

Short-term turnover statistics based on VAT and Monthly Business Survey

Li-Chun Zhang (L.Zhang@soton.ac.uk)

Alison Pritchard (alison.pritchard@ons.gov.uk)

Outline

- Some general remarks
- A prediction framework
- Highlights of empirical findings
- Way forward

- Alternative approaches, roughly speaking:

1. BR + MBS (all units)
2. BR + VAT/MBS (super units) + MBS (rest units)
3. BR + MBS (largest units) + VAT (rest units)
4. *BR + VAT (all units; in retrospect)*

- Remarks

- Target population *and* classification based on combined sources
- Coherence/compatibility/constraint btw survey and register data
- Timeliness *vs.* burden/resource

- Decomposing development

- prediction of VAT mature turnover total using VAT early and historic reports
- harmonisation and reconciliation btw VAT mature totals and MBS estimates

- Longitudinal progressive register data

- repeated reports of values over time
- available value for a given reference point in time may evolve over time

NB. Delays and changes in registers; distinct feature compared to sample / census

- Necessary with prediction and modelling

- Target population $U(t)$: *VAT-active* units in *statistical* period t
 - Target total turnover: $Y(t) = Y_{U(t)} = \sum_{i \in U(t)} y_i(t)$
 - Available value by *measurement* time s : $y_i(t; s)$ for $s \geq t$
- NB.* $y_i(t; s) = 0$ if no value is available by s ; possibly $y_i(t; s') \neq y_i(t; s)$ for $s' > s$
- Population- t measured at s : $U(t; s) = \{i; y_i(t; s) \neq 0\}$
 - Turnover- t at s : $Y(t; s) = Y_{U(t; s)} = \sum_{i \in U(t; s)} y_i(t; s)$
 - Two useful assumptions, at least initially:

$$y_i(t) = \lim_{s \rightarrow \infty} y_i(t; s) := y_i(t; \infty) \quad (1)$$

$$y_i(t) = \min_{s: y_i(t; s) \neq 0} y_i(t; s) \quad (2)$$

NB. maturing after e.g. 12 months: $U(t) = U(t; t + 12)$ and $Y(t) = Y(t; t + 12)$

A prediction framework: Composition of $U(t)$

- $t + d$: *estimation* time point for $Y(t)$, with $d > 0$ being the *production lag*
- $I_i(t; t + d)$: 1 if $y_i(t; t + d) \neq 0$, and 0 otherwise
- At $t + d$, divide $U(t)$ into 3 disjoint sub-sets:

$$U(t) = U_1(t; t + d) \cup U_2(t; t + d) \cup U_0(t; t + d)$$

$$\text{VAT-reports: } U_1(t; t + d) = \{i; I_i(t; t + d) = 1\}$$

$$\text{VAT-delays: } U_2(t; t + d) = \{i; I_i(t; t + d) = 0$$

$$\cap I_i(t; \infty) = 1 \cap \sum_{j=1}^{\infty} I_i(t - j; t + d) \geq 1\}$$

$$\text{VAT-birth delays: } U_0(t; t + d) = \{i; I_i(t; \infty) = 1 \cap \sum_{j=0}^{\infty} I_i(t - j; t + d) = 0\}$$

