Measuring inflation through different sampling designs implemented on scanner data

A. Guandalini, A. Bernardini, C. De Vitiis, F. Inglese and M.D. Terribili
ISTAT – ITALY

Meeting of the Group of Experts on Consumer Price Indices
2 - 4 May 2016
Geneva, Switzerland

Istat

Summary

1. Aims of the presentation
2. The current sample strategy of the CPI survey
3. The new general sampling design for CPI survey
4. Different sampling strategies from Scanner Data
5. Results
6. Open issues and conclusions

1. Aims of the presentation

Comparing properties
of weighted and un-weighted higher level elementary price indices
when different sampling scheme are implemented
on scanner data (SD)

- Preliminary results:
  - Probabilistic versus Non-probabilistic
- New results:
  - Different probability sampling strategies
    - Allocation
    - Selection scheme
    - Aggregation price index formula

2. The current sample strategy of the CPI survey

1st stage

Municipalities

Forced by law

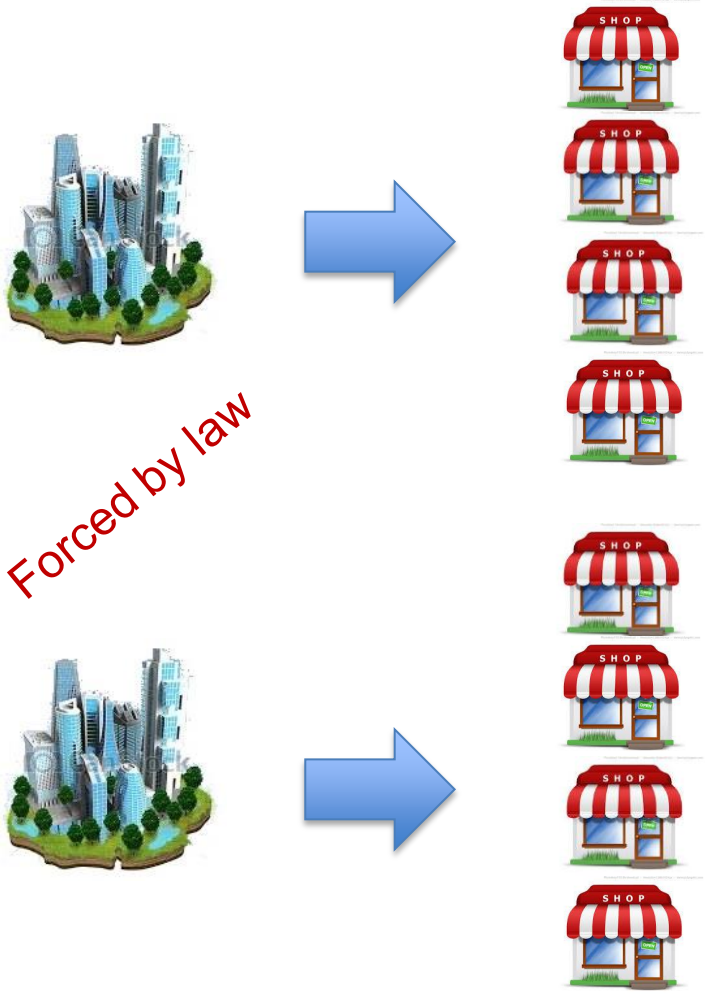
2. The current sample strategy of the CPI survey

1st stage

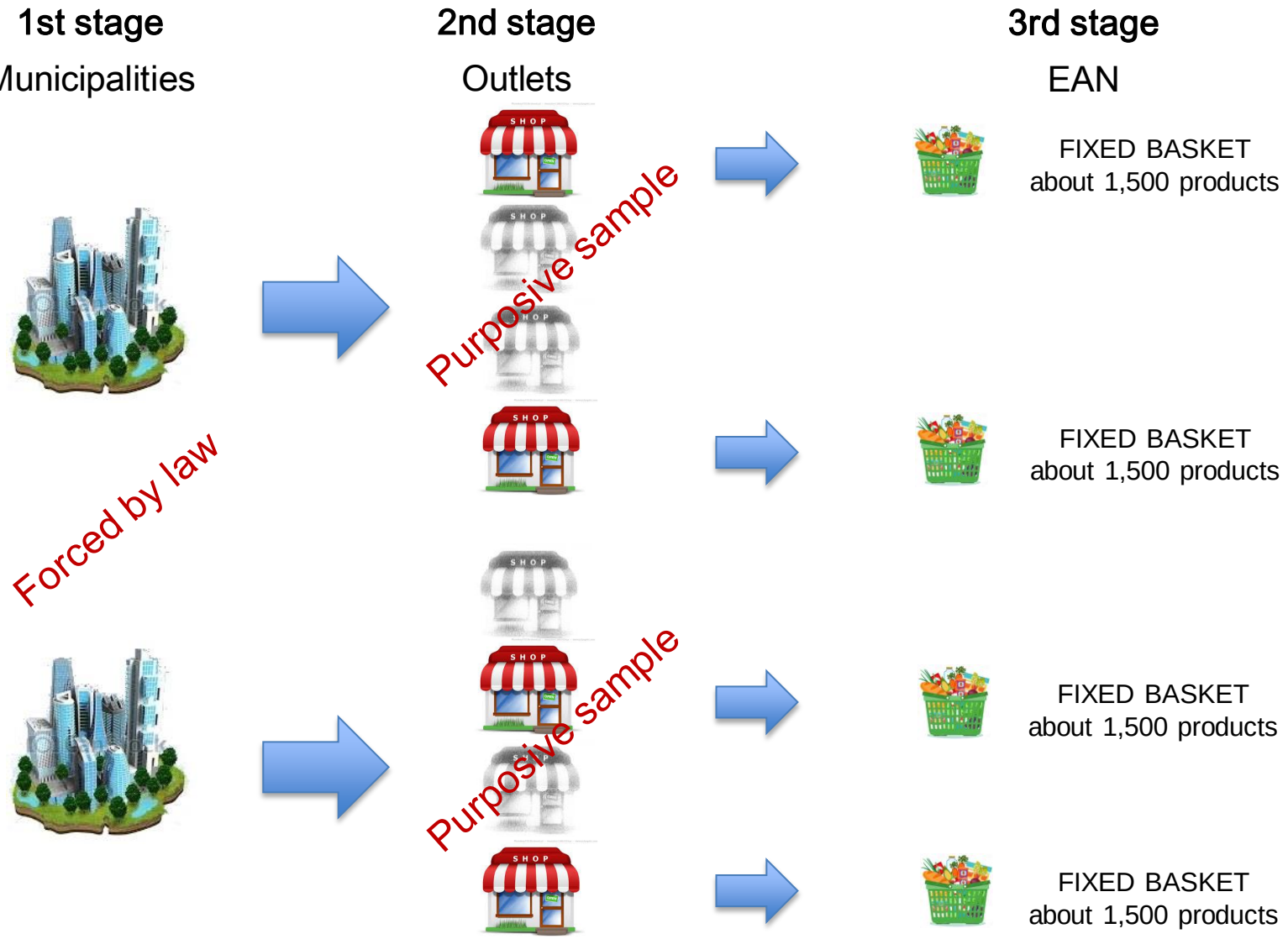
Municipalities

2nd stage

Outlets



2. The current sample strategy of the CPI survey



Istat

3. The new general sampling design for CPI survey

ASIA-LU

Register of

Local Unit Active Enterprices

3. The new general sampling design for CPI survey

ASIA-LU

Register of

Local Unit Active Enterprices

Modern distribution

Traditional distribution

3. The new general sampling design for CPI survey

ASIA-LU

Register of

Local Unit Active Enterprices

Probability sampling

Modern distribution

Food & Grocery

Other retail sector

Traditional distribution

3. The new general sampling design for CPI survey

ASIA-LU

Register of

Local Unit Active Enterprices

Modern distribution

Food & Grocery

Probability sampling

Other retail sector

Probability sampling

Traditional distribution

4. Different sampling strategies from Scanner Data

Operational framework

- Turin province
- 2014
- Six consumption segments (COICOP-6)
- Relevant week (one of the first three full weeks in a month)
- Cleaned dataset
  - Formal check (completeness of data)
  - Quality check (quantities sold and turnover)
  - “Mobile trimming” (inadmissible unit prices)
- Permanent series (only series always observed from December 2013 to December 2014)

