1. Introduction

Internet sales statistics measure the value of Internet Sales of British retailers on a monthly basis. They are released as part of the Retail Sales Index, a key economic indicator used to estimate consumer spending on retail goods and the outputs of the retail sector. Further information on the RSI can be found in the quick guide to retail sales¹.

Internet sales statistics provide an insight into the performance of the Internet arm of the retail sector. These statistics provide an estimate of the value of Internet sales in seasonally and non-seasonally adjusted form.

2. Defining Internet Sales

Internet sales statistics are collected through the Monthly Business Survey – Retail Sales Index, as such the industries included in the suite of Internet Sales Statistics are the same and are defined by the UK Standard Industrial Classification (SIC) 2007². Therefore Internet sales statistics published in the Retail Sales Statistical release are presented for all retailing excluding automotive fuel and for three main retail sector groupings:

- Predominantly food
- Non-food
- Non-store retailing.

The Automotive fuel sector is not included as there is little to no Internet sales activity within this industry.

The non-food sector is further broken down to provide statistics on:

- non-specialised stores or department stores
- textiles, clothing and footwear
- household goods stores
- other specialised stores.

Table 1 provides a full description of each of these categories, the categories highlighted have Internet sales statistics published in the retail sales tables³. Lower level data is unavailable at present but ONS are reviewing the possibilities of publishing lower level Internet Sales data.

2.1 Difference between Non-store retailing and Internet Sales

Internet sales statistics cover the whole of the retail industry as such data is available on the Internet sales of for example department stores and supermarkets. In contrast to this Non-store retailing is a subset of the retail industry and includes only those retailers that sell predominantly online or through mail order.

---

## Internet sales index categories and their percentage weights in 2013

<table>
<thead>
<tr>
<th>Publication category</th>
<th>Percentage of All retailing</th>
<th>RSI Aggregate</th>
<th>SIC**</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predominantly food stores</td>
<td>15.2</td>
<td>Agg1</td>
<td>47.11</td>
<td>Non-specialised stores with food, beverages or tobacco predominating (e.g. supermarkets)</td>
</tr>
<tr>
<td>Non-specialised food stores</td>
<td>14.7</td>
<td>Agg26</td>
<td>47.21</td>
<td>Fruit and vegetables</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>47.22</td>
<td>Meat and meat products</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>47.23</td>
<td>Fish, crustaceans and molluscs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>47.24</td>
<td>Bread, cakes, flour confectionery and sugar confectionery</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>47.29</td>
<td>Other food, beverages and tobacco in specialised stores</td>
</tr>
<tr>
<td>Retail sale of alcoholic drinks, other beverages and tobacco</td>
<td>0.2</td>
<td>Agg27</td>
<td>47.25</td>
<td>Alcoholic and other beverages</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>47.26</td>
<td>Tobacco products</td>
</tr>
<tr>
<td>Predominantly non-food stores</td>
<td>37.3</td>
<td>Agg12</td>
<td>47.19</td>
<td>Non-specialised stores where sales of food, beverages and tobacco is not predominant (e.g. department stores)</td>
</tr>
<tr>
<td>Non-specialised stores</td>
<td>8.4</td>
<td>Agg5</td>
<td>47.51</td>
<td>Textiles</td>
</tr>
<tr>
<td>Textile, clothing and footwear stores</td>
<td></td>
<td></td>
<td>47.71</td>
<td>Clothing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>47.72</td>
<td>Footwear and leather goods</td>
</tr>
<tr>
<td>Household goods stores</td>
<td>4.9</td>
<td>Agg7</td>
<td>47.52</td>
<td>Hardware, paints and glass</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>Agg31</td>
<td>47.54</td>
<td>Electrical household appliances</td>
</tr>
<tr>
<td></td>
<td>2.1</td>
<td></td>
<td>47.59</td>
<td>Furniture, lighting equipment and household articles not elsewhere classified</td>
</tr>
<tr>
<td></td>
<td>1.7</td>
<td></td>
<td>47.63</td>
<td>Audio and video equipment</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td></td>
<td>47.64</td>
<td>Music and video recordings</td>
</tr>
<tr>
<td>Other non-food stores</td>
<td>10.6</td>
<td>Agg13</td>
<td>47.73</td>
<td>Dispensing chemists</td>
</tr>
<tr>
<td>Pharmaceutical, medical, cosmetic &amp; toilet articles</td>
<td>0.1</td>
<td>Agg32</td>
<td>47.74</td>
<td>Medical and orthopaedic goods</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td></td>
<td>47.75</td>
<td>Cosmetic and toilet articles</td>
</tr>
<tr>
<td>Other retail in specialised stores</td>
<td>2.2</td>
<td>Agg35</td>
<td>47.41</td>
<td>Computers, peripheral units and software</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td></td>
<td>47.42</td>
<td>Telecommunications equipment</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td></td>
<td>47.61</td>
<td>Books</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td></td>
<td>47.62</td>
<td>Newspapers and stationery</td>
</tr>
<tr>
<td>Other retail sales in specialised stores not elsewhere classified (NEC)</td>
<td>3.5</td>
<td>Agg33</td>
<td>47.64</td>
<td>Sporting equipment</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td></td>
<td>47.65</td>
<td>Games and toys</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td></td>
<td>47.76</td>
<td>Flowers, plants, seeds, fertilisers, pet animals and pet food</td>
</tr>
<tr>
<td></td>
<td>1.7</td>
<td></td>
<td>47.77</td>
<td>Watches and jewellery</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td></td>
<td>47.78</td>
<td>Other retail sales of new goods in specialised stores</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>47.79</td>
<td>Second-hand goods in stores</td>
</tr>
<tr>
<td>Other non-store retailing</td>
<td>47.5</td>
<td>Agg14</td>
<td>47.91</td>
<td>Mail order houses (including internet retailers)</td>
</tr>
<tr>
<td>Mail order houses</td>
<td>46.1</td>
<td></td>
<td>47.81</td>
<td>Stalls and markets</td>
</tr>
<tr>
<td>Other non-store retailing</td>
<td>14.4</td>
<td>Agg11</td>
<td>47.99</td>
<td>Other non-store retail sale</td>
</tr>
</tbody>
</table>

