



Office for
National Statistics
Swyddfa
Ystadegau Gwladol

Challenges in Integrating Administrative VAT Data into UK Short-term Output Statistics

Markus Gintas Šova

Overview

- Background
- Issues with VAT data
- Current research
- Future plans

Background

Growing interest in using admin data to partially replace surveys.

In the UK, 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, survey costs and improve quality.

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

...

How to Produce Monthly Data?

Currently:

Divide quarterly data by 3, annual data by 12.

Alternatives:

in proportion to the number of days

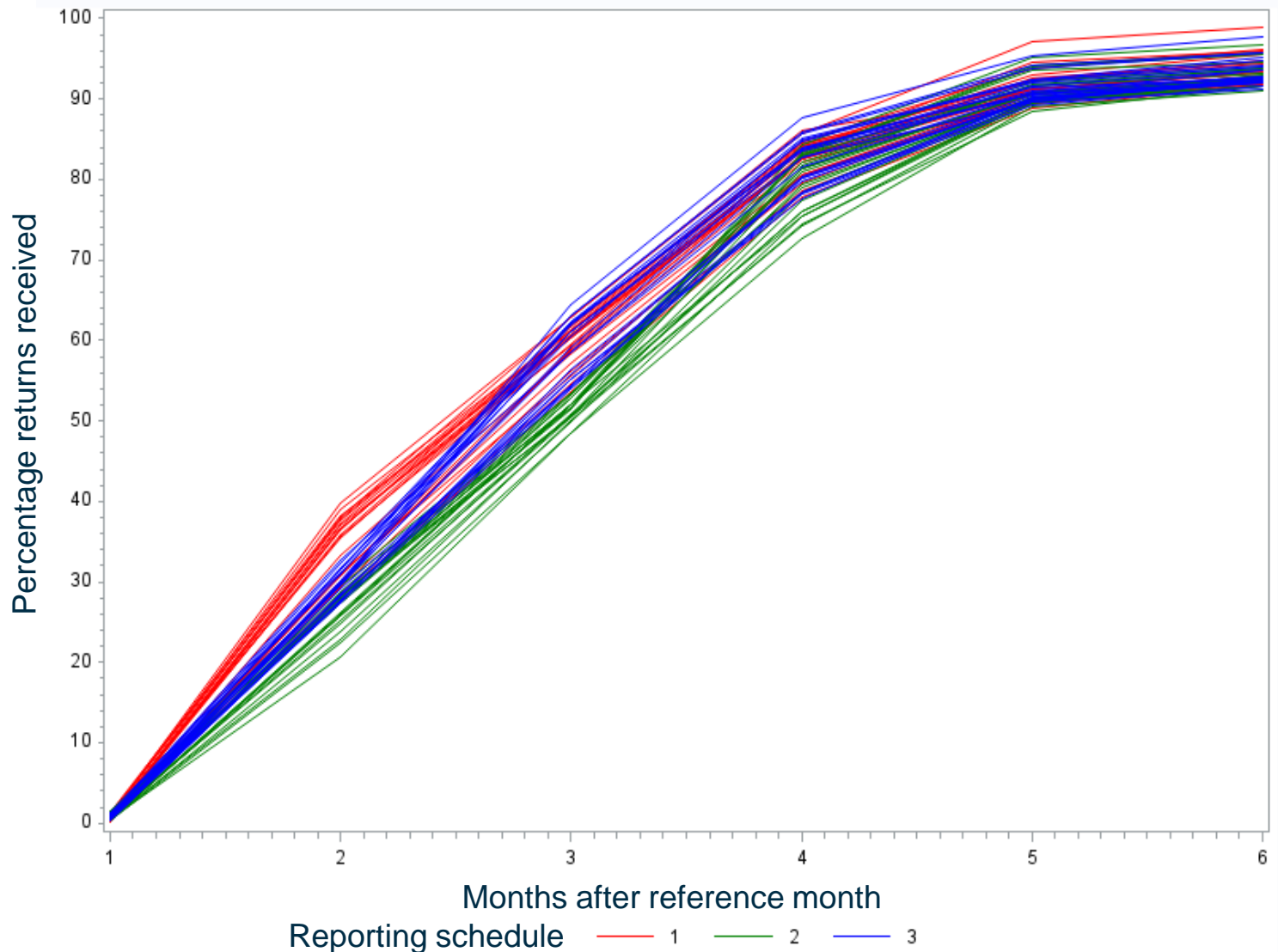
in proportion to the number of working days

