Feasibility study into the use of HMRC turnover data within Short-term Output Indicators and National Accounts

Author: John Allcoat, Office for National Statistics

Date: 14 August 2015

Executive summary

This article provides an update into the early work examining the benefits of incorporating turnover data collected by Her Majesty's Revenue and Customs (HMRC) as part of businesses' Value Added Tax (VAT) returns, within short-term output indicators and National Accounts.

A method has been devised that allows HMRC turnover data to be attributed to individual businesses or reporting units. The method takes turnover data at VAT registration unit level and apportions to the consistent ONS reporting unit based on employment. These data are available monthly and on a UK Standard Industry Classification 2007 basis from January 2007 onwards. Early analysis has shown at the period five months from the reference period month the coverage is approximately 99% in terms of turnover and 96% in terms of number of reporting units. This could have the potential to significantly improve the coverage of businesses in comparison to current data sources.

These data are being analysed to consider their suitability to supplement or replace data collected by ONS turnover surveys. This has significant potential saving implications which are estimated to be 40% in ONS running costs for the relevant surveys and 50% in the cost of completing these surveys to business.

This is the first in a series of articles which aim to share the progress of the work with users.

1   Introduction

1.1   Project Overview

Since 2012, we have been reviewing the data sources used in the output approach to measuring Gross Domestic Production (GDP(O)) as part of the continuous GDP(O) improvement programme. In recent months, we have been exploring the potential uses for turnover data from Her Majesty's Revenue and Customs (HMRC) in the compilation of GDP estimates. These include to quality assure and validate existing data sources, to supplement current data sources or to replace existing data sources. As a result of our early findings, the potential to use HMRC data within short-term output indicators and National Accounts has become a key priority for ONS.

Previous updates on the work have been communicated to users via the; GDP(O) improvement report – September 2014 (see section 3.4), GDP(O) improvement report – June 2013 (see section 3.1), and user seminars held on 19 June 2014 and 2 October 2013 which focused on the improvements to the estimation of the output approach to the measurement of Gross Domestic Product (GDP).

Businesses which meet the VAT turnover threshold must provide data to HMRC in a VAT return, responding to the "Total value of sales and all other outputs excluding any VAT" field. This is comparable to turnover data as collected by the Monthly Business Survey (MBS). Both data sources have been collecting turnover data for many years. This project aims to address the comparability and compatibility of these data sources and will explain how ONS plans to continue to resolve its outstanding barriers to implementation over the forthcoming years.

1.2   Data Overview

Turnover data are collected by HMRC through VAT returns, which captures the businesses' response to the "Total value of sales and all other outputs excluding any VAT" field. This information is provided by businesses on a monthly, quarterly or annual basis dependent on their VAT reporting pattern. It is also important to note that businesses can report any schedule of reporting quarters or annual pattern and therefore there are 16 possible reporting staggers (12 annual reporting staggers, 3 quarterly reporting staggers and a monthly reporting stagger). Taken as a whole, the information is a census of all businesses which are registered to pay VAT (The current VAT threshold for businesses for the financial year ending 2016 is a turnover of more than £82,000 per annum). It is therefore comparable with data collected by the MBS which asks for businesses to provide a response to the question "What was the value of the business's turnover excluding VAT".

The data are supplied by HMRC to the Inter Departmental Business Register (IDBR) on a monthly basis via a monthly, secure electronic feed. These data are firstly matched to a snapshot of the IDBR based on the unique identifier of VAT registration unit number. This allows HMRC turnover data to be assigned to an IDBR enterprise. From the IDBR enterprise level, HMRC turnover is then apportioned to the reporting unit, by the business area, based on the employment of the reporting unit. The data are also validated using rules from the ESSNet project looking into the use of Administrative and Accounts data for Business statistics. The final output is a quality assured HMRC turnover dataset at the reporting unit level. Further information on the process will be covered in future updates on the project. (see section 6 – Next steps).

