

**El acceso a los microdatos administrativos públicos: la nueva frontera  
de la investigación económica y social**

**Manuel Arellano**  
CEMFI

XVIII Aula de Verano "Ortega y Gasset"  
UIMP, Santander, 20-21 Agosto 2018

## Introducción

- La recogida y mantenimiento de bases de datos es una parte integral de la actividad de las administraciones públicas y los organismos supervisores.
- Por ejemplo, los registros de afiliados a la Seguridad Social, los de contribuyentes de la Agencia Tributaria, o los de préstamos en el Banco de España.
- También mantienen grandes bases de datos las administraciones públicas de Educación, Sanidad y Justicia, entre otras.

- Esta actividad no solo no es nueva sino que está en los orígenes remotos de la propia escritura.
- La escritura cuneiforme de los sumerios no narra hechos épicos o literarios sino que se trata de registros burocráticos sobre bienes pagados, raciones entregadas a trabajadores o productos agrícolas distribuidos.



- Lo que es nuevo es el crecimiento exponencial de los últimos años en la capacidad de almacenamiento de datos digitalizados y el desarrollo de las tecnologías de acceso y gestión de gigantescos almacenes de información digital.
- Microchips en la nube en vez de legajos en los sótanos.

- Un ejemplo singular del almacenamiento de documentos administrativos en nuestra historia es el Archivo General de Indias.



- Las nuevas tecnologías de la información han creado oportunidades de expandir el conocimiento en formas inimaginables hasta hace poco tiempo.
- A medida que se ha desarrollado la capacidad para almacenar y analizar datos, los gobiernos del mundo han creado archivos de datos a gran escala sobre tributación, empleo, programas asistenciales, salud y educación.
- Aunque estos datos se archivan electrónicamente con fines puramente administrativos, ahora es técnicamente posible utilizarlos también para el análisis estadístico.
- El acceso a los datos administrativos anonimizados también supone un control de nuevo tipo sobre la calidad y limpieza de la administración pública y en definitiva una profundización de la democracia.

## Plan de la presentación

### *Los datos administrativos en la investigación*

- La economía como ciencia empírica
- Sistemas de acceso a los datos administrativos
- Problemas
- La situación en España

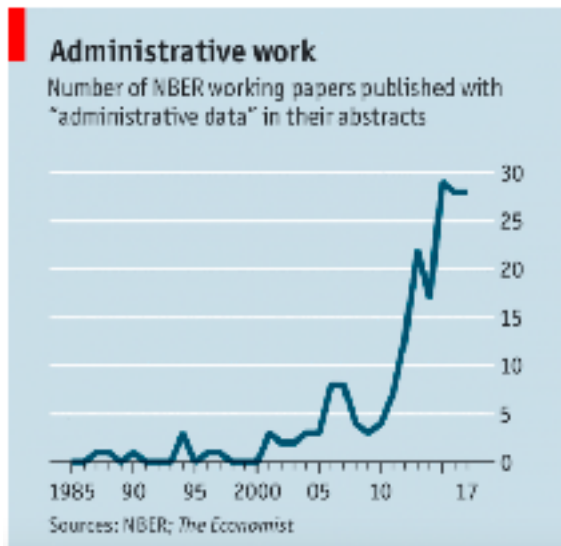
### *Ilustraciones*

- Madurez y rendimiento escolar
- El impacto de la política monetaria sobre el riesgo que asumen los bancos en su concesión de préstamos
- El efecto a largo plazo de los buenos maestros
- Los factores biológicos en la desigualdad intergeneracional de la renta
- Barrios étnicos y éxito de los inmigrantes

## La economía como ciencia empírica

- Las mejoras masivas en la computación han convertido a la “ciencia lúgubre” en una ciencia cada vez más empírica, a la vez que se desdibujan los contornos entre la economía y las otras ciencias sociales.
- La proporción de artículos puramente teóricos en las principales revistas de las ciencias económicas disminuyó del 58% en 1983 al 19% en 2011 (Hamermesh 2013).
- La investigación empírica incluye trabajos que utilizan datos disponibles públicamente, como las encuestas de hogares, los que utilizan datos experimentales, y los que se basan en datos administrativos conseguidos a través de una compleja negociación.
- Estos últimos son los que actualmente dominan la investigación puntera.
- En el siglo XX dominaban los estudios con datos de encuestas norteamericanas.
- Actualmente, la disponibilidad diferencial de datos administrativos ha llevado a que se estudien países europeos como Suecia, Noruega, Dinamarca, Alemania o Austria.
- Últimamente las infraestructuras de acceso y cruce de datos administrativos se están desarrollando en EE.UU. y en otros países como Portugal, pero no así en España.