more sophisticated time series methods

Different methods may be suitable for different industries.

Issue: Data Timeliness

Response Timeliness for NACE Group 107



Impact of Data Timeliness

VAT reporting timeliness varies by reference period.

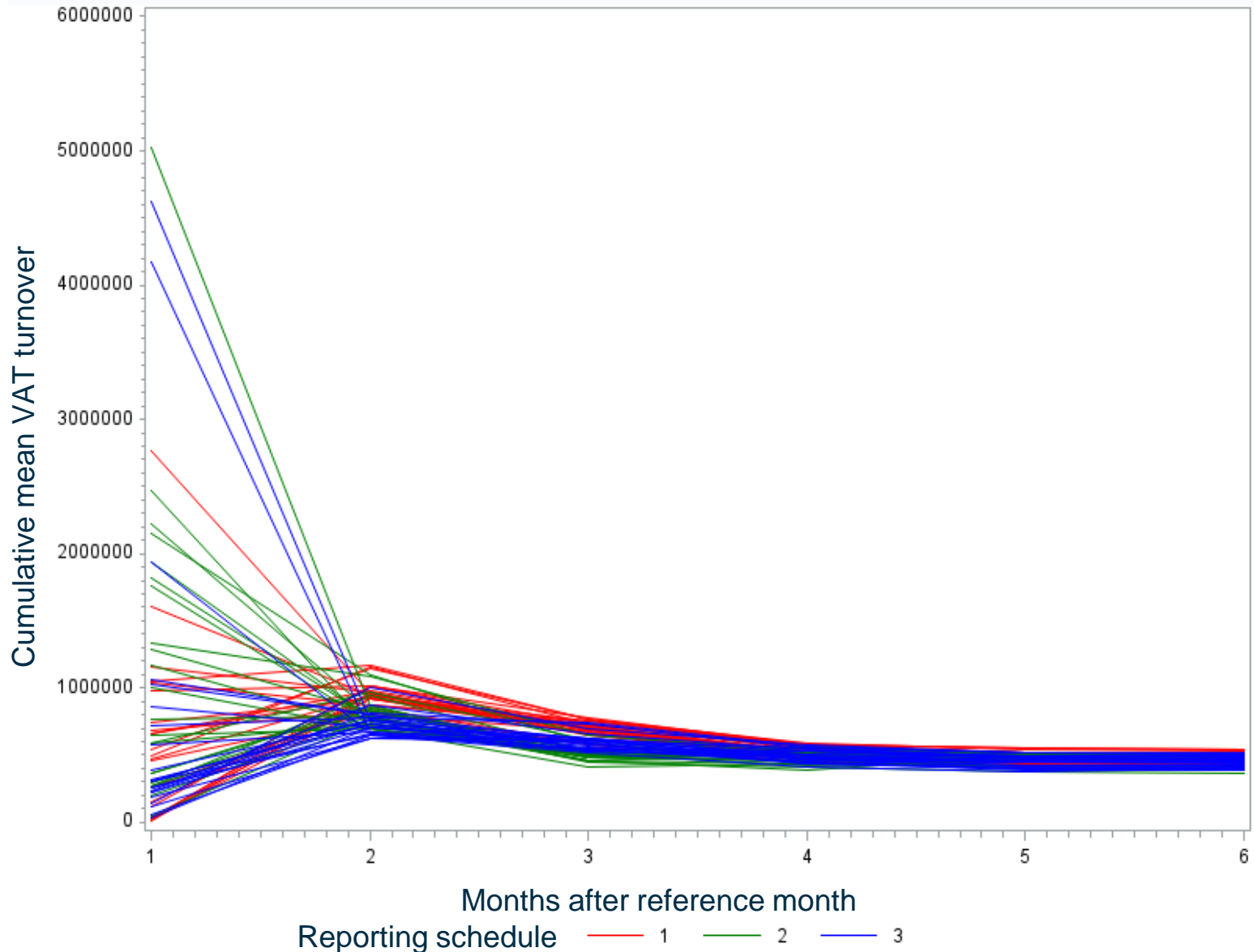
Also varies by reporting schedule.

