

Los retos de la econometría en el progreso de la estadística

Manuel Arellano
CEMFI

XXIII Dia de l'Estadística de Catalunya

Facultat d'Economia i Empresa, Universitat de Barcelona

20 octubre de 2023

- Quiero empezar dando las gracias a los organizadores de esta edición del Día de la Estadística de Catalunya por invitarme y a Montse Guillén en particular.
- Esta es una ocasión especial para mi.
- En primer lugar, porque estoy en mi universidad y mi facultad; donde estudié y empecé mi actividad docente como universitario.
- En segundo lugar, porque a pesar de mi doble identidad como economista y estadístico, por mi afiliación académica tengo menos ocasiones de interactuar con estadísticos que con economistas.
- En tercer lugar, por la oportunidad de hablar sobre econometría y sus retos con miembros de una vibrante comunidad de académicos y profesionales de la estadística como es la Societat Catalana d'Estadística.

Plan de la conferencia

- Me gustaría tratar cuatro puntos:
- Primero, me referiré a la relación entre la econometría y la estadística.
- A continuación, hablaré sobre la situación actual de la econometría y sus retos de futuro, especialmente aquellos que tienen que ver con los métodos estadísticos.
- En tercer lugar, comentaré uno de mis últimos trabajos, utilizándolo para ilustrar algunas cuestiones metodológicas.
- Por último, concluiré con unos comentarios sobre el acceso a los datos administrativos como uno de los grandes retos de la estadística económica.

I. Econometría y estadística

- El desarrollo del método estadístico es uno de los aspectos más fascinantes de la historia de la ciencia en los últimos 100 años, quizá el que más si pensamos en su impacto en el conjunto de todas las ciencias.
- La fascinación está en que el propio método ha ido tomando forma a la vez que sus aplicaciones en campos muy diversos con interacciones de ida y vuelta.
- En los años 50, cuando el método estadístico ya era una disciplina coherente y unificada y aparecieron los primeros departamentos de estadística, nos damos cuenta de que en su formación habían confluído desde matemáticos a economistas y demógrafos, pasando por biólogos, astrónomos o ingenieros, entre otros.
- Esta *doble existencia de la estadística*, como método común y a la vez parte integral del desarrollo de ciencias sustantivas con necesidades muy diversas, no tiene comparación con las matemáticas.
- Es lo que explica no sólo la multiplicidad de comunidades de investigadores: estadísticos matemáticos, económetras, bioestadísticos, psicómetras, etc. sino también la existencia de flujos de métodos y aplicaciones en todas las direcciones.

I. Econometría y estadística (cont.)

- Por ejemplo, en la ciencia económica el análisis de series temporales y de las respuestas causales no experimentales han sido centrales desde sus orígenes.
- El resultado es que desde la econometría han surgido contribuciones importantes a la teoría estadística de series temporales y a la metodología de inferencia causal.
- Estos desarrollos se han trasladado al conjunto del conocimiento estadístico y, a su vez, en la econometría han influido los avances en otros ámbitos de la estadística.
- Una cuestión distinta (de la que a veces nos lamentamos) es que de resultados de la compartimentación de las ciencias, vivimos en comunidades de investigadores que desarrollan sus propias culturas, terminologías y ámbitos de publicación.
- Lo que se convierte a veces en fuente de duplicaciones y retrasos en la transmisión del conocimiento.

I. Econometría y estadística (cont.)

- Harold Hotelling, distinguido estadístico y economista y uno de los primeros presidentes de la Econometric Society, escribió hace más de 70 años un enfático ensayo que les recomiendo sobre el lugar de la estadística en la universidad.
- A Hotelling le preocupaba la descentralización de la estadística, de la docencia en particular, y el difícil equilibrio entre la dedicación a los métodos y a sus aplicaciones.
- Sin embargo, el tiempo transcurrido desde entonces muestra que el policentrismo de la estadística, a caballo entre el método y las ciencias, está en la base de su pujanza.

II. Los retos de la econometría

- Estamos en medio de cambios trascendentales. Cualquier aspecto de la vida pública y privada va a ser capaz de convertirse en datos y, por tanto, susceptible de un escrutinio estadístico sistemático.
- Esto es cierto en el caso de textos, imágenes y mapas; de registros administrativos como impuestos, seguridad social o créditos; de las transacciones entre empresas y consumidores; y de los archivos hospitalarios, educativos o judiciales.
- Ante esta avalancha de información, estamos desesperadamente escasos de métodos econométricos y computacionales relevantes para los nuevos tipos de datos y las preguntas que los acompañan.
- Me referiré en concreto a las relaciones causales y a la medición descriptiva en economía.

