sus componentes no lo son, razón por la cual deben encontrarse proxy adecuadas para ellos (Hanchane & Mostafa, 2010). Postulando para como proxy para X_j^B :

$$\overline{X}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ij} \qquad (4.4.1.2)$$

Entretanto para el segundo componente X_{ij}^W lo es desviación de la media del clúster o covariable:

$$\widetilde{X}_{ij} = X_{ij} - \overline{X}_j \qquad (4.4.1.3)$$

4.4.2 Modelo "entre escuelas"

Si se desea obtener los efectos "between" de las covariables, se promedia la variable dependiente y las explicativas para cada escuela *j* sobre los *i* estudiantes y se realiza la regresión sobre los resultados en medias:

$$\overline{Y}_{.j} = \beta_{00} + \beta_{10}\overline{X}_{.j} + u_j + \overline{e}_{.j} \tag{4.4.2.1}$$

Siendo $\overline{Y}_{.j}$ el puntaje medio en matemáticas para la escuela j, $\overline{X}_{.j}$ es la media para la covariable ESCS (índice socioeconómico y cultural) y $\overline{e}_{.j}$ es la media de los residuos de nivel 1. En este modelo cualquier información de la variabilidad dentro de las escuelas es eliminada, y los coeficientes de las covariables que no varían entre escuelas son absorbidos por el intercepto (Hesketh & Skrondal 2012). El efecto "entre" β^B puede ser interpretado como la diferencia en el puntaje promedio de matemáticas comparando dos diferentes colegios, en función el promedio del nivel socioeconómico.

4.4.3 Modelo "al interior de la escuela"

Las variables tanto dependiente como explicativa se han centrado alrededor de las respectivas medias de clúster. Las covariables que no varían dentro de los clúster han sido omitidas de la ecuación al igual que el término u_j. El beneficio de esta estimación HLM "al interior" reside en la estimación de β_{1j} proveyendo importantes pistas de como el desempeño escolar de los estudiantes varían dentro de la escuela. Si se quiere obtener los efectos "al interior" de la escuela, se estima el modelo

$$Y_{ij} - \overline{Y}_{.j} = \beta_{10}\left(X_{ij} - \overline{X}_{.j}\right) + \left(e_{ij} - \overline{e}_{.j}\right) \tag{4.4.3.1}$$

La estimación de los efectos "al interior" β^{w} puede obtenerse mediante mínimos cuadrados ordinarios (MCO) cuyos errores estándar son más grandes que las de las estimaciones "entre", pues solamente están basadas en la variabilidad dentro de los clúster.

Una alternativa para la estimación de los efectos "al interior" puede ser obtenido incluyendo un intercepto fijo, α_j para cada escuela mediante la inclusión de una variable dicotómica por cada escuela y excluyendo el intercepto β_{00}[29]. Denotando como d_{jk} la variable dummy para la k escuela (k=1…264), el modelo de efector fijos estimado por MCO puede ser reescrito como:

$$Y_{ij} = \beta_{10}X_{ij} + \sum_{k=1}^{264} d_{kj}\,\alpha_k + e_{ij} \qquad (4.4.3.2)$$

Este modelo con efectos específicos de la escuela reflejados en los coeficientes α_k, deja la explicación de la variación dentro de la escuela a las covariables. En este caso, los coeficientes de nivel 2 no pueden ser estimados. En la práctica esta estimación cae en desuso por la excesiva perdida de grados de libertad consecuencia de la estimación de 264 interceptos.

4.4.4 Relación entre el estimador de efectos "al interior de la escuela", el estimador del modelo "entre escuelas" y el modelo con efectos aleatorios en el intercepto

Similar al estimador del modelo de datos agrupados (MCO combinado), el modelo con intercepto aleatorio puede ser considerado como un promedio de los estimadores "al interior" y "entre".

[29] Previamente el intercepto de cada escuela era representado por $\beta_{00} + u_j$

Existe un escenario en el cual la media del desempeño académico difiere entre escuelas debido a un efecto composición del nivel socioeconómico; por ejemplo dentro de las escuelas, altos niveles socioeconómicos están asociados a altos desempeños académicos lo cual explica completamente por qué la escuela con gran nivel medio socioeconómico tiene una gran media de desempeño académico. Adicionalmente existe un efecto denominado contextual $\beta^{B} - \beta^{w}$, el cual puede deberse a la asignación no aleatoria de estudiantes con alto nivel socioeconómico a las mejores escuelas (las cuales tienen mejor financiación, esto es las privadas) como bien directamente por los "efectos par".

Con el fin de probar el supuesto de exogeneidad de nivel 2 se utiliza la aproximación $E(u_j|X_{ij})$ dada en ecuación (4.4.3) dentro del modelo general:

$$Y_{ij} = \beta_{00} + \beta_{10}X_{ij} + \eta\overline{X}_j + w_j + e_{ij} \qquad (4.4.4.1)$$

Siguiendo la aproximación de Mundlak (1978) se postula:

$$\eta = \beta_{20}^{B} - \beta_{10}^{W} \qquad (4.4.4.2)$$

Hsiao (1999) sostiene que ignorar la diferencia entre los efectos "al interior" y "entre" resulta en una incorrecta especificación del modelo guiando a estimaciones inconsistentes, sesgadas e ineficientes de β_{10}. Se partirá entonces de la ecuación (4.4.4.1) e incluyendo la expresión dada en (4.4.4.2):

$$Y_{ij} = \beta_{00} + \beta_{10}X_{ij} + \left(\beta_{20}^{B} - \beta_{10}^{W}\right)\overline{X}_j + w_j + e_{ij} \qquad (4.4.4.3)$$

Relajando el supuesto de que los efectos "interior" y "entre" son los mismos para una variable en particular, digamos X_{ij} (nivel socioeconómico, ESCS), se tiene el modelo:

$$Y_{ij} = \beta_{00} + \beta_{10}X_{ij} + \beta_{20}^{B}\overline{X}_j - \beta_{10}^{W}\overline{X}_j + w_j + e_{ij} \qquad (4.4.4.4)$$

$$Y_{ij} = \beta_{00} + \beta_{10}^{W}\left(X_{ij} - \overline{X}_j\right) + \beta_{20}^{B}\overline{X}_j + w_j + e_{ij} \qquad (4.4.4.5)$$

La covariable centrada en la media$\left(X_{ij} - \overline{X}_{.j}\right)$ dado que no varía entre clúster no está correlacionada con $u_j$mientras que u_j no varia al interior de los clúster.

Recordando que $X_{ij} - \overline{X}_j$ y $\overline{X}_j$ son las respectivas proxys de X_{ij}^W y X_j^B:

$$Y_{ij} = \beta_0 + \beta_{10}^W X_{ij}^W + \beta_{20}^B X_j^B + e_{ij} \qquad (4.4.4.6)$$

Si $\beta_{10}^W = \beta_{20}^B = \beta_{10}$ luego el modelo (4.4.4.5) se convierte en el modelo con intercepto aleatorio planeado en ecuación (4.4.1). La ecuación (4.4.4.5) también puede verse como el relajamiento del supuesto de que X_{ij} no esta correlacionado con el intercepto aleatorio si se este se define como $\beta^B \overline{X}_{.j} + w_j$ (Hesketh & Skrondal 2012).

Análogo a datos de panel, la aproximación de Mundlak (1978) adaptada a modelos multinivel mantiene que ignorar la diferencia entre los efectos "al interior" y "entre" resulta de una incorrecta especificación del modelo. Las hipótesis sobre la diferencia de efectos son[30]:

$$H_0: \beta_{10}^W - \beta_{20}^B = 0 \qquad (4.4.4.7)$$

$$H_A: \beta_{10}^W - \beta_{20}^B \neq 0$$

4.4.5 La Solución de Mundlak

Para solucionar el problema de endogeneidad de nivel 2 ocasionado por la diferencia entre los estimadores $\hat{\beta}_{10}^W$ y $\hat{\beta}_{20}^B$, Mundlak (1978) propone incluir las medias de todas la variablesde nivel 1. Si existe el nivel 1 de exogeneidad (no correlación entre las variables de nivel 1 y e_{ij}) la

[30]Hesketh y Skrondal (2012) afirman que el realizar un test de sobre para la hipótesis nula de que el coeficiente de la media de la covariable a nivel del clúster sea cero, es asintóticamente equivalente al test de Hausman adaptado para modelos multinivel. La variante del test de Hausman que los autores proponen compara los estimadores FGLS y "within" y aunque pudiera haberse hecho utilizando la comparación de los estimadores "within" y "between", esta opción no está implementada en STATA ni en HLM 7.0

inclusión de las medias de los clúster asegura las consistencia de los estimadores de los efectos "al interior".

Debe anotarse que chequear directamente endogeneidad de nivel 1, esto es chequear si los residuos de nivel 1, e_{ij} esta correlacionado con covariables de nivel 1 o 2 no se posible; requiriendo del uso variables instrumentales, lo que está por fuera de los objetivos de la presente investigación.

4.5 Tamaño Muestral y su relación con el sesgo y la potencia en modelos multinivel

En este apartado se dispone de un análisis del tamaño de la muestra en los distintos niveles dada su relación con el posible sesgo en las estimaciones. Mok (1995) señala que el número de observaciones de segundo nivel tienen un gran impacto en la precisión de las estimaciones[31].

Mediante simulaciones llega a la conclusión de que con una muestras de tamaño pequeño (menor a 800, lo cual puede ocurrir con j=40 e i=20) las estimaciones de los efectos fijos son menos sesgadas cuando el número de grupos se incrementa a expensas de una reducción del número de observaciones dentro del grupo[32] (Stapleton & Tomas, 2008).

En su trabajo Mass & Hox (2005) muestran que aunque los estimadores de los efectos fijos y de los componentes de la varianza son insesgados (bajo condiciones que se consideran apropiadas para el número de grupos y tamaños dentro del grupo), los errores estándar de los coeficientes de nivel 2 tienden a estar sesgados hacia abajo cuando se consideran menos de 30 unidades de segundo nivel. En consonancia con los anteriores trabajos, estos autores encuentran

[31] En la misma vía el autor concluye que las estimaciones de los componentes de la varianza son menos sesgadas con un mayor número de grupos manteniendo el tamaño de la muestra constante

[32] A la misma conclusión llega Snijders (2005).

que para la estimación de los efectos fijos de primer nivel son más importantes los tamaños muéstrales a nivel de grupo que los tamaños muéstrales a nivel individual.

Una manera de explicar el impacto del tamaño en cada nivel es expresar el efecto de diseño en función del coeficiente de correlación intraclase (ICC)

$$EF = 1 + (n - 1) \times ICC \tag{4.5.1}$$

Correspondiendo *n* al tamaño promedio de los clúster, es inmediato que manteniendo el tamaño muestral constante, al aumentar el ICC aumenta el efecto de diseño. Si se tienen grupos homogéneos internamente, en este caso, que los exámenes de los estudiantes sean similares dentro del mismo grupo, en comparación a los puntajes de individuos pertenecientes a otros grupos, luego no se gana mucha información agregando grupos adicionales. En este caso la potencia para detectar efectos fijos (de primer nivel) decrece entre tanto ICC aumente manteniendo el tamaño muestral constante en todos los niveles. Así las cosas, la potencia (1- probabilidad de no rechazar H_0 cuando esta es falsa o 1- error tipo II) para detectar efectos de segundo nivel es mas sensible al número de grupos, en oposición al número de unidades de primer nivel dentro de los grupos. Surge entonces la necesidad de contar con muestras grandes en cada uno de los niveles, sugiriendo que la potencia necesaria para encontrar efectos de interacción cruzados entre niveles requiere un tamaño de al menos 30 unidades en cada uno de los niveles[33] (Stapleton & Tomas, 2008).

---

[33] Todo el análisis del número de grupos necesarios pasa por el hecho de que el número de grupos escogidos sea representativo. Este requisito es cubierto dado que en un diseño de tres etapas como lo es el marco muestral de PISA, se realiza un muestreo estratificado donde las observaciones (grupos) caen dentro de cada estrato a diferencia de un muestro del muestreo por clúster (diseño e dos etapas) en el que no todos los grupos caigan dentro de la muestra y no sean representativos.

5. Estimación del modelo multinivel

5.1 Estimación del Modelo Nulo

La estimación de los modelos utiliza información de los estudiantes ubicados en el nivel ISCED 3, el cual hace parte de la variable ISCEDL la cual clasifica los programas de educación en cinco niveles, siendo el de nivel 3, correspondiente a secundaria superior, esto es grados decimo y once.. Utilizar el índice ISCED permite alcanzar uno de los objetivos del estudio PISA, el cual es comparar los desempeños educativos de estudiantes de distintos países. En efecto ISCED permite hacer comparables la estructura de sistemas educativos los cuales varían ampliamente entre los países, lo cual es requisito para la producción de estadísticas educativas comparables internacionalmente (OECD, 1999).

La estimación del modelo nulo como primer paso, constituye una justificación de la utilización de modelos multinivel donde parte de la varianza de la variable dependiente es explicada por las variables de segundo nivel[34]. Siendo Y_{ij} el desempeño en matemáticas del estudiante i en la escuela j:

Modelo nivel 1:

$$Y_{ij} = \beta_{0j} + e_{ij} \qquad (5.1.1)$$

Modelo nivel 2:

$$\beta_{0j} = \beta_{00} + u_{0j} \qquad (5.1.2)$$

La estimación del modelo nulo multinivel por máxima verosimilitud se consigna a continuación:

[34] Se estima el siguiente modelo utilizando el software HLM 7.0 el cual tiene en cuenta los valores plausibles.

TABLA 1. Modelo nulo del desempeño en matemáticas (PVMATH)

Efectos Fijos				
	Coeficiente	Error st.	T-est.	Prob
Intercepto(β_{00})	406.371906	2.667160	152.062	0.000
Efectos Aleatorios				
	Varianza	G.d.l	Chi-cuadrado	Probabilidad
Variabilidad entre escuelasσ^2_{u0}	1621.77842	263	2915.45128	0.000
Variabilidad dentro de la escuela σ^2_e	3156.90144			
Correlación intraclase	0.339			
Confiabilidad , β_{0j}	0.888			
Desviance[35] (2 parámetros est.)	55276.44			

Fuente: PISA OECD. Estimación con el software HML 7.0

La significancia estadística de la varianza "entre escuelas" indica que el promedio del desempeño académico en matemáticas varía a través de las escuelas lo cual justifica la inclusión de variables de nivel 2. A esta conclusión también se llega a calcular el coeficiente de correlación intraclase:

$$\rho = \frac{\sigma^2_{u0}}{\sigma^2_{u0}+\sigma^2_e} = \frac{(1621.77842)}{(1621.77842 + 3156.90144)} = 33.9 \qquad (5.1.3)$$

[35] Para calcular este coeficiente en HLM 7.0 debe deshabilitarse la opción de valores plausibles. En este caso solo se utiliza el primer valor plausible de la prueba de lenguaje PV1MATH. Véase en Anexos Modelo 1_modelo nulo.

En efecto el 33.9% de la varianza total en el desempeño en matemáticas es atribuible a las escuelas, mientras que 66.1% es atribuible a los estudiantes. Con el fin de disminuir la variabilidad deben ingresarse al modelo covariables de nivel 2 (en el modelo definitivo después de incluir covariables de nivel 2, el coeficiente de intracorrelacion cae a 16.6%)

Recordando que el intercepto β_{0j} indica el promedio del desempeño en matemáticas para las *J* escuelas (j=1,2,…,J) varía a través de las escuelas, el coeficiente de confiabilidad para β_{0j}(acotado en el intervalo 0-1) mide el grado en el cual se puede discriminar el promedio del desempeño académico entre escuelas; la representación de este estadístico es como sigue:

$$\frac{1}{J}\sum_{j=1}^{J}\frac{\sigma_{u0}^2}{\sigma_{u0}^2+\sigma_{00j}^2} \qquad (5.1.4)$$

El coeficiente de confiabilidad mide el grado de variación de las estimaciones MCO de nivel 1 *(i)* a través del conjunto de unidades de nivel 2, *(J).* Un valor bajo para este coeficiente sugeriría que es difícil discriminar entre escuelas sobre la base de su promedio de desempeño académico (X.MA et al 2008). Con un valor de 0.888 el modelo indica que el desempeño en la prueba de matemáticas puede discriminarse entre escuelas.

De otro lado, el coeficiente de desviación estadística equivalente -2 log likehood (-2LL) constituye un criterio de selección de que permite comparar modelos más complejos. En este caso cuando se comparan modelos más complejos que anidan modelos sencillos como el modelo nulo, se utilizan test de diferencias de los valores “desviance” los cuales constituyen test de ratio de verosimilitud.

