

Use of VAT administrative data in publishing short-term economic indicators

Craig McLaren and Matthew Whipple

Ceri Lewis, Rhian Murphy, Mark Stephens

Surveys and Economic Indicators, Office for National Statistics

May 2019

This talk will focus on...

- Our approaches for using VAT administrative data for producing economic statistics
- Methods we have used to deal with characteristics of the VAT data set
- Confronting differences between sample surveys and VAT

ONS short term business surveys

- Four key sectors of the economy with monthly National Statistics outputs relating to:
 - Construction; Production; Retail; Services
 - Feeds directly into GDP measures
- Compiled primarily using sample surveys, but increasingly we are supplementing with administrative data

ONS short term business surveys

- Existing sample sizes:
 - Construction (~8000); Production (~6000); Retail (~5000); Services (~27000)
- Expensive to source compared to some administrative data
- Production, Retail and Services have now moved on-line already saving ~38000 paper despatches a month
- VAT Turnover identified as an alternative data source

Gains in scale

- Since December 2017, VAT Turnover used in output estimates for smaller employment sampling sizebands of selected industries

	Production	Services	Construction	Total
Survey units replaced	1800	4000	2400	8200
VAT units	88 000	528 000	100 000	632 000
% Sector from VAT	12.4	6.4	14.3	7.7

- Helps address sample rotation volatility in smaller employment strata

Bonus! Administrative data in Construction

- Since 2017 VAT Turnover used in construction output estimates for selected industries
- Now use VAT Turnover for ~84 000 businesses to replace 2400 survey returns in monthly sample
- Helps address sample rotation volatility in smaller employment strata
- Also using a detailed commercial dataset from Barbour ABI (only for New Orders data and sub-national sectoral detail)

Sampling approach

- Monthly surveys are sampled from the Inter-Departmental Business Register (IDBR), e.g. sample frame
- We use employment as a sampling variable with five size-bands
 - Size band 1 has between 0-4 or 0-9 employment
 - Size band 2 and 3 are medium sized
 - Size band 4 has between 100+ or 150+ or 249+ employment
 - Size band 5 is for businesses in size band 2 or 3 where turnover is large

VAT data supply

- Monthly data feed from HRMC gives turnover returns at VAT enterprise level
- Businesses are legally obliged to submit a VAT return if they meet a reporting threshold (e.g. £85000 per annum for 2017/18 fiscal year)
- Important: VAT Turnover is NOT the amount of VAT. It is 'total value of sales and all other outputs excluding VAT' which is then recorded in a VAT return.

VAT data collection

- We use box 6 data from the HMRC VAT dataset

1. VAT due in this period on sales and other outputs	<input type="text"/>
2. VAT due in this period on acquisitions from other EU Member States	<input type="text"/>
3. Total VAT due (the sum of boxes 1 and 2)	<input type="text"/>
4. VAT reclaimed in this period on purchases and other inputs (including acquisitions from the EU)	<input type="text"/>
5. Net VAT to be paid to HMRC or reclaimed by you (difference between boxes 3 and 4)	<input type="text"/>
6. Total value of sales and all other outputs excluding any VAT. Include your box 8 figure	<input type="text"/>
7. Total value of purchases and all other inputs excluding any VAT. Include your box 9 figure	<input type="text"/>
8. Total value of all supplies of goods and related costs, excluding any VAT, to other EU Member States	<input type="text"/>
9. Total value of all acquisitions of goods and related costs, excluding any VAT, from other EU Member States	<input type="text"/>

Characteristics: VAT data supply

- VAT is returned on a monthly, quarterly or annual basis
- Includes all monthly, quarterly and annual VAT returns delivered in the reference month including revisions to previous periods.
- Returns can be submitted on a variety of different staggers (dates for what the return refers to)

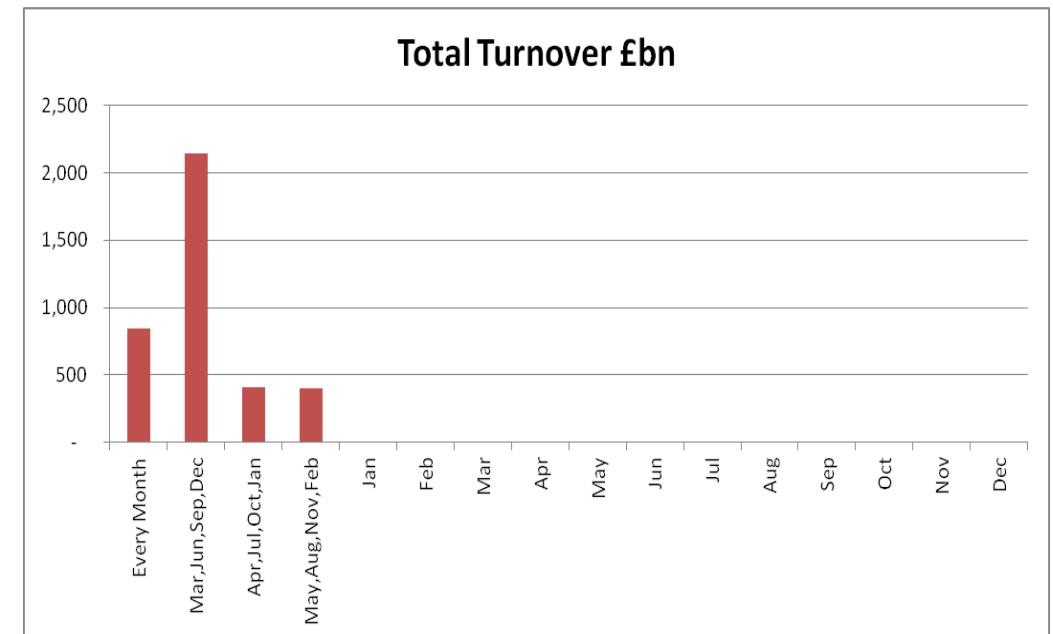
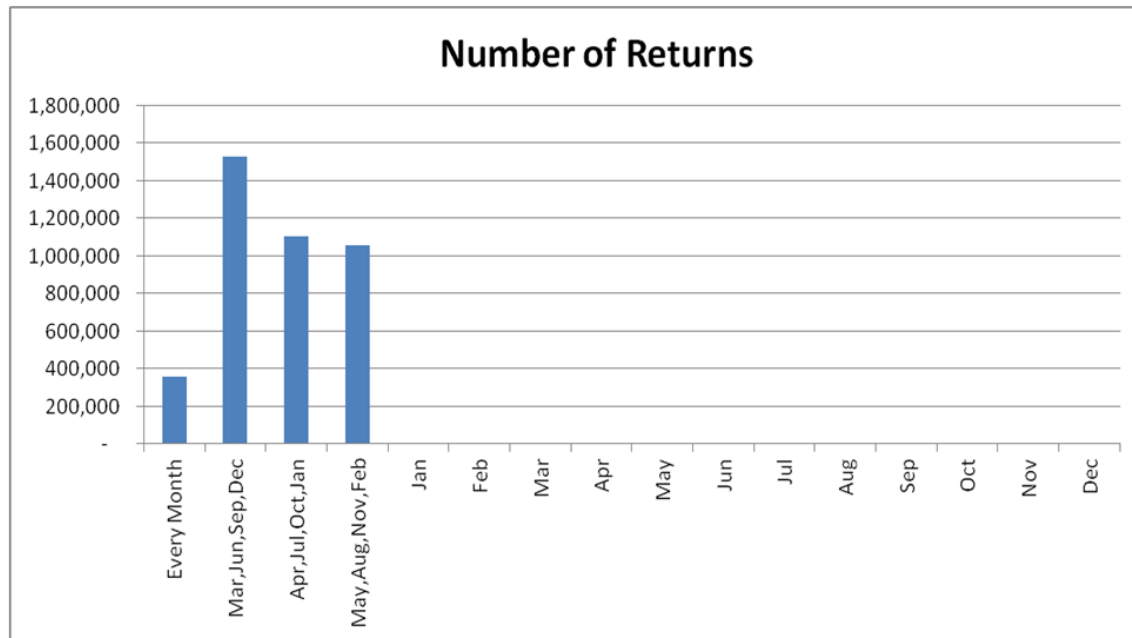
Characteristics: VAT returns

VAT returns by number and size of turnover by return frequency over a six month period.

