

Annual Purchases Survey: technical report

Detailed quality and methodology information about the technical procedures used to produce the Annual Purchases Survey (APS) estimates.

Contact:
Jon Gough
purchases.survey@ons.gov.uk

Release date:
13 May 2019

Next release:
To be announced

Table of contents

1. [Introduction](#)
2. [Questionnaire design](#)
3. [Sampling procedure](#)
4. [Data collection](#)
5. [Converting respondent data into published estimates](#)
6. [Publication and data dissemination](#)
7. [Disclosure control and data confidentiality](#)
8. [Revisions policy](#)
9. [Future work](#)
10. [Contact and other information](#)

1 . Introduction

This guide describes the methodological and technical procedures used by the Office for National Statistics (ONS) to produce the Annual Purchases Survey estimates. The report is aimed at users who want to know more about the background and history, uses and users, and concepts and statistical methods underlying the survey. It includes information about questionnaire development, sample design, data collection, results processing, publications and quality issues.

This technical guide will be revised in line with any major future survey developments.

1.1 Overview

The Annual Purchases Survey (APS) was reinstated by the Office for National Statistics (ONS) in 2015 to meet a range of user demands and to fulfil the recommendations for its reinstatement from two independent reviews; one by [Dame Kate Barker and Art Ridgeway \(PDF, 570KB\)](#) and the other by [Professor Sir Charles Bean](#). A previous survey, entitled the Purchases Inquiry, was suspended in 2006 due to insufficient quality in the data and to reduce the ONS costs and burden on UK businesses (final reference period being 2004). However, given the survey provides important information on the products that UK businesses purchase, it was decided that it should be reintroduced. The new survey aims to strengthen the estimates of the intermediate consumption structure of the UK economy.

The primary aim of the APS is to provide a comprehensive picture of the goods and services used or transformed in the production process and running of UK businesses, otherwise referred to as intermediate consumption. This product-level information is used in supply and use tables (SUTs), which are an integral part of the measurement of gross domestic product (GDP). The Eurostat Manual of Supply Use and Input-Output Tables recommends that benchmarked supply and use tables are produced at least every five years based on updated source data. The reintroduction of this survey will help the ONS to adhere to international best practice outlined in the [European Systems of Accounts: ESA 2010](#) and [Balance of Payments Manual: BPM6](#). The APS collects information on companies' intermediate consumption, a national accounts concept defined within the ESA 2010 manual as:

"Intermediate consumption consists of goods and services consumed as inputs by a process of production, excluding fixed assets whose consumption is recorded as consumption of fixed capital. The goods and services are either transformed or used up by the production process."

The survey collects information about businesses' expenditure on energy, services, goods and materials that are used up or transformed by the business activity. It includes raw materials, power and fuel, rental on buildings and business services such as advertising, recruitment consultancy and cleaning. It specifically excludes fixed assets or capital investment, staff costs and goods and services bought for resale without further processing.

In 2016, Annual Purchases Survey questionnaires were sent by the ONS to approximately 31,000 businesses in the UK. In the UK, it is a compulsory survey that is administered under the statutory powers of the Statistics of Trade Act 1947 for Great Britain and under the Employment (Northern Ireland) Order 1988 for Northern Ireland.

1.2 Main uses and users of the data

Intermediate consumption is required as part of the process to set the annual level of gross domestic product (GDP), which is an essential statistic for informing fiscal and monetary policy decisions. Broadly speaking, GDP is calculated by adding up the value of the output of firms less the goods and services that are used in the production of that output. This includes raw materials, power and fuel, rental on buildings and business services such as advertising, recruitment consultancy and cleaning. The value of these goods and services used in production is called "intermediate consumption".

The data collected from the APS feed into the supply and use framework, which is a central component of the national accounts balancing process and sets the annual level of nominal GDP. The [Eurostat Manual of Supply Use and Input-Output Tables](#) recommends that benchmarked supply and use tables are produced at least every five years based on updated source data. These source data should also be updated periodically to reflect the changing composition of businesses' intermediate consumption in various industries.

There was a strong demand for up-to-date data on purchasing patterns between industries as this helps users understand inter-dependencies between, and characteristics of, UK businesses. This is vital for understanding how the economy is likely to respond to economic policy and economic shocks.

Additional uses include detailed coverage of water and energy as well as the proportion of total purchases as imports.

There are a wide range of users for APS data. Users include those from government, both internal with the ONS and external in other government departments such as:

- the Department for Business, Energy and Industrial Strategy (BEIS)
- the Department for International Trade (DIT)
- the Department for Digital, Culture, Media and Sport (DCMS)
- the Department for the Environment, Food and Rural Affairs (Defra)

Devolved administrations such as the Scottish and Welsh Governments, as well as local authorities, also constitute main users of the Annual Purchases Survey outputs.

Annual Purchases Survey Stakeholder Working Group meetings are held regularly. This provides an opportunity for any changes or developments to the survey to be discussed directly with its government users, so that, where possible, their requirements can be met.

1.3 History

The original Purchases Inquiry (PI) ran from the 1950s to 2006. Data were collected at five-yearly intervals, then annually from 1999 onwards. The sample size and industry coverage were expanded between 1999 and 2001, from 3,000 businesses (covering partial production, distribution and sales industries) to 28,000 (covering full production, construction, distribution and services industries). In its fullest form, the PI contained around 500 variations of questionnaires and 1,400 questions, with an average of 40 questions going to each respondent.