*Based on aggregated annual data from the Retail Sales Inquiry, 2013.
** UK Standard Industrial Classification 2007
3. Scope and coverage

Internet sales statistics are compiled using data from the Monthly Business Survey – Retail Sales Index (MBS-RSI). The MBS-RSI covers businesses which are engaged in retail trade and which operate in Great Britain; that is England, Scotland and Wales. It does not include Northern Ireland, the Isle of Man or the Channel Islands.

Internet sales covers spending on goods, as such spending on services such as holidays, air fares & train tickets, insurance & banking and digital downloads are not covered and retailers are asked to separate out turnover from services from their return.

4. Data collection and sampling

4.1 Questionnaire

Internet sales statistics are collected using the MBS-RSI questionnaire. Retailers are asked to provide the value of Internet sales for a given accounting period.

The accounting period follows a 4-4-5 week pattern. As a result, over the course of a 365 day calendar year (or 366 days in a leap year), one (or two) trading days a year are lost. To compensate every five or six years an extra week is added to the cycle, usually in January. The most recent five week January 2014.

Most businesses are able to report data for these accounting periods, however, some provide data on a calendar month basis. The questionnaire asks respondents to indicate the dates to which the returned turnover data relates, that is, accounting period or calendar month.

4.2 Mode of collection

For the main retail sales sample, data are collected using telephone data entry (TDE). Respondents are sent a letter specifying the questions and instructions on how to return their data through the TDE system.

In order to ensure a good response rate, approximately 64% each month, response chasing is conducted by trained staff and allows respondents to return data over the phone.

4.3 Sampling

Retail sales data are collected from a sample of approximately 5,000 retailers across Great Britain. The sample represents the whole retail sector and includes all large retailers and a representative sample of smaller businesses. The known retail industry population is approximately 200,000 businesses and while the sample represents 2.5% of this population in terms of number of businesses the sample covers approximately 93% of all known turnover in the retail industry.

The Inter-Departmental Business Register (IDBR) is used as the sampling frame for the MBS-RSI. A sampling frame is a list of all units in the population of interest, in this case a list of all businesses in the UK that we can obtain information from.

