



Office for
National Statistics
Swyddfa
Ystadegau Gwladol

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

M G Šova

P J Broad

C B Orchard

Overview

- Background
 - Issues with VAT data
 - Methods of incorporating VAT data in STS
 - Results
 - Future plans
- 

Background

Growing interest in using admin data to partially replace surveys.

Her Majesty's Revenue & Customs (HMRC) collects Value Added Tax (VAT) data from all VAT-registered enterprises.

Its use for short-term statistics (STS) could reduce respondent burden and survey costs.

But there are some issues...

Issue: Reporting Schedules

Enterprises report to HMRC according to one of 16 schedules.

Monthly (10%):

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Quarterly:

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec

Annual (0.2%):

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec

...

Issue: Timeliness

Responses received by HMRC as % of turnover:

40% of turnover within 30 days

94% of turnover within 40 days

100% of turnover within 188 days

The data then have to be compiled and transmitted.

Have to either use mature data (i.e. from the wrong period) or forecast for the survey period.

Issue: Matching to Business Register

Some large multi-site enterprises are split into several, sometimes many, reporting units on the business register.

For such enterprises, the VAT data needs to be apportioned appropriately.

Issue: Data Quality

Needs cleaning!



The Survey Used for Testing

Monthly Inquiry into the Distribution and Services Sector

Methods tested for NACE Rev. 1.1 divisions.

Each method keeps the sample *as-is* for large enterprises and complex enterprises.

Method 1: Replacement

No sampled data below an employment threshold – use forecast or mature VAT data for such strata instead.

For forecast VAT data, thresholds of 100 & 250 employment tested.

For mature VAT data (6 & 12 months before survey period) a threshold of the lower limit of employment band 4 was used (usually 100 employees).

Method 2: Rescaling from Larger Ents

As method 1 (replacement) but uses rescaled VAT data (forecast or mature).

Rescaling factor calculated for each combination of NACE division & month:

$$\frac{\text{survey estimate for larger enterprises}}{\text{VAT turnover total for larger enterprises}}$$

Method 3A: Rescaling from Reduced Samp

Reduced sample below an employment threshold – use rescaled mature VAT data for such strata instead.

Stratum sample sizes reduced to 5%, 10%, 25%, 50%, 75% (subject to minimum of 5).

3 rescaling factors tested:

- median of turnover ratios
- ratio of turnover totals
- trimmed mean of turnover ratios

Method 3B: Alt Reduced Samp Rescaling

As method 3A (rescaling from reduced sample) but applying 2 rescaling factors.

First factor relates reduced sample turnovers to the mature data in the mature data period using:

- median of turnover ratios
- ratio of turnover totals
- trimmed mean of turnover ratios

Second factor:

$$\frac{\text{survey estimate from reduced sample in survey period}}{\text{survey estimate from reduced sample in mature data period}}$$

Outline of Analysis

Focus on where methods perform well.

Use Root Mean Square Difference (RMSD) between survey estimates and estimates including VAT admin data.

Define “performing well” as $\text{RMSD} \leq 5.0\text{pp}$

First Conclusions

For 1-month growths:

- Method 1 (replacement) nearly always performs better than method 2 (rescaling from larger enterprises).
- Method 3B (alternative scaling from reduced sample) performs badly nearly everywhere.

From now on we shall ignore methods 2 & 3B.

Colour-coded Table of Results (1-m growths)

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	■	■		■	■	■	■	■	■	■	■	■
51	■	■			■	■	■	■	■	■	■	■
55			■	■								
60	■	■		■	■	■	■	■	■	■	■	■
63	■	■	■	■	■	■	■	■	■	■	■	■
64	■	■	■	■	■	■	■	■	■	■	■	■
71	■	■	■	■						■	■	■
72	■	■			■	■	■	■	■	■	■	■
73	■	■	■	■			■	■	■		■	■
74												
80	■	■	■				■	■	■	■	■	■
85	■	■										
90	■	■						■	■	■	■	■
92	■	■	■	■						■	■	■
93								■	■	■	■	■

performs well for 1-month growths

Colour-coded Table of Results (12-m growths)

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	Green	Green	Green	Green	Green	Green	Green	Green	Green
51	Green	Green			Green	Green	Green	Green	Green
55			Green	Green					
60	Green	Green		Green			Green	Green	Green
63									
64			Green	Green					
71			Green	Green					
72			Green						
73				Green					
74									
80	Green	Green	Green	Green			Green	Green	Green
85		Green							
90	Green	Green	Green				Green	Green	Green
92									
93	Green	Green							

