



european conference
on quality in
official statistics

Alternative panel data approaches: longitudinal representativity

Silvia Biffignandi – University of Bergamo

Alessandro Zeli - Istat

Outline

- Objective
- Panel construction approaches
- Methodology
- Comparison and evaluation criteria
- Results
- Further research

Objective

Context

- users need to widen statistical information
- NSIs needs
- new generation of databases

Research objective

- Longitudinal databases and M&A

Objective: relevance

Changes may affect:

- the dimensional structure (number of persons employed)
- the economic activity
- the legal status (firm name, fiscal code).

Moreover many events can imply the death of an enterprise.

Panel construction: three approaches

- firms transformation events. Example Nardecchia et al.(2010)
- simple code panel
- cross section approach, i.e. a series of cross-sections bound together, for instance Grazzi et al.(2009,2014)

Proposed methodology

Approach: firms transformation events.
Complex database.

Advantages :

- continuity of a single firm
- maximisation of the representativity of panel.

Proposed methodology

Databases involved

- ASIA (Italian Business Register)
- Events Register
- Financial reports database (Chamber of Commerce)

Proposed methodology

Events Register

Code1: origin firm code (for instance the seller),

Code2: destination firm code (for instance the buyer),

Event Code

Event Date: month and year

Quality Code

Proposed methodology

The firms involved in events are grouped together by using the Event Register.

Cluster procedure: combination of the origin and destination code and by means of the event code.

All M&A events are used to cluster the involved firms by means of a special code (cid);

Proposed methodology

Super-entity scope

built a super-entity (i.e. a cluster of firms involved in events during the panel period).

This super-entity may be traced along all panel period unlike the single firms.

Proposed methodology

Steps:

- 1) set of clusters of grouped firms by events; all firms having some events in the period are clustered together even for the years before or after the event;
- 2) for a single year the clusters of firms involved in M&A is obtained by linking the set of clusters with the corresponding year of the Business register ASIA.;

Proposed methodology

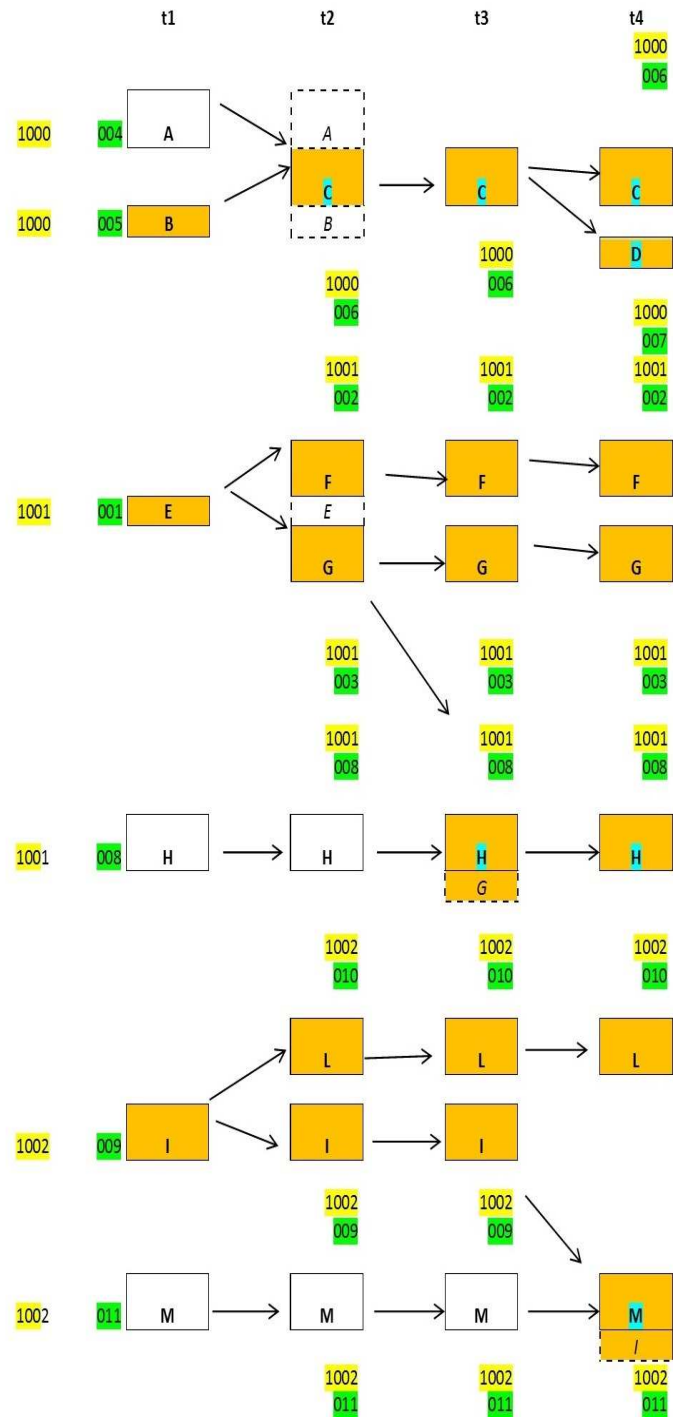
Steps:

