Exploitation of HMRC VAT data

Author: Mark Stephens and John Allcoat, Office for National Statistics
Date: 7 October 2015

Executive summary

This article is the next update on the work to exploit data collected by Her Majesty’s Revenue and Customs (HMRC) from Value Added Tax (VAT) returns as an administrative data source for short-term output indicators (STOI) and National Accounts. The STOI in scope are the Index of Production (IoP), Index of Services (IoS) and Output in the Construction Industry.

This article:

• describes the use of HMRC turnover data within STOI and National Accounts in an international context and explains the UK history of the use of HMRC turnover data (section 2);
• considers how HMRC turnover sourced from a VAT return can be used within STOI and National Accounts (section 3.1);
• describes the potential of exploiting other HMRC variables beyond turnover (sections 3.2 and 4);
• provides more detailed information regarding HMRC turnover data and how we process and derive it at the micro-level (section 5).

This is the second in a series of articles to update you on our progress; the first article was published on 14 August 2015 entitled Feasibility study into the use of HMRC turnover data within Short-term Output Indicators and National Accounts.

1 Background

The first article set out the plans and intentions for the project and explored some of the benefits and challenges. ONS have already engaged with main users such as Her Majesty’s Treasury (HMT), the Bank of England and the Office for Budgetary Responsibility through the short-term output indicators stakeholder group held on the 13 August 2015 and the team conducting an independent review of economic statistics. A common theme has been to describe and explain our place in an international context and provide further detail about the method and process to obtain the micro-data.

2 The international context and the history of UK practice of VAT turnover data for short-term output indicators

2.1 The international context

Our European counterparts use more turnover data sourced from a VAT return in National Accounts and short-term output indicators than we do at present, with Scandinavian countries being the most advanced. However, the UK position is influenced by some important differences.
It should be understood that the UK is at a disadvantage in the availability of monthly VAT data. For instance the UK VAT return has 9 variables that must be completed which is less than for example, Italy. But crucially the UK only has some 10% of HMRC turnover reported on a monthly basis (with 89% quarterly and less than 1% annually), one of the lowest in the European Union (EU). In contrast Finland (Efficient use of administrative data in the production of economic statistics in Finland) and Germany (The integrated system of editing administrative data for STS in Germany) have more than 90% of turnover reported from enterprises on a monthly basis. However, with access to VAT micro-data we are now able to construct “monthly” estimates from staggered quarterly VAT reports to overcome this longstanding issue, that is, for the latest month, data will be available from businesses reporting monthly and the one third of businesses which cover this latest month on the quarterly stagger (see section 5.2 and table 1 on the quarterly stagger and the move from quarterly to monthly returns). This will ensure that larger amounts of data would be available for estimates of Gross Domestic Product (GDP).

The first estimate of UK GDP is one of the timeliest in the world. In order to achieve this it is based exclusively on the output approach to the measurement of GDP (GDP(O)) as survey data from our Monthly Business Survey (MBS) is timelier than VAT data. The UK is one of a few countries which lead with the output approach, most lead with the expenditure approach to measuring GDP. This means European countries are afforded more time to assemble data from VAT returns and use these at later stages of the quarterly GDP cycle. Moreover many counterparts take advantage of a different culture of businesses providing data to government agencies and/or the sharing of this administrative data. For example, in Finland there is significant sharing of administrative data with monthly VAT returns, covering some 93% of turnover from enterprises, used in business statistics since 1999.

Many countries have a longstanding monthly index of industrial production, but few have a well developed monthly index of services. At European level this is in the process of change, with plans for a monthly index of services to become an EU requirement under the Short Term Statistics regulation. In this context the UK is very much a world leader. The article entitled The Challenges of Measuring the UK Service Sector highlights this further. In May 2007, the UK monthly index of services became a National Statistic due in part to the development of timely monthly turnover inquiries to cover much of the service sector. To further highlight the UK’s position in international comparisons for measuring the service sector, we are hosting an EU training event for Member States in May 2016 to share practical experiences in developing and maintaining the index of services. This builds on our contributory role with the Organisation for Economic Co-operation and Development Compilation manual for an index of service production published in 2007.