In 2006 (reference year 2005), the PI sample size was cut by 50% to reduce costs and burden on businesses. Consequently, the results were considered to be of insufficient quality and the survey was suspended. The last dataset dates from 2005 and covers reference year 2004. In the absence of a PI from 2006 onwards, patterns from the 2004 reference year have been used to apportion total industry data from the [Annual Business Survey \(ABS\)](#).

The industry totals used in the supply and use tables are provided by the ABS. However, product-level estimates in the supply and use tables are apportioned using the patterns of the 2004 PI data. The 2004 PI data will now be outdated due to changes in businesses' purchases patterns, for example, due to significant technological advances. The reintroduction of the APS aims to collect detailed, up-to-date, product-level information to replace the outdated 2004 proportions.

1.4 Timeline

The APS process from sample selection through to the publication of the final estimates is shown in Table 1.

Please note that for the 2015 period there were two questionnaire dispatches; tranche 1 in month three and tranche 2 in month six. From the 2016 period onwards, all questionnaires were dispatched at the same time during the month three of the following year.

This timetable is currently not adhered to as the Annual Purchases Survey is in the developmental stage, but this is the ideal timetable which is hoped to be reached in 2019 for 2018 reference period.

Table 1: Summary of the survey process

Month						
1	Questionnaire preparation					
2		Sample selection				
3			Questionnaire dispatch			
4				Receipt of survey questionnaires and business level		
5						
6						
7						
8						
9						
10				Quality assurance of aggregate level estimates		
11						
12						Delivery of 'provisional' estimates to National Accounts
13						
14						
15						Delivery of 'final' estimates to National Accounts

Download this table

[.xls](#)

2 . Questionnaire design

2.1 Overview

The Annual Purchases Survey (APS) currently has 109 questionnaire types (corresponding to the 109 industries), with approximately 53 questions on each. The survey questions were produced for the APS using the Statistical Classification of Products by Activity (CPA) version 2.1.

The CPA is the classification of products (goods as well as services) at the level of the European Union (EU). Product classifications are designed to categorise products that have common characteristics. They provide the basis for collecting and calculating statistics on the production, distributive trade, consumption, international trade and transport of such products.

A set of 35 core products are considered crucial to all industries. The remaining CPA product questions are subsets relating to specific industries. However, not wanting to restrict businesses to an assumed list of products, a list of the remaining products not specified on the main body of the questionnaire was included. This could be found at the back of the questionnaire for the reference periods 2015, 2016 and 2017.

From the 2018 reference period onwards, these remaining products can be found at the end of each section. This allowed businesses the potential for selection of additional products where appropriate. Analysis of these additional products may inform changes to the questionnaire form tailoring and industry product assumptions in the future.

2.2 Questionnaire types

The APS questionnaire has two versions:

- manufacturing
- non-manufacturing

The difference between the two versions is that on the manufacturing questionnaire, the following yes or no question is asked:

"Did your business carry out work on behalf of a customer when you provided only labour and no materials?"

There are 109 questionnaires in total: 44 manufacturing and 65 non-manufacturing.

2.3 Questionnaire development

The development of the APS was essential in meeting the regulations set out in the Eurostat Manual of Supply, Use and Input-Output Tables. Requirements from stakeholders such as national accounts, the Department for Business, Energy and Industrial Strategy (BEIS), and devolved administrations were also gathered and prioritised.

All requests for inclusions and exclusions or changes to the questionnaires had to be agreed with the APS Project Board. Any such requests were assessed and costed before being agreed, including the relevant compliance costs (the costs incurred by businesses through responding to the survey). Once provisional agreement to any change had been obtained, the required changes were tested to ensure that responding businesses understood the proposed wordings and were able to supply the relevant information.

2.4 Questionnaire review

The APS is a relatively complex survey and balance needs to be struck between questions that businesses can respond accurately to and the detail required to underpin national accounts estimates. Therefore, the APS questionnaire will be reviewed alongside response rates, validation, sample design and the returned data in readiness for future publication and use in the national accounts supply use tables.

A high-level review of the APS questionnaire was conducted in 2018, in preparation for reference year 2018 (sent out in 2019). Working closely with the Behavioural Insight Unit within the Office for National Statistics, significant changes were made to the questionnaire format. Some of the changes to note are the use of a checklist to ensure all stages of the questionnaire are completed, additional guidance on how to provide a best estimate, moving the list of additional products from the end of the questionnaire to the end of each section, and using “please” and “thank you”.

A review of the changes made to the 2018 questionnaire form will also be conducted in 2019.

2.5 Variables collected

The main variables that are collected on the survey are described in Table 2.

Table 2: Description of main survey variables

Section Variable	Description
A Reporting period	Reporting period start and end dates
Total purchases	2015 to 2017: Total expenditure on energy, services, goods and materials of which: <ul style="list-style-type: none">• total expenditure on energy, services, goods and materials purchases for resale without further processing?*
B	<ul style="list-style-type: none">• total expenditure on energy, services, goods and materials used up or transformed by your business?* 2018 onwards: Total purchases of energy, services, goods and materials used up or transformed
C Purchases of energy, water and waste	• Expenditure by the business on specific products, based on its industry assumptions
D Purchases of services	• Space for additional expenditure on specific products, not relating to industry assumptions
E Purchases of goods, materials and related services	• Total expenditure used up or transformed by the business • Expenditure from suppliers located in the UK or outside of the UK
F Summary of purchases	This has a total of sections C, D and E and should equal the intermediate consumption value in Section B
G Completion time and contact	Comments*, contact name, time taken to complete; hours and minutes*, cost of completion*, telephone number, fax number, email

Source: Annual Purchases Survey, Office for National Statistics

Notes

1. The asterisk (*) signifies that these questions were removed for the 2018 reference period and onwards.
[Back to table](#)

3 . Sampling procedure

3.1 Sampling frame

The Inter-Departmental Business Register

A sampling frame is a complete list of all the members of a population being studied, from which the sample is drawn. The sampling frame for the Annual Purchases Survey (APS) is the list of UK businesses on the Inter-Departmental Business Register (IDBR).