5.2 Son los efectos "al interior" y "entre"? Probando endogeneidad de nivel dos

A continuación se estudia la posible presencia de endogeneidad de nivel 2, esto es, la correlación del término de error de nivel 2 con covariables de nivel 1. Particularmente en este caso se sospecha que la variable índice de nivel socioeconómico (ESCS) es endógena al estar correlacionada con el término de error u_{0j}. Postulando que la variable omitida de nivel 2 es el "efecto par" (MEANESCS, media del índice ESCS) está relacionada con el índice ESCS, la correlación entre variables de distintos niveles conllevan a una inconsistencia y un sesgo hacia arriba en la estimación del coeficiente de primer nivel, pues al postular que la correlación entre ambas variables es positiva. Para probar endogeneidad se estima la ecuación dada en (4.4.4.1):

$$Y_{ij} = \beta_{00} + \beta_{10}X_{ij} + \eta\bar{X}_j + w_j + e_{ij}$$

Siguiendo la aproximación de Mundlak (1978) se postula:

$$\eta = \beta_{20}^{B} - \beta_{10}^{W} \tag{5.2.1}$$

Siendo η, la diferencia entre las pendientes "al interior" y "entre", la presencia de endogeneidad de nivel 2 se desprende de una incorrecta especificación al omitir la proxy de $\mathrm{X}_{\mathrm{j}}^{\mathrm{B}}$, asumiendo que los efectos "entre" y "al interior" son iguales cuando en realidad son diferentes. Teniendo en cuenta que la diferencia $\mathrm{X}_{\mathrm{ij}} - \bar{\mathrm{X}}_{\mathrm{j}}$ es la proxy de $\mathrm{X}_{\mathrm{ij}}^{\mathrm{W}}$ y que la proxy de $\mathrm{X}_{\mathrm{j}}^{\mathrm{B}}$ es $\bar{\mathrm{X}}_{\mathrm{j}}$, luego la diferencia entre ambas es $\bar{\mathrm{X}}_{\mathrm{j}}$. Postulando al índice socioeconómico y cultural ESCS, como la covariable endógena, se incluyen como proxys de $\mathrm{ESCS}_{\mathrm{ij}}^{\mathrm{W}}$ y $\mathrm{ESCS}_{\mathrm{j}}^{\mathrm{B}}$ a $\mathrm{ESCS}_{\mathrm{ij}} - \overline{\mathrm{ESCS}}_{\mathrm{j}}$ y $\overline{\mathrm{ESCS}}_{\mathrm{j}}$ respectivamente[36].

[36] $ESCS_j^B$, $\overline{ESCS}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} ESCS_{ij} = MEANESCS_j$

Si el coeficiente η es significativamente diferente de cero, luego los efectos "entre" y "al interior" son diferentes, existiendo una incorrecta especificación, existiendo endogeneidad[37]. A continuación se estima el modelo que implementa el test de incorrecta especificación[38]:

$$\mathrm{PVMATH_{ij}}=\beta_{0j} + \beta_{1j}\left(\mathrm{ESCS_{ij}} - \overline{\mathrm{ESCS_j}}\right) + \beta_{1j}(MEANESCS_j - \overline{MEANESC}) + \beta_{3j}GENERO_{ij} + \mathrm{e_{ij}}$$

(4.2.2)

$$\beta_{0j} = \beta_{00} + u_{0j}$$

$$\beta_{1j} = \beta_{10}$$

$$\beta_{2j} = \beta_{20}$$

$$\beta_{3j} = \beta_{30}$$

$$e_{ij} \sim (0, \sigma_e^2)$$

$$u_{0j} \sim (0, \sigma_{u0}^2)$$

[37] El software HLM permite implementar este test al contemplar dentro de sus opciones de estimación el hacerlo con variables centradas en la media general o en la media de grupo.
[38] Véase en Anexos Modelo 2_ Probando endogeneidad de nivel 2 en el modelo

TABLA 2. Modelo con intercepto aleatorio que implementa un test endogeneidad

	Efectos Fijos			
	Coeficiente	Error st.	T-est.	Prob
Intercepto (β_{00})	393.654328	2.067321	187.412	0.000
ESCS*(β_{10})	10.509855	0.867524	12.002	0.000
MEANSESC**(β_{20})	34.543237	2.404379	10.165	0.000
GENERO(β_{30})	36.079328	1.748800	20.631	0.000
	Efectos Aleatorios			
	Varianza	G.d.l	Chi-cuadrado	Probabilidad
Intercepto σ^2_{u0}	27.10816	734.85260	1527.46414	0.000
Variabilidad dentro de la escuela σ^2_e	52.74142	2781.65760		

*Variable Centrada en la media de grupos **Variable Centrada en la media general

Fuente: PISA OECD. Estimación con el software HML 7.0

Siendo $\beta_{20} \neq 0$ $(\eta \neq 0)$ se concluye que los estimadores “within” y “between” son diferentes. Existiendo endogeneidad en el modelo la solución mediante la aproximación de Mundlak (1978) incorpora variables contextuales de la escuela (efecto par).

5.3 Variables del Estudiante

Postulando la orientación vocacional como uno de los principales determinantes del desempeño académico a nivel de escuelas, se incluye como variable de primer nivel dentro del modelo jerárquico. La construcción de esta variable dicotómica se basa en la variable ISCEDO, la cual realiza una clasificación del nivel ISCED 3 (secundaria superior, esto es grados decimos y once) en dos categorías: tipo 1 o general y tipo 3 o vocacional (técnica).

Como ya se mencionó la variable orientación vocacional de naturaleza dicotómica se define de la siguiente manera:

$$VOCACION_{ij} = \begin{cases} 1 & Orientacion\ Tecnica \\ 0 & Orientacion\ Academica \end{cases}$$

Según ISCED, la orientación general (o académica) no está diseñada explícitamente para preparar a los estudiantes para una ocupación o para entrar a posteriores programas técnicos. Entretanto, la educación vocacional (o técnica) prepara a los estudiantes para entrar directamente sin entrenamiento adicional, a ocupaciones específicas (OECD, 1999).

Como puede observarse en el grafico 1, entre mayor sea el índice socioeconómico mayor es el desempeño en la prueba de matemáticas, advirtiendo que las escuelas con mayor puntaje en la prueba son aquellas con orientación académica.

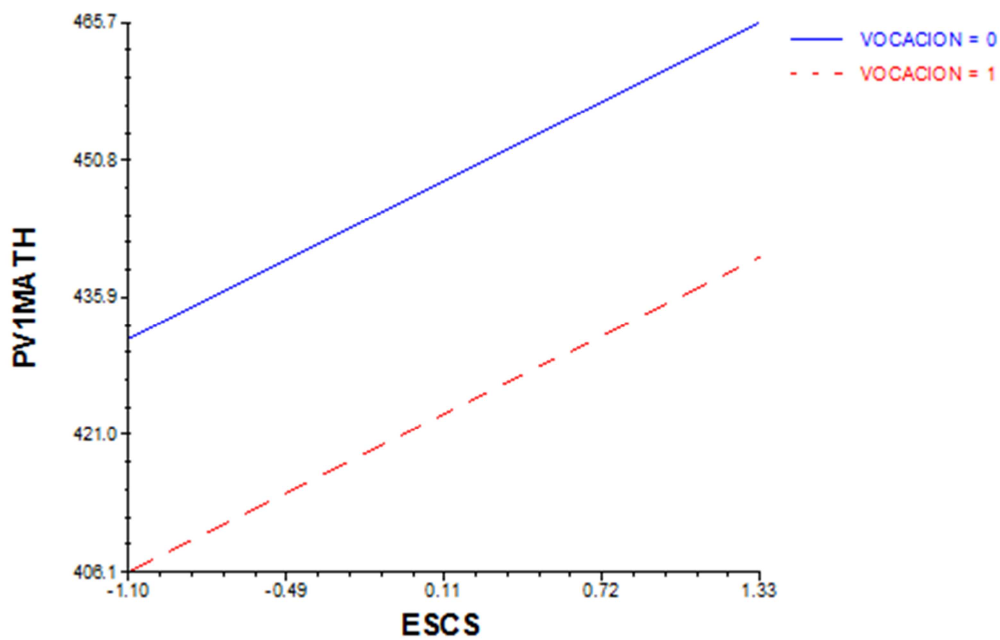

Figura No 1: Relación entre el índice ESCS y el puntaje de matemáticas moderando por la orientación vocacional Estimación con el software HML 7.0

Se postulan como covariables adicionales de nivel 1, el género y el índice socioeconómico ESCS[39].Tradicionalmente se sostiene que el rendimiento del estudiante tiene una relación directa con los recursos económicos y posición social de la familia. Aunque estas variables no son controlables directamente mediante políticas educativas, si pueden moderarse ofreciendo igualdad de oportunidades a los estudiantes, como por ejemplo mediante la asignación de pares mediantes el sistema de "vouchers". Debe anotarse que PISA construye tres índices que son usados para medir el ambiente en el hogar. El primero de ellos es un índice de recursos del hogar para la educación (HEDRES); derivado del cuestionario de los estudiantes, refleja la

[39]Posteriormente se modelan efectos par y no linealidades asociadas al índice socioeconómico del estudiante incluyendo en el primer caso el promedio de la variable (MEANESCS) y en el segundo su varianza (VARESCS) de manera que permita modelar el impacto de la diversidad social dentro de la escuela.

disponibilidad del número de diccionarios, un lugar adecuado para estudiar, un escritorio de estudio, libros de texto, diccionario y calculadoras.

El segundo índice se refiere a las posesiones culturales (CULTPOSS) que se elabora a partir de la información relativa a la disponibilidad de literatura clásica, libros de poesía y obras de arte[40]. Este índice constituiría una medida del interés de los padres en los niños, o directamente en temas de educación o interés en general. Un último índice se refiere a las posesiones del hogar (HOMEPOSS). En la especificación del modelo se incluye solo el índice socioeconómico ESCS, al integrar los anteriores tres índices.

Dado que se argumenta que en las pruebas de matemáticas, el mejor desempeño está a favor de los hombres en relación a las mujeres, se incluye el género como variable de nivel 1. A continuación se observa la anterior regularidad empírica. Denotando la variable dicotómica género:

$$\mathrm{GENERO_{ij}} = \begin{cases} 1 & \text{Hombre} \\ 0 & \text{Mujer} \end{cases}$$

[40]Un índice de actividad cultural también está disponible. Es un índice de cuan a menudo los estudiantes visitan un museo o una galería de arte en el año precedente a la prueba, atendieron a una ópera, ballet, concierto de música clásica o una obra de teatro.

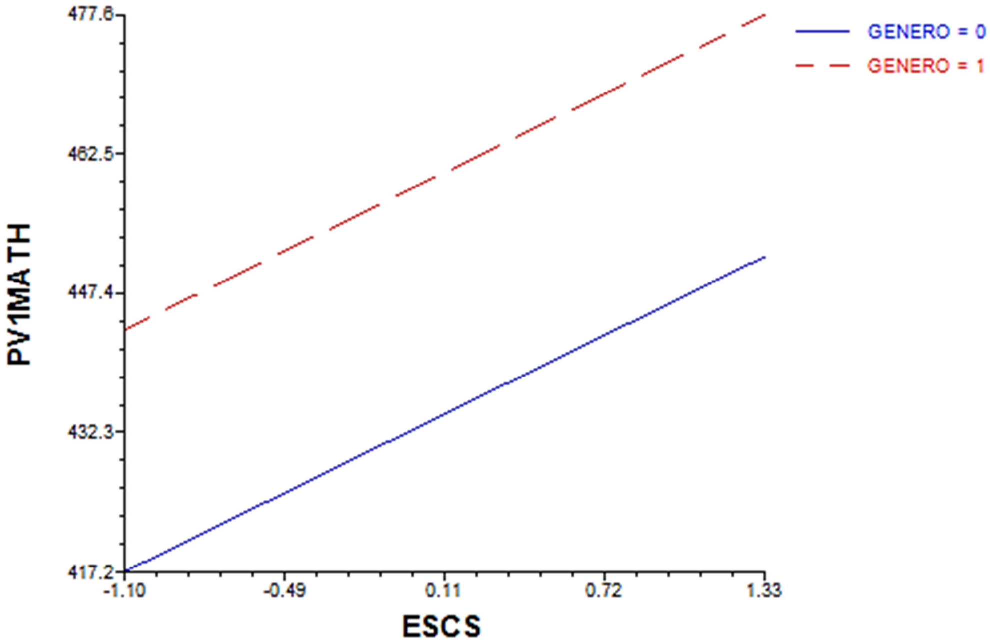

Figura No 2: Relación entre el índice ESCS y el puntaje de matemáticas moderando por género (PVMATH) Estimación con el software HML 7.0

Se observa un mayor puntaje promedio (dado por el intercepto del modelo) a nivel de escuelas para los hombres respecto a las mujeres

La justificación de ingresar los efectos par en este caso la media del índice socioeconómico ESCS, $MEANESCS_j$ dentro de las variables de nivel 1,va de la mano con la aproximación de Mundlak (1978) la cual intenta eliminar el sesgo tanto como sea posible introduciendo en el modelo de efectos aleatorios dado en ecuación (4.4.4.1) las medias de las covariables de nivel que se sospechan endógenas. Si los "efectos par" resultan significativos luego el proceso de aprendizaje en la escuela depende de cómo los efectos contextuales se combinan con las variables de nivel 2.

5.4 Variables de la escuela

La presencia de una estructura jerárquica en los datos produce una heterogeneidad no observada, en el que la media de la variable dependiente varía a través de los clúster. Con el fin de modelar esta regularidad, variables de nivel 2 deben ser incluidas pues la variación en el desempeño en matemáticas no puede atribuirse solo a variables del estudiante.

La necesidad de explicar las diferencias en los puntajes discriminando en los niveles individuales (estudiantes) y grupales (escuela) justifica la utilización de la técnica de análisis multinivel. Como bien se examinará posteriormente, las pendientes del modelo resultaron fijas siendo el intercepto el único parámetro con variación aleatoria, lo cual abre la opción de incluir variables de nivel de la escuela para su modelación.

Se postulan dos componentes que explican el promedio a nivel de la escuela. De un lado se incluye la variable tipo de escuela (TIPOESC) esto es si es pública o privada:

$$\mathrm{TIPOESCS_j} = \begin{cases} 1 & \text{Publica} \\ 0 & \text{Privada} \end{cases}$$

La otra variable se postula afecta el rendimiento promedio en la escuela *j* es el efecto contextual cuya proxy es el promedio del nivel socioeconómico de la escuela (MEANESC).

Se modela el coeficiente que acompaña al efecto par en el nivel 1 de manera no aleatoria al colocar la función de la variable de segundo nivel (TIPOESC). Esto responde ala hipótesis de trabajo de que el efecto contextual (MEANESC) esta moderado por el tipo de escuela; el “efecto par” tendría un impacto diferenciado dependiendo si se trata de una escuela pública o privada. A continuación el siguiente grafico da cuenta de esta diferencia.

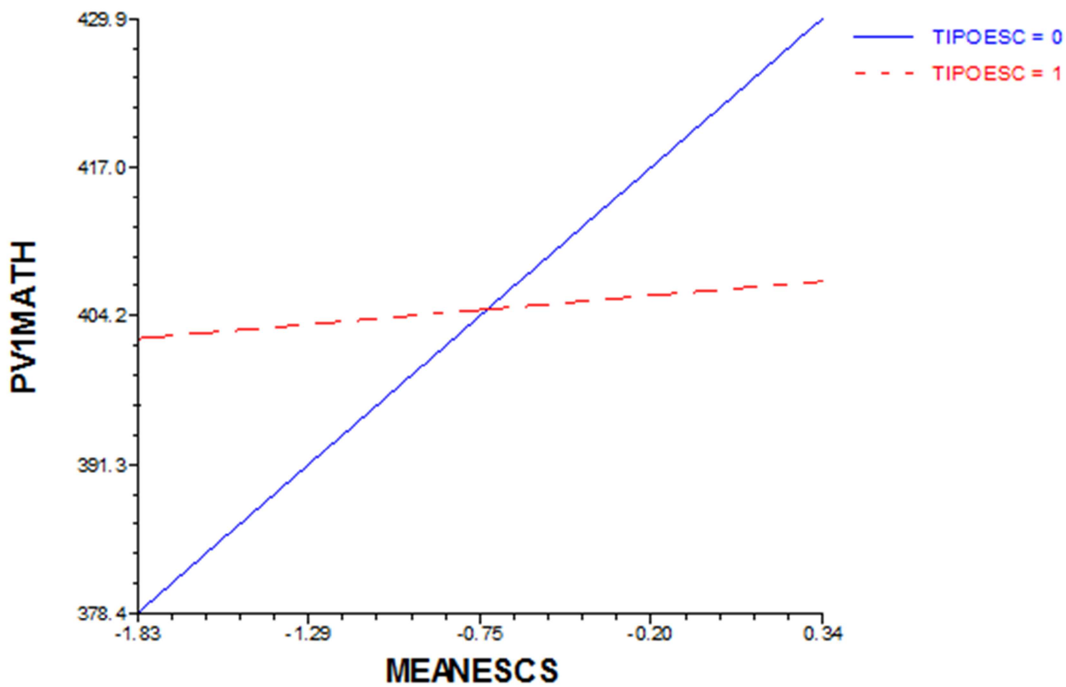

Figura No 3: Relación entre el efecto contextual o "efecto par" y el puntaje de matemáticas moderando por el tipo de escuela (TIPOESC). Estimación con el software HML 7.0

Se observa una pendiente más pequeña para las escuelas públicas, indicando que el "efecto par" tiene un impacto relativamente pequeño sobre el desempeño académico en matemáticas respecto a una escuela privada.

La inclusión de la variable MEANESCS en el nivel 2 para modelar intercepto es intuitivo, toda vez que las brechas socioeconómicas entre las escuelas tienen un impacto sobre el promedio del desempeño académico[41].La primera recomendación antes de especificar el modelo definitivo es analizar si el efecto individual varía por tipo de escuelas, es decir si realmente existen efectos aleatorios (Hesketh & Skrondal, 2012).

[41] Variables contextuales como la media del índice socioeconómico (MEANESCS) y la varianza de este índice (VARESCS) no están disponibles en la base inicial de PISA; en vez de ello se obtienen agregando a nivel desde el archivo escuela, para luego realizar un "merge" con el archivo de estudiantes utilizando variables clave como el código del estudiante y de la escuela.

