Multilateral indices and the relaunch problem

Product clustering and alternative solutions

Jacco Daalmans

8 June 2022

Contents

- Relaunch problem
- Product clustering
- Alternative solutions: imputation and matching
- Conclusions

Relaunch problem

Relaunches

- Product replaced by (almost) identical one
- Price change often larger than quality change
- Matched-model methods miss price changes due to relaunches (if nothing is done)
- Here focus on: Geary-Khamis, TPD and GEKS-Törnqvist

Product clustering

Method

- Items combined into product clusters
- Cluster prices computed as unit values
- Unit value bias for heterogeneous items
- Crucial: Suitable cluster definition

Example: Unit value bias

	Price (0)	Price (1)		Quantity (0)	Quantity (1)
Product 1	10	12		100	300
Product 2	18	20		300	100
Unit value price	**16**	**14**			

Clustering: pros and cons

Pros

- Broad applicability
- Well established
- Easy

Cons

- Unit value bias (heterogeneous strata)
- Analysis below cluster level impossible (loss of details)

Imputation

Method

- Price estimated for non-sold items

- Aim: solve relaunch problem rather than to complete data -> might give interpretation problems

  e.g. imputed price for a nonseasonal item

- Non-trivial choice between imputation methods

- Methods available for some indices (e.g. GEKS-Törnqvist) but less well known for others (e.g. TPD)

Relation imputation and clustering

Imputation methods exist that give the same results as product clustering

- GEKS- Törnqvist:

  clustering same as imputing each price with unit value

  (unobserved **and** observed prices)

- TPD and Geary-Khamis:

  Same as for GEKS- Törnqvist, but also quantities need to be imputed

Relation imputation and clustering

- Unit value bias in clustering corresponds to a rigorous imputation approach

- In paper: new imputation methods that:
  - mimic product clustering
  - less rigorous replacement of prices

If all prices are observed -> No adjustment
(contrary to clustering)

If many prices are missing -> Similar results as clustering

Product matching

Method

- Each new product matched with a disappeared product
- Replacement and replaced products should be similar
- Semi-automatic procedures proposed: text mining and manual analysis
- Automatable methods needed for transaction data
- Solution needed if number of new and disappeared products are unequal
- In paper: a simple procedure with arbitrary choices

Matching

Pros

No adjustment of prices and quantities

(contrary to clustering and imputation)

Cons

Arbitrary choices (matching procedure)

Interpretation of the matched products

Simulation

Setup

- 12 month data TV's, chocolate and potato products
- Fixed population: products that have been sold each month
- Simulated relaunches (change of product ID)
- Monte carlo simulation (100 replicates)
- Comparison:
  no correction, clustering, imputation, matching versus 'true' index without simulated relaunches
- Criterion: median abs difference of index values

Scenario 1: random relaunches

TV's : Median distance from true index

	GEKS-Törnqvist	Geary Khamis	TPD
No correction	1.18	1.32	0.74
Imputation	0.65	4.74	4.83
Matching	1.38	1.48	2.33
Clustering	4.56	5.52	5.63

Chocolates: Median distance from true index

	GEKS-Törnqvist	Geary Khamis	TPD
No correction	0.37	0.16	0.20
Imputation	0.47	0.94	1.47
Matching	0.46	0.58	0.98
Clustering	0.78	1.07	1.03

Scenario 1: random relaunches

Potatoes: Median distance from true index

	GEKS-Törnqvist	Geary Khamis	TPD
No correction	1.96	1.40	1.22
Imputation	1.78	1.31	1.15
Matching	0.48	0.53	0.51
Clustering	1.27	1.37	1.37

- Best correction method:
  - For TV's and Chocolates: 'No correction'
  - For Potatoes: 'Matching'

- Matching always better than clustering

Scenario 2: Non-random relaunches

- Relaunches occur at one time period for randomly selected 75% of all items.

- Each relaunch goes along with a simulated, permanent price increase by 20%.

Scenario 2: Non-random relaunches

TV's : Median distance from true index

	GEKS-Törnqvist	Geary Khamis	TPD
No correction	11.14	7.39	10.54
Imputation	1.24	3.81	5.11
Matching	2.10	1.44	3.62
Clustering	5.63	6.66	6.70

Chocolates: Median distance from true index

	GEKS-Törnqvist	Geary Khamis	TPD
No correction	14.77	11.90	14.61
Imputation	1.83	0.34	3.04
Matching	0.20	4.09	0.54
Clustering	0.74	0.85	0.81

Scenario 2: Non-random relaunches

Potatoes: Median distance from true index

	GEKS-Törnqvist	Geary Khamis	TPD
No correction	13.33	10.42	13.40
Imputation	0.83	2.06	1.28
Matching	0.95	1.58	1.05
Clustering	1.10	1.31	1.20

- Correction method necessary ("no correction" gives large errors)

- Best correction method depends on data set and index method:
  - Matching (5 cases)
  - Imputation (3 cases)
  - Clustering (1 case)

- Matching better than clustering for 7 out of 9 cases

Conclusion

- Correction for relaunches needed (given a price increase)

- Clustering easy, well-understood, broad applicable, but unit value bias for heterogeneous strata

- Imputation and matching mostly give better results in a simulation study (especially matching)

- Drawbacks: arbitrary choices, unnatural to ‘add’ values to transaction data, interpretation difficulties.

Thank you!