Big Data: ideas, desafios y oportunidades para la economia

Walter Sosa Escudero

Universidad de San Andrés y CONICET

wsosa@udesa.edu.ar • @wsosaescudero • waltersosa.weebly.com

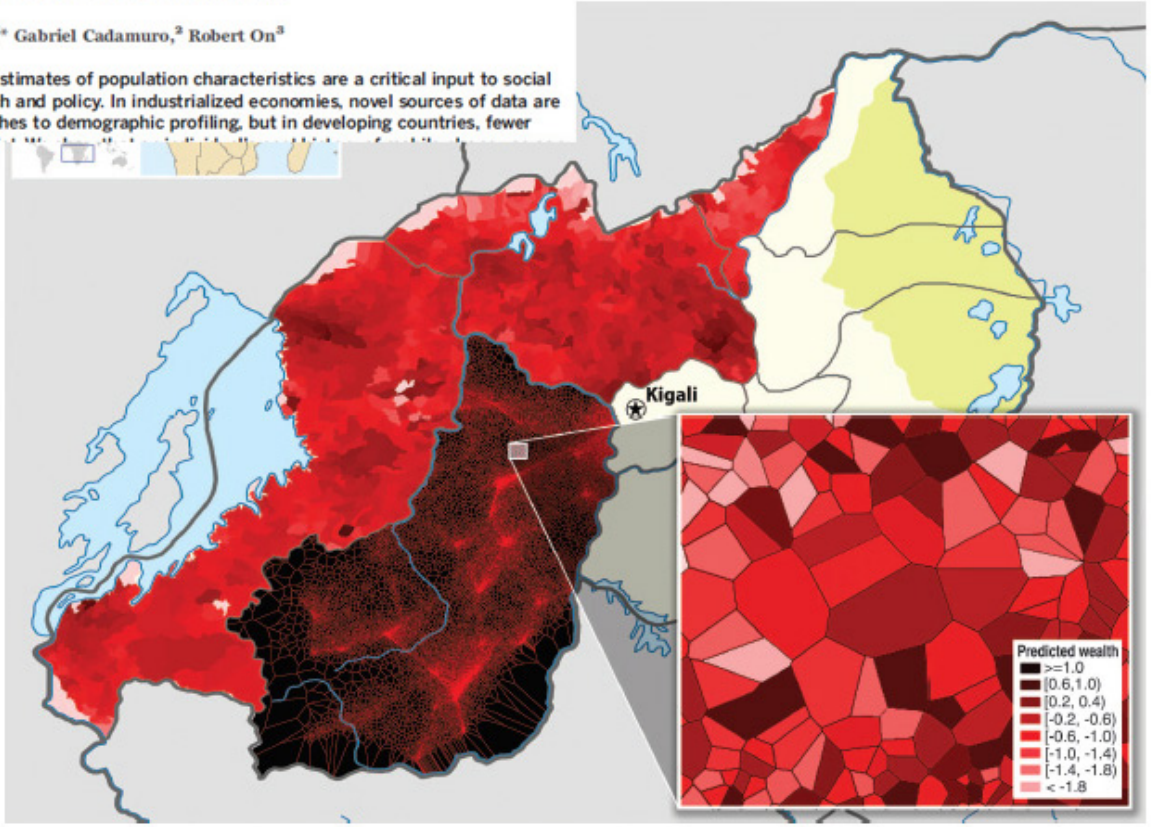
Pobreza en Rwanda (predecir)

ECONOMICS

Predicting poverty and wealth from mobile phone metadata

Joshua Blumenstock,[1*] Gabriel Cadamuro,[2] Robert On[3]

Accurate and timely estimates of population characteristics are a critical input to social and economic research and policy. In industrialized economies, novel sources of data are enabling new approaches to demographic profiling, but in developing countries, fewer



Precios en Argentina (medir)

