BRITISH BUSINESS STATISTICS - PUBLIC SECTOR AMBIVALENCE AND PRIVATE SECTOR OPPORTUNITY?

Stephen Penneck

1. Today's theme has been 'Statistics for Business, About Business', and we have had contributions which have illustrated:

- some of the technical challenges of sampling and estimation;
- the importance of good communications to maximise the utility of the statistics;
- how to use data to improve decision making in the private sector;
- the progress that the Open Data Institute has made;
- and the work that the Bank of England does to extract the key message.

2. What I want to do in my contribution in bringing some of these themes together is to suggest that if you trace the history of official business statistics in the UK, government has been ambivalent in its approach, sometimes seeing business statistics as no more than a building block for macroeconomic statistics. To some extent this reflects changes in the emphasis given to industrial policy, but the effect has been, at times, to limit the usefulness of official business statistics for business purposes. More recently the private sector has developed extensive transactional databases used for marketing and customer focus, which could be more widely used and exploited. And the future for users of business microdata may be in the growing opportunities that private sector data present.

3. So my talk is in three parts. Firstly, I want to set out what makes business statistics special; how business data differ from other data. And I will be doing that using a historical perspective, showing how these differences emerged.

4. I then want to show how the balance between business statistics for micro and macro-economic uses changed during the second half of the 20th century.

5. And finally, some reflections on Open data and its implications for public sector and private sector business data.

What makes business statistics special?

6. There are a number of features of official business data that set them apart from other statistical data, and have an impact on how they are treated statistically and how they can be used.

<u>Their origin from administrative records</u>

7. The early history of British business statistics is a history of statistics derived from administrative records. Customs data, the earliest form of business statistics, which together with price statistics and population censuses form the tripod on which early statistical systems were based, were first collected in the UK in the 17th century when trade protection became a major instrument of government policy. They were not regularly published until the 1830s. Together with tables of statutory returns of corn sales and prices, information about shipping, railways, wages and employment, these formed the stable of 19th century economic statistics. The genesis of business statistics in administrative data reflects the growing interest of government (particularly in the 19th and early 20th centuries) in the protection and regulation of industry and trade.

8. Although there were agricultural censuses from 1866 and a census of wages from 1886, it was not until the early 20th century, with the first UK census of production (1907) that business data

collection became a statistical rather than an administrative function. The origin of business statistics in administrative functions has left a legacy in the terminology used for business surveys: until quite recently – ONS business survey staff frequently baffled their social survey colleagues with reference to 'returned data' instead of 'survey responses' and 'forms' rather than 'questionnaires'.

9. Things changed in the early 20th century with the first of a pattern of censuses of production followed by the growth of short term surveys, and was given greater impetus during the Second World War. Throughout the middle of the 20th century there were a series of government reports which criticised the inadequacies of economic statistics and, by the 1970s an almost wholly survey based approach to business statistics had been developed. That was a very necessary development, but some of the benefits of administrative sources, have only recently again been fully appreciated.

The use of a register

10. The use of administrative data sources for early business statistics naturally led to the development of business registers. These have been far more fully developed than for population statistics, for example. In the early days each administrative source was essentially also a register and there was a plethora of limited and inconsistent registers for statisticians to try to use, based on returns under the Factories Acts, company registrations, trade directories, some field enumeration for retail trades done in conjunction with the population census, and information from employment exchanges. It was not until the 1970s that a comprehensive statistical register began to be developed, based on continuously updated tax data – VAT and PAYE) and recognising the possibility of a family of different register units (establishments, enterprises, local units, etc).

The quantitative, accounting nature of the data

11. As the title of today's seminar makes clear, business statistics are data about business. Analysts are interested in structural data, which essentially shows the contribution of businesses and industries to the economy; in product data, giving information about sales of specific products; and in short term data, which charts business performance. These include data on turnover, inventories, purchases, wages and salaries, capital expenditure, research and development expenditure etc.

12. Many of these data are found in company accounts, though usually not to the definitions or timeliness needed for economic analysis. Although company accounts are not a primary statistical source, the nature of business data does mean that it is quantitative, and that there are accounting relationships apparent within the datasets that are collected. So it is possible to add and compare the cells in a questionnaire for each business, allowing detailed validation checks at the micro level. This aspect makes the use of web based data instruments attractive, allowing validation and correction of data to take place at the point of data entry, by the respondent, improving accuracy and reducing statistical agency costs.

