

THE DEVELOPMENT OF THE UK ANNUAL BUSINESS INQUIRY

Gareth Jones¹

ABSTRACT

This paper describes the ABI project to integrate ONS' main annual surveys of businesses across all economic sectors and also to bring together surveys of employment and financial data. The ABI is also capable of generating a wide range of sub-national analyses, another objective of the development. Methodological aspects covered by the paper include sample design, estimation and outlier treatment, apportionment of data from reporting units to local units (individual sites) and the methodology for sub-national and small area estimation. The sub-national methodology involves the use of logistic and loglinear models.

KEY WORDS: Survey integration; Coherence; Sub-national analyses; Logistic and loglinear models.

1. INTRODUCTION

This paper describes the development of the UK Annual Business Inquiry (ABI), a new integrated survey of employment and accounting information from businesses and other establishments in most industry sectors of the economy. The ABI replaces the following annual survey systems:

- Annual Employment Survey (AES)
- Annual Censuses of Production and Construction (ACOP/C), which include the Purchases Inquiry (PI)
- The six annual Distribution and Services (DSI) inquiries, viz.
 - Annual Wholesale Inquiry
 - Annual Retail Inquiry
 - Annual Motor Trades Inquiry
 - Annual Catering Inquiry
 - Annual Property Inquiry
 - Annual Service Trades Inquiry

In addition to the restructuring and integration of inquiries, major improvements in methodology have been implemented in a standardised way across the ABI. The ABI provides more coherent and consistent annual business statistics covering a range of variables for the whole economy. Various outputs from the ABI have helped to improve the quality of the national accounts. Also, the quality of "per head" type statistics, for example output per head, have been improved considerably.

2. BACKGROUND

The ABI project took place against the background of two significant developments in business statistics:

- The successful implementation of the Inter-departmental Business Register (IDBR) during 1994 and 1995. This laid the foundation for a more integrated "whole economy" inquiry.
- The introduction by Eurostat (the Statistical Office of the European Communities (EC)) of a new EC Regulation on Structural Business Statistics (SBS) which extended the data requirements of earlier EC Directives in a number of ways.

¹ Dr D Gareth Jones, UK Office for National Statistics, Room 1.202 Gov't Bldgs, Cardiff Road, Newport, NP10 8XG, UK. E-mail: gareth.jones@ons.gov.uk

In particular the SBS regulation covered the Services sectors of the economy for the first time. Its full provisions came into force for the 1999 survey year.

The systems replaced by the ABI are listed in the introduction above. The Annual Employment Survey collected information on number of employees (with a four way split between male/female, full/part time) and industry classification according to the Standard Industrial Classification, SIC (92). These data were collected separately for each site belonging to businesses in the survey sample. The AES (which began in 1995) was used primarily for the annual estimates of employee jobs and itself replaced earlier censuses of employment which had been carried out less frequently in earlier years. The data obtained from the AES also provided important information for the updating of the business register, particularly with regard to the internal structure of enterprises, the so-called "local units" or individual sites at which enterprises operate. Such information is important in producing regional or other geographical analyses of business data.

The Annual Censuses of Production and Construction and the Annual Distribution and Services Inquiries were sample surveys collecting annual accounting information from businesses, usually in more detail than is obtainable from company accounts. The Annual Census of Production also collected some data on employment, though on a slightly different basis to the AES. The accounting data obtained from these surveys was used for a variety of purposes. One of the most important was the estimation of Gross Value Added which is used in the National Accounts for the estimation of Gross Domestic Product (GDP). The Purchases Inquiry was a subsample of the ACOP sample from which was collected more detailed information on the breakdown of purchases of goods and services by businesses. Approximately one fifth of industries in the Production sector were surveyed each year on a rotating five-yearly basis.

3. OBJECTIVES

The objectives of the ABI project can be categorised under four broad headings:

- Consistency, standardisation and improved methodology
- Meeting international requirements
- New analyses, especially in the Services sectors
- Quality assessment and improved maintenance of the business register.

