

www.makswell.eu

Calculating regional price indices to improve poverty measurement

Monica Pratesi & Paul A. Smith



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 770643

PARTNERS

Outline

- Improving poverty studies
 - Spatial Price indices
 - Data: scanner data and HES
 - Methods: PPPs (CPD model) and Small area estimation of poverty rates
- The impact of cost of living on poverty incidence
- Regional price indices
 - Temporal and spatial
 - Concepts and challenges in estimation
 - Small area estimation of expenditure weights
- Quality measures for price indices

Improving poverty indices

Why:

SDGs monitoring- Policy actions

- *measures related to the places where people live*
- *measurement of cost of living and its spatial variation at local level: SPLs, PPPs*
- *extending the geographical notation of poverty indices to have local measures*

Who:

ASESD Centre “Camilo Dagum”: Caterina Giusti, Stefano Marchetti, Luigi Biggeri, Tiziana Lauretillaria Benedetti, Gaia Bertarelli, Francesco Schirripa Spagnolo, , Federico Polidoro and price unit staff (Istat),

What:

The proposed **methodology** is applicable in European countries as it is based on *current sample surveys* as EU-SILC and HES to produce local indicators of monetary relative poverty and on *scanner data on prices of RTCs*, that are generally available for NSIs in western countries

Model-based methodology

CPD models – SAE models

sub-regional level - NUTS3 level in European classifications

Experiment: Italy 2018 – NUTS3: Italian Provinces

Scanner data – ACNielsen - prices

- *Comparability*: scanner data of the modern distribution: millions of prices for thousands of products (GTIN Global Trade Item Number code).
- *Coverage*: supermarkets and hypermarkets, especially for food, beverages and personal and home care products.
- *Representativeness*: turnover weights

GTIN, prices are calculated taking into account turnover and quantities:

$$\text{weekly price} = \text{weekly turnover} / \text{weekly quantities}$$

Monthly prices are calculated with arithmetic mean of weekly prices weighted with quantities.

Italy: 2017, 2018, 103 over 107 provinces (not rural areas), *sample of outlets (1800)*, 95% of modern retail chain distribution that covers 55.4% of total retail trade distribution for this category of products, no perishables and seasonal products, 10,5% in terms of the total expenditures of families for the consumption. In addition, this share is not uniform across the Italian territory

Scanner Data

- **Strengths:**
 - impressive coverage of transactions,
 - information on: sales; expenditure; quantities; and quality,
 - very detailed information on characteristics of products sold (brand, size and type of outlet) provided at GTIN (GlobalTrade Item Number) code level,
 - millions of prices for thousands of products identified by the GTIN code.
- **Weaknesses:**
 - deep data cleaning is needed,
 - several outliers,
 - unit value price available only after dividing total turnover for that item by the total quantity sold,
 - only 5 out of 12 ECOICOP,
 - local markets and small shops are excluded.

In particular, in the next application at SAE level we use only ECOICOP 01 because it is possible to find the same products in HES only for ECOICOP 01.

HES– ESS-Istat – Consumption expenditure

- *Comparability*: conducted in all countries members of the European Union and by ISTAT in Italy .
- *Coverage*: all the expenses incurred by resident families to purchase goods and services intended for consumption
- *Representativeness*: more frequent expenditures referred to large consumption products

Two-weeks Diary: 7 outlets (traditional shop, open market and street vendors, hard discount, hypermarkets and supermarkets, department, stores and outlet chains, farm or direct producer, internet) and 25 products.

In Italy HES is the survey used to produce the **official estimates of absolute and relative poverty**.

Italy: 2017, sample survey 20000 units in about 540 municipalities, two stages (systematic sampling 1st stage, quota sampling 2nd stage), proportional to the universe of the adult Italian population, stratified by geographical distribution and size of the municipalities of residence – CAPI mode

HES – Household Expenditure Survey

Poverty Line: At national level the PL for households of two components is set equal to the per-capita mean Monthly Consumption Expenditure (MCE) at country level. To take into account the existence of economies of scale in consumption within households, the poverty line is then adjusted by using the Carbonaro scale (Istat, 2010).

Head Count Ratio: relative measure of poverty incidence, by defining for each household an indicator variable which takes value 1 if the Monthly Consumption Expenditure (MCE) of the household is less or equal the poverty line, value 0 otherwise. The values are then averaged by using the sample weights.

$$nPL = \frac{\sum_{i=1}^m \sum_{j=1}^{n_j} CE_{ij} w_{ij}}{\sum_{i=1}^m \sum_{j=1}^{n_j} a_{ij} w_{ij}}$$

w_{ij} : sample weights, a_{ij} hh size

$$HCR_{ij} = I(CE_{ij} \leq PL \cdot s_{ij})$$

s_{ij} : values of Carbonaro scale

$$HCR_i = \frac{\sum_{j=1}^{n_j} HCR_{ij} w_{ij}}{\sum_{j=1}^{n_j} w_{ij}}$$

Methods – toward local PPPs, SPIs, poor specific indexes

WB approach: comparability «like to like», aggregation of price data at item level, which leads to price comparisons at BH level. To compute provincial PPPs, we used as already mentioned the CPD model according to the approach followed by the World Bank. To aggregate the results from BH level comparisons to higher level aggregates Fisher price index



PPP produces an inflation rate that is equal to the price of the basket of goods at one location divided by the price of the basket of goods at a different location
 PPPs for the first quintile of the price distribution at regional level
 Reference Italy=100 national average

ASESD approach: The approach considers the unit value prices from the consumer side (or point of view), weighted mean price per gr. or ml. for products in ECOICOP-8-digit j and province I, basket of elementary products that the consumers of each province have really purchased , CPD model estimated by weighted LS



Country Product Dummy Model – Dagum proposal

CPD model for mean price:

$$\ln p_{ij} = \sum_{i=1}^m \alpha_i D_i + \sum_{j=1}^k \beta_j G_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