Businesses are added to the IDBR if they are:

- registered for Value Added Tax (VAT) and Pay As You Earn (PAYE) with Her Majesty's Revenue and Customs (HMRC)
- an incorporated business registered at Companies House

The IDBR covers businesses in all parts of the economy, except:

- some very small businesses
- the self-employed
- businesses without employees
- businesses with low turnover
- and some non-profit making organisations

There are 2.6 million businesses on the IDBR; covering nearly 99% of UK economic activity. It is used by government departments, including the Office for National Statistics (ONS), as the sampling frame for most business surveys.

To keep the information on the IDBR up-to-date, administrative data from HMRC are supplemented by data from ONS surveys, such as the Business Register and Employment Survey (BRES).

Further information about the IDBR can be found on the [ONS IDBR webpages](#).

Reporting units

The business unit to which questionnaires are sent is called the reporting unit. The response from the reporting unit can cover the enterprise as a whole, or parts of the enterprise identified by lists of sub-units (called local units). Other than for a minority of larger businesses or businesses that have a more complex structure, the reporting unit is the same as the enterprise. For this reason, the APS reporting unit counts are presented as enterprise counts.

An enterprise may consist of one or more local units, for example, the head office for a group of shops. An enterprise may therefore have local units at different locations and may carry out more than one type of economic activity.

Currently, the APS does not produce regional estimates.

Standard Industrial Classification (SIC)

Each enterprise is classified according to the [Standard Industrial Classification of Economic Activities \(SIC\)](#) system. The UK is required by European legislation to have a system of classification consistent with the European Union's industrial classification system. The system underwent a major review in 2007. APS data have been collected and published on the SIC 2007 system.

UK SIC 2007 is divided into 21 sections, each denoted by a single letter from A to U. The letters of the sections can be uniquely defined by the breakdown to the divisions (denoted by two digits), which are then broken down into groups (three digits), then into classes (four digits) and, in some but not all cases, again into subclasses (five digits).

For example, in SIC 2007:

- section C manufacturing (comprising divisions 10 to 33)
- division 13 manufacture of textiles
- group 13.9 manufacture of other textiles
- class 13.93 manufacture of carpets and rugs
- subclass 13.93/1 manufacture of woven or tufted carpets and rugs

The full structure of SIC 2007 consists of 21 sections, 88 divisions, 272 groups, 615 classes and 191 subclasses.

Each local unit is assigned a single SIC code, which corresponds to the unit's principal activity. Where more than one type of economic activity is carried out by a local unit or enterprise, its principal activity is the activity in which most of the people are employed, though this does not necessarily account for 50% or more of the total employment of the reference unit.

For example, if a reference unit contains four local units (for example, four local shops), and each specialises in different things then they would be under different SICs. If one local unit contains more employees than the others, then the local unit with the most employees is where the reference unit would be classified to. The proportion of employees in each local unit could be 30%, 25%, 25% and 20% of all employees, so the reference unit would be placed under the SIC that holds 30% of the employees. There are detailed rules for determining SIC for multiple-activity economic units, including situations where measures of value added are not available.

Re-classification of a business can occur due to a relatively small change to the nature of its operation, and this can have a significant effect on APS estimates by industry. In addition, the correction of misclassification of businesses can lead to bias, particularly where there is systematic movement from one industry to another. This is because, where classification updates are identified via survey returns, it is only units in the survey sample that are updated.

All surveys that do not cover the whole business population, such as the APS, have the potential for some underestimation of output variables due to the re-classification of units moving out of the APS survey population but never into it. However, such underestimation is likely to be small. In the APS, this effect is corrected for by adjusting the weights of the businesses that remain in the sample.

The exact inclusions and exclusions of industries in the APS are detailed in this section.

Inclusions

- Agriculture, forestry and fishing (section A) (Standard Industrial Classification 01.6 to 01.7)
- Mining and quarrying (section B)
- Manufacturing (section C)
- Electricity, gas, steam and air conditioning supply (section D)
- Water supply; sewerage, waste management and remediation activities (section E)
- Construction (section F)
- Wholesale and retail trade; repair of motor vehicles and motor cycles (section G)
- Transport and storage (section H)
- Accommodation and food service activities (section I)
- Information and communication (section J)
- Financial and insurance activities (section K)
- Real estate activities (section L)
- Professional, scientific and technical activities (section M)
- Administrative and support service activities (section N)
- Education (section P)
- Human health and social work activities (section Q)
- Arts, entertainment and recreation (section R)
- Other service activities (section S)

Exclusions

- Agriculture, forestry and fishing (section A) (Standard Industrial Classification 01.1 to 01.5)
- Public administration and defence; compulsory social security (section O) (note that the Annual Business Survey (ABS) covers all legal statuses, except the public sector, while the APS covers only legal statuses 1 to 3: company, sole proprietor and partnership)
- Activities of households as employers; undifferentiated goods – and services – producing activities of households for own use (section T)
- Activities of extraterritorial organisations and bodies (section U)

3.2 Sample design

Data are collected by the ONS from approximately 31,000 businesses in the UK (excluding the Channel Islands and the Isle of Man). The area covered is in line with other business surveys produced by the ONS. Sample selection is carried out using a stratified random sample design. Groups of reporting units (strata) are defined by three variables:

- employment size band
- SIC
- geographical region

Sample selection occurs independently for each stratum. When the sample is designed, the size of the sample in each stratum is determined by an algorithm, which distributes the sample amongst the cells to give the lowest estimated variance (uncertainty). This design is significantly more efficient (that is, it gives a much more accurate estimate for the same sized sample) than a simple, unstratified random sample.