Causalidad

- Es de esperar que la econometría seguirá contribuyendo métodos estadísticos para el análisis causal de datos longitudinales como ha hecho hasta ahora.
- En el estudio de respuestas dinámicas se empiezan a abrir camino los enfoques que combinan la modelización con aspectos de diseño o datos sobre intenciones.
- El análisis causal en redes presenta retos importantes:
 - Por un lado, el diseño de estrategias experimentales y observacionales para identificar la transmisión causal de impactos a través de las redes.
 - Por otro, el desarrollo de una teoría estadística para datos con *dependencias complejas* como las que aparecen en las redes.
- Otras áreas de investigación que se están desarrollando rápidamente:
 - Propiedades de estimadores por simulación y algoritmos de aprendizaje automático.
 - Herramientas para que el análisis de políticas macroeconómicas dependa menos de la teoría y más de los datos, combinando micro y macro datos.

Medición descriptiva

- Con las nuevas fuentes de datos, la prominencia de los estudios innovadores de medición descriptiva ha ido en aumento.
- Tradicionalmente, muchos éxitos de la econometría han sido análisis descriptivos que no tratan de estimar efectos causales ni modelos estructurales. Se trata de estudios que idean nuevos conceptos que se han de medir y las técnicas para hacerlo.
- Las aportaciones más notables proporcionan mediciones que desafían la sabiduría convencional sobre el estado del mundo o arrojan luz sobre nuevas perspectivas.
- En tiempos recientes ha habido una explosión de estudios descriptivos esclarecedores.
- Por ejemplo, la descomposición de la variación salarial entre componentes de empresa y de trabajador utilizando registros de la seguridad social; los estudios de la desigualdad económica utilizando los registros tributarios; o la medición del valor añadido de los maestros o de la geografía de la movilidad intergeneracional.
- Asociados con estos estudios hay retos metodológicos en los que la interacción entre econométricas, psicométricas, estadísticas y científicas de la computación puede dar lugar a grandes avances.
- Por ejemplo, en los *problemas de deconvolución* asociados a los modelos multidimensionales con variables latentes.

III. Datos de panel y expectativas

- En varios proyectos recientes, he intentado medir las *expectativas de las personas o los hogares* utilizando diversas fuentes de microdatos longitudinales o de panel.
- Las expectativas son un elemento clave de las decisiones económicas.
- La mayoría de las decisiones involucran incertidumbres y estas pueden ser determinantes para las decisiones tanto o más que las preferencias.
- Por ejemplo, la incertidumbre (o inseguridad) sobre la renta futura tienen un gran impacto en cómo se reparte la renta corriente entre el consumo y el ahorro.

La desigualdad en la seguridad económica

- Una vez sabemos medir el grado de seguridad económica de una persona podemos estudiar la variabilidad de dicha seguridad entre unas personas y otras.
- Se trata de un tipo de desigualdad a la que se ha prestado escasa atención. No es lo mismo tener certeza sobre la renta futura que tener una gran inseguridad, ya que esta te impide hacer planes.
- Con Stéphane Bonhomme, Micolé De Vera, Laura Hospido y Siqi Wei desarrollamos una metodología para medir la desigualdad en la seguridad económica.
- Utilizamos micro datos de panel obtenidos enlazando los historiales laborales de la Seguridad Social con los registros del impuesto sobre la renta y el censo para una muestra del 4% de la población española de afiliados a la seguridad social de 2005 a 2018 (Muestra Continua de Vidas Laborales, MCVL).

La desigualdad en la seguridad económica (cont.)

- Construimos medidas de riesgo de renta anual individual que validamos con las preguntas de expectativas subjetivas de la Encuesta Financiera de las Familias (EFF) del Banco de España.
- Los resultados muestran que más de la mitad de los españoles saben lo que van a ganar en el próximo año con gran certeza; pero hay una masa muy grande de gente, sobre todo jóvenes, que tiene una inseguridad tremenda, con desviaciones esperadas respecto al promedio de su renta en un año del 30 % y más.

¿Cómo medir el riesgo individual de renta? Tres posibilidades

- 1) Preguntar a las personas por sus expectativas en una encuesta; esto es, por las probabilidades de que sus rentas sean unas u otras (por ej. en la EFF).
 - Utilizamos esta opción para validar los resultados.
- 2) Construir un modelo de las decisiones de consumo y ahorro para determinar indirectamente el riesgo de renta como aquel que permite reconciliar los comportamientos observados con los predichos por el modelo.
 - Este ejercicio puede mejorar nuestra percepción del riesgo, pero el resultado dependerá de cuán confiable sea el modelo.
- 3) Seguir un enfoque de predicción, obteniendo el riesgo de renta a partir del cálculo de las probabilidades de que las rentas futuras de un conjunto de individuos con ciertas características, una vez observadas, sean unas u otras.
 - En este caso es necesario disponer de la mayor parte de la información relevante que los individuos utilizan implícitamente en sus propias predicciones.