CPD model for quantiles of the price distribution:

$$\ln Q_{q,ij} = \sum_{i=1}^m \gamma_i D_i + \sum_{j=1}^k \beta_j G_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

p_{ij} : mean price of goods type j in area i , the goods type are aggregate using the COICOP 8 digit, the prices are computed in euro per gram or milliliter, the mean price is computed taking into account the sample design

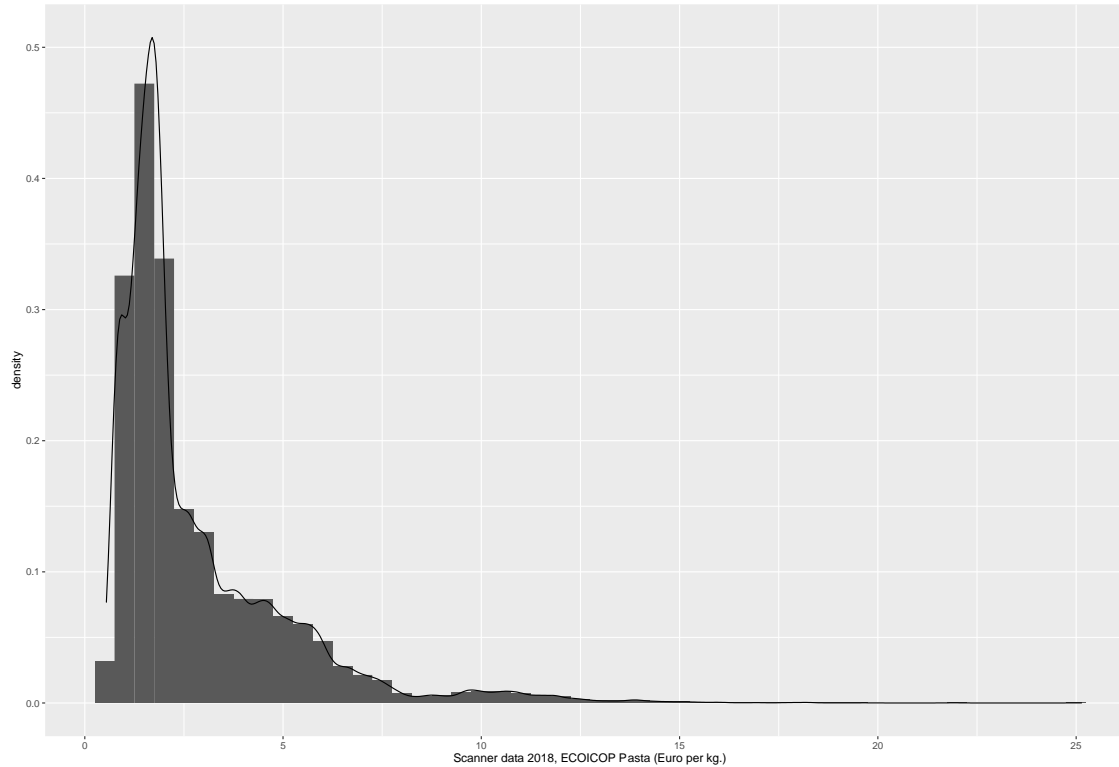
D_i : dummy vector for area i

G_j : dummy vector for goods of type j

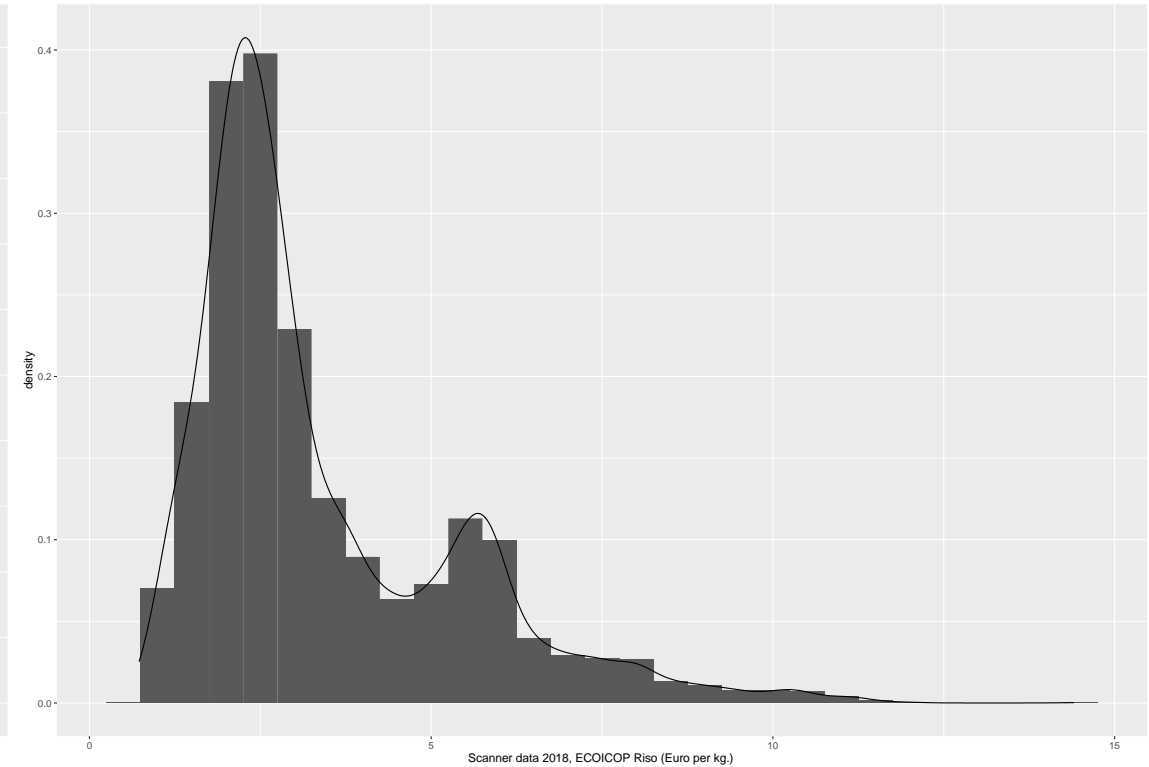
e^{α_i} is the Spatial Price Index (SPI) for area i respect to the base area, SPIs reference area is Italy

e^{γ_i} is the modified SPI for area i respect to the base area, it is an index for the low part of the distribution of prices