The variables defining the strata are:

- six employment size bands: 0 to 9, 10 to 19, 20 to 49, 50 to 99, 100 to 249, and 250 and over
- industry class: four-digit [UK Standard Industrial Classification 2007: SIC 2007](#)
- region: England, Scotland, Wales and Northern Ireland

Please note, the APS does not publish to this level of disaggregation.

The sample is designed so that a sample for a stratum will generally be selected for two years and the units in that sample will largely not be reselected for at least two years after that selection. The random sample selection uses permanent random numbers (PRN), a unique nine-digit identifier that is randomly assigned to each unit when it is added to the IDBR. The sample for each stratum is constructed using consecutive PRNs from within that stratum until the sample size required has been reached.

For the APS, each sample is generally selected for two years, and there is a year-to-year overlap of half the sample. That is, in any year, half of the sample will be newly selected, and half will have been selected in the previous year as well. This is illustrated in Table 3, for a sample of four units taken from a stratum containing 10 units (note that these are not real PRNs). This design means that, for half the sample, returns are available from the same businesses in consecutive years, and this helps to maintain the quality of editing and validation, imputation and outlier detection (see Section 5 for more information).

Table 3: Example of permanent random numbers sampling method

Year 1	Year 2	Year 3
843*	843	843
1,390 *	1,390	1,390
2,639 *	2,639 *	2,639
2,718 *	2,718 *	2,718
2,817	2,817 *	2,817 *
3,445	3,445 *	3,445 *
4,824	4,824	4,824 *
5,685	5,685	5,685 *
5,824	5,824	5,824
7,208	7,208	7,208

Source: Office for National Statistics

Notes:

1. The asterisk (*) signifies that these units were selected in this year.

Download this table

[.xlsx](#)

Table 3 shows that in the first year, units 843, 1,390, 2,639 and 2,718 were selected. In the second year, the first two units were dropped (843 and 1,390) and the last two units were retained (2,639 and 2,718). Additionally, two new units were added into the sample (2,817 and 3,445). This process was then repeated for the third year. When the last PRN in the stratum is reached, selection rolls around to the smallest again.

However, there are a few exceptions to this design. If a selected unit then moves to another stratum, for example, by changing SIC classification (see Section 3.1), then it may be selected for a second two-yearly period. Also, if there are fewer units within a stratum, the likelihood of consecutive selection will increase. For these reasons, there is never a guarantee that a business will only be selected for two years.

A further exception arises for the strata within the largest and smallest employment size bands.

For the largest size bands, containing businesses with employment of 250 or more, all the enterprises are selected every year. This is because these strata tend to have fewer enterprises in them, and yet, as they are large enterprises, they are dominant respondents to estimated total values. Including all the largest enterprises significantly reduces uncertainty on the estimated total values.

For most businesses with employment of 0 to 9, Osmotherly rules apply. These rules state that when a business with 0 to 9 employment has been selected in a survey, it will only be selected for a single year and it will not be reselected for at least three years following selection (provided they complete and return the questionnaire). There are a few exceptions to these rules, but in general, they are implemented to reduce the burden on small businesses, which may not have much resource for completing survey questionnaires.

4 . Data collection

4.1 Timetable of questionnaire dispatch

Questionnaires are sent to collect information from businesses relating to the previous 12 months, which is known as the reference year. The questionnaires are required to be returned to the Office for National Statistics (ONS), in a pre-paid envelope, within six weeks.

Approximately 31,000 businesses are selected to receive the Annual Purchases Survey questionnaire. In the 2015 reference year, the sample was split into two tranches for dispatch (around 50% in each). The first dispatch was sent out in February 2016 (Tranche 1) and the second was sent in May 2016 (Tranche 2). From the 2016 reference year onwards, questionnaires were sent in one dispatch (in the February after the reference year).

4.2 Welsh questionnaires

The ONS Welsh Standards give an option for Welsh business respondents to request a Welsh language version of the questionnaire. This option is clearly shown and written in Welsh on the front page of the Annual Purchases Survey questionnaire. For example, there were eight Welsh forms requested in 2017.

4.3 Euro respondents

Respondents who prefer to provide their purchases values in the Euro currency are provided with a Euro questionnaire upon request. For example, there were 19 Euro forms requested in 2017.

These responses are converted to pounds sterling (£) using the [universal currency converter](#).

4.4 Expected questionnaire receipt

To meet the minimum accuracy standards required by users, the Annual Purchases Survey (APS) response rate target is 75% by the November following the reference year. For large businesses (those with employment of 250 or more), the minimum response rate required is 90%.

The ONS also has specific targets for the most economically important businesses that are selected to complete the APS. These economically important businesses are referred to as “key respondents” and are those businesses that are important to either an industry or a specific product. The targeted response rate for “key respondents” is set at a minimum of 95% but the ONS strives to achieve 98%.

