Big data and central banks

By David Bholat of the Bank's Advanced Analytics Division.[1]

1   Introduction

On 2 and 3 July 2014, the Bank of England hosted an event titled 'Big Data and Central Banks'. The event was organised by the Advanced Analytics and the Statistics and Regulatory Data Divisions, under the umbrella of the Bank's Centre for Central Banking Studies (CCBS). Its purpose was to discuss the IT estate, analytical techniques, governing arrangements, and strategic objectives central banks are developing around 'big data', that is, data sets that are granular, high frequency and/or non-numeric. Invited speakers from central banks and financial regulatory bodies addressed around 50 Bank staff and approximately 25 delegates from foreign central banks.[2]

The purpose of this report is to articulate the motivations for the event, key themes emerging from it, and explain why big data is likely to become of increasing importance to central banks in the years ahead. The report proceeds as follows. The first section explains how the event fits with other activities at the Bank aimed at expanding its data sources and enhancing its data analysis capabilities. Relevant related activities include a predecessor CCBS event, the creation of the Bank's Data Lab and Data Community, and, most significantly, the release of the Bank's Strategic Plan. The second section then summarises presentations made at the event. The event was held under Chatham House Rule meaning opinions are not attributed to identifiable speakers. So rather than recap each individual presentation, this report instead identifies major themes emerging across them. Among these themes are the benefits for central banks in having standardised granular data, the importance of legal considerations in enabling and constraining the scope of granular data collections, and the development of inductive analytical approaches to complement deductive approaches that traditionally have held sway in central banks. The report then concludes by speculating that big data might not only change how central banks operate, it also might be transforming how financial firms and other economic agents do business. To the extent that this transformation is occurring, it is important for central banks to understand how big data is changing the structure of the economy in ways that might impact monetary and financial stability, as well as economic growth and employment.

2   Background

One standard definition of big data is that it is data displaying one or more of the following characteristics:

(1) These data are high volume, often because data are reported on a granular basis, that is, item-by-item, for example, loan-by-loan or security-by-security.[3]

(2) These data are high velocity, because these data are frequently updated and, at the limit, collected and analysed in real time.

(3) These data are qualitatively various, meaning they are either non-numeric, such as text or video, or they are extracted from novel sources, such as social media, internet search records or biometric sensors.

Judged by this definition, the Bank of England traditionally has not dealt much with big data. Data volume historically has been low because its primary sources have been summary financial statements and aggregate statistics compiled by the Bank's Statistics and Regulatory Data Division and the Office for National Statistics. Data velocity also has been slow because these statistics and financial statements are reported in the main at most on a quarterly frequency and revised with lags. And although the Bank does have a history of gathering qualitative data through surveys and interviews with external contacts, overall the variety of data has been minimal because most formal analysis undertaken in the Bank uses structured data sets, that is, numeric data stored in relational databases, ie row and column format.

More recently, however, the Bank has been one of the leading central banks when it comes to research using higher volume, higher velocity and qualitatively more various data sets. For example, McLaren and Shanbhogue (2011) used Google data as an indicator of UK labour and housing market conditions. Benos and Sagade (2012) used equity transaction data to understand the consequences of high-frequency trading on stock markets, while Benos, Wetherilt and Zikes (2013) used transactional trade repository data to investigate the structure and dynamics of the UK credit default swap market. And both Davey and Gray (2014) and Merrouche and Schanz (2009) used high-value, high-velocity payment systems data to analyse banks' intraday liquidity management.

The steady increase in research done by the Bank using big data in part reflects its greater availability. This is because the financial crisis of 2007–08 prompted a number of statutory and supervisory initiatives that require greater disclosure by financial firms of their data to central banks and regulators. These include the reporting of firms' large exposures on a counterparty-by-counterparty basis on Common Reporting (COREP) templates; security-by-security reporting of insurers' assets mandated by Solvency II scheduled to come into force in 2016; and the reporting of transactional derivatives data as required by European Market Infrastructure Regulation. Such granular data became more readily available to the Bank of England after it assumed supervisory and regulatory responsibilities last year with the establishment of the Prudential Regulation Authority.

