Construction of Regional Consumer Price Indices using Small Area Estimation

Nora Würz, Timo Schmid, Paul Smith & Nikos Tzavidis
Freie Universität Berlin & University of Southampton

1. Motivation

- **Consumer Price Indices** (CPIs) are important indicators for monetary stability and inflation. For UK they are (normally) calculated on national level, but there is the need for regional CPIs.
- **Regional CPIs** show differences in the rate of regional inflation, for this calculation different regional baskets of good and services and their weighting pattern are needed.
  - Regional weights for UK has only been produced for 111 main product classes and 5 aggregated regions by the Office for National Statistics (2011), however finer subdivisions are desired.

- **Problem:** Small sample sizes lower the accuracy of regional baskets and their weighting pattern.
- **Research proposal:** Apply Small Area methods to produce regional baskets with higher accuracy.
  - To overcome the problem of **small sample sizes** within each region.
  - Archive **longitudinal consistency of the baskets** by reducing the changes caused by small sample sizes, so the comparability of CPIs from different years increases.

The **CPI for the UK** is calculated as a chained Laspeyres type index.
- Prices are collected monthly.
- A national basket of goods and services and its weighting pattern is updated annually.

2. Database

The Living Cost and Food (LCF) survey for UK [2] is used to estimate the **basket composition** and their **weighting pattern**.

- Continuous survey: Interviews and diaries are spread over the whole year.
- Sampling design: Multi-stage stratified random sample with clustering.
- Includes 5133 to 5593 households for the three available years (2012, 2013 and 2014).

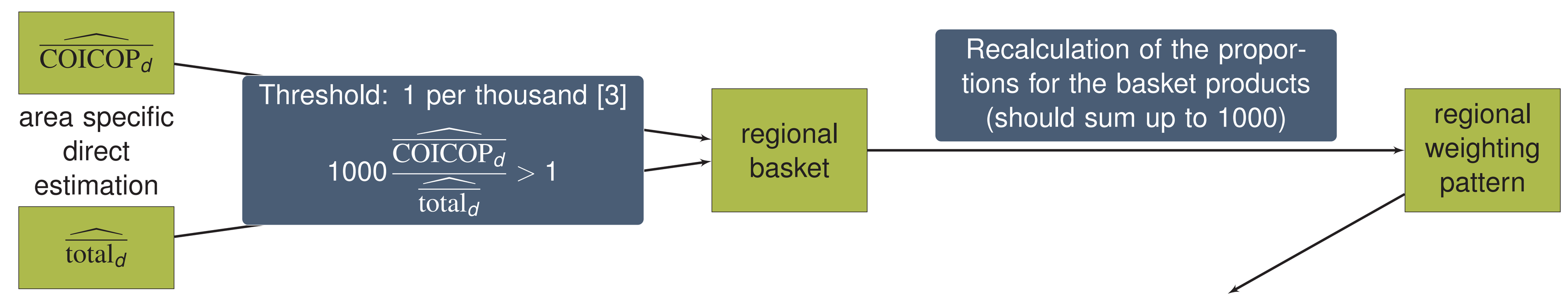
Diary dataset
- Each individual has to keep an expenditure diary for two weeks.
- The expenditures of the household members are aggregated on household level.
- The products are coded according to the United Nations **C**lassification **O**f **I**ndividual **CO**nsumption by **P**urpose **(COICOP)**.

Household dataset
- 1946 variables
- Area specific direct estimated means can be used as auxiliary variables.

By COICOP all product classes assign to a hierarchical 5 digit code:

1.1.1.2.1	bread	1.1.1	products with main ingredient grain	1	food & non-alcoholic beverages
1.1.1.2	bread, buns & biscuits	1.1	food except non-alcoholic beverages		