Little data available until 2 months after reference period ($t+2$).

For many industries not until 4 months after reference period ($t+4$).

Quite an issue for a short-term indicator!

Issue: Mean Responses by Timeliness



In Other Words...

We have informative reporting.

Early data does not have the same distribution as later data.

Issue: Definitions

VAT units not always the same as statistical units.

Reported turnover definition differs from STS definition of turnover.

Issue: Data Quality

Needs cleaning!

Cannot verify surprising data with enterprises.

Issue: Data Revisions

Some VAT returns are subsequently revised
... sometimes more than once
... sometimes by more than an order of magnitude.

>99.9% of returns unchanged 12 months after reference month ($t+12$).

Assume that revisions improve accuracy.

Effect of large revisions may be reduced when initial returns are cleaned.

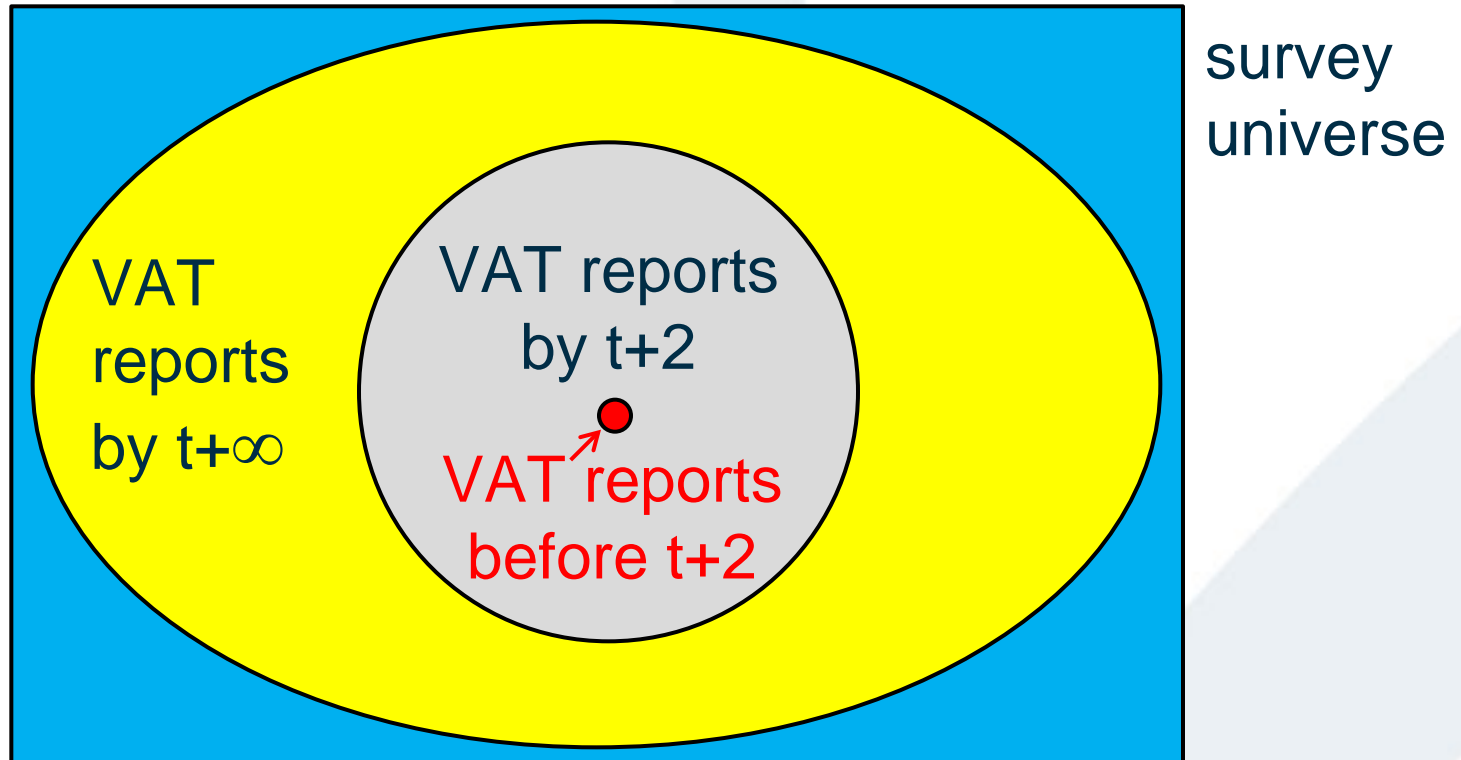
Issue: Undercoverage

Enterprises have to be VAT-registered if annual sales exceed a threshold, currently £82 000.

There are some exemptions.

Enterprises with sales below the threshold can register voluntarily.

Issue: Undercoverage



Two-phase non-random informative sample.
Includes some self-selection.

Adjust using register employment or turnover?

Current State of Research

Focus so far has been on using VAT turnover returns to at least partially replace MBS (Monthly Business Survey) turnover.

Predicting VAT turnover data to overcome timeliness issue is proving difficult.

Another Possibility?

Demand for monthly turnover estimates at class-level (4-digit NACE).

But MBS industrial stratification at a higher level
– sometimes divisions (2-digit).

Results in some classes having little or no data.

A sample redesign would either increase costs or reduce accuracy for higher-level estimates.

Could we use VAT returns to apportion MBS stratum turnover estimates to classes?

What about Timeliness?

If class turnover proportions within a stratum are stable over time, then we can use VAT returns from earlier periods.

Also consider register data:

- same administrative source (HMRC)