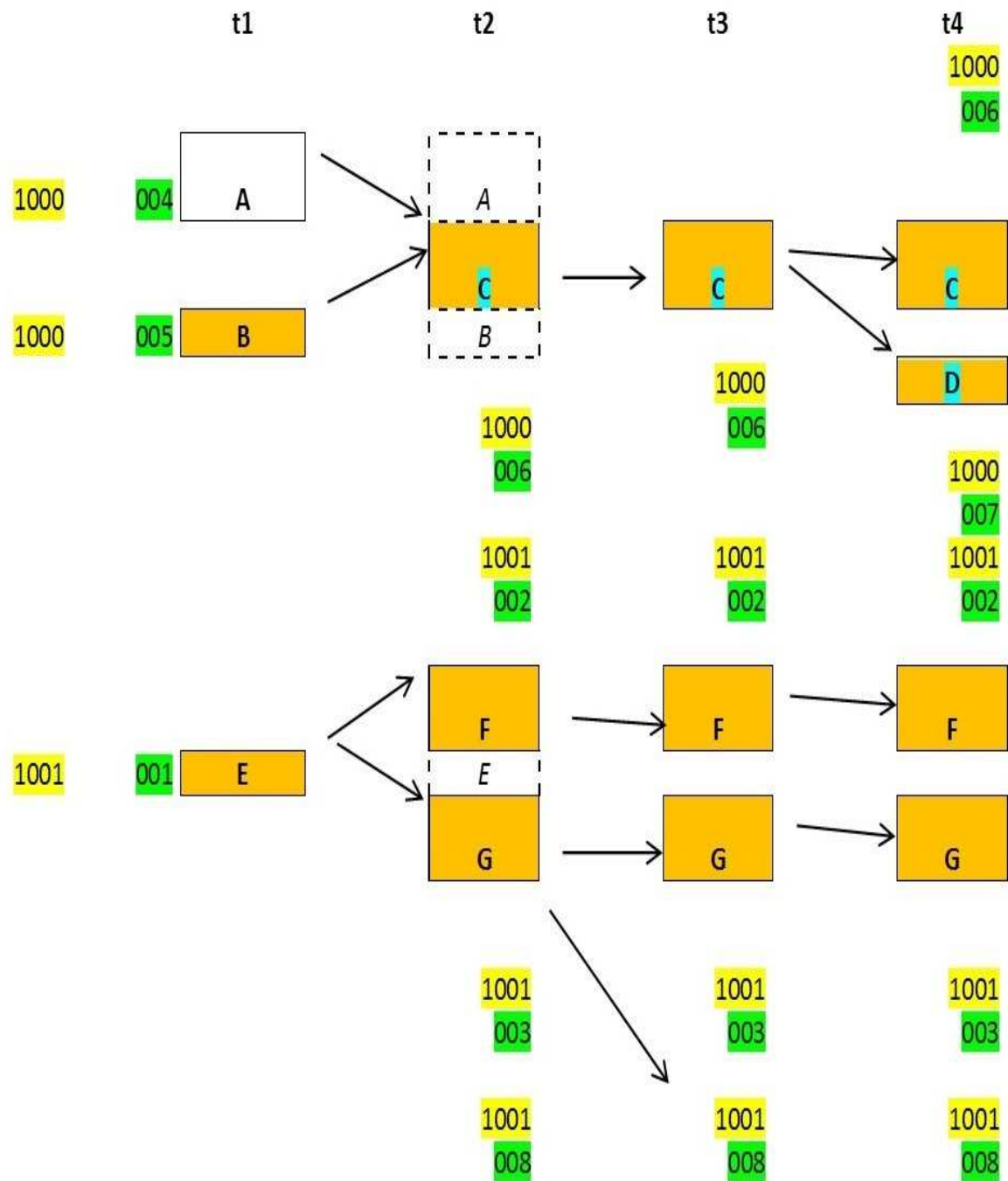
- 3) having linked the set of clusters with the Business Register at year t , all firms (codes) existing in the year t that have, had or will have events in the panel period is obtained;