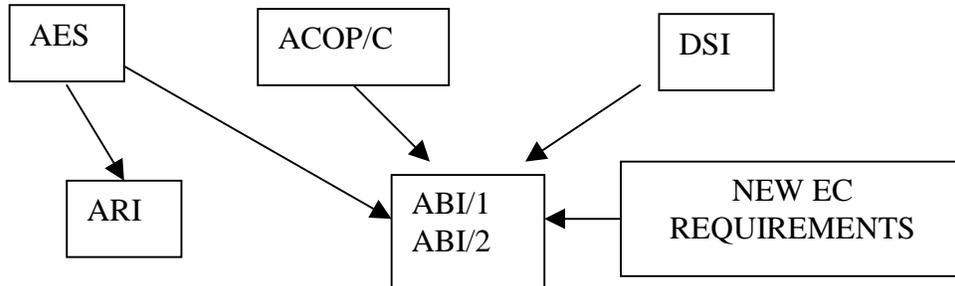
By combining the collection of employment and accounting data into the same survey the validity of productivity measures is improved since value added and employment are estimated on a consistent basis through the use of common reporting units and industry classifications. The lack of comparability between employment and value added had been a serious problem with the previous systems. The integration of industry sectors into one survey avoids the problems of omissions and double counting which were previously possible. A common methodology is also applied across all industry sectors, including standardisation of definitions. Also, the ABI has UK coverage as required for the National Accounts and Eurostat whereas the AES did not cover Northern Ireland.

The EC SBS Regulation set additional requirements which have largely been met by adding questions to the 1999 ABI. The Regulation also requires the use of standard definitions for international comparability. A set of standard analyses is also specified and has been adopted for the ABI. These analyses include regional and size breakdowns as well as the more traditional industry sector analyses. Many of the EC requirements as well as other analyses available from the ABI system cover all industry sectors, including services. In particular, regional analyses of value added in the service industries are now available

As will be explained below, a second inquiry known as the Annual Register Inquiry (ARI) has been developed in parallel with the ABI. The ARI serves the purposes of annually measuring various aspects of register quality and also improving the updating of the register and consequently its quality. It is designed in such a way as to meet the requirements for register updating laid down in the EC Business Registers Regulation and provides information on size (measured by employment), industry classification and location of both complete enterprises and the individual sites at which they operate (local units).

4. SYSTEM STRUCTURE

The diagram below shows the relationships between the previous systems and the ABI and ARI.



Employment data are now collected along with accounting data in ABI and the sectoral inquiries ACOP/C and DSI are merged. The ABI form is in two parts, one dealing with employment data (ABI/1) and the other with accounting data (ABI/2). The employment data relate to a December reference date and both parts of the form are despatched around the end of the year. The main reason for splitting the form into two parts is that the employment data are available from businesses much earlier than their accounting information. Data collection for ABI/1 is closed down at the end of March of the following year whereas the closedown for ABI/2 is about six months later in order to allow businesses to provide information from their own annual accounts. The early return of the ABI/1 part of the form allows publication of employment estimates to follow a timetable similar to that which has existed for the AES. Another reason for using a two-part form is that employment information and accounting information are often provided by different parts of an organisation and the two sections of the form can be addressed to different individuals.

The sample size of ABI/1 in 1998 was approximately 78,500 enterprises. ABI/2 is a subsample of ABI/1 in which some of the industry sectors are not covered and has a sample size of approximately 75,000. ABI/1 covers divisions 2-93 inclusive of SIC (92) - i.e. all divisions except Agriculture, Private Households with employed persons and Extra-territorial organisations. In 1998 ABI/2 covered the above industries except for divisions 2 (Forestry), 5 (Fishing), 65-67 (Financial Services), 75 (Public Administration) and the public sector in Education (division 80) and in Health and Social Work (division 85). Doctors, dentists and charitable organisations in division 85 were also omitted. It is expected that the industry coverage of ABI/2 will be expanded in future years (e.g. SIC divisions 2 (Forestry) and 5 (Fishing) have been covered in the 2000 ABI/2). The Purchases Inquiry has been temporarily retained as a sub-sample of ABI/2, the additional questions on purchases detail being added to the ABI/2 forms for the relevant industry sectors. A phased expansion of the PI is taking place. Current plans are to expand the industry coverage each year until all ABI/2 industries are covered every year from 2001 onwards. However it may be decided to make the Purchases Inquiry a separate survey to avoid overloading the ABI/2 questionnaire.