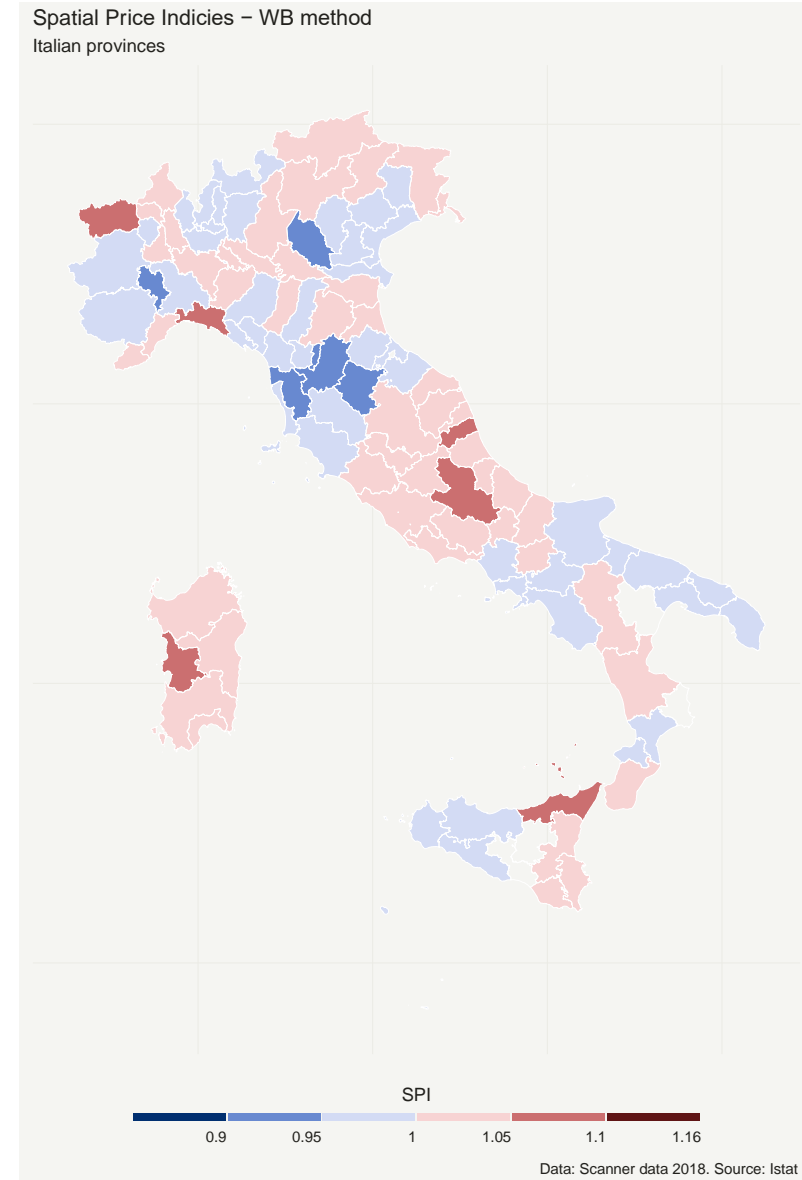
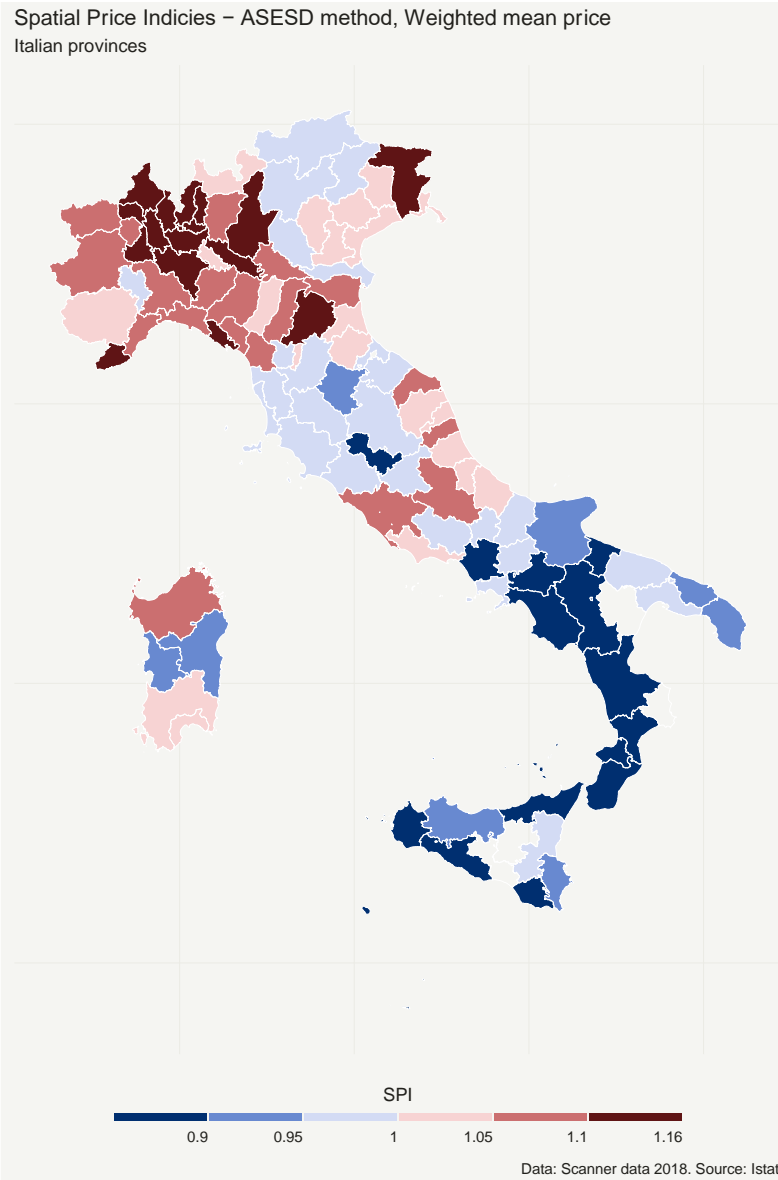
Ex. Distribution of prices