PRELIMINARY EXPERIMENTS

Probability versus non-probability selection schemes

Comparison among:

- Cut-off selection of series based on thresholds of covered sales :

  60-80% coverage of all expenditure of the consumption segment in previous year, 2013

- Reference method

  most sold elementary items for each representative product (current fixed basket approach)

- Probability sampling – independent samples in consumption segments

PRELIMINARY EXPERIMENTS

Probability versus non-probability selection schemes

Comparison among:

- Cut-off selection of series based on thresholds of covered sales :

  60-80% coverage of all expenditure of the consumption segment in previous year, 2013

- Reference method

  most sold elementary items for each representative product (current fixed basket approach)

- Probability sampling – independent samples in consumption segments

MAIN RESULTS

- Probability sampling always brings more efficient estimates than non-probability selection scheme (i.e. most sold and cut-off)
- Sample scheme is not neutral with respect to the choice of aggregation formulas
- Sampling error varies among consumption segments

FURTHER EXPERIMENTS
Probability selection schemes

- Elementary price indices computed for 13 months with three aggregation formulas:
  - Jevons (unweighted),
  - Fisher
  - Lowe
- Montecarlo simulation (500 replications)
- Comparison among aggregation formulas performances under different sampling designs in terms of:
  - *Bias* $B(\hat{\theta}) = E\left[\hat{\theta}\right] - \theta$
  - *Variance – Relative sampling error* $RE(\hat{\theta}) = \frac{\sqrt{Var(\hat{\theta})}}{\hat{\bar{\theta}}}$

FURTHER EXPERIMENTS

Probability selection schemes

***Population* size** 50,000 EAN
121 outlets

Sample size 7,400 EAN

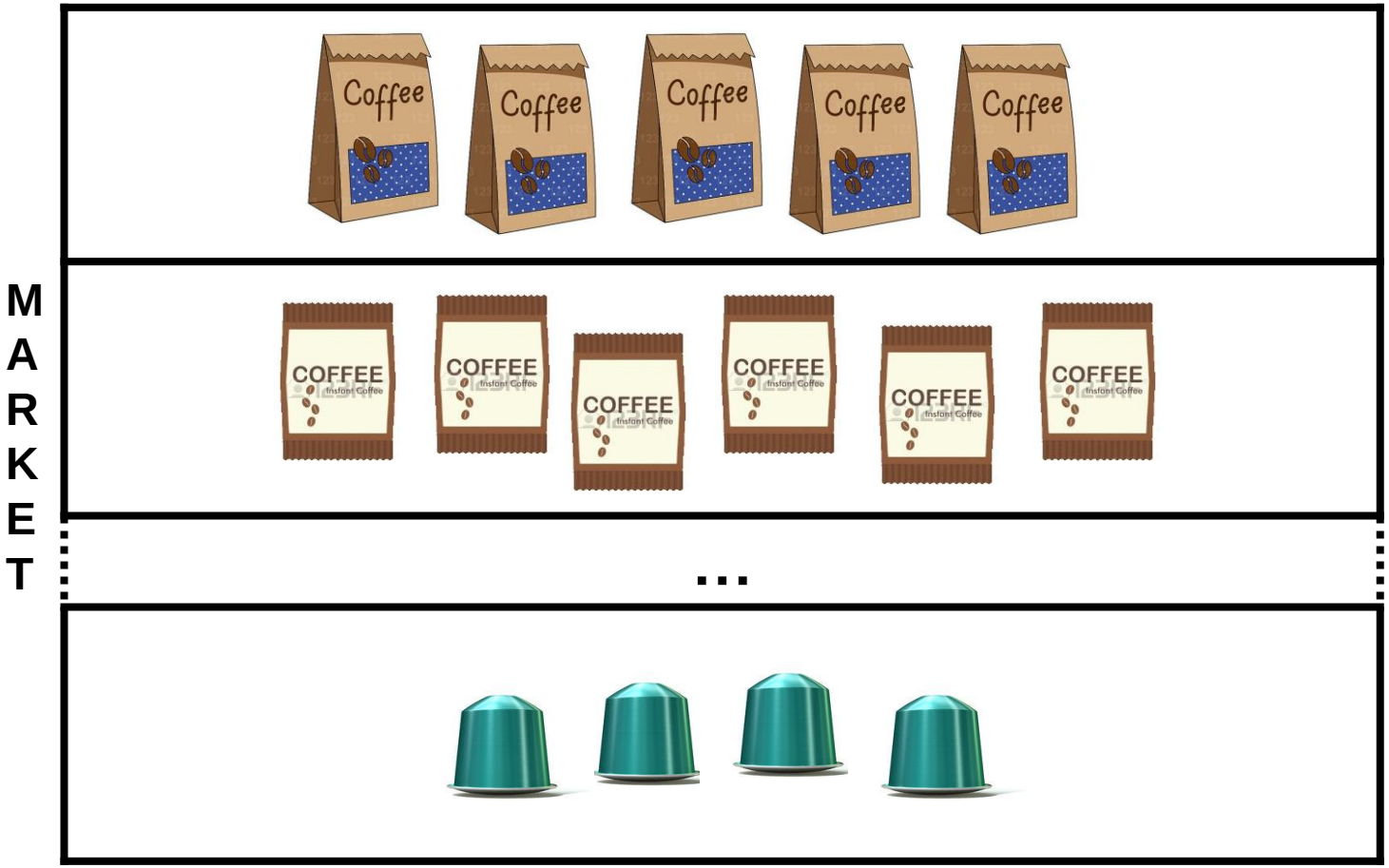
Sampling design	Level	Stratification	Allocation	Selection
Stratified	EAN	Market (ECR4)	Neyman	SRS
				PPS (turnover)
Stratified 1 stage	Outlet	Chain * type	Neyman	SRS
				PPS (turnover)
			Proportional	SRS
				PPS (turnover)
Stratified 2 stages	Outlet	Chain * type	Proportional	PPS (turnover)
	EAN	Market		SRS PPS (turnover)

Istat

Stratified sampling

- Stratification of EAN by market (ECR4)
- Neyman allocation with respect to price relative variability in 2013 in each stratum