- Según *The Economist*,<sup>1</sup> al menos 28 artículos publicados por el NBER en 2017 utilizaron datos administrativos. Antes de 2000 casi ninguno lo hizo.



<sup>1</sup>“Government data are ever more important to economic research,” *The Economist*, May 26th 2018.



## La economía como ciencia empírica (cont.)

- Las áreas en las que los datos administrativos son más utilizados incluyen:
  - Mercado de trabajo
  - Salud
  - Educación
  - Patentes/Innovación
  - Creación de empresas y su crecimiento
  
- Ejemplos de datos administrativos archivados en el dominio público:
  - Administración de la Seguridad Social
  - Agencia Tributaria – datos de ingresos de personas y empresas
  - Administración de la Salud – servicios recibidos por los beneficiarios
  - Administración de las escuelas – historiales educativos, clases, profesores, notas.
  - Registros de las transacciones inmobiliarias.

## *Ventajas de los datos administrativos*

- Los datos administrativos ofrecen tamaños muestrales mucho más grandes y presentan menores errores de medida y no respuesta que los datos de encuestas.
- Los datos administrativos son esenciales para la evaluación de las políticas públicas.  
Ejemplos:
  - Efectos de las prestaciones por desempleo.
  - Efectos de los impuestos sobre la oferta de trabajo.
  - Efectos de las políticas medioambientales sobre la salud y la mortalidad utilizando registros de nacimientos y fallecimientos y registros hospitalarios.
- Los archivos administrativos tienen una estructura longitudinal que permite seguir a los individuos a lo largo del tiempo.
- Y así estudiar cuestiones de relevancia política como las consecuencias a largo plazo de la pérdida del empleo o el grado de movilidad de la renta durante la vida.

## La economía como ciencia empírica (cont.)

- La mera descripción de patrones básicos en los datos administrativos puede tener profundas consecuencias en el debate público. Einav y Levin (2014) llaman la atención sobre dos casos notables:
- Piketty y Saez (2003) utilizaron datos fiscales del IRS<sup>2</sup> para calcular series históricas de la participación en la renta total de los perceptores en los percentiles superiores de los hogares en EE.UU.
- Este trabajo ha tenido una gran influencia al convertir el aumento de la fracción de renta detentada por los mayores perceptores en el centro de los debates sobre la desigualdad económica.
- John Wennberg y co-autores utilizaron grandes muestras de prestaciones de Medicare<sup>3</sup> para demostrar que existe una gran variación en el gasto por afiliado que no se puede atribuir a diferencias en el estado de salud o a los precios.
- Esta investigación recibió gran atención durante el debate sobre la Affordable Care Act en 2009 y se ha convertido en la evidencia principal sobre la ineficiencia del sistema de salud en EE.UU. (*Dartmouth Atlas of Health Care*).

---

<sup>2</sup>El IRS (Internal Revenue Service) es la agencia tributaria de los Estados Unidos.

<sup>3</sup>Medicare es un seguro médico público para personas mayores de 65 años o discapacitadas.

## Metodología estadística

- Los datos administrativos difieren de otros datos, ya que no se recogen originalmente para fines de investigación.
- Los investigadores no suelen participar en el diseño, la estructura y el contenido de los datos administrativos.
- Estas bases de datos son grandes, aunque no tanto como los tipos de *big data* recogidos en las redes sociales, los datos de texto, el rastreo por GPS o las transacciones de supermercados.
- Por lo general estos datos son más complejos que las encuestas a las que los investigadores sociales están acostumbrados. Connelly et al (2016) introducen estas cuestiones desde la perspectiva de la investigación sociológica.
- Los datos administrativos a menudo están desestructurados y su uso requiere inventiva y una gestión sustancial para organizarlos en el formato requerido para el análisis.
- A menudo se requieren grandes equipos de investigadores en una labor colaborativa que se asemeja a la de los laboratorios de las ciencias naturales.
- Métodos de la estadística, la econometría y las ciencias de la computación (machine learning): una síntesis en ciernes.