1.3   Legislation

The HMRC turnover data are supplied to us under section 91 of the VAT Act 1994. This legislation allows us access to the data on the basis of "disclosure of information for statistical purposes". It is important to highlight the act stresses that data which allows identification of individual businesses cannot be disclosed. This means data would only ever be allowed to be shared at the aggregated industry level. Also if data are deemed suitable and can be used as a data source within short-term output indicators and/or National Accounts, again they would only ever be published at the aggregated level with disclosure checks conducted to ensure individual businesses could not be identified.

1.4   UK Statistics Authority Strategy – "Better Statistics, Better Decisions"

The project is completely consistent with the UK Statistics Authority strategy for UK statistics – "Better Statistics, Better Decisions".

In particular, the project is closely linked with:

Objective 6 – Develop and implement innovative methods

Objective 7 – Demonstrate value for money

Objective 6 covers innovation and exploitation of new methods and data sources. Recent legislation has been approved which allows the data sharing between ONS and HMRC. Therefore this new data source and method to ONS has the capacity to illustrate the benefits of cross-working between departments and avoid unnecessary duplication.

Objective 7 identifies how Government can make better and more efficient use of public resources. HMRC turnover data could be used to supplement or replace our data sources. This has the potential to save money for ONS, but also has the benefit of reducing the burden on businesses that have completed a VAT return, no longer having to complete a turnover-based questionnaire from us. (see section 3.1 – Potential savings)

2   Data suitability of HMRC turnover data

2.1   Main messages

The coverage of HMRC turnover data is at least 99% based on turnover and 96% based on reporting units at the point of five months after the reference period.

Analysis of 2014 HMRC turnover data shows that on average the availability of data at each reference point from the reference period:

Time from reference VAT period (t)	Availability of HMRC turnover data available based on the latest monthly snapshot (July 2015)
t+1	0.1%
t+2	38.9%
t+3	70.3%
t+4	93.6%
t+5	100.6%

N.B. "t" equates to the VAT period reference month. The data are delivered on the first Tuesday of the month for the previous month. For example, January 2014 (t) data were delivered on Tuesday 4 February 2014. The February 2014 snapshot (t+1) can then be produced.

2.2   Timeliness of the data

One important characteristic of any data source is its timeliness. This is especially important for a short-term output indicator perspective, as these are published to tight deadlines as demanded by our domestic users and as specified by the EU in the Short Term Statistics regulation (STS).

The UK has a legal obligation under the STS regulation to transmit turnover data shortly after the reference period end. For the production and services industries the UK must supply turnover data 2 months after the reference period month (t+2). For retail industries the UK must supply turnover data to a shorter deadline of 1 month after the reference period month (t+1).

The UK also has a legal obligation to supply output data to the EU by a shorter deadline for the production and construction industries. These must be transmitted 1 month and 10 days after the reference period for the production industries and 1 month and 15 days after the reference period for construction industries. Work is also underway by the EU to make it a legal requirement for the flash estimate of GDP to publish at t+29 days after the reference period. The deadlines for these transmissions are set by legislation. As a consequence, the timeliness of HMRC turnover data is a critical consideration for the current feasibility study.

Monthly snapshots using HMRC turnover data have been taken by the business area since January 2014 and we are therefore able to see how the data has evolved over time.

Some important points to note with the analysis are:

- For any one snapshot only the latest 25 months remain open to revision. This has been set to mirror the revision policy of short-term output indicators which fall in line with the National Accounts revision policy. Short-term output indicators have the latest 13 months open for revision. In order for the cleaning methodology to work, the previous 12 months are also required and are subject to minimal revisions, hence

the requirement for 25 months to be open for revisions at each monthly snapshot of the data.

- 100% is taken as the value of the data at the July 2015 snapshot that is, the latest available snapshot of data. For example January 2014 has therefore had 18 revisions (February 2014 to July 2015), with each vintage of data since that month. It follows that the January 2014 data will therefore not be finalised until February 2016.

Figure 1 illustrates how each monthly snapshot for 2014 has developed in terms of data coverage based on HMRC turnover.

