Granular data and advanced analytics

Paul Robinson, Head of Advanced Analytics, Bank of England
Financial Information Forum of Latin American and Caribbean Central Banks
5 May 2016
Why are we interested in Big Data?

• What do we mean by the term
  – Very loose meaning, covering data, techniques and attitude
  – Granular data crucial

• Why are we interested?
  – Change of responsibilities
    • The arrival of the PRA
  – Change of opportunity
    • More data, increased computing power, technical advances
  – Change of circumstances
    • Lessons from the financial crisis
  – Change of philosophy
    • Inductive vs deductive reasoning
What are we interested in?

• Gaining a richer understanding of the phenomenon of interest
  – Can help disentangle cause and effect…
  – …and identify the underlying issue that needs to be addressed
• Getting a speedier reading of developments in the economy and financial system
  – ‘Nowcasting’ and ‘nearcasting’
  – This might be particularly important when the system is undergoing rapid changes
• Quantifying previously purely qualitative data
  – Eg text
Loan-to-income multiple $\geq 4.5$

Source: Data are based on the Bank of England’s internal Product Sales Database collected by the FCA.

Advanced Analytics at the Bank of England
Discount from last asking price by buyer type

**Sources:** WhenFresh (Zoopla listings), Land Registry Price Paid, Land Registry Cash/Mortgage data, FCA Product Sales Data on mortgages, ONS Postcode Directory.

Advanced Analytics at the Bank of England
Sources: WhenFresh (Zoopla listings), Land Registry Price Paid, Land Registry Cash/Mortgage data, FCA Product Sales Data on mortgages, ONS Postcode Directory.
EMIR Data

Positions in outstanding CHF-denominated FX derivatives positions on 15/1/15
Issues encountered

- Identifying the purpose of the trade (hedging vs speculation)
- Cross-border issues
- Identifying counterparties (only ~ 50% had a LEI)
- Consolidation of institutions
- Direction of trades
- Identifying the initiator of the trade
- Separating swaps from other forms of derivative
Anonymised CHAPS payments between banks
Issues with analysing ‘Big Data’

• Example: CPI micro-data
• The ONS has produced a data set comprising:
  – 215 months (Feb 1996-Dec 2013)
  – ~110,000 prices collected per month (not the same number each month)
  – 1,113 items (not the same items each year)
  – 71 COICOP classes
  – various other meta-data (eg type of shop, region etc)
  – in total: 24,442,988 records with 25 fields
  – 611,074,700 pieces of data
Issue 1: the stability of annual inflation

UK CPI inflation 12m rate

Percentage change over 12 months


0 1 2 3 4 5 6
Issue 2: explaining non-linear functions

- Try explaining the intuition behind this relationship to busy policy makers…
Issue 3: Stability

- An issue that is closely linked to over-fitting is the stability of the models
- This is a particularly important issues when there is no strong *a priori* reason to think that the world works in this way
- (Though *a priori* thinking can also be misleading at times)
Issue 5: Confidentiality / ‘Big Brother’ state

• This was not relevant to the CPI work
• In general, the more detailed and granular the data set is, the more likely it is to contain confidential information
• We must ensure that:
  – we only use data for appropriate reasons
  – the minimum number of people are able to see any confidential data given the needs of the situation
  – data are stored securely and professionally