Consequently the Bank hosted an event in 2013 titled 'The Future of Regulatory Data and Analytics'. Its focus was on how to best integrate the Bank's new supervisory and regulatory data collections with existing statistical compilations. Much of the discussion at the event centred on developing a new post-crisis data strategy for central banks, prefiguring the Bank's new Strategic Plan (Bank of England (2014)). Although the Strategic Plan has many facets, one of its major points of emphasis is data. This is particularly evident in the three strategic initiatives that fall under the 'Analytical Excellence' pillar. The 'One Bank Research Agenda' initiative commits the Bank to supplying new data sets to the public in order to crowd-source solutions to challenging policy questions.[1] The 'New Approach to Data and Analysis' initiative created an Advanced Analytics Division with the objective of establishing a centre of excellence for the analysis of big data. And the 'One Bank Data Architecture' initiative is to be overseen by the first Chief Data Officer in the Bank's history, with the goal of integrating data across the Bank, partly through the enforcement of metadata standards to ease information sharing.[2]

Since announcing these strategic initiatives, the Bank has made strides toward their achievement. Three milestones are worth highlighting. The first is the establishment of a

Data Lab. The Lab is a room in the Bank with computers that are uploaded with state-of-the-art IT tools. Bank employees who visit the Lab are supported by a small team of IT experts who help them store, manipulate, visualise and analyse granular and unstructured data. A second and related development is the formation of a new Bank-wide Data Community. The Data Community is a group that organises fora for staff about big data issues, including monthly seminars, a new Bank intranet site with information on novel ways staff can use data, and a Data Art Gallery event exhibiting examples of innovative approaches for visualising data. Finally a third key milestone was the convening of the 'Big Data and Central Banks' event. More information about that event follows in the next section.

3   CCBS event

An increasing number of events about big data are being organised at universities and by private sector bodies. But central banks bear a unique set of legal powers, media and legislative scrutiny, and public responsibilities that make their use of big data different from that in other kinds of organisations. Hence the focus for this year's CCBS event was on 'Big Data *and* Central Banks', with emphasis on the conjunction. Unlike other big data themed events, its focus was less on the cutting edge of what is possible with big data, and more on highlighting what is being done already by central banks and financial regulatory authorities.[3] Big data was thus invoked in relative rather than absolute terms.

The event took place over two days. The first day consisted of three panels bookended by two speeches. The three panels examined in turn each of the three traits — volume, velocity and variety — associated with big data. Unlike many big data related events where only analytical results are presented, panellists were advised not to smooth over legal, technical or operational challenges they had encountered. Instead panellists were encouraged to highlight these challenges so other participants could learn from them. In advance of the event, panellists were sent a set of key questions intended to elicit presentations covering their big data projects end to end. These questions are reproduced in the appendix. They may be useful prompts for others thinking through the details of planning big data projects.

The second day was split in two. During the first half, participants were introduced to the topic of data mining. Data mining refers to a methodology, tools and set of algorithms used to detect patterns in data sets. Three families of algorithms were highlighted:

(1) www.bankofengland.co.uk/research/Pages/onebank/datasets.aspx.
(2) Bholat (2013) recaps the predecessor event.
(3) Compare with the ECB's recent workshop on using big data for forecasting and statistics: www.ecb.europa.eu/events/conferences/html/20140407_call_international_institute_of_forecasters.en.html.

- Classification algorithms such as decision trees and neural networks used to make predictions.

- Segmentation algorithms such as clustering methods used to find groups within data.

- Association algorithms used to analyse linkages between different attributes in data.

Following the data mining tutorial in the morning, the second half of the day started with another panel session. While the first three panels focused on one-off big data projects, or spotlighted unusually innovative areas of central banks using big data, the focus of the fourth panel was on delivering enterprise-wide change in the way data is managed and used across an entire central bank. As a corollary, the focus of this panel was less on how big data approaches can help central banks analyse their external environment, and more on how these approaches can help central banks improve the efficiency of their internal operations; for example, through smart document management systems that proactively push content to staff based on their past reading preferences. The key questions posed to speakers on this panel are also in the appendix.

Following the fourth panel, a moderated roundtable took place to give event participants the chance to reflect on lessons they had learned over the two days. The event then concluded with a speech made by one of the Bank's executives.