As mentioned above, the introduction of the ABI has also been accompanied by the development of another inquiry, the Annual Register Inquiry (ARI) which began operation in July 1999. The ARI subsumes previous register proving activity and also replaces the AES in updating the local unit structures of enterprises and is designed for this latter purpose rather than for the direct measurement of employment. Large enterprises are covered every year and medium sized enterprises (20-99 employment) every four years. Below the employment threshold of 20 there is no systematic proving of enterprises due to the very large numbers of such businesses. However, part of the ARI sample is allocated to quality assessment of various register characteristics (especially industry classification). Another part of the sample is available for targeting, on an ad hoc basis, areas of the register where quality problems are known to exist. The ARI is thus a very flexible tool for updating the IDBR and for assessing and improving its quality. Approximately 400,000 local units in 68,000 enterprises are covered per year.

5. PROJECT TIMETABLE

The phased introduction of the ABI system is represented by the outline timetable below. The dates shown here are the survey years to which the data relate and not the dates when results become available.

- 1996 AES, ACOP/C, DSI
- 1997 AES, modified ACOP/C and DSI for parallel run
- 1998 AES, ABI
- 1999 ARI, ABI (with full EC requirements)

The first changes were made for the 1997 inquiries. The ACOP/C and DSI inquiries were modified to collect employment information on a basis which would be comparable with the AES for 1997. In the case of ACOP/C this meant an extension of existing employment questions and a change in definition from a year average figure to one at a given reference date. The same questions were introduced into DSI which had not previously contained questions on employment. The purpose of these changes was to meet the requirements of customers for a parallel run of the ABI methodology against that of the previous systems. The ABI results and analysis software was written for first use on the 1997 data even though the data were collected via the previous survey systems. The parallel run was intended both to evaluate the new methodology and also to estimate any discontinuities introduced in time series by making the methodological changes.

Other changes made in 1997 included a move in DSI from the use of turnover to the use of employment for size stratification of the sample design, putting DSI on the same basis as ACOP/C and the filling of some minor gaps in the industry coverage of ACOP/C and DSI. Some work on the quality of the IDBR in these industries was undertaken to support this extension of coverage. Additionally, the data collection software for ACOP/C was brought into line with DSI which already used standard data collection software.

The ABI was introduced for the 1998 inquiry year and completed the integration of the ACOP/C and DSI systems. This involved a full review of sample design, form types, questions and definitions with the perspective of a single system rather than a set of separate sectoral inquiries. The AES was retained for 1998 alongside ABI to provide a more complete parallel run for the employment questions. The transition to the new ABI and ARI systems was completed for the 1999 survey year with the launch of the ARI in July 1999 and the addition of extra questions to the 1999 ABI to meet the full requirements of the SBS Regulation. The expansion of the Purchases Inquiry will not however be complete until 2001 as described above.

The official adoption of estimates of employee jobs from ABI/1 in preference to those from AES was complicated by the fact that 1998 ABI/1 estimates proved to be about one million higher than those of the 1998 AES. An extensive research programme was needed to establish the reasons behind this difference and concluded that the ABI estimates were the more reliable. One consequence is that there is now much closer agreement between business survey and Labour Force Survey employment estimates for the UK.

6. FORM TYPES, QUESTIONS AND DEFINITIONS

In the run up to the launch of the ABI at the end of 1998 a thorough review was undertaken of all form types to be used, the questions they contained and the definitions of those questions. An attempt was made to minimise the number of form types while also maintaining form content which would appear relevant to the businesses which received the forms. The form types are customised for industry sectors and sub-sectors and in 1998 there were 3 basic form types for ABI/1 and 21 for ABI/2. In addition, for each basic form type there is usually a short form equivalent which collects information on the main totals but not the more detailed breakdowns which appear on the basic form types. Short forms are used to reduce the burden

of form filling on businesses. In ABI/2 a proportion of businesses in the sample receive the short form, this proportion increasing as the size of businesses becomes smaller. The businesses receiving the short forms are subsampled randomly from the total sample in each stratum. In ABI/1 a different approach is taken, the short forms being used for businesses which have also been sampled for the ARI or the fourth quarter short-term employment survey and also for all Northern Ireland businesses. Imputation methods are used to expand the information on short forms to the detailed breakdowns required for analysis.