## Sistemas de acceso a los datos administrativos

### *La experiencia nórdica*

- En Dinamarca, Statistics Denmark proporciona datos anonimizados que combinan información de distintas bases de datos administrativas para proyectos de investigación aprobados.
- Los datos pueden ser accedidos por los investigadores en remoto desde cualquier ordenador a través de un servidor seguro.
- Los investigadores solicitan el acceso a los datos por medio de centros acreditados en las principales universidades, y el acceso se proporciona por medio de un proceso competitivo basado en el mérito científico.

## *Datos anonimizados*

- Como atestiguan numerosas experiencias internacionales, es posible proporcionar acceso seguro a los datos administrativos anonimizados a los investigadores.
- Hasta cierto punto se trata de compartir la custodia de estos datos ejercida hasta ahora exclusivamente por los funcionarios responsables de las agencias respectivas con los miembros de la comunidad científica.
- Es posible mantener los estándares más exigentes de privacidad a la vez que se permite a los investigadores el acceso directo a los registros individuales.

## Sistemas de acceso a los datos administrativos (cont.)

- Según Card, Chetty, Feldstein y Saez (2010) en su libro blanco dirigido al NSF,<sup>4</sup> un programa de acceso a los datos que sea sostenible y eficiente debe reunir las condiciones siguientes:
  - 1) Competencia abierta por el acceso a los datos basada en el mérito científico.
  - 2) Suficiente ancho de banda para acomodar proyectos múltiples de forma simultánea.
  - 3) Inclusión de investigadores jóvenes y estudiantes de postgrado en los equipos de investigadores con acceso a los datos.
  - 4) Modo de acceso: Acceso directo a los micro datos anonimizados in situ o preferiblemente a través de conexiones remotas.
    - El acceso a datos procesados en vez de los registros brutos o el envío de códigos ejecutables a los funcionarios de las agencias no resuelve el problema.
    - La segunda alternativa elimina en la práctica la posibilidad del análisis exploratorio interactivo de los datos.

---

<sup>4</sup>El NSF (National Science Foundation) es la agencia nacional de investigación de los Estados Unidos.

## Sistemas de acceso a los datos administrativos (cont.)

### *Cruce de datos*

- En algunos países del norte de Europa, como Noruega, Finlandia y Suecia, ya existe una infraestructura eficaz para vincular e intercambiar datos administrativos con fines de investigación.
  - Por ejemplo, enlaces de registros escolares con datos de la vida laboral.
- La fusión de datos es muy valiosa para la investigación. Es más fácil de conseguir con una agencia centralizada (modelo danés). En el caso de agencias múltiples (EE.UU.) requiere de la cooperación.
- La disponibilidad de números de identificación únicos (como el DNI) en los países nórdicos ha sido una gran ventaja.
- En países como el Reino Unido y EE.UU., la falta de números de identificación consistentes en todos los sistemas administrativos dificulta los cruces.
- Compartir datos entre agencias puede requerir la reforma de los requisitos legales de confidencialidad.



## Sistemas de acceso a los datos administrativos (cont.)

### *Sostenibilidad del acceso a los micro datos administrativos en la perspectiva política*

- La divulgación es esencial. La mayoría de los gestores y ciudadanos –incluyendo muchos economistas– no son conscientes de los beneficios de la investigación que utiliza datos administrativos.
- Hay campos como la medicina o la química en los que los avances generados en otros países pueden ser fácilmente importables.
- En el caso de la investigación social no es necesariamente así. Cada sociedad tiene particularidades importantes que resultan de su historia y de la inercia de sus instituciones.
- Las propias agencias gubernamentales podrían hacer un uso más efectivo de sus datos como guía en la toma de decisiones.
- Dar acceso a los datos a los investigadores en la expectativa de que los utilicen de manera que acaben redundando en mejoras en el funcionamiento de las agencias.
- Un círculo virtuoso que comienza poniendo recursos para el acceso a los datos, sigue con los resultados de la investigación y da pie a su explotación en beneficio público.