Table 4 shows the response rate obtained for the 2015 reference period. If you are interested in the response rates for another period, please see the relevant statistical publication.

Table 4: Percentage of businesses selected in each size band and the response rates obtained for the 2015 reference period

Sizeband Selected Returned Cleared Returned (%)

Sizeband	Selected	Returned	Cleared	Returned (%)
0-9	13,712	10,860	8,723	79.2
10-19	2,743	2,250	2,005	82
20-49	3,395	2,732	2,530	80.5
50-99	2,066	1,666	1,523	80.6
100-249	1,818	1,472	1,352	81
250+	7,379	6,640	6,360	90
Total	31,113	25,618	22,493	82.3

Source: Annual Purchases Survey, Office for National Statistics

4.5 Reminder letters

Up to two reminder letters are sent to businesses who have not returned a completed questionnaire by the end of April.

All non-responders with employment of 1,000 or more are sent a Chief Executive Letter (CEL), and a duplicate questionnaire, rather than a second reminder as their impact on provisional estimates is the greatest. The CEL is a stronger reminder to inform the chief executive or managing director that the business has not responded and is a reminder of the legal requirement to respond. The CEL further outlines the non-compliance penalties prior to any enforcement procedures.

4.6 Response chasing

During the data collection period, the APS response rates for returned questionnaires are monitored regularly. A manual exercise is undertaken during the data collection cycle to identify industries with low response rates.

Telephone response chasing starts after the second reminders have been dispatched (start of June) and continue, if necessary, up to the final result run (November following the reference year). It is intended to encourage the completion of the questionnaire and address any respondent issues in a timely and efficient manner, which all leads to the production of a quality output.

4.7 Enforcement strategy

The APS carries out enforcement action under the [Statistics of Trade Act 1947](#). Enforcement action is used to maintain response rates, and hence the quality of the survey. It is used only as a last resort, after attempts to encourage businesses to complete the survey following telephone calls and reminder letters.

If enforcement action is carried out, the business will be issued with a summons to court. If this happens, the business can still choose to respond to the survey, and the case will be withdrawn. This option is only allowed once. If the business becomes subject to enforcement a second time the business will be prosecuted. Businesses can be fined up to a maximum of £2,500 for non-response.

5 . Converting respondent data into published estimates

5.1 Editing and validation

Questionnaires are sent to businesses by post, along with detailed instructions on how to complete and return them. When responses are received, they are entered into the processing system electronically. The questionnaires go through different phases of cleaning and processing to improve data quality:

Step 1: Questionnaires are electronically scanned into the data store.

Step 2: Data are then transferred to the processing system. Initial structural validation checks are carried out on the returned data.

Data will fail these structural validation checks if (for example):

- the reference already exists for that reference period (duplicate records)
- the questionnaire is not complete (not all pages of the questionnaire have been returned)
- more than one copy of the same questionnaire page had been returned
- an unknown product reference is returned (the product reference returned is not present in the survey selection file)

Step 3: Once the structural validation checks have been completed and any failures have been dealt with, returned data are passed through a further set of validation checks.

Data will fail these validation checks if (for example):

- components of the questionnaire do not add to the reported totals
- the returned dates fall outside of the determined thresholds
- the returned data contain negative values
- data are not returned where a question is compulsory
- for the 2016 reference period onwards, if a large change year-on-year is identified (for example, if a business' total intermediate consumption figure increases or decreases by 20% from one year to the next)

Step 4: Following these automatic editing and validation check, manual quality assurance of the data is conducted. As a final quality assurance of the data, high-level aggregates are sent to experts in other government departments as part of peer appraisal (for example, the Department for Business, Energy and Industrial Strategy (BEIS)).

5.2 Imputation

Imputation techniques are used to estimate the value of missing data due to non-response, whether partial (item non-response) or full (unit non-response). Item non-response occurs when a business returns a value for its section total (for example, total amount spent on energy, water and waste) but is unable to break the total down to a product level. Unit non-response occurs when a business does not respond to the questionnaire at all.

Imputation is designed to give better results than deletion, in which all subjects with any missing values are omitted from the analysis. This is important because a suitable level of accuracy needs to be achieved (see Section 4.4).

A review into the imputation techniques was conducted to strengthen the way the APS estimates for missing values. Imputation methods usually require the previous year's data to assign a methodologically sound figure for making these estimates. As the APS was new for the reference year 2015, there were no previous results to calculate these imputations from. Therefore, different imputation methods were used for reference year 2015 and reference years 2016 and onwards, because previous data were now available from this reference period.

Approach for 2015 reference period

Unit non-response

For non-responding large businesses (those with employment of 250 or more), the figures were constructed using the relevant business' Annual Business Survey (ABS) data. Data collected for the ABS cover business transactions, including a comparable figure of intermediate consumption, which is a good estimator of the overall figure collected as part of the APS.

This total intermediate consumption figure from the ABS was then split out into the various components needed by the APS by imputing based on similar businesses. This is calculated using information from businesses who responded to the APS who fall within the same industry group as the non-responding business. The "average" proportion across these businesses for each component is then applied to the non-responding business' data.

For non-responding small businesses, imputation was not carried out, and totals were estimated using weighting (see Section 5.3 for a discussion of weighting).

Item non-response

For businesses that could provide section totals but could not break them down into product-level detail, product breakdowns were used from a randomly selected similar business.