5.5 Modelo Definitivo: Pendientes de nivel 1 con variación aleatoria o no aleatoria?

La existencia de los efectos aleatorios se debe a que los estimadores de nivel 1 contendrán el verdadero efecto causal a nivel individual adicional a una fracción del efecto del determinante a nivel de grupo. De ser la variación aleatoria significativa, una covariable de nivel 2 puede ser ingresada para modelar los efectos individuales, no obstante las pendientes de nivel 1 pueden tener variación no aleatoria. Si la variación es no aleatoria se incluirán variables de nivel 2 mas no efectos aleatorios u_{0j}omitiendo los términos de error u_{pj}.

EL modelo con efectos aleatorios en las pendientes:

$$\mathrm{PVMATH}_{ij} = \beta_{0j} + \beta_{1j}\mathrm{ESCS}_{ij} + \beta_{2j}\mathrm{MEANESCS}_{ij} + \beta_{3j}\mathrm{GENERO}_{ij} + \beta_{4j}\mathrm{VOCACIONAL}_{ij} + e_{ij}$$

(5.5.1)

$$\beta_{0j} = \beta_{00}$$

$$\beta_{1j} = \beta_{10} + u_{1j}$$

$$\beta_{2j} = \beta_{20} + u_{2j}$$

$$\beta_{3j} = \beta_{30} + u_{3j}$$

$$\beta_{4j} = \beta_{40} + u_{4j}$$

$$e_{ij} \sim (0, \sigma_e^2)$$

$$u_{0j} \sim (0, \sigma_{u0}^2)$$

Se presentan los resultados de la estimación con efectos aleatorios en cada una de las pendientes del modelo (véase en Anexos Modelo HLM Covariables nivel 1 con efectos aleatorios).

TABLA 3. Modelo con efectos aleatorios

	Efectos Aleatorios[42]			
	Varianza	G.d.l	Chi-cuadrado	Probabilidad
ESCSσ^2_{u1}	9.96164	28	28.17212	*0.456*
MEANESCS σ^2_{u2}	580.05781	28	58.50048	*0.001*
GENEROσ^2_{u3}	205.97541	28	20.06453	*>.500*
VOCACIONσ^2_{u4}	387.41986	28	42.07360	*0.042*
Desviance (16parametrosest)	54464.4376			

Fuente: PISA OECD. Estimación con el software HML 7.0

Los niveles de probabilidad de los efectos aleatorios resultaron no son significativos para las variables ESCS y GENERO, contrario a lo que pasa para las variables MEANESCS y VOCACION cuyas variaciones aleatorias son significativas al 5%. Estos resultados proporcionarían el primer indicio sobre la existencia de impactos diferenciados del "efecto par" y de la orientación vocacional, postulando como variable explicativa en ambos casos el tipo de escuela.

La estimación del modelo jerárquico definitivo, arrojo que las variables MEANESCS y VOCACION, al descomponer su variación en aleatoria y no aleatoria (en función de TIPOESC), resulta significativa la última mas no la primera; es decir la variación aleatoria de estos coeficientes deja de ser significativa cuando se introducen variables de la escuela para modelarlos.

[42] Estos resultados se obtuvieron utilizando la metodología PISA con los cinco valores plausibles. Para el cálculo del coeficiente de "desviance" se estimó el modelo utilizando full máxima verosimilitud con el objetivo de realizar un test que sirva como criterio de selección de modelos (posteriormente se explicara la metodología de este test).Véase en Anexos Modelo 3_Modelo con efectos aleatorios y valores plausibles.

La especificación final del modelo multinivel:

$$\mathrm{PVMATH_{ij}} = \beta_{0j} + \beta_{1j}\mathrm{ESCS_{ij}} + \beta_{2j}\mathrm{MEANESCS_{ij}} + \beta_{3j}\mathrm{GENERO_{ij}} + \beta_{4j}\mathrm{VOCACIONAL_{ij}} + e_{ij}$$

(5.5.2)

$$\beta_{0j} = \beta_{00} + \beta_{01}MEANESCS_j + \beta_{02}\mathrm{TIPOESC} + u_{0j}$$

$$\beta_{1j} = \beta_{10}$$

$$\beta_{2j} = \beta_{20} + \beta_{21}\mathrm{TIPOESC}$$

$$\beta_{3j} = \beta_{30}$$

$$\beta_{4j} = \beta_{40} + \beta_{41}\mathrm{TIPOESC}$$

$$e_{ij} \sim (0, \sigma_e^2)$$

$$u_{0j} \sim (0, \sigma_{u0}^2)$$

Cada unidad de análisis se interpreta en función del nivel pudiendo ser un estudiante y una escuela. De este modo se tienen estos datos de partida 5021 estudiantes (grados decimo y once) y 264 escuelas o unidades de nivel 2.

El modelo mixto queda especificado a continuación:

$$\begin{aligned}\mathrm{PVMATH_{ij}} &\\ &= \beta_{00} + \beta_{01}\mathrm{MEANESCS}_j + \beta_{02}\mathrm{TIPOESCS}_j + \beta_{10}\mathrm{ESCS_{ij}} \qquad (5.5.3)\\ &+ \beta_{20}\mathrm{MEANESCS_{ij}} + \beta_{21}\mathrm{MEANESCS_{ij}} * \mathrm{TIPOESCS_j} + \beta_{30}\mathrm{GENERO_{ij}}\\ &+ \beta_{40}\mathrm{VOCACIONAL_{ij}} + \beta_{41}\mathrm{VOCACIONAL_{ij}} * \mathrm{TIPOESCS_j} + u_{0j} + e_{ij}\end{aligned}$$

Se escoge esta presentación pues es más ilustrativa al incluirlos efectos fijos y aleatorios sin pasar por coeficientes intermedios de cada nivel. A continuación se interpretan los coeficientes del modelo finalmente escogido para su estimación:

β_{00}: Media del puntaje de la prueba de matemáticas para todas las escuelas

σ_e^2: Varianza dentro de la escuela

e_{ij}: Error asociado con cada estudiante

β_{10}: Diferencial de la media del puntaje de la escuela pública respecto a una escuela privada

β_{02}: Cambio en el puntaje promedio de matemáticas por un cambio de una unidad en el promedio del índice ESCS

β_{20}: Cambio en el puntaje en matemáticas ante un cambio en el "efecto par" del estudiante.

β_{21}: Efecto diferencial del "efecto par" de una escuela pública respecto a una escuela privada.

β_{30}: Diferencial en el puntaje de matemáticas de un estudiante hombre respecto a un estudiante mujer.

β_{40} y β_{41} permiten calcular el promedio del puntaje de matemáticas para las diferentes combinaciones de orientación vocacional con tipo de escuela, privada o publica.

u_{0j}: efecto asociado con cada cluster.

σ_{u0}^2: es la la varianza entre escuelas

Visto como un modelo de regresión con un intercepto especifico a la escuela $\beta_{00} + u_{0J}$, el intercepto aleatorio u_{0J} puede ser considerado como una variable latente que no es estimada a través de los parámetros fijos β_{01} hasta β_{40} pero cuya varianza σ_{u0}^2 es estimada junto con la varianza de los residuos de primer nivel, σ_e^2. Este modelo lineal con intercepto aleatorio con covariables constituye un ejemplo de un modelo lineal mixto con efectos fijos y aleatorios (Hesketh & Skrondal 2012).

5.6 Reportando resultados del análisis multinivel

A continuación se consignan los resultados del modelo jerárquico de dos niveles por restringida máxima verosimilitud, teniendo en cuenta los cinco valores plausibles de la prueba.

TABLA 4. Modelo final que incluye variables del estudiante y de la escuela con efectos aleatorios y no aleatorios[43]

Efectos Fijos				
	Coeficiente	Error st*.	T-est.	Prob
Intercepto (β_{00})	436.284007	3.954395	110.329	0.000
MEANESCS (β_{01})	15.852550	7.470371	2.122	0.001
TIPOESC$_j$ (β_{02})	-22.395684	5.976559	-3.747	0.000
ESCS(β_{10})	10.525716	0.899390	11.703	0.000
MEANESCS (β_{20})	24.008589	8.354902	2.874	0.000
MEANESCS*TIPOESCS (β_{21})	-21.222611	5.738215	-3.698	0.001
GENERO (β_{30})	36.179268	1.704301	21.228	0.000
VOCACION (β_{40})	-21.467914	9.506690	-2.258	0.024
VOCACION*TIPOESC (β_{41})	26.213557	10.144764	2.584	0.010
Efectos Aleatorios				
	Varianza	G.d.l	Chi-cuadrado	Probabilidad
Variabilidad entre escuelas σ^2_{u0}	549.71746	261	1187.07520	0.000
Variabilidad dentro de la	2761.16568			

[43] Véase en Anexos Modelo 4_Modelo Final.

escuela σ_e^2	
Correlación intraclase	0.166
Confiabilidad , β_{0j}	0.765
Desviance *(11 parámetros est)*	54371.326870

Fuente: PISA OECD. Estimación con el software HML 7.0. *Errores estándar robustos.

Después de controlar por variables del estudiante y de la escuela se encuentra una correlación intraclase de 0.166 esto es, las escuelas explican la variación del desempeño en matemáticas en un 16.6% (respecto al 33.9% del modelo nulo).

Se encontró que el nivel socioeconómico del estudiante resulta significativo en la explicación del desempeño educativo. Con signo positivo indica $\hat{\beta}_{10}$ que a mayor nivel socioeconómico del estudiante, mayor puntaje obtiene en la prueba de matemáticas. De otro lado la presencia de la media de la variable ESCS, esto es MEANESCS es justificada metodológicamente atendiendo la aproximación de Mundlak (1978). Significativo y con signo positivo, el papel "efecto par" dentro del proceso de aprendizaje, arroja que a mayor nivel socioeconómico de los "pares" mayor es el desempeño académico individual.

El "efecto par" diferenciado, con coeficientes de 24.01$(\hat{\beta}_{20})$ y 2.79 $(\hat{\beta}_{20} + \hat{\beta}_{21})$ para las escuelas privadas y públicas respectivamente, confirma la hipótesis de que el "efecto par" tiene un significativo impacto en las primeras y un efecto imperceptible en las segundas. Se concluye entonces la existencia del primer efecto diferenciador entre las escuelas públicas y privadas más allá de una simple comparación promedio de puntajes, a saber que los efectos contextuales o también denominados "efecto par" tienen un comportamiento disímil entre ambas escuelas.

Dos de los resultados esperados para este tipo de trabajos se confirman; el primero, las escuelas públicas tienen menor puntaje promedio que las escuelas privadas $(\hat{\beta}_{01} < 0)$, y el segundo, los estudiantes hombres tienen mejor rendimientos en estas pruebas respecto a sus “pares” mujeres con un diferencial de 36 puntos.

El segundo efecto diferenciador que se postula entre ambas escuelas es el impacto del tipo de orientación vocacional, técnica o académica, por tipo de escuela. Resulta adecuado modelar desempeño académico promedio por tipo de escuela, discriminando por la orientación vocacional del estudiante. Empezando con las escuelas privadas, el desempeño promedio para los estudiantes con orientación vocacional académica es 436.13 puntos(β_{00}) mientras que para los estudiantes con orientación vocacional técnica el puntaje es de 414.81puntos $(\beta_{00} + \beta_{40})$.

Se observa una diferencia importante en el desempeño académico dentro de las escuelas privadas según la orientación vocacional del estudiante, no ocurriendo así en las escuelas públicas. En efecto, el desempeño promedio del estudiante de una escuela pública con orientación académica es de 413.88 $\left(\hat{\beta}_{00} + \hat{\beta}_{02}\right)$ mientras que para un estudiante con orientación vocacional dentro de una escuela pública es de 418.62 $\left(\hat{\beta}_{00} + \hat{\beta}_{02} + \hat{\beta}_{40} + \hat{\beta}_{41}\right)$.

6. Test de correcta especificación del modelo

6.1 Seleccionando el número de niveles

Opdenakker & Van Damme (2001) encuentran que ignorar niveles en este tipo de análisis tiene consecuencias importantes en las estimaciones no solo de los efectos fijos, sino también en la estimación de los componentes de la varianzas. Al respecto un modelo de tres niveles con las ciudades como variable de tercer nivel fue estimado[44]. Inicialmente la estimación de un modelo nulo permite descomponer la varianza del puntaje en tres niveles con significancia estadística incluso al 1%. Sin embargo al incluir variables de nivel del estudiante, por ejemplo nivel socioeconómico (ESCS) y efecto contextual (MEANESCS) se concluye como no significativa la participación de un tercer nivel en el análisis del desempeño académico en favor de un modelo jerárquico de dos niveles: estudiantes y escuela.

Cada unidad de análisis se interpreta en función del nivel, especificando una ecuación para el estudiante, la escuela y la región. En un primer nivel la ecuación para el estudiante se especifica:

$$\mathrm{PVMATH}_{\mathrm{ijk}} = \beta_{0\mathrm{jk}} + \beta_{1\mathrm{jk}}\mathrm{ESCS}_{ijk} + \beta_{2\mathrm{jk}}\mathrm{MEANESCS}_{ijk} + \mathrm{e}_{\mathrm{ijk}} \qquad (6.1.1)$$

Dónde:

$\mathrm{PVMATH}_{\mathrm{ijk}}$: Es el desempeño en Matemáticas de un estudiante i en la escuela j de la ciudad k.

$\beta_{0\mathrm{jk}}$: Es la media de puntuaciones en Matemáticas de la escuela j de la ciudad k.

$\beta_{1\mathrm{jk}}$: Es el efecto del nivel socioeconómico del estudiante sobre el desempeño académico en la escuela j en la ciudad k (se asume como efecto fijo= β_{100})

[44] Para este modelo se tomó toda la muestra de estudiantes sin discriminar por tipo de escuela ni grados. Para el nivel 3 se tienen 8 ciudades dentro de la muestra tomada por PISA.

β_{2jk}: Es el impacto del “efecto par” sobre el desempeño académico en la escuela j en la ciudad k (se asume como efecto fijo= β_{200})

e_{ijk}: Es la diferencia que tiene el puntaje de cada estudiante respecto al promedio de la escuela a la que pertenece.

En este modelo el desempeño de cada estudiante se obtiene sumando el efecto aleatorio individual al puntaje promedio de su respectiva escuela. Esta desviación se distribuye normal con media cero y varianza σ_e^2. Para el nivel 2 el promedio del desempeño académico de la escuela puede representarse por la siguiente ecuación:

$$\beta_{0jk} = \beta_{00k} + \beta_{01k}\mathrm{TIPOESC} + u_{0jk} \qquad (6.1.2)$$

Dónde:

β_{00k}:El desempeño académico promedio en la ciudad k

β_{01k}: El impacto sobre el desempeño académico de la escuela pública respecto a la privada en la ciudad k.

u_{0jk}:Es la desviación que tiene cada escuela respecto al promedio de la ciudad.

Los efectos aleatorios del segundo nivel se consideran con media cero y una varianza σ_{u0}^2. El desempeño académico de cada escuela se obtiene al sumar el efecto aleatorio más el desempeño académico promedio de todas las escuelas de la ciudad. Finalmente en el tercer nivel, la variabilidad entre ciudades se representa mediante la siguiente ecuación[45]:

$$\beta_{00k} = \beta_{000} + u_{00k} \qquad (6.1.3)$$

$$\beta_{01k} = \beta_{010}$$

[45] Se considera β_{1jk} un efecto fijo

β_{000}: Es el desempeño promedio general de toda la muestra de estudiantes

u_{00k}: Es el efecto aleatorio o desviación del promedio académico de cada ciudad respecto del promedio académico general. Estos efectos aleatorios se distribuyen normal con media 0 y varianza σ^2_{u0k}.

TABLA 5. Modelo con tres niveles incluyendo solo un predictor a nivel individual

	Efectos Fijos			
	Coeficiente	Error st.	T-est.	Prob
Intercepto (β_{000})	460.342900	2.241752	205.350	0.000
TIPOESC(β_{010})	-20.370957	7.281919	-2.797	0.006
ESCS (β_{100})	11.680120	0.952948	12.257	0.000
MEANESCS (β_{200})	22.756289	5.881207	3.869	0.000
	Efectos Aleatorios			
	Varianza	G.d.l	Chi-cuadrado	Probabilidad
Variabilidad dentro de la escuela σ^2_e	3024.69374			
Variabilidad entre escuelas σ^2_{u0}	613.06761	109	540.97147	0.000
Variabilidad en el tercer nivel σ^2_{u0k}	0.30397	7	6.86251	>.500
Confiabilidad , β_{0jk}	0.773			

Fuente: PISA OECD. Estimación con el software HML 7.0

Se observa que la varianza del efecto aleatorio u_{00k} no es significativa, esto es no existe variabilidad de tercer nivel[46].

6.2 Evaluando el ajuste del modelo elegido: Test de diferencias chi cuadrado

El procedimiento de máxima verosimilitud además de arrojar las estimaciones de los parámetros también provee el coeficiente una desviación estadística que permite comparar el log de máxima verosimilitud de un modelo anidado contra uno saturado. Este coeficiente puede asimilarse a una medida de ajuste del modelo, al describir que tan mal desempeño tiene el modelo estimado respecto al mejor modelo posible.