SEGMENT = COFFEE

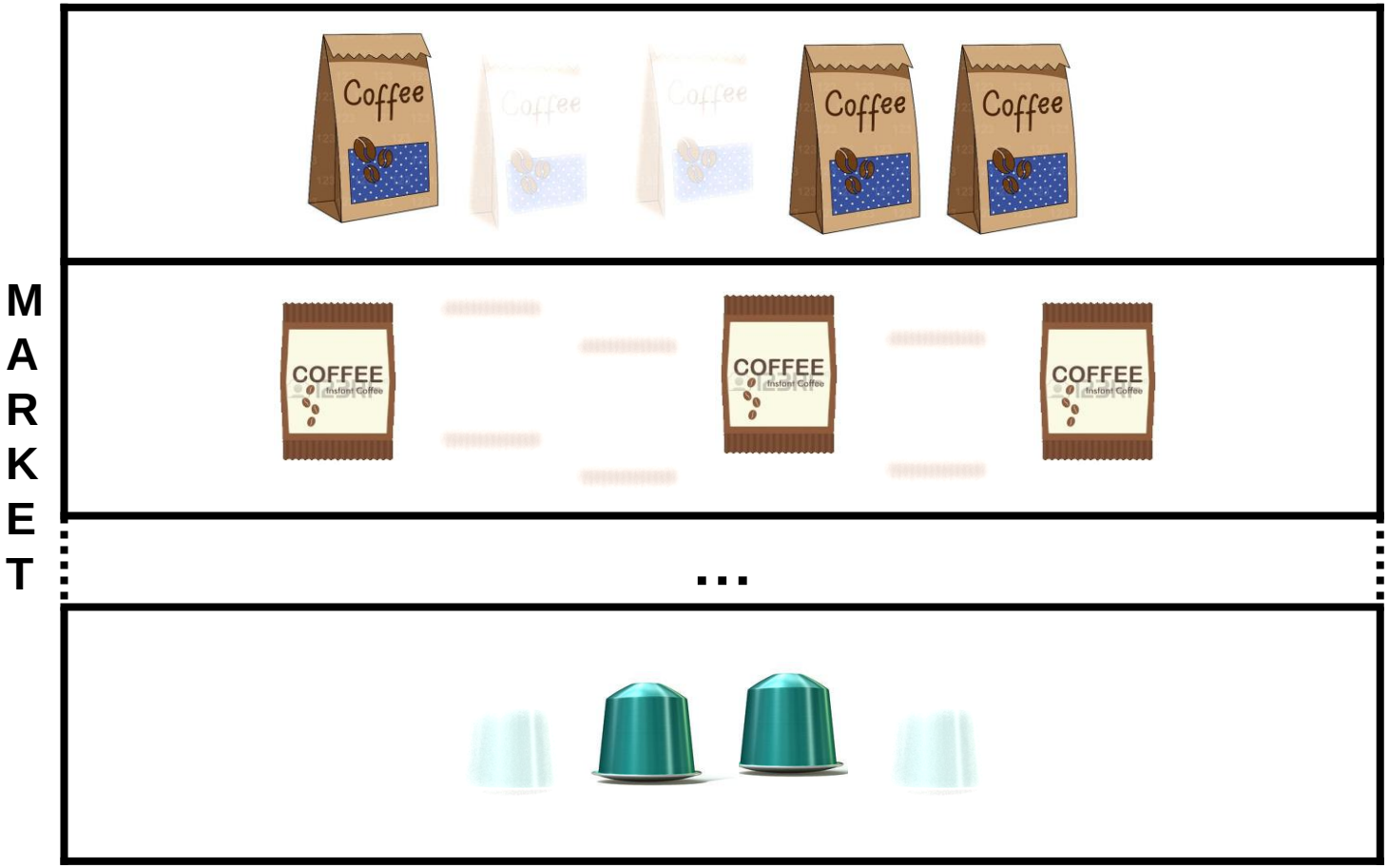


Istat

Stratified sampling

- Stratification of EAN by market (ECR4)
- Neyman allocation with respect to price relative variability in 2013 in each stratum