A stratified random sample is used to ensure small and large businesses are represented according to the retail trade population structure. First retailers are classified to their industrial grouping as defined by SIC 2007. Five bands or strata are then created based on employment size:

- Band one, 0-4 employment
- Band two, 5–9 employment

- Band three, 10–99 employment
- Band four, 100+ employment
- Band five, employment between 10–99 with turnover greater than £60 million a year.

A random sample is taken from bands one, two and three. Retailers in bands four and five are continuously sampled.

5. Editing, validation and imputation

5.1 Editing and validation
The MBS-RSI uses an editing and validation approach known as selective editing. Selective editing is an internationally recognised method that uses a data-based approach to assess the influence of business estimates on the aggregate outputs.

The selective editing approach means that the editing process should be more efficient and effective since it will only edit potential errors that have a significant impact on final outputs.

Under selective editing key variables on the questionnaire are defined and scores derived for these. The scores compare the value provided on the questionnaire with expected values, where the expected values are generally estimated using past data or other available information related to the variable, for example, from administrative sources. The scores from ‘key’ variables are then combined to derive a score for every questionnaire. This derived, single score for the business’s return is then tested against a methodologically set threshold. If the score is higher than the defined threshold then the questionnaire will fail and be flagged for manual editing. Thresholds have been derived and set to ensure minimum bias is introduced from not editing values that may have been edited under the previous edit all returns system. Thus selective editing does result in an adverse impact on output quality.

In order to ensure that all errors are captured, before questionnaires pass through selective editing they are subjected to automatic editing and then a number of user-defined checks. For example, these will include checks to detect changes to reporting periods and implausible returns such as turnover being less than zero.

5.2 Imputation
Each month approximately 64% of respondents complete and return the MBS-RSI questionnaire, equating to approximately 93% coverage by turnover. Ratio imputation, explained below, is used to provide estimates for non-respondents.

For those respondents who do not return their data in time for publication of the retail sales aggregates, the total turnover and total internet sales are imputed using a ratio imputation method known as ratio of means\(^5\).

---

6. **Estimation**

Estimation is the process of approximating some characteristic of a population, in this case retail sales in Great Britain, when information is available only for a sample of the population. In order to provide estimates of the full retail sector, ratio estimation is used. Ratio estimation is used when the ratio between the returned turnover over the registered turnover variable on IDBR is roughly constant and the variance of the returned turnover can be assumed to be proportional to over the registered turnover. The estimate for retail sales in Great Britain is then given by multiplying the estimate of the ratio by the number of retail business on the IDBR.

Unlike the method used to estimate the full retail sector an expansion or number raised estimator is used to estimate the value of Internet retail sales. Number raised estimation works by ‘expanding’ the sample total up to the size of the population by multiplying the sample total by the ratio of the number of retailers on the IDBR divided by the number of business within the sample.

7. **Indexing**

Estimates for Internet sales are published in both index and monetary form. Unlike estimates of retail sales only value indices are produced. ONS are currently reviewing the possibility of creating volume indices and are researching what price statistics are available for them to use in order for these indices to be created.

7.1 **Reference period**

The reference period is the year for which the index is scaled to equal 100. Internet value indices are currently published in index form with a reference year of 2011 equal to 100.

To keep RSI in line with the UK National Accounts, the reference year is updated each year when weights are available from the Blue Book. The change to the reference year is a simple calculation and does not impact on the movements of the series.

7.2- **Deflation and volume estimates**

It is not possible to create volumes estimates of Internet sales due to the availability of suitable prices to use as deflators. ONS have looked at the possibility of using commodity Consumer Price Indices in the same way as is done for Retail Sales, however, there is concern that these prices may not fully reflect the prices of goods sold online and thus this work is ongoing.

8. **Time series analysis**

Seasonally adjusted estimates are derived by estimating and removing calendar effects (for example Easter moving between March and April) and seasonal effects (for example increased spending in December as a result of Christmas) from the non-seasonally adjusted (NSA) estimates. Seasonal adjustment is performed each month, and reviewed each year, using the standard, widely used software, X-12-ARIMA. Before adjusting for seasonality, prior adjustments are made for calendar effects (where statistically significant) such as: returns that do not comply with the standard trading period, bank holidays, Easter, and the day of the week Christmas occurs.