Como ya se mencionó los modelos fueron estimados mediante máxima verosimilitud restringida (REML) cuya reconocida ventaja es la insesgadez de los componentes de la varianza. No obstante cuando se trata de comparar el ajuste de modelos anidados, el método REML solo puede ser usado para comparar ajustes de modelos si estos tienen el mismo número de efectos fijos y solo difieren en el número de efectos aleatorios. Entretanto el método de full máxima verosimilitud (FILM) no tiene esta restricción al permitir comparar modelos con diferente número de efectos fijos (D.B Maccoach & Black, 2008).

En este orden de ideas se estiman los modelos mediante FIML con el fin de utilizar los coeficientes de desviación estadística y construir así un test que permita comparar grados de ajuste entre modelos.

Partiendo de la función de verosimilitud concentrada que se obtiene al tener en cuenta los estimadores de máxima verosimilitud (EMV):

[46] Véase en Anexos, Modelo 5_Modelo con tres niveles

$$L_c^* = \log L = -\frac{n}{2}\left(1 + \log(2\pi)\right) - \frac{n}{2}\log\left(\frac{SCR}{n}\right) \qquad (6.2.1)$$

Denotado por LL, y multiplicado por -2 se obtiene la desviación estadística (desviance) del modelo multinivel equivalente a -2LL

$$D = n\left(1 + \log(2\pi)\right) + n\log\left(\frac{SCR}{n}\right) \qquad (6.2.2)$$

Denotando como D_2 al coeficiente del modelo más complejo con variables en los dos niveles (ver tabla 5) y como D_1 el correspondiente al modelo con solo covariables de nivel 1la diferencia $\Delta D = D_1 - D_2$puede esperarse positiva; en efecto el modelo más simple tendrá al menos tan alta varianza como el más complejo y a su vez este último generalmente mostrará una varianza más baja que el modelo simple.

$$\Delta D = D_1 - D_2 \qquad (6.2.3)$$

$$\Delta D = 54464.43 - 54371.32 = 93.11$$

En grandes muestras, las diferencia entre ambas desviaciones se distribuirá ji-cuadrado con grados de libertad igual a la diferencia de parámetros entre los dos modelos estimados[47] (De Leeuw, 2008). Si el modelo con mayor número de parámetros no logra disminuir la varianza respecto al modelo más parsimonioso luego la elección es este último.

Las hipótesis del test se especifican a continuación:

$$H_0\text{: Modelo parsimonioso}$$

$$H_0\text{: Modelo actual}$$

[47]Véase en Anexos Modelo 6_Modelo con variables del estudiante y efectos aleatorios y Modelo 7_Modelo Final Full MV sin plausibles.

Los resultados del test se consignan en la siguiente tabla[48]:

TABLA 6. Test de comparación de modelos

Chi-cuadrado	93.11082
Grados de libertad	5
Valor p	0.001

Fuente: PISA OECD. Estimación con el software HML 7.0

Una disminución del coeficiente de desviación estadística en 93.11 reflejaría un mejor ajuste con un modelo con covariables en ambos niveles respecto al modelo con solo variables de nivel 1.

Una crítica a la anterior técnica de selección de modelos es que en grandes muestras la hipótesis nula es rechazada, produciendo modelos más complejos (Weaklim, 2004). La utilización de criterios de selección de modelos Akaike y el criterio Bayesiano, los cuales tienen como base el coeficiente de desviación estadística, convergerán hacia la elección del mismo modelo[49] (D.B Maccoach & Black, 2008).

[48] Véase en anexos Modelo 7_Modelo Final Full MV sin plausibles.
[49] Akaike y el criterio bayesiano propone escoger el modelo con el menor valor para cada criterio, cuyas formas respectivamente son: AIC=$\mathrm{D}+2\mathrm{p}$ y BIC=$\mathrm{D}+\ln(\mathrm{n})*\mathrm{p}$, siendo p el numero de parámetros. Este último criterio, además de penalizar el ingreso de parámetros al modelo también tiene en cuenta el impacto del tamaño de la muestra. No obstante no existe un consenso en la literatura sobre si n corresponde a las unidades de primer nivel, de segundo nivel o al total ($\mathrm{n_j} \times \mathrm{J}$).

6.3 Inferencia estadística respecto a los componentes de la varianza

Para realizar inferencia estadística respecto a los componentes de la varianza se utiliza el test de ratio de máxima verosimilitud el cual compara el ajuste de los dos modelos (Snijders & Bosker, 2012). El primer modelo no restringe ninguno de los componentes de varianza, produciendo una función de verosimilitud L_1, entretanto el segundo modelo restringe el componente de la varianza a cero, produciendo una verosimilitud de L_0. Con una distribución asintótica ji-cuadrado con 1 grado de liberta, el estadígrafo $2[\ln(L_1) - \ln(L_0)]$ indica que los componentes de la varianza son significativos al 5% (véase Tabla 4).

6.4 Testeando los supuestos en el modelo multinivel

6.4.1 Evaluando el supuesto de homocedasticidad de nivel 1.

Los supuestos subyacentes en el modelo multinivel son similares a los de un modelo de regresión que ignora estructuras jerárquicas: normalidad y distribución normal de los errores. Es conocido en la econometría tradicional, que la violación del supuesto de homocedasticidad (en este caso en el nivel 1) afecta la eficiencia de los estimadores de nivel 1 (Snijders & Bosker, 2012).

Debe anotarse que aunque en modelos multinivel, el supuesto de homocedasticidad se refiere a todos los niveles, se hace énfasis en la homocedasticidad de nivel 1 dado el interés en obtener estimadores eficientes de los efectos fijos. Constituyendo una medida de incorrecta especificación del modelo, la heterocedasticidad surge cuando variables de nivel 1 siendo omitidas, están distribuidas con varianza no constante a través de las unidades de nivel 2[50] (Raudenbush & Brick, 1999).

TABLA 7. Testeando homocedasticidad en los residuos de nivel 1[51]

Chi-cuadrado	266.75
Grados de libertad	261
Valor p	0.390

Estimación con el software HML 7.0

Examinando el nivel de probabilidad del test de homocedasticidad, se concluye la variabilidad del entre las 264 unidades de nivel 2 (escuelas) en términos de los residuales de nivel 1, "dentro de la escuela" es constante obteniendo estimadores eficientes.

6.4.2 Normalidad de residuos de nivel 1

Respecto al supuesto de normalidad en los errores de nivel 1, debe anotarse que el software HLM 7.0 permite guardar los residuos de la estimación en los dos niveles. Comenzando con el nivel 1, una primera aproximación es el grafico para este residuo de la densidad de kernel y compararlo con una función de densidad normal[52].

[50]Siento esta la causa más probable de Heterocedasticidad, otra forma de incorrecta especificación seria considerar un efecto fijo cuando en realidad constituye un efecto aleatorio

[51]Véase en Anexos Modelo 8_Test de Homocedasticidad Nivel 1

[52] HLM 7.0 permite entre sus opciones guardar los residuos de nivel 1 y nivel 2 en diferentes formatos como SPSS y STATA. Esto facilita realizar test sobre estos términos de error. Véase en Anexos Graph1.

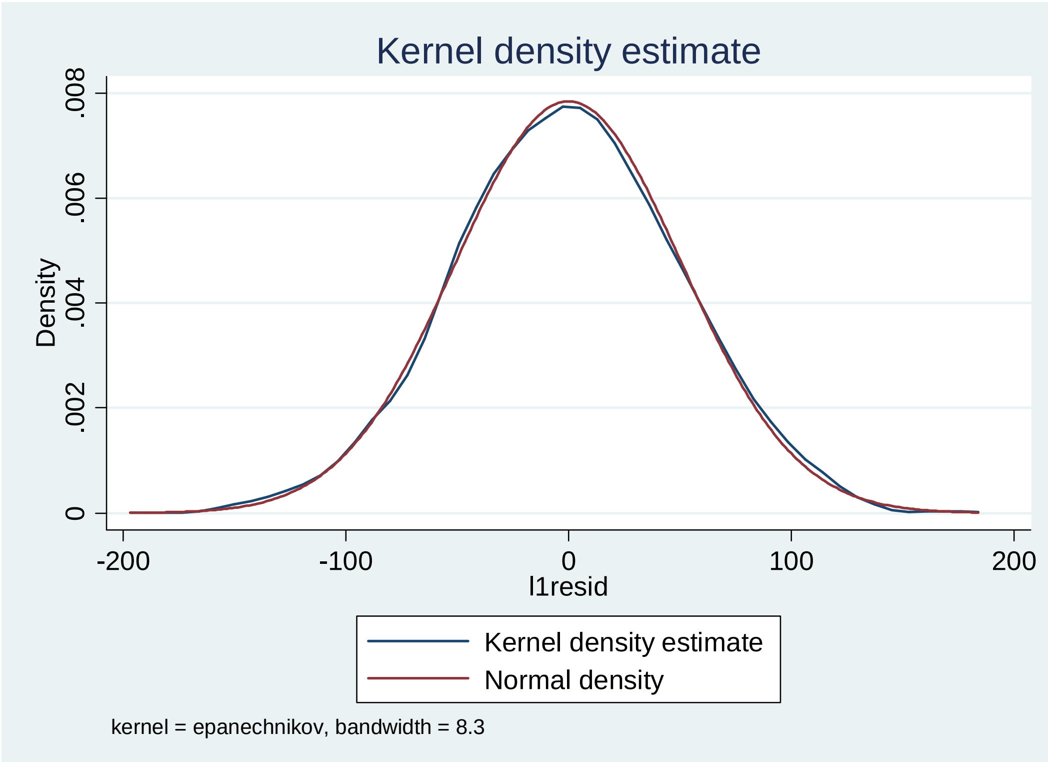

Figura No 4: Función de densidad de kernel para los residuos de nivel 1. Estimación con el software STATA 11.0

Se observa un buen ajuste del grafico de Kernel a la función de densidad normal, lo cual constituye un indicio de normalidad del error. Un test formal confirma este resultado[53]:

TABLA 8. Testeando normalidad en los residuos de nivel 1

Coeficientes	Valor p
Chi-cuadrado	0.6348
Asimetría	0.463
Kurtosis	0.544

Estimación con el software STATA 11.0

[53] Véase el archivo residuosnivel1.dta y en anexos test de normalidad residuos de nivel 1 . Los test de normalidad en STATA puede obtenerse con los comandos sktest y la función de densidad de kernel, kdensity.

En efecto, con un nivel de probabilidad de 0.458 la hipótesis nula de normalidad no es rechazada.

6.4.3 Normalidad en los residuos de nivel 2: Utilizando el estimador empírico de Bayes

Antes de introducirse a la prueba de normalidad de los residuales de nivel 2 se expone brevemente el estimador empírico de Bayes, el cual sirve como insumo para el test en cuestión. El estimador empírico de Bayes (EB) de los coeficientes de nivel 1 β_{pj} es una combinación ponderada de de la estimación de $\hat{\beta}_{pj}$ y de $\hat{\beta}_{p0} + \sum_{q=1}^{Q} \hat{\beta}_{pq} W_{qj}$

La ventaja del estimador EB que mejora el desempeño frente a estimaciones separadas, reside en que los estimadores de nivel 1 no son solo exclusivamente función de las unidades de nivel 2. En efecto el estimador EB es una función de los efectos fijos, los cuales son estimados extrayendo información de las unidades de nivel 1 aun cuando existan pocos datos sobre estas unidades (Steenbergen & Jones, 2002).

La prueba de normalidad de los residuos de nivel 2 utiliza dos variables; de un lado la distancia de Mahalanobis (mdist) como una medida de la distancia del estimador empírico de Bayes de los efectos cuyos valores ajustados son dados por el modelo multinivel estimado. La segunda variable utilizada (chiptc) son los valores teóricos de una χ^2 con *v* grados de libertad igual al número de factores aleatorios. Una aproximación a un test de normalidad de nivel 2 lo constituye un diagrama de dispersión entre las dos variables[54].

[54] Bajo el supuesto de normalidad, esta medida debería distribuirse aproximadamente χ^2 con *v* grados de libertad, siento *v* el número de efectos aleatorios. Véase el archivo residuosnivel2.

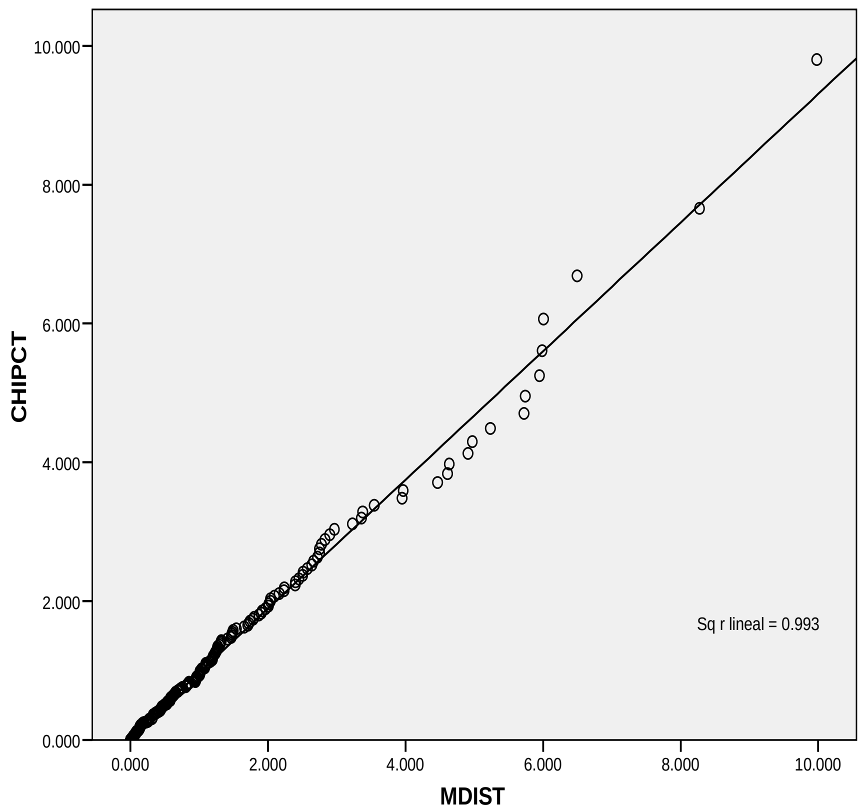

Figura No 5: Test de normalidad de los residuos de nivel 2: Una aproximación gráfica. Estimación con el software SPSS 15.0

Bajo el supuesto de normalidad en los residuos de nivel 2, el diagrama de dispersión debería mostrar una línea con 45 grados de inclinación, como ocurre en este caso, indicando normalidad. Maas & Hox (2004) muestran que la violación del supuesto de normalidad en los residuos de nivel 2 tiene un efecto pequeño en la estimación de los efectos fijos.

7. Modelo logístico multinivel

7.1 ¿Tutorías en matemáticas? Modelando la elección del estudiante mediante un modelo logístico multinivel

Asumiendo que quienes toman clases o tutorías remediales de matemáticas son aquellos estudiantes con pobre desempeño en esta área, a continuación se estima un modelo de elección binaria que permita identificar que variables a nivel individual y de la escuela soportan dicha elección. Dada la estructura jerárquica de los datos se espera que las variables a nivel de la escuela tengan un efecto en la elección individual, requiriendo de modelos logísticos dentro de un esquema multinivel. También denominados modelos jerárquicos lineales generalizados (HGLM), la aplicación de modelos logísticos en un modelo multinivel es paralelo a sus aplicaciones en modelos en diseños lineales[55].

7.2 Formulación del modelo lineal generalizado

De la misma manera que los modelos con variable dependiente continua, se está interesado en la esperanza (media) de la respuesta en función de covariables tanto de nivel 1 como de nivel 2. Siendo la variable dependiente dicotómica:

$$\text{REM\_MATE}_{ij}: \begin{cases} 1 & \text{Toma clases remediales de matematica} \\ 0 & \text{No toma clases remediales en el area} \end{cases}$$

[55] También denominados modelos lineales generalizados mixtos (Fielding, 2003; McCulloch & Searle, 2001). Los modelos HLM son un caso especial de los modelos HGLM cuando se utiliza la distribución normal para describir la distribución de los residuos. En el caso generalizado, la variable dependiente binaria son analizadas usando una distribución logística.

Definiendo:

$$\Pr(\mathrm{Y_i} = 1) = \mathrm{p_{ij}} \qquad (7.2.1)$$

$$\Pr(\mathrm{Y_i} = 0) = 1 - p_{ij}$$

La esperanza de una variable binaria se define como el valor que toma por su respectiva probabilidad:

$$\mathrm{E(Y_i) = (P_i)1 + (1 - P_i)0 = P_{ij}} \qquad (7.2.2)$$

En este caso la esperanza de una variable binaria, es la probabilidad de que su respuesta sea 1. Dado que las probabilidades estimadas deben caer en el intervalo $[0,1]$ se especifican funciones no lineales en las variables. Para fines de ilustración se asume un predictor:

$$\Pr(\mathrm{Y_i} = 1|\mathrm{X_i}) = \mathrm{h}\left(\beta_{0j} + \beta_{1j}\mathrm{X_i}\right) \qquad (7.2.3)$$

$$\mathrm{g}\{\Pr(\mathrm{Y_i} = 1|\mathrm{X_i})\} = \beta_{0j} + \beta_{1j}\mathrm{X_i} = \eta_{ij} \qquad (7.2.4)$$

Siendo equivalentes ambas formulaciones si y solo si la función $h(.)$ es la inversa de la función $g(.)$, esta última es conocida como “link function” entre tanto $h(.)$ es conocida como la función inversa de “function link” (Hesketh & Skrondal, 2012).