3. Generating regional baskets and the respective weighting pattern



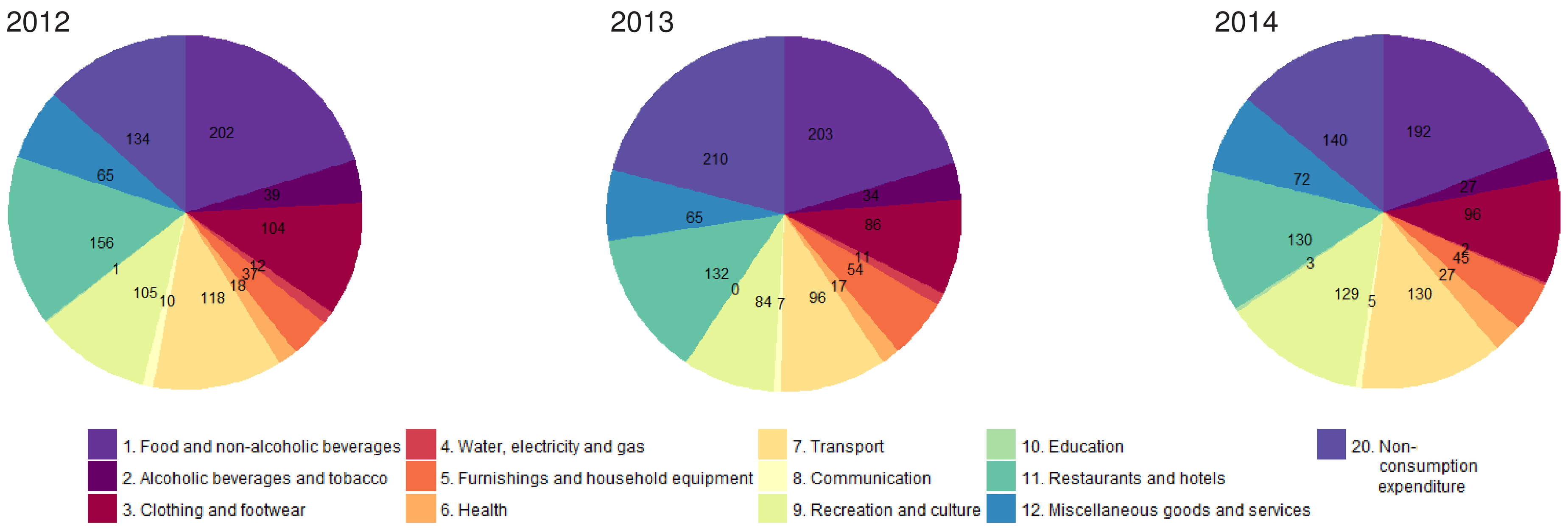


Figure: Weighting pattern for London from direct estimated expenditures for the different years (in part per thousand).

On regional level sample sizes are small and often only a few numbers of households buy a specific product.
- low accuracy of direct estimation, baskets and their weighting pattern

Table: Sample sizes of the 12 Governmental Official Regions for the available three years.

year	North East	North West	York. & the H.	East Midl.	West Midl.	Eastern	London	South East	South West	Wales	Scotland	Nor. Ire.
2012	262	623	521	425	513	563	490	783	493	266	483	171
2013	251	585	462	424	526	497	480	681	429	246	412	151
2014	255	588	459	440	470	498	407	740	468	222	434	152

The direct estimated regional baskets (cf. table below) and their weighting pattern (cf. pie charts above) are not stable over time.

Table: Summary for the proportion of equal products over the different regions: The basket of 2013 and 2014 is compared to the basket of 2012 respectively 2013.

year	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2013	0.85	0.89	0.89	0.89	0.90	0.93
2014	0.84	0.86	0.87	0.87	0.88	0.91

4. Small Area Estimation using Fay-Herriot models

Fay & Herriot (1979) introduced a model for small area estimation on area-level. It improves the estimation (lower MSE) by including census information as auxiliary variable and accepting a bias.

- For domain d the Fay-Herriot (FH) model is specified as
$$\hat{\theta}_d = x_d^t\beta + u_d + e_d \qquad d = 1, \dots, D \ ,$$
where $\hat{\theta}_d$ is the direct estimator, u_d the random effect, e_d the sampling error and D the number of domains. x_d are area-specific auxiliary variables linked by β to $\hat{\theta}_d$.
- The empirical best linear unbiased prediction for this model is obtained by
$$\hat{\theta}_d^{FH} = x_d^t\hat{\beta} + \hat{u}_d = \hat{\gamma}_d\hat{\theta}_d + (1-\hat{\gamma}_d)x_d^t\hat{\beta} \qquad d = 1, \dots, D \quad .$$
The shrinkage factor $\hat{\gamma}_d = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_{e_d}^2}$ indicates how much weight is put on the direct estimator and the synthetic part due to the (estimated) variances of u_d and e_d.