Figure 1: Timeliness of HMRC turnover data from the reference period for 2014 for the whole economy
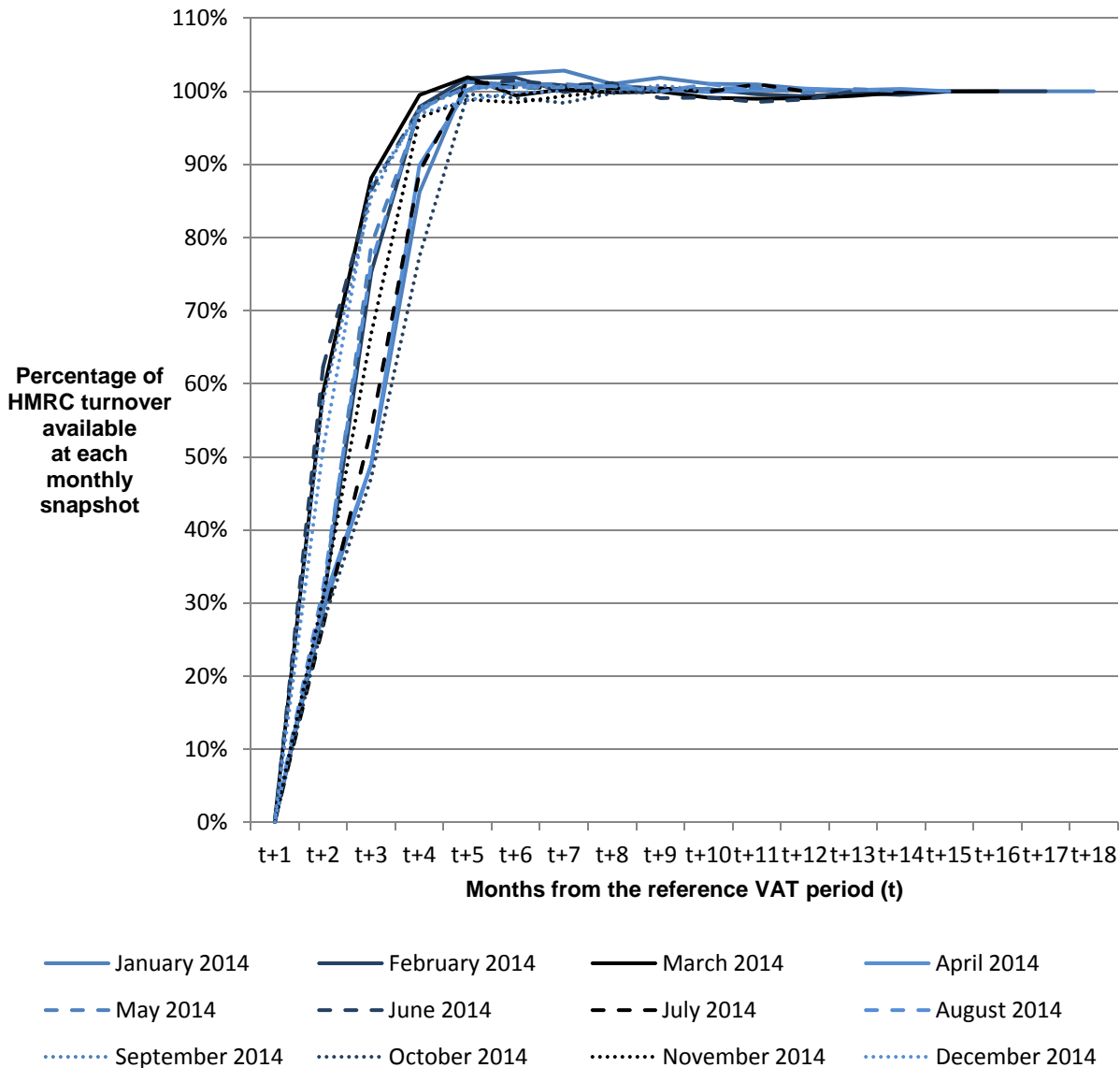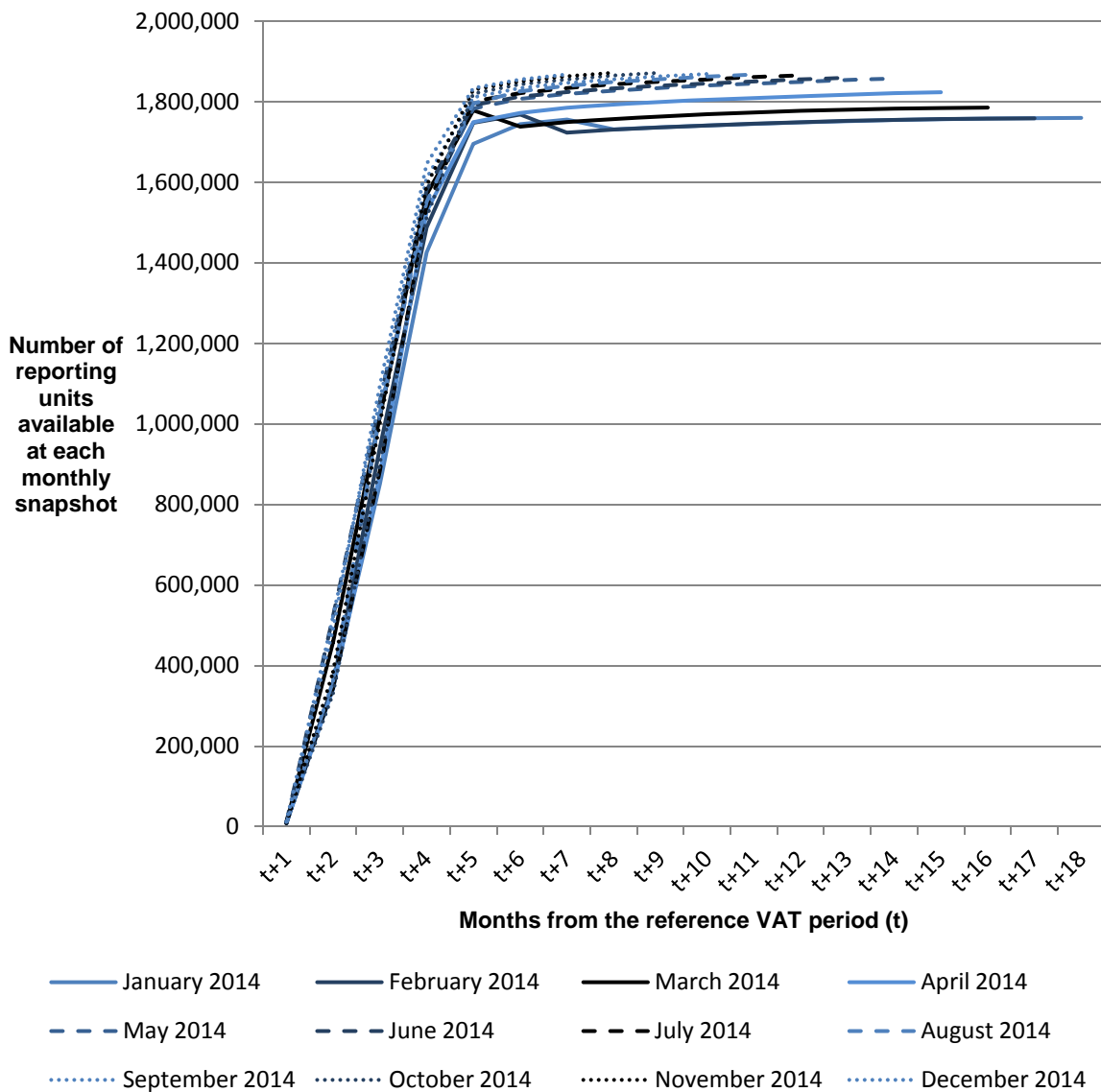