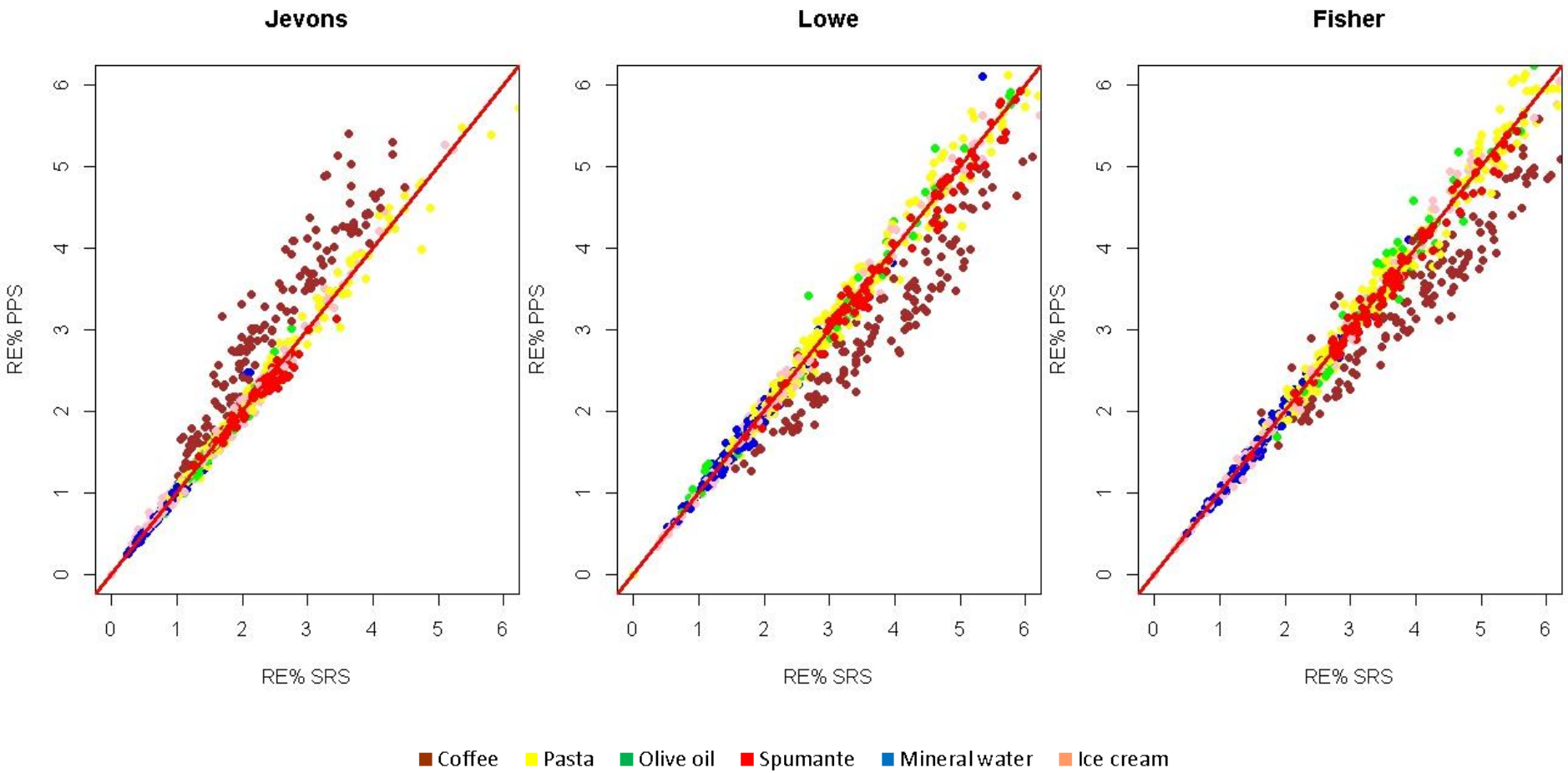
SEGMENT = COFFEE



Istat

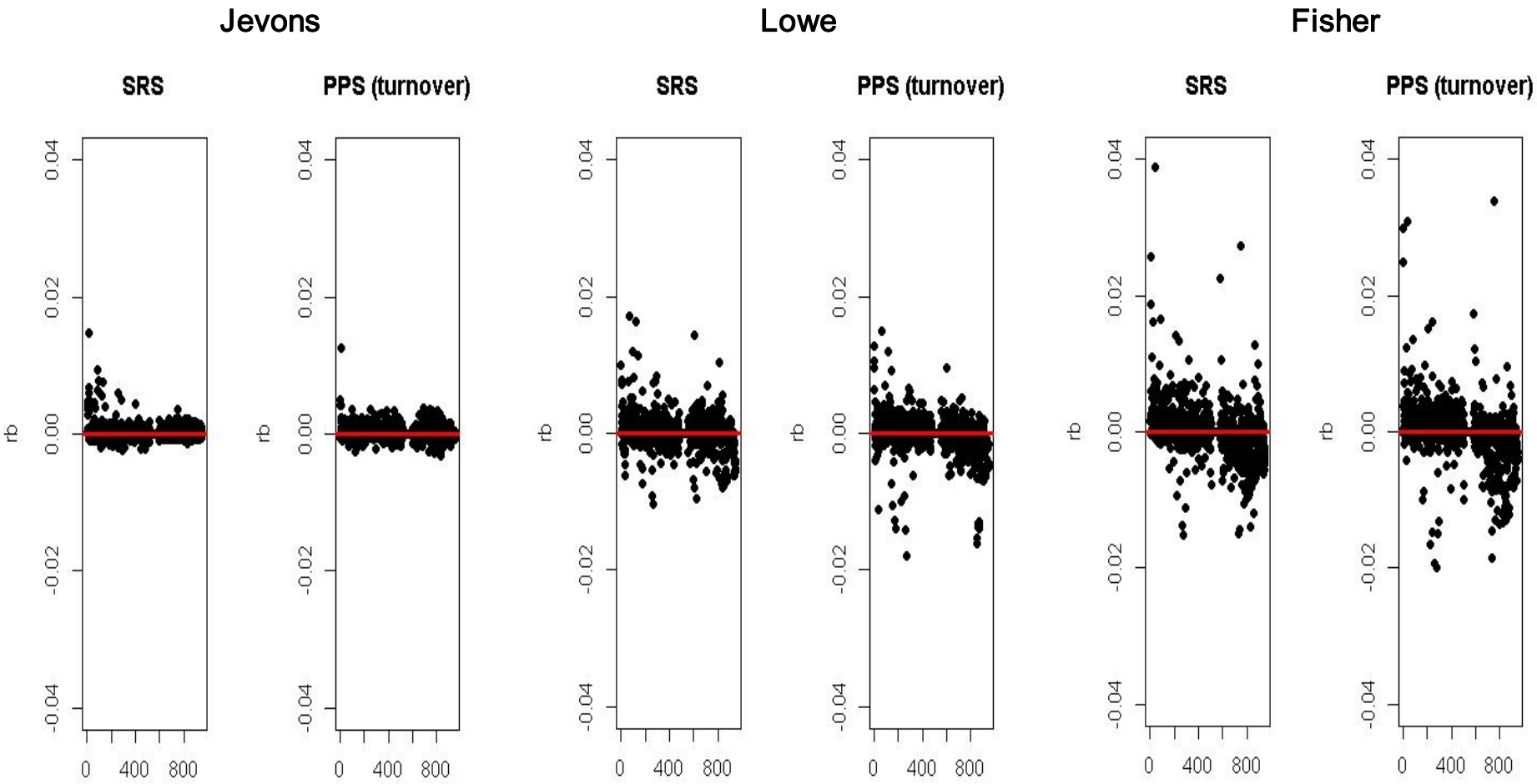
Stratified sampling

Comparison of RE% under SRS and PPS (size=turnover) for different consumption segments (COICOP-6)



Stratified sampling

Relative bias of monthly index estimates when EAN are selected with SRS or with PPS