The most common European model is to use both administrative and survey data with survey data covering larger enterprises and administrative (VAT) data covering small and medium enterprises (See the paper The use of incomplete admin data for monthly and quarterly estimates: findings after two year ESSnet Admin data). With good quality survey estimates of turnover available for larger businesses for the preliminary GDP estimate and the comprehensive coverage of small and medium sized businesses from HMRC, we plan to construct a new model for estimating business activity in consultation with users. We wish to adopt this model as it will ease the burden on small and medium-sized businesses and
improve our efficiency, while the quality of the output will be improved by the significant coverage of business activity that HMRC turnover delivers.

2.2 The historical use of VAT data for short-term output indicators

Data from VAT sources are not currently used in deriving STOI estimates. However, for completeness it should be noted that this has not always been the case. In the 1990s, when quarterly turnover inquiries covering service sector industries, were launched and mirrored our monthly surveys covering the production sector, HMRC turnover data was used to benchmark estimates. With the move towards monthly surveys for the service sector in the late 1990s to enable a monthly index of services, HMRC turnover data sourced from a VAT return was considered as a data source as opposed to our monthly surveys. The main issues with VAT data at that time were the lack of timeliness and the divergence between HMRC aggregate data constructed with industrial classifications, which were different from the classifications we used through the National Accounts. Even so, some HMRC data was used as an indicator. However, as the monthly surveys were steadily developed the use of VAT as a benchmark and a data source has declined. In 2009 a new MBS replaced the previous HMRC supply for the final industry (taxi operations) and HMRC turnover has not been used within STOI since that time.

In summary, historical issues which prevented the use of HMRC turnover data have been overcome in part by having access to HMRC micro-data at business level to construct monthly data (as outlined in section 2.1 – The international context and in more detail in section 5.2 – Overview of the process). Moreover, by feeding the micro-data through our register of businesses, the Inter-Departmental Business Register (IDBR), we can achieve a more consistent approach in the reporting of businesses according to their standard industrial classifications across HMRC and ONS and, more importantly, across the integrated system of National Accounts.

3 Improving the quality of short-term output indicators and the National Accounts

3.1 Turnover

HMRC turnover at micro-data level is available across the whole economy and provides a census of activity for businesses registered to pay VAT. This includes all businesses exceeding the VAT turnover threshold and businesses that are under the VAT registration threshold but are registered for VAT (the current VAT threshold for businesses for the financial year ending 2016 is a turnover of more than £82,000 per annum). This is a potential qualitative improvement in comparison with our surveys such as the MBS. In the first article it was stated that efficiency savings could be made in cutting the number of MBS questionnaires for small and medium-sized enterprises and this remains an important aim of the project. The MBS covered 43% of the economy as a proportion of the weight or gross value-added (outputs less inputs) of each industry within the overall economy in 2012.
However, it should be noted that the MBS is one of many data sources used to provide turnover, or current price, estimates of activity as described in the GDP(O) source catalogue.

- estimates from General Government Final Consumption Expenditure (GGFCE) from HMT estimates of public sector expenditure, account for approximately 17% of the economy
- estimates from Household Final Consumption Expenditure (HHFCE) includes imputed rent, actual rent and gambling, cover approximately 11% of the economy
- estimates from Financial Corporations (FINCO), derived primarily from the Bank of England, account for approximately 6% of the economy
- further estimates from our inquiries, covering retail and construction, both each account for approximately 6% of the economy respectively and other sources account for approximately 4%

In addition, approximately 8% of the economy does not have a turnover or current price data source. These are industries where volume estimates are used to measure activity (for example energy and mining volumes and jobs data). The turnover of the industry is “derived” from reflating volume data with appropriate price indices.

As part of this project we will consider the suitability of HMRC turnover estimates replacing other sources. In addition, we will be able to use HMRC turnover estimates to benchmark or sense check estimates, for example construction output.

### 3.2 Expenditure

HMRC also collect expenditure data on a VAT return which is similarly available across the whole economy. Although this is not yet a focus for the project, access to expenditure data presents an opportunity for a qualitative improvement to National Accounts. Expenditure data includes business purchases of both capital assets and other input costs, or purchases. Data on purchases, when modelled as a separate series, could potentially function as an estimator of intermediate consumption. As the data are available on a monthly and quarterly basis this could improve estimates which are currently only available on an annual basis from our Annual Business Survey. This could potentially help towards the long-term aspiration to develop double deflation (the separate deflation of both inputs and outputs for GDP(O)) and quarterly supply and use tables. Annual supply and use tables are integral to National Accounts as they bring together the output, expenditure and income measures of GDP and, when balanced, provide a single measure of current price GDP which integrates the components of gross value-added, inputs and outputs, and final demand. This preferred approach would be timelier through quarterly tables.