Period-end month	Count	Count %	Turnover (£billions)	Turnover %
Monthly returns				
Every Month	358,809	9%	846	22%
Rolling 3 month returns				
Mar, Jun, Sep, Dec	1,529,868	38%	2,146	56%
Apr, Jul, Oct, Jan	1,102,561	27%	412	11%
May, Aug, Nov, Feb	1,058,678	26%	402	11%
Rolling 12 month returns				
Annual, all months	11,182	~0%	~0	~0%

Data challenges- Frequency of VAT returns

- VAT data report over a variety of staggers
- Large challenge to assign to monthly path
- Note scale and the yearly by month reporters

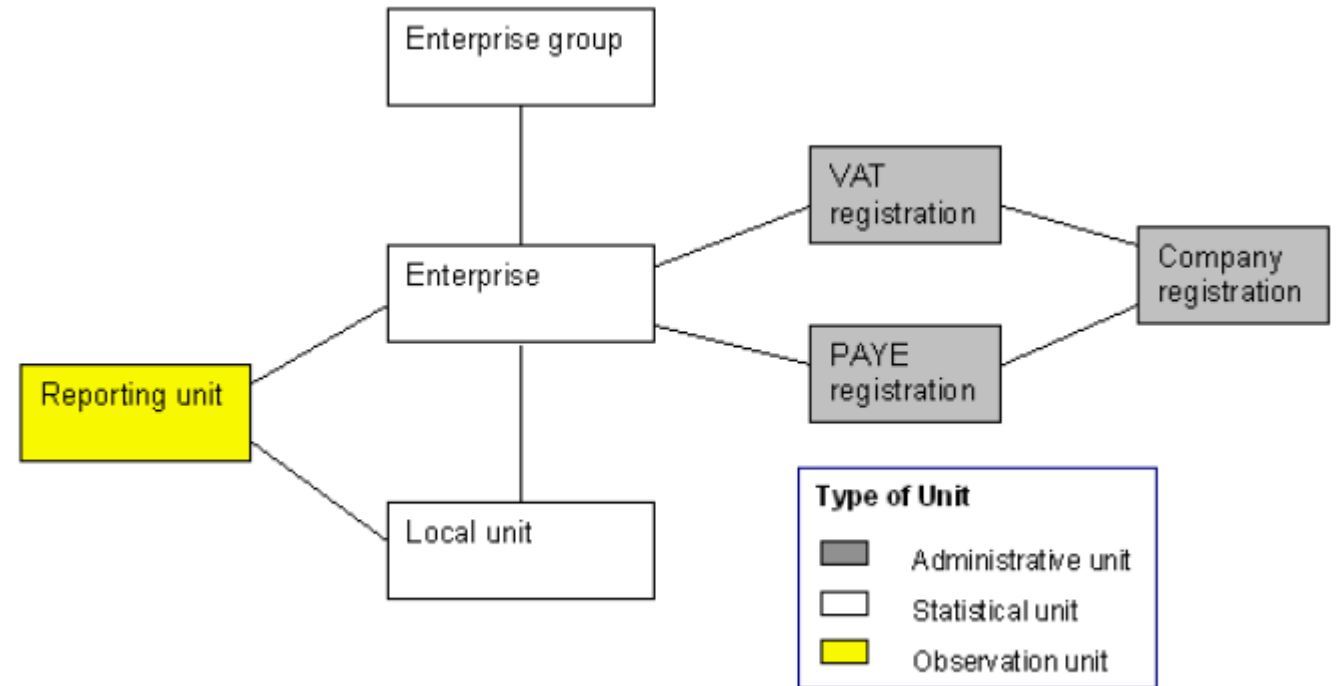


Response timeliness – data returns

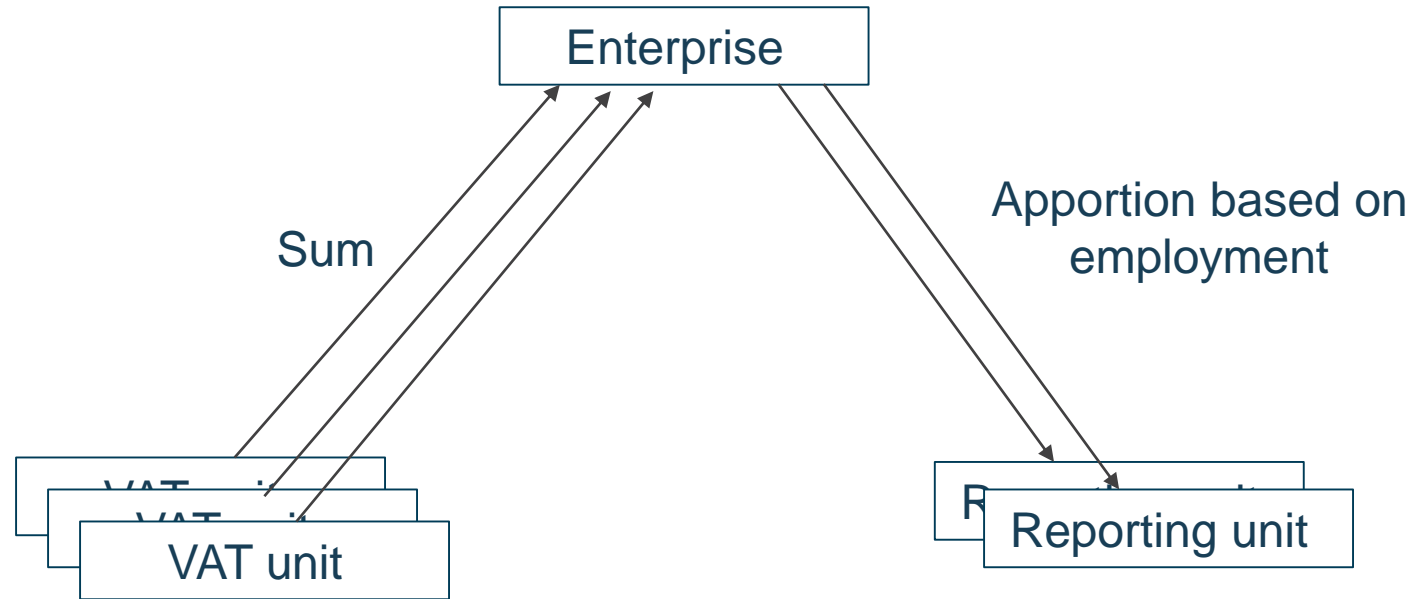
- Responses will vary over time
- Driven by the reporting requirements and deadlines, so returns can come in at particular time points

Characteristics: Units

- Non-direct reporting relationships
- VAT units are not always identical to ONS reporting units
- VAT units are matched to enterprises, and aggregated up



Dealing with complex units



How are we using VAT Turnover in practice

- A benchmark alongside survey data
- Replacement for survey data
- Replace other non-survey data

Using VAT data as a benchmark

- VAT data used to create a quarterly index (primarily supplied over a 3 month stagger) benchmarked to the business survey monthly path
- VAT data only replaced selected small and medium sized businesses for the MBS within Construction, Production and Services
- Will always retain survey data for large businesses due to
 - Quicker response
 - Potential for more business intelligence
 - Ability to query data

How we have selected industries for benchmarking

1. Fit – How VAT Turnover compares to MBS data at the corresponding level
2. Impact – What is the impact of changing data source at the sampling and industry level
3. Revisions – what is the impact on the revision performance by source change
4. Data journey processing – Administrative data can inform us about business complexity
5. Micro data – How VAT returns compare to survey returns at individual level

Current benchmark position

- Construction, Production and Services
- VAT benchmarks for selected industries used since December 2017
- Use of VAT was expanded for Blue Book 2018
- Use VAT for 261 of the 488 sampling cells in scope (53%)
- VAT from 734,000 businesses now used to benchmark 10,120 returns
- An [article](#) published in March 2018 gives more detail on the use of VAT data in National Accounts

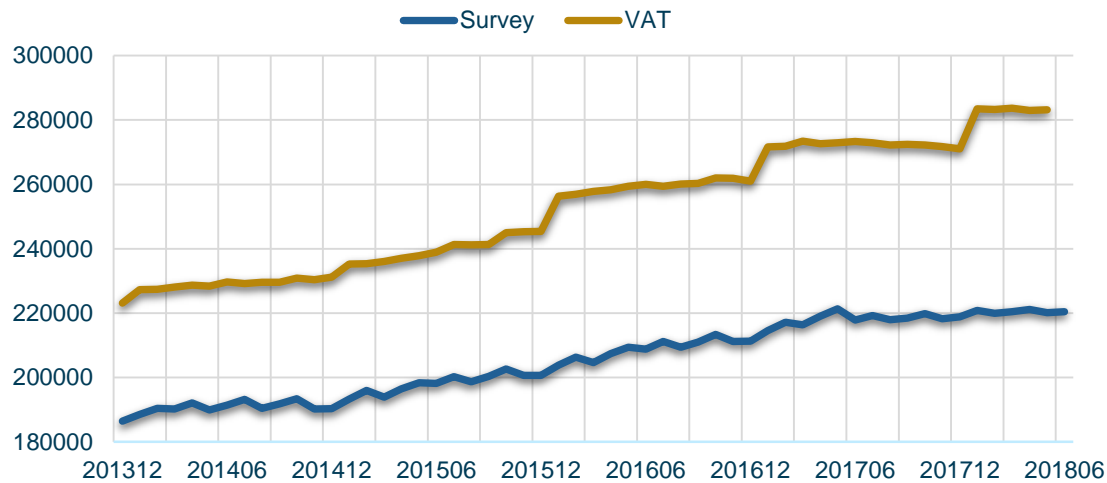
Implementing VAT changes

- New processing systems being developed
- New strategic technologies integrated into the results processing with significant improvement over existing systems. Processing times cut from 2 days to 2 hours
- New methods to process the data
 - **Matching and linking** -> 99% rates to the IDBR; **Cleaning** – ability to clean at sources; **Estimation** – now covers lags in data; **Calendarisation** – seasonal trading components