Prices for pasta (€/Kg)				
Min.	1st Q.	Med.	3rd Q.	Max.
0.54	1.42	1.9	3.6	25.14



Prices for rice (€/Kg)				
Min.	1st Q.	Med.	3rd Q.	Max.
0.73	2.1	2.7	4.43	14.4



Methods: Local PLs and local HCRs

Local poverty lines

national PL modified by SPIs:

$$nPL_i^* = nPL \times (\lambda_i SPI_i + 1 - \lambda_i)$$

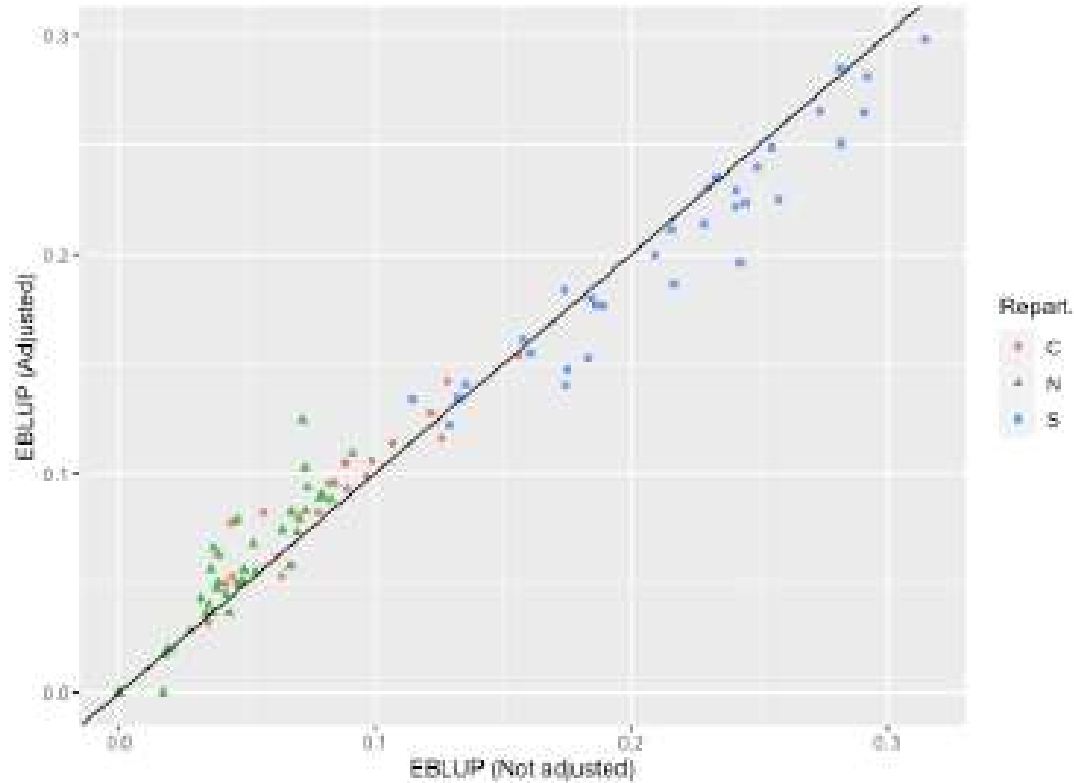
λ_i is the estimated (SAE) share of food consumption in province i (HES)

Local HCRS

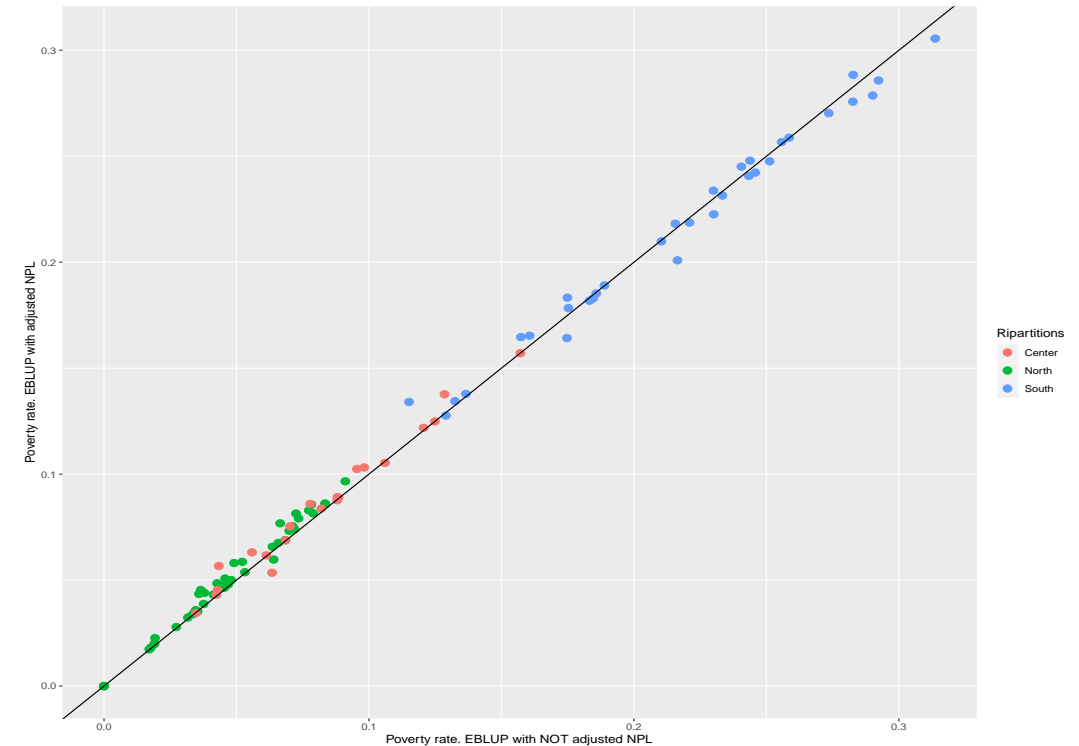
SAE Fay and Herriot area level model for the HCR at provincial level using the adjusted nPL_i^*