While the number of trading days in a set period remains the same, the standard reporting periods of 4-4-5 weeks introduce a ‘phase-shift’ effect, associated with the fact that the standard periods do not match calendar months and move slightly each year relative to the calendar month. In a typical year, the standard periods total 52 weeks or 364 days. As a result, the standard periods fall back one day each year (or two days for a leap year).

---


Every five or six years, the reporting year is brought back into line with the calendar year by adding an extra week, normally to January.

The seasonally adjusted estimates also have corrections for bank holidays; Easter effects which are caused when Easter falls late in March or early in April and the day of the week Christmas occurs. Prior connections are applied as necessary. Corrections are estimated and applied where there is a statistically significant effect.

The calculation of the RSI has an adjustment to compensate for calendar effects which arise from the difference in the reporting periods.

9. Confidentiality and disclosure
The Retail Sales Index is compliant with the Code of Practice for Official Statistics8, Principle 5.

Principle 5, Practice1, states:
Ensure that official statistics do not reveal the identity of an individual or organisation, or any private information relating to them, taking into account other relevant sources of information.

All necessary steps are taken to protect the confidentiality of data collected from retail businesses. This includes statistical disclosure controls to ensure that individual businesses are not identified in the published statistics.

When data are shared with other bodies, for example Eurostat and the Scottish Government it is done so under legislation and using secure electronic file transfer methods.

10. Dissemination
Statistics covering Internet sales estimates are disseminated monthly through:

- The Retail Sales Statistical Bulletin
- Retail Sales Tables

The Retail Sales Statistical Bulletin presents the key messages within the data, analysis of growth rates and the proportion of online sales of total retail sales (excluding automotive fuel).

Internet sales publication dates coincide with the publication dates for RSI and can be found on the UK National Statistics Publication Hub release calendar www.statistics.gov.uk/hub/release-calendar/index.html.

11. Reliability of estimates
There are two possible types of error in estimates of retail sales, sampling error and non-sampling error.

- Sampling error occurs because a sample, rather than the entire population, is surveyed. It is the difference between the true value for the population and the estimated value. One way of measuring this difference is through standard errors9. While standard errors have been calculated for retail sales statistics, standard errors for Internet sales are under review and as such are unavailable.

---


• Non-sampling error arises from inaccuracies in collecting, recording and processing the data. The most significant of these errors are: misreporting of data items, deficiencies in coverage, non-response, and processing errors. Every effort is made to minimise reporting error by the careful design of questionnaires, intensive training and supervision of editing and validation staff, and efficient data processing procedures.

12. Revisions
Internet sales non-seasonally adjusted data is revised as needed. This typically occurs due to late data returns, updated respondent information, replacing adjustments with actual data, and reclassifications of respondents to the appropriate category either within or out of retail. Changes are not made to any non-seasonally adjusted data prior to 2001. Revisions to non-seasonally adjusted data will directly impact on the seasonally adjusted estimates.

Revisions to all published seasonally adjusted estimates are not constrained or restricted as they are for other short term indicators and the national accounts. Revisions are allowed to occur naturally each month along the full length of each time series as a consequence of the direct seasonal adjustment of the component time series. Aggregate seasonally adjusted estimates will then be derived. Any revisions from the directly seasonally adjusted time series will then occur as part of the higher level aggregation. This approach ensures that the RSI uses all available information and provides the most up-to-date estimates for recent periods.

Internet sales revisions are covered by the RSI revisions policy10.

13. Comparability with other retail sales indicators
There are other indicators of Internet sales available. These include the British Retail Consortium’s (BRC)11 United Kingdom Retail Sales Monitor and IMRG and Capgemini e-Retail Sales Index. There are differences between the indicators in relation to coverage and compilation that can lead to discrepancies between published figures. Notably the RG and Capgemini e-Retail Sales Index includes the amount spent online on services such as holidays and music downloads.

11 http://www.brc.org.uk/brc_home.asp