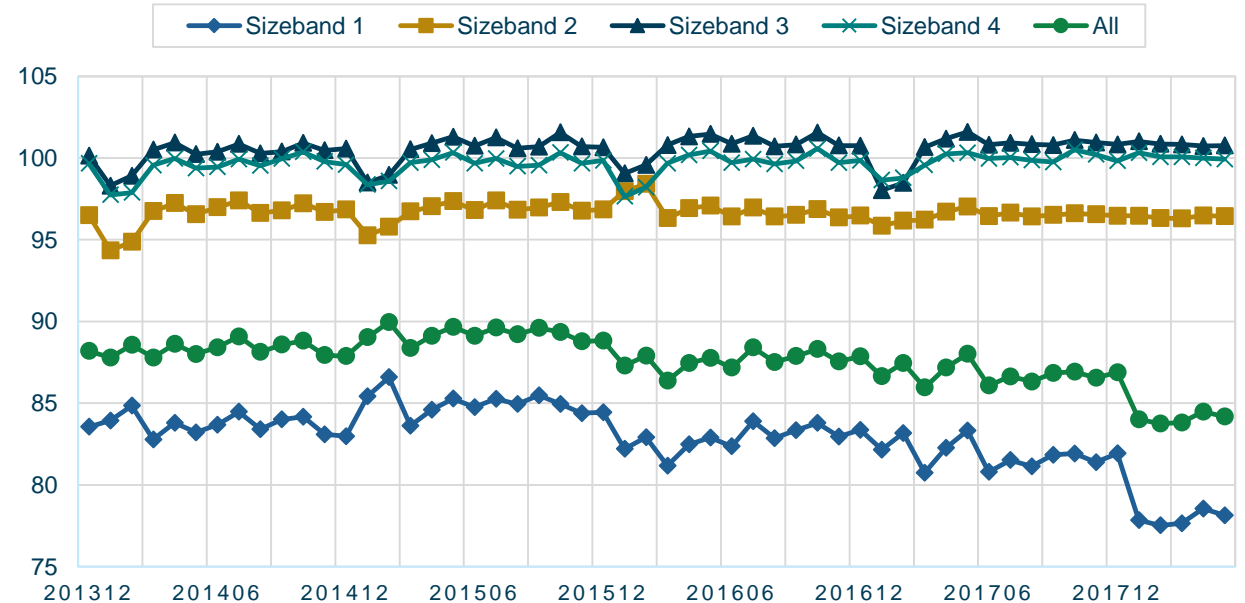
Example universe analysis

- Understanding the impact of the VAT threshold

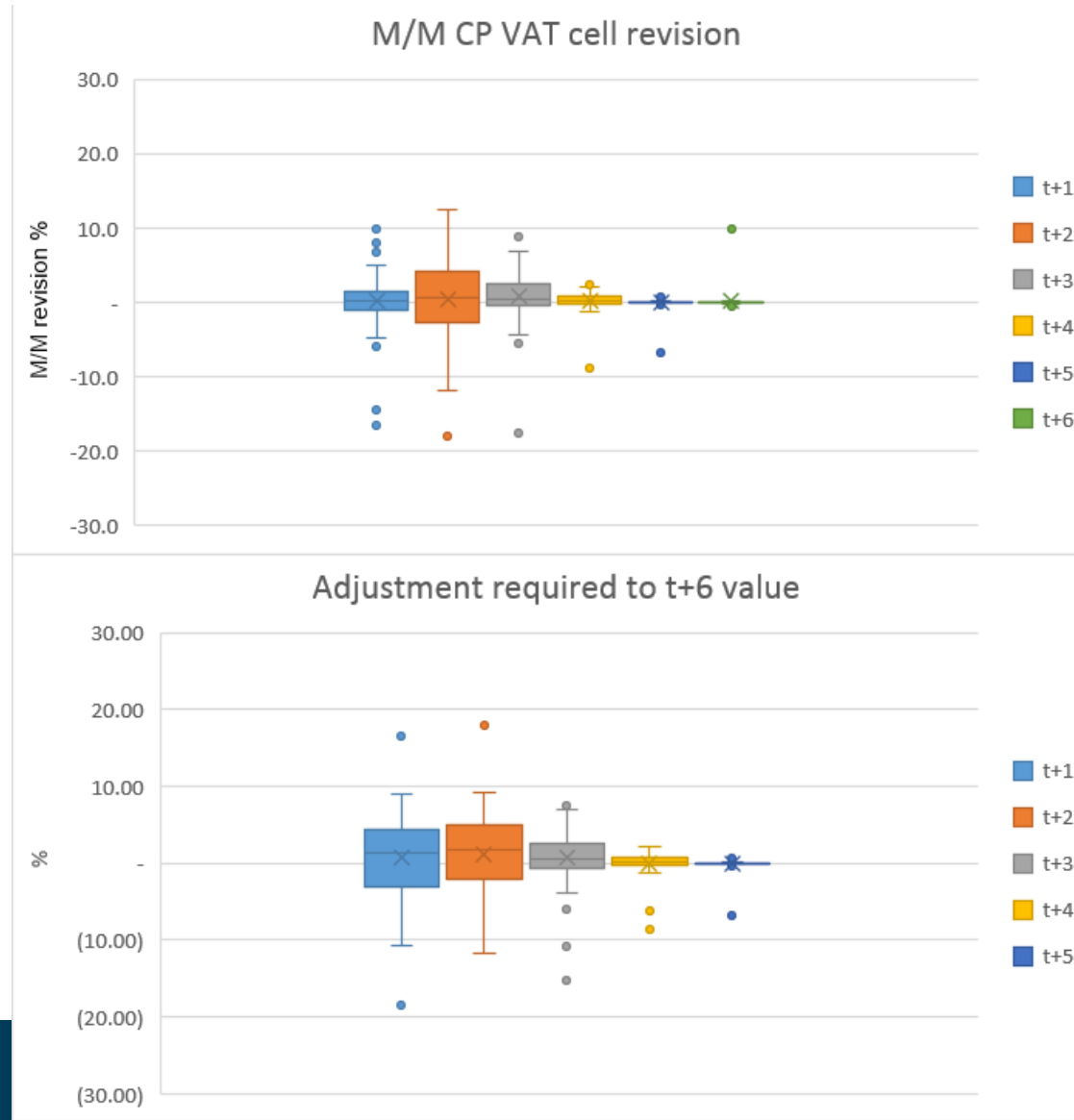
Example Universe



EXAMPLE SECTOR- VAT UNIVERSE PROPORTIONS

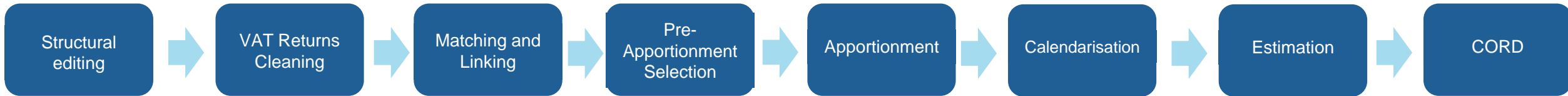


Example revisions analysis



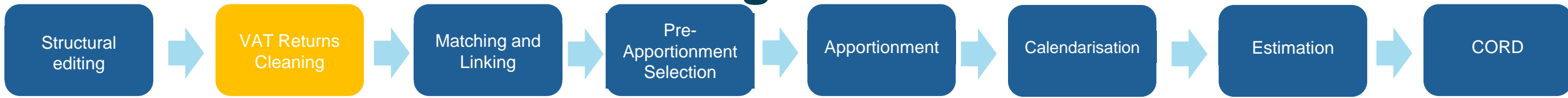
- Most recent data subject to cell level revisions
- Most recent two months can be forecast, to account for initial response
- Example sector based on 6 years of data

Strategic methods



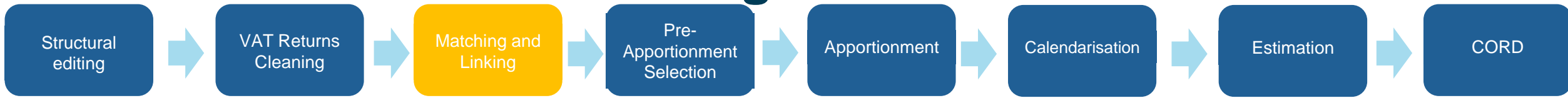
- Processing administrative data (VAT) and survey data, will require different methods; and the underlying data will likely display different properties, e.g. in terms of timeliness and methods parameters
- Consistent and reproducible methods are important

Strategic methods



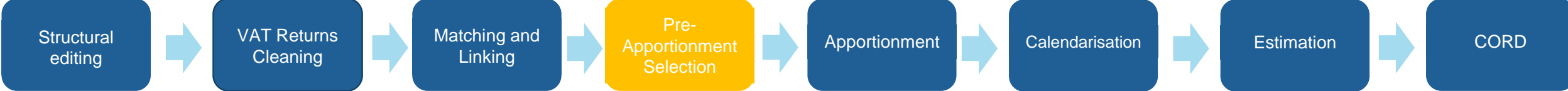
- Validation (check that row is valid)
- Duplicate processing (check if there is a duplicate row)
- Thousand pound rule
- Quarterly pattern rule
- Suspicious turnover rule
- Identify first returns
- Confirm dataset