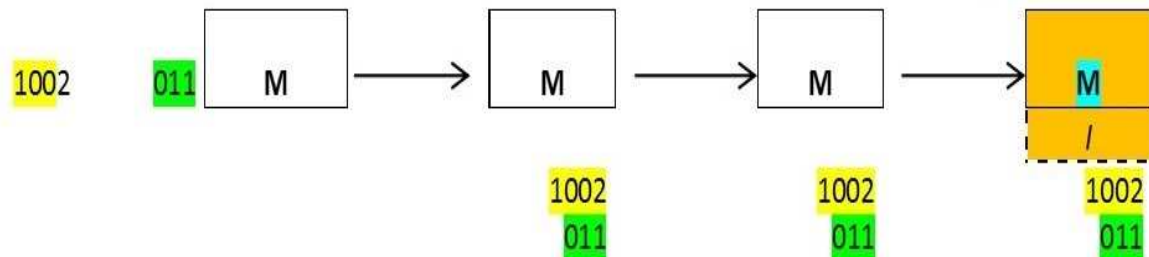
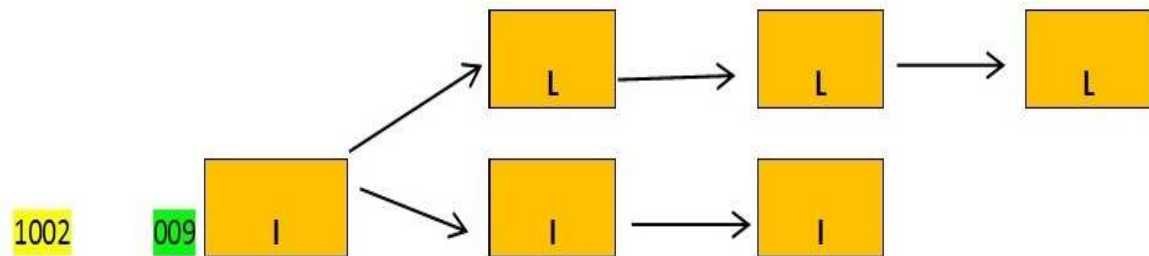
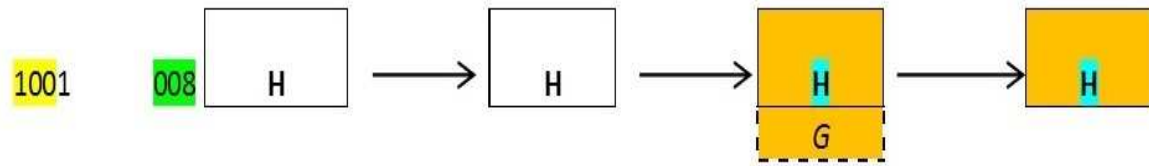
Proposed methodology

Steps:

- 4) codes of point 3 are linked with the file of the financial reports of limited enterprises coming from Chamber of Commerce register
- 5) relevant economic variables value are computed.