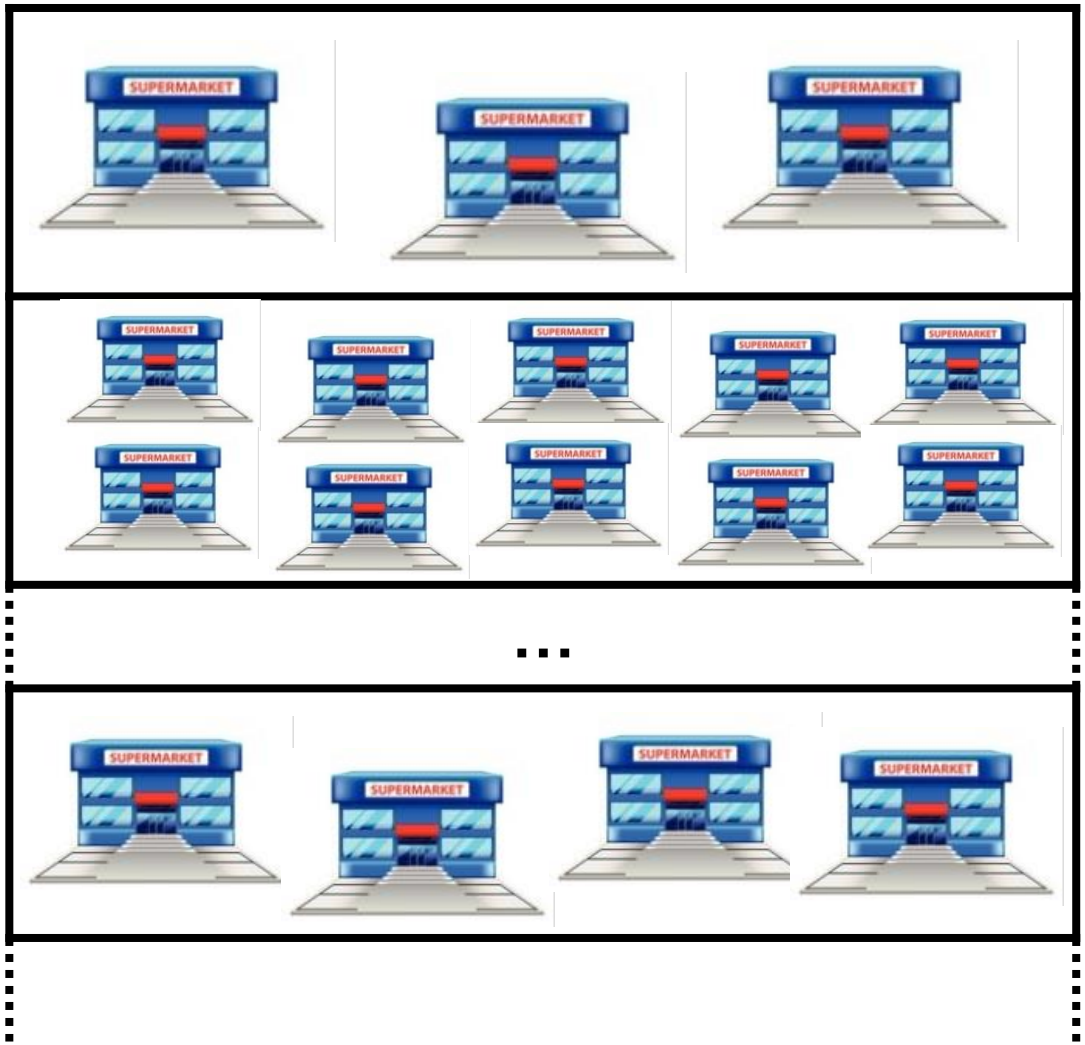


Istat

Stratified 1 stage

- Stratification of outlets by chain and type

Chain	Type
coop	IPER MARKET
	SUPER MARKET
…	…
	IPER MARKET
	…

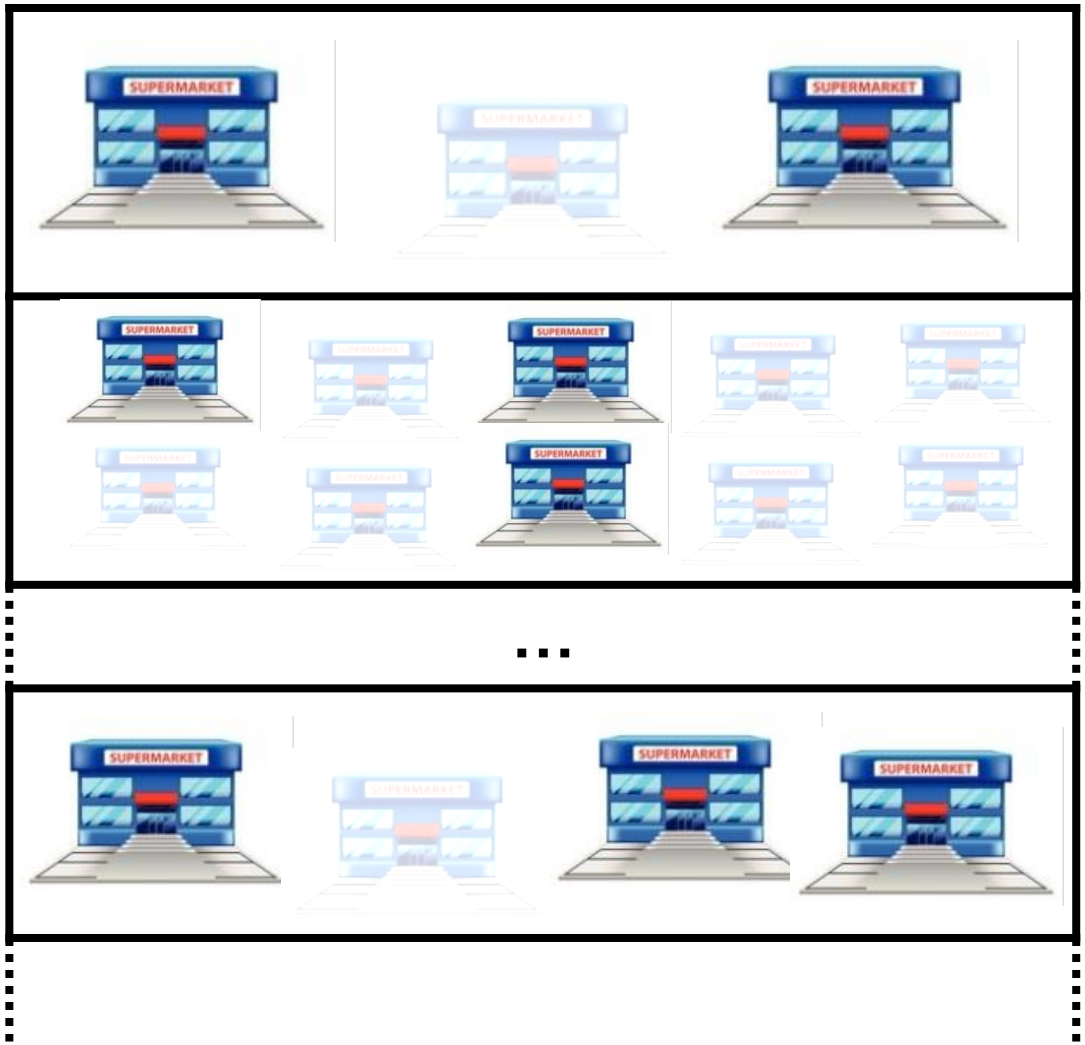


Istat

Stratified 1 stage

- Stratification of outlets by chain and type

Chain	Type
coop	IPER MARKET
	SUPER MARKET
…	…
	IPER MARKET
	…



Istat

Stratified 1 stage

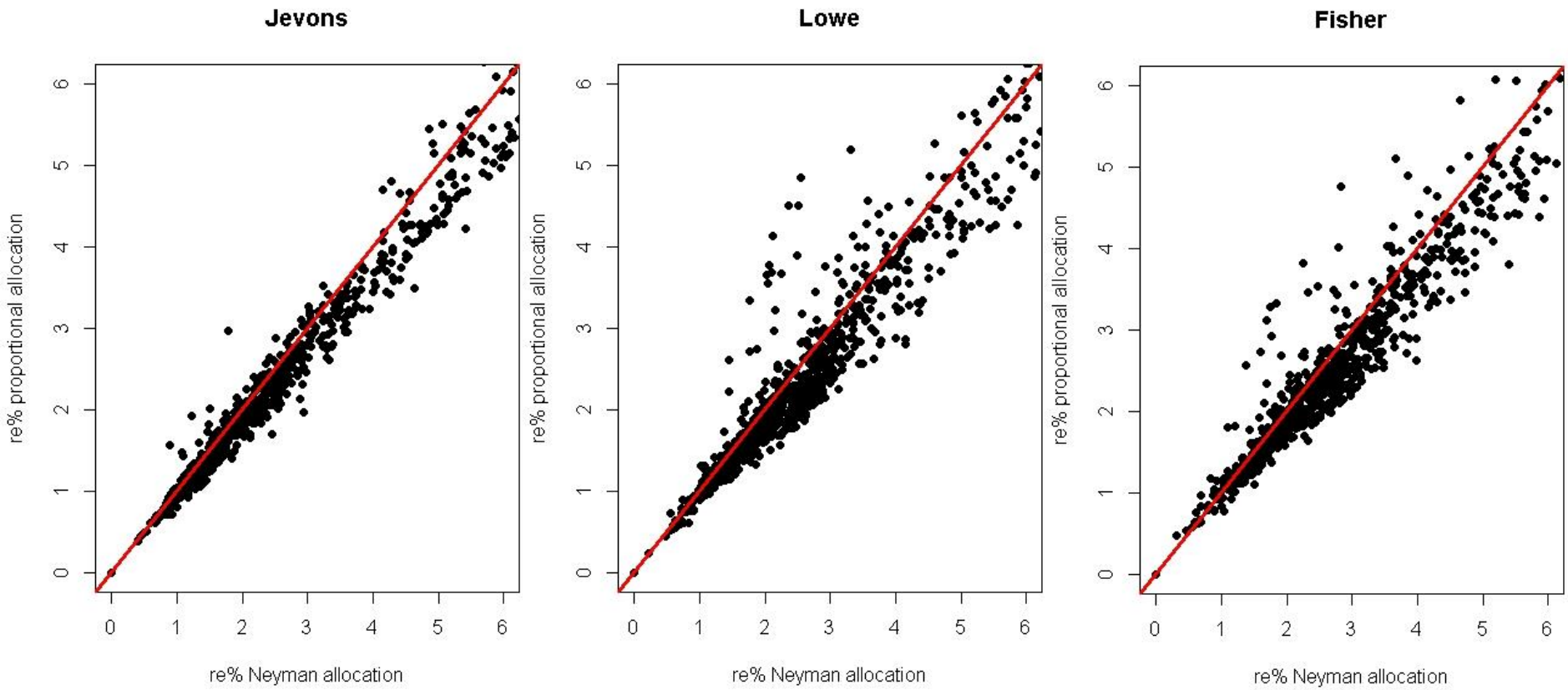
- Stratification of outlets by chain and type