The main questions asked in ABI/1 are the number of employees with a four-way split between male/female, full/part time (as in the AES) and also the number of working proprietors/partners and the number of other unpaid workers (e.g. family workers) as well as total employment which includes all of the above. Also the definitions used are consistent with the AES definition of employees.

The ABI/2 forms in general contain many more questions than ABI/1 and the range of questions is more variable across industry sectors. However there is a core set of questions covering turnover (i.e. sales of goods and services), employment costs, purchases of goods and services, taxes and subsidies, inventories and capital investment which occur on nearly all form types. The basic form types also usually contain more detailed breakdowns of these aggregates. Definitions of these quantities have been reviewed and standardised. In general the definitions used are now consistent with the requirements of the SBS Regulation which in turn are usually consistent with the European System of Accounts (ESA).

Additionally, ABI/2 contains a number of “filter questions” which are used to identify businesses with particular types of activity so as to improve the sampling frame for other more detailed inquiries. Filter questions for Research and Development activity and International Trade in Services have been included and questions on E-Commerce were included on the ABI 2000 forms.

7. ANALYSES AVAILABLE

The table below summarises the range of analyses potentially available from ABI, subject to considerations of accuracy and confidentiality.

Collected	Derived	Analysed by
Employment	Gross Value Added (GVA)	SIC Industry
Turnover	Total Output	Geography
Employment Costs	Operating Surplus	Size (Employment or Turnover)
Purchases	GVA/Turnover	Legal Status
Taxes/Subsidies	Labour Productivity	Country of Ownership
Inventories	Unit Wage Costs	Time (i.e. time series)
Investment	Inventory Ratios	
	Investment Ratios	

For most of the main variables collected, breakdowns are also available. For example, employment costs are subdivided into wages and salaries, pension contributions, social security contributions and redundancy payments. The breakdowns available for turnover vary by industry sector and in the retail sector there is a detailed breakdown by product. In addition, the Purchases Inquiry provides considerable purchases detail for the industries it covers.

A number of important variables can be derived from the data collected, notably gross value added and total output, both at basic prices, operating surplus, labour productivity and unit wage costs. Both the collected and derived variables can be analysed by a number of other variables mostly taken from the IDBR. In addition to analyses by SIC industry and size, businesses can be classified and analysed by country of ownership and by legal status (i.e. whether they are companies, partnerships, single proprietor businesses, public corporations, non-profit making bodies, central or local government). Geographical

analyses are also possible according to the EC NUTS hierarchical classification system. NUTS has five levels in the hierarchy which approximately follow the organisation of local government:

- NUTS1 Region
- NUTS2 Group of Counties or Unitary Authorities
- NUTS3 County or Unitary Authority
- NUTS4 District or Unitary Authority
- NUTS5 Ward

In addition to the above it is also possible to compare different years and build up time series as required.

8. METHODOLOGY OF THE ABI

8.1 Sample Design

The ABI is sampled from the population of reporting units on the business register (IDBR). A reporting unit (RU) is usually a single complete business but in some cases arrangements have been made with businesses to collect data either for several businesses combined or for parts of businesses of heterogeneous nature. Each RU on the IDBR also consists of one or more local units (i.e. individual sites) at which it operates. Local units are particularly important when considering any kind of geographical analysis.

The sample size of ABI/1 in 1998 was approximately 78,500. The ABI/2 sample size is slightly lower at about 75,000 because of the industry sectors not covered. Sampling is done for ABI/1 and the ABI/2 sample then automatically results by excluding the appropriate industry sectors. The sample design is a stratified random one with three stratification dimensions. Strata are defined in terms of:

- six employment sizebands (1-9, 10-19, 20-49, 50-99, 100-249, 250+).
- region (viz. England & Wales combined/Scotland/Northern Ireland).
- SIC industry.