### 3.3 Value-added or margins

In deriving data for turnover (or output) and purchases (intermediate consumption) there is potential for us to isolate value-added (outputs less intermediate consumption), as stated in section 3.2. For STOI the turnover of a business is an effective proxy of its output. However, for some industries the turnover of a business is not the most appropriate indicator. For the wholesale (UK SIC2007 division 46), retail (UK SIC2007 division 47), motor trades (UK SIC2007 division 45) and travel agency industries (UK SIC2007 division 79) the output we wish to measure is the service provided in the selling of goods and services (although
consumers may not necessarily recognise this as being the case). For example, the retailer
does not normally produce the good that is sold – the retailer buys goods and then trades
them at a different price. It is the difference between the prices that is the helpful in
understanding the value-added of the retailer. (It is important to note that by EU regulation
the Retail Sales Index is a measure of the retail of goods and specifically excludes the
turnover generated by other activities including the sale of services. It is for this reason that
HMRC turnover data would not be conceptually consistent with the required regulation).
Similarly with travel agents, it is not the total turnover of the travel agent we wish to measure,
but the commission or fee that the business receives for arranging the transaction.
At present MBS turnover series and Retail Services Inquiry data are used for these 4
industries. We will explore the data for these industries to identify if improvements can be
made that would materially impact how these industries are treated as part of STOI.
The new data may also allow us to consider its use in the construction sector. Here there are
continuing issues in the measurement of prices and output which may be improved by a
better understanding of the margins of individual businesses. We will consider this issue as
part of a future article.

4  The potential of HMRC data

This project is actively considering the use of turnover and expenditure data, but other data
are also potentially within scope. Specifically, we will explore the availability of export, import
and Pay As You Earn (PAYE) data from HMRC. If we have the legal framework to use the
data for statistical purposes, while fulfilling important requirements on, disclosure for
example, then access to such a wide array of business information could be a powerful
analytical tool for policymakers. This would be enhanced with the use of such data at a
regional level.

This is an area of technological exploitation that ONS will actively pursue as a pathfinder for
the exploitation of administrative data. It has significant overlap with the flow of funds project
and so both projects will proceed on a joint basis to maximise the potential advantages
offered by a more scientific approach to the capture, storage, use and analysis of large
datasets.

5  Our process to obtain HMRC data

We have established and run a monthly process that allows the creation of micro-data from
the HMRC turnover dataset. The process allows matching of differing HMRC and ONS
business reporting structures – the HMRC VAT reporting unit and the ONS reporting unit.
This creation of matched micro-data is an important feature of the project and a significant
development over previous attempts to use the data within STOI.

5.1  Inter-Departmental Business Register structure and key features

Before describing the process it is important to understand the structure and procedure to
update the IDBR. This will explain some important terms and improve understanding of the
issues which define IDBR unit structures.
The IDBR is updated from a variety of different data sources which provide a range of different units dependent on these data sources. These data sources provide information to update business structures along with beneficial auxiliary information. These include:

- HMRC VAT – for VAT registrations, business identification details and HMRC turnover
- HMRC PAYE – for PAYE registrations, business identification details and PAYE employment
- Companies House – for Limited Companies registrations and business identification details
- Department for Environment Food and Rural Affairs – for farm holdings registrations and business identification details
- ONS Surveys – for turnover, employment and geographical information and industrial classification

The range of different data sources used to update the IDBR is to ensure the information and sampling frame is as complete and up to date as possible.

Figure 1 shows the main issue within the project. The monthly VAT turnover data from HMRC is delivered to us at VAT reporting level. For most VAT traders their VAT registration and reporting structures are the same. However, some businesses form a group for VAT reporting purposes and one registered VAT trader reports the turnover for the whole VAT group. Therefore a VAT trader could be matched to more than one enterprise unit from IDBR.