Strategic methods

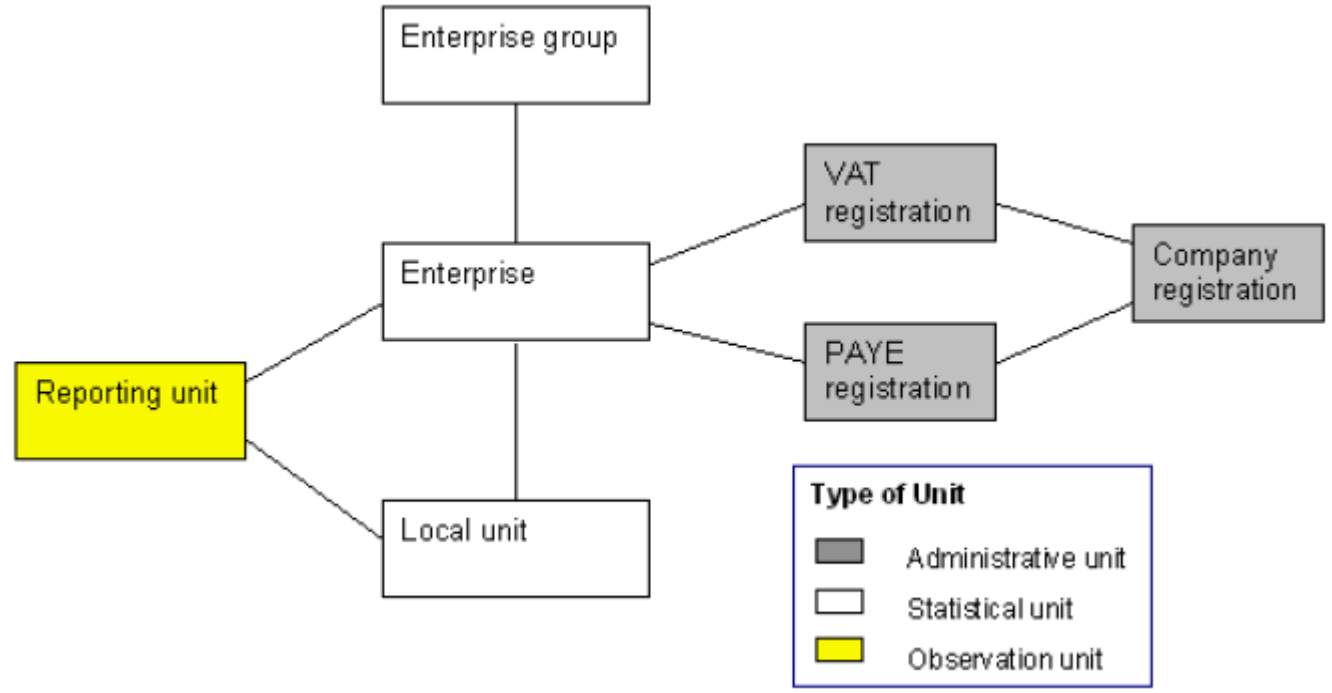


- Join with IDBR records
- < 99% Vat data can be matched

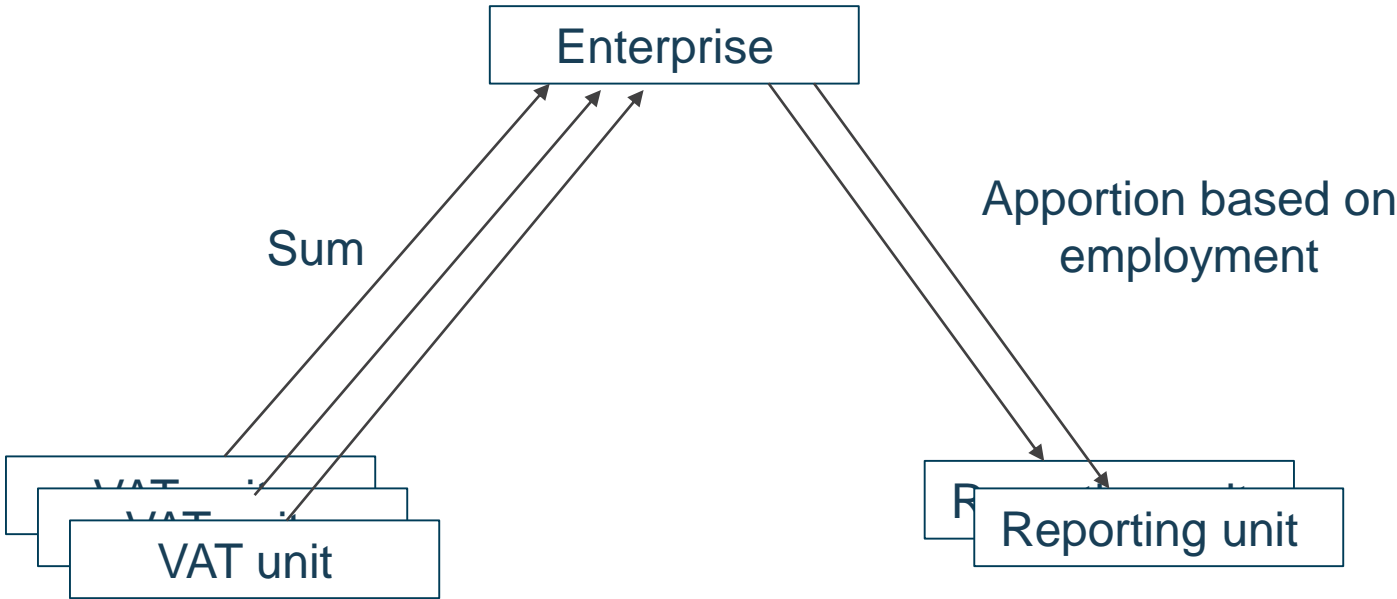
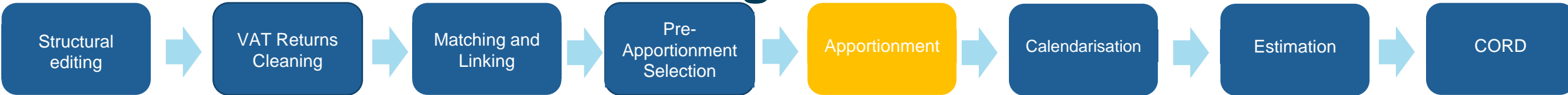
Strategic methods



- Non-direct reporting relationships
- VAT units are not always identical to ONS reporting units
- VAT units are matched to enterprises, and aggregated up

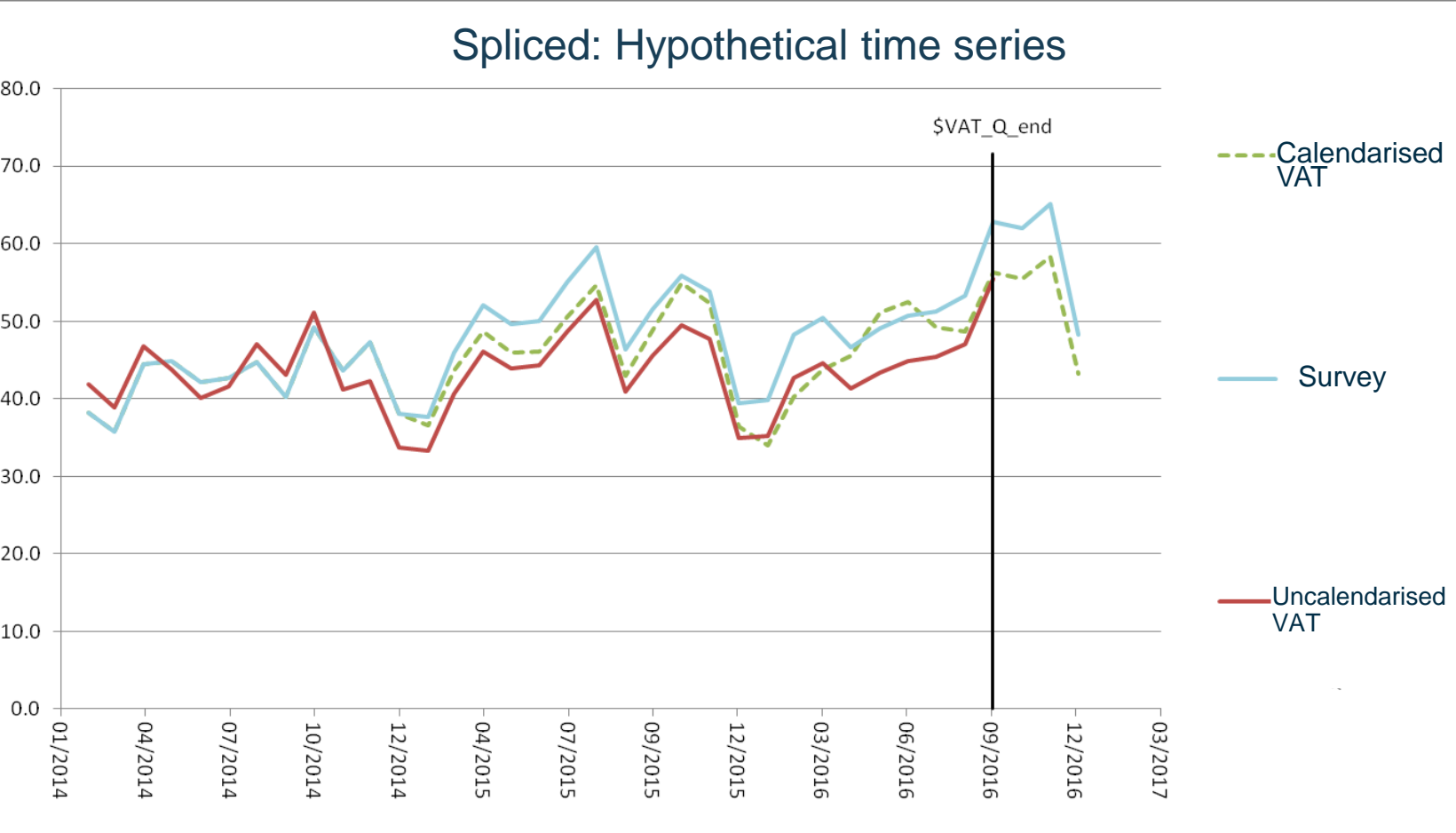
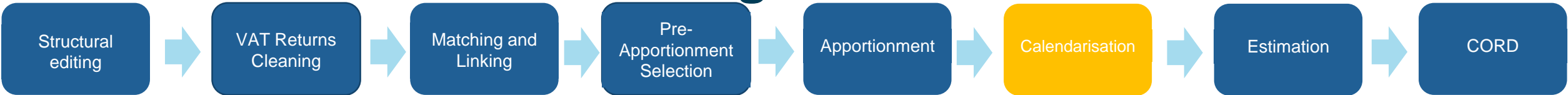


