

# Virtual Worlds

## Modelling A Local Authority Using the FRS

# Why?

- Localism agenda: Benefits increasingly devolved to local authorities;
- Social Care;
- Born of failure: we were asked to bid for a project to do this – we didn't get it!

# How?

- The FRSs we have don't include LA identifiers
- In any case, samples would be small, and sample is stratified.
- Instead: find weights so the sample looks like the council

# How?

- Technique is an extension of the method used to create the basic FRS weights included in the dataset
- Council populations differ in (e.g)
  - Proportions employed and unemployed
  - Racial composition
  - Age Structure
  - Numbers on Benefits
- And many others

# How?

- Get Data on these things broken down by council from Nomis, HMRC, DWP...
- Make a dataset with the same variables from FRS
- Find a set of weights such that the FRS data sums to the council totals
- Infinitely many sets of weights are possible
- Pick the set that minimises the distance (in some sense) between the final weights and some initial (usually uniform) set of weights.
- Problem is then a constrained minimisation problem familiar in economics

# How: Choosing Weights?

- A variety of distance measures are possible
- See Creedy (2006) and Deville and Sarndal (1992) (Atkinson (1989) for a discussion;
- Obvious one: sum of squared deviations ('Chi-square') – doesn't work because weights can be negative
- Various 'entropy' measures proposed, but these can be unstable
- 'Constrained chi' works well, usually – minimise sum of squared deviations between uniform and new weights but place constraints on how far new weights can deviate from initial weights - e.g. must be positive.

# Limitations

- Can only fix composition differences – distributions of (e.g.) wages, housing costs will be corrected only to the extent that they are due to differences in composition (a council has disproportionate number of professionals, large houses, etc.);
- Composition variables not in the targets set may be made worse by the new weights;
- Program doesn't always converge

# Implementation

- Written in Ada
- Data in relational database
- Council target data assembled from
  - Nomis Census Data
  - HMRC Tax Credit Data
  - Benefit Receipts from DWP and Nomis
- FRS data from 2009/10 and 2010/11 (other datasets possible)
- Getting NOMIS and FRS consistent is fiddly
- Minimisation code based on Creedy
- Generates Household-Level weights
- Has (very ugly) web interface <http://projectsvr.virtual-worlds-research.com/la/>
- All downloadable from <http://virtual-worlds-research.com/downloads/>



# (Ugly!) User Interface

LARW :: - Mozilla Firefox

Distance ... TMA 04: ... DD309-... PDF File ... ICA Institute ... Inbox - ... Facebook LAR...

projectsrv.virtual-worlds-research.com/la/?code=E06000042 Andy Murray Verdasco

Generate FRS Weights

### Possible Targets

Age in ranges: 0-4, 5-10, 11-15, 16-24, 25-34, 35-44, 45-54, 55-59, 60-64, 65-74, 75+

	LA Totals	Uniform-Weighted FRS
0-4	19,908	15,537
5-10	19,644	17,014
11-15	15,840	14,553
16-24	25,565	21,322
25-34	40,556	25,997
35-44	38,588	31,040
45-54	33,941	30,229
55-59	14,099	13,367
60-64	13,169	15,284
65-74	15,467	23,157
75+	12,044	17,378

Aggregated Ages: ☒

Genders: ☐

Ethnic Group: ☐

Genders (M/F) All residents

	LA Totals	Uniform-Weighted FRS
Male	123,116	108,013
Female	125,705	116,863

Ethnic Group: Aggregated (white, mixed, asian, black, other)

	LA Totals	Uniform-Weighted FRS
White	200,673	205,612
Mixed	8,235	3,813
Asian	22,782	10,845
Black	17,131	4,606

LARW :: - Mozilla Firefox

Distance ... TMA 04: ... DD309-... PDF File ... ICA Institute ... Inbox - ... Facebook LAR...

projectsrv.virtual-worlds-research.com/la/?code=E06000042

Andy Murray Verdasco

	4	17,760	20,148
	5	25,158	26,507
	6	16,914	19,947
	7	9,736	10,215
	8	7,165	5,870
	9+	7,765	5,227

Occupation (aggregated: white\_collar,blue\_collar)

	LA Totals	Uniform-Weighted FRS
White Collar	70,316	52,360
Blue Collar	57,467	41,671

Occupation (Standard Occupational Classification 2000 disaggregated)

	LA Totals	Uniform-Weighted FRS
Managers, Directors and Senior Officials	13,871	14,365
Professional	23,219	12,749
Associate Professional and Technical	17,189	14,660
Administrative and Secretarial	16,037	10,586
Skilled Trades	11,298	9,921
Caring Leisure and Other Service	10,035	8,667
Sales and Customer Service	11,910	6,327
Process Plant and Machine Operatives	8,484	6,596
Elementary Occupations	15,740	10,161

Economic Activity (ISO) Aggregated

	LA Totals	Uniform-Weighted FRS
Active	137,717	102,481

Occupation Aggregated: ☐

Occupation Disaggregated: ☐

Economic Activity Aggregated: ☐

(L

distance functions' are available.

Chi-square (unconstrained) just minimises the sum of squares; this is not recommended as it can produce negative weights. Chi-Squared (constrained) is similar but puts upper and lower bounds on the weights. See the Deville and Sarndal paper for a description of the other measures; note that these frequently fail on this data and sometimes produce negative weights (indicative of a possible programming error). It's probably best to leave the distance function type as "constrained chi".

The upper and lower bounds represent how far from uniform the weights are allowed to go when using a constrained distance function; a lower-bound 0.1 means the weights cannot be less than 1/10th of the uniform weight, and an upper bound of 10 means that no household can have a weight greater than 10x uniform.

The weights generation can fail if the council population characteristics are very different from the FRS population. If this happens try widening the lower and upper bounds.

Weights can be generated so that the whole of the FRS sample grosses up to the Council's totals, or so that a subset of the FRS does - just England, or just the Government region containing the council. Using a subset might help ensure that the dataset better resembles the council in other ways, such as wage or rent levels.

Distance Function Type:

Lower Bound:

Upper Bound:

FRS Subset:

Generate FRS Weights

Download Weights [here](#).

**Version Information**

\$Revision: 16193 \$  
\$Date: 2013-06-03 11:39:35 +0100 (Mon, 03 Jun 2013) \$

# Example Output

YEAR	SERNUM	WEIGHT
910	1	1.999
910	2	2.087
910	3	2.882
910	4	1.092
910	5	1.113
910	6	1.542
910	7	1.324
910	8	3.538
910	9	1.521
910	10	1.542
910	11	1.627
910	12	2.087
910	13	3.019
910	14	1.777
910	15	2.369
910	16	1.753

# Do do

- Add code to correct wages, housing costs, other incomes, given composition:
  - Quite good disaggregated data on wages from LFS
  - Housing costs data seems less good (regional only?)
  - Haven't investigated other incomes
  - Probably never be able to capture some councils properly (Westminster)
- Support other datasets;
- Wider range of targets;
- Different aggregations (regions, police authorities, metropolitan areas..);
- Build directly into tax-benefit model;
- Nicer interface.

- Run the program:  
<http://projectsvr.virtual-worlds-research.com/la/>
- Download the code:  
<http://virtual-worlds-research.com/downloads/>
- Further Reading:

[Creedy \(2003\) “Survey Reweighting for Tax Microsimulation Modelling”](#)

Deville, J.-F. and Sarndal, C.-E. (1992) Calibration estimators in survey sampling.  
Journal of the American Statistical Association 87