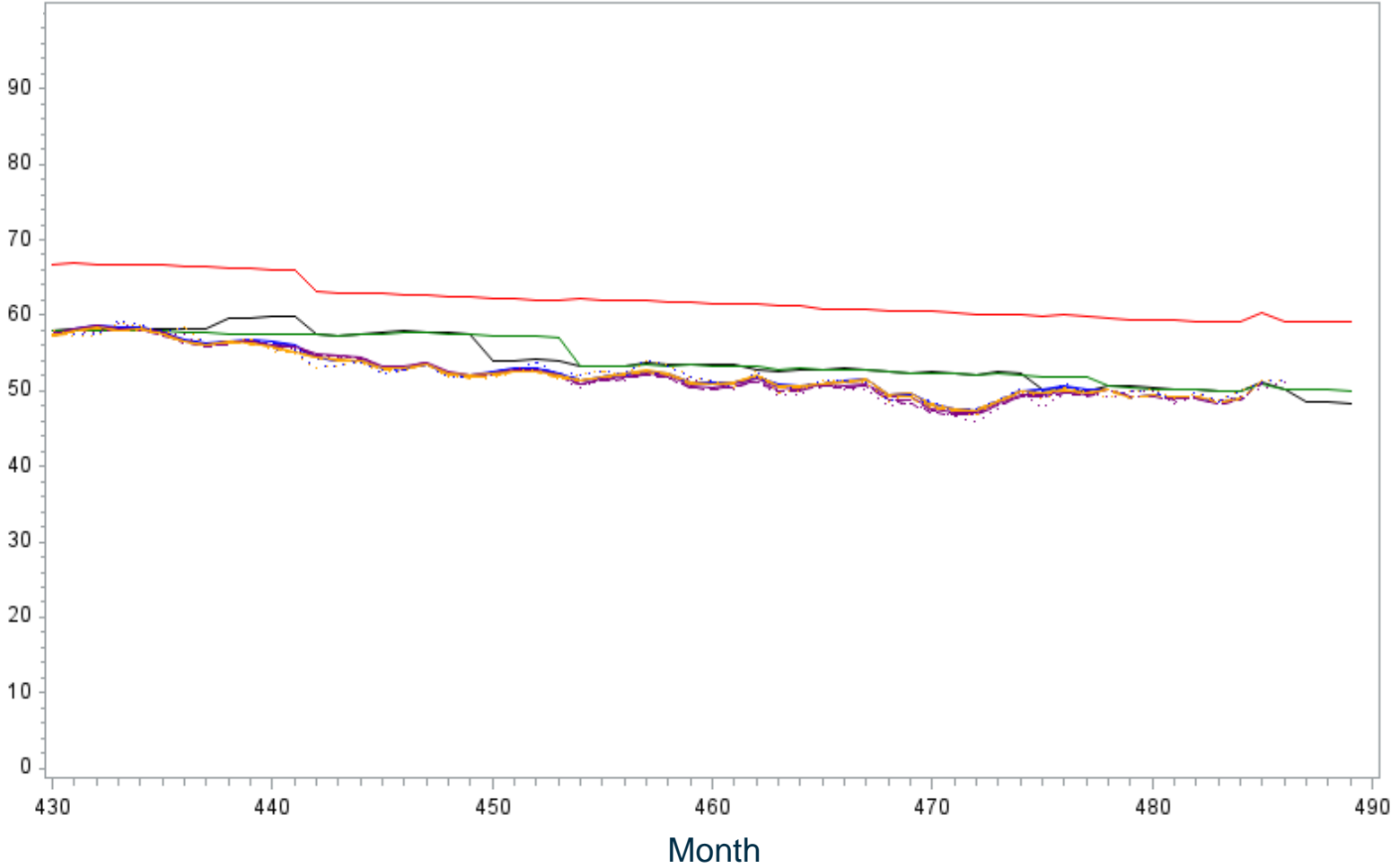
- complete register coverage

- but older than most recent VAT returns and covers 12 months.

So, are the proportions stable?