Each of these enterprise units can then be represented by one or more statistical reporting units. In order for comparison purposes for our business surveys and for the UK to meet its legal requirement to provide data under the Short Term Statistics regulation (STS), data are required to be produced at the kind-of-activity unit (KAU). In terms of the IDBR the reporting unit level satisfies the requirements of the KAU. This therefore requires a procedure
(described in 5.2 – Overview of the process) to match the VAT reporting unit level data to the reporting unit for statistical analysis and comparison purposes.

There is further information about IDBR sources, structure and updating for publications located on our website.

5.2 Overview of the process

The process conducted each month by the business area is as follows.

- HMRC deliver the dataset of all VAT returns that have been returned for the previous month, via a secure, electronic dataset each month to the IDBR. It is important to highlight that these relate to various (current and previous) reference periods in 16 different staggers, as businesses can report their VAT on a monthly, quarterly or annual basis.
- these are a single monthly stagger, 3 quarterly staggers and 12 annual staggers and businesses can report any stagger to fit in with their business needs. (Table 1 in the annex shows all the possible reporting staggers for a VAT return). For example, VAT period 502 relates to January 2015. This represents the monthly VAT returns for January 2015, the quarterly VAT returns for the different quarterly staggers where January 2015 falls (as the first, second or third of the three months) and annual VAT returns for the different annual staggers where January 2015 falls.
- the HMRC turnover data are delivered on the first Tuesday of each month.
- the HMRC data files at the VAT reporting unit level are uploaded to the IDBR the following weekend.
- the following Monday the data are extracted by the business area. Along with extracting the HMRC dataset at VAT reporting unit. A snapshot of the IDBR is also taken by the business area.

A suite of programs are then run in order to match the data from VAT reporting unit to the statistical reporting unit level.

a) Reading in the monthly VAT files from HMRC

The first program reads in the monthly VAT files into SAS to allow data manipulation and interrogation by the business area. It firstly orders the VAT returns by the reporting period they refer to. Some validation checks are conducted at this point of the process. This ensures only 1 VAT return is attributed to a VAT reporting unit per VAT reporting period. Also, where revised VAT returns have been submitted for a VAT period, the program ensures the latest return is used within the dataset. It is assumed this is a revision from the VAT reporting unit and perceived to be a more accurate reflection of their turnover than previous returns for the period.

b) Preparation of the IDBR extracts

The next SAS program prepares the IDBR extracts for the matching to the VAT reporting units. It establishes the most relevant (time wise) structure between VAT registration unit – statistical enterprise unit – statistical reporting unit(s) and aggregates them to VAT reporting unit level. During this process the proportions between the statistical reporting unit’s employment to the enterprise employment and each individual enterprise employment to the
sum of employment in the group of enterprises associated with 1 VAT reporting unit are calculated. These proportions are then used in the third SAS program “(c) Linking VAT reporting unit with reporting units” to apportion the HMRC turnover to the corresponding statistical reporting units.

The method to apportion HMRC turnover to the reporting unit on the basis of reporting unit employment percentage within the enterprise, has been tested and agreed as the best method to apportion the data. This was part of the work from European Statistical Service Network (ESSNET) project, the use of administrative and accounts data for business statistics.

c) Linking VAT reporting unit with reporting units

The next part of the process takes HMRC turnover from its VAT reporting unit and links this turnover to its corresponding enterprise. It is then linked again on the basis of the reporting unit employment proportions within the enterprise calculated within “b) Preparation of the IDBR extracts”.

It is worthwhile to note that, as HMRC data is a valuable and important component of updating the IDBR, all VAT registration units can find a match in terms of enterprise. From the enterprise the reporting units employment proportions are then used to assign HMRC turnover to the reporting unit. All HMRC turnover can be attributed to reporting units at this stage. However, reporting units with zero employment mean HMRC turnover cannot be assigned on the basis of reporting unit employment. This means an element will not be linked and cannot be used in the analysis. Instances in which this could be improved are highlighted within 5.3 – Potential areas of improvement to the process.

