

Arteaga, Paredes²: Elección de colegios y segregación

Francisco Gallego
Instituto de Economía PUC-Chile y J-PAL

CEP, 11 de diciembre de 2010

El paper

- Paper empírico. Idea central: Estudiar la decisión de elección de colegio de las familias y con esas preferencias cuantificar la importancia de diferentes factores en explicar la segregación educaciones en Chile.
- Estrategia empírica. Datos de estudiantes secundarios 2006-2009. Usar información de decisiones (preferencias reveladas) y características de colegios en el contexto de un modelo que permite explotar heterogeneidad observable e inobservable de las familias. Además innovan definiendo de un modo diferente el "choice set".
- Conclusiones: Similar a la literatura previa que usa modelos de elección discreta: roles de distancia, precios y "calidad". Educación secundaria \Rightarrow Modalidad (TP vs CH importa).

- Novedad: modelación de heterogeneidad no observable (algo) más flexible. Dudas sobre choice-set.
- No hay un buen tratamiento de temas de causalidad.
- Potencialmente útil pero falta mucho para derivar implicancias para política públicas.

Mis Comentarios

1. No fácil: el paper y la audiencia
2. Comentarios Académicos Centrales y Sugerencias
3. Comentarios Académicos menores
4. Implicancias en Políticas Públicas

El paper y la audiencia

1. ¿Cómo leo este paper? Como académico.
2. ¿Cómo la audiencia lo ve? Como herramienta de políticas públicas.
3. Mis comentarios apunta a 1. e implican que para 2. podemos aprender poco todavía.
4. Mencionaré muy brevemente 1. (los autores ya tienen los comentarios) y argumentaré respecto de 2.

Comentarios Académicos

- **Me centraré fundamentalmente en el modelo de elección de colegios y no en las simulaciones de efectos sobre segregación.**
- Modelo de elección discreta: identificar las preferencias profundas de las personas sobre la base de sus elecciones (preferencias reveladas).
Clave: (i) atributos de las elecciones, (ii) características de las personas, (iii) estrategia de identificación (causalidad).
- "Sus" elecciones:

- Los colegios pueden elegir (especialmente relevante en media)→ Autores agregan indicadores de selección en los colegios como atributo de elección. Atajo complicado conceptualmente.
- * Modelo "verdadero" tiene elección por demanda y oferta. Altamente no lineal. Aproximación del trabajo demasiado simple: mejor interactuar atributos del colegio valorados por la familia con estos "controles"
- * ¿Interpretación causal de la valoración de las familias por estos atributos de selección? Muy complicado. Muestra que debe haber variables omitidas.
- Comportamiento estratégico de los padres: ¿qué está en su choice set? ¿a dónde postulan?

- * Diferentes fuentes apuntan a que postulan a muy pocos colegios y que son rechazados en muy pocas partes (referencias en tesis de Felipe)...
- * ...y al mismo tiempo hay un excesivo cambio de establecimientos (en básica). Datos administrativos sugieren que en un año (entre primero y sexto) en promedio 12% de los niños se cambian de colegio y cerca de un 40% ha pasado por más de un colegio (Larroulet, 2012).
- * Demasiado parecido a los resultados de Alvin Roth, Parag Pathak y otros en elecciones de padres en sistemas centralizados que permiten comportamiento estratégico.
- * Modelo dinámico muy complejo de estimar e interpretar (Grau, 2014 lo intenta).

⇒ ¿Qué identifican los autores? Forma "semi-estructural" (comentarios de Tartari, 2009 a Gallego y Hernando, 2009), pero para eso necesitamos identificar bien los parámetros (causalidad). Sino sujeto a crítica de Lucas y por tanto los ejercicios contrafactuales que se hagan no informan de efectos potenciales de políticas

– Innovación sobre el choice set:

- * Idea interesante pero con problemas conceptuales: ¿por qué hacer esto en dos etapas?. Relacionado con el trabajo de Chumacero, Gómez y Paredes (2011) y con otros trabajos.
- * Econométricamente: ¿qué ganamos con el procedimiento que un método bien hecho no identifique automáticamente? ¿Ganamos consistencia? No es nada de obvio. Ganamos eficiencia, puede

ser...PERO tenemos que testear formalmente. En otros contextos se aplica el test de McFadden y Hausmann para comparar estimaciones con diferentes alternativas.

- * Académicamente: ¿cuáles son las alternativas? ¿cuáles son los supuestos? ¿bajo qué condiciones son plausibles? En general la investigación en ciencias sociales no es como la ingeniería: yo no quiero soluciones a los problemas, quiero entender los supuestos de los métodos, entender las alternativas y entender qué gano con cada una de ellas.
- * De hecho si uno mira los resultados de las tesis de Felipe cuando compara modelo con choice set restringido y modelo con todas las opciones, los resultados parecen ser muy parecidos. ¿Ganamos eficiencia entonces?