Granular data

One recurring theme raised during the event was the benefits to central banks in having access to granular data from financial firms. As noted earlier, central banks typically have collected aggregate data from firms using reporting returns structured like standard financial statements. These returns tend to translate end-user requirements literally. For example, an end-user might request data on firms' liquid assets, defining that term as deposits and government securities with original maturities less than one year. A regulatory return might then be sent to firms containing a field for reporting a single 'liquid assets' figure, as defined by the end-user. However, liquidity is a notoriously fluid concept. Assets that ordinarily might be sold with minimal loss to the seller may no longer be so under changed conditions; for example, if the credit profile of the issuer of the securities held by financial firms has changed. Furthermore, the definition of liquid assets can vary across analysts and central bank areas. For example, some end-users might conceive of shares held in money market funds as liquid assets, while others might not. To the extent that each conception of liquid assets gives rise to discrete data collections, multiple returns might have to be filed by financial firms, even though the data reported is highly repetitive except at the margin. Yet the duplication of reported data can

still leave data gaps since the aggregated nature of the figures precludes end-users from drilling down and asking more detailed questions as circumstances require. For example, if an end-user needs to assess whether government securities held by a firm are issued by the central government or municipal authorities. Circumstances might then require the costly *ad hoc* collection of this data from firms at short notice.

An alternative approach to multiple returns is to collect granular data once. A number of speakers at the CCBS event advocated this approach. A representative view was that different returns often specify varying consolidation bases, valuation methods, accounting conventions, definitions and report at different frequencies, making it difficult to stitch data together. When one speaker's country suffered a financial crisis, it was discovered that the aggregate data available to the central bank was incomplete and incompatible for pinpointing financial fragilities. This prompted the central bank to introduce exposure-by-exposure reporting. According to the speaker, such granular data now makes possible the coherent mapping of the banking system, enabling the central bank to better spot systemic risk and manage it with macroprudential policy.

The idea that analysing microdata makes it easier to discover macro-patterns might appear paradoxical at first glance. On the contrary, it might make intuitive sense to think that having more data would make it harder to identify the wood from the trees. To paraphrase information economist Herbert Simon, a wealth of information might create a poverty of attention, leading to worse decision-making (quoted in Haldane (2013)). However, a number of speakers noted that granular data becomes analytically tractable if overlaid with visual analytic tools. Instead of eyeballing millions of rows and columns of granular data, end-users are able to quickly picture the data at different units of analysis, drilling down to identify granular fault lines which might be otherwise concealed at an aggregate level.[1]

The benefits in gaining access to granular data might not only accrue to central banks. They also may accrue to the financial firms central banks regulate. For instance, one central banker noted that by having access to the granular detail on loans pre-positioned by firms with central banks as collateral, this may result in reduced haircuts and thus cheaper emergency liquidity for firms because central banks can then better judge the quality of the underlying loans against which they are lending.

However, greater data granularity in itself is not a panacea. If granular data collections are introduced unevenly across a central bank and managed in end-user silos, then the organisation runs the risk of reproducing the inconsistencies

---

(1) Flood *et al* (2014) provides an overview of some of these tools.

and inefficiencies of the current approach of collecting aggregate data using multiple returns. According to one speaker, one way to prevent this from occurring is to harmonise and enforce common definitions of granular data attributes across the organisation. In loan-by-loan databases, common attributes include the original value of the loan, the currency in which it is denominated, its purpose, outstanding balance, original and residual maturity, the repayment schedule, the interest rate, the reference rate (if applicable), and information on financial firms' counterparties, such as the income of the obligor, their domiciled country, and credit rating.

Legal considerations

As the foregoing list of attributes indicates, existing granular data sets collected by some central banks tend to have good coverage of information on counterparties and contracted cash flows. However, one area where the scope of these collections could be extended is in capturing more detail on legal terms and conditions. Even apparently minor legal provisions can have major systemic consequences. Consider the recent Greek sovereign debt crisis. The absence of collective action clauses that would have permitted bondholders to write down Greece's outstanding debt with binding effect on credit minorities, coupled with the fact that most of the bond issues named Greek courts as arbitrators in case of dispute, are key factors explaining the ECB's decision to purchase Greek sovereign debt (Grant *et al* (2014)). More generally, key clauses in financial agreements which central banks may want to capture include provisions limiting personal liability, creditors' ranking in bankruptcy proceedings, breach of contract and default triggers, and protective covenants.[1]

However, several speakers at the CCBS event observed that embracing big data did not necessarily require that central banks enlarge their data collections. For these speakers, the key task facing central banks is not getting more data. Rather they argued that it is doing more with the data central banks have already. One speaker cited payment systems data as a good example. Historically, these real-time data have been used to monitor operational risks. For example, by looking at the data, central banks might observe that a payment system is too reliant on a small number of settlement banks. The central bank might then address this excess concentration by inviting indirect participants accessing the payment system through these settlement banks to become direct participants (Finan, Lasaosa and Sunderland (2013)). But this same data also can be useful for other purposes. For example, these data might be used to monitor the settlement behaviour of individual firms. Such information might provide a timelier indicator of firms' liquidity position than liquid assets figures submitted on regulatory returns. Payment systems data also might be linked with other data to achieve new insights. For example, payments data might be blended with loan-by-loan

mortgage origination data to identify shortfalls in mortgage repayments much sooner than those shortfalls are reported as arrears by firms in their quarterly financial disclosures.