Proposed methodology

At the end of the events grouping is available:

- the whole ASIA register classified according also to the events classifications
- list of enterprises linked to the economic data from financial reports of limited enterprises laid down to Chamber of Commerce
- values of main economic variable

Comparison and evaluation criteria

Benchmark population: balance sheet database.

Comparison between M&A db (cid) and simple code db (cod)

Analyses:

- Preliminary information: coverage
- Statistical tests

Comparison and evaluation criteria

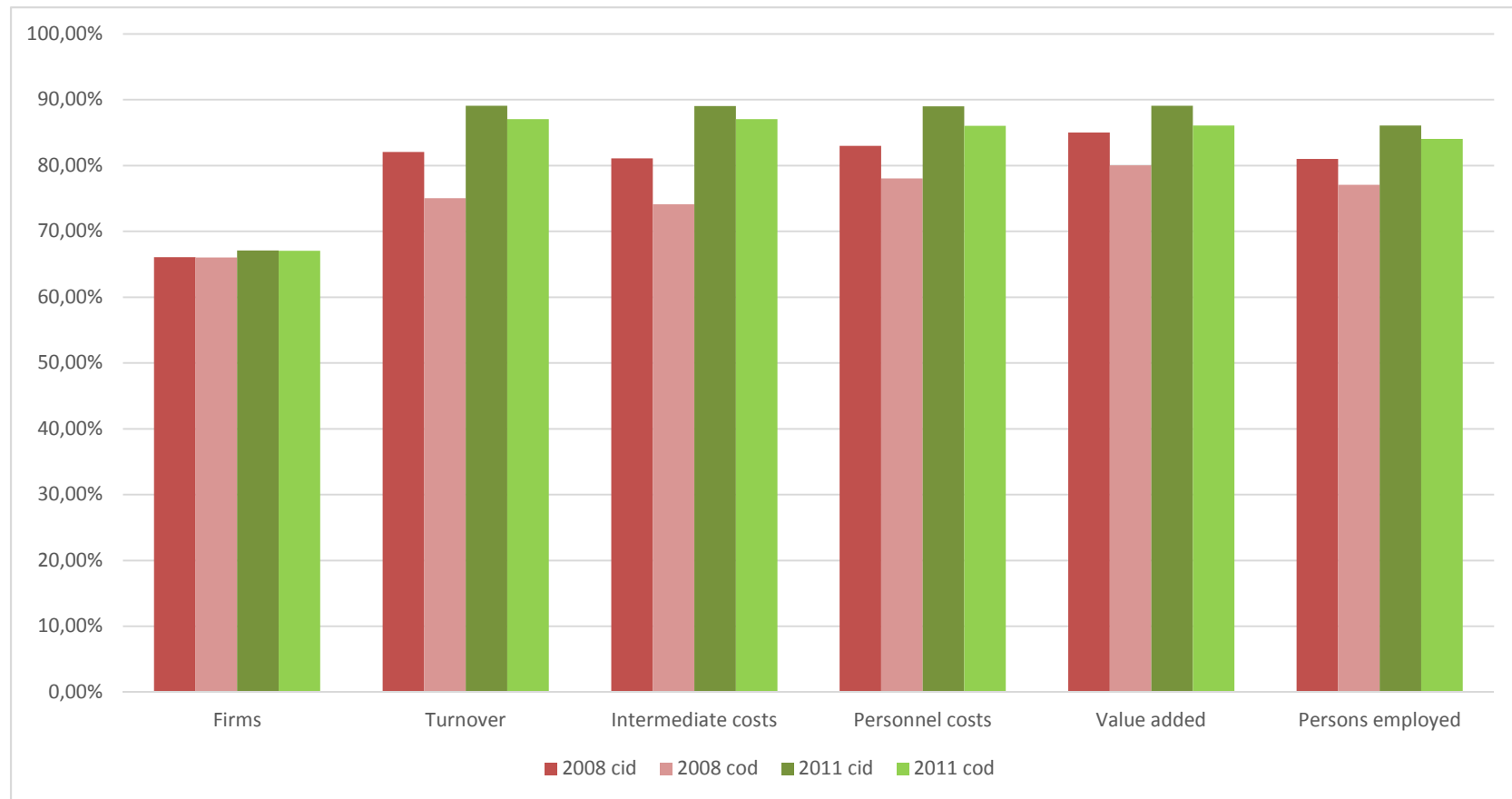
Method:

- Spearman correlation
- Fligner Policello test
- R-indicator

Coverage rate of the two panel with respect to target population – Years 2008-2011.