## Problemas

### *Cuestiones legales y éticas*

- Dado que los datos administrativos no se recogen principalmente con fines de investigación, el público puede tener dudas sobre su privacidad, la vinculación de sus datos de diferentes fuentes y el uso por parte de los investigadores.
- Estos temas se estudian en Laurie y Stevens (2014) desde la perspectiva de la jurisprudencia médica.

### *Confianza en el gobierno*

- En términos generales, la confianza de los ciudadanos en que las actuaciones de sus gobiernos sean las correctas varía sustancialmente de un país a otro (OCDE Government at a Glance 2013).
- La desconfianza del gobierno afecta la capacidad de implementar programas eficientes de acceso a los datos administrativos.
- La confianza también está relacionada con la seguridad de los datos: preocupación por la fuga de registros individuales después del procesamiento y la fusión de registros.

## Problemas (cont.)

### *Replicabilidad y confidencialidad*

- Datos administrativos: las preocupaciones con la confidencialidad pueden imposibilitar la replicabilidad.
- La solicitud de datos se ha convertido en el código de conducta estándar en las revistas académicas. Pero las revistas tienden a otorgar excepciones por motivos de confidencialidad de los datos.

### *Desigualdad entre investigadores*

- No todos están contentos de depender de fuentes gubernamentales.
- Algunos economistas se quejan de que los investigadores prominentes tienen más posibilidades de tener las conexiones necesarias para obtener esos datos.
- Esta línea de investigación podría ampliar la desigualdad dentro de la profesión, porque las superestrellas académicas tienen más oportunidades para realizar investigaciones singulares.
- Es importante disponer de un ancho de banda suficiente que permita el acceso a todos los equipos de investigadores que lo merezcan.

## La situación en España

- España está por detrás de otros países en el acceso de los datos administrativos.
- No obstante, ha habido iniciativas importantes para facilitar el acceso de muestras aleatorias de registros individuales de la seguridad social e IRPF a los investigadores.
- La disponibilidad de estos datos permite llevar a cabo valiosas investigaciones.
- Sin embargo, su potencial es limitado por comparación con las bases de datos que contienen el universo de afiliados o declarantes, disponibles en otros países.
- El universo de afiliados y declarantes es esencial para estudiar la dinámica de las interacciones entre empresas y trabajadores.
- Por lo que respecta a los datos de educación o del sistema sanitario, a excepción de algunas iniciativas recientes, está casi todo por hacer.

## La situación en España (cont.)

- A pesar de nuestro retraso, se dan las condiciones para dar un gran salto adelante.
- Hay que crear infraestructuras que posibiliten en España el análisis estadístico de los datos generados por la administración pública acorde con las posibilidades tecnológicas de nuestro tiempo.
- Tenemos abundante experiencia internacional en la que apoyarnos.
- La evidencia empírica es importante para el diseño y la evaluación de las políticas públicas. En la medida en que las nuevas tecnologías nos ofrecen la oportunidad de tener más y mejor evidencia, sería irresponsable desaprovecharla.
- Un aspecto adicional es lo que se denomina el *efecto observador* en física: el hecho de que la mera observación de un fenómeno pueda cambiar el fenómeno observado.
- Cabe esperar que el análisis concienzudo de los datos administrativos por los investigadores redunde en mejoras en la propia actividad de la administración pública.

## La situación en España (cont.)

### *La Muestra Continua de Vidas Laborales (MCVL)*

- Desde 2004 está disponible para los investigadores una base de datos administrativa ofrecida por la Seguridad Social, la MCVL, que incluye el 4% de las personas que tienen relación con la Seguridad Social, sea como cotizantes (asalariados) o como beneficiarios (pensionistas) (1,1 millón de personas).
- La MCVL se ha utilizado principalmente en estudios laborales y económicos, entre ellos los que evalúan las tendencias de entrada a la jubilación y la duración de las relaciones laborales.
- Recientemente se ha empezado a utilizar la MCVL en la investigación en salud pública por parte de los epidemiólogos.

## La situación en España (cont.)

### *Datos fiscales*

- A través del Instituto de Estudios Fiscales, la Agencia Tributaria proporciona una muestra del 4% de los declarantes del IRPF. No hay datos disponibles sobre sociedades.
- Los datos fiscales de los miembros de la MCVL están enlazados a partir de cierto momento.
- También se enlazan los datos fiscales de los miembros de las encuestas que forman parte de Eurostat. En concreto, la Encuesta de Población Activa y la Encuesta de Condiciones de Vida.