13. Product data was an important component of business data collection in the early years, using standard lists of products agreed with businesses in advance. Their importance reflected the policies of the time – the high profile of discussions on trade protection policy and the interest in strategically important industries.

14. The complexity of business data and their derivation from accounting sources made them particularly suited to collection by postal questionnaire, rather than face-to-face interview as became the predominant form of social survey data collection. This required the development of paper handling and scanning capabilities.

Their confidentiality

15. Given that many business data are about the performance of businesses and their relative competitive position, they are regarded by businesses as highly confidential, revealing much about the costs, sales, profits and investment of individual businesses. Concerns about confidentiality

played a very large role in the public debates about the growth of the business statistics system in the 20th century, and feature strongly in the parliamentary debates for the 1907 census and the 1947 Statistics of Trade Act, which continues to be the legislation governing business statistics, and requires data to be kept confidential.

16. Business data are only made available for research purposes under controlled conditions, and any publicly released findings have to be subject to complex disclosure control.

The skewed nature of the data

17. The skewedness of most business data marks them out from many other data:

Employment	No. of units (thousand)	Proportion of total turnover (%)
0-1	814	2.3
2-5	911	5.9
6-10	41	5.0
11-20	18	4.4
21-100	21	12.9
>100	21	69.5

Source: the UK business register, November 2011.

18. The largest 21,000 out of 1.8m businesses produce 70% of UK business turnover. At time, the concentration of British industry has been a focus of government policy, leading to the development of concentration ratio analysis.

19. Disclosure control, the checking of unit record or aggregate data to ensure they do not reveal information about individual businesses, is more difficult because of the skewed nature of most business data, and can render public analysis almost impossible. In most industries there is some concentration of economic activity in a relatively few businesses, so simply removing identification details from data sets does not make them non-disclosive. Despite these difficulties, statistical offices go to some length to make business data available for research where this is possible.

The use of sampling

20. The move to surveys did not immediately mean the introduction of sampling. The early surveys were complete censuses, in the sense that they aimed to collect information from all businesses in the UK, although the very smallest businesses were not included. Sampling was not introduced into business surveys until 1952, and has generally been through the use of stratified random samples, with industry and size as the main stratification variables. This produces efficient samples that are straightforward to manage. The introduction of comprehensive registers made sampling easier, and made it possible to co-ordinate sampling across surveys.

The availability of a framework

21. The fact that many business statistics come from business accounting systems means that they can readily be fitted into a national statistical framework: the national accounts. This defines relationships between statistical entities at a national level, improving coherence between statistical measures and enhancing statistical analysis.

22. The main development of national accounts took place in the early 20th century – the coming of the censuses of production, the development of macro-economic thinking, and the

statistical needs of the Second World War provided important stimuli, and led to an international standard (the SNA) which aims to ensure comparability of economic statistics across the world. The first census of production led to the publication of a first rudimentary output measure of UK GDP in 1912 – perhaps the first in the world?

The changing balance between micro-economic and macro-economic uses for business statistics

23. Now let me turn to the changing balance between micro- and macro-economic statistics. As we have seen, the initial driver for the development for business statistics in the 19th and early 20th centuries was government industrial and trade policies. This led a system of business statistics that provided detailed product statistics, and statistics of capital expenditure (albeit at infrequent intervals) which enabled assessments of trade protection policy and industrial capacity during wartime to be made. Some of the analysis published from the census from the 1930s included concentration ratios and operating ratios.

24. A number of reports on economic statistics - Balfour (1927), Macmillan (1931), Nelson and Hopkins – highlighted their inadequacies. But it took the Second World War and the move to economic planning for post war reconstruction and stability to put economic statistics on a firm legislative footing. The White Paper on Employment Policy of 1944 had significant implications for economic statistics: *'The Government are prepared to accept in future the responsibility for taking action at the earliest possible stage to arrest a threatened slump.'* This was having noted the depression that followed the First World War. *'This involves a new approach and a new responsibility for the state.' 'Many of the decisions . . will depend on quick and accurate diagnosis.' 'It is therefore vital for (the government) to obtain, more fully and much more quickly than they have in the past, exact quantitative information about current economic movements.'* The White Paper called for short term statistics on employment, production and consumption, etc and an annual census of production. This White Paper led to the 1947 Statistics of Trade Act which is still the legal basis of business statistics in the UK, giving the government broad powers to collect a range of business and employment statistics.