Contents lists available at SciVerse ScienceDirect

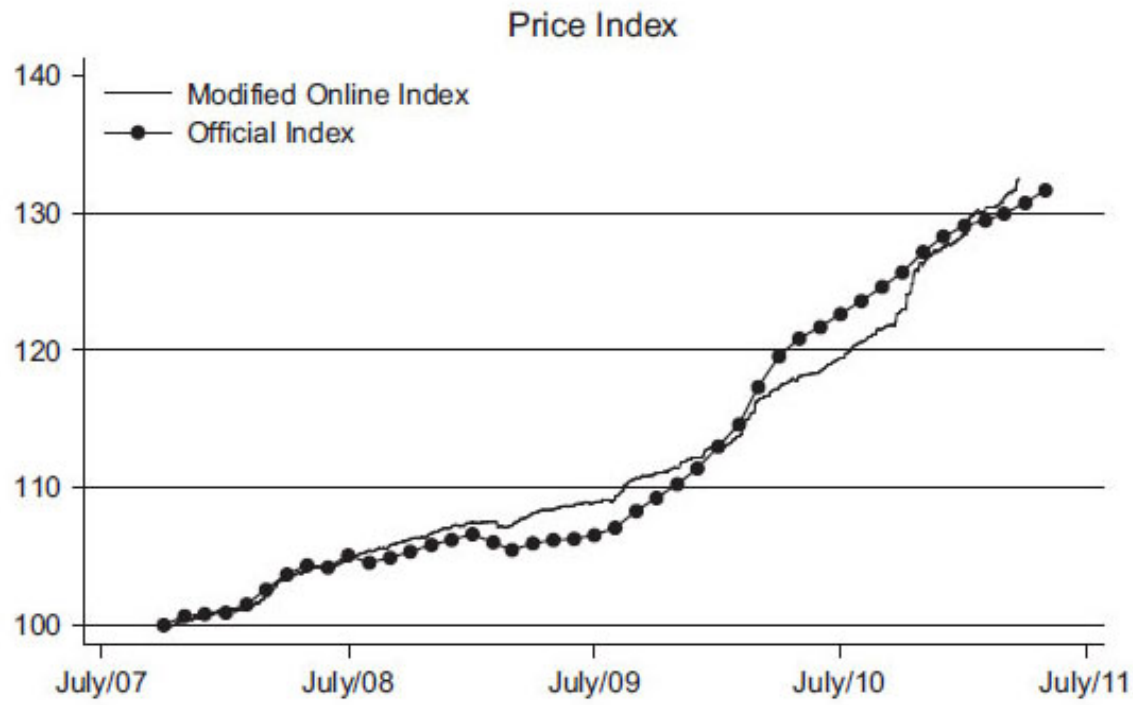
Journal of Monetary Economics

journal homepage: www.elsevier.com/locate/jme

Online and official price indexes: Measuring Argentina's inflation

Alberto Cavallo*

Massachusetts Institute of Technology, Sloan School of Management, 77 Massachusetts Ave E62-512, Cambridge, MA 02139, USA



Impuesto a las ventas en EEUU (efecto causal)

Sales Taxes and Internet Commerce

Liran Einav

Dan Knoepfle

Jonathan Levin

Neel Sundaresan

AMERICAN ECONOMIC REVIEW
VOL. 104, NO. 1, JANUARY 2014
(pp. 1-26)

Small vs. Big data

Small data (estadistica clasica)

- Como extraer lo maximo de pocos datos?
- Solucion: estructruar los datos (muestreo)
- Enfoque: muestreo complejo que aproxima el muestreo al azar (lento y caro).

Big Data

- Muchos datos (Volumen)
- Muchos datos no estructurados (Variedad)
- Muchos datos no estructurados e immediatos (Velocidad)
- 'Condicionalmente baratos'.

Machine learning vs estadistica/econometria

$$Y = f(X) + u$$

Estadistica clasica

- Interes en $f(.)$. Efecto causal
- Modelo? Teoria, experimento.
- Probabilidades (error estandar, tests)

Machine learning

- Interes en Y: predecir, clasificar, medir.
- Modelo? No hay modelo. Lo aprendemos.
- Estimacion puntual (no inferencia).