Strategic methods



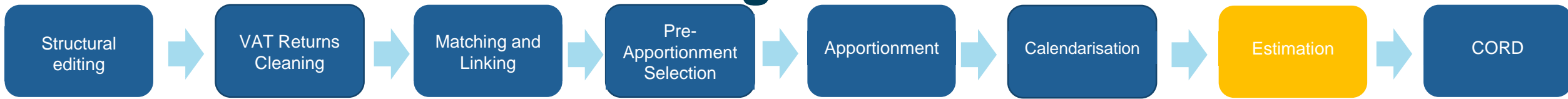
- Vat units report on different staggers
- Can lead to close assessment for the reporting unit figures based on complex units

Strategic methods



Need to consider a consistent series over time to enable comparison and analysis

Strategic methods



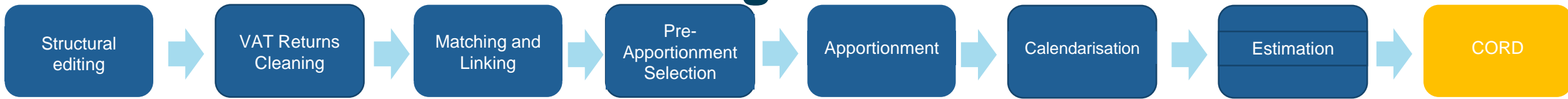
$$R_t = \frac{\sum_{impclass} y_{i,t}}{\sum_{impclass} y_{i,t-1}}$$

Ratio of means imputation, summing over matched RU's in the previous period.
Impclass defined by industry and sizeband

$$y_{i,t}^* = R_t \times y_{i,t-1}$$

y^* is the imputed value

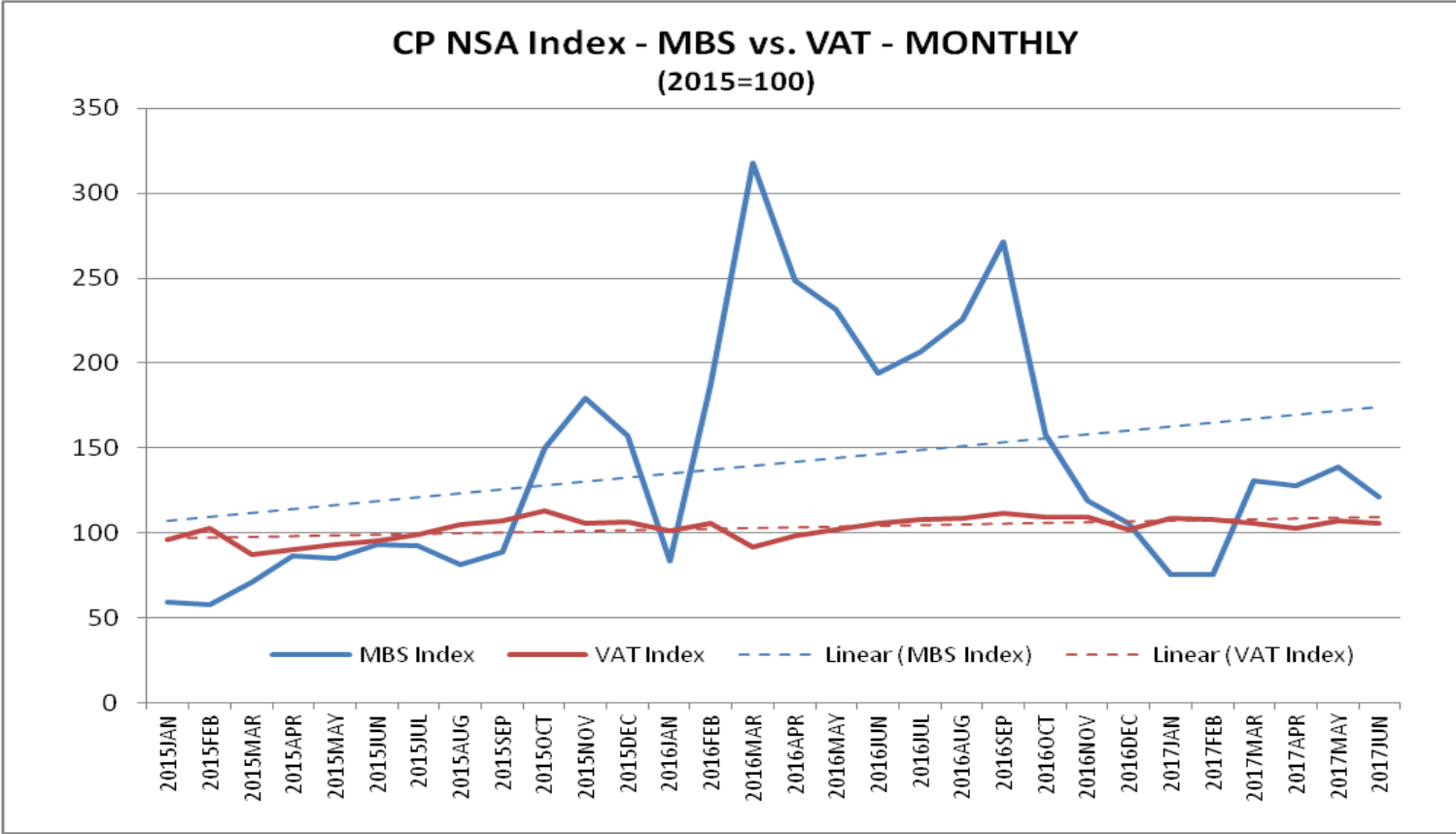
Strategic methods



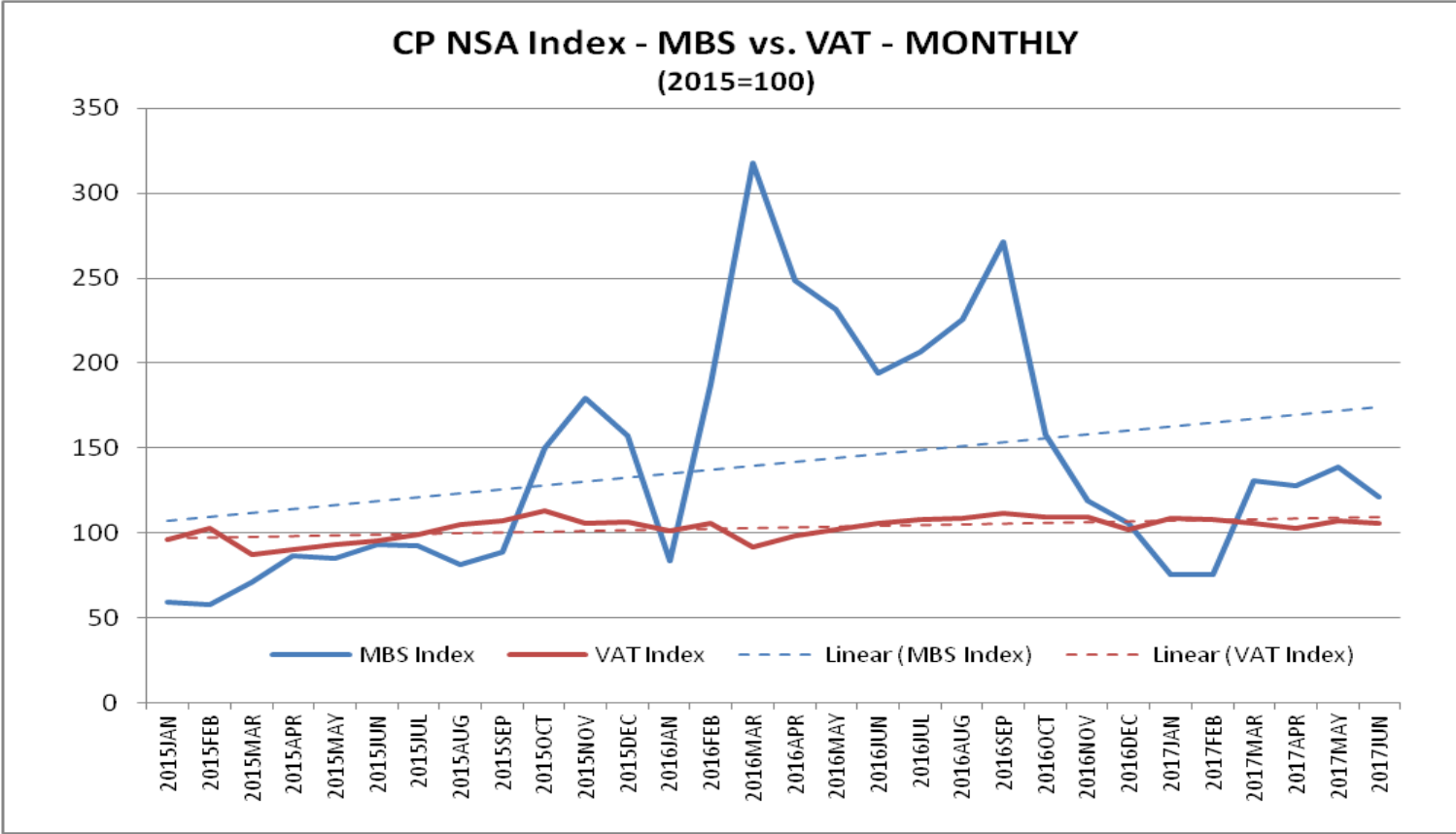
- Aggregate by industry and sizeband, to output to aggregation processing system
- Aggregated with survey data via a partition approach
- **Important: we can control where VAT data is used. Partition approach allows us to use VAT to replace surveys for small businesses, while keeping survey contact for large businesses**
- Deflate (where needed) and apply seasonal adjustment
- Publish