Approach for 2016 reference period onwards

Unit non-response

For large businesses (250 employment or more) that have responded in the previous reference period, unit non-response can be imputed for. This uses the business' data from the previous reference period, and an average growth, to impute the missing data in the current reference period for the business. This method, called the "ratio of means" technique, is explained in this section.

The ratio of means technique assigns each business with an imputation class (a group of similar businesses), which are based on the two-digit standard industrial classification and employment brackets of 0 to 99 and 100 and over. Businesses in the imputation class that have responded in both the previous and current period will be used to calculate an "imputation link".

The imputation link is calculated as follows:

$$R_{class,t} = \frac{\sum_{class} y_{i,t}}{\sum_{class} y_{i,t-1}}$$

where:

- $class$ is the set of all businesses in the imputation class that have a response at both time t and $t-1$ (the previous and current year)
- $y_{i,t}$ is the section total for business i in the current period
- $y_{i,t-1}$ is the section total for business i in the previous period

The section total for business j is imputed as follows:

$$\hat{y}_{j,t} = R_{class,t} \times y_{j,t-1}$$

The component values in the section for business j are imputed as follows:

$$\hat{y}_{j,k,t} = R_{class,t} \times y_{j,k,t-1}$$

where:

$y_{j,k,t-1}$ is the value of purchases for product k in the previous year.

It is important to note:

- only returned values are used in the calculation of $R_{class,t}$ – imputed values are excluded
- the link is applied to returned values for the previous year, or imputed values that were imputed using this method in the previous year

For non-responding small businesses, the same weighting method as the 2015 reference period was used.

Item non-response

When a business has provided us with a high-level total for each of energy, goods and services, but cannot't break it down in to detailed products, then the breakdown will be automatically imputed based on other similar businesses.

For a product k , and a given imputation class, the proportion of a section total that it will be assigned is calculated as follows:

$$p_{class,k,t} = \frac{\sum_{class} y_{i,k,t}}{\sum_{class} y_{i,t}}$$

where:

- $class$ is the set of businesses that have provided a breakdown for the section at time t
- $y_{i,k,t}$ is the value of purchases for product k for business i at time t
- $y_{i,t}$ is the section total for business i at time t

The imputed value of purchases for business j for product k is calculated as:

$$\hat{y}_{j,k,t} = p_{class,k,t} \times y_{j,t}$$

5.3 Estimation of totals

It is not possible to collect data on every UK business every year (a census), because:

- the burden on businesses would be too great
- the cost of running such a census would be prohibitive
- a well-designed sample survey can produce better estimates than a census with a poor response rate

Therefore, the APS collects information from a sample of the UK business population each year. The sample design is described in Section 3.2. This section describes how returns from the sample are used to estimate totals for the whole population.

Weighting

To calculate the estimates for an entire population from data collected from a sample, the APS uses standard statistical weighting methods. Specifically, the results received from the sample are multiplied by three weights:

- the design weight, known in ONS as the a-weight
- the calibration weight, known in ONS as the g-weight
- the o-weight or outlier weight

Design weight

Known in ONS as the a-weight, this accounts for the sample design so that a business' probability of selection is properly reflected. So, for example, a business with a small probability of being selected for the survey will have a large design weight.

Calibration weight

Known in ONS as the g-weight, this makes a correction for an unrepresentative selected sample. For example, in a random selection of five businesses out of a population of 10, it is possible that the five businesses selected are, by chance, towards the upper boundary of the employment size band, as opposed to an even distribution of businesses across the size band. If no correction is made, the population total would be over-estimated.

Auxiliary information, that is information not collected by the survey, which acts as a proxy for the variable of interest, is used to correct for this effect. The weight is the ratio of the actual population total for the auxiliary variable to the population total estimated from the sample's auxiliary variables is calculated. For the APS, the auxiliary variable is the employment value found on the IDBR.

Outlier weight

The o-weight, or outlier weight, identifies potential outliers in the sample, reducing them in line with comparable businesses from the same cell. For example, businesses reporting intermediate consumption much higher than similar businesses of the same size and within the same industry, would be reduced in line with these to reduce volatility in the overall population estimate. These are weighted at product level.

The weighted value is then calculated as:

$$\text{weighted value} = \text{returned value of the variable} \times a - \text{weight} \times g - \text{weight} \times o - \text{weight}$$

Estimates of population totals are then found by simply summing the weighted values across the whole sample.

Calculating the a-weights

The a-weight is calculated for each stratum in the sample, which is a group of businesses defined by their size and industry (see Section 3.2). In its simplest form, the a-weight, a , for each stratum is calculated as:

$$a = \frac{N}{n}$$

where:

- N is the total number of businesses in the cell (the population)
- n is the number of businesses in the returned sample

For example, to estimate the weight of a pile of fifty 50 bricks, 10 bricks could be weighed. N , the total number of bricks is 50; n , the sample size, is 10; and a is therefore 50 divided by 10 equals 5.

Calculating the g-weights

G-weights are calculated for groups of strata within the same industry, but across several size bands. Generally, regions (England, Scotland, Northern Ireland and Wales) are grouped for the calculation of g-weights. These groups are called g-weight bands.