Auxiliary data from Italian tax agency database 2017

The impact of cost-of-living on poverty incidence

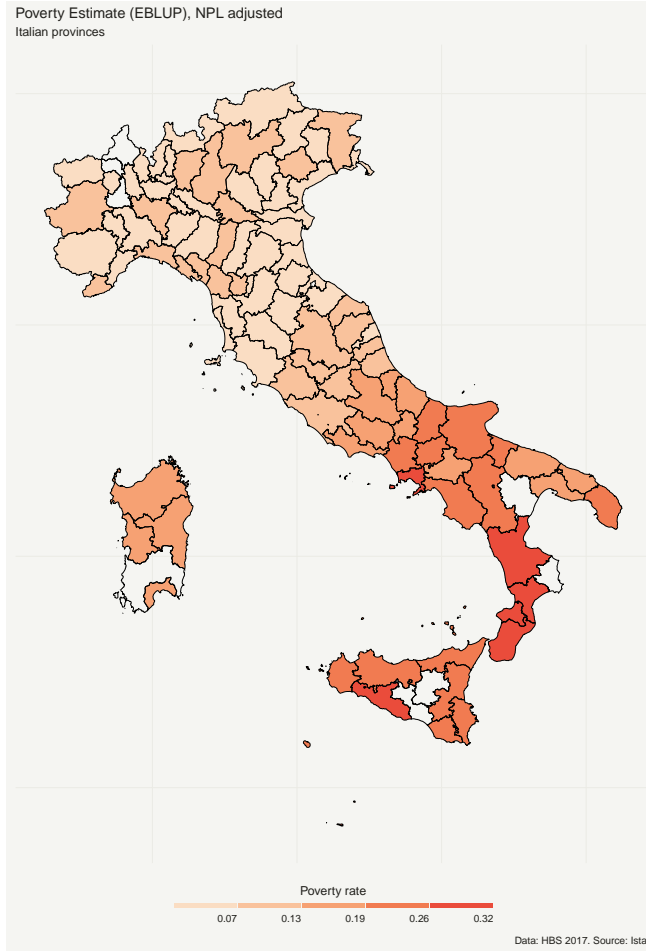


Poverty rate at provincial level in Italy: provincial EBLUPs estimates using the national PL adjusted with the SPI($Q_{0.2}$) and SRIs vs EBLUPs estimates using the not adjusted national PL.



Poverty rate at provincial level in Italy: provincial EBLUPs estimates using the SPI($Q_{0.2}$) adjusted vs not adjusted national PL.

The impact of cost-of-living on poverty incidence



Poverty estimates by FH model – NPL adjusted

- 1 - **price distributions** for groups of items (BHs or other groupings) obtained by scanner data are a valuable information. The first quantiles (quintiles, or deciles) of these distributions are mimicking the prices paid by the poor.
- 2 - SPIs for **housing Rents** (SPI Rs) obtained by HES or by archives of Revenue and Tax Agencies.
- 3 - the prices from scanner data may be affected by the **price policy of the RTCs**, this can have a uniform effect at the country level, smoothing the subregional variability of the prices. – first price products
- 4 - sources of **uncertainty** affecting the whole process, probability and not probability surveys

References

- Eurostat** and OECD (2012). Eurostat-OECD Methodological Manual on Purchasing Power Parities. Luxembourg: Publications Office of the European Union.
- Biggeri**, L. and T. **Laureti** (2018). Publications, experiments and projects on the computation of spatial price level differences in Italy. Technical report, Paper presented at the 3rd Taskforce meeting the ICP, World Bank held the 27th September 2018, Country case studies: Italy.
- Fay**, R. and R. **Herriot** (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association* 74, 269-277.
- T. **Laureti** and F. **Polidoro**. Testing the use of scanner data for computing sub-national purchasing power parities in Italy. In *Proceeding of 61st ISI World Statistics Congress*, Marrakech, 2017
- Marchetti**, S., G. **Bertarelli**, L. **Biggeri**, G. **Giusti**, M. **Pratesi**, and F. Schirripa Spagnolo (2019). Small area poverty indicators adjusted using local price indexes. Italian Conference on Survey Methodology (ITACOSM), 5-7 June, 2019, Florence.
- Pratesi M.**, (2016) (ed) *“Analysis of Poverty Data by Small Area Estimation”*, New York, Wiley.
- Renwick**, T., B. **Aten**, E. **Figuerola**, and T. **Martin** (2014). Supplemental poverty measure: A comparison of geographic adjustments with regional price parities vs. median rents from the American community survey. Technical report, Bureau of Economic Analysis.
- Suits**, D. (1984). Dummy variables: Mechanics v. interpretation. *Review of Economics and Statistics* 66, 177-180.
- Istat (2010). La differenza nel livello dei prezzi al consumo tra i capoluoghi delle regioni italiane. Roma, Italy: Istat, Italian national statistical office.