Examples

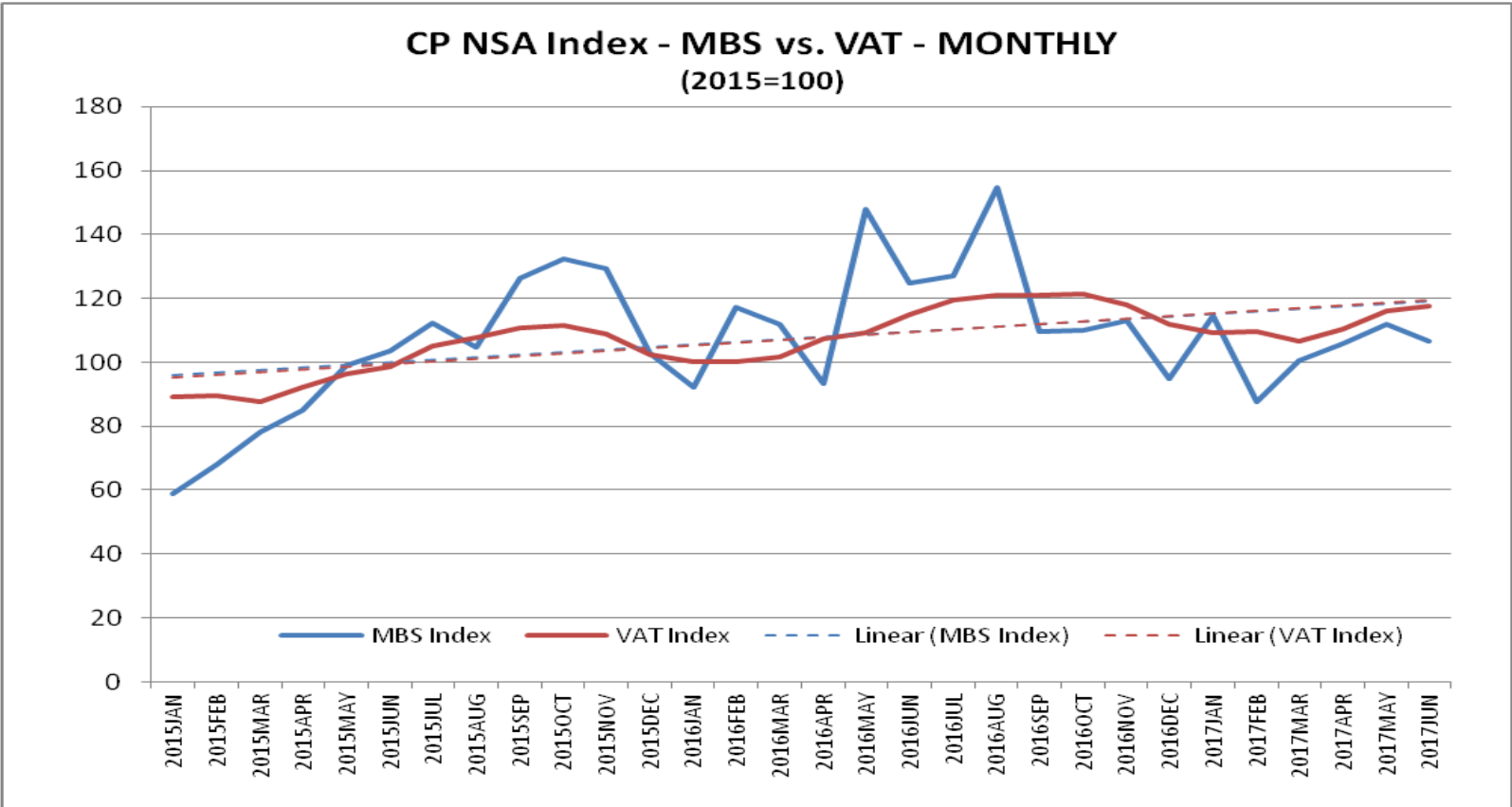
SIC07 – 43.10 – SB1 – Emp 0-4 - Demolition



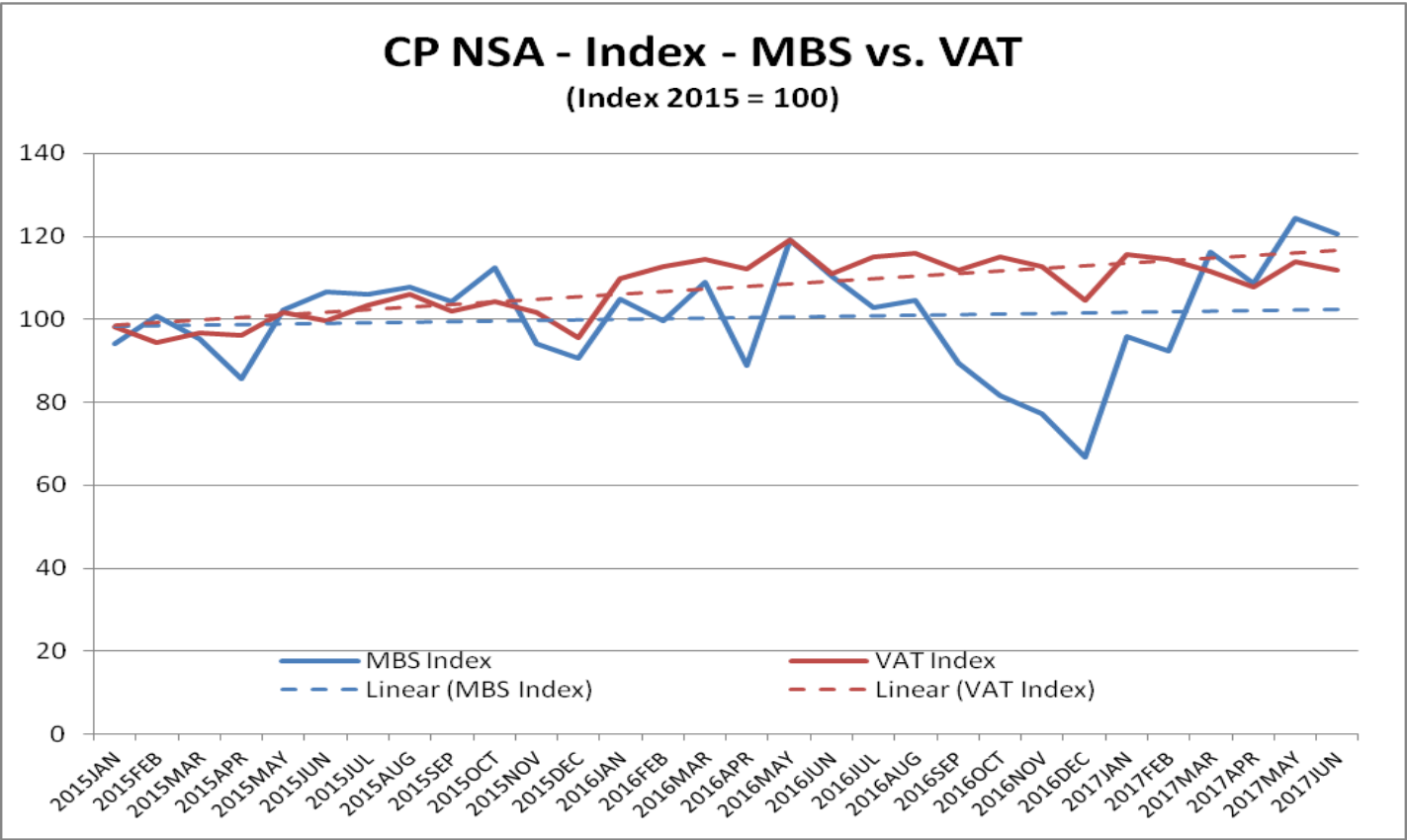
SIC07 – 43.10 – SB1 – Emp 0-4 - Demolition