5. Application

- **Selection of auxiliary variables** from household data: Lasso regression is used.
- **Variance estimation of the direct estimator:** According to a conducted simulation study, naive non-parametric bootstrapping is chosen.

Product-specific Fay-Herriot models:
- The R-package *sae* (Molina & Marhuenda, 2015) is applied.
- Fay-Herriot models bundling product classes on different levels are used.
  - **Reason:** Overcoming the small number of domains (12) for each product class.
  - **Implementation:** Product classes are pooled according to their COICOP code.
  - **Input for FH model:** The direct estimations of each product class in each bundle are treated as independent observations and build together the input vector $\hat{\theta}_d$.

  Example: Pooling of the product classes

level 1:	1.x.x.x.x	identical first COICOP digit	bundling of 65 product classes
level 2:	1.1.x.x.x	identical first two COICOP digits	bundling of 58 product classes
⋮	⋮	⋮	⋮
level 5:	1.1.1.2.1	identical first five COICOP digits	individual product class (bread)

- From these estimated expenditures baskets and weighting patterns are generated.

6. Results

Coefficient of Variation for the different FH models & the direct estimation

- The CV $\left(\frac{\sqrt{\mathrm{MSE}(\hat{\theta}_d^{FH})}}{\hat{\theta}_d^{FH}}\right)$ decreases with increasing FH level.
- No explicit trend between the FH fitting on level 1 and direct estimation.

Table: Quantiles of the CV for different FH models among the three years.

year	2012			2013			2014		
Fitting	1st Qu.	Median	3rd Qu.	1st Qu.	Median	3rd Qu.	1st Qu.	Median	3rd Qu.
direct	15.15	30.31	59.05	15.79	30.92	60.93	15.76	30.62	58.75
FH level 1	15.34	31.24	58.77	15.77	32.00	62.90	14.71	30.44	61.67
FH level 2	13.46	26.37	52.51	12.82	24.60	46.49	13.68	26.23	48.02
FH level 3	12.53	24.31	46.08	12.87	24.19	43.76	13.38	25.07	44.65
FH level 4	11.37	22.58	44.23	11.55	22.53	42.39	12.15	23.32	43.30
FH level 5	10.28	18.53	36.61	10.42	18.83	35.78	10.92	19.31	32.58

According to the CV, a fitting of the FH model on detailed COICOP level is recommended.