A Nice Example

NACE Class 9003 as a Percentage of Stratum 6451

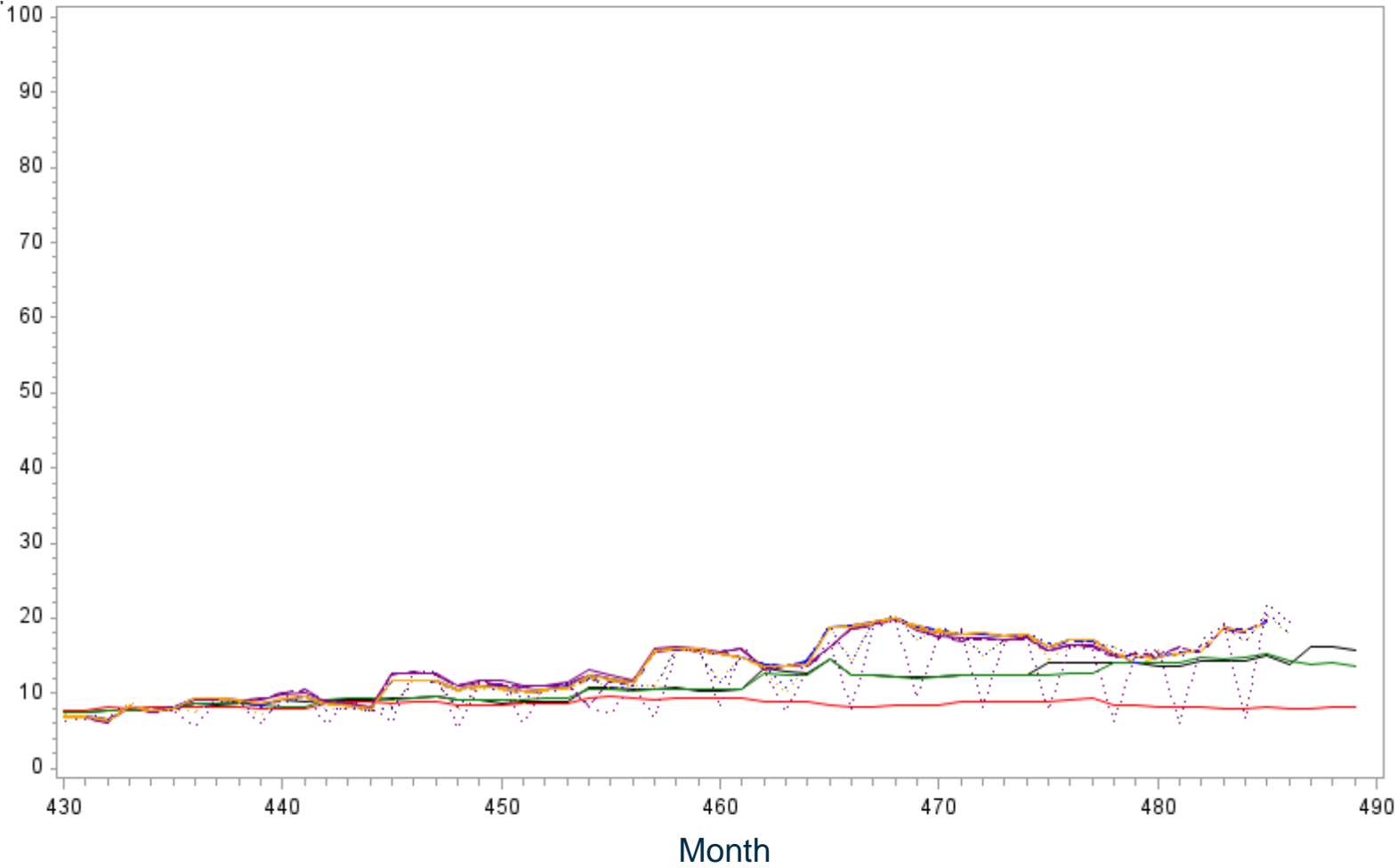


Method

- pidbrct
- pidbre
- pidbrft
- pvatct12
- pvatct4
- pvatct5
- pvatct6
- pvate12
- pvate4
- pvate5
- pvate6
- pvatft12
- pvatft4
- pvatft5
- pvatft6

Timeliness Can Still Be an Issue

NACE Class 1084 as a Percentage of Stratum 5091



- Method
- pidbrct
 - pidbre
 - pidbrft
 - pvatct1 2
 - pvatct4
 - pvatct5
 - pvatct6
 - pvate1 2
 - pvate4
 - pvate5
 - pvate6
 - pvattf1 2
 - pvattf4
 - pvattf5
 - pvattf6

Does it Matter...?

... if the proportions are not stable?