### *Datos del sistema educativo*<sup>5</sup>

- Disponibles en Cataluña a través de la iniciativa DataResSS desde 2015.
- Datos de pruebas estandarizadas disponibles en la Comunidad de Madrid.

### *Datos de préstamos*

- Se han utilizado en la investigación pero no hay vías de acceso a los datos establecidas para investigadores ajenos al Banco de España.

---

<sup>5</sup>Figlio et al (2016) presenta una lista internacional de datos utilizados para estudiar cuestiones educativas.

- DataResSS (Data for Research in Social Sciences o Datos para la Investigación en las Ciencias Sociales) es una iniciativa conjunta de la Barcelona GSE y el IDESCAT para poner los datos administrativos de Cataluña a disposición de los investigadores.
- Para acceder a los datos administrativos, un investigador debe enviar una propuesta. Las propuestas son evaluadas primero por un comité académico que tiene en cuenta la calidad del proyecto y consideraciones éticas.
- A continuación un comité ejecutivo decide los detalles del acceso a los datos. Esta etapa puede requerir la firma de un convenio entre el IDESCAT y los investigadores o sus instituciones.
- Tras esta fase, el investigador puede acceder a los datos sujeto a un compromiso de no divulgación/no identificación.
- El modo de acceso a los datos se determina caso por caso, bien sea el acceso directo a los datos o el acceso restringido dentro de las instalaciones de IDESCAT.
- El envío de propuestas está abierto a los investigadores de todo el mundo.



## **Ilustraciones**

## **Madurez y rendimiento escolar en los datos del sistema educativo de Cataluña**

*Caterina Calsamiglia y Annalisa Loviglio (2016)*

- Estudian el efecto de la edad sobre el rendimiento académico de los niños.
- Centran su análisis en los estudiantes matriculados en escuelas públicas de Cataluña desde el año escolar 2009-2010 hasta el 2013-2014.
- Cada año unos 79.000 niños se matriculan en el primer grado de una escuela primaria.
- El 67% asiste a una escuela pública, el 30% a una concertada y el resto a una privada.
- Encuentran que los niños nacidos en la segunda mitad del año tienen peores resultados en las evaluaciones, tasas de abandono escolar más altas, menores transiciones al itinerario académico y mayores problemas de aprendizaje.

*Los datos del sistema educativo de Cataluña*

- Combinan datos de diferentes fuentes que les proporcionan información detallada sobre la matrícula, la progresión escolar, los resultados académicos y las características sociodemográficas de los estudiantes.
- La Consejería de Educación les proporcionó los registros individuales de matrícula en todas las escuelas desde preescolar hasta la secundaria.
- La infraestructura de tecnología de la información que permite la recogida automática de datos fue adoptada por prácticamente todas las escuelas a partir de 2010.

## *Los datos del sistema educativo de Cataluña (cont.)*

- La información básica (edad del niño, escuela y clase a la que asistió) está disponible para todas las escuelas. Para las escuelas públicas se recogen características sociodemográficas adicionales (como sexo, nacionalidad, necesidades especiales).
- Para los niños matriculados en escuelas públicas observan las evaluaciones internas que reciben al final de cada ciclo escolar en cada una de las materias cursadas.
- El Consejo de Evaluación de Cataluña les proporcionó los resultados de las pruebas estandarizadas realizadas por todos los estudiantes del último año de la escuela primaria y del último año de la secundaria.
- Estos exámenes se llevaron a cabo desde 2009-2010 para la escuela primaria y desde 2011-2013 para la ESO. Evalúan las competencias básicas en matemáticas y lenguas y tienen un propósito puramente estadístico.
- Finalmente, recogen información sobre la familia del estudiante, en concreto sobre la educación de los padres de acuerdo con el Censo de 2002 y los datos del Padrón.
- Todas las fuentes de datos fueron fusionadas y anonimizadas por el IDESCAT.