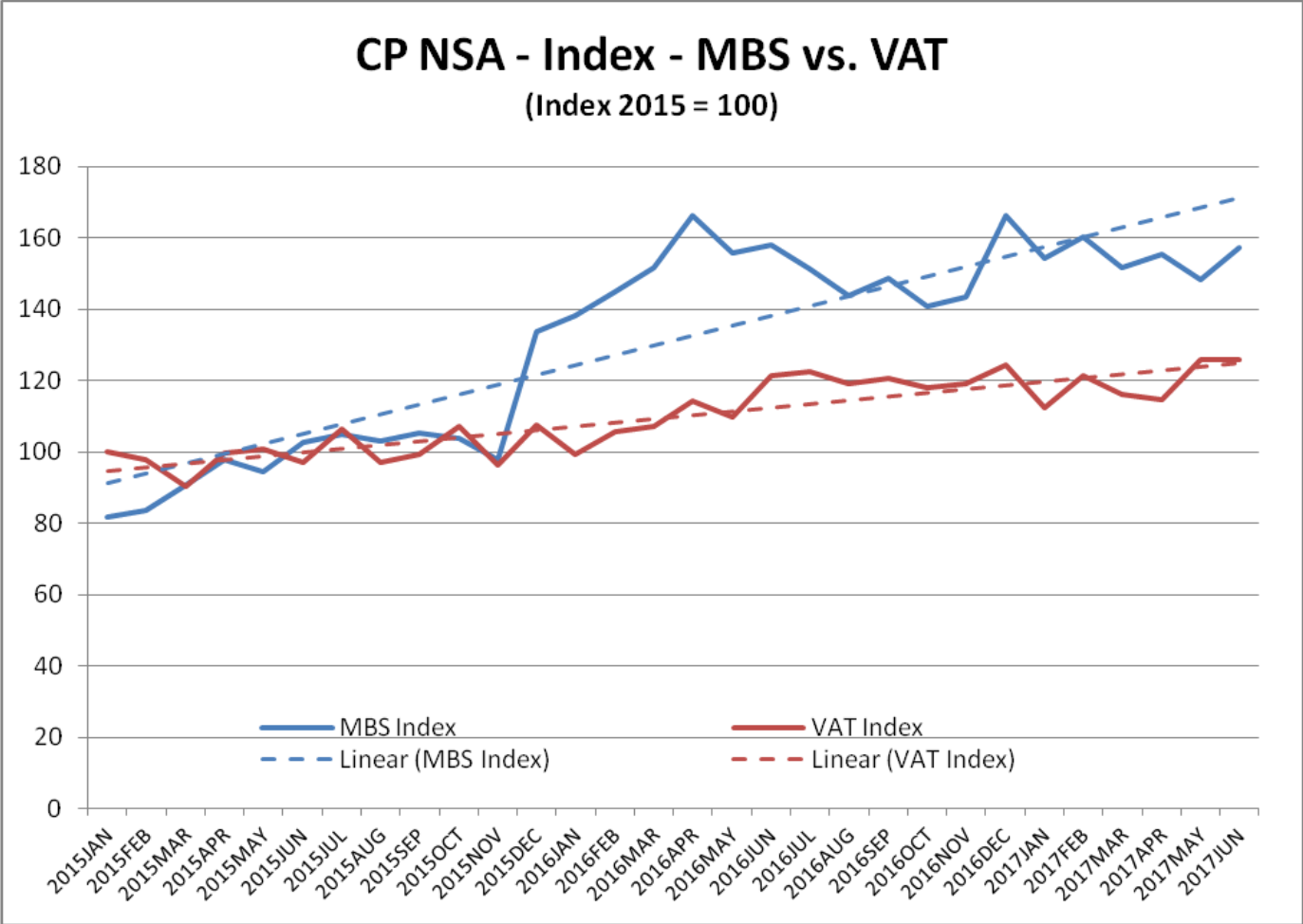
SIC07 – 43.33 – SB1 – Emp 0-4 – Floor and wall covering



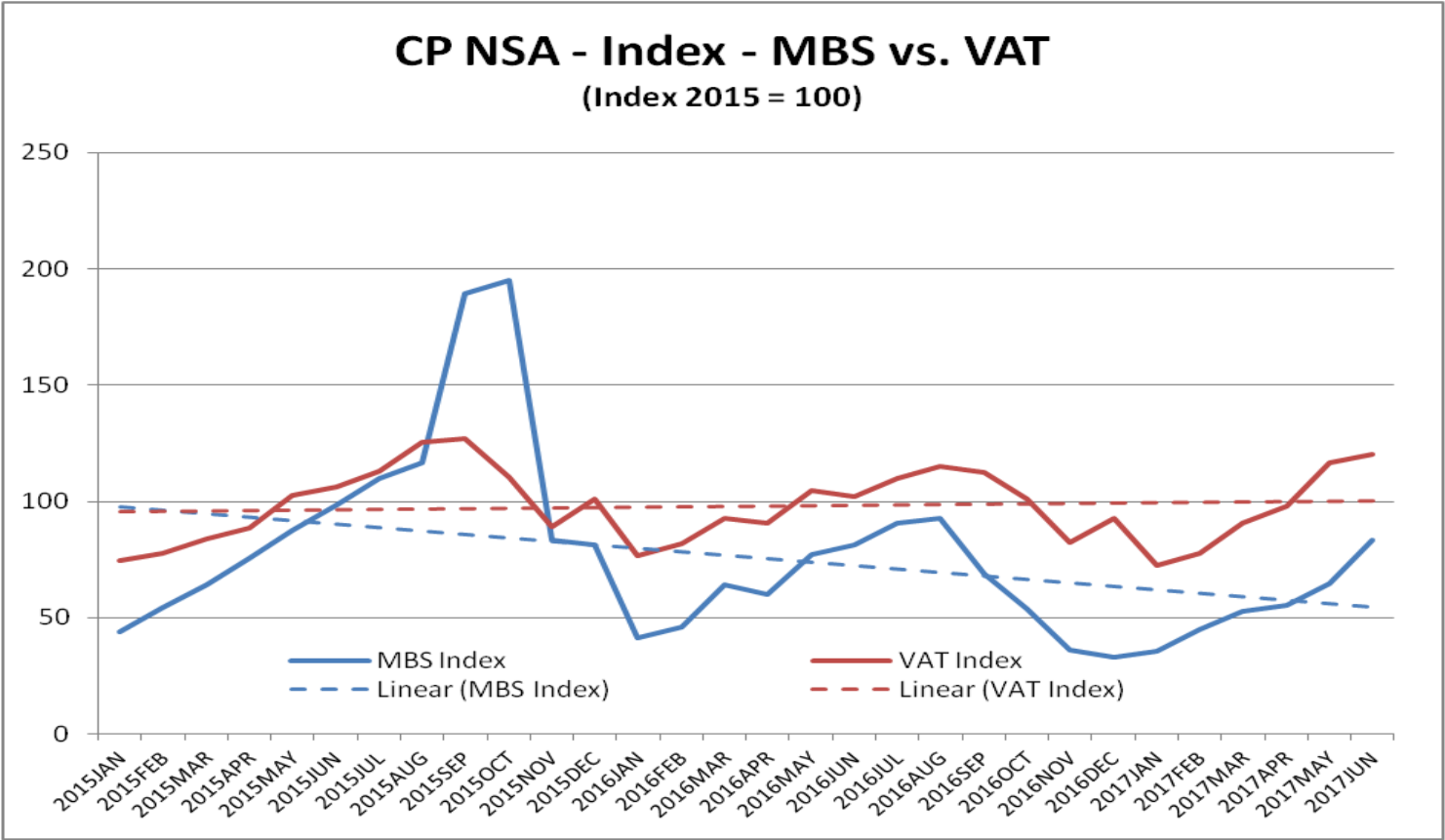
SIC07 – 75 – SB1 – Vets



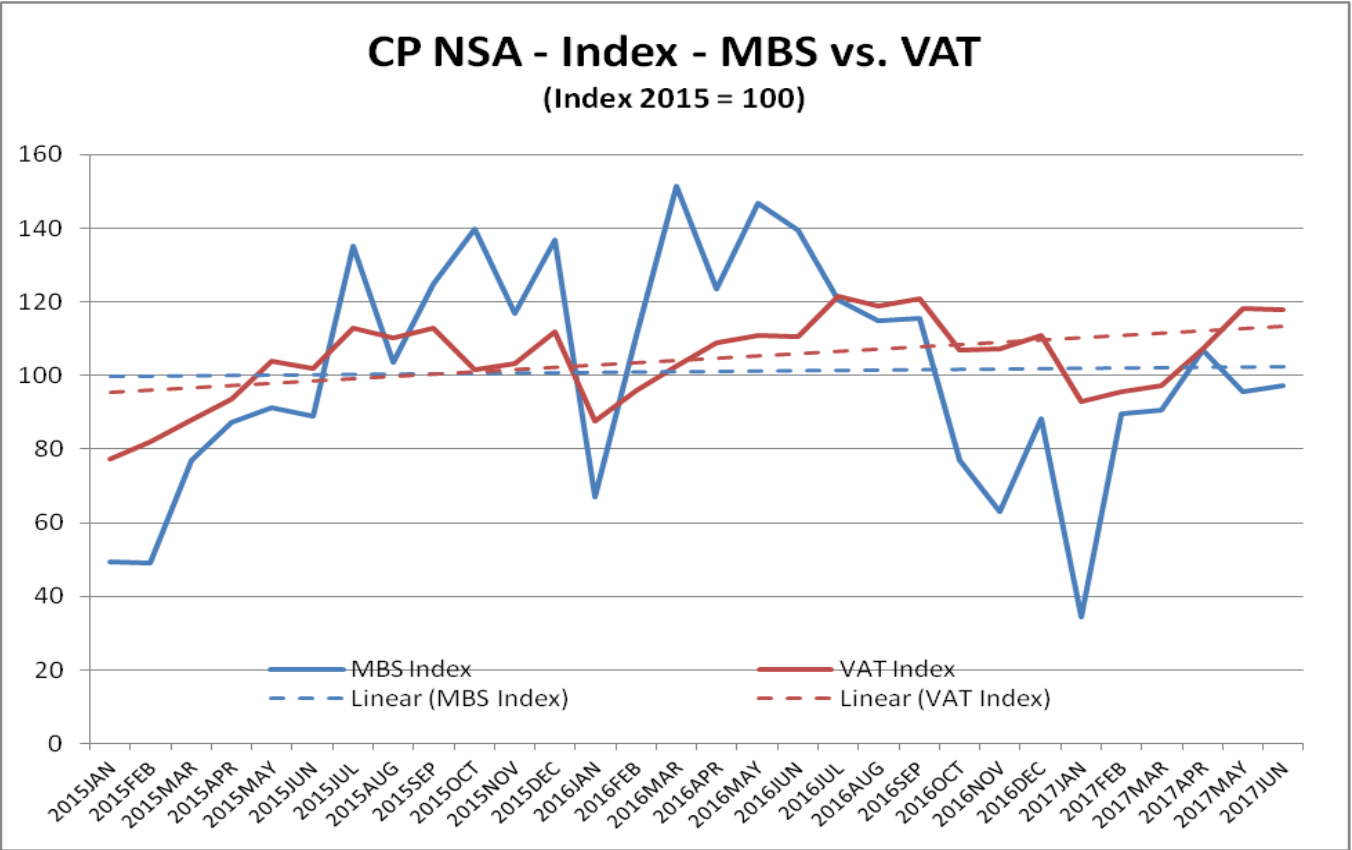
SIC07 – 96.02 – SB1 – Hairdressers and beauty treatment