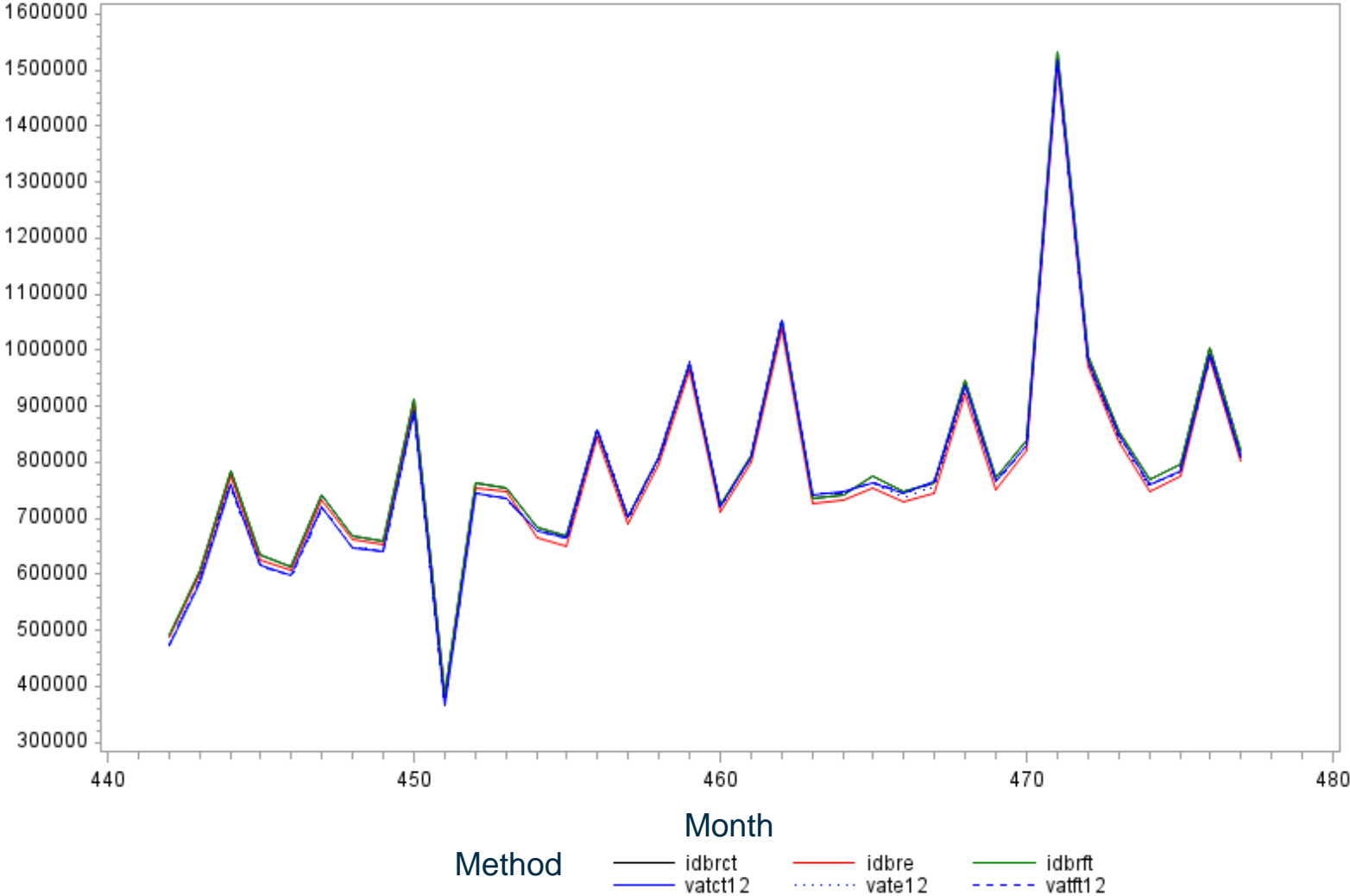
... if the different methods give notably different proportions?

The impact on the class estimates is more important than the impact on the class proportions in each stratum.

So what is the impact on monthly class turnover estimates?

