On the Potential Use of Administrative VAT Data for Estimating Short-term Output Growth in the UK

Markus Gintas Šova[1], Peter John Broad[2] and Craig Brailsford Orchard[3]

[1]Office for National Statistics, UK, e-mail: markus.sova@ons.gov.uk
[2]Office for National Statistics, e-mail: peter.broad@ons.gov.uk
[3]United Kingdom Statistics Authority, e-mail: craig.orchard@statistics.gov.uk

Abstract

The use of administrative VAT data to partially replace survey data for the estimation of short-term output growth offers the prospect of reducing respondent burden. We examine the challenges involved in the UK context and compare various methods of incorporating the VAT data. We conclude that for some divisions VAT data can be used in the estimation of 1-month output growths.
Keywords: Administrative VAT data, short-term output

1 Introduction

There is a growing interest in the potential use of administrative data to partially replace surveys for the estimation of short-term statistics. One such application is the use of Value Added Tax (VAT) data in the estimation of monthly output. The attraction of VAT data is that they are available for all VAT-registered enterprises at no additional cost either to the enterprises or to the National Statistical Institute. Furthermore, protocol 3 of the UK's Code of Practice for Official Statistics (UK Statistics Authority, 2009) states that "Administrative sources should be fully exploited for statistical purposes, subject to adherence to appropriate safeguards." However, the use of VAT data in the compilation of short-term statistics presents certain challenges not encountered when using business surveys. Depending on regulations governing VAT, these may include issues such as timeliness, collection periods and data quality. This paper presents some initial research into the use of VAT data in the UK context. In the following section we discuss the challenges and ways to address them. The methods for using VAT data are described in section 3. Results are presented in section 4, and we conclude with a discussion of our findings.

2 Complicating Factors

In the UK, enterprises are expected to report VAT to Her Majesty's Revenue and Customs (HMRC) according to one of 16 schedules. These are monthly (1 schedule), quarterly (3 schedules, each with a different set of starting months) and annual (12 schedules, each with a different starting month). Approximately 10% of enterprises report monthly, and 0.2% report using an annual schedule (see Orchard, 2010). Parkin (2010) tested several methods of converting the quarterly VAT data into monthly series, but none were found to be superior to simply apportioning the quarterly figures equally into each month covered by the quarter. We have not explored using VAT data reported to an annual schedule because the suitability

of annual data to produce useful monthly series for short-term statistics is highly questionable, and because so few enterprises report to an annual schedule.

HMRC receives 100% of VAT returns within 188 days of the end of the reference period. Returns accounting for 40% of turnover are received within 30 days and 94% of turnover within 40 days of the end of the reference period. Parkin (2010) tested various methods of forecasting mature VAT data and recommends the Holt-Winters method. We shall compare this with using mature VAT data directly.

For VAT data to be used they first need to be matched to reporting units (RUs) on the business register. Most RUs are whole enterprises, but some large multi-site enterprises are split into several RUs for statistical purposes. Orchard (2010) describes how VAT data has been apportioned to RUs for such enterprises.

Finally, VAT data can suffer from data quality issues. Lewis (2012) describes how the VAT data has been cleaned.

3 Methods of Incorporating VAT Data

A selection of methods were used to incorporate cleaned VAT data into the UK's Monthly Inquiry into the Distribution and Services Sector (now superseded by the Monthly Business Survey) covering 2005-2008. Their performances were evaluated for NACE Rev.1.1 divisions. Each method keeps the sample *as-is* for larger enterprises and for complex enterprises. The methods are described below.

Method 1 (replacement):
This method assumes that enterprises below a specified employment threshold are not sampled. The cleaned VAT data is used directly for this part of the population. This method was applied using mature VAT data (referring to 6 months and 12 months before the survey period) and forecast VAT data. For the mature data the threshold used was the lower limit of the fourth employment band (usually 100 employees). For the forecast data thresholds of 100 and 250 employees were used.

Method 2 (rescaling from larger enterprises):
This method is similar to method 1, but the cleaned VAT data are first rescaled by a factor equal to the ratio of the survey estimate for larger enterprises to the VAT total (mature or forecast, as appropriate) for larger enterprises. The ratio was calculated for each combination of month and NACE division. The same thresholds were used as for method 1.

Method 3A (rescaling from reduced sample)
This method assumes that the sample size for enterprises below the employment threshold is reduced, but not to zero. The sample reduction was carried out by stratum and tested on sample sizes of 5%, 10%, 25%, 50% and 75% of the existing stratum sample sizes, subject to each stratum retaining a minimum sample size of 5. This allows the cleaned VAT data to be rescaled by a factor comparing the reduced sample turnovers to the VAT data. The factors were calculated in 3 ways: the median of turnover ratios; the ratio of turnover totals; and the trimmed mean of turnover ratios. These factors were calculated for each combination of stratum (for those strata below the employment threshold) and month. This method was only applied to mature VAT data.

Method 3B (alternative rescaling from reduced sample)
Method 3B is the same as method 3A, except that the cleaned mature VAT data is multiplied by two factors. The first factor compares the reduced sample turnovers from the mature data reference period to the mature data (using either the median of turnover ratios, the ratio of turnover totals or the trimmed mean of turnover ratios). The second factor is the ratio of the reduced sample turnover estimate for the survey period to the reduced sample turnover estimate for the mature data reference period.