At this stage the reporting unit is assigned a complexity marker of either “simple” or “complex”. These are based on the relationship between VAT reporting unit, enterprise and reporting unit. If 1 VAT reporting unit can be matched to 1 enterprise and then matched to 1 reporting unit this is described as a “simple” relationship. Based on early preliminary findings across 2014, approximately 90% of reporting units within the economy have this relationship. The remaining 10% of reporting units within the economy have a “complex” relationship. As expected these “complex” businesses tend to be the larger, more detailed business structures. Early analysis has shown that these “complex” businesses account for approximately 78% of the total turnover in the economy, based on 2014 data. A “complex” relationship could therefore be either of the following relationships:

i. 1 VAT reporting unit to many enterprises to many reporting units
ii. 1 VAT reporting unit to 1 enterprise to many reporting units

d) Validating and cleaning

The next step in the process validates and cleans the apportioned data. The data received from HMRC are administrative, therefore a number of processes are required to make sure that simple rules can be applied to improve the quality of the data where appropriate. The cleaning rules have been devised as part of the European Statistical System Network programme on the use of administrative data. (Editing and Imputing VAT Data for the Purpose of Producing Mixed-Source Turnover Estimates and Issues to consider when turning to the use of administrative data: the UK experience have further background information about the issues and possible solutions for validating and cleaning turnover data from administrative data sources).
The cleaning rules are:

**Thousand pounds rule**
This rule is applied to businesses who have possibly incorrectly reported their VAT return in thousand pounds instead of pounds. This rule works by calculating a ratio between current VAT return in comparison to their previous VAT return. If the ratio falls between the values 0.00065 and 0.00135 it is multiplied by 1000.

**Quarterly patterns rule**
This rule is only applied to businesses who report their VAT on a quarterly basis. There are 3 variations for the quarterly patterns rule. The aim is to try to understand whether these businesses are reporting true quarterly HMRC turnover, by identifying suspicious quarterly patterns. Therefore, the cleaning under this rule is based on:

- reporting units having exactly the same positive values for any 4 consecutive quarters; this implies that the business is actually reporting annual values allocated equally between the 4 quarters
- reporting units having exactly the same positive values in any 3 consecutive quarters and then a different value for the fourth quarter; this implies that the business is assessing its annual value and allocating it between the 4 quarters. The fourth quarter therefore is allocated the residual value to sum to the annual value
- reporting units having zero values in any 3 quarters and then a positive value in the fourth quarter; this implies the business is returning an annual value.

In all these situations the annual value is re-assigned to the corresponding quarters using the median proportions between the quarters calculated from the businesses in the same class with genuine reported values. A class is defined as the industry section as per the UK SIC2007 industrial classification and employment sizeband. (Table 2 within the annex has a description of employment sizebands (“empband” variable).)

**Suspicious turnover rule**
This rule identifies reporting units that have suspicious turnover for a VAT return. A return is deemed suspicious by firstly matching their current VAT return to their previous VAT return. (This applies for all the reporting schedules). A value is deemed suspicious as a reporting unit will be compared to all reporting units within that employment sizeband at the UK SIC 2007 class level. (The employment sizebands have been set at employment groupings of 0 to 9, 10 to 49, 50 to 99 and 100 and over). Once the data has been stratified by class and employment sizeband the reporting unit’s current and previous VAT returns will be tested in comparison to the median value of the class and reporting stagger. A set of criteria in terms of scores is then produced and if it falls outside these scores it will be deemed as a suspicious turnover value. This is then replaced by a value which is the ratio of the current period sum of VAT turnover divided by previous period sum of VAT turnover for the total UK SIC 2007 class and employment sizeband multiplied by the reporting unit previous period VAT turnover figure.

A clean marker is then applied to the reporting unit to ascertain whether the data has been cleaned or not (“0” = no cleaning applied “1 to 5” = cleaning has taken place on the reporting
unit). Further analysis is planned to understand the effects of the cleaning rules on the dataset.

e) Dividing HMRC turnover into monthly data

Once the data has gone through steps a to d we then produce monthly data from the quarterly and annual values at the reporting unit level. We take quarterly returns and divide by 3 and assign this amount of HMRC turnover to the VAT period months for which the quarterly stagger relates to. Likewise an annual return is divided by 12 and HMRC turnover is assigned to the VAT period months for the annual stagger relates to.