NB. VAT-existent universe: VAT-reports & VAT-delays

NB. reported 0 turnover: VAT-inactive; $I_i(t) = 0$; excluded from $U(t)$

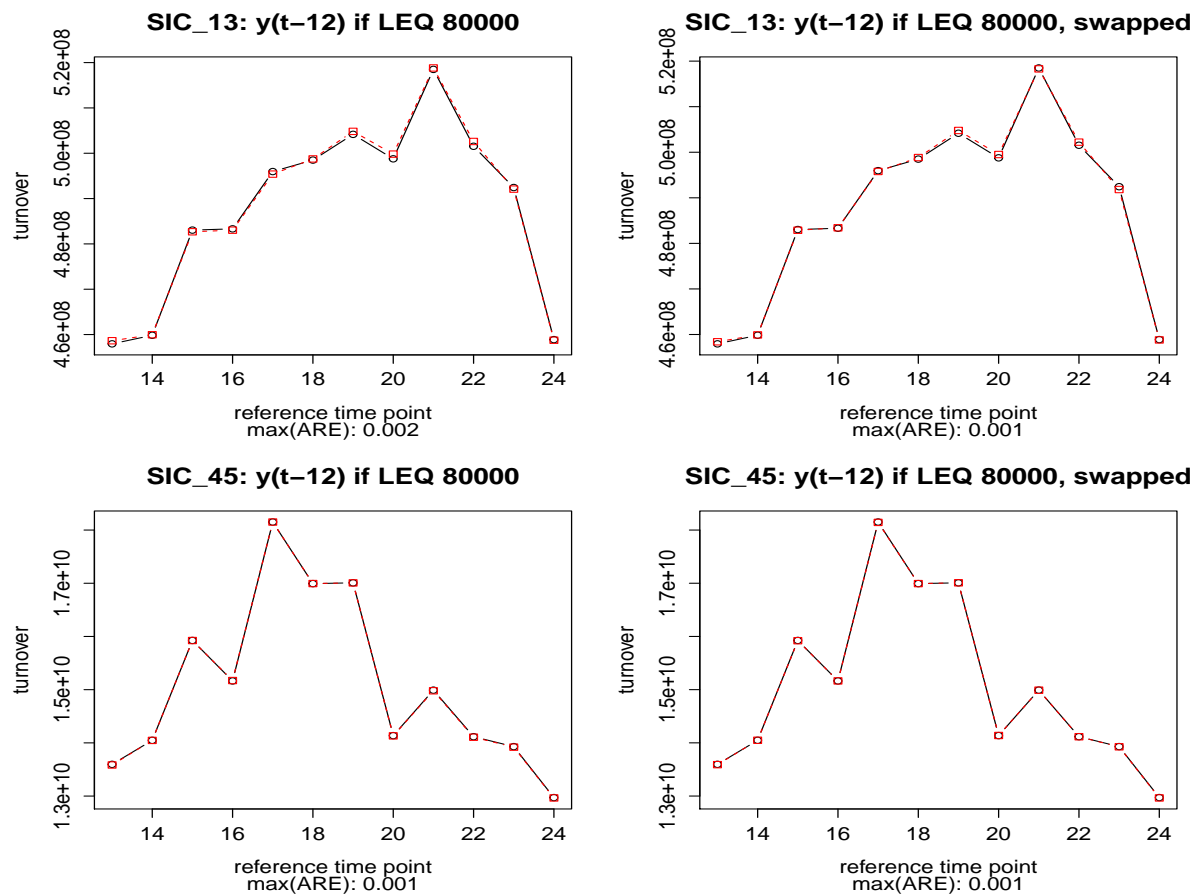
$$\hat{Y} = Y_{U_1(t;t+d)} + \hat{Y}_{U_2(t;t+d)} + \hat{Y}_{U_0(t;t+d)} \quad (3)$$

- $Y_{U_1(t;t+d)}$ directly given by assumption (2): $y_i(t) = \min_{s:y_i(t;s) \neq 0} y_i(t; s)$
- $\hat{Y}_{U_2(t;t+d)}$ is the predicted VAT-delay total
- $\hat{Y}_{U_0(t;t+d)}$ is the predicted VAT-birth delay total

