Administrative Data in Household Finance: Credit File Data

Stefania Albanesi

University of Pittsburgh, NBER and CEPR

NBER: Cambridge, November 30 2018

ADMINISTRATIVE DATA: CREDIT FILES

- Source: Experian

- Coverage: 2004Q1-2015Q4

- Sample: 1-1.1 million borrowers (anonymous)
  70% representative
  30% ever delinquent oversample

- Variables:
  all consumer debt except payday loans
  number of trades, balance, credit limit, missed payments
  public records (bankruptcy, foreclosure, court judgements)
  credit scores (Vantage Score 3.0)
  zip code
  individual and household income estimates
  approx 250 attributes in total

CREDIT FILE DATA: PROJECTS

- Insolvency and the 2005 Bankruptcy Reform, joint with J. Nosal

- Credit Growth and the Financial Crisis: A New Narrative, joint with G. De Giorgi and J. Nosal

- Predicting Consumer Default: A Deep Learning Approach, joint with D. F. Vamossy

- The Role of Investors in the 2007-2009 Housing Crisis: An Anatomy

- Real Estate Investors: A Quantitative Analysis

- Economic Uncertainty and the 2016 Election: The Role of Financial Distress, joint with B. Montano and M. Morelli

CREDIT FILE DATA: CHALLENGES

1 Awareness of the data:

   data can be purchased!

   without prior use, hard to understand structure and scope

   vendors more practiced with commercial clients with limited data needs

   (small samples, limited time range, only few attributes)

2 Cost:

   very high for individual researcher

   legal hurdles make cost sharing problematic

   combining sources of funding even for individual researcher may be difficult

   property rights of licensee (researcher) and funding institution sometimes
   unclear

3 Independence and intellectual freedom:

research needs to be vetted by data vendor

may require that their name not be used or certain results not included if potentially detrimental to their image or reputation

4 Publishing

most academic journals have data availability policy

need to make an exception for proprietary data, requires approval by editorial board

usually just require that the data may be available for purchase for other researchers

Translating Research into Policy

- Current projects on personal bankruptcy, mortgage default and credit scoring have direct policy implications

- Reaching policy makers is always hard

- Media can be used as a filter

→ Personal bankruptcy:
   Too Broke for Bankruptcy? *ProPublica* series

→ Real estate investors:
   NPR All Things Considered

POTENTIAL TO SUPPORT RESEARCH

- Access by more researchers would increase the number of questions addressed and may also provide useful insights on how to improve the data

- More users imply larger networks with possibility of interaction and peer support

- Data vendor can develop more experience in dealing with academic clients and their needs

  Example: 7 year limit to storing data, sample stratification etc.

- Linking with other datasets, such as social security or IRS data, property deeds data etc.