- Latent class logit: modelo que permite identificar "clases" de individuos con preferencias diferentes. Las clases son desconocidas pero se supone que dependen de atributos de las personas. En este caso se usa log del ingreso y educación de los padres: ¿Por qué sólo eso? ¿Alternativas? ¿Por qué no poner interacciones? ¿Cómo interpretamos a cada grupo?
- * ¿Por qué domina en términos empíricos a alternativas? No se cita literatura de econometría que compara modelos, por ejemplo recién esta semana salió Gregolo y Verboven (2014, Restat). Vuelvo a insistir: mejor presentar modelos con supuestos alternativos y mostrar en qué y por qué cambian los resultados.
- * ¿Cuál es el costo de suponer iguales preferencias entre pocas clases? Solución: volver a interactuar en la segunda etapa. Pero, ¿es suficiente? ¿es mejor? ¿por qué no usar modelos de utilidad

aleatoria?: El componente no observables de valoración de preferencias es probablemente clave y eso sería una innovación en la estimación para Chile (comparar con Grau).

- * CLAVE: Explicar esto en el paper (hay que ver la tesis de Felipe) e interpretar los coeficientes dentro de cada clase que son muy diferentes. ¿Tienen sentido los valores estimados? ¡¡¡Elasticidad precio positiva en modelos de personas de ingresos altos!! Obviamente hay variable omitida.
 - * ¿Potencial simulación? ¿qué pasa si las preferencias de la clase 5 se aplican al resto de las clases? Más en general, ¿cuánto de la variación cambia por variación de preferencias entre clases?
- Características de las opciones y de las personas: ¿por qué raíz de

distancia? (efecto no lineal, pero por qué?). Efecto pares? (Gallego y Hernando, 2008), otras?

- Estrategia de identificación: básicamente es por forma(s) funcional(es). Hay varias cosas que afectan variables omitidas relevantes y que afectan varios cálculos. Por ej. (i) elasticidad precio positiva para algunos grupos (correlación positiva con no-observables), (ii) eliminar selección empeora la segregación (en tesis de Felipe).
 - Solución: variables instrumentales. Sin embargo, algunos ejercicios de robustez en la tesis de Felipe sugieren que en el caso de copago no es sólo esto (usando la estrategia que usamos en Gallego y Hernando, 2008). ¿Cómo son las primeras etapas?

- Error de medición en otros atributos también puede producir sesgo de atenuación. Por ejemplo, distancia puede estar medida con error.

Comentarios Menores

- Paper hay que escribirlo mucho mejor. Para entender partes muy importantes hay que ir a la tesis de Felipe.
- Citar literatura internacional relevante: Hastings, Kaine, Staiger (2006) para EEUU; Carneiro, Das, Reis (2010) para Pakistan; **Grau (2014)**; Neilson (2014) para Chile, otros.
 - Notar que todos los papers que usan modelos estructurales disponibles: Hastings et al. (2005, 2006), Bayer y McMillan (2005); Gallego y Hernando (2008) tienen bastantes años y no hemos podido publicarlos en revistas buenas.

- Los efectos marginales estimados tienen demasiadas estrellas. ¿Están bien calculados los errores estándar? ¿reconocen estructurales grupales?
- Un par de cosas sobre simulaciones de segregación:
 - ¿La asignación de niños a colegios en las simulaciones es robusta en el sentido de Gale-Shapley? En Gallego-Hernando (2008b) esto es importante.
 - ¿Cómo se trata la selección de los colegios?
 - Agregar errores estándar es clave (no es fácil, pero es posible).

Implicancias para Políticas Públicas

- ¿Qué podemos aprender para informar políticas públicas?
- Las simulaciones son informativas pero no de cambios de políticas. No sabemos si los efectos son causales ni si pueden surgir otros efectos de equilibrio general.
- **Una analogía:** en macroeconomía hacemos descomposiciones del crecimiento entre contribución de factores (capital físico, capital humano, trabajo, productividad). No hay noción de causalidad subyacente pero puede informar (Así varios economistas notaron que el milagro de varios tigres asiáticos en los 90s ya no era tal; también así hemos notado cosas que han sucedido en Chile)

- Este paper puede ayudar en esta línea: ¿podemos descomponer "contablemente" factores correlacionados con la segregación? Creo que sí. Esta evidencia confirma que hay bastante heterogeneidad de preferencias que explica parte importante de la asignación de estudiantes a colegios y por ello es difícil atribuir a una cosa la segregación, etc. Esto es relevante para entender lo que sucede en equilibrio.
- Una interpretación (plausible) de la heterogeneidad en preferencias es que las familias tienen problemas importantes en decisiones de educación. Esto es consistente con mucha evidencia disponible en el mundo (www.povertyactionlab.org) y sugiere que mejoras importantes en las políticas de admisión y selección (y, con ello, en la segregación) requerirán mucho más trabajo que simplemente prohibirlas o llevar a un sistema centralizado (ausente en el proyecto actual).

Conclusiones

- Trabajo analiza un área importante del estudio del sistema educacional chileno.
- Trabajo con alto potencial académico luego de trabajo adicional.
- Trabajo presenta evidencia descriptiva sugestiva e interesante, que conviene estudiar en detalle adicionalmente.