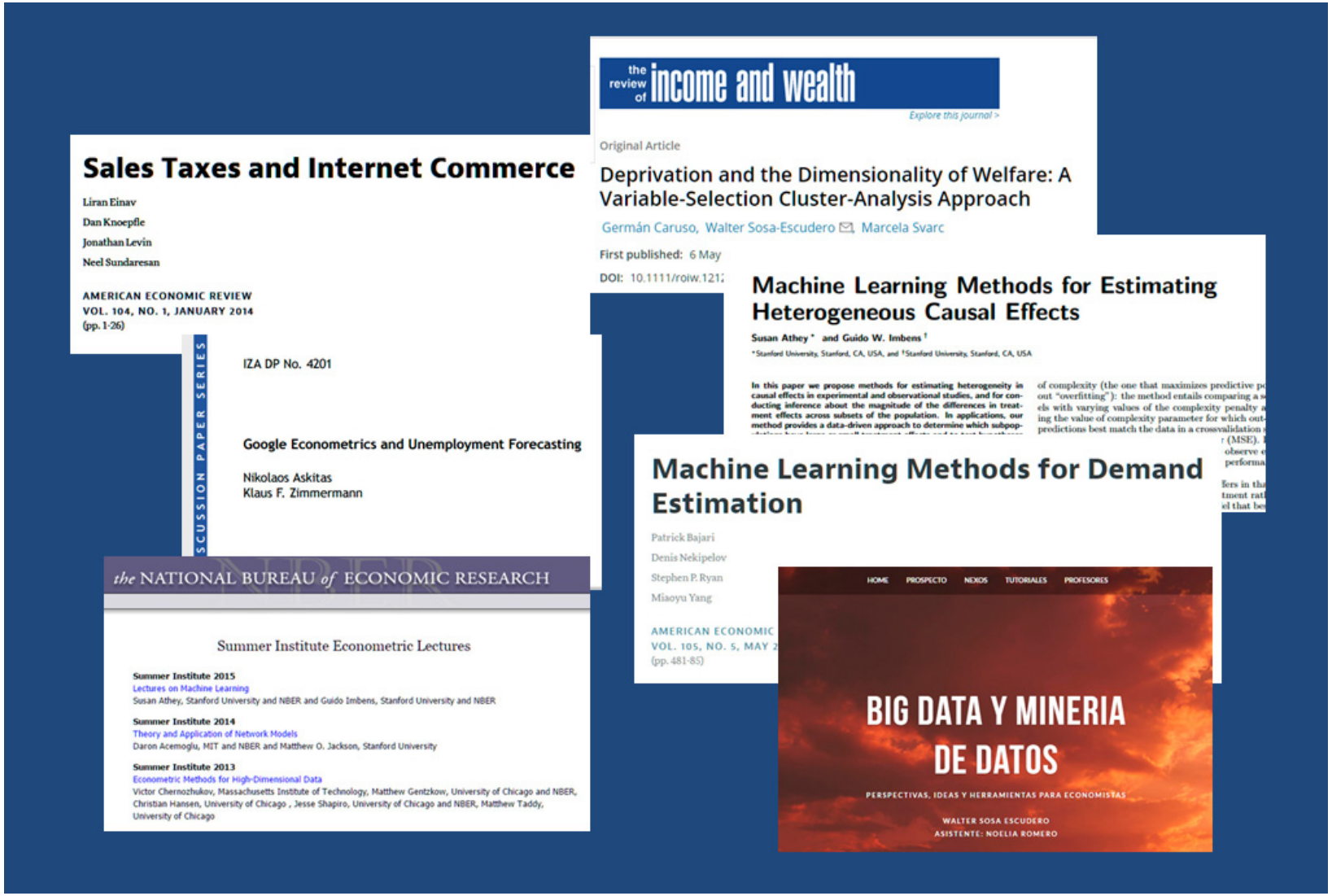
Letra chica

- Dependencias (realmente tenemos big data?. Trump effect)
- Choice based sampling.
- Contrafactuales nunca se observan (podemos tener *todos* los datos?).
- Falacia de la correlacion.
- Transparencia / privacidad.
- Comunicabilidad. Caja negra (deep learning, forests, etc.)
- Consenso social/politico.