- Remarks

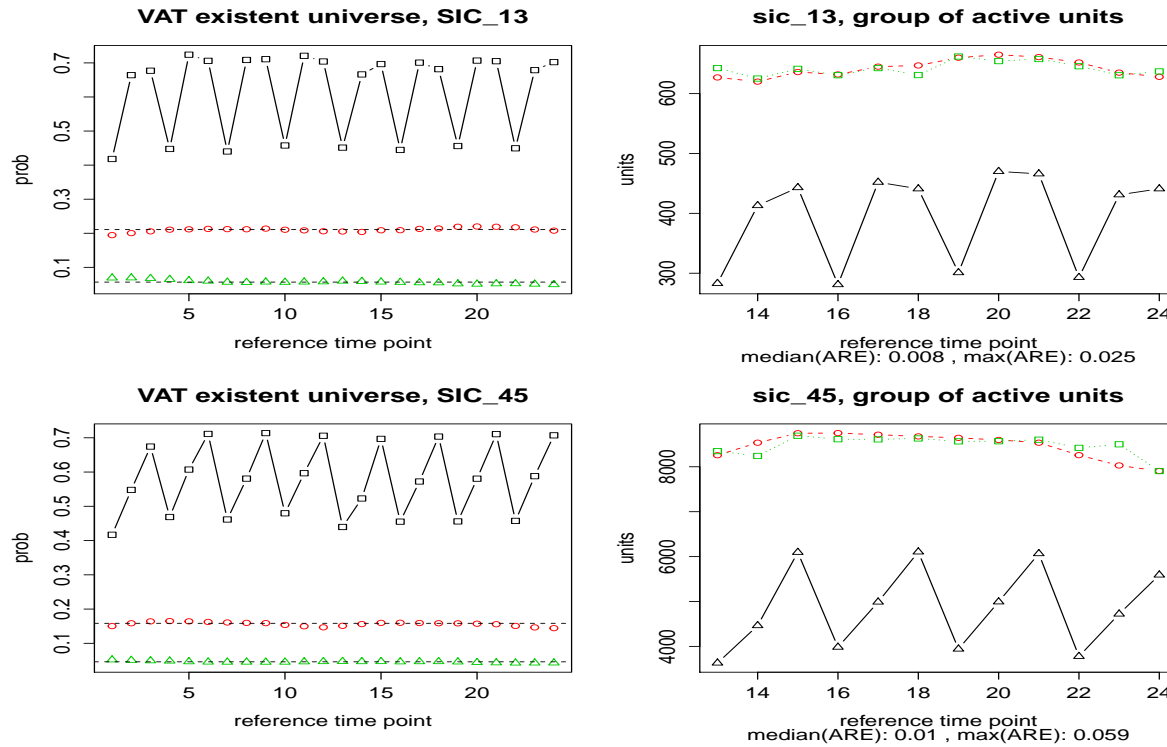
- avoid *ad hoc* ‘rules’ for I_i in practice; interplay with BR
- modelling of $\{(I_i, y_i); i \in U(t)\}$; an intriguing research area
- $U_1(t; t+d)$ a ‘sample’ by self-selection; extending informative sampl/nrs theory

Empirical finding (I): A substitution exercise



Turnover total of existent units (circle), *band-wise* substitution of $t - 12$ values for non-large units (square) for SIC-13 (80% units) and SIC-45 (85% units), with and without swapping

Empirical finding (II): Prediction of the number of large units



Left: reporting rate at $t + 3$ (square), proportion of units with turnover ≥ 80000 (circle), proportion of inactive units (triangle); reference time points in 2010 and 2011.

Right: number of reporting existent units at $t + 3$ (triangle), population size (circle) and predicted size (square) of units with turnover ≥ 80000 ; reference time points in 2011.

Way forward: An outline of a two-part solution based on combined data sources

- Monthly MBS sample of the largest self-representing units

NB. cut-in threshold; emerging in-scope units; outdated units

- Remaining turnover total by prediction and forecast/substitution

- VAT-birth delay total may be produced by projection (Zhang, 2013)

- VAT existent units (80%+ units) that are below a cut-off threshold

- VAT existent units that fall between the cut-off and cut-in thresholds

NB. timeliness vs. accuracy; different NACE-groups; band-wise strata

- Better uses of the VAT register can be made for
 - construction of the target population
 - selection and maintenance of the cut-in self-representing sample
 - exemption of survey compliance for the majority of units
 - prediction of the remaining non-self-representing non-cut-off units
- Future developments
 - harmonisation btw VAT register and BR: units and classification
 - VAT data delivery scheme: time point and frequency
 - maintenance of monthly sample; use of outlier-robust methods
 - evaluation and monitoring of results: cross-sectional and over-time