REGIONAL TEMPORAL PRICE INDICES AND QUALITY MEASURES

Paul A. Smith, University of Southampton

Outline

- Regional temporal price indices
 - Temporal and spatial
 - Concepts and challenges in estimation
 - Small area estimation of expenditure weights
- Quality measures for price indices

Spatial and temporal price indices

- Italian work based on PPP framework
 - Spatial indices – “region K = 100”
 - Measure difference in price level between regions
 - Important to adjust poverty lines based on purchasing power
- This work based on CPI framework
 - Temporal indices – “2019 = 100”
 - Measure differences in inflation between regions

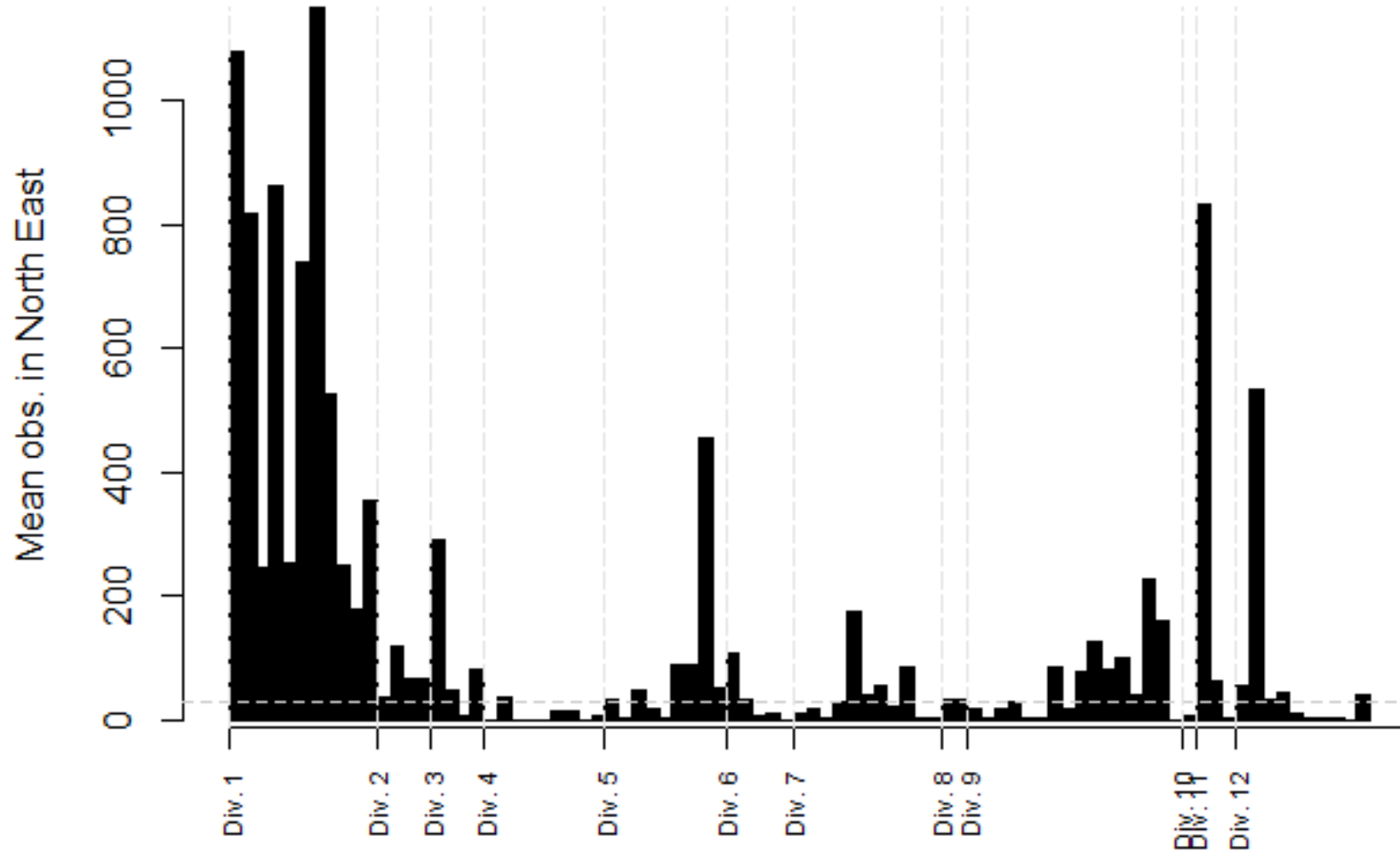
Concepts and estimation challenges

- Regional index: requires
 - regional basket: from HFS (=LCF in UK)
 - regional weights: from HFS
 - regional prices: from CPI price collection
- Regional sample sizes small
 - stability of basket and weights year to year
 - sufficient price quotes

Strata with few price quotes

	1	2	3	4	5	6	7	8	9	10+
North East	600	584	415	488	803	1,098	1,133	323	204	2,867
North West	381	497	413	251	151	128	151	169	226	6,560
Yorkshire and the Humber	381	397	303	209	264	351	387	438	502	5,714
East Midlands	422	247	164	127	226	361	525	717	794	4,929
West Midlands	307	357	373	375	338	446	324	399	618	5,235
East	616	200	327	449	209	172	260	308	378	5,979
London	219	97	192	275	302	336	465	492	469	6,365
South East	356	523	269	273	136	142	184	184	168	6,931
South West	316	220	316	444	460	293	349	367	442	5,707
Wales	838	247	328	397	723	1,251	1,001	400	193	2,917
Scotland	707	387	219	230	222	153	221	225	317	5,654
Northern Ireland	580	919	784	1,031	1,435	592	366	257	368	2,230
Total	5,723	4,675	4,103	4,549	5,269	5,323	5,366	4,279	4,679	61,088
Total percentage	5.4%	4.5%	3.9%	4.3%	5.0%	5.1%	5.1%	4.1%	4.5%	58.1%