25. The changing focus of economic policy, away from trade protection and industry regulation towards macro-economic management, together with the rapid rise to pre-eminence of national accounting in the second half of the 20th century led to a change in emphasis of business statistics with the needs of national accounts steadily becoming the more important driver.

26. Further government reports, Verdon Smith (1954) and Radcliffe (1959), led to piecemeal improvements in statistics in the 1950s and early 1960s. The House of Commons Estimates Committee reported on government statistics in 1966, commenting on the contribution of statistics to a range of government activities:

1. Short term forecasting: national income and balance of payments - forecasting came originally from wartime needs of adjusting demand to match available supply, but had moved to a need to determine short term stabilisation measures.
2. Manpower planning - need for better knowledge of the stock of qualified manpower to devise manpower policies
3. Productivity prices and incomes - information to measure the impact of government policies was poor.
4. Regional planning - policies were still in their infancies. Need to promote balanced national growth
5. National plan - The 1965 national plan required estimates of output, exports, employment and investment by industry. Consistent statistics were not available.

27. The Committee commented that 'It is not inevitable, and in your Committee's views it is undesirable, that the GSS should regard industry primarily as a source of information rather than as a consumer of information.' They explained further: 'The need for more disaggregation and greater comparability of statistics is a consequence of the development of micro-economic regulators to supplement the overall control of the economy by budgetary and other means.' The Committee concluded by calling for more timely statistics, in greater detail and greater coherence. It particularly called for government to produce the statistics that industry needed.

28. In 1968 Jack Stafford, then Director of Statistics at the Board of Trade commented: ... *'it has become clear that industry is looking to the Government to provide more up to date statistics in greater commodity detail.... to permit the analyses which industrial economists are increasingly wishful to make. The government's own requirements (reflect) the growing need to be informed about the developing structure of industry, the pattern of its activities, and markets, and its changing location. The Report from the Estimates Committee underlines the desirability of developing industrial statistics in these directions.'*

29. In response to the Estimates Committee Report a new system of industrial statistics was set up in the early 1970s, based around the new Business Statistics Office in Newport. The essential features were a new Central Register of Businesses, an Annual Census of Production, and new quarterly sales inquiries (which contained extensive product detail). The vision was for a continuous flow of quarterly and annual information which could be integrated through the common register and provide an information service to industry. The system would provide quarterly figures of the type which had formerly only been available from five yearly censuses.

30. It would enable substantial improvements to be made to the index of industrial production, through more complete data, better imputation techniques and a better conceptual basis.

31. The Estimates Committee report had emphasised the need for more up to date statistics analysed by industry, commodity and occupation, and in a form that would permit detailed micro-analysis by industrial economists and planners. Indeed the plan was for the new Central Register to provide a key to enable statistics on related subjects to be brought together, and to provide a databank for analytical purposes. A primary objective of the new system was to improve the quantity and quality of the short term product statistics available to users in industry.

32. In a 1971 article Laurie Berman (the new Director of Statistics at the DTI) was able to say: *'I think it is fair to say that there have been no really spectacular developments in <u>macro-economic statistics</u> in recent years* [he was talking about the previous five years]; *(but) there has been a considerable improvement in <u>micro-economic statistics</u> and more are on the way'*. Berman was able to conclude that 'the pioneering work of setting up the basic economic accounts has now ended. The main task now is to extend the coverage of existing statistics, the frequency of surveys, improve the sampling frame, fill information gaps and improve quality.'

33. The 1970s then, in retrospect, may be seen as the heyday of detailed official business statistics. But government's interest in industrial planning did not last and the building of the first computer model of the economy led to a focus on forecasting at the macroeconomic level. Government's use of economic statistics became increasingly a focus on the national accounts. The contribution of commodity industrial detail to this picture was limited. So when the 1979 Rayner review stated that government statistics should be produced primarily for government, government's needs by then were clearly macro-economic.

34. The Rayner report set a target of 20% for reductions in staff working on DTI statistics. The report questioned the need for the quarterly sales inquiries and for an annual census of production.