## **El impacto de la política monetaria sobre el riesgo que asumen los bancos en su concesión de préstamos**

### *La Central de Información de Riesgos (CIR) del Banco de España*

- La CIR es un servicio público que gestiona una base de datos en la que constan todos los préstamos que las entidades financieras conceden a sus clientes.
- Las entidades de crédito tienen la obligación de informar sobre los créditos, operación a operación.
- Los datos declarados permiten al BdE conocer los créditos totales concedidos por todas las entidades, lo que facilita el ejercicio de sus competencias de supervisión bancaria.
- Por su parte, las entidades que declaran sus riesgos a la CIR reciben mensualmente información agregada del riesgo contraído por las personas físicas y jurídicas para las que han hecho una declaración, lo que facilita el seguimiento de sus riesgos de crédito.
- No obstante, cualquier entidad puede pedir información concreta de un titular si éste le solicita la concesión de un crédito.
- Cualquier persona física o jurídica puede solicitar los datos declarados en la CIR a su nombre, siguiendo un procedimiento pautado.

- Cuantifican el efecto de la política monetaria sobre la composición de la oferta de crédito en relación con el riesgo que asumen los bancos.
- Su variable de política monetaria es un tipo de interés a corto plazo que responde al tipo de interés que fija el Banco Central Europeo.
- Sus datos van del 2002 al 2008. Tienen unas 240.000 solicitudes de préstamos a bancos por parte de empresas no financieras.
- Cruzan cada préstamo con información acerca de los balances del banco y de la empresa que intervienen.
- El 86% de los préstamos corresponden a empresas con relaciones con más de un banco.
- Esta circunstancia les permite separar el efecto de la política monetaria en tres partes: un efecto de empresa, un efecto de banco y un efecto de la interacción entre ambos.
- Clave en este estudio es el cruce de los datos de la CIR, que contiene el stock de préstamos concedidos por los bancos a las empresas, con los datos de peticiones de información por parte de los bancos acerca de solicitudes de crédito de las empresas.
- Por tanto, observan solicitudes individuales de préstamo, observan si son concedidas o no y en el caso de ser concedidas observan la cuantía prestada.

## *Resultados*

- Jiménez et al encuentran que tipos a corto más bajos inducen a los bancos a prestar más a empresas con riesgo.
- El riesgo de las empresas lo miden por la presencia de un historial de préstamos morosos.
- El efecto es mayor para los bancos poco capitalizados que para los más capitalizados.
- La reducción de 1 punto porcentual en el tipo a corto aumenta la probabilidad de que un préstamo sea concedido a una empresa con una mala historia crediticia por un banco poco capitalizado (relativo a uno más capitalizado) en un 8%.
- Sus resultados muestran que una política monetaria de tipos bajos aumenta el riesgo que adquieren los bancos en su concesión de créditos.
- Por tanto, la política monetaria tiene efectos sobre la asunción de riesgo que van más allá de los efectos separados sobre el balance de las empresas y de los bancos.

## El efecto a largo plazo de los buenos maestros

*Raj Chetty, John Friedman y Jonah Rockoff (2014)*

- Este estudio combina datos administrativos de 2,5 millones de escolares de la ciudad de Nueva York junto con sus ingresos en su época adulta 20 años después.
- La pregunta que se hacen es si los alumnos de maestros con una medida de valor añadido alta tienen ingresos más altos como adultos.
- El "valor añadido" de un maestro lo miden por su contribución al aumento de las notas en exámenes estandarizados.
- Los resultados son muy llamativos.
- Encuentran que reemplazando un maestro en el 5% más bajo de la distribución por un maestro promedio aumenta los ingresos acumulados en la vida adulta en un cuarto de millón de dólares en términos de valor actual.

## **El efecto a largo plazo de los buenos maestros (cont.)**

- El estudio muestra varias de las ventajas de los datos administrativos. Primero, los autores consiguen enlazar datos de notas escolares con datos de impuestos pagados por las mismas personas muchos años después.
- Segundo, la naturaleza de largo plazo de los datos impositivos permite captar tanto los ingresos de los estudiantes en la edad adulta como información acerca de sus padres cuando eran menores de edad.
- Finalmente, la naturaleza granular de los datos sobre los tests les permite verificar un supuesto clave para la identificación de un efecto causal, a saber, que los estudiantes más capaces no se emparejan con los mejores maestros.