		Firms	Turnover	Intermediate costs	Personnel costs	Value added	Persons employed
2011cid		67.8	89.7	89.5	89.1	89.6	86.7
2011cod		67.3	87.6	87.3	86.5	86.9	84.4

Coverage rate of *cid* and *cod* with respect to target population — Year 2008 and 2011



Coverage rates

		Firms	Turnover	Intermedia te costs	Personnel costs	Value added	Persons employed
2008	cid	66.8	82.4	81.9	83.0	85.0	81.0
2008	cod	66.5	75.6	74.9	78.4	80.1	77.5
2009	cid	70.6	84.4	83.6	84.6	85.9	83.0
2009	cod	70.2	77.6	76.7	79.6	80.5	79.2
2010	cid	62.3	88.4	87.8	88.2	89.2	84.0
2010	cod	61.8	83.6	83.0	84.6	85.4	81.5
2011	cid	67.8	89.7	89.5	89.1	89.6	86.7
2011	cod	67.3	87.6	87.3	86.5	86.9	84.4

Results

Spearman correlation

(two-digit NACE (economic divisions))

- close values for *cid* and *cod* panel
- *cid* Spearman coefficients are higher than *cod* ones for any variables in both years

The ranking of economic sectors does not vary very much from one approach to the other.

Cid improves representativity at least in terms of ranking of economic activity.

Spearman correlation coefficients between total population and *cid* - *cod* panel for Nace 2-digit level of some relevant economic variables – Years 2008 and 2011

	2008		2011	
	cod	cid	cod	cid
persons employed	99.6	99.6	99.6	99.6
	<.0001	<.0001	<.0001	<.0001
intermediate costs	95.6	96.2	99.5	99.6
	<.0001	<.0001	<.0001	<.0001
profits	86.9	88.3	97.8	98.6
	<.0001	<.0001	<.0001	<.0001
personnel costs	99.1	99.2	99.5	99.6
	<.0001	<.0001	<.0001	<.0001
value added	99.2	99.3	99.7	99.8
	<.0001	<.0001	<.0001	<.0001
turnover	95.5	96.3	99.5	99.6
	<.0001	<.0001	<.0001	<.0001

Results: FP test

- no significant difference between the two panels and the original population of firms
- the cid value are in every case (but the person employed for the year 2008) smaller than cod ones. Considering FP values as a "distance" between distributions cid db improves representativity with respect to the cod one

FP-test on distributions of selected relevant economic variables

		Turnover	Personnel costs	Intermediate costs	Value added	Persons employed
2008	Cid	0.95	0.87	0.99	0.78	1.55
		0.17	0.19	0.16	0.21	0.05
	Cod	1.08	0.88	1.09	0.8	1.46
		0.13	0.19	0.13	0.21	0.07
2011	Cid	0.67	0.74	0.71	0.76	0.77
		0.25	0.22	0.23	0.22	0.22
	Cod	0.82	0.9	0.88	0.94	0.93
		0.2	0.18	0.19	0.17	0.27

Results: R-indicators

economic sector indicators:

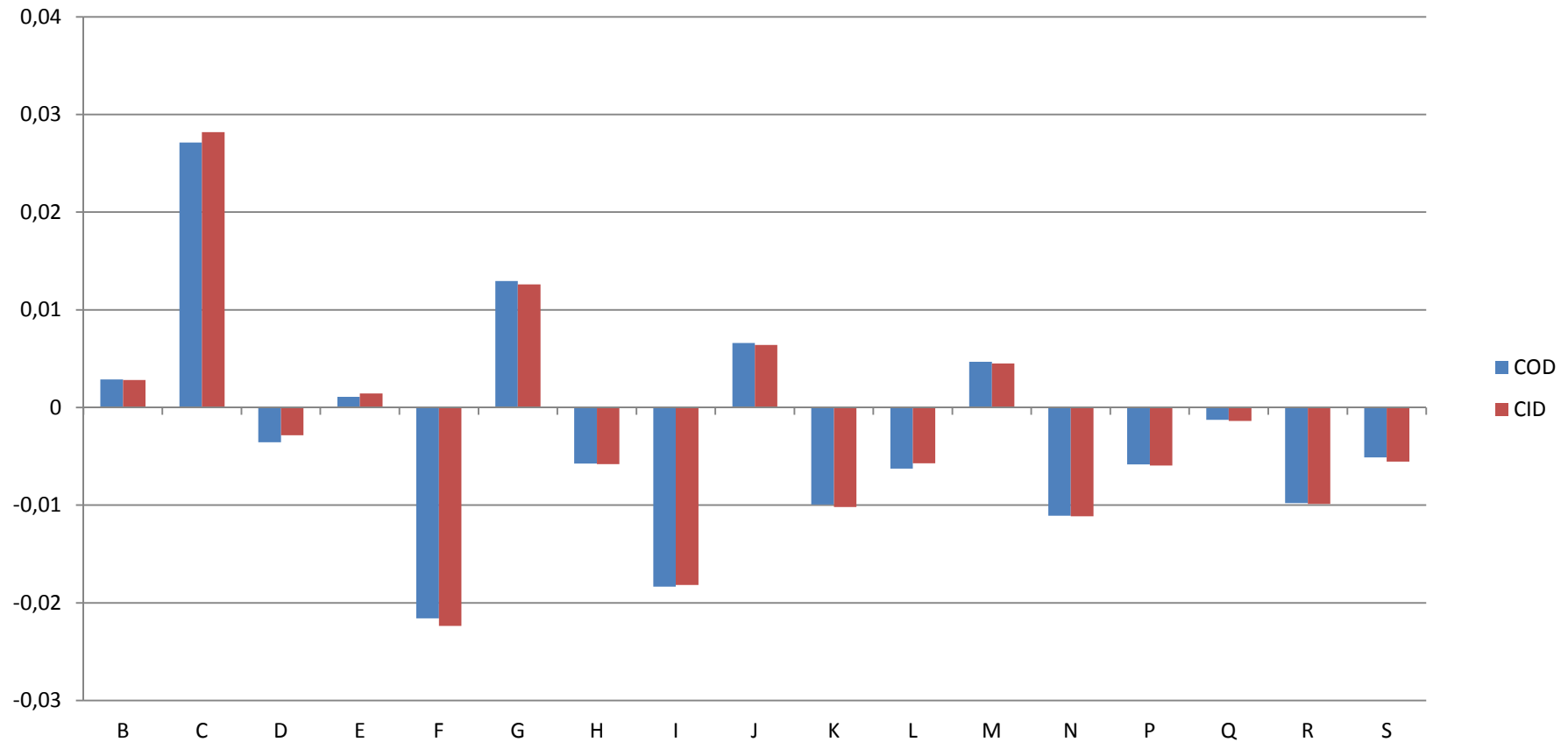
- mining, manufacturing, trade in both type of panel: over-represented
- the other sectors: under-represented