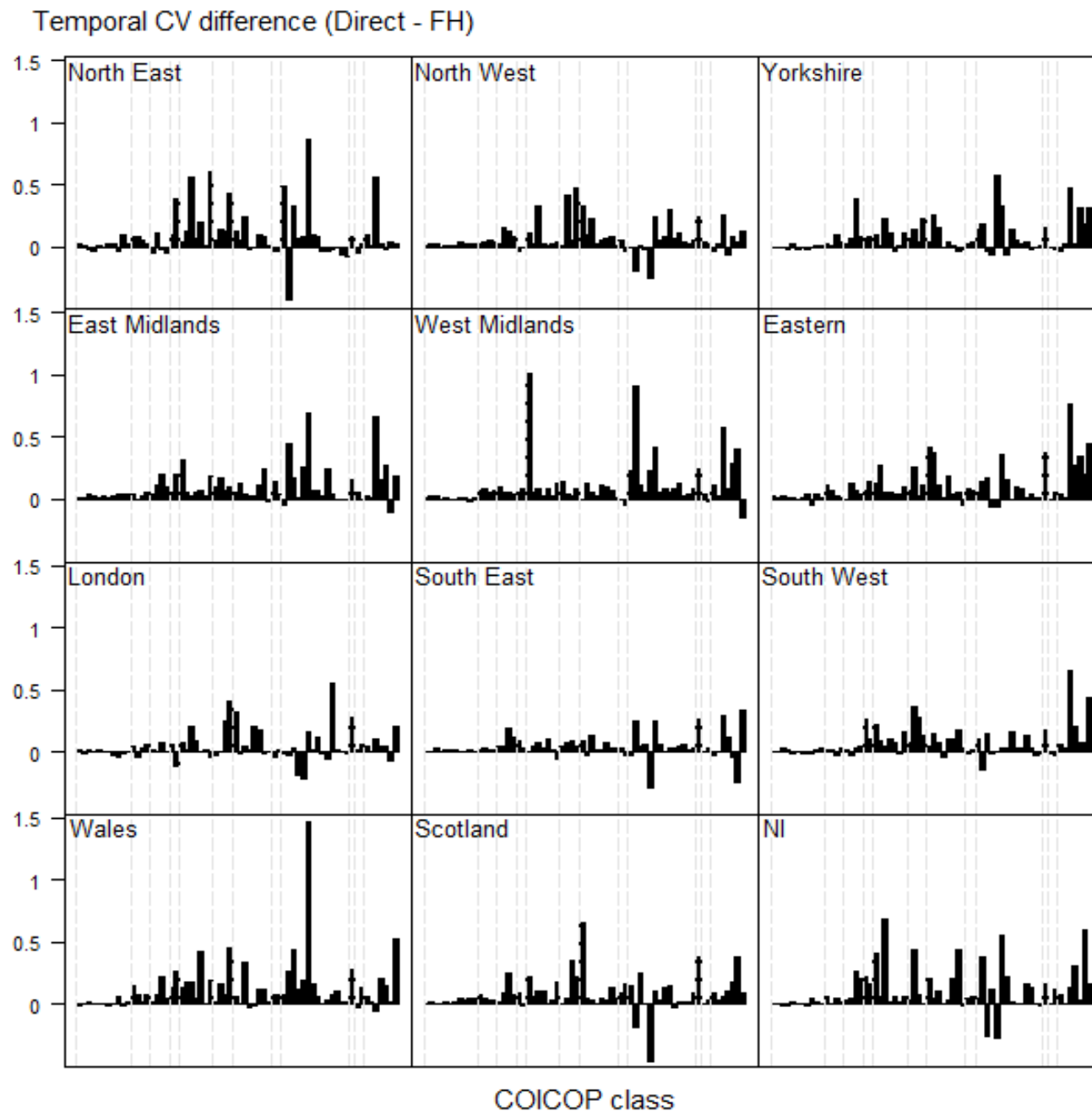
HFS observations by COICOP class

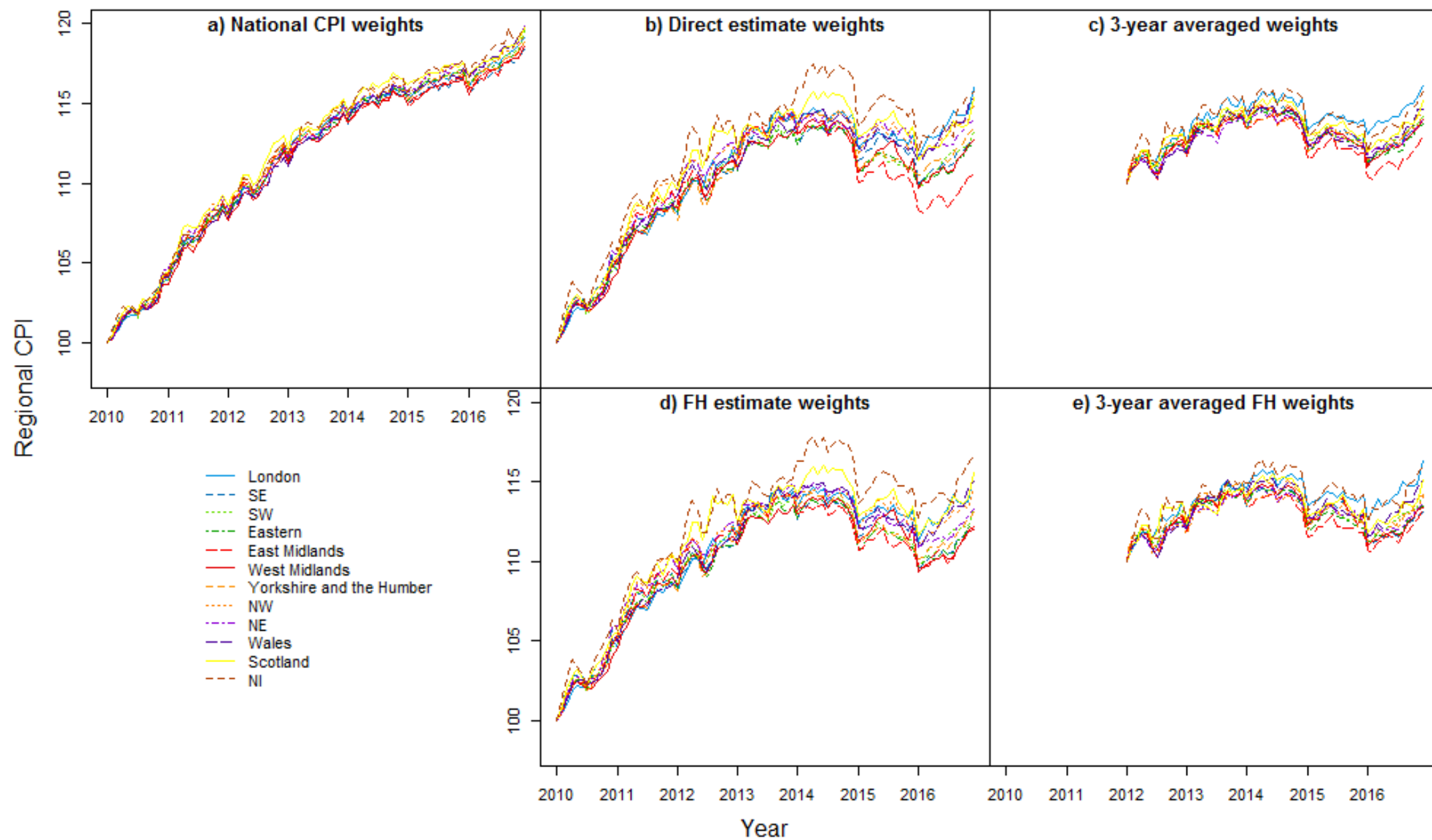


Small area models for expenditure

- Fay-Herriot (area-based) models
 - Many predictor variables from survey
 - Few (regional) observations – risk of overfitting

Improvement in cv for weights





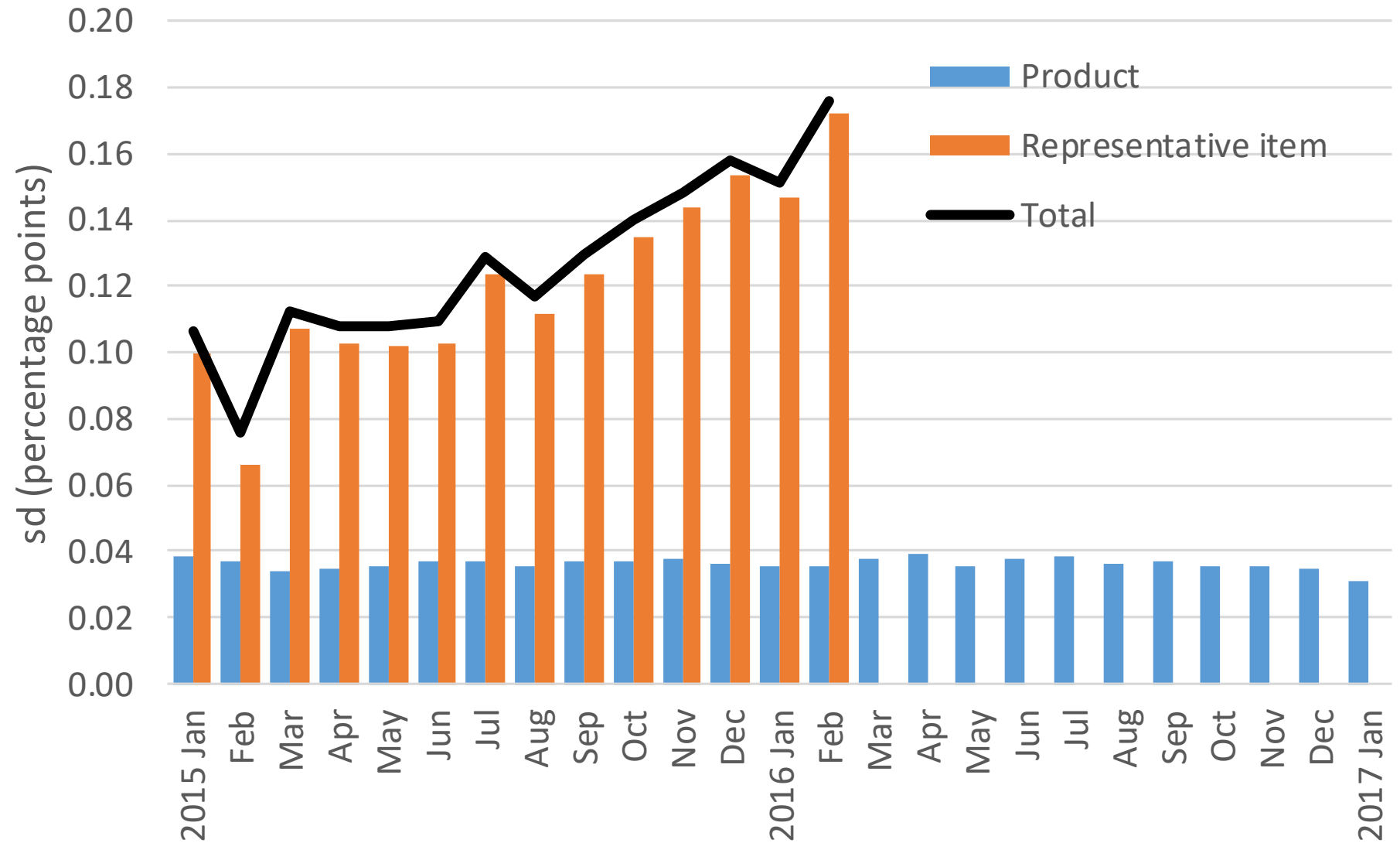
Quality measures for price indices

- Andersson *et al.* (1987b): “[t]he Swedish CPI has not been regarded, by its users, as an estimate of an unknown parameter. Rather, there seems to be a widespread agreement that the published value of the CPI is, by definition, the truth.”

Approaches to estimating CPI variance

- Design-based with Taylor linearisation
- Replication based
 - Balanced half samples (US)
 - Jackknife and/or bootstrap
- Model-based
 - Detailed modelling for best fit (Kott, Valliant in US)
 - Broader approach for robustness (Zhang in Norway)

Replication-based estimated sds for components of the UK CPI (provisional)



Total error for price indices

- Biggeri & Giommi (1987) framework:
 - estimated index
 - defined goal (“true value”) of the index
 - ideal goal

⇒ variances + biases ⇒ mse
- Calculation of components challenging
- Only Dalén (1995) calculates range of components for same index
 - Declines to calculate overall mse

Future research directions

- Small area estimation for baskets and weights
 - compositions, GSPREE estimator
 - Benchmarking to national indices
- Explore spatiotemporal index
 - Many complexities
- Comparison of variance approaches, esp model variances
- Apply quality measurement to regional indices
 - Total quality for national and regional indices