Una característica atractiva de los modelos lineales generalizados es que ellos involucran predictores que se asemejan a los modelos lineales. Esta flexibilidad permite trabajar con variables dicotómicas como explicativas, interacciones entre variables, polinomios entre otros, tal como se haría en un modelo lineal.

Las opciones disponibles para la “link function” con variables dependientes binarias son el Probit y el Logit. Dadas unas propiedades deseables en la transformación logit, para estimar una regresión logística, la especificación “logit link” es como sigue[56]:

[56] La primera propiedad deseable de la función “logit link” es que elimina la asimetría inherente en las estimaciones de las *Odds Ratio* (Agresti, 1996). Otra propiedad es que el rango del logit va desde infinito negativo hasta infinito positivo, lo cual elimina el problema de acotamiento asociados con el *Odds Ratio.* Por último, la

$$logit\{Pr(Y_i = 1|X_i)\} = ln\left\{\underbrace{\frac{Pr(Y_i=1|X_i)}{1-Pr(Y_i=1|X_i)}}_{Odds(Y_i=1|X_i)}\right\} = \eta_{ij} = \beta_{0j} + \beta_{1j}X_i \qquad (7.2.5)$$

Se tiene entonces que el logit η_{ij} representa el log de las probabilidades de “éxito” para la ith persona condicional a un conjunto de predictores, presentando propiedades deseables como el ser continuos y lineales en los parámetros.

La expresión dentro del paréntesis representa la probabilidad de que $Y_i = 1$ dado X_i, el numero esperado de éxitos (1) por fallas (0). El modelo logit puede expresarse alternativamente como una función exponencial del ratio de probabilidades (Odds).

$$Odds(Y_i = 1|X_i) = exp(\beta_{0j} + \beta_{1j}X_i) \qquad (7.2.6)$$

Dada la razón entre las probabilidades y la ratio de probabilidades se tiene que:

$$Odds = \frac{Pr(Y_i=1|X_i)}{1-Pr(Y_i=1|X_i)} \qquad (7.2.7)$$

Y la ratio de probabilidades

$$Pr = \frac{Odds}{1+Odds} \qquad (7.2.8)$$

$$Pr(Y_i = 1|X_i) = logit^{-1}(\beta_{0j} + \beta_{1j}X_i) = \frac{exp(\beta_{0j}+\beta_{1j}X_i)}{1+exp(\beta_{0j}+\beta_{1j}X_i)} \qquad (7.2.9)$$

Una aproximación básica para estimar los parámetros de la ecuación (6.2.9), es linealizar la función exponencial de manera que asuma la forma de un modelo jerárquico lineal de dos niveles, para luego de aplicar cuasi-verosimilitud usando el supuesto de distribución binomial y definir los determinantes de la elección (Goldstein & Rasbash, 1996).

función “logit link” es lineal en los parámetros, lo cual significa que los efectos de las variables explicativas en el *log-Odds* son aditivos ((O’ Connell, Golstein, Rogers and Peng, 2009).).

7.3 Penalizada cuasi-verosimilitud (PQL)

Los métodos más comunes para la estimación de modelos multinivel logísticos se basan en máxima verosimilitud. Entre las opciones se encuentran Marginal Cuasi verosimilitud, MQL (Goldstein & Rasbash, 1996) y Penalizada Cuasi Verosimilitud PQL (Breslow & Clayton, 1993). Siendo esta última metodología incorporada en las rutinas de HLM 7.0 a continuación se hace una descripción del método para un modelo de regresión logístico de dos niveles[57].

Siguiendo el desarrollo de Khan y Shaw (2011), considere la variable dependiente de nivel 1:

$$\mathrm{Y_{ij}} = \mathrm{REM\_MATE_{ij}}: \begin{cases} 1 & \text{Toma clases remediales de matematica} \\ 0 & \text{No toma clases remediales en el area} \end{cases}$$

$\mathrm{Y_{ij}}$ toma el valor de 1 con probabilidad condicional Pr_{ij}. La variable dependiente $\mathrm{Y_{ij}}$ se asume con distribución Bernoulli, cuya media $\mathrm{E}(\mathrm{Y_{ij}}) = \mathrm{Pr_{ij}}$ se relaciona con los predictores lineales tanto de primer nivel como de segundo nivel. El modelo en su primer nivel puede ser reescrito como:

$$\mathrm{logit_{ij}} = \mathrm{x_{ij}^T} \beta_{\mathrm{j}} \qquad (7.3.1)$$

Donde el vector $\mathrm{X_{ij}}$ de dimensión $(P+1) \times 1$ contiene las variables de nivel 1 y adicionalmente el intercepto; entretanto β_j es un vector columna que contiene los coeficientes de efectos fijos, a partir de los cuales se da origen a las variables de segundo nivel. Considerándolos en este caso con variación aleatoria:

$$\beta_{\mathrm{j}} = \mathrm{z_j} \gamma + \mathrm{u_j} \qquad (7.3.2)$$

[57] Desde su investigación en 1995, el modelo preferido por los investigadores al momento de estimar un modelo logit mediante modelos multinivel es el método PQL.

Donde z_j constituye un vector columna de observaciones sobre un conjunto de predictores a nivel de la escuela *j*, siendo γ el vector de coeficientes de nivel 2. Se probará entonces si los parámetros de nivel 1 son aleatorios, es decir si las preferencias de los estudiantes varían de acuerdo a atributos de la escuela o si por el contrario sus preferencias solo están condicionadas por variables de nivel 1.

El modelo mixto que combina ambos niveles es:

$$\mathrm{logit}_{ij} = x_{ij}{}^{T} z_j \gamma + x_{ij}{}^{T} u_j \tag{7.3.3}$$

$$\mathrm{logit}_{ij}\{\Pr(Y_{ij} = 1 | Xj_{ij})\} = \ln\left\{\frac{\Pr(Y_{ij}=1|X_{ij})}{1-\Pr(Y_i=1|X_{ij})}\right\} = \eta_{ij} = x_{ij}{}^{T} z_j \gamma + x_{ij}{}^{T} u_j \tag{7.3.4}$$

Empezando con un modelo de un solo nivel[58]:

$$Y_{ij} = p_{ij} + e_{ij} \tag{7.3.5}$$

Pudiendo linealizarse utilizando series de expansión de primer orden de Taylor. En la iteración *s* se obtiene:

$$p_{ij} \approx p_{ij}{}^{(s)} + \frac{dp_{ij}}{d\eta_{ij}}\left(\eta_{ij} - \eta_{ij}{}^{(s)}\right) \tag{7.3.6}$$

Evaluando la derivada en $p_{ij}{}^{(s)}$:

$$\frac{dp_{ij}}{d\eta_{ij}} = p_{ij}(1 - p_{ij}) = \omega_{ij} \tag{7.3.7}$$

Sustituyendo la aproximación lineal para p_{ij} en ecuación (2.1) se llega a:

$$Y_{ij} = p_{ij}{}^{(s)} + \omega_{ij}{}^{(s)}\left(\eta_{ij} - \eta_{ij}{}^{(s)}\right) + e_{ij} \tag{7.3.8}$$

Manipulando algebraicamente de manera que las cantidades conocidas queden en lado izquierdo de la ecuación, se tiene:

[58] El enfoque PQL puede ser tratado como un modelo no lineal de regresión (Khan & Shaw, 2011)

$$\frac{Y_{ij}-p_{ij}{}^{(s)}}{\omega_{ij}{}^{(s)}}+\eta_{ij}{}^{(s)}=\eta_{ij}+\frac{e_{ij}}{\omega_{ij}{}^{(s)}} \tag{7.3.9}$$

Denotando:

$$Y_{ij}{}^{*(s)}=\frac{Y_{ij}-p_{ij}{}^{(s)}}{\omega_{ij}{}^{(s)}}+\eta_{ij}{}^{(s)} \tag{7.3.10}$$

$$\varepsilon_{ij}=\frac{e_{ij}}{\omega_{ij}{}^{(s)}} \tag{7.3.11}$$

$$\varepsilon_{ij}\sim N\left(0,\omega_{ij}{}^{(s)-1}\right) \tag{7.3.12}$$

$$\eta_{ij}=x_{ij}{}^{T}z_j\gamma+x_{ij}{}^{T}u_j \tag{7.3.13}$$

Se observa entonces que la ecuación toma la forma de un modelo lineal jerárquico de dos niveles:

$$Y_{ij}{}^{*(s)}=x_{ij}{}^{T}z_j\gamma+x_{ij}{}^{T}u_j+\ \varepsilon_{ij} \tag{7.3.14}$$

Este esquema es conocido como penalizada cuasi-verosimilitud (PQL) al ser obtenido involucrando solo primeras y segundas derivadas, con un término de penalización sobre los efectos aleatorios. La estimación de $\eta_{ij}{}^{(s)}$ se denota a continuación:

$$\eta_{ij}^{(s)}=x_{ij}{}^{T}z_j\gamma^{(s)}+x_{ij}{}^{T}u_j^{*(s)} \tag{7.3.15}$$

Donde $u_j^{*(s)}$ resulta de la siguiente aproximación:

$$u_{oj}{}^{*(s)}\cong\left(z_j{}^{T}w_j{}^{(s)}z_j+T^{(s)-1}\right)^{-1}z_j{}^{T}w^{(s)}\left(Y_j^{*(s)}-X_{ij}^{T(s)}\hat{\gamma}^{(s)}\right) \tag{7.3.16}$$

$$w_j{}^{(s)}=diag\left\{\omega_{ij}{}^{(s)},\dots,\omega_{n_j j}{}^{(s)}\right\} \tag{7.3.17}$$

Aunque la estimación PQL tiene como ventaja la capacidad de converger, no obstante en su versión restringida como no restringida tiene como inconveniente el no calcular una desviación estadística que permita comparar entre modelos. De otro lado aunque las estimaciones a través de la metodología alternativa de Laplace producen una confiable desviación estadística que permite comparar entre modelos, puede presentar problemas para converger[59].

7.4 Modelo nulo

La utilidad de un modelo nulo sin variables a nivel del estudiante o de la escuela permite obtener información de la variabilidad de la probabilidad de elegir asistir a tutorías remediales en matemáticas a través de las escuelas.

En primera instancia se estima un modelo nulo, este definido por el siguiente conjunto de ecuaciones:

$$\eta_{ij} = \beta_{0j} \tag{7.4.1}$$

Utilizando el logaritmo de la razón de probabilidades o logit, el modelo incluye un efecto aleatorio con el fin de analizar si las preferencias resumidas en los coeficientes de nivel 1, varían por escuela.

$$\beta_{0j} = \beta_{00} + u_{0j} \tag{7.4.2}$$

$$\eta_{ij} = \beta_{00} + u_{0j} \tag{7.4.3}$$

$$ln\left[\frac{p_{ij}}{1-p_{ij}}\right] = \beta_{0j} = \eta_{ij} \tag{7.4.4}$$

Sin variables explicativas, β_{0j} representa el "log odds" de la probabilidad de que un estudiante elija tomar clases remédiales de matemáticas a no tomarlas en la *jth* escuela. El

[59] HLM 7.0 incorpora la transformación de Laplace entre las opciones de estimación la cual no convergió en algunos modelos. Un requisito para el cálculo de la desviación estadística bajo la transformación de Laplace es la inclusión efectos aleatorios u_{0j} en la estimación de los parámetros de nivel 1 en función de las variables de nivel 2.

logaritmo de la razón de probabilidades promedio estimado por máxima verosimilitud es de 0.3410. Este valor aproximadamente constituye la ratio de frecuencias de los estudiantes que asisten a tutorías remediales en matemáticas respecto a quienes no las toman:

$$\frac{p_{ij}}{1-p_{ij}} = \frac{24.7}{73.1} = 0.3317$$

A continuación se consignan los resultados del modelo HGLM nulo (véase en anexos Modelo 9_HGLM nulo)

TABLA 9. Estimación del modelo logístico multinivel sin covariables

Modelo de unidades especificas					
Efecto fijo					
	Coeficiente	Error st.	T-est*.	OR	Prob
Intercepto(β_{00})	-1.103414	0.045815	-24.084	0.331736	0.000
Efecto Aleatorio					
Estimación de los componentes de la varianza					
	Error St	Varianza	Gdl**	Chi-cuadrado	Prob
Error de nivel 2, u_{0j}	0.49722	0.24723	263	488.77030	0.000

Estimación con el software HML 7.0. **Errores estándar robustos. **Grados de libertad

Un modelo logístico multinivel sin covariables en ambos niveles, al igual que su equivalente en un modelo lineal justifica la inclusión de variables a nivel de la escuela; la significancia de la varianza del termino de error de segundo nivel indicaría que el logaritmo de la razón de probabilidades promedio de la escuela *jth* de elegir tutorías remediales en matemáticas varia a través de las escuelas.

7.5 Modelando la probabilidad con covariables: Modelo lineal generalizado mixto (HGLM)

Al igual que en los modelos lineales generalizados, los modelos HGLM tienen implícito el supuesto de que las respuestas están distribuidas independientemente Bernoulli, con probabilidades determinadas con X_{ij} y u_{0j} (Hesketh & Skrondal 2012):

$$Y_{ij} \sim Bernoulli(\varphi_{ij}) \quad (7.5.1)$$

Así como en un modelo HLM estándar, los coeficientes de nivel 1 de un modelo HGLM pueden ser especificados como fijos, con variación aleatoria o con variación no aleatoria[60] (O´Connell et al. 2009). Para ilustrar la diferencia entre estos dos últimos considere un modelo con un predictor de primer nivel, el nivel socioeconómico (ESCS) y cultural y el género (GENERO).

$$logit(\varphi_{ij}) = \eta_{ij} \quad (7.5.2)$$

$$\eta_{ij} = \beta_{0j} + \beta_{1j}\mathrm{ESCS}_{ij} + \beta_{2j}\mathrm{GENERO}_{ij}$$

Proponiendo un intercepto aleatorio, se considera como predictor de segundo nivel el efecto contextual $(\mathrm{MEANESCS_j})$ y un efecto aleatorio u_{0j}[61].

$$\beta_{0j} = \beta_{00} + \beta_{01}\mathrm{MEANESCS_j} + u_{0j} \quad (7.5.3)$$

$$\beta_{1j} = \beta_{11}$$

$$\beta_{2j} = \beta_{21}$$

Combinando ambos niveles se llega a un modelo jerárquico lineal generalizado (HGLM)[62]:

[60] En la literatura sobre logit jerárquicos, los modelos sin variación aleatoria son mejor conocidos con el nombre de “modelo promedio poblacional” y “modelos especifico a las unidades” para los modelos con variación aleatoria (que incluyen un término de error u_{0j})

[61] El intercepto aleatorio puede ser pensado como el reflejo de la omisión de variables a nivel de la escuela.

$$logit\{Pr(Y_{ij} = 1|X_{1ij}, X_{2ij,}\mathrm{u}_{0\mathrm{j}})\} = \beta_{00} + \beta_{01}\mathrm{MEANESCS_j} + \beta_{11}\mathrm{ESCS}_{ij} + \beta_{2\mathrm{j}}\mathrm{GENERO}_{ij} + \mathrm{u}_{0\mathrm{j}} \quad (7.5.4)$$

En el anterior modelo denominado "modelo específico a las unidades" los efectos aleatorios son asumidos independientes a través de las j escuelas e independientes de las covariables X_{ij}.

En contraste con los modelos lineales de efectos aleatorios, las estimaciones consistentes en modelos de efectos aleatorios logísticos requieren que la parte aleatoria del modelo está correctamente especificado adicional a la parte fija[63].

Entre tanto un modelo con variación no aleatoria contemplaría la siguiente estructura (modelo lineal generalizado):

$$logit\{Pr(Y_{ij} = 1|X_{ij}, \mathrm{u}_{0\mathrm{j}})\} = \beta_{00} + \beta_{01}\mathrm{MEANESCS_j} + \beta_{11}\mathrm{ESCS}_{ij} + \beta_{21}\mathrm{GENERO}_{ij} \quad (7.5.5)$$

Dependiendo de la inclusión del término aleatorio$\mathrm{u}_{0\mathrm{j}}$ en los modelos logit jerárquicos puede establecerse una distinción entre los resultados de lo que la literatura ha denominado modelo "modelo promedio poblacional" de los resultados del modelo denominado "unidades específicas"[64].

El modelo dado en ecuación (6.5.5) llamado "población promedio" al omitir $\mathrm{u}_{0\mathrm{j}}$ no provee información sobre las variaciones de nivel 2, siendo solo es útil para hacer inferencia de los efectos poblacionales promedio.

---

[62] Los modelos HLM son una extensión de modelo jerárquico lineal generalizado HGLM.
[63] Diferentes modelos fueron estimados estimando las pendientes del modelo con variación aleatoria y no aleatoria para asegurarse una correcta especificación del modelo.
[64]Hesketh y Skrondall (2012). Ambos tipos de resultados están disponibles en STATA y HLM 7

Entretanto el modelo denominado de “modelos especifico a las unidades” en ecuación [6.5.4] incluye además de las covariables, el término u_{0j}. En este último modelo es posible estimar un cambio en la probabilidad de respuesta para cualquier unidad de nivel 2, siendo la probabilidad una función no lineal, puesto que el cambio dependerá de u_{0j}; en otras palabras este modelo provee información de cómo las covariables a nivel de grupo varían entre las escuelas (Diggle et al, 1994)[65]. Surge entonces el interrogante ¿Cuál de los dos modelos debe utilizarse?