Work has been undertaken to investigate the issue of using a primarily quarterly administrative data source for the compilation of monthly STOI indicators. This explored how to interpolate the monthly path from quarterly and annual VAT returns (The internal ONS paper “Interpolation and extrapolation of turnover from value added tax returns” covered this further). Using a select sample of industries it was found that no single method could be applied for all industries to interpolate the monthly path, therefore a simple method of allocating quarterly and annual returns equally to the months they correspond to was proposed.

Work and analysis is planned to improve this process and is further discussed under 5.3 – Potential areas of improvement to the process.

This process described in steps a to e takes around one day to run each month. After which, a rich and detailed dataset by reporting unit is produced. For each dataset a wide range of variables are produced. Table 2 within the annex has a full list of variables.

5.3 Potential areas of improvement to the process

Whilst an established and repeatable process, which is able to overcome the difficulties of data as collected for one purpose on a different business unit structure, has been developed; there are still some improvements that could be made to the process that would improve the quality of the dataset. We aim to find out how these could be implemented over the coming months.

We are looking at how we can improve the method to produce monthly data from quarterly and annual returns. At present the method is very simplistic to produce monthly data from the quarterly and annual returns. It is worth emphasising that the majority of VAT returns are quarterly (approximately 90%) and therefore the process is having to interpolate data into a monthly path. This is an important requirement for STOI as IoP, IoS and the Output in the Construction industry are all monthly indicators. Possible ways in which improvements could be made is to implement a calendar or working day adjustment rather than just simply dividing by 3 or 12 dependent on whether it is a quarterly or annual return and attributing to the appropriate reporting staggers. Another more complex method is to apply a cubic spline to the quarterly data.

We will be investigating and analysing the cleaning rules. Further work is required to fully understand the effects of the cleaning and validating process. As businesses cannot be contacted by us to understand unusual data points this is the only process available to clean
and validate the data. However, the extent to which the data could be cleaned requires further investigation.

We will also consider the apportionment rules that apportion HMRC turnover to reporting units. A limitation of the method is that we are aware of instances where employment is a less optimal predictor of turnover, for example, businesses involved in wholesale and financial services where reporting units with small employment should correctly reflect significant turnover values.

6 Next steps

A further article to update users as to progress with the project is planned for publication November 2015.

7 Contact details

Name: John Allcoat
Email: stoi.development@ons.gsi.gov.uk
Telephone: 01633 456616
8 References

Allcoat, J (2015) “Feasibility study into the use of HMRC turnover data within Short-term Output Indicators and National Accounts” Office for National Statistics


Lewis, D and Woods, J (2013) “Issues to consider when turning to the use of administrative data” ESSNET on the Use of Administrative and Accounts Data for Business Statistics working paper


Parkin, N (2010) “Interpolation and extrapolation of turnover from value added tax returns” ONS internal paper

9 Annex

Table 1 – List of different reporting staggers for VAT return

The table below illustrates the 16 different reporting staggers that a business could report their VAT to HMRC.

<table>
<thead>
<tr>
<th>Stagger type number</th>
<th>Description of the different types of VAT reporting staggers</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Monthly returns</td>
</tr>
<tr>
<td>1</td>
<td>Quarterly returns due in March</td>
</tr>
<tr>
<td>2</td>
<td>Quarterly returns due in January</td>
</tr>
<tr>
<td>3</td>
<td>Quarterly returns due in February</td>
</tr>
<tr>
<td>4</td>
<td>Annual returns with year ending January</td>
</tr>
<tr>
<td>5</td>
<td>Annual returns with year ending February</td>
</tr>
<tr>
<td>6</td>
<td>Annual returns with year ending March</td>
</tr>
<tr>
<td>7</td>
<td>Annual returns with year ending April</td>
</tr>
<tr>
<td>8</td>
<td>Annual returns with year ending May</td>
</tr>
<tr>
<td>9</td>
<td>Annual returns with year ending June</td>
</tr>
<tr>
<td>10</td>
<td>Annual returns with year ending July</td>
</tr>
<tr>
<td>11</td>
<td>Annual returns with year ending August</td>
</tr>
<tr>
<td>12</td>
<td>Annual returns with year ending September</td>
</tr>
<tr>
<td>13</td>
<td>Annual returns with year ending October</td>
</tr>
<tr>
<td>14</td>
<td>Annual returns with year ending November</td>
</tr>
<tr>
<td>15</td>
<td>Annual returns with year ending December</td>
</tr>
</tbody>
</table>