Stability of the regional baskets over time

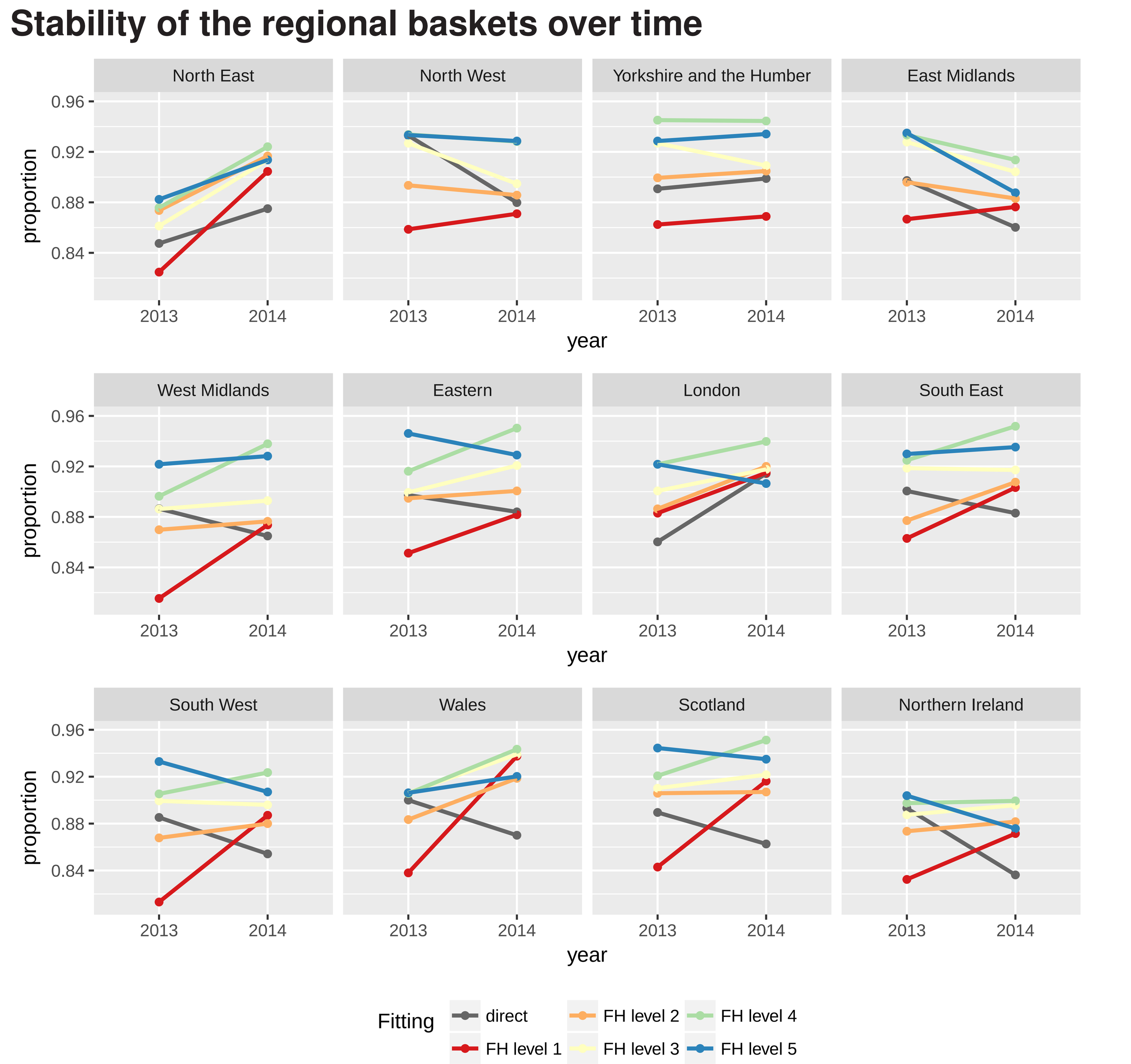


Figure: The lineplots show the proportion of equal products in the regional baskets for each Governmental Official Region in UK for 2013 and 2014 compared to the basket of 2012 respectively 2013.

To achieve high longitudinal consistency a fitting of the FH model on COICOP level 5 or 4 seems recommendable.

7. Discussion and Outlook

- This project shows that FH models should be applied, when regional baskets are calculated due to naturally much smaller sample sizes for regions.
- The models should be fitted on a detailed COICOP level according to the CV and the stability of the resulting baskets.

Further research:
- Improving the FH models by taking the uncertainty of the auxiliary variables into account (Ybarra & Lohr, 2008).
- Including benchmark approaches to constrain the small area estimates of the mean expenditures in each region to the national mean.
- Improve the FH algorithm for small number of domains.
- Analyse the prices & calculate regional CPIs for UK.

References

[1] Office for National Statistics (2011), UK Relative Regional Consumer Price levels for Goods and Services for 2010
[2] Office for National Statistics, Bulman, J., Kubascikova-Mullen, J. & Whiting, S. (2014), Living Costs and Food Survey Technical Report for survey year: January - December 2013, Great Britain and Northern Ireland
[3] THE COUNCIL OF THE EUROPEAN UNION (1998), Council Regulation (EC) No 1687/98 of 20 July 1998 amending Commission Regulation (EC) No 1749/96 concerning the coverage of goods and services of the harmonised index of consumer prices. Official Journal of the European Communities, L 214(41): 12-22
[4] Fay, R.E. & Herriot, R.A. (1979), Estimates of income for small places: an application of James-Stein procedures to census data. Journal of the American Statistical Association, 74(366a):761-766.
[5] Molina, I. & Marhuenda, Y. (2015), sae: An R Package for Small Area Estimation. The R Journal, 7(1): 81-98
[6] Ybarra, L.M.R. & Lohr, S.L. (2008), Small area estimation when auxiliary information is measured with error. Biometrika, 95(4): 919-931

For further information

Freie Universität Berlin

Nora Würz, Timo Schmid
nora.wuerz@fu-berlin.de,
timo.schmid@fu-berlin.de
Department of Economics
Garystr. 21 D-14195 Berlin

Paul Smith, Nikos Tzavidis
p.a.smith@soton.ac.uk,
n.tzavidis@soton.ac.uk
Department of Social Statistics & Demography
SO17 1BJ Southampton