Cuando las características entre las unidades de nivel 2 varían, luego este modelo es el más adecuado, teniendo que el término aleatorio u_{0j} es significativo asi como su varianza. Este modelo es más adecuado cuando se está interesado en tener en cuenta la variación entre las escuelas (en sus características) en la probabilidad de “éxito.

Tal como lo menciona O´Connell et al (2009), la diferencia entre ambos modelos puede ser grande dependiendo de cuan diferentes sean las unidades de nivel 2 respecto a las variables a nivel de grupo[66]. En efecto en este caso particular donde se trabaja con una muestra de escuelas (unidades de nivel 2) relativamente heterogéneas (escuelas privadas y escuelas públicas) se esperaría observar alguna diferencia significativa entre ambas estimaciones, lo cual efectivamente sucede en este caso.

Aun cuando la variación entre las unidades de nivel 2 sea nula de entrada existirá por mínima que sea una diferencia entre los efectos del modelo “población-promedio” y los efectos del modelo de “unidades específicas”, debido principalmente a que el promedio de una función no lineal no es equivalente a una función no lineal del promedio (Hesketh & Skrondal, 2012)

[65] Cuando se incluye u_{0j} se esta frente a un modelo que la literatura también denomina “modelo clúster-especifico” Hosmer y Lemeshow (2000).En este modelo los efectos pueden asimilarse al efecto “promedio de los agentes” encontrados en la estimación de un modelo logit estándar. El “promedio de los agentes” promedia todos los efectos marginales de los individuos adicionando un efecto aleatorio u_{0j}.

[66] Cuando la variación entre las escuelas en sus características no es significativa, se omite el término de error u_{0j} con la consecuencia de que ambas estimaciones *“unidades específicas”,* y *“población-promedio”* son iguales.

Las probabilidades del modelo población promedio con intercepto aleatorio o “subject specific” (ecuación 6.5.4) puede ser obtenidas promediando las probabilidades de unidades específicas sobre la distribución del intercepto aleatorio (Hesketh & Skrondal 2012). Toda vez que los interceptos aleatorios son continuos, este promedio se obtiene integrando:

$$\Pr\left(Y_{ij}=1\middle|X_{ij},Z_j\right)=\int \Pr\left(Y_{ij}=1\middle|X_{ij},Z_j,u_{0j}\right)\phi\left(u_{0j};0,\psi\right)d\,u_{0j} \qquad (7.5.6)$$

Calculando el promedio de una función no lineal (promediando las probabilidades de las unidades específicas)[67]:

$$\Pr\left(Y_{ij}=1\middle|X_{ij},Z_j\right)=\int \frac{\exp\left(\beta_{00}+\beta_{01}Z_j+\beta_{11}X_{ij}+u_{0j}\right)}{1+\exp\left(\beta_{00}+\beta_{01}Z_j+\beta_{11}X_{ij}+u_{0j}\right)}\phi\left(u_{0j};0,\psi\right)d\,u_{0j} \qquad (7.5.7)$$

Calculando el efecto del modelo con efecto promedio –población:

$$\neq \frac{\exp\left(\beta_{00}+\beta_{01}Z_j+\beta_{11}X_{ij}\right)}{1+\exp\left(\beta_{00}+\beta_{01}Z_j+\beta_{11}X_{ij}\right)} \qquad (7.5.8)$$

Como ya se mencionó, el promedio del inverso del logit del predictor lineal:

$$\beta_{00}+\beta_{01}Z_j+\beta_{11}X_{ij}+u_{0j} \qquad (7.5.9)$$

Esto es el promedio de una función no lineal dado en ecuación [6.5.7] no es lo mismo que el inverso del logit del promedio del predictor lineal dada en ecuación [6.5.8]:

$$\beta_{00}+\beta_{01}Z_j+\beta_{11}X_{ij} \qquad (7.5.10)$$

dando origen a una diferencia entre ambos efectos[68].

---

[67] Donde $\phi\left(u_{0j};0,\psi\right)$ es la función de densidad normal con media cero y varianzaψ.

[68] En STATA se comprueba la existencia de tal diferencia con un simple ejercicio: display (invlogit(1)+invlogit(2))/2= 0.8059 para el promedio de una función no lineal es diferente que el inverso del logit del promedio del predictor, invlogit(1+2)/2)=0.8175 (Hesketh& Skrondall,2012)

7.6 Modelo a estimar y resultados de la estimación

La especificación del modelo HGLM queda de la siguiente manera:

$$ln\left[\frac{p_{ij}}{1-p_{ij}}\right] = \beta_{00} + \beta_{01}\text{MEANESCS}_j + \beta_{11}\text{ESCS}_{ij} + \beta_{21}\text{GENERO}_{ij} + u_{0j} \qquad (7.6.1)$$

Constituyendo un ejemplo de un modelo mixto lineal generalizado (GLMM) con efectos fijos y efectos aleatorios, los resultados de la estimación se consignan a continuación (Véase en anexos Modelo 10_Modelo HGLM):

Cuadro 10. Estimando la probabilidad de tomar clases remediales de matemáticas mediante un Modelo logístico multinivel

Estimación de los efectos fijos

Modelo "unidades específicas" con errores robustos estándar

Efectos Fijos						
	Coeficiente	Error st**.	T-est.	OR	Prob	*Intervalo de Conf.*
Intercepto (β_{00})	-1.5153	0.080078	-18.924	0.219	0.000	(0.188 , 0.257)
MEANESC(β_{01})	-0.2378	0.060723	-3.917	0.788	0.000	(0.699 , 0.887)
ESCS(β_{10})	-0.0851	0.035342	-2.410	0.918	0.016	(0.860 , 0.984)
GENERO(β_{20})	0.2327	0.069491	3.349	1.262	0.001	(1.099 , 1.444)

Efectos aleatorios					
	Error St	Varianza	Gdl	Chi-2	Prob
Error de nivel 2 u_{0j}	0.4409	0.19445	262	434.396	0.000

Estimación con el software HML 7.0. **Errores estándar robustos

Si realmente hay una estructura jerárquica dentro de la elección del estudiante, parecería natural incorporarla en el modelo mediante la inclusión de u_{0j}, y que fuera significativoel componente de la varianza asociado a u_{0j}, como efectivamente sucede. En este sentido se podría afirmar que el modelo "unidades específicas", constituye un modelo multinivel puesto que la estructura jerárquica está explícitamente especificada[69].

7.7 Inferencia estadística vía Intervalos de confianza

Dada que es más útil la interpretación de los coeficientes OR que la de sus respectivos coeficientes (logaritmo natural de los coeficientes OR), a continuación se aborda la manera de hacer inferencia con los coeficientes "Odds ratio". La hipótesis nula para los coeficientes OR usualmente es:

$$H_0: OR = 1$$

$$H_A: OR \neq 1$$

Utilizando un nivel de significancia $\alpha = 0.05$ el correspondiente intervalo de confianza para los coeficientes OR es obtenido exponenciando los límites de intervalo de confianza:

$$\exp\{\hat{\beta} - t_{\alpha/2}\widehat{SE}(\hat{\beta})\} \;,\; \exp\{\hat{\beta} + t_{\alpha/2}\widehat{SE}(\hat{\beta})\} \qquad (7.7.1)$$

En este caso si el intervalo de confianza contiene el valor de los coeficientes bajo H_0, se puede concluir que la correspondiente variable no es significativa. Hay que tener especial cuidado con la estimación de estos intervalos al utilizar el software HLM 7.0 toda vez que no utiliza los errores estándar robustos, lo cual si hace al momento de estimar los t-estadísticos de

[69] Véase en Anexos Modelo 9_Modelo HGLM

los efectos fijos. En este caso para el coeficiente OR asociado a la variable GENERO, el intervalo de confianza utilizando errores robustos es:

$$\exp\{0.2327 - 1.960(0.0694)\}\ ,\ \exp\{0.2327 + 1.960(0.0694)\}$$

$$IC=[1.099,\ 1.444]$$

Con lo cual se concluye que el coeficiente OR asociado a GENERO es significativo dado que no incluye el valor de OR bajo Ho.

Respecto a la segunda covariable de nivel uno ESCS, el intervalo de confianza es:

$$\exp\{-0.0851 - 1.960(0.035342)\}\ ,\ \exp\{-0.0851 + 1.960(0.0353422)\}$$

$$IC=[0.860,\ 0.984]$$

Por último, la variable de nivel 2, el efecto contextual también resulta significativa, su intervalo de confianza no contiene el parámetro bajo H_0:

$$\exp\{-0.2378 - 1.960(0.060723)\}\ ,\ \exp\{-0.2378 + 1.960(0.060723)\}$$

$$IC=[0.699,\ 0.887]$$

7.8 Interpretación de los resultados e inferencia estadística

El coeficiente OR para la variable GENERO, mide el impacto sobre la probabilidad de elegir de una variable explicativa. Por ejemplo si se quiere comparar las probabilidades de que un estudiante toma clases remediales de matemáticas entre estudiantes hombres (GENERO=1) y estudiantes mujeres (GENERO =0), se calcularía la siguiente ratio:

$$OR = \frac{odds(REM\_MATE_{ij}=1|Vocacional_{ij})}{odds(REM\_MATE_{ij}=1|Academica_{ij})} = \frac{\frac{Pr(REM\_MATE_{ij}=1GENERO_{ij}=1)}{1-Pr(REM\_MATE_{ij}=1GENERO_{ij}=1)}}{\frac{Pr(REM\_MATE_{ij}=1GENERO_{ij}=0)}{1-Pr(REM\_MATE_{ij}=1GENERO_{ij}=0)}} \quad (7.8.1)$$

Respecto a la interpretación, un *OR* de 1.0 indica que una variable explicativa no tiene efecto en los *Odds* de elegir tomar clases remediales, es decir, los *Odds* de elegir es el mismo para estudiantes hombres y estudiantes mujeres.

Es de anotar que los *Odds ratio* tienen como cota inferior cero mas no cota superior, la cual puede ir a infinito. Valores pequeños para los coeficientes *OR* (<1) indican que los *Odds* de elegir para las personas con el valor de X_{ij} en el denominador (X_{ij}=0) es más grande que los *Odds* de elegir para las personas con el valor más alto usado en el numerador (para la opción X_{ij} =1).Lo contrario es cierto cuando el coeficiente *OR* es mayor que la unidad. En el caso de la variable género, su respectivo coeficiente OR con valor de 1.26, significativo al 5%, arroja un resultado no esperado, a saber que los estudiantes hombres tienen más probabilidad de tomar tutorías remediales en matemáticas respecto a las mujeres.

De otro lado con un coeficiente OR para el nivel socioeconómico de 0.9183, se concluye que la probabilidad de tomar clases remediales de matemáticas es más alta para estudiantes con bajos niveles socioeconómicos respecto a los estudiantes con mayores niveles para esta variable.

Respecto al efecto contextual, MEANESCS, es significativo al 5% y con un coeficiente OR de 0.7883, se encuentra que la probabilidad promedio de que los estudiantes asistan a tutorías en matemáticas, es más alta para escuelas con bajo promedio de nivel socioeconómico respecto a escuelas con alto promedio socioeconómico. Se confirma con este modelo logístico multinivel la regularidad empírica hallada mediante el modelo multinivel lineal, a saber que el efecto contextual cumple un rol fundamental en el desempeño académico de los estudiantes, incluso más importante si se comparan los coeficientes del efecto contextual y del nivel socioeconómico, siendo relativamente mayor el segundo en ambos modelos, el lineal y logístico multinivel.

8. Conclusiones

La presente investigación representa un enfoque alternativo a los ya tradicionales estudios que se limitan a la comparación de puntajes académicos entre las escuelas públicas y privadas. Al tener en cuenta dinámicas diferenciales al interior de cada tipo de escuela sea pública o privada, se logra tener otra mirada sobre los determinantes del desempeño académico a nivel de secundaria. Inicialmente se puede confirmar las regularidades empíricas en los tradicionales estudios, a saber que el género y el tipo de escuela tienen impacto sobre el desempeño académico individual. En efecto los estudiantes hombres aventajan a sus similares mujeres en la prueba de matemáticas, mientras que pertenecer a una escuela pública impacta negativamente el mencionado desempeño.

Se postularon como dinámicas diferenciales al interior de cada escuela, el impacto de la orientación vocacional y del "efecto par" o efecto contextual en el desempeño académico. Respecto al primero se encuentra que al interior de la escuela privada, los estudiantes con orientación vocacional académica tienen puntajes significativamente más altos que sus pares con orientación vocacional técnica. Entretanto al interior de las escuelas públicas, el anterior resultado se revierte con una ligera ventaja, aunque significativa de los estudiantes con orientación vocacional técnica sobre sus pares con orientación académica; en realidad podría afirmarse que existe muy poca diferencia en el promedio entre los estudiantes de escuelas públicas con orientación técnica respecto a aquellos con orientación académica.

Esta regularidad empírica ignorada en anteriores estudios podría tener su explicación en la composición de los tipos de orientación al interior de cada escuela. Mientras que en la escuela pública la composición de las orientaciones esta distribuida en un 64% en el enfoque comercial y un 25% enfoque industrial, en las escuelas privadas la concentración se encuentra casi exclusivamente en el área comercial con un 92%.

La concentración de la orientación vocacional en la escuela privada a favor del enfoque comercial estaría explicada por una mejor adaptación de las escuelas privadas a un cambiante entorno económico en el cual la terciarizacion es cada vez más acentuada. Para cerrar la disertación alrededor del efecto de la orientación vocacional, debe anotarse que sin entrar a diferenciar por tipo de escuela, se encuentra que los estudiantes con orientación vocacional técnica muestran un desempeño académico inferior a los que tienen orientación vocacional académica. Este resultado se explica por el menor tiempo que tienen los estudiantes con orientación vocacional técnica para dedicar a las actividades académicas, toda vez que al menos el 25% del total de horas de clase deben ser destinadas al aprendizaje y practica de saberes laborales adquiridos durante la práctica.

Respecto al segundo efecto diferencial al interior de cada tipo de escuela, el efecto contextual o efecto par, medido por el promedio del nivel socioeconómico de la escuela, al interior de las escuelas públicas aunque significativo es apenas perceptible. Se encuentra entonces que el nivel socioeconómico promedio de la escuela, tiene un impacto relativamente pequeño sobre el desempeño académico individual en comparación con el nivel socioeconómico del estudiante. Este hallazgo se explicaría por la relativa homogeneidad socioeconómica existente al interior de una escuela pública. Dicha regularidad empírica cambia cuando se traslada a la escuela privada, en el cual el efecto contextual tiene gran impacto en el puntaje obtenido en la prueba de matemáticas, incluso comparado con el impacto nivel socioeconómico individual.

Se concluye entonces que existen unos efectos diferencial es por escuela sea pública o privada cuando se tienen en cuenta los impactos sobre el desempeño académico de las variables orientación vocacional y el efecto contextual o “efecto par”. En el primer caso la interacción de los roles dela escuela como preparación para la universidad y para el trabajo, impactan el desempeño académico medido por los puntajes de la prueba PISA de manera diferenciada. En el segundo caso, el efecto contextual va muy de la mano con la composición socioeconómica al interior de cada escuela.

La inclusión del "efecto par" no solo obedeció a la hipótesis planteada por la mayoría de trabajos en el área sino también como un resultado "natural" propio de la corrección de endogeneidad de nivel 2. En efecto uno de los aciertos metodológicos de la presente investigación es la detección de endogeneidad y su posterior corrección mediante la adaptación de la aproximación de Mundlak (1978) de datos de panel a modelos multinivel. Luego la inclusión del efecto contextual con su correspondiente significancia estadística se justifica desde lo teórico y metodológico.

Se logra encontrar un esquema que permite justificar de manera consistente y lógica la utilización de la metodología multinivel para el tema propuesto. Gran parte del desarrollo del trabajo intenta llevar a feliz término esta tarea. Si bien la mayoría de trabajos en el área justifican desde el diseño muestral por clúster o estratificado la utilización de técnicas multinivel, realmente no se logra visualizar un trabajo que logre defender el modelo finalmente propuesto. Como lo mencionaba Cox (1987) "todos los modelos son falsos pero algunos son útiles", la utilidad del modelo tienen intrínsecamente que ver con la "defensa" del modelo.

Es claro que la naturaleza jerárquica de los datos se adapta muy bien a los modelos jerárquicos, no obstante detrás de estos modelos existen una serie de supuestos que deben ser contrastados. Debiendo probarse si se violan los supuestos de homocedasticidad, normalidad y exogeneidad para luego incorporar su posible solución, se encuentra que los dos primeros se mantienen mientras el último no, requiriendo de la mencionada aproximación de Mundlak.

Abordar la posible violación de los supuestos no es suficiente con miras a la construcción del modelo jerárquico, pues también requiere justificar la escogencia del número de niveles dentro así como la clasificación de los coeficientes de primer nivel en efectos fijos o aleatorios. En el primer caso la omisión de niveles dentro del modelo llevaría a estimaciones sesgadas de los efectos fijos, razón por la cual no es conveniente omitir sin justificación un tercer nivel en el análisis, lo cual no es el caso encontrando apropiado la utilización de un modelo de dos niveles.

Por ultimo dentro del esquema de "defensa" del modelo se encuentra la justificación de un modelo con intercepto aleatorio y con efectos fijos en cada una de las pendientes. Detrás de correcta especificación de los coeficientes de nivel 1 se encuentra el supuesto de homocedasticidad de estos coeficientes; el ignorar la variación de carácter aleatorio tanto en el intercepto como en las pendientes de nivel 1 conduciría a la violación del anterior supuesto.