Oportunidades

- More data: Big data no es solo muchos sino mas datos (pocos o muchos).
- Identificar la complejidad. Maldicion de la dimensionalidad.
- Rapido (crucial para la politica). Goggle Flu Trends. Price scrapping.
- Oportunidad para diseñar experimentos causales.
- Complementa a las estadisticas oficiales (no reemplaza).
- Cobertura (Rwanda). Rural vs. urbano, etc..

Nexos


Sales Taxes and Internet Commerce
Liran Einav
Dan Knoepfle
Jonathan Levin
Neel Sundaresan
AMERICAN ECONOMIC REVIEW
VOL. 104, NO. 1, JANUARY 2014
(pp. 1-26)
the review of income and wealth
Explore this journal >
Original Article
Deprivation and the Dimensionality of Welfare: A Variable-Selection Cluster-Analysis Approach
Germán Caruso, Walter Sosa-Escudero Marcela Svarc
First published: 6 May
DOI: 10.1111/roiw.121
Machine Learning Methods for Estimating Heterogeneous Causal Effects
Susan Athey * and Guido W. Imbens †
IZA DP No. 4201
Google Econometrics and Unemployment Forecasting
Nikolaos Askitas
Klaus F. Zimmermann
Machine Learning Methods for Demand Estimation
Patrick Bajari
Denis Nekipelov
Stephen P. Ryan
Miaoyu Yang
AMERICAN ECONOMIC
VOL. 105, NO. 5, MAY
(pp. 481-85)
the NATIONAL BUREAU of ECONOMIC RESEARCH
Summer Institute Econometric Lectures
Summer Institute 2015
Lectures on Machine Learning
Susan Athey, Stanford University and NBER and Guido Imbens, Stanford University and NBER
Summer Institute 2014
Theory and Application of Network Models
Daron Acemoglu, MIT and NBER and Matthew O. Jackson, Stanford University
Summer Institute 2013
Econometric Methods for High-Dimensional Data
Victor Chernozhukov, Massachusetts Institute of Technology, Matthew Gentzkow, University of Chicago and NBER, Christian Hansen, University of Chicago , Jesse Shapiro, University of Chicago and NBER, Matthew Taddy, University of Chicago
HOME PROSPECTO NEXOS TUTORIALES PROFESORES
BIG DATA Y MINERIA DE DATOS
PERSPECTIVAS, IDEAS Y HERRAMIENTAS PARA ECONOMISTAS
WALTER SOSA ESCUDERO
ASISTENTE: NOELIA ROMERO

Referencias

- Survey I: Varian, H. R., 2014, Big data: New tricks for econometrics, Journal of Economic Perspectives, 28(2), 3-27.
- Survey II: Einav, L., and Levin, J., 2014, Economics in the age of big data. Science, 346(6210).
- Simple book: James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). An introduction to statistical learning (Vol. 6). New York: Springer.
- Book: Murphy, K., 2012, Machine Learning: a Probabilistic Perspective, , MIT Press, Cambridge.
- Predictive: Blumenstock, J., Cadamuro, G., and On, R., 2015, Predicting poverty and wealth from mobile phone metadata. Science, 350(6264), 1073-1076.

- Measurement: Cavallo, A. and Rigobon, R. 2016, The Billion Prices Project: Using Online Prices for Measurement and Research, Journal of Economic Perspectives, Vol. 30, 2, pp.151-78.
- Causal: Athey, S., and G. Imbens (2015), Machine Learning Methods for Estimating Heterogeneous Causal Effects, NBER working paper.
- Policy: Kleinberg, J., Ludwig, J., Mullainathan, S., and Obermeyer, Z., 2015, Prediction policy problems. The American Economic Review, 105(5), 491-495.
- Econometrics: Belloni, A., V. Chernozhukov, and C. Hansen, 2014, High-Dimensional Methods and Inference on Structural and Treatment Effects, Journal of Economic Perspectives, 28(2): 29-50.