Figure 1 illustrates that once t+5 is reached, nearly all the data which are going to be received would have been collected. Even at t+4 we have on average 94% of data across the months for 2014. For example, the January 2014 data are nearly finalised in terms of turnover coverage by the June 2014 snapshot. (The June 2014 snapshot was delivered to the IDBR on 1 July 2014.)

From a short-term output indicator perspective it is also useful to look at data availability at t+1, t+2 and t+3 timeframes. A small amount of HMRC turnover data are available at t+1 so

these could not solely be used in the preliminary estimate of GDP, meaning our turnover data would still need to be collected to meet user demand for estimates in the immediate short-term. At t+2 and t+3, while the data coverage has a significant element of the data has yet to be returned, there is still a considerable number of reporting units which have an available VAT return.

Figure 2 illustrates the number of reporting units that have HMRC turnover data at each monthly snapshot from its reference point. For example, at t+2, even though on average across 2014 the data availability is just under 40% based on turnover, this equates to HMRC turnover data for just over 400,000 businesses. This compares favourably with the survey data from MBS which covers 32,000 businesses in the production and services industries. An aim of the project will be to understand whether the businesses which have responded at t+2 and t+3 are representative of businesses which are yet to respond. If so, it may be possible to use these early VAT returns.

Figure 2: Timeliness of HMRC data from the reference period for 2014 for the whole economy based on number of reporting units

2.3   Main benefits from previous attempts to use HMRC turnover data

This rich and detailed data source has previously been explored as a potential source to assist in the production of short-term output indicators.

For example the National Statistics Quality Review in Short-Term Output Indicators (2000) highlighted that VAT data should be used to validate estimates within the short-term output indicators (see Task 10 – Validating the estimates).

The Review of Statistics for Economic Policymaking (2003) conducted by Christopher Allsopp also recommended that "greater access for the ONS to administrative records, including tax data, could improve both regional and national data, while offering important savings in the compliance burden from survey forms sent to businesses."

More recently the National Statistics Quality Review: National Accounts and Balance of Payments (2014) conducted by Dame Kate Barker and Art Ridgeway suggested the following development recommendations relating to the use of administrative data. (pages 63–65):

> Development Recommendation 14: ONS should actively pursue access to identifiable micro data from HMRC so that that it can apply more rigorous statistical techniques to the estimation of these important variables. Access to micro data should enable improvements in sample size and estimation such as stratification by size and region of domicile should be pursued. The pursuit of such micro data will be dependent on wider discussions on appropriate access to confidential information but the strong track record of ONS with regard to guarding the confidentiality of the information it collects from businesses and persons should be considered by the relevant Government departments in assessing the risk of broader access for the ONS.

> Development Recommendation 15: It is recommended that the ONS pursue opportunities for expanded access to administrative data in general. This should include attention to the following aspects: maximum access to identifiable micro data to allow more rigorous statistical techniques and quality assurance; opportunities to have administrative systems capture additional or improved data, particularly when administrative processes are undergoing reviews. This requires the cooperation of other administrative organizations to ensure ONS are aware of these opportunities.

It is clear that this work is not a new phenomenon; however, some of the previous barriers to using HMRC turnover data have been addressed as part of the current feasibility study. Many of these issues were highlighted by an internal ONS paper published by Jones and Moore (2000). This report discussed a similar piece of work to the current feasibility study and looked at how HMRC turnover data could be used to replace ONS inquiry turnover data.

The issues that were previously barriers to implementation that have now been addressed through the current study include:

- The difference in the business unit structure between HMRC and ONS. HMRC collect data at the VAT registration unit level. This is different from ONS which collects survey data at the reporting unit level. Part of the current feasibility study has delivered the creation of reporting unit data by linking VAT registration unit data from HMRC to a snapshot of the IDBR. The method links the VAT registration unit to its corresponding IDBR enterprise. From this the method then apportions the HMRC turnover to the reporting unit based on the reporting unit employment proportion of the IDBR enterprise. Most units within the economy have a "simple" relationship. This is where the VAT registration unit is the same as the IDBR enterprise and is the same as reporting unit. For these units, no apportionment based on employment is required. Based on early findings from the feasibility study, on average across 2014, 90% of businesses in the overall economy

displayed a "simple" relationship. This means that the data from the remaining 10% of businesses are subjected to the apportionment methodology. These are described as "complex" businesses. To highlight the point that "complex" businesses are larger businesses, the 10% of "complex" businesses in terms of number of businesses accounted for 78% of the total turnover in the economy on average across 2014.

- Previous attempts to use HMRC turnover data within ONS have been at the aggregated industry level. This posed problems as the ability to interrogate and understand drivers behind movements between data points was limited. This has been rectified in the current project, because microdata are produced which can be interrogated with the ability to understand which reporting units are drivers behind any movements in the series.