Chain	Type
coop	IPER MARKET
	SUPER MARKET
…	…
	IPER MARKET
	…

Istat

Stratified 1 stage sampling

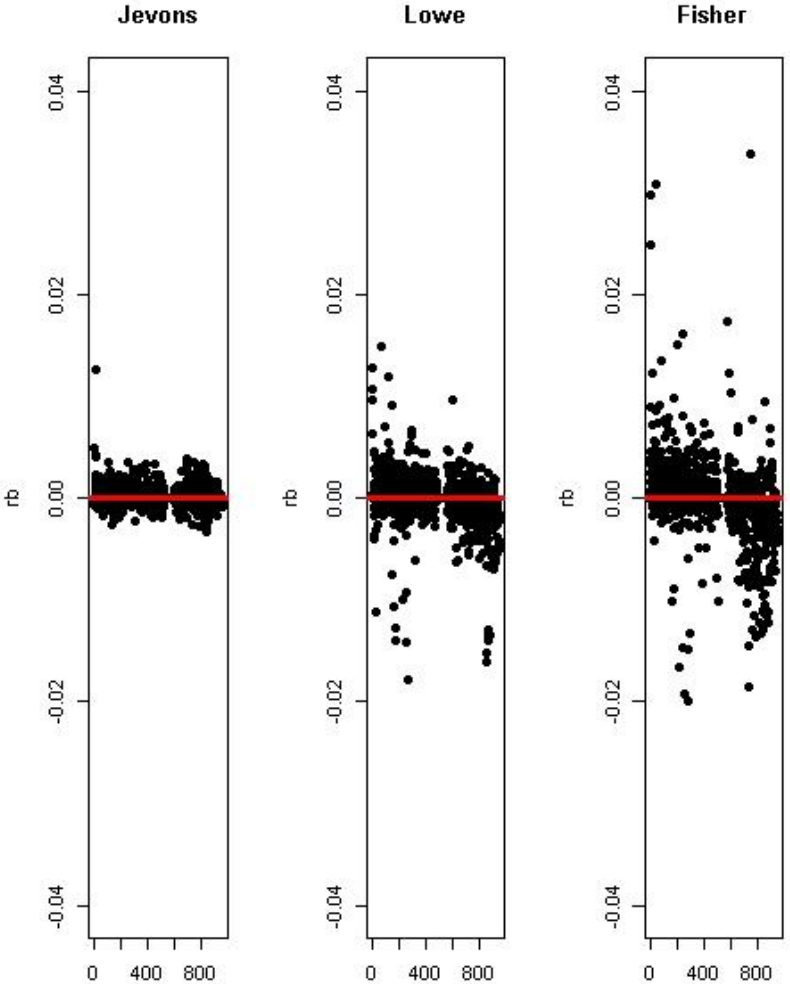
Comparison of RE% for PPS selection of outlets allocated with Neyman and Proportional allocation



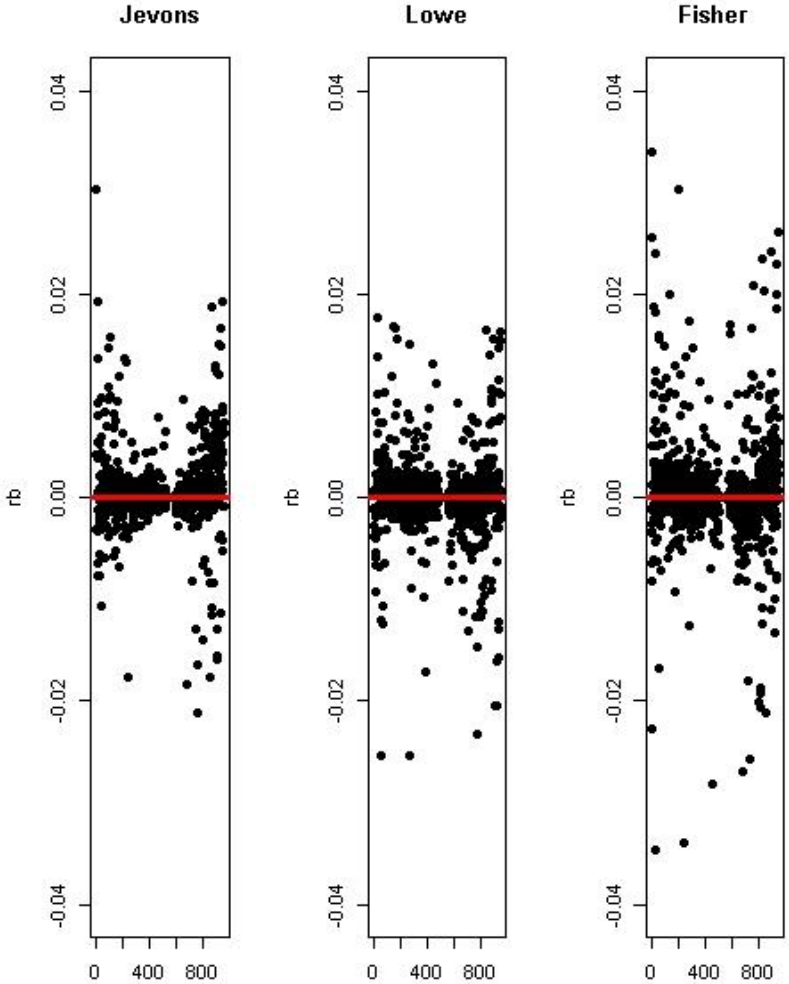
Istat

Comparison between the relative bias in Stratified sampling and in Stratified 1 stage sampling