## Los factores biológicos en la desigualdad intergeneracional de la renta

*Anders Björklund, Mikael Lindahl y Erik Plug (2006)*

- Su objetivo es descomponer la asociación intergeneracional entre componentes pre- y post nacimiento.
- Encuentran que ambos tipos de factores son importantes tanto para el nivel educativo como para la renta.
- Gracias a la disponibilidad del universo de una población es posible estudiar estadísticamente subpoblaciones minoritarias que en el caso de encuestas sería imposible.
- Un ejemplo es la muestra de adoptados en Suecia entre 1962 y 1966 que estudian Björklund y coautores.

- El registro de población sueco contiene información acerca de los padres biológicos y adoptivos de los adoptados. Esta información aparece como resultado de las características del proceso de adopción en Suecia durante el periodo de estudio.
- Los autores cruzan esta información con la proveniente de los registros educativos y de impuestos para obtener información acerca de las características de los padres y de los hijos en la edad adulta.
- El registro tributario les proporciona rentas y salarios de los padres en 1970, 75, 80, 85 y 90. Con ello calculan una renta promedio de los padres a lo largo de 20 años.
- Miden la renta y los estudios de los hijos en 1999 cuando estos tenían entre 33 y 37 años de edad.
- Entre 1962–1966 nacieron en Suecia 5292 niños que fueron adoptados en su gran mayoría durante los seis primeros meses de vida.
- La muestra de Björklund et al contiene 2125 adoptados sobre los que se dispone de información completa en todas las dimensiones –casi 2000 adoptados tenían padre biológico desconocido.

## La persistencia intergeneracional en renta o educación

- La persistencia intergeneracional en renta o educación se suele medir con un modelo de la forma

$$Y_j^{bc} = \beta_0 + \beta_1 Y_j^{bp} + error$$

donde  $Y_j^{bc}$  representa el logaritmo de la renta de largo plazo o el nivel educativo,  $j$  es un índice de familia,  $bc$  y  $bp$  denotan hijo biológico y padre biológico, respectivamente.

- El parámetro  $\beta_1$  mide la fuerza de la asociación intergeneracional, representando una mezcla de efectos biológicos, prenatales y del entorno en el que creció el niño.
- Björklund et al utilizan sus datos de adoptados para descomponer  $\beta_1$  en dos partes:

$$Y_j^{ac} = \alpha_0 + \alpha_1 Y_j^{bp} + \alpha_2 Y_j^{ap} + error$$

donde  $ac$  y  $ap$  denotan hijos adoptados y padres adoptivos, respectivamente.

- Los coeficientes  $\alpha_1$  y  $\alpha_2$  captan factores pre-nacimiento y factores post-nacimiento, respectivamente. De nuevo, se trata de medidas de asociaciones predictivas.
- Las medidas son imperfectas porque los adoptados no se asignan a las familias adoptivas al azar y porque no empiezan a vivir con la familia adoptiva justo tras nacer.

## Resultados

- Los resultados para años de educación son  $\hat{\beta}_1 = 0.240$  (en una muestra de no adoptados) y  $\hat{\alpha}_1 = 0.113$  y  $\hat{\alpha}_2 = 0.114$  en la muestra de adoptados.
- Separando el nivel educativo de padre ( $f$ ) y madre ( $m$ ), se obtiene  $\hat{\beta}_1^f = 0.170$  y  $\hat{\beta}_1^m = 0.158$  para no adoptados, mientras que para adoptados los resultados son  $\hat{\alpha}_1^f = 0.094$  y  $\hat{\alpha}_2^f = 0.094$  por un lado y  $\hat{\alpha}_1^m = 0.101$  y  $\hat{\alpha}_2^m = 0.021$  por otro.
- La suma de las contribuciones de padres biológicos y adoptivos es comparable a la persistencia obtenida en los datos de no adoptados.
- En una extensión consideran interacciones entre el componente innato y el adquirido.
- Encuentran que estas interacciones son significativas y positivas en algunos casos, lo que sugiere que no solo *natura* y *nurtura* importan sino que además la importancia de la *nurtura* es distinta en distintas partes de la distribución de *natura*.
- Estos resultados son importantes para informar el diseño de las políticas orientadas a mejorar la igualdad de oportunidades.