Within England and Wales industry stratification is at 4 digit SIC level. Within Northern Ireland it is at 2 digit SIC level and within Scotland at a hybrid 2/3/4 digit level. All stratification variables are taken from the IDBR. Special arrangements have been agreed with the Scottish and Northern Ireland Offices to allocate larger than proportional sample sizes to these two regions (viz. 9,000 and 3,000 respectively).

Subject to the sample size constraints above, the sample has been allocated to strata using the Neyman optimum allocation method which minimises the expected variance of total turnover over all strata. This results in the strata corresponding to the largest businesses being completely enumerated in all industries. In most cases some of the strata in sizebands below this are also completely enumerated. In addition, cases with high register turnover and low register employment (defined as £50m or greater turnover and less than 10 employment) are also completely enumerated. This provides a limited form of stratification by the turnover variable and significantly reduces the expected variance of estimates derived from the sample. In 1999 this procedure was extended to provide complete enumeration of all businesses with high register turnover regardless of their register employment. The sample allocation is scheduled for a full review now that two full years' data are available from the ABI.

Below the threshold for complete enumeration, the sample is rotated as follows. Businesses with less than 10 employment are completely replaced each year. Businesses with employment from 10 up to the threshold for complete enumeration are given a rotation rate of 50 per cent (i.e. half are replaced each year). This system of rotation is designed to spread the form filling burden on businesses while retaining a reasonable degree of matching of the sample between consecutive years which improves the accuracy of estimates of change between years.

8.2 National Analyses

Data obtained from the sample are grossed up to the population of the IDBR using the combined ratio estimator, strata corresponding to different sizebands being combined if the sample size would otherwise be too small. This estimator makes use of register auxiliary variables to improve the accuracy of estimation.

The auxiliary variable used in grossing the employment data in ABI/1 and the employment costs data in ABI/2 is the IDBR employment. For all other data in ABI/2 the auxiliary variable used is register turnover. The choice of auxiliary variable was determined by examination of the size of expected sampling errors using different auxiliary variables.

In order to improve the robustness of the above estimator a procedure for dealing with atypical observations (outliers) has been adopted. This consists of creating additional post-strata into which the outliers are placed as though they had been completely enumerated at the sampling stage (they are also removed from their "natural" strata). The criteria for marking businesses as outliers have been determined empirically so as to provide robust estimates. In ABI/1 a business is treated as an outlier if the ratio of its reported employment to its register employment exceeds 20. In ABI/2 a business is treated as an outlier if either it is an outlier for ABI/1 or if the ratio of its reported turnover to its register turnover exceeds 50. If a business is marked as an outlier it is treated as such for all variables collected in the survey.

The extent of the sampling error which can be expected with the above sample size/design and associated estimator is as follows. The coefficient of variation of the estimate of total employment across all industries is about 0.4 per cent and the coefficient of variation of total value added across all industries is about 1.5 per cent. This means that the 95 per cent confidence interval for total employment is ± 0.8 per cent and the 95 per cent confidence interval for total value added is ± 3 per cent.

8.3 Local unit apportionment

Since the data collected in ABI is for reporting units (usually complete enterprises), the first stage in any analyses at sub-national level is to apportion reporting unit data across the local units (LUs) within the reporting unit. This apportionment is carried out by modelling the ratio $R = X$ divided by register employment for each variable of interest X . The model is in two parts. Firstly, the probability of a non-zero value is assumed to be given by a logistic linear model with independent variables register employment sizeband, register SIC and register region code. Secondly, given a non-zero value, R is assumed to follow a log-linear model with the same set of independent variables as for the logistic model. The parameters estimated for these models are then used to derive proportions for each local unit (adding to unity across all local units in a reporting unit) which determine the apportionment of the reporting unit total across the local units. The models are fitted to the reporting unit data and then applied to local units to conduct the apportionment (the independent variables being available on the IDBR for both reporting units and local units). In order to make this switch from reporting unit to local unit as reliable as possible, the larger reporting units were excluded from consideration in fitting the model (in fact, only reporting units with less than 100 employment and less than 3 local units were used to fit the model).