Stratified sampling



Stratified 1 stage sampling

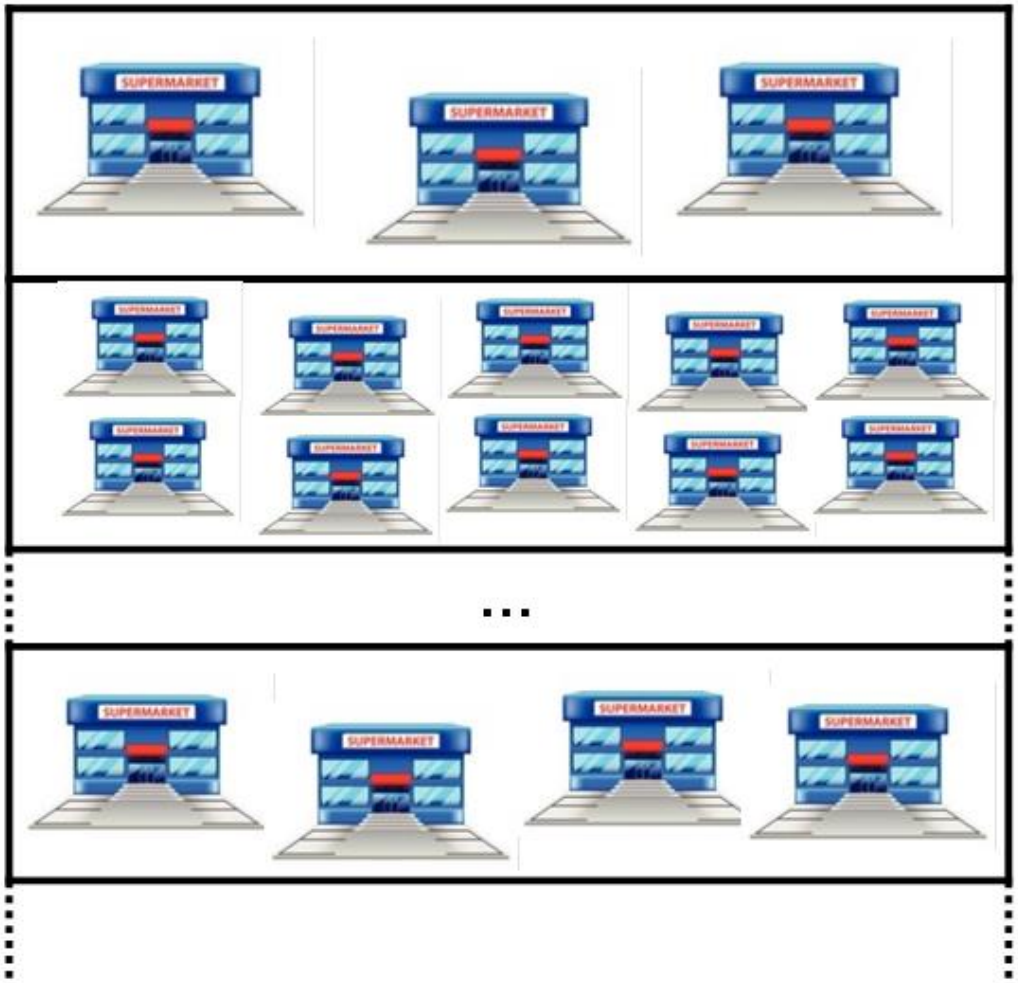


Istat

Stratified 2 stages sampling

- Stratification of outlets by chain and type
- Stratification of EAN by market (ECR4)

Chain	Type
coop	IPER MARKET
	SUPER MARKET
...	...
Carrefour	IPER MARKET
	...



Istat

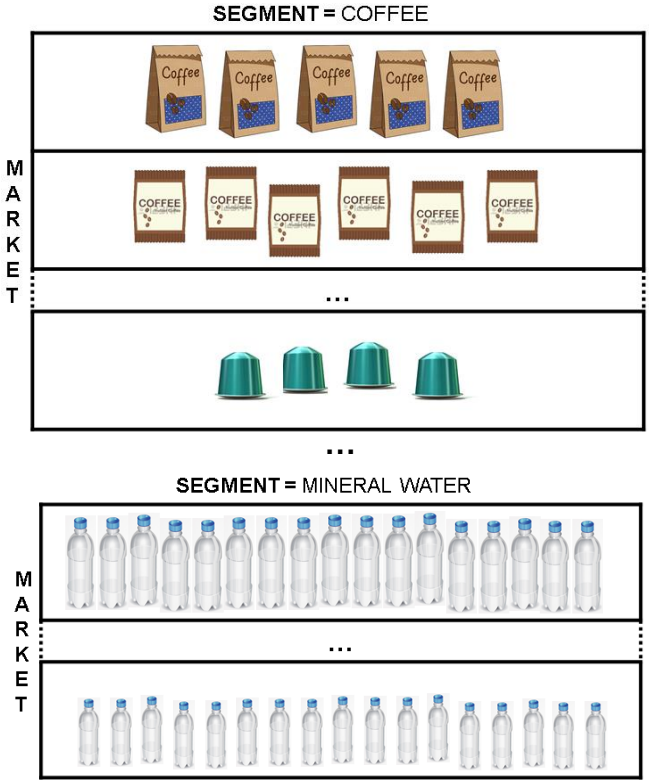
Stratified 2 stages sampling

- Stratification of outlets by chain and type
- Stratification of EAN by market (ECR4)

Chain	Type
coop	IPER MARKET
	SUPER MARKET
...	...
Carrefour	IPER MARKET
	...

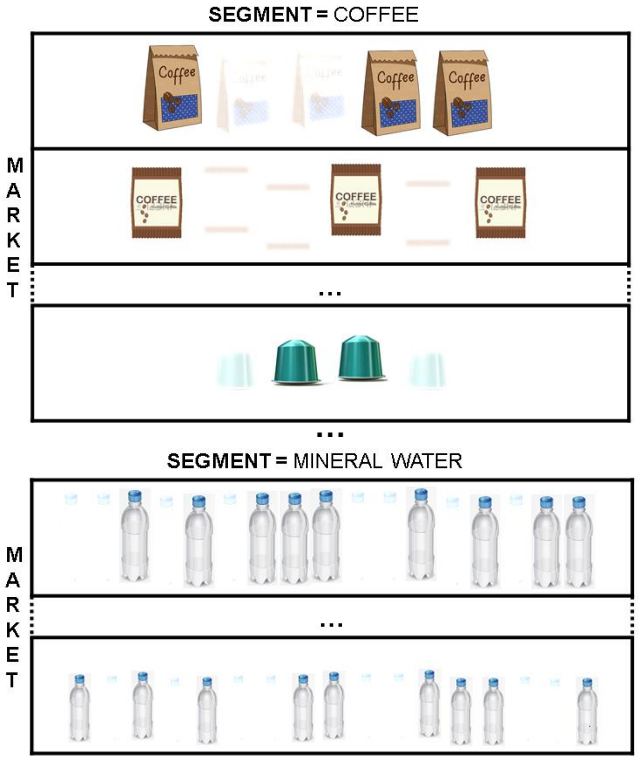
Stratified 2 stages sampling

- Stratification of outlets by chain and type
- Stratification of EAN by market (ECR4)