Se confirma que solo el intercepto tiene variación aleatoria, es decir, el promedio del desempeño académico en matemáticas varia a través de las 264 escuelas de la muestra. Adicional al efecto aleatorio de nivel 2 se encuentra que el tipo de escuela y el efecto contextual impactan positivamente el desempeño académico promedio. Respecto a los coeficientes de nivel 1 si bien se habla de efectos fijos o aleatorios, resulta más adecuado referirse a estos coeficientes en términos de variación aleatoria o no aleatoria, toda vez que el hablar de "efectos fijos" solo se circunscribe a denotarlos como una constante.

El modelo finalmente escogido presenta variación no aleatoria en dos de las pendientes de nivel 1. El postular el tipo de escuela como variable explicativa para modelar los coeficientes asociados al efecto contextual o efecto par y a la variable orientación vocacional siendo significativos en ambos casos justifica dos de las principales hipótesis de trabajo, esto es, la existencia de impacto diferencial del efecto par y de la orientación vocacional por tipo de escuela.

Por último, se confirma mediante la estimación de un modelo logístico multinivel, que el ambiente en que se desenvuelven los estudiantes tiene un significativo impacto en su desempeño académico. Modelando la probabilidad de que un estudiante elija tomar clases adicional o tutorías para remediar el bajo desempeño académico se encuentra que a mayor nivel socioeconómico individual y promedio, es menor la probabilidad de tomar esta clase de tutorías.

De naturaleza fija el coeficiente que acompaña al nivel socioeconómico individual (ESCS) apoya lo encontrado en el modelo multinivel lineal en el cual este coeficiente no varía entre escuelas. Con la estimación final de este modelo se concluye que las preferencias de los estudiantes respecto a su elección y que se ven reflejadas en los coeficientes de nivel 1, no cambian a través de las escuelas, no obstante la estructura jerárquica que justifica la elección de un modelo logístico multinivel reside en la naturaleza aleatoria del intercepto del modelo, lo cual comparte su análogo lineal.

Bibliografia

Agresti, A (2002). *Categorical Data Analysis*. New Jersey: John Wiley & Sons.

Bettinger, E; Kremer, M & Saavedra, J. (2010). Are Educational Vouchers Only Redistributive? *The Economic Journal*, *120*(546), 204–228.

Breslow, N.E. & Clayton, D.G. (1993). Approximate inference in generalized linear models. *Journal of American Statistical Association*, *88*, 9-25

Caldas, S (1999). Multilevel examination of student, school, and district-level effects on academic achievement. *Journal of Educational Research*, *93*, 91-100

Dempster, A, Laird, N & Rubin, D (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*. *Series B (Methodological)*, 39 (1),1-38.

Diggle. P, Liang, K. & Zeger, S. (1994). *Analysis of longitudinal data*. Oxford: Clarendon.

Eide E, & Showalter M (1998). The effect of school quality on student performance: A quantile regression approach. *Economic Letters 58*, 345-350.

Ferron, J., Hogarty, K., Dedrick, R., Hess, M., Niles, J., & Kromrey, J. (2008). Reporting results from multilevel analyses. In A. O'Connell & B, McCoach (Eds.), *Multilevel analysis of educational data* (pp. 391-426). Charlotte, NC: Information Age Publishing.

Goldstein, H (1986). Multilevel mixed linear model using iterative generalized least squared. *Biometrika*, *73*, 43-56.

Goldstein, H & Rasbash, J (1996). Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society: Series A*, *159*(3), 505-513.

Goldstein, H (2002). *Multilevel statitiscal models*. London: Arnold Publishing.

Hanchane, S & Mostafa, T (2010). Endogeneity problems in multilevel estimation of education production functions: an analysis using PISA data. *Centre for Learning and Life Chances in Knowledge Economies and Societies. LLAKES, 14,* 1-45. *Retrieved from* http://www.llakes.org/wp-content/uploads/2010/11/HanchaneMostafa-14-final-online.pdf

Hanushek, E. (1996). Measuring Investment in Education. *The Journal of Economic Perspectives*, *10*(4), 9-30.

Hesketh, S & Skrondal, A (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, *128*(2), 301–323.

Hesketh, S & Skrondal, A (2012). *Multilevel and longitudinal modeling using Stata*. College Station: Stata Press.

Hsiao, C (2003). *Analysis of Panel Data*. Cambridge: Cambridge University Press.

Hosmer, D & Lemeshow, S (2004). *Applied Logistic Regression*. New Jersey: John Wiley & Sons.

Hox, J (2010). *Multilevel Analysis, Techniques and Applications*. New York: Routledge.

Jacinto, C (2009). *Consideraciones sobre estrategias de inclusión con calidad en la escuela secundaria*. Retrieved September 12, 2013, from SITEAL website: http://www.siteal.iipe-oei.org

J. de Leeuw & E. Meijer (2008). Introduction to multilevel analysis. In J. de Leeuw & E. Meijer (Eds), *Handbook of Multilevel Analysis*, (pp. 84-116). New York: Springer.

Khan, R & Shaw, E. (2011). Multilevel Logistic Regression Analysis Applied to Binary Contraceptive Prevalence Data. *Journal of Data Science*, 9, 93-110. Retrieved from: *http://ssrn.com/abstract=2019344*

Kref, G., De Leew, J., & Aiken, L (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Lineal Models*, *30*, 1-22.

Lavy, V (2010). Do differences in school´s instruction time explain international achievement gaps in math, science, and reading? Evidence from developed and developing countries. *National Bureau Economic Research. Working paper 16227*, from Nber website *http://www.nber.org/papers/w16227*.

Maas, C., & Hox, J. (2005).Sufficient sample sizes for multilevel modeling. *Methodology*, *1*, 86-92.

Maas, C., & Hox, J. (2004). The influence of violations of assumptions on multilevel parameter estimates and their standard errors. *Computational statistics & data analysis*, *26*, 427-440.

Maccoach, D and Black, A (2008). Evaluation of Model Fit and Adequacy. In A.A. O'Connell & D. B. McCoach (Eds.), *Multilevel modeling of educational data* (pp. 245-271). Charlotte, NC: Information Age Publishing.

Ma, X. ,Ma, L., Bradley K, (2008). Using multilevel modeling to investigate school effects. In A.A. O'Connell & D. B. McCoach (Eds), *Multilevel modeling of educational data* (pp. 245-271). Charlotte, NC: Information Age Publishing.

Mislevy, R, Johnson, E., G., & Muraki, E. (1992). Scaling procedures in NAEP. *Journal of Educational Statistics, 17*, 131-154

Moerbeek, M (2004). The consequence of ignoring a level of nesting in multilevel analysis. *Multivariate Behavioral Research, 39*, 129-149.

Mundlak, Y (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica*, *46*(1), 69-85.

O'Connell, A., Goldstein, J., Rogers, H., & Peng, C. (2008).Multilevel Logistic Models for Dichotomous and Ordinal Data. In A.A. O'Connell, A.A. & D.B. McCoach (Eds.), *Multilevel Modeling of Educational Data*. Charlotte, NC: Information Age Publishing.

OECD (1999). *Classifying Educational Programmes Manual for ISCED-97 Implementation in OECD Countries*. Organization for Economic Co-operation and Development.

Opdenakker, M. & Van Damme, J. (2001). Relationship between school composition and characteristics of school process and their effect on mathematics achievement *.British Educational Research Journal*, *27*, 407–432.

Raudenbush S. (1995). Reexamining, reaffirming, and improving of Hierarquical models. *Journal of Educational and Behavioral Statistics*, *20*(2), 210-220.

Raudenbush, S & Brick, A (2002). *Hierarquical lineal models: Applications and data analysis methods*. Thousand Oaks, CA: Sage Publications.

Rangvid, B (2008). Source country differences in test score gaps: evidence from Denmark. *University Press of Southern Denmark, 22,* 1-61.

Saavedra, J.E y Medina, C (2012). Formación para el Trabajo en Colombia. *Documentos CEDE*, *35*, 1-84.

Scheeweis, N & Winter-Ebmer, R (2005). Peer effects in Austrian Schools. *Reihe Ökonomie / Economics Series*, Institut für Höhere Studien, *170*, 1-24

Snijders, T. (2005). Power and sample size in multilevel modeling. In B.S Everitt & D.C Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science.* (pp. 1570-1573). Chicester: Wiley.

Snijders, T & Bosker, R (2012). *Multilevel Analysis: An introduction to basic and advanced multilevel modeling*. London: Sage Publishers

Steenbergen, M. & Jones, B (2002). Modeling multilevel data structures. *American Journal of Political Science*, *46*(1), 218-237.

Swaminathan H. y Rogers J. (2008). Estimation procedures for hierarchical linear models. In A.A. O'Connell and D. B. McCoach (Eds), *Multilevel modeling of educational data* (pp. 469-517). Charlotte, NC: Information Age Publishing.

Tian, M (2006). A quantile regression analysis of family background factor effects on mathematical achievement. *Journal of Data Science*, *4*, 461-478.

Zimmerman, D and Winston, G (2003).Peer Effects in a Higher Education. In Hoxby, M Caroline (Ed.), *College Choices: The Economics of Where to Go, When to Go, and How Pay For It*. Chicago: University Chicago Press.

Anexos

Modelo 1_modelo nulo.

```
  The maximum number of level-1 units = 5021
  The maximum number of level-2 units = 264
  The maximum number of iterations = 100
Method of estimation: restricted maximum likelihood
This is part of a plausible value analysis using the following variables:
     PV1MATH
     PV2MATH
     PV3MATH
     PV4MATH
     PV5MATH

Summary of the model specified (in hierarchical format)
 ---------------------------------------------------

Level-1 Model

        PV1MATH = B0 + r

Level-2 Model
        B0 = G00 + u0
Mixed Model
    PV1MATH = G00  + u0+ r
THE AVERAGED RESULTS FOR THIS PLAUSIBLE VALUE RUN
 sigma^2 =   3156.90144
tau
 INTRCPT1,B0   1621.77842

 ----------------------------------------------------
  Random level-1 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1, G0                        0.888

Final estimation of fixed effects:
 ----------------------------------------------------------------------------
                                     Standard             Approx.
   Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------------------------------
For       INTRCPT1, B0
   INTRCPT2, G00         406.371906   2.672411   152.062       263     0.000
 ----------------------------------------------------------------------------

The outcome variables are:  PV1MATH, PV2MATH, PV3MATH, PV4MATH, PV5MATH

Final estimation of fixed effects
(with robust standard errors)
 ----------------------------------------------------------------------------
                                     Standard             Approx.
   Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------------------------------
For       INTRCPT1, B0
   INTRCPT2, G00         406.371906   2.667160   152.361       263     0.000
 ----------------------------------------------------------------------------

inal estimation of variance components:
 -----------------------------------------------------------------------------
Random Effect           Standard      Variance     df    Chi-square  P-value
                        Deviation     Component
 -----------------------------------------------------------------------------
INTRCPT1,       u0       40.27131    1621.77842   263    2915.45128    0.000
 level-1,       r        56.18631    3156.90144
```

Modelo 2_ Probando endogeneidad de nivel 2 en el modelo

```
  The maximum number of level-1 units = 5143
  The maximum number of level-2 units = 264
  The maximum number of iterations = 100

  Method of estimation: restricted maximum likelihood

  This is part of a plausible value analysis using the following variables:
     PV1MATH
     PV2MATH
     PV3MATH
     PV4MATH
     PV5MATH
Summary of the model specified (in hierarchical format)
 ---------------------------------------------------

Level-1 Model

        PV1MATH = B0 + B1*(ESCS) + B2*(MEANESCS) + B3*(GENERO) + r
Level-2 Model
        B0 = G00 + u0
        B1 = G10
        B2 = G20
        B3 = G30

ESCS has been centered around the group mean.
MEANESCS has been centered around the grand mean.



The outcome variables are:  PV1MATH, PV2MATH, PV3MATH, PV4MATH, PV5MATH

Final estimation of fixed effects
(with robust standard errors)
----------------------------------------------------------------------------
                                       Standard             Approx.
   Fixed Effect         Coefficient   Error       T-ratio   d.f.     P-value
----------------------------------------------------------------------------
For       INTRCPT1, B0
   INTRCPT2, G00          393.654328   2.100472   187.412       263     0.000
For      ESCS slope, B1
   INTRCPT2, G10           10.509855   0.875691    12.002       206     0.000
For MEANESCS slope, B2
   INTRCPT2, G20           34.543237   3.398258    10.165       846     0.000
For   GENERO slope, B3
   INTRCPT2, G30           36.079328   1.748800    20.631       423     0.000
----------------------------------------------------------------------------


Final estimation of variance components:
-----------------------------------------------------------------------------
Random Effect           Standard      Variance     df    Chi-square  P-value
                        Deviation     Component
-----------------------------------------------------------------------------
INTRCPT1,       u0       27.10816      734.85260   263   1527.46414    0.000
 level-1,       r        52.74142     2781.65760
-----------------------------------------------------------------------------
```

Modelo 3_Modelo con efectos aleatorios y valores plausibles

```
  The maximum number of level-2 units = 264
  The maximum number of iterations = 100

  Method of estimation: restricted maximum likelihood
  This is part of a plausible value analysis using the following variables:
     PV1MATH
     PV2MATH
     PV3MATH
     PV4MATH
     PV5MATH

 Summary of the model specified (in hierarchical format)
 ---------------------------------------------------
Level-1 Model

        PV1MATH = B0 + B1*(ESCS) + B2*(MEANESCS) + B3*(GENERO) + B4*(VOCACION) + r

Level-2 Model
        B0 = G00
        B1 = G10 + u1
        B2 = G20 + u2
        B3 = G30 + u3
        B4 = G40 + u4
Mixed Model
    PV1MATH = G00
    + G10*ESCS
    + G20*MEANESCS
    + G30*GENERO
    + G40*VOCACION
     + u1*ESCS   + u2*MEANESCS   + u3*GENERO
     + u4*VOCACION + r
THE AVERAGED RESULTS FOR THIS PLAUSIBLE VALUE RUN
 sigma^2 =    2754.28484

tau
     ESCS,B1       9.96164     -51.56253      -7.03549      -3.14822
 MEANESCS,B2     -51.56253     580.05781     106.20770     206.95475
   GENERO,B3      -7.03549     106.20770     205.97541     120.84355
 VOCACION,B4      -3.14822     206.95475     120.84355     387.41986

tau (as correlations)
     ESCS,B1  1.000 -0.678 -0.155 -0.051
 MEANESCS,B2 -0.678  1.000  0.307  0.437
   GENERO,B3 -0.155  0.307  1.000  0.428
 VOCACION,B4 -0.051  0.437  0.428  1.000
```

Random level-1 coefficient	Reliability estimate
ESCS, G1	0.055
MEANESCS, G2	0.420
GENERO, G3	0.211
VOCACION, G4	0.247

The outcome variables are: PV1MATH, PV2MATH, PV3MATH, PV4MATH, PV5MATH

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	428.206338	2.264651	189.083	67	0.000
For ESCS slope, B1					
INTRCPT2, G10	10.693850	0.927282	11.532	116	0.000
For MEANESCS slope, B2					
INTRCPT2, G20	29.401084	2.471030	11.898	177	0.000
For GENERO slope, B3					
INTRCPT2, G30	36.488139	1.911590	19.088	263	0.000
For VOCACION slope, B4					
INTRCPT2, G40	-1.356971	3.827986	-0.354	254	0.723

The outcome variables are: PV1MATH, PV2MATH, PV3MATH, PV4MATH, PV5MATH

Final estimation of fixed effects
(with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	428.206338	3.330328	128.578	317	0.000
For ESCS slope, B1					
INTRCPT2, G10	10.693850	0.920546	11.617	113	0.000
For MEANESCS slope, B2					
INTRCPT2, G20	29.401084	2.998653	9.805	263	0.000
For GENERO slope, B3					
INTRCPT2, G30	36.488139	1.798190	20.292	263	0.000
For VOCACION slope, B4					
INTRCPT2, G40	-1.356971	3.817040	-0.356	251	0.723

Modelo 4_Modelo Final.

```
The maximum number of level-1 units = 5021
The maximum number of level-2 units = 264
The maximum number of iterations = 100

Method of estimation: restricted maximum likelihood

The outcome variables are:  PV1MATH, PV2MATH, PV3MATH, PV4MATH, PV5MATH


Summary of the model specified (in hierarchical format)
---------------------------------------------------
Level-1 Model

        PV1MATH = B0 + B1*(ESCS) + B2*(MEANESCS) + B3*(GENERO) + B4*(VOCACION) + r

Level-2 Model
        B0 = G00 + G01*(MEANESCS) + G02*(TIPOESC) + u0
        B1 = G10
        B2 = G20 + G21*(TIPOESC)
        B3 = G30
        B4 = G40 + G41*(TIPOESC)

Mixed Model
    PV1MATH = G00 + G01*MEANESCS + G02*TIPOESC
    + G10*ESCS
    + G20*MEANESCS + G21*TIPOESC*MEANESCS
    + G30*GENERO
    + G40*VOCACION + G41*TIPOESC*VOCACION
     + u0+ r
THE AVERAGED RESULTS FOR THIS PLAUSIBLE VALUE RUN
 sigma^2 =    2761.16568


tau
 INTRCPT1,B0     549.71746

 Final estimation of fixed effects
 (with robust standard errors)
```