The modelling process also required decisions concerning the most appropriate degree of aggregation of sizebands, SIC codes and region codes for use as independent variables. It was decided to work at as high a degree of aggregation as possible without sacrificing a significant degree of "goodness of fit" of the model. This has led to the following groupings which have been used for all survey variables:

- Employment Sizebands: 1-2, 3-4, 5-9, 10-19, 20-49, 50-99, 100-249, 250+.
- Industries: 2 digit SIC for the logistic model and 3 digit SIC for the log linear model
- Regions: NUTS2 level

For the questions on employees in ABI/1 (i.e. total employees and the 4-way split) there are other data sources which may be used to apportion from RU to LU level, viz. the ARI in Great Britain and the 2 - yearly Censuses of Employment in Northern Ireland. The information on local unit distribution of employees from these sources is potentially better than that obtained from the above modelling process and so is used in preference to modelling where available.

8.4 Sub-national Analyses

Sub-national analyses of ABI data are obtained by a combination of conventional and synthetic estimation methods. At high levels of aggregation conventional estimation is used, while at lower aggregation levels

conventional methods are supplemented by synthetic estimation using IDBR data. These procedures are described below.

A post-stratification of all local units in the RU sample and population is used. The post-stratification variables are the LU region, the parent RU sizeband and the parent RU SIC industry of the LU. The parent RU sizeband is the same as is used for stratification in national analyses. Otherwise the aggregation levels are LU NUTS1 region and 2 digit parent RU SIC, except that some miscellaneous aggregation of post-strata (both in terms of region and SIC) is necessary to provide adequate sample sizes. The estimator used is the standard (uncombined) ratio estimator with register LU employment as the auxiliary variable.

The use of a different post-stratification for sub-national analyses to that used for national analyses gives rise to a problem of consistency between national and sub-national results in so far as the total of all regional results for the UK may not equal the UK value obtained from the national estimation system. To overcome this problem scaling factors have been introduced for each survey variable. These are applied to the weights arising from the sub-national estimation system and are set so as to ensure that additivity as described above is achieved. Finally it should be noted that totals and other forms of derived variables are calculated after grossing up the sample data and are not grossed separately in the sub-national system as this would give rise to another form of non-additivity or incoherence.

In principle the above methodology will provide estimates for any required sub-population (or domain, in the terminology used below) , including very small areas such as local authority wards. However, in practice, the sampling errors for very small domains are extremely large. Also some domains will contain no sampled units so that the resulting estimate for them will be zero. Although a zero estimate may be within sampling error expectations, it is presentationally troublesome since it may be known in advance that the correct domain estimate should be greater than zero.

It is not therefore realistic to use this estimation methodology for very small domains and a set of "minimum domains" for which it may be used has been imposed. These minimum domains are two-dimensional in the sense that they are defined in terms of a given level of SIC disaggregation and a given level of geographical disaggregation, each of which may be different for any particular minimum domain. The minimum domains have been defined taking into account both the sampling error associated with them and their sample size in terms of the number of local units they contain.

Below the level of minimum domains a form of synthetic estimation is used to obtain the required estimates. Essentially, if a domain is a subset of a minimum domain, estimates for it are obtained by prorating on the basis of the total register employment of the domain of interest and the total register employment of the minimum domain. Sampled units are removed before prorating and added back afterwards. If a domain straddles two or more minimum domains it is split into smaller domains which do not cross minimum domain boundaries. These are then estimated separately and the results added. The use of synthetic estimation for very small domains means that estimates of error are not available. Sampling error does not increase beyond minimum domain level, but other forms of error may be introduced and these are not quantifiable.

9. FURTHER DEVELOPMENTS

Four further areas of work are planned:

- Refinements in form types and questionnaire design with a view to reducing the compliance burden, especially for small businesses.
- Investigation of whether the accuracy of employment estimates could be further improved by using administrative payroll data as auxiliary information in estimation.
- Research into a better method of outlier treatment, especially with a view to using a two-sided rather than one-sided method.
- Possible separation of the Purchases Inquiry from ABI/2 in order to avoid questionnaire overload.