As a result, the number of product headings in the quarterly inquiries was cut substantially, and sample sizes were reduced. Similar reductions were made to the Annual Census of Production.

35. In a speech to the ISI in 1983, Sir John Boreham, then head of the GSS said: *'in the UK the GSS now concentrates heavily on serving government. This is a shift in emphasis since 1981; before that, serving society (which includes industry) received more weight; it still gets some weight but less than before 1981.'*

36. He pointed out that although the Statistics of Trade Act provides a power to provide a statistical service to industry, it was not a requirement.

37. Ray Ash, Director of the BSO in an 1985 article on the Business Statistics Office after 15 years, reported that*: 'The climate which favoured the expanded collection of official statistics began to change in the mid 1970's with rising concern about cost of public expenditure and complaints from small firms about the costs of providing statistical information. The change in the climate became more pronounced during the last five years with the government's policies of less intervention in industry and of cutting back the civil service.'*

38. Businesses as form fillers were becoming more vociferous. The need to complete detailed product headings on a quarterly basis had not been popular, and the Armstrong Rees review was commissioned by DTI in 1988 to investigate this issue. It recommended a further halving of the number of quarterly product headings, and further sample reductions. The review pointed out that production industries only accounted for 25% of GDP [it is now down to 16%]. The new system of industrial statistics of the early 1970s had lasted barely a decade. It had been based on a vision of industrial planning which had not lasted and provided a focus on an industrial structure soon to decline.

39. By the late 1980s the vision of the early 1970s had not been realised, and criticism of the national accounts (principally the size of statistical discrepancies and revisions) led to the 1989 Pickford Report. But Pickford said: 'our impression is that much of the product detail collected by DTI is no longer necessary. Users in the Treasury confirmed that little regular use is made of input-output tables for forecasting purposes.'

40. The Pickford report made the case for statistical improvements for macro-economic purposes and led to all business statistics being brought into the then Central Statistical Office, cementing a closer bond between business statistics (previously collected by DTI) and the national accounts (produced by CSO).

41. Two 'Chancellors' initiatives' in 1990 and 1991 provided new resources for economic statistics, but mainly to improve macro-economic statistics. Product statistics received a boost with a new product survey Prodcom, set up in 1993 to meet a new European regulation. But although since then there have been statements culminating in the Statistics and Registration Services Act 2007 recognising that statistics are produced for all users, in reality budget constraints have meant that while, in an article published in 2000, ONS were able to claim a 'decade of improvements in economic statistics', this was based on greater investment in macro-economic statistics, with little direct benefit for business users.

Open data – the challenge to all data producers

42. So that is how, as I see it, business statistics have developed, most recently to meet macro-economic needs. But the need for statistical analysis to inform a number of current issues requires analysis of data at much lower levels than can be done from macro-economic data.

43. Questions such as the relationship between output growth and productivity, investment and growth, and the benefits from innovation require statistical modelling and analysis of microdata. For example, the recent study by Field and Franklin at the ONS used microdata from the Annual Business

Survey show a widening of the productivity gap between the more productive larger businesses and less productive small businesses. ONS has had procedures in place for several years to allow access to microdata for research purposes under secure conditions. The Virtual Microdata Laboratory has had over 400 academic and government users since 2004, researching topics on productivity, earnings, skills investment and business demography, using a variety of business surveys and the business register.

44. The move towards more Open data respects the need for data to be protected where it is confidential. The essentially confidential nature of business data has been explained earlier, and ONS has squared the circle of maximising its public utility while preserving its confidentiality through a series of access arrangements. The ONS allows:

- Publication of detailed local area and disaggregated statistics on our Neighbourhood Statistics and NOMIS web sites.
- Unrestricted access to anonymised census data sets which are available on request
- Access to anonymised social survey data sets to researchers, by registration with the ESRC Data Archive
- Access to identifiable but not identified business and social survey datasets to approved researchers on premises and with technology licensed by ONS, with disclosure checks provided on resulting outputs. Each request for access is considered by the ONS Microdata Release Panel.

45. In response to a user request, ONS now publishes the price quotations it collects for the consumer prices index. Although these are collected in confidence from retailers, confidentiality has been maintained by aggregating geographical data. A Data Explorer function is being developed for the web site, which will enable users to access data sets and perform analysis on the web site, and also an Application Programming Interface (API) which will support open standard formats and allow third parties to pull data out of the web site to use on their own web site.