Results: R-indicators

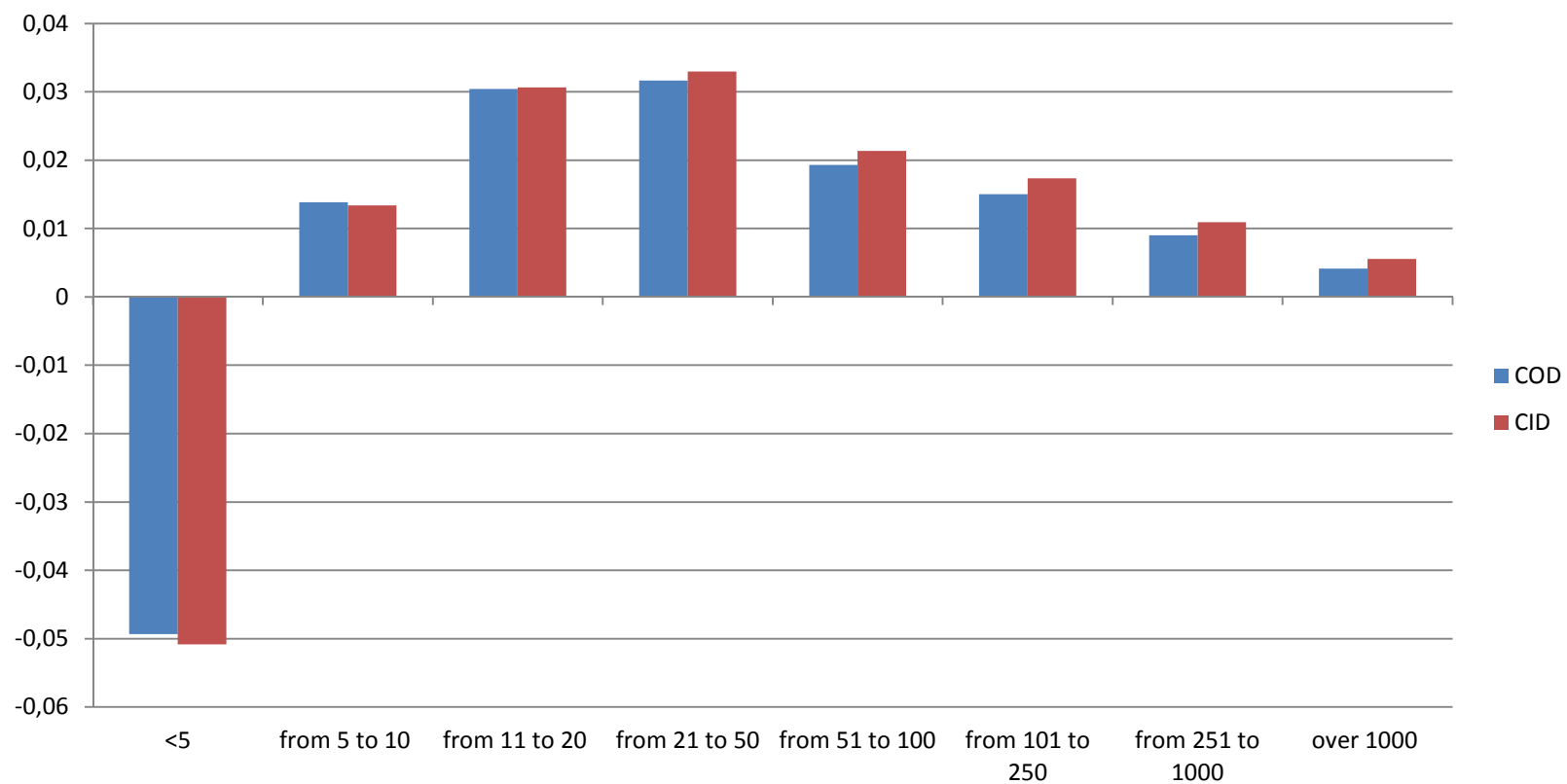
size class partial indicators:

- in general, small firms under-represented (less than 5 employees) and larger enterprises over-represented
- for *cid* panel, greater over-representativity for larger enterprises with more than 10 persons employed, whereas *cod* panel less under-representativity for the smallest firms (with less than 10 persons employed) in 2011.

Category Level Unconditional Partial Indicators (*economic sectors*) – Year 2011



Category Level Unconditional Partial Indicators (size class) – Year 2011



Concluding remarks and further research

- ranking and distribution are well represented by each of the two approaches
- **cid superior in many ways**
- cid better for medium-large enterprises

Cid: good performance, but further investigation on the impact of the economic values