- HMRC industry classification process. Previously it was identified the HMRC Standard Industry Classification coding was inconsistent to that held by ONS. This therefore had implications when trying to produce estimates at an industry level. This problem has been resolved by linking VAT registration unit to its corresponding IDBR enterprise and using the IDBR industrial classification.

- The establishment of a regular and secure monthly feed of data between HMRC and ONS to the IDBR. This has been improved with the [statutory instrument 2878 of 2011 made under s47 of the Statistics and Registration Services Act 2007](#) which was passed through parliament November 2011.

- The cleaning process. It is known that the data supplied to ONS are not cleaned by HMRC. However, as part of the European Statistical System Network (ESSNet) project on the use of administrative data a set of cleaning rules have been created which allows some cleaning of the HMRC turnover data at reporting unit level. ONS previously only had access to aggregated HMRC data and our knowledge of the data quality was limited. Now ONS have the microdata we can create statistical processes to validate them. These are currently under development so the data will not be shared with users until we are satisfied it is of sufficient quality.

3   Proposed method – to use HMRC data in combination with ONS survey data

The proposed method for the use of these data is to supplement our current survey turnover data with HMRC turnover data using the strengths of both data sources. For MBS this is the timeliness and guaranteed coverage of larger businesses. The MBS also guarantees monthly data collection and therefore a monthly path can be produced. For HMRC turnover, the advantages are the significantly larger coverage at the t+4 and t+5 time frames in terms of both turnover coverage and number of businesses, as well as reduced burden for smaller and medium sized businesses. Another advantage of the combined ONS survey and HMRC turnover approach is the ability to validate and quality assure data between the 2 data sources.

Due to the pressures of timeliness for short-term output indicators, it is envisaged that larger businesses, based on employment, will still be sampled as part of the MBS. Without these MBS data, ONS short-term indicators would be exposed to the risk that larger businesses' data would not be available in time for publication of the key outputs. However, in order to reduce the burden on businesses and reduce ONS costs, HMRC turnover data could be used to capture information for small and medium sized businesses.

The method proposed requires further work and investigation to understand how much the HMRC data could be used to reduce the MBS sample. However it does have the significant advantages of reducing costs and improving quality through better coverage if it proves successful.

3.1   Potential savings

The project has the potential to provide significant savings in the terms of costs. Whilst it is difficult to assign an exact figure in terms of the savings of the switch to a combined MBS and HMRC methodology to data collection, below tries to provide some estimates in terms of cost reductions for both ONS and businesses. It is also worth noting that a separate project has been underway within the office which explores the use electronic data collection instead of paper collection. Therefore cost savings are a combined approach of the 2 projects.

Progress with Electronic Data Collection project and the use of administrative data from HMRC could potentially deliver significant net savings for the MBS by 2020. This is estimated to be of the order of 40% in our running costs of the survey and 50% in the cost of completing the survey to business. The most significant savings to business would be for small and medium size businesses, as if the proposed method of collecting all larger businesses based on employment via survey data and a smaller sample of small and medium sized businesses' data is proved successful there would be no need for us to collect turnover estimates from the smaller businesses as this information would be sourced from HMRC.

As part of the Electronic Data Collection programme, under current plans, businesses selected for the MBS could be completing returns online, leading to significant savings, by spring 2017. The use of administrative turnover data from HMRC will allow the MBS to focus on larger businesses to satisfy user demand and EU legislative requirements for timely survey estimates (see section 2.2 – Timeliness of the data). The challenges and the potential needs to be explored more fully but it is estimated, at this stage, that a 40% net cash saving could be made to running costs by 2020 - although this would also be dependent on the ability to source new administrative data for the quarterly employment survey. The quarterly employment survey data are collected via the MBS in the third month of each quarter. If we decide to replace some of the MBS sample with HMRC turnover data, then we would require another method to collect these data.