46. ONS has looked at the potential use of Google Trends data on frequency of searches to validate official data from the retail sales index survey. This weekly data has potential for quality assuring and supplementing official survey data. The results showed that Google data were especially useful for non-frequent purchases.

47. Private sector organisations now use the web as a data source. It is trawled daily by a number of organisations for huge numbers of prices, for price comparison web sites, but also regularly put together by organisations such as State Street's PriceStats Index. They produce daily price indices for 19 countries, and, even though they only cover the large supermarket chains, the results mirror official consumer price indices very closely, not surprisingly perhaps given the high degree of concentration in retailing.

48. Companies routinely use data from web searches to refine their web sites to ensure customers are offered services related to their buying history. Loyalty card data help businesses decide where to open new stores or which advertising campaigns to run. Forty years ago retailers would be looking to government to collect the data on which such decisions could be based. Interestingly, Tesco has announced an initiative (Clubcard Play) to open up their Clubcard data to their customers.

49. So, data from the private sector make a growing contribution to the picture of British businesses. The Open Data Institute recognises this. They argue that many organisations do not recognise the potential of the data that they hold, but can frequently get direct benefits from publishing it as open data. So, for example, publishing open data can:

- increase custom for an organisation's main (paid-for, non-data) product or service
- ease information sharing within the organisation and with partners and customers

- help provide a wider audience for published information
- provide incentives for others to help maintain datasets, reducing the maintenance cost for the organisation
- encourage innovation around the organisation's data (eg new analysis techniques they can take advantage of)
- help third parties to develop skills and services which the organisation can contract rather than developing in house

50. Much data produced by private sector organisations has commercial value with access controlled through copyright laws. Advocates of Open data argue that this is contrary to the common good, and the Open Data Institute aims to help organisations identify data they own that can be opened up to realise these benefits.

51. Although private sector transactional data - be they of payments, service usage, etc - are clearly confidential between the two parties, they can easily be anonymised by removing identifying fields, and this does not generally reduce the value of the data. In addition, transactional data covers whole populations and is not based on samples, so the impact of sampling on analysis of small attributes does not arise.

52. So what does the future hold and what questions arise as businesses begin to look more towards their own data holding for business analysis, rather than to government?

- How can we encourage private sector data (which have a high market value) to be more widely available?
- How far are these developments common across Europe and does Eurostat have a role in facilitating them?
- As the focus moves to private sector data collection, is the government missing anything, by not collecting the data themselves?
- And if the Estimates Committee met again, what would they make of this and would they now see business needs for data as largely being met by businesses themselves?

References

Armstrong A and Rees D, Review of DTI statistics, DTI, Dec 1988

Ash R, The Business Statistics Office after 15 years, Statistical News No 68, Feb 1985

Ayoubkhani D and Swannell M, Using Internet Search Terms to Help Quality Assure Movements in Retail Sales Survey. Survey Methodology Bulletin No. 70. March 2012

Berman L S, The Central Register of Businesses, Statistical News, CSO, 1969

Berman L S, Recent improvements in official economic statistics, JRSS, Series A, 1971

Chamberlin G, Googling the Present, ONS, ELMR, Dec 2010

Fessey M, Some developments in Economic Statistics since 1934, The Statistician, 1969.

Field S and Franklin M, Microdata perspectives on the UK productivity conundrum, ONS, January 2013

Jenkinson G and Brand M, A decade of improvement in economic statistics, Economic Trends, May 2000

Rayner D, Review of Government Statistical Services, 198

Tily G, John Maynard Keynes and the development of national accounts in Britain, 1895 to 1941, Review of Income and Wealth, June 2009

Pickford S et al, Government economic statistics, a scrutiny report, 1989

Ritchie F, Secure access to confidential micro-data: four years of the Virtual Microdata Laboratory, ONS ELMR, May 2008

Smith P and Penneck S, 100 years of the Census of Production in the UK, ONS. GSS Methodology Series No 38, April 2009.

Sorrell A, New statistics and the Index of industrial production, Statistical News No 15, Nov 1971.

Stafford J, The development of industrial statistics, Statistical News, CSO, Statistical News No 1, 1968