4 Results

The motivation for this work has been to find methods (or ideally a single method) of using VAT data to at least partially replace a survey of short-term output, specifically 1-month growth. We shall therefore focus on where the methods have performed well (where they have performed badly it would be clearly inappropriate to implement them). We arbitrarily define performing well as the method producing a root mean square difference (RMSD) of less than 5.0 percentage points for 1-month growths.

Table 1: Root mean square differences (in percentage points) of 1-month growths by method and division where 5.0 or less. The 3 rows under the method 3A columns refer to the factor calculation methods (the median of turnover ratios, the ratio of turnover totals, and the trimmed mean of turnover ratios, respectively).

NACE Rev.1.1 Division	Method 1 (12 month old mature data)	Method 1 (6 month old mature data)	Method 1 (forecast data, employment threshold 250)	Method 1 (forecast data, employment threshold 100)	Method 3A (12 & 6 month old mature data, 5% sample retained)		Method 3A (12 & 6 month old mature data, 10% sample retained)		Method 3A (12 & 6 month old mature data, 25% sample retained)		Method 3A (12 & 6 month old mature data, 50% sample retained)		Method 3A (12 & 6 month old mature data, 75% sample retained)	
50	4.9	4.6		3.3	3.1	3.4	2.6	2.7	**2.1**	**2.1**	**1.9**	**1.8**	**1.8**	**1.7**
											3.9	4.3	3.2	3.3
								4.0		2.6	2.8	**2.0**	**2.0**	**1.9**
51	2.4	2.4			**1.8**	**2.0**	**1.8**	**1.9**	**1.7**	**1.8**	**1.6**	**1.7**	**1.6**	**1.7**
													4.5	4.9
						2.6		**2.4**		**2.1**		**1.9**	4.7	**1.8**
55			4.0	3.6										
60	3.8	3.8		4.8	4.9		4.0	4.5	3.3	3.5	2.9	3.0	2.7	2.8
													4.3	4.2
												3.3		2.9
63	3.0	3.8	3.4	3.2	4.7	5.0	4.4	4.7	3.6	3.8	3.2	3.3	3.0	3.1
													3.9	4.3
												4.4		3.6
64	**1.8**	**2.4**	**1.5**	**1.3**	4.6		3.8		2.7	4.3	**2.3**	3.3	**2.1**	2.9
														4.8
71	4.6	4.5	4.5	3.7										
													5.0	4.8
72	4.0	4.0						4.5	3.3	3.7	2.7	3.0	**2.5**	2.7
										5.0		3.4		2.8
73	3.4	3.0	4.7	3.8										
													5.0	
74														
80	**2.4**	**2.4**	**2.5**						4.7		2.9	3.1	**2.4**	**2.4**
85	3.0	**2.1**												
90	2.7	2.6									3.2	3.1	2.7	**2.4**
											4.2	4.5	3.0	3.2
													3.4	4.1
92	4.6	4.7	4.4	3.9										
93											4.6	4.9	3.9	3.9

Based on the definition of performing well, method 1 outperforms method 2 for all divisions using forecast data and 6-month old mature data, and for nearly all divisions for 12-month old mature data. Method 3B performs poorly nearly everywhere. Table 1 shows the RMSDs for the other methods where they are 5.0 percentage points or less, with RMSDs of 2.5 percentage points or less are shaded. It is immediately clear that no single method is best for all divisions, although method 1 appears to do well for most. For method 3A, the median of turnover ratios generally provides the best factor for rescaling the VAT data. No method has been found that works well for division 74.

5 Discussion

This project has demonstrated that VAT data can be used to partially replace surveys for measuring short-term (1-month) output growth for some divisions. However, the most appropriate method to use varies by division and for at least one division no suitable method has been identified. Broad (2012) extends this analysis to cover 12-month output growths and output levels, which any recommendations for implementation will need to take into account. Our arbitrary definition of good performance, a root mean squared difference of no more than 5.0 percentage points, may at first sight seem large, but it should be regarded in the context of the standard errors of the existing survey at division level.

Once recommendations are made as to which methods (if any) to use for which divisions, the impact at the top level of aggregation will need to be evaluated. Further research should be based on NACE Rev.2 and be extended to include the production industries. Finally, it is thought that there is some scope for a more refined apportionment of quarterly VAT returns into months.

Acknowledgements

The authors are grateful to ONS colleagues Ann Langford and Mila Teneva for their important contributions to this work. Part of this work was carried out under work package 4 of the ESSNet project "The Use of Administrative and Accounts Data for Business Statistics."

References

Broad, P.J. (2012). *Use of administrative VAT data in the Monthly Business Survey.* ONS internal paper, available from the author on request.

Lewis, D.J. (2012). *Cleaning VAT Turnover data for use with MBS mixed-source estimates.* ONS internal paper, available from the author on request.

Orchard, C.B. (2010). *File Preparation, Data Matching, and Distribution Analysis of VAT Turnover*. ONS internal paper, available from the author on request.

Parkin, N. (2010). *Interpolation and Extrapolation from Value Added Tax Returns.* ONS internal paper, available from the author on request.

UK Statistics Authority (2009). *Code of Practice for Official Statistics*, edition 1.0. UK Statistics Authority, available at:
http://www.statisticsauthority.gov.uk/assessment/code-of-practice/code-of-practice-for-official-statistics.pdf