4   Other uses for the data

Apart from the potential to use these data to quality assure, supplement and/or replace our current turnover data sources, there are also other potential uses for the dataset that are being explored as part of the feasibility study. These include:

- The regional dimension of the dataset. As the dataset has a NUTS1[1] level marker included at the reporting unit, analysis will be undertaken to see if regional breakdowns could be produced based on HMRC turnover data. This would likely involve extending the current method to apportion HMRC turnover based on local unit employment rather than reporting unit employment. This work has significant potential advantages, in particular the timeliness of regional data. Work will also be conducted to see to how this compares with current regional economic outputs.
- The use of purchases data. In addition to turnover, HMRC also collect data on firm level purchases which are available for statistical analysis purposes. The information relates to a businesses' response to the "Total value of purchases and all other inputs excluding any VAT" field on a VAT return. It is proposed that the same methodology, as used in the HMRC turnover estimates – linking from VAT

---

[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

registration unit to IDBR enterprise and then apportioning to reporting unit based on reporting unit employment – could be applied to this data source. This would have significant benefits because, not only would it provide an estimate of purchases at the reporting unit level, but it would also provide an approximate figure of intermediate consumption at the reporting unit level. From this an approximate figure of gross value added could be derived at the reporting unit. This would be beneficial for industries where the short-term assumption of fixed output to value added ratios is weakest.

5   Challenges

Whilst there are obvious potential benefits to the use of this detailed administrative dataset, it also poses challenges which need to be addressed before implementation within short-term output indicators and National Accounts. Sova et al (2013) addressed many of these challenges under the broad headings of: timeliness, reporting schedule, under coverage and data quality; many of which have already been discussed within this article.

Some further challenges specific to this feasibility study also include:

- The need to build a system to store and analyse the data. A system is required firstly to store and warehouse the data but also to have the capability to interrogate and record changes when amendments are required from month to month.
- The timetable for delivery is driven by the STS regulation and European System of Accounts (ESA2010). This is tight and has significant implications in terms of data availability of HMRC turnover by the time of publication for the short-term output indicators and for the proposed work to produce the flash estimate of GDP at t+ 29 days. At present the STS regulation is a legal requirement for all EU Member States to provide these data and could potentially have financial penalties should deliveries not be made.
- The assumption that HMRC turnover follows reporting unit employment. This is an important assumption of the methodology and fundamental to the production of reporting unit level estimates. This requires work to understand if the relationship is more robust for some industries than others.
- The development of the method to convert quarterly and annual HMRC turnover data to a monthly path. At the moment the method simply divides a quarterly return by 3 and an annual return by 12 for the corresponding months to which it relates to.
- Each month, an element of HMRC turnover cannot be matched from VAT registration unit to an IDBR enterprise. Work is required to try to reduce this and to understand if this unmatched HMRC turnover displays any bias in terms of its characteristics.

6   Next steps

Over the coming months the project will look to investigate and further understand the data suitability for use within short-term output indicators and National Accounts. This will specifically cover the following:

- Validating and understanding the data at the aggregated industry and reporting unit level. If this proves satisfactory we will share results at a disclosure checked aggregated industry level with users.
- The production of a HMRC turnover dataset apportioned by local unit employment for regional accounts purposes.
- To investigate the possibility of using the same method to produce a HMRC purchases dataset apportioned by reporting unit employment.

In terms of the next steps for the project, the work makes up part of the GDP(O) improvement programme. An update of the GDP(O) improvement programme will be published on 30 September 2015. It is planned further technical information with regards to the project will be released alongside this article.

7   Contact details

Name: John Allcoat

Email: stoi.development@ons.gsi.gov.uk

Telephone: 01633 456616

8   References

Her Majesty's Treasury (2004) Review of Statistics for Economic Policymaking, HM Treasury: London

Office for National Statistics (2000) National Statistics Quality Review (NSQR) Series Report Number 1: Review of short-term output indicators, ONS: London

Office for National Statistics (2014) National Statistics Quality Review, (NSQR) Series (2) Report No.2: Review of National Accounts and Balance of Payments, ONS, Newport

Sova, M, Teneva, M and John, M (2013): Challenges in Using Administration VAT Data in the estimation of Short-term Output Growth in the UK. Paper presented at the "2013 International Methodology Symposium", Ottawa, Canada, 16 October 2013

Stephens, M (2013) "GDP Output Improvement Report" Office for National Statistics

Stephens, M and Doody R (2014) "GDPO Improvement Report – 2014" Office for National Statistics