Predictores para medir el riesgo de renta individual y su desigualdad

- Seguimos un *enfoque de predicción* aprovechando la disponibilidad de registros continuos de historiales laborales completos en España, con los que intentamos *imitar* el conjunto de información de los individuos.
- Utilizamos técnicas de regresión y de aprendizaje automático.
- Construimos medidas individuales de riesgo de renta en función de la historia laboral pasada, la renta, la demografía del hogar y factores macroeconómicos.
- Nuestra medida preferida es un *coeficiente de variación* (CV) que representa la desviación esperada respecto a la renta promedio en un año.
- Una vez disponemos de coeficientes de variación individuales, podemos estudiar su desigualdad, su persistencia, y cómo cambian con la edad y el ciclo económico.

Posibles problemas

- El primero es que aunque el conjunto de información sea adecuado, el predictor no lo sea.
 - Examinamos esta posibilidad utilizando técnicas de inteligencia artificial (redes neuronales). Los resultados no cambian.
- Otro es que los individuos tengan más información que nosotros.
 - Permitimos heterogeneidad en los predictores utilizando efectos fijos por grupos obtenidos a partir de algoritmos de clasificación. De nuevo los resultados son similares.
- Un tercer problema es que los individuos perciban el riesgo de forma distinta al problema de predicción que formulamos.
 - Para abordar este punto recurrimos a los datos de expectativas subjetivas de la EFF. Obtenemos una coincidencia notable entre ambos enfoques.

Interpretación económica del coeficiente de variación

- El CV es aproximadamente proporcional a una medida económica estándar del coste en términos de bienestar del riesgo de renta, lo cual ayuda a su interpretación.
- No obstante, esta aproximación tiene dos limitaciones.
- Primero, si ricos y pobres no son igualmente reacios a fluctuaciones proporcionales en su renta (porque el bienestar dependa de un nivel mínimo de subsistencia), el CV subestimaré el coste del riesgo tanto más cuanto la renta de un individuo esté próxima al nivel de subsistencia.
- Además, si los riesgos altos se encuentran entre las rentas bajas, el impacto de la renta de subsistencia en el bienestar tenderá a subestimar la desigualdad del riesgo en la economía.

Interpretación económica del coeficiente de variación (cont.)

- Segundo, si hay asimetrías en la distribución de la renta (en logs.) el coste del bienestar no sólo dependerá del riesgo medido por la dispersión sino también de los riesgos asociados con la asimetría.
- Estas consideraciones sugieren el interés de un enfoque más estructural que se apoye explícitamente en una función de bienestar.
- No obstante, aquí queremos enfatizar la simplicidad de una medida estadística lo más transparente posible, adoptando la *perspectiva de la estadística económica*.

IV. El acceso a los datos administrativos

- La capacidad de innovar en la medición de características relevantes de la economía española depende del acceso de los investigadores a los microdatos administrativos públicos.
- El estudio anterior no habría sido posible sin la existencia de la MCVL, cuya aparición supuso un hito en el potencial de conocimiento sobre nuestra economía.
- No obstante, aprenderíamos mucho más si se pudieran llevar a cabo estudios basados en el universo de afiliados a la Seguridad Social cruzados con las declaraciones fiscales de personas físicas y a su vez con las declaraciones de sociedades.
- Por ejemplo, podríamos estudiar con garantías no sólo el riesgo de renta a un año, sino también a cinco o diez años, que pueden variar de forma importante con la calidad de las empresas en las que trabajan los individuos, especialmente los jóvenes.

IV. El acceso a los datos administrativos (cont.)

- Observar el universo de datos administrativos proporciona indirectamente una gran riqueza de información sobre las empresas españolas, que son las grandes desconocidas en la información estadística actual.
- Es esta una carencia gravísima cuando estamos comprobando en los datos de otros países las inmensas diferencias que hay entre unas empresas y otras en aspectos cruciales como la productividad, los modelos de negocio, las amenities, el poder de mercado, o las políticas de formación de sus trabajadores.
- En relación con este punto, conviene dejar claro que la anonimización efectiva de los micro registros es imposible cuando se trata de las empresas.
- Por ello hay que desarrollar los mecanismos necesarios para que equipos autorizados de investigadores puedan tener acceso a los datos de las empresas en las mismas condiciones de confidencialidad garantizadas por contrato que se exigen a los funcionarios que desarrollan su trabajo en torno a estos registros.