Indeed, blending data gives central banks a 'third way' between the false dichotomy of either buying all their data from external parties or building all data-capture systems in-house. For example, one speaker presented an analysis of a country's housing market that blended different data sets — some proprietary, others open source, others still purchased from commercial vendors. The combined data set contained mortgage origination information blended with up-to-date information on property prices and obligors' credit scores from credit rating agencies. The speaker noted that the value of this combined data set was greater than the sum of its individual parts. At the same time, because each part had been collected by an organisation with a comparative advantage in its provision, the speaker claimed that the cost-benefit calculus had been optimised.

However, there are technical and legal obstacles to blending different data sets. The technical obstacle is that data are often stored in different formats so it can be laborious to transform them into a common technical type. The legal obstacle is that the use of data collected by central banks is sometimes restricted only to those purposes explicitly expressed in legislation. That might mean, for example, data collected for supervisory purposes may have restrictions on its use by monetary economists situated in a different part of the central bank, even though the data might be useful for secondary purposes. One speaker argued that these types of legal strictures should be relaxed if central banks are to achieve economies of scope when collecting data and so also reduce the regulatory reporting burden borne by financial firms. In a similar vein, another speaker felt more multilateral agreements were needed to allow greater cross-border sharing of data between regulators. The speaker noted that although their central bank has access to highly granular data on the domestic counterparty exposures of their banks, it does not have access to similarly detailed data on the exposures of these domestic banks to foreign counterparties.

New approach to data and analysis

If these types of technical and legal obstacles to sharing and blending data can be overcome, the resulting quantitative increase in data might add up to a qualitatively new approach for analysing the economic and financial system (Varian (2014)). In recent decades, the dominant analytical approach inside central banks has been deductive. A deductive approach starts from a general theory and then seeks particular data to evaluate it. Suppose an analyst starts

---

(1) An example of a protective covenant prohibiting certain behaviour is a negative pledge. A negative pledge prohibits an obligor from pledging specified assets to other creditors.

by positing an accounting identity that the product of the quantity of money (*M*) and its velocity (*V*) is equal to the product of the price level (*P*) and expenditures on goods and services in the economy (*Q*).[1] If the analyst further assumes that the velocity of money is stable, then an increase in money might be hypothesised to result in inflation.[2] The analyst might then seek to test the validity of the theory using money and price data over a particular period of time.

An alternative research strategy is induction. An inductive approach starts from data and then seeks to generate theoretical explanation of it. Induction may mitigate confirmation bias, that is, the tendency to seek data which confirms *ex-ante* assumptions. For instance, one speaker at the event noted that many commentators have simply assumed that the recent wave of defaults in US sub-prime mortgages was caused when adjustable rate mortgages reset to a higher level. According to the speaker, however, careful analysis of granular mortgage data actually revealed that subprime mortgage defaults spiked before their initial, so-called 'teaser' rate expired.

Of course, deduction and induction are ideal types. In reality, explanatory approaches are always mixed.[3] Nevertheless, the reason why a more self-consciously inductive approach was advocated by some event participants is that the recent crisis punctured many previously dominant deductive models which purported to explain how economies and financial systems work universally. So in the aftermath of the crisis, a space has emerged for a form of induction called abduction. In other words, inferring the best explanation for a particular puzzle given patterns in the data, without pretence to making generalisable theoretical claims. Hence why some commentators have read the new emphasis on big data as 'an epistemological change' for economics from 'a science based on the notion of the mean and standard deviation from the 'normal' to one based on individual particularity' (Taylor, Shroeder and Meyer (2014)).[4]

4   Conclusion

This report has summarised the Bank's recent 'Big Data and Central Banks' event and placed it within the context of the organisation's new strategic approach to data analysis. In brief, and to paraphrase one event participant, the new approach involves a shift in tack from analysing structured, aggregated sample data collected with a specific set of questions in mind, to analysing data that is more heterogeneous, granular and complete such that these data are fit for multiple purposes. Throughout the report, emphasis has been placed on the ways bigger and better data might enhance the Bank's analytical toolkit and improve its operational efficiency, with the end goal being to promote the good of the people of the United Kingdom by maintaining monetary and financial stability.