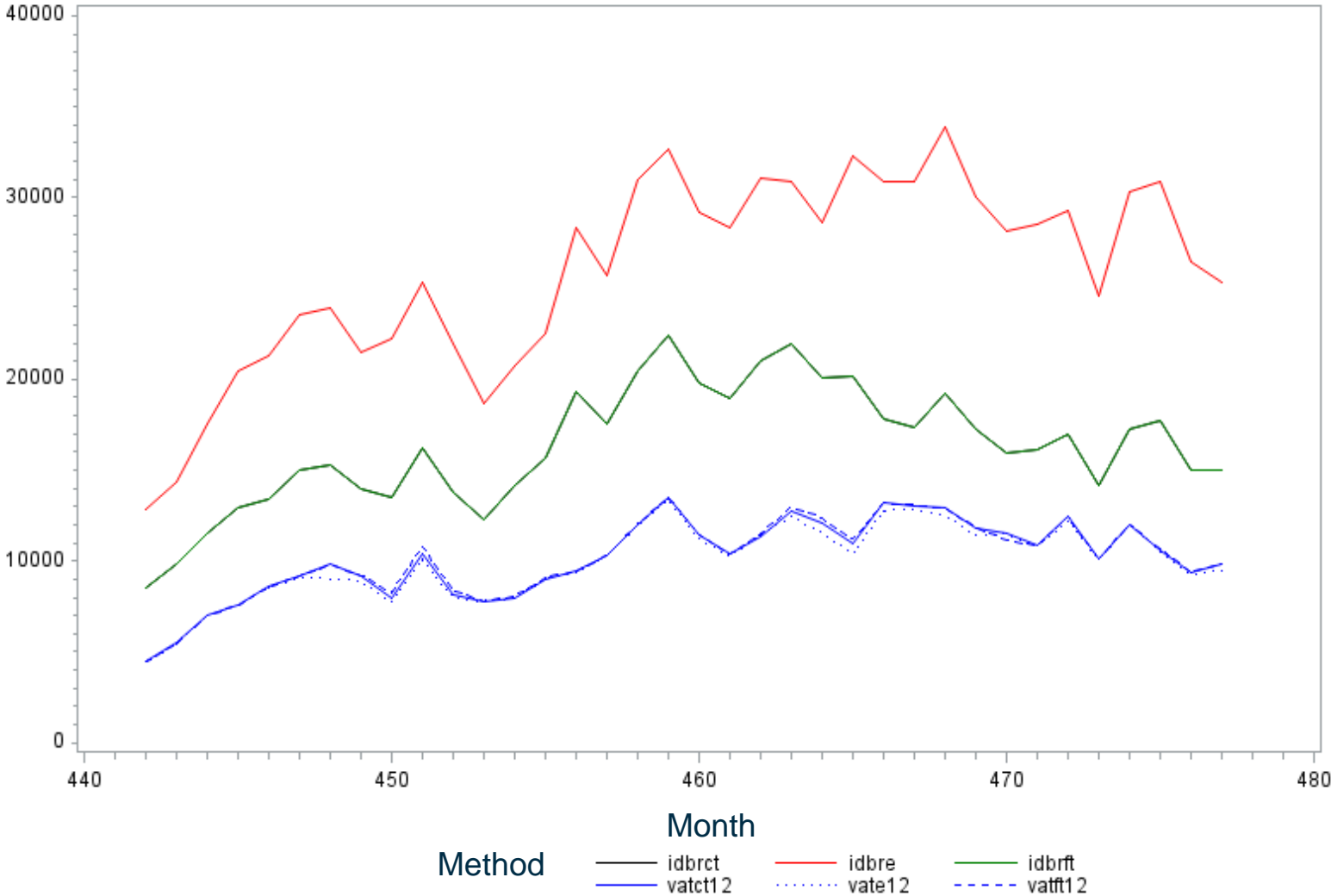
A Nice Example

NACE Class 2811 Turnover Estimates



A Not So Nice Example

NACE Class 3900 Turnover Estimates



Future Plans

- Seek to explain why apportionment results sometimes diverge for different methods.
- Test performance in a live environment.
- Continue longer-term research for partial replacement of MBS.