Table 2 – List of variables for each reporting unit

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ruref</td>
<td>Reporting unit (RU) reference number.</td>
</tr>
<tr>
<td>refperiod</td>
<td>The month for which the VAT turnover refers.</td>
</tr>
<tr>
<td>rx_period</td>
<td>The file where the RU was found to be linked to the VAT trader through an enterprise.</td>
</tr>
<tr>
<td>arrivalperiod</td>
<td>The month when the VAT turnover had become available to ONS.</td>
</tr>
<tr>
<td>complexity</td>
<td>Complexity of the relation between VAT trader and the reporting unit. Simple means 1 VAT trader linked to one enterprise with one RU. Any other variation of the relationship between the 3 units is labelled complex.</td>
</tr>
<tr>
<td>legal_status</td>
<td>The legal status of the enterprise in IDBR.</td>
</tr>
<tr>
<td>live_lu</td>
<td>Number of live Local Units associated with the RU at the refperiod/rx_period.</td>
</tr>
<tr>
<td>frozen_empment</td>
<td>Frozen employment for the RU at the refperiod/rx_period taken from the IDBR.</td>
</tr>
<tr>
<td>current_empment</td>
<td>Current employment for the RU at the refperiod/rx_period taken from the IDBR.</td>
</tr>
<tr>
<td>empband</td>
<td>Calculated as follows: when (0&lt;=current_empment&lt;10) empband='1'; when (10&lt;=current_empment&lt;50) empband='2'; when (50&lt;=current_empment&lt;100) empband='3'; when (current_empment&gt;=100) empband='4';</td>
</tr>
<tr>
<td>frozen_SIC07</td>
<td>Frozen industry classification for the RU at the refperiod/rx_period taken from the IDBR.</td>
</tr>
<tr>
<td>current_SIC07</td>
<td>Current industry classification for the RU at the refperiod/rx_period taken from the IDBR.</td>
</tr>
<tr>
<td>section</td>
<td>Derived from the current industry classification for the RU at the refperiod/rx_period.</td>
</tr>
</tbody>
</table>
Table 3 – List of NUTS1 regions
The concatenation of the gor||ssr will give what is meant by NUTS1¹ region codes in the table below:

<table>
<thead>
<tr>
<th>Code</th>
<th>Region</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>North East</td>
<td>England</td>
</tr>
<tr>
<td>BA</td>
<td>North West</td>
<td>England</td>
</tr>
<tr>
<td>BB</td>
<td>North West</td>
<td>England</td>
</tr>
<tr>
<td>CB</td>
<td>Merseyside</td>
<td>England OLD - Now part of Northwest (BB)</td>
</tr>
<tr>
<td>DC</td>
<td>Yorkshire and Humberside</td>
<td>England</td>
</tr>
<tr>
<td>ED</td>
<td>East Midlands</td>
<td>England</td>
</tr>
<tr>
<td>FE</td>
<td>West Midlands</td>
<td>England</td>
</tr>
<tr>
<td>GF</td>
<td>Eastern</td>
<td>England</td>
</tr>
<tr>
<td>GG</td>
<td>Eastern</td>
<td>England</td>
</tr>
<tr>
<td>HH</td>
<td>London</td>
<td>England</td>
</tr>
<tr>
<td>JG</td>
<td>South East</td>
<td>England</td>
</tr>
<tr>
<td>KJ</td>
<td>South West</td>
<td>England</td>
</tr>
<tr>
<td>WW</td>
<td>Wales</td>
<td>Wales</td>
</tr>
<tr>
<td>XX</td>
<td>Scotland</td>
<td>Scotland</td>
</tr>
<tr>
<td>YY</td>
<td>Northern Ireland</td>
<td>Northern Ireland</td>
</tr>
<tr>
<td>ZZ</td>
<td>Non-UK</td>
<td>Non-UK</td>
</tr>
</tbody>
</table>

¹ NUTS stands for the Nomenclature of Units for Territorial Statistics (NUTS) and provides a single uniform breakdown for the production of regional statistics for the EU.
NUTS1: Wales, Scotland, Northern Ireland and the 9 English regions
NUTS2: 40 sub-regions – mainly groups of counties and unitary authorities
NUTS3: 173 local areas – principally individual counties and unitary authorities