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	436.284007	3.954395	110.329	261	0.000
MEANESCS, G01	15.852550	7.470371	2.122	261	0.035
TIPOESC, G02	-22.395684	5.976559	-3.747	261	0.000
For ESCS slope, B1					
INTRCPT2, G10	10.525716	0.899390	11.703	134	0.000
For MEANESCS slope, B2					
INTRCPT2, G20	24.008589	8.354902	2.874	4019	0.004
TIPOESC, G21	-21.222611	5.738215	-3.698	598	0.000
For GENERO slope, B3					
INTRCPT2, G30	36.179268	1.704301	21.228	700	0.000
For VOCACION slope, B4					
INTRCPT2, G40	-21.467914	9.506690	-2.258	4751	0.024
TIPOESC, G41	26.213557	10.144764	2.584	4751	0.010

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	u0	23.44605	549.71746	261	1187.07520	0.000
level-1,	r	52.54680	2761.16568			

Modelo 5_Modelo con tres niveles

```
The maximum number of level-1 units = 2286
The maximum number of level-2 units = 118
The maximum number of level-3 units = 8
The maximum number of iterations = 100

  Method of estimation: full maximum likelihood
The outcome variables are:  PV1MATH, PV2MATH, PV3MATH, PV4MATH, PV5MATH
Summary of the model specified (in equation format)
 ---------------------------------------------------
Level-1 Model
        PV1MATH = P0 + P1*(ESCS) + P2*(MEANESCS) + e
Level-2 Model
        P0 = B00 + B01*(TIPOESC) + r0
        P1 = B10
        P2 = B20
Level-3 Model
        B00 = G000 + u00
        B01 = G010 + u01
        B10 = G100
        B20 = G200
```

Final estimation of fixed effects
(with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, P0					
For INTRCPT2, B00					
INTRCPT3, G000	460.333107	2.247301	204.838	7	0.000
For TIPOESC, B01					
INTRCPT3, G010	-20.389450	7.287194	-2.798	7	0.027
For ESCS slope, P1					
For INTRCPT2, B10					
INTRCPT3, G100	11.679898	0.952996	12.256	332	0.000
For MEANESCS slope, P2					
For INTRCPT2, B20					
INTRCPT3, G200	22.745985	5.880098	3.868	1656	0.000

Final estimation of level-1 and level-2 variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	r0	24.75695	612.90653	76	423.34786	0.000
level-1,	e	54.99736	3024.71007			

Note: The chi-square statistics reported above are based on only 3 of 8 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

Final estimation of level-3 variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1/INTRCPT2,	u00	1.12145	1.25766	2	0.89394	>.500
INTRCPT1/ TIPOESC,	u01	1.31096	1.71861	2	0.73078	>.500

Modelo 6_Modelo con variables del estudiante y efectos aleatorios

```
  The maximum number of level-1 units = 5021
  The maximum number of level-2 units = 264
  The maximum number of iterations = 100

  Method of estimation: full maximum likelihood

  The outcome variable is  PV1MATH

Summary of the model specified (in hierarchical format)
 ---------------------------------------------------

Level-1 Model

        PV1MATH = B0 + B1*(ESCS) + B2*(MEANESCS) + B3*(GENERO) + B4*(VOCACION) + r

Level-2 Model
        B0 = G00
        B1 = G10 + u1
        B2 = G20 + u2
        B3 = G30 + u3
        B4 = G40 + u4

Mixed Model
    PV1MATH = G00
    + G10*ESCS
    + G20*MEANESCS
    + G30*GENERO
    + G40*VOCACION
     + u1*ESCS  + u2*MEANESCS  + u3*GENERO
     + u4*VOCACION + r

Iterations stopped due to small change in likelihood function
******* ITERATION 23 *******


 sigma^2 =   2723.41617

 Standard error of sigma^2 =     58.17976

tau
     ESCS,B1      5.36521    -39.88966     -0.64493      9.66616
 MEANESCS,B2    -39.88966    572.86403     96.23744    160.45784
   GENERO,B3     -0.64493     96.23744    210.33652    -58.99031
 VOCACION,B4      9.66616    160.45784    -58.99031    495.07783


 ----------------------------------------------------
  Random level-1 coefficient   Reliability estimate
 ----------------------------------------------------
      ESCS, G1                        0.031
  MEANESCS, G2                        0.420
    GENERO, G3                        0.217
  VOCACION, G4                        0.295
 ----------------------------------------------------
Note: The reliability estimates reported above are based on only 29 of 264
units that had sufficient data for computation.  Fixed effects and variance
components are based on all the data.
The value of the log-likelihood function at iteration 23 = -2.723222E+004
 The outcome variable is  PV1MATH
```

```
   Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
----------------------------------------------------------------------------
For       INTRCPT1, B0
   INTRCPT2, G00         429.842247   1.977965   217.315      3964     0.000
For      ESCS slope, B1
   INTRCPT2, G10          11.004293   0.820892    13.405       263     0.000
For MEANESCS slope, B2
   INTRCPT2, G20          30.854623   2.275627    13.559       263     0.000
For   GENERO slope, B3
   INTRCPT2, G30          36.567438   1.859162    19.669       263     0.000
For VOCACION slope, B4
   INTRCPT2, G40          -0.114787   3.725639    -0.031       263     0.975
----------------------------------------------------------------------------

The outcome variable is  PV1MATH

Final estimation of fixed effects
(with robust standard errors)
----------------------------------------------------------------------------
                                      Standard             Approx.
   Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
----------------------------------------------------------------------------
For       INTRCPT1, B0
   INTRCPT2, G00         429.842247   3.188747   134.800      3964     0.000
For      ESCS slope, B1
   INTRCPT2, G10          11.004293   0.821149    13.401       263     0.000
For MEANESCS slope, B2
   INTRCPT2, G20          30.854623   2.868695    10.756       263     0.000
For   GENERO slope, B3
   INTRCPT2, G30          36.567438   1.733572    21.094       263     0.000
For VOCACION slope, B4
   INTRCPT2, G40          -0.114787   3.726990    -0.031       263     0.975
----------------------------------------------------------------------------


Final estimation of variance components:
-----------------------------------------------------------------------------
Random Effect           Standard      Variance     df    Chi-square  P-value
                        Deviation     Component
-----------------------------------------------------------------------------
    ESCS,         u1      2.31629       5.36521    28      23.19585    >.500
MEANESCS slope, u2       23.93458     572.86403    28      49.06807    0.008
  GENERO slope, u3       14.50298     210.33652    28      17.52243    >.500
VOCACION slope, u4       22.25034     495.07783    28      39.92388    0.067
 level-1,         r      52.18636    2723.41617
-----------------------------------------------------------------------------

Statistics for the current model
---------------------------------------------------
Deviance                           = 54464.437687
Number of estimated parameters = 16
```

Modelo 7_Modelo Final Full MV sin plausibles.

```
 The maximum number of level-1 units = 5021
 The maximum number of level-2 units = 264
 The maximum number of iterations = 100

 Method of estimation: full maximum likelihood

 The outcome variable is  PV1MATH


Summary of the model specified (in hierarchical format)
---------------------------------------------------

Level-1 Model

        PV1MATH = B0 + B1*(ESCS) + B2*(MEANESCS) + B3*(GENERO) + B4*(VOCACION) + r

Level-2 Model
        B0 = G00 + G01*(MEANESCS) + G02*(TIPOESC) + u0
        B1 = G10
        B2 = G20 + G21*(TIPOESC)
        B3 = G30
        B4 = G40 + G41*(TIPOESC)


Mixed Model
    PV1MATH = G00 + G01*MEANESCS + G02*TIPOESC
    + G10*ESCS
    + G20*MEANESCS + G21*TIPOESC*MEANESCS
    + G30*GENERO
    + G40*VOCACION + G41*TIPOESC*VOCACION
     + u0+ r

Iterations stopped due to small change in likelihood function
******* ITERATION 7 *******

sigma^2 =   2724.81049

 Standard error of sigma^2 =     55.84576

tau
 INTRCPT1,B0    564.13823

Standard error of tau
 INTRCPT1,B0     63.23161
```

Random level-1 coefficient	Reliability estimate
INTRCPT1, G0	0.771

The value of the log-likelihood function at iteration 7 = -2.718566E+004
The outcome variable is PV1MATH
Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	437.061774	4.083957	107.019	261	0.000
MEANESCS, G01	18.195419	4.567623	3.984	261	0.000
TIPOESC, G02	-21.291282	5.866930	-3.629	261	0.000
For ESCS slope, B1					
INTRCPT2, G10	10.815047	0.802206	13.482	4751	0.000
For MEANESCS slope, B2					
INTRCPT2, G20	23.737943	6.713094	3.536	4751	0.000
TIPOESC, G21	-21.512858	5.922180	-3.633	4751	0.000
For GENERO slope, B3					
INTRCPT2, G30	36.077720	1.597813	22.579	4751	0.000
For VOCACION slope, B4					
INTRCPT2, G40	-19.688042	10.321698	-1.907	4751	0.057
TIPOESC, G41	26.060792	10.891669	2.393	4751	0.017

The outcome variable is PV1MATH Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	437.061774	3.950440	110.636	261	0.000
MEANESCS, G01	18.195419	7.752003	2.347	261	0.020
TIPOESC, G02	-21.291282	6.042222	-3.524	261	0.001
For ESCS slope, B1					
INTRCPT2, G10	10.815047	0.802312	13.480	4751	0.000
For MEANESCS slope, B2					
INTRCPT2, G20	23.737943	8.717172	2.723	4751	0.006
TIPOESC, G21	-21.512858	5.870828	-3.664	4751	0.000
For GENERO slope, B3					
INTRCPT2, G30	36.077720	1.657687	21.764	4751	0.000
For VOCACION slope, B4					
INTRCPT2, G40	-19.688042	10.465814	-1.881	4751	0.060

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	u0	23.75159	564.13823	261	1242.17756	0.000
level-1,	r	52.19972	2724.81049			

Statistics for the current model

Deviance = 54371.326870
Number of estimated parameters = 11

Model comparison test

Chi-square statistic = 93.11082
Degrees of freedom = 5
P-value = <0.001

Modelo 8_Test de Homocedasticidad Nivel 1

```
Statistics for current covariance components model
--------------------------------------------------
Deviance                       = 54365.434061
Number of estimated parameters = 2

Test of homogeneity of level-1 variance
----------------------------------------
Chi-square statistic   =    266.75429
degrees of freedom     =    261
P-value                =      0.390
```

Test de Normalidad sobre los residuos de nivel 1

Skewness/Kurtosis tests for Normality

Variable	Pr(Skewness)	Pr(Kurtosis)	adj chi2(2)	joint Prob>chi2
l1resid	**0.463**	**0.544**	**0.91**	**0.6348**

Test de comparación de Modelos (A partir del modelo estimado por Full Maxima Verosimilitud)

```
Statistics for the current model
--------------------------------------------------
Deviance                       = 54371.326870
Number of estimated parameters = 11

Model comparison test
----------------------------------
Chi-square statistic   =     93.11082
Degrees of freedom     =      5
P-value                =      <0.001
```

Modelo 9_HGLM nulo

```
 The maximum number of level-1 units = 5021
 The maximum number of level-2 units = 264
 The maximum number of micro iterations = 14

 Method of estimation: restricted PQL
 Maximum number of macro iterations = 100

 Distribution at Level-1: Bernoulli

 The outcome variable is REM_MATE


Summary of the model specified (in hierarchical format)
---------------------------------------------------

Level-1 Model

        Prob(Y=1|B) = P

        log[P/(1-P)] = B0

Level-2 Model
        B0 = G00 + u0

Level-1 variance = 1/[P(1-P)]


Mixed Model
    eta = G00  + u0
The value of the log-likelihood function at iteration 11 = -2.896043E+003
RESULTS FOR NON-LINEAR MODEL WITH THE LOGIT LINK FUNCTION:
 Unit-Specific Model, PQL Estimation - (macro iteration 6)

tau
 INTRCPT1,B0      0.24723

 ----------------------------------------------------
  Random level-1 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1, G0                         0.444
 ----------------------------------------------------
The value of the log-likelihood function at iteration 2 = -7.070569E+003
 The outcome variable is REM_MATE

 Final estimation of fixed effects: (Unit-specific model)



                                      Standard             Approx.
   Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------------------------------
 For       INTRCPT1, B0
   INTRCPT2, G00         -1.103414   0.045909   -24.035       263    0.000
 ----------------------------------------------------------------------------

 ----------------------------------------------------------------------
                                           Odds          Confidence
   Fixed Effect         Coefficient        Ratio         Interval
 ----------------------------------------------------------------------
 For       INTRCPT1, B0
   INTRCPT2, G00         -1.103414       0.331736     (0.303,0.363)
 ----------------------------------------------------------------------
```

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	u0	0.49722	0.24723	263	488.77030	0.000

RESULTS FOR POPULATION-AVERAGE MODEL

The value of the log-likelihood function at iteration 2 = -7.050352E+003
The outcome variable is REM_MATE

Final estimation of fixed effects: (Population-average model)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	-1.075678	0.045394	-23.696	263	0.000

Fixed Effect	Coefficient	Odds Ratio	Confidence Interval
For INTRCPT1, B0			
INTRCPT2, G00	-1.075678	0.341067	(0.312,0.373)

The outcome variable is REM_MATE

Final estimation of fixed effects
(Population-average model with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	-1.075678	0.044936	-23.938	263	0.000

Fixed Effect	Coefficient	Odds Ratio	Confidence Interval
For INTRCPT1, B0			
INTRCPT2, G00	-1.075678	0.341067	(0.312,0.373)

Modelo 10_Modelo HGLM

```
  The maximum number of level-1 units = 5021
  The maximum number of level-2 units = 264
  The maximum number of micro iterations = 14
Method of estimation: restricted PQL
  Maximum number of macro iterations = 100

  Distribution at Level-1: Bernoulli

  The outcome variable is REM_MATE
Summary of the model specified (in hierarchical format)
 ---------------------------------------------------

Level-1 Model

        Prob(Y=1|B) = P

        log[P/(1-P)] = B0 + B1*(ESCS) + B2*(GENERO)

Level-2 Model
        B0 = G00 + G01*(MEANESCS) + u0
        B1 = G10
        B2 = G20

Level-1 variance = 1/[P(1-P)]


Mixed Model
    eta = G00 + G01*MEANESCS
    + G10*ESCS
    + G20*GENERO
     + u0
The value of the log-likelihood function at iteration 11 = -2.883599E+003
RESULTS FOR NON-LINEAR MODEL WITH THE LOGIT LINK FUNCTION:
  Unit-Specific Model, PQL Estimation - (macro iteration 10)

tau
 INTRCPT1,B0      0.19445

 ----------------------------------------------------
  Random level-1 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1, G0                         0.388
 ----------------------------------------------------
The value of the log-likelihood function at iteration 2 = -7.079619E+003
 The outcome variable is REM_MATE
```

The outcome variable is REM_MATE

Final estimation of fixed effects: (Unit-specific model)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	-1.515379	0.077210	-19.627	262	0.000
MEANESCS, G01	-0.237867	0.063362	-3.754	262	0.000
For ESCS slope, B1					
INTRCPT2, G10	-0.085161	0.035172	-2.421	4755	0.016
For GENERO slope, B2					
INTRCPT2, G20	0.232727	0.068624	3.391	4755	0.001

Fixed Effect	Coefficient	Odds Ratio	Confidence Interval
For INTRCPT1, B0			
INTRCPT2, G00	-1.515379	0.219725	(0.189,0.256)
MEANESCS, G01	-0.237867	0.788308	(0.696,0.893)
For ESCS slope, B1			
INTRCPT2, G10	-0.085161	0.918364	(0.857,0.984)
For GENERO slope, B2			
INTRCPT2, G20	0.232727	1.262037	(1.103,1.444)

The outcome variable is REM_MATE

Final estimation of fixed effects
(Unit-specific model with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	-1.515379	0.080078	-18.924	262	0.000
MEANESCS, G01	-0.237867	0.060723	-3.917	262	0.000
For ESCS slope, B1					
INTRCPT2, G10	-0.085161	0.035342	-2.410	4755	0.016
For GENERO slope, B2					
INTRCPT2, G20	0.232727	0.069491	3.349	4755	0.001

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	u0	0.44097	0.19445	262	434.39611	0.000

RESULTS FOR POPULATION-AVERAGE MODEL
The value of the log-likelihood function at iteration 2 = -7.091699E+003
The outcome variable is REM_MATE

Final estimation of fixed effects: (Population-average model)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	-1.488339	0.076491	-19.458	262	0.000
MEANESCS, G01	-0.233030	0.062971	-3.701	262	0.000
For ESCS slope, B1					
INTRCPT2, G10	-0.083463	0.034744	-2.402	4755	0.016
For GENERO slope, B2					
INTRCPT2, G20	0.229298	0.067908	3.377	4755	0.001

Fixed Effect	Coefficient	Odds Ratio	Confidence Interval
For INTRCPT1, B0			
INTRCPT2, G00	-1.488339	0.225747	(0.194,0.262)
MEANESCS, G01	-0.233030	0.792130	(0.700,0.897)
For ESCS slope, B1			
INTRCPT2, G10	-0.083463	0.919926	(0.859,0.985)
For GENERO slope, B2			
INTRCPT2, G20	0.229298	1.257717	(1.101,1.437)

The outcome variable is REM_MATE
Final estimation of fixed effects
(Population-average model with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	-1.488339	0.079210	-18.790	262	0.000
MEANESCS, G01	-0.233030	0.059981	-3.885	262	0.000
For ESCS slope, B1					
INTRCPT2, G10	-0.083463	0.034485	-2.420	4755	0.016
For GENERO slope, B2					
INTRCPT2, G20	0.229298	0.067991	3.372	4755	0.001