Stratified 2 stages sampling

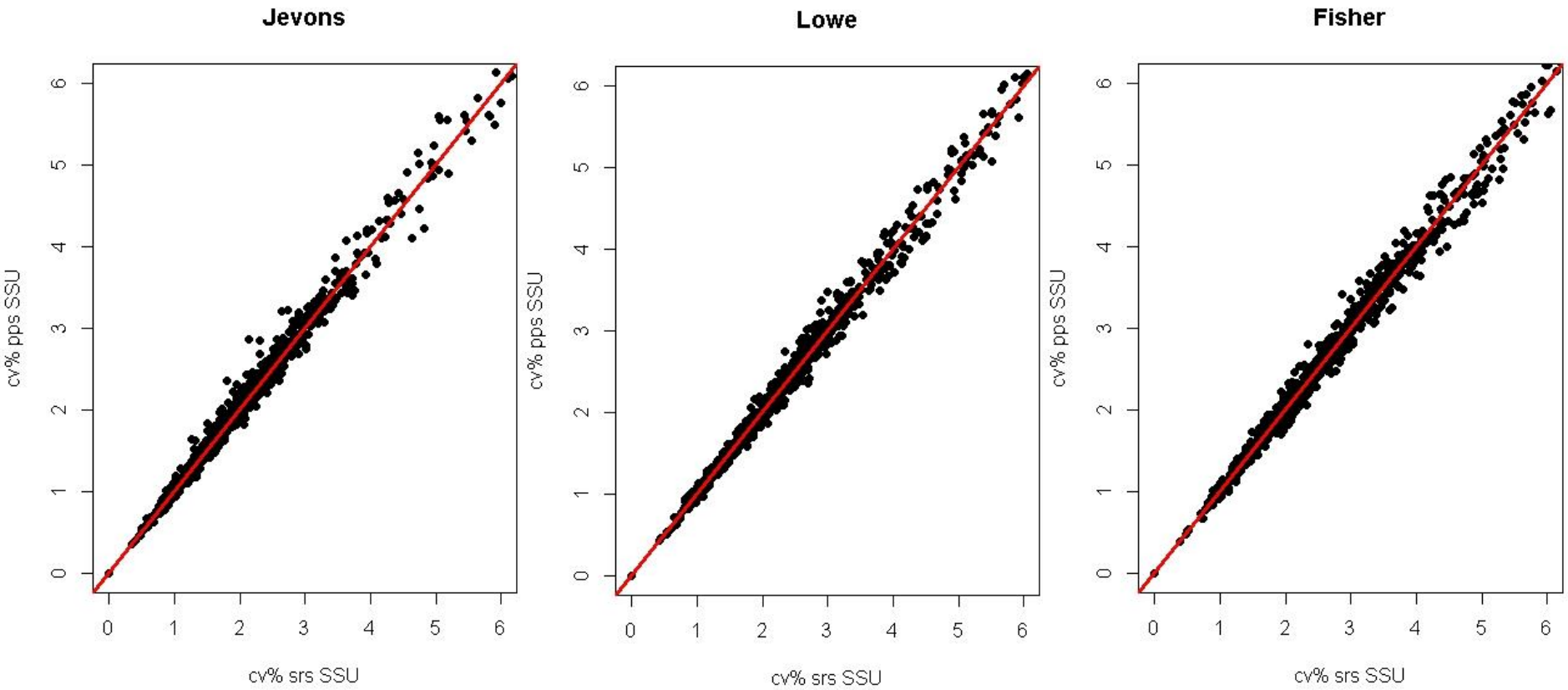
- Stratification of outlets by chain and type
- Stratification of EAN by market (ECR4)



Istat

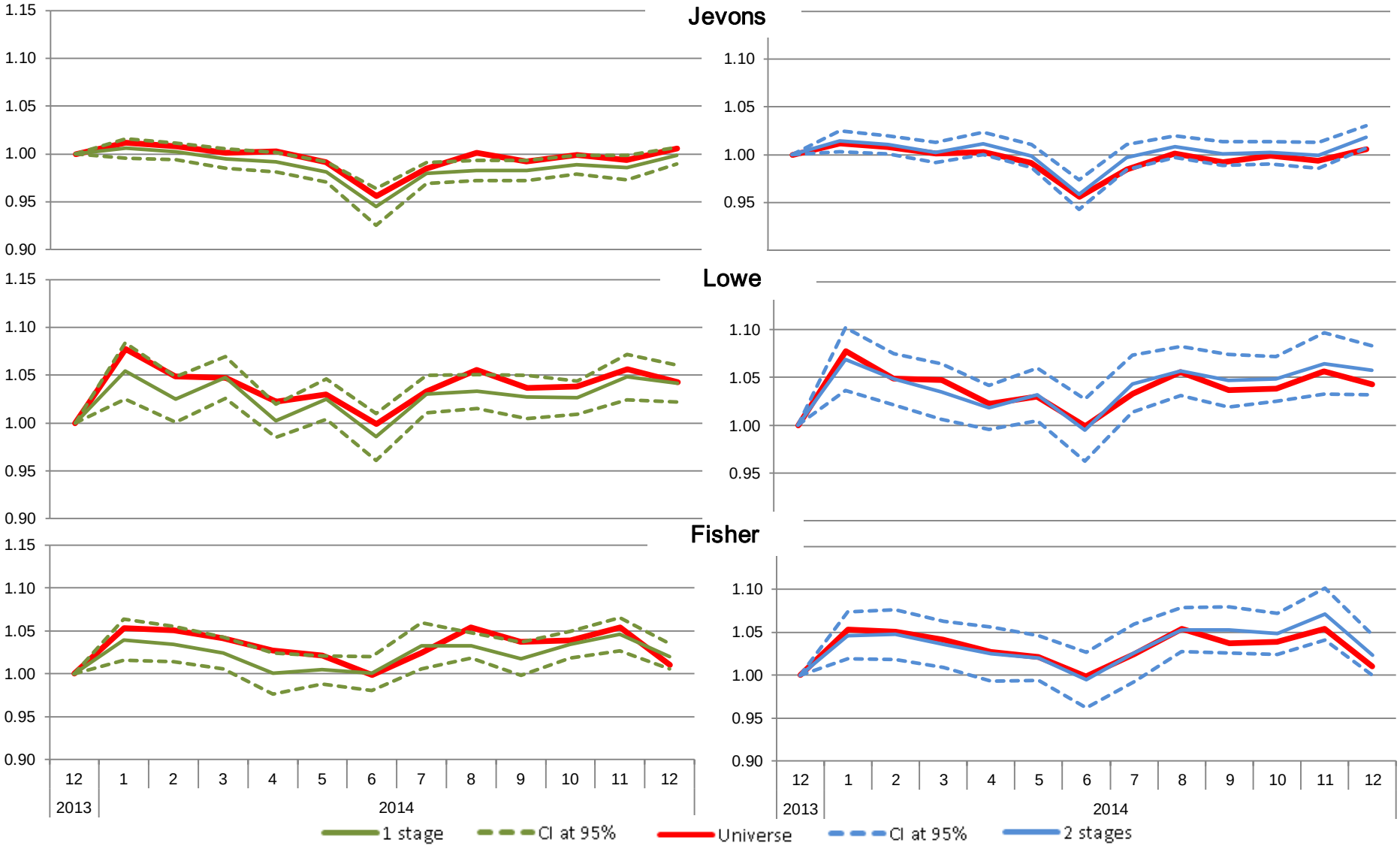
Stratified 2 stages sampling

Comparison of RE% when Second Stage Units (EAN) are selected with SRS or PPS (size=turnover)



Istat

Stratified 1 stage vs. Stratified 2 stages (Turin, Coffee 2014)



Istat

7. Conclusions

Future developments

- Evaluation of the impact of attrition and substitution of elementary items during the yearly data collection
- Implementation of monthly chained matched-item index

Open issues

- Integration between scanner data and traditional data for index compilation
- Different hypothesis are under evaluation:
  - Combining indices obtained with different approaches
  - Gradually stop manual collection, at least for food and grocery, considering the high expenditure coverage of modern distribution
  - Aim at define and realise a probability sampling also for traditional distribution