## Barrios étnicos y éxito de los inmigrantes






*Per-Anders Edin, Peter Fredriksson y Olof Åslund (2002)*

- Interés en las consecuencias para el éxito laboral de vivir en barrios de alta concentración étnica o nacional. En Suecia el 11% de la población ha nacido en el extranjero. De éstos, más de un 40% viven en un barrio étnico.
- En principio, el efecto causal de vivir en un barrio étnico para los inmigrantes es ambiguo.
- Por un lado, la segregación residencial puede reducir la tasa de adquisición de conocimientos propios del país de adopción (eg. la lengua) y de ahí dificultar el acceso a mejores empleos.
- Por otro, el enclave puede representar una red que aumenta las oportunidades de sus miembros diseminando información a nuevos inmigrantes.
- Si se comparan directamente, los ingresos de los inmigrantes que viven en barrios étnicos son un 5% inferiores a los que no, teniendo en cuenta diferencias en edad, estudios, sexo, situación familiar, lugar de procedencia y año de inmigración.
- El problema es que esta asociación estadística puede no tener valor causal si la decisión de vivir en un barrio étnico está asociada con las oportunidades esperadas.

## Barrios étnicos y éxito de los inmigrantes (cont.)

- Entre 1985 y 1991 el gobierno sueco asignó un lugar de residencia inicial a inmigrantes refugiados. La motivación de la política era que la dispersión de inmigrantes facilitaría su integración.
- Edin et al. utilizan un indicador de la asignación inicial (8 años antes de la residencia en el momento del estudio) como *variable instrumental* para obtener un efecto causal.
- Las estimaciones dan lugar a una ganancia para inmigrantes de baja cualificación del 13% asociada con el aumento de una desviación estándar en la concentración étnica.
- Para los inmigrantes de alta cualificación no se observa ni ganancia ni pérdida.
- Cuando la acción de una política crea una situación de este tipo se habla de *experimento natural*, en el sentido de que se crea el tipo de variación exógena propio de los experimentos controlados.

-  Björklund, A., M. Lindahl y E. Plug (2006): "The Origins of Intergenerational Associations: Lessons from Swedish Adoption Data," *Quarterly Journal of Economics* 121, 999–1028.
-  Calsamiglia C. y A. Loviglio (2016): "Maturity and School Outcomes in an Inflexible System: Evidence from Catalonia," CEMFI y Barcelona GSE.
-  Card, D., R. Chetty, M. Feldstein y E. Saez (2010): "Expanding Access to Administrative Data for Research in the United States," National Science Foundation SBE 2020 White Paper.
-  Chetty, R., J. Friedman y J. E. Rockoff (2014): "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review* 104, 2633–2679.
-  Connelly, R., C.J. Playford, V. Gayle y C. Dibben (2016): "The role of administrative data in the big data revolution in social science research," *Social Science Research*, 59, 1–12.
-  Edin, P-A., P. Fredriksson y O. Åslund (2003): "Ethnic Enclaves and the Economic Success of Immigrants - Evidence from a Natural Experiment," *Quarterly Journal of Economics* 11, 329–57.
-  Einav, L. y J. Levin (2014): "The Data Revolution and Economic Analysis," *Innovation Policy and the Economy* 14, 1-24.

-  Figlio, D. N., K. Karbownik y K. G. Salvanes (2016): "Education Research and Administrative Data." En: E. Hanushek, S. Machin y L. Woessmann (eds.) *Handbook of the Economics of Education*, Vol. 5, Capítulo 2.
-  Hamermesh, D. (2013): "Six Decades of Top Economic Publishing: Who and How," *Journal of Economic Literature* 51, 162–72.
-  Jiménez, G., S. Ongena, J-L Peydró y J. Saurina (2014): "Hazardous Times for Monetary Policy: What Do Twenty-Three Million Bank Loans Say About the Effects of Monetary Policy on Credit Risk-Taking?" *Econometrica* 82, 463–505.
-  Laurie, G. y L. Stevens (2014): "The Administrative Data Research Centre Scotland: A Scoping Report on the Legal & Ethical Issues Arising from Access & Linkage of Administrative Data," Edinburgh School of Law Research Paper 2014/35.
-  Piketty, T. y E. Saez (2003): "Income Inequality in the United States, 1913–1998." *Quarterly Journal of Economics* 118, 1–39.