Viewed in isolation, central banks' increasing interest in big data might be viewed as a conjunctural phenomenon, that is, as a response to the recent financial crisis. However, viewed more broadly, it appears instead to reflect a more fundamental structural shift toward the exploitation of big data by other economic agents (Bholat (2013)). This broader embrace of big data has both supply and demand sources. On the supply side, increases in the volume, velocity and variety of data have been driven by technological advances that have increased storage capacity and processing power while lowering costs. And on the demand side, there is increasing interest from economic agents in understanding how analysis of their data might enhance productivity and profits (Bakhshi, Mateos-Garcia and Whitby (2014), Brown, Court and McGuire (2014) and Einav and Levin (2013)).

To date, big data tools and techniques have had less of an impact on financial services than they have had on other sectors of the economy such as the information and communications industry. However, the situation appears to be changing rapidly. Some of the largest and most established banks are now taking a fresh look at their customers' transactional data to tailor their customer offer and to enhance early detection of fraud (Davenport (2014)). At the same time, some of the new 'challenger' financial services providers are using innovative risk models that exploit novel sources of data like social media (King (2014)). Taken in sum, these developments may have a positive impact on financial stability over the long term if they improve the financial decisions made by firms and their counterparties. But there are also nearer-term risks if early adopters of big data significantly disrupt the business models and profitability of incumbent firms. The impact of big data on the wider economy may be similarly double-edged. While it might boost productivity and lower costs, it may also alter the productive structure of the real economy and wealth distribution in ways that are difficult to forecast and measure (Rifkin (2014)). Given the possible pace and depth of these changes, central banks likely will need to make further advances to 'nowcast' (Bell *et al* (2014)) the economy, building on existing initiatives to exploit timelier data on prices, employment and output.

In sum, big data is likely to become a topic of increasing interest to central banks in the years ahead. This is because it is likely to change both the internal operations of central banks, and transform the external economic and financial systems central banks analyse.

(1) The equation of exchange: *MV = PQ*.
(2) The quantity theory of money.
(3) For instance, the way data are measured and the metadata categories used to record them are laden with theoretical assumptions, and the causal interpretation of correlated data is always made with some priors in mind. For a different perspective and stronger argument for induction see Anderson (2008).
(4) Mullainathan (2014) provides concrete examples of inductive and abductive approaches to economic analysis.

Appendix

Key questions for first three panels

Background

1 Briefly describe the nature of the data, eg the number of observations and the attributes.
2 If relevant, how is metadata developed and maintained?
3 How frequently is the data received?
4 Who are the main users of the data?
5 Why was a big data approach more appropriate than a conventional data approach?
6 If applicable, what other information is blended with this data?

Processes and problems

7 What have been the main operational and/or analytical problems encountered?
8 How much has this big data project cost or saved the organisation?
9 Who can access this data and under what conditions?
10 How is data quality checked?
11 What procedures exist to deal with non-reporting or the misreporting of data?

Technology

12 How is the data stored?
13 How is the data secured?
14 What hardware and software is used?
15 How was the hardware and software chosen?

Analysis and results

16 What analytical techniques are used on the data?
17 If applicable, how is the data visualised?
18 Briefly describe some of the main analytical results from exploiting the data.
19 How has the data enabled better policy decisions?

Key questions for fourth panel

1 Why was a new approach to data required?
2 What criteria determine whether data is purchased or directly collected by the central bank?
3 How is data integrated across the organisation?
4 How have staff's skills been developed to deal with big data?
5 What strategies were used to get senior leadership buy-in around big data?
6 What data governance arrangements exist in the organisation?
7 What have been the main legal and security issues encountered?

References

Anderson, C (2008), 'The end of theory', *Wired Magazine*, available at http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory.

Bakhshi, H, Mateos-Garcia, J and Whitby, A (2014), 'Model workers: how leading companies are recruiting and managing their data talent', available at www.nesta.org.uk/sites/default/files/model_workers_web_2.pdf.

Bank of England (2014), 'Strategic Plan: background information', available at www.bankofengland.co.uk/about/Documents/pdfs/stratplanback.pdf.

Bell, V, Co, L W, Stone, S and Wallis, G (2014), 'Nowcasting UK GDP growth', *Bank of England Quarterly Bulletin*, Vol. 54, No. 1, pages 58–68, available at www.bankofengland.co.uk/publications/Documents/quarterlybulletin/2014/qb14q106.pdf.

