UNIVERSITY OF Southampton

An application of weighting approaches to assess the sensitivity of business survey estimates to “the unit problem”

1

Paul A. Smith

Southampton Statistical Sciences Research Institute (S3RI)

Outline

- the units problem
- estimation at different levels of units hierarchy
  - different estimation models
- estimation at different times
  - evolution of business structures
  - longitudinal estimation of changes
- discussion

The units problem

- businesses have complex interrelationships between their parts
- for statistical purposes we impose a model of these relationships
  - allows estimation using standard procedure
  - provides consistency
- model (and its application) are not unique representations ("the units problem")
  - what is the error in taking one model rather than another?

Units hierarchy

- Enterprise group > enterprise > (reporting unit) > local unit > local KAU
- estimates of variables of interest vary according to level
- for example, take reporting unit (RU) > local unit (LU)
  - ONS has examples of statistics based on both levels
- in (closed) population *A* have $diff_{pop} = \sum_{RU \in A} 1.x_j - \sum_{LU \in A} 1.x_l = 0$
- …but for domains, non-closed populations, differences are ≠ 0

Estimating differences

- Typically can't *enumerate* information on both structures
- Have sample data
  - assume sample data available on *both* structures
  - then can calculate $\sum_{RU \in A} w_j x_j$ and $\sum_{LU \in A} w_l x_l$
- the difference estimates population difference from measuring with different structures
- …but what to choose for w_j and w_l?

weights

- w_j typically based on RU auxiliary variables and totals
- w_l typically based on LU auxiliary variables and totals
- differences due to estimation model (cf Hedlin *et al*. 2001 JOS)
  - so how much difference from model choice and how much from 'units problem'?
- Lemaître & Dufour (LD) (1987) introduced 'integrated weighting' for households and people
  - can apply to two-level weighting for businesses
  - in my illustrative example need to drop two (very sparse) constraints

Comparison of weightings
(small illustrative example)

Method	Register employment (known)	Turnover (estimated)
population total (known)	5958	65431
RU estimation	5941 *	65766
LU estimation	5958	65724
Lemaitre-Dufour RU & LU estimation	5958	66249

Domain estimates

NACE	1	2	3	4	5
RU pop total	18378	9005	37650	253	146
RU	18260	9431	37910	164	0
L-D RU NACE	17956	9454	38528	311	0
LU pop total	16018	8446	38429	372	2165
LU	15814	8487	38927	276	2220
L-D LU NACE	15684	8618	39425	230	2293

Estimation using longitudinal business structures

- Business structures not static
  - eg LUs change hands (RUs), born, die, buy-out etc
- (Generalised) weight share (Lavallée 2007) idea takes weights from originally selected units and shares them among units at later time
- eg LU estimation
  - calculate weights at t_0
  - share weights among units in structure at t_1
  - use both sets of weights on s*ame* period's data to assess impact of changing structures

Comparison of weightings
(small illustrative example)

Method	Register employment (known)	t_0 turnover (estimated)
population total (known)	5958	65431
t_0 LU estimation wts	5958	65619
t_1 LU estimation wts	5958	65641

Domain breakdown

NACE	1	2	3	4	5
population total	16018	8446	38429	372	2165
t_0 LU estimation	15991	8454	38695	276	2203
t_1 LU estimation	15990	8454	38721	273	2203

Discussion

- No *error*, just *difference*
- Weights give one way to estimate differences
  - (suspect Bayesian methods may allow incorporation of model uncertainty in a more structured way)
- Need clarity over what population parameter is estimated, and how it can be interpreted in context of units problem
- Needs simulation (not just single example) and real data

- Paul Smith
- p.a.smith@soton.ac.uk