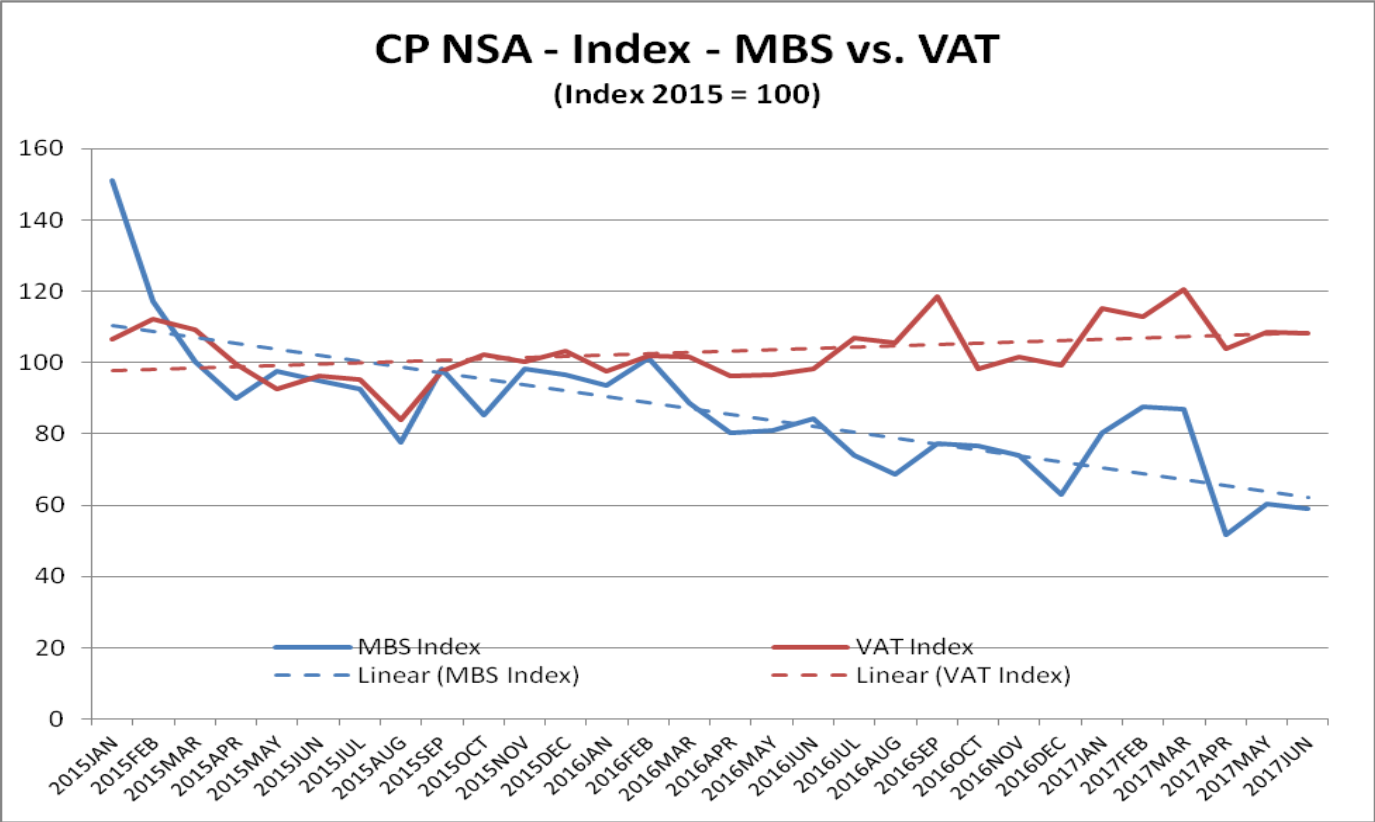
SIC07 – 55.1 – SB1 – Hotels and similar accommodation



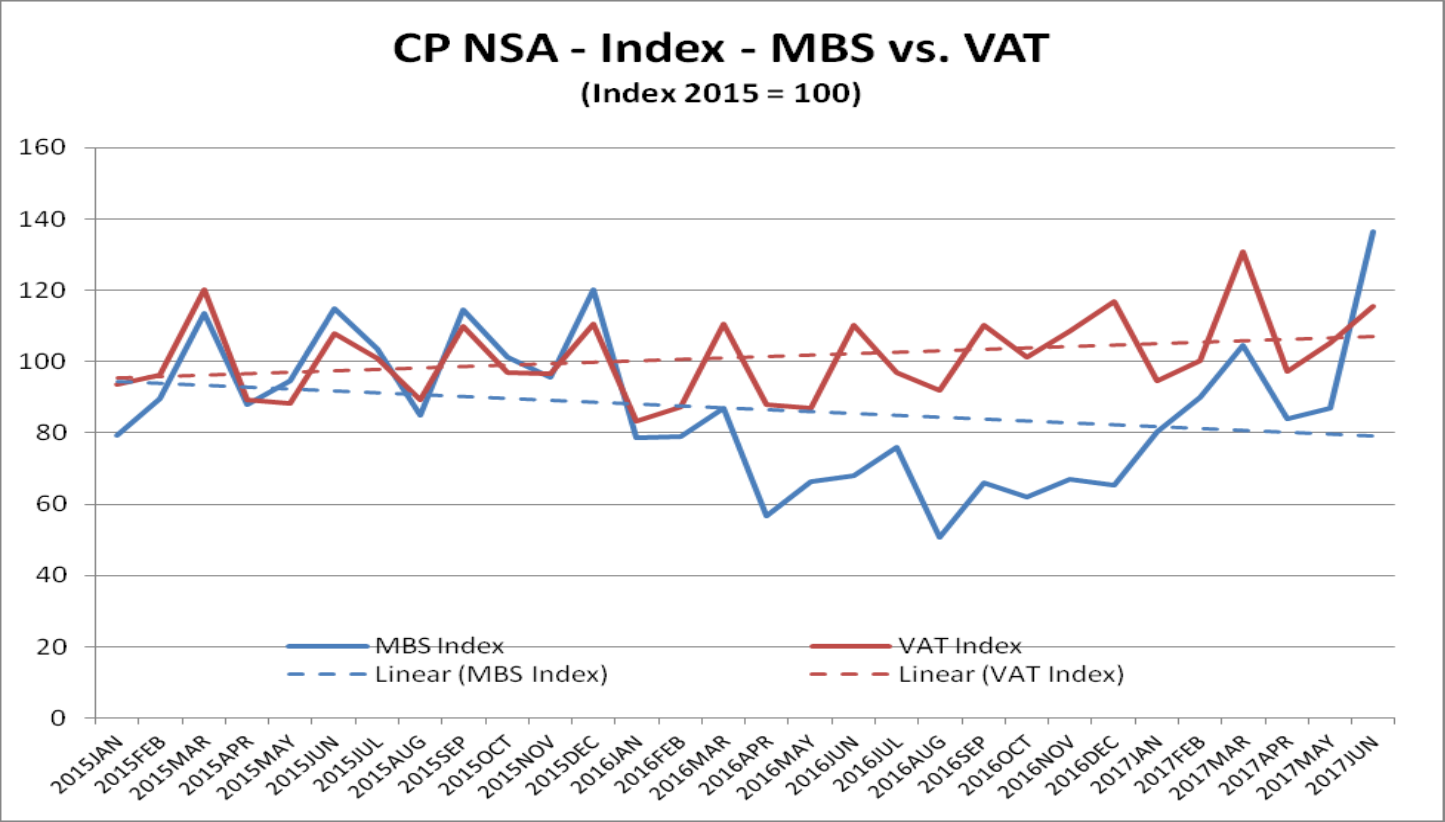
SIC07 – 11.02 to 11.05 – SB1 – Manufacture of wine, cider, beer and other non-distilled fermented beverages



SIC07 – 96.03- SB3 - Funerals



SIC07 – 26.5 – SB3 – Manufacture of instruments and appliances for measuring, testing and navigation; watches and clocks



Summary

- We have built a platform to process VAT turnover data
- Converts VAT unit level data into Reporting Unit level aggregates
- Allows for reducing the number of survey forms and comprehensive dataset for coverage
- Further work and analysis is ongoing

Questions

- Get in touch:
craig.mclaren@ons.gov.uk (twitter: @chmclaren) or
matthew.whipple@ons.gov.uk
- Please check out the poster from the Data Science Campus:
“Faster indicators of UK economic activity”

References

- [VAT turnover data in National Accounts: background and methodology](#)
- [VAT turnover implementation into national accounts](#)
- [Future uses of VAT](#)
- [Quality assurance of administrative data \(QAAD\) report for Value Added Tax turnover data](#)