Benos, E and Sagade, S (2012), 'High-frequency trading behaviour and its impact on market quality: evidence from the UK equity market', *Bank of England Working Paper No. 469*, available at www.bankofengland.co.uk/research/Documents/workingpapers/2012/wp469.pdf.

Benos, E, Wetherilt, A and Zikes, F (2013), 'The structure and dynamics of the UK credit default swap market', *Bank of England Financial Stability Paper No. 25*, available at www.bankofengland.co.uk/research/Documents/fspapers/fs_paper25.pdf.

Bholat, D (2013), 'The future of central bank data', *Journal of Banking Regulation*, Vol. 14, No. 3–4, pages 185–94, available at www.palgrave-journals.com/jbr/journal/v14/n3/full/jbr20137a.html.

Brown, B, Court, D and McGuire, T (2014), 'Views from the front lines of the data-analytics revolution', *McKinsey Quarterly*, March, available at www.mckinsey.com/insights/business_technology/views_from_the_front_lines_of_the_data_analytics_revolution.

Davenport, T (2014), 'Big data at work: dispelling the myths, uncovering the opportunities', *Harvard Business Review*, Boston.

Davey, N and Gray, D (2014), 'How has the Liquidity Saving Mechanism reduced banks' intraday liquidity costs in CHAPS?', *Bank of England Quarterly Bulletin*, Vol. 54, No. 2, pages 180–89, available at www.bankofengland.co.uk/publications/Documents/quarterlybulletin/2014/qb14q207.pdf.

Einav, L and Levin, JD (2013), 'The data revolution and economic analysis', *NBER Working Paper No. 19035*, available at www.nber.org/papers/w19035.

Finan, K, Lasaosa, A and Sunderland, J (2013), 'Tiering in CHAPS', *Bank of England Quarterly Bulletin*, Vol. 53, No. 4, pages 371–78, available at www.bankofengland.co.uk/publications/Documents/quarterlybulletin/2013/qb130408.pdf.

Flood, M, Lemieux, V, Varga, M and Wong, B L (2014), 'The application of visual analytics to financial stability monitoring', *OFR Working Paper No. 14*, available at www.treasury.gov/initiatives/ofr/research/Documents/OFRwp2014-02b_FloodLemieuxVargaWong_ApplicationVisualAnalyticsFinancialStabilityMonitoring_revised.pdf.

Grant, J, Zibin, J, Bholat, D and Boukaddour, S (2014), 'Conference on Monetary and Financial Law', *Bank of England Quarterly Bulletin*, Vol. 54, No. 3, pages 330–34, available at www.bankofengland.co.uk/publications/Documents/quarterlybulletin/2014/qb14q307.pdf.

Haldane, A (2013), 'Why institutions matter (more than ever)', available at www.bankofengland.co.uk/publications/Documents/speeches/2013/speech676.pdf.

King, B (2014), *Breaking banks: the innovators, rogues and strategists rebooting banking*, Wiley, Singapore.

McLaren, N and Shanbhogue, R (2011), 'Using internet search data as economic indicators', *Bank of England Quarterly Bulletin*, Vol. 51, No. 2, pages 134–40, available at www.bankofengland.co.uk/publications/Documents/quarterlybulletin/qb110206.pdf.

Merrouche, O and Schanz, J (2009), 'Banks' intraday liquidity management during operational outages: theory and evidence from the UK payment system', *Bank of England Working Paper No. 370*, available at www.bankofengland.co.uk/research/Documents/workingpapers/2009/wp370.pdf.

Mullainathan, S (2014), 'Big data and the inductive method to theory testing: a framework with applications', speech delivered at the Royal Economics Society Annual Conference, University of Manchester, Manchester, available at http://live.wavecast.co/res-annual-conference/session/hahn-lecture-big-data-and-the-inductive-method-to-theory-testing-a-framework-with-applications/.

Rifkin, J (2014), *The zero marginal cost society: the internet of things, the collaborative commons, and the eclipse of capitalism*, St. Martin's Press, New York.

Taylor, L, Schroeder, R and Meyer, E (2014), 'Emerging practices and perspectives on big data analysis in economics: bigger and better or more of the same?', *Big Data & Society*, July–September, pages 1–10, available at http://bds.sagepub.com/content/1/2/2053951714536877.full.

Varian, H (2014), 'Big data: new tricks for econometrics', *Journal of Economic Perspectives*, Vol. 28, No. 2, pages 3–28.