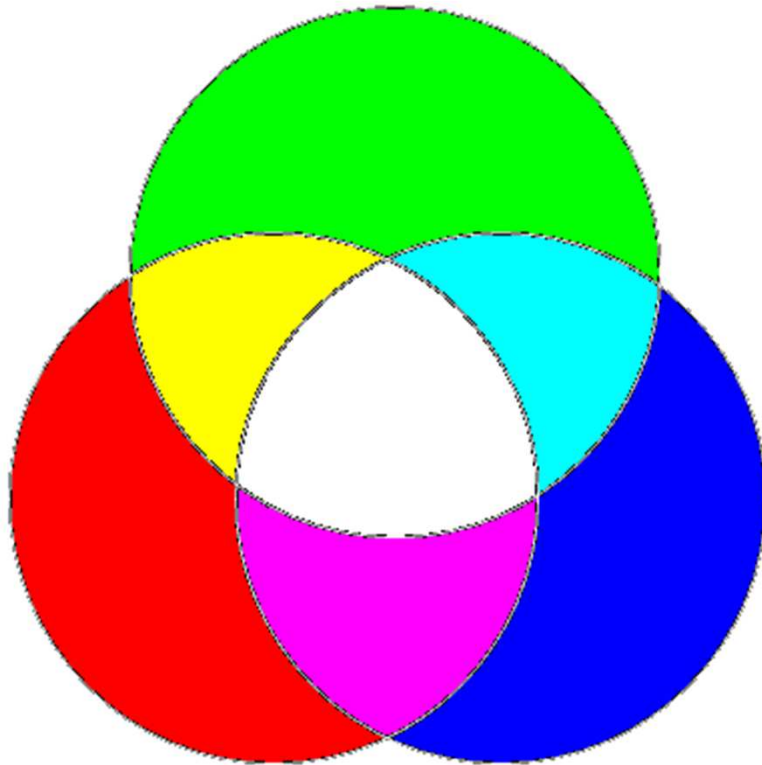
performs well for 12-month growths

Colour-coded Table of Results (levels)

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	■	■		■	■	■	■	■	■
51	■	■			■	■	■	■	■
55								■	
60	■	■			■	■	■	■	■
63	■	■			■	■	■	■	■
64							■	■	■
71	■	■			■	■	■	■	■
72	■	■			■	■	■	■	■
73	■	■						■	■
74	■						■	■	■
80					■	■	■		■
85									
90	■	■			■	■	■	■	■
92			■	■				■	
93									

performs well for levels

Combine Using Primary Colours of Light



performs well
for 12-month
growths

performs well
for levels

performs well
for 1-month
growths

Colour-coded Table of Results (all)

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			Green		Red	Red	Red	Red	Red	Cyan	Pink	Cyan	Pink
51					Pink	Red	Red	Red	Red	Red	Red	Pink	Pink
55			Cyan	Cyan					Red				
60				Cyan	Pink	Pink	Pink	Pink	Red	Red	Red	Pink	Pink
63	Pink	Pink	Blue	Blue	Pink	Pink	Pink	Pink	Red	Red	Red	Pink	Pink
64	Blue	Blue	Cyan	Cyan	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
71	Pink	Pink	Cyan	Cyan	Red	Red	Red	Red	Red	Red	Red	Pink	Pink
72	Pink	Pink	Green		Red	Red	Red	Pink	Pink	Pink	Pink	Pink	Pink
73	Pink	Pink	Blue	Cyan								Blue	Blue
74	Red				Red	Red	Red	Red	Red	Red	Red	Red	Red
80	Cyan	Cyan	Cyan	Green	Red	Red	Red	Blue	Red	Cyan	Cyan	Cyan	Cyan
85	Blue	Cyan			Red	Red	Red	Red					Red
90			Green		Red	Red	Red	Red	Red	Red	Pink	Pink	Pink
92	Blue	Blue	Pink	Pink								Red	Red
93	Green	Green							Blue	Blue	Blue	Blue	Blue

performs well for 1-month growths

performs well for 12-month growths

performs well for levels

Future Plans

- Test most promising methods in live environment.
- Extend to all industries and use NACE Rev.2.
- Revisit method for apportioning data into months.
- Seek to explain why methods perform well/badly in different industries.