

VAT part 2 - additional indicators, bias and anomalies

2018-06-13

VAT part 2

This project covers further work that the Campus does on the VAT (expenditure and turnover data). The Project will:

- explore the use of VAT returns as an early economic indicators (e.g. expenditure, births and deaths, number and timing of returns, modelling economic statistics such as profits etc) – this may include standalone indicators, inputs into improving national economic statistics, and insights into quality assurance of national economic statistics
- explore bias and quality, for example within different reporting periods, assumptions about reporting level (enterprise, group etc.), types of return, editing etc.
- develop methods for anomaly detection in the VAT data which can:
 - be used for quality assurance and economic analysis
 - be applied to other large administrative data sets *dynamic systems, network analysis, agent based models
- can we model intermediate consumption from VAT TO & EXP, and subtract to get an estimate for cap ex?

Suggested priorities for discussion 16 Nov, for stage 1 of this, to happen over the next 3 months

All of these for the reference quarter, but could compare with later revised values. Use raw data - caveat for regional, SIC, employment breakdowns

- update vindicator and publish article & data ASAP
- births & deaths - number of each, growth rate of time series, difference (B - D)
- number of returns in the reference quarter, and by (cumulative) month
- can explore by SIC, region and employment
- expenditure - as for turnover

- clustering? including all variables, maybe drop TO / exp? does this tell us anything, and does the mix of reporting types change? -> VAT reporting behaviour

And some of the areas that could be explored include: * apportionment to regions - checking summing to totals, better methodology * aggregation to quarterly time series, including exploring biases within different reporting periods * biases & relationships - reporting unit, enterprise, enterprise group etc. - errors in assumptions of reporting level? * match < 100% QOPS – why, too small to match VAT? births & deaths? timing where VAT not reported? * HMRC VAT registrations - daily delivery * other VAT info - change to staggers, type of return e.g. rebate, etc - clustering, growth / decline, revisions – how can all of these be used to better understand what’s going on in the economy, or as early indicators that something new is happening * QA – identifying *real* outliers / anomalies, identifying anomalous behaviour in groups with certain characteristics? * Can we improve survey aggregation? Particularly for volatile / badly behaved indicators like capital investment? * Can this be generalised to other large admin datasets (e.g. PAYE)?

team members

- Louisa Nolan
- Jonathan Gillard (Reader in Statistics, Cardiff Uni)
- Emily O’Riordan (PhD student, anomaly detection, Cardiff Uni)
- Luke Shaw

stakeholders

- Economic stats: Rob Kent-Smith, Andrew Sutton, Richard Heys, James Scruton, Rob Doody etc.,
- Economic Experts Working Group
- economics users
- NSIs - speak to Mark Stephens
- Martin Weale
- Duncam Elliot, methodology time series

Updates

2018-11-19T16:52:54Z Project is resourced jointly with Data scientists and Economists.