In its simplest form, the g-weight is the ratio between the total of the auxiliary variable estimated from the sample and the actual population total for the auxiliary variable. The g-weight will therefore be greater than one when the total auxiliary estimated from the sample is less than the total auxiliary in the population, and less than one when the total auxiliary estimated from the sample is more than the total auxiliary in the population. If response is representative, all the g-weights should be close to one. The g-weight therefore helps correct for any imbalances in the selected sample that arise through random chance or non-response. This is calculated as follows:

$$g = \frac{T_{pop}}{(T_{samp} \times N/n)}$$

where:

- T_{pop} is the sum of IDBR employment over all businesses in the population
- T_{samp} is the sum of IDBR employment over all businesses in the returned sample
- N is number of businesses in the population
- n is number of businesses in the returned sample
- $T_{samp} \times N/n$ is the total for the auxiliary estimated from the sample

Calculating the o-weights

The o-weights are calculated for each product separately, to determine if it represents an outlier or not. The o-weights are calculated using L-values, which set the parameters for the returned data, setting an “upper” and “lower” limit of acceptable values, excluding extreme values through winsorisation.

The o-weights are calculated using the L-values to determine an individual weight for each product using the following formula.

For positive returns ($y_{RU} > 0$), the outlier weight for each reference unit for each question (y) should be:

$$owt_{RU,y} = \min \left[1, \left(\frac{1}{a_{RU} g_{RU,y}} \right) + \left(\frac{1}{y_{RU}} \right) \left(\beta_c x_{RU} + \frac{L_{pos}}{a_{RU} g_{RU,y} - 1} \right) \left(1 - \frac{1}{a_{RU} q_{RU,y}} \right) \right]$$

For zero returns ($y_{RU} = 0$), the outlier weight should be calculated as:

$$owt_{RU,y} = 1$$

where:

- owt is the o-weight
- y is the question number

$owt_{RU,y}$ is the o-weight for a given RU for the returned figure for a specific question (y).

Weighted values will be calculated as the product of the value for each question and the weight for that record or variable.

6 . Publication and data dissemination

6.1 Publication of the Annual Purchases Survey results

Estimates from the Annual Purchases Survey (APS) are published on the [Office for National Statistics' website](#). The first period of collecting data was for the 2015 survey period; these were published as proportions, constrained to the Annual Business Survey (ABS) totals.

Following from the second year of the APS (2016 reference period), the estimates were published as proportions and values constrained to the ABS. These were shown as datasets and were supported by a statistical bulletin. APS data are published constrained to the ABS totals because the ABS is an established survey, which has a much larger sample than the APS (73,000 businesses compared with 31,000 businesses respectively). This therefore provides a more robust estimate at the total intermediate consumption aggregate level and ensures comparability and consistency.

The proposed publication cycle from the 2018 reference period onwards is:

- provisional results are published in the November following the reference period
- final results are published in the April after the provisional results are published

See Section 1.4 for a full timeline of the survey.

The APS publishes industry and product information from approximately 31,000 businesses covering 110 grouped products. Details on these can be found in Section 3.1.

High-level variables included within the publication are the total intermediate consumption, as well as this split into three sections:

- energy, water and waste
- services
- goods, materials and related services

6.2 Dissemination of the Annual Purchases Survey results

Estimates from the APS will be provided to the National Accounts Supply and Use team within the Office for National Statistics. These figures will be included as part of the calculations for the annual supply and use tables. Estimates provided will be consistent with those included within any publication.

7 . Disclosure control and data confidentiality

7.1 Confidentiality protection requirement by law and Government Statistical Service (GSS) policy

The need to keep records of individuals, businesses or events used to produce official statistics confidential is enshrined in law. However, this does not prevent the release of anonymised or aggregated data.

The [Code of Practice for Statistics](#) and the National Statistician's guidance: [Confidentiality of Official Statistics](#) provides the Government Statistical Service (GSS) policy framework for official statistics in this regard. The Code of Practice guarantees confidentiality to those who provide private information for the production of official statistics.

Furthermore, the ONS surveys are conducted on behalf of the UK Statistics Authority, and all outputs are subject to Section 39 of the [Statistics and Registration Service Act 2007 \(SRSA\)](#).

Business surveys operating within the UK are governed under the [Statistics of Trade Act 1947](#). This makes participation in the surveys compulsory, and confidentiality requirements that relate to published data are specific in Section 9 of the act. It also states that tables should not be published that would disclose any information relating to an individual business, unless there is expressed consent in writing from that business. In addition, data should not be published that would reveal the exact number of respondents contributing to a cell if there are fewer than n respondents, as detailed in Section 7.4.

The Office for National Statistics is now fully compliant with the GDPR regulation.

7.2 The ONS confidentiality pledge

The confidentiality pledge is an assurance of confidentiality given to survey respondents and states:

"All the information you provide is kept strictly confidential. It is illegal for us to reveal your data or identify your business to unauthorised persons."

7.3 Statistical disclosure control and the ONS

The [Statistical Disclosure Control Policy](#) sets out the standards for safeguarding the information provided in confidence to the ONS. "Statistical disclosure control" refers to the methods that reduce the risk that confidential information is published in any official statistics. These methods are applied if ethical, practical or legal considerations require the data to be protected.

Statistical disclosure control involves modifying data so that the risk of identifying individuals is reduced, but at the same time attempts to find a balance between improving confidentiality protection and maintaining an acceptable level of quality in the published data. Statistical disclosure control would be applied to the Annual Purchases Survey (APS) data before publication.

7.4 Identifying disclosive data for the Annual Purchases Survey

The design of the APS means that totals can be estimated for each industry and product. However, these totals are proposed to be aggregated for publication purposes, for example, to all businesses in an industry, or to higher-level industry groups. Combining totals like this improves the statistical quality of the estimates and reduces the risk of disclosure. It is at the aggregated level that the statistical disclosure control will be carried out. The first step is to identify whether data could be disclosive, that is whether there is a risk that information about an individual business could be identified.

In the discussion in this section, a "cell" refers to an element of a published table, containing the aggregated data (as described previously), not to the sampling strata described in Section 3.2. For tables of total values published by the APS, there are two criteria that must be met for the published value to be deemed non-disclosive. These criteria are:

- minimum threshold rule: this rule states that there must be at least n reporting units (businesses) in a cell
- $p\%$ rule: this rules states that the total contribution of the m largest respondents in the cell aggregated total must be less than $p\%$ of the total in that cell

The values of n , m and p should remain confidential. Knowing these values could allow information on individual businesses to be calculated.

In this example, there are 10 businesses in a cell, of which four have returned their total turnover estimates, and $n = 3$, $m = 1$, and $p = 95\%$.

Table 5: Example of disclosure control

Business	A	B	C	D	E	F	G	H	I	J
Total Purchases (£'000s)	20	30	5				1,500			

Source: Annual Purchases Survey, Office for National Statistics

The following two criteria are applied to the data:

- threshold rule: there are four businesses that have reported values – the minimum threshold, n , is three, so the cell is not disclosive under this rule
- $p\%$ rule: total returned turnover of the cell = £(20+30+5+1,500) thousand = £1,555 thousand. Since $m = 1$, and the largest respondent is business G, with a total turnover of £1,500 thousand, the percentage contribution of business G to the total turnover in the cell is calculated as follows:

$$\left(\frac{1,500}{1,555} \right) \times 100 = 96.5\%$$

In this case, 96.5% is greater than 95%, so under this rule, the cell is disclosive.

As the cell has not met both criteria, it is identified as a disclosive cell, and disclosure control methods must be applied before the data can be published.

7.5 Disclosure control methods

Standard techniques for controlling statistical disclosure are used for the release of the APS results. These are described in this section.

Cell suppression

Cell suppression is the standard method used to protect tables with disclosive cells. The disclosive cells are suppressed, that is, they are not published. This is known as primary suppression. Other, non-disclosive cells must sometimes also be suppressed, to prevent the values of the primary suppressed cells from being calculated by subtraction of all the other cells from the total. These are known as secondary suppressions. This is the method used by the APS (and other surveys) to suppress disclosive values.

Merging of cells and cell aggregation

Cells may also be combined to prevent publication of disclosive data. For example, where there are very few industries in a specific sector a higher industry classification will be used instead. This has not been used for the 2015 and 2016 reference years.

Rounding

Monetary estimates in any standard release will be rounded to the nearest £ thousand, £ million or £ billion. Percentages or rates displayed in any release will be derived from the unrounded values and then the percentages rounded to one decimal place.

8 . Revisions policy

The Office for National Statistics (ONS) ensures that published estimates are as accurate as possible. However, if significant changes are made to source data after publication, then estimates will be revised. The ONS has a clear [policy on how revisions are handled](#) across the organisation and the specific procedure for the Annual Purchases Survey (APS) is outlined in this section.

8.1 Planned revisions

Planned revisions usually arise from either the receipt of additional data from late responding businesses or the correction of errors to existing data by businesses responding. Revisions to future published APS data will be dependent on the defined publication schedule. Based on the understanding that APS estimates will be published annually, it is anticipated that alongside figures for the current year, revised estimates will be published for the previous year's estimates.

8.2 Unplanned revisions

In addition to planned revisions to the current and previous survey years, additional unplanned revisions may be published if they are considered to be large enough and of sufficient interest to users such that a delay until the next standard release is not justifiable, or if they affect data in more than just the current and previous survey years. The timing with which these revisions are released will take into account:

- the need to make the information available to users as soon as is practicable
- the need to avoid two or more revisions (to the same data items) in quick succession, where this might cause confusion to users

All unplanned revisions will be released in compliance with the same principles as other new information.

9 . Future work

9.1 Questionnaire review

Another review of the Annual Purchases Survey (APS) questionnaire is planned. This will be a detailed review of the questions themselves – specifically, their wording and ordering. Additionally, a review of which questions are included for each industry group's questionnaire form will be conducted – reviewing whether the businesses consistently answer the same questions and if questions need to be added or reviewed.

In the future, the Office for National Statistics (ONS) will introduce online data collection, which will further aid question filtering and the potential for business-specific bespoke questionnaires.

9.2 Methods

Further review of the imputation methods may be required in the future. A review of the APS sample is being conducted to ensure we are sampling businesses in the most efficient and effective way.

9.3 Regional estimates

Currently the APS does not produce regional estimates. However, it is hoped in the future it will. If, and when, the APS produces regional estimates, they would be produced in the following way.

The geography assigned to the enterprise is based on a postcode that is generally the registered office for the business. If this information is used to produce regional estimates it could lead to bias, as the enterprise address given is generally the head office, and head offices can be over-represented in big cities such as London and Edinburgh. Therefore, estimates will be apportioned to regions based on local unit information held on the Inter-Departmental Business Register (IDBR) when producing regional estimates in the future.

10 . Contact and other information

A full analysis of the [Development of the Annual Purchases Survey](#) was published in December 2017.

[Quality and Methodology Information for the Annual Purchases Survey](#) was published in February 2019.

For further information, please contact the Annual Purchases Survey Team by email at purchases.survey@ons.gov.uk.