národní
úložiště
šedé
literatury

Common mistakes in the application of bibliometric information in the Czech Republic

Jurajda, Štěpán
2011

Dostupný z http://www.nusl.cz/ntk/nusl-112980

Dílo je chráněno podle autorského zákona č. 121/2000 Sb.

Licence Creative Commons Uveďte autora-Neužívejte dílo komerčně-Nezasahujte do díla 3.0 Česko

Tento dokument byl stažen z Národního úložiště šedé literatury (NUŠL).

Datum stažení: 03.10.2024

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní nusl.cz .

Common mistakes in the application of bibliometric information in the Czech Republic

Štěpán Jurajda

CERGE EI

Knowledge-Research-Education Conference

Prague Sep 2011

How to achieve excellence in R&D?

Allocation mechanisms and built-in incentives of evaluation methodologies (often based on bibliometric data) determine how scientists spend their time and whether they choose to work in a country, given its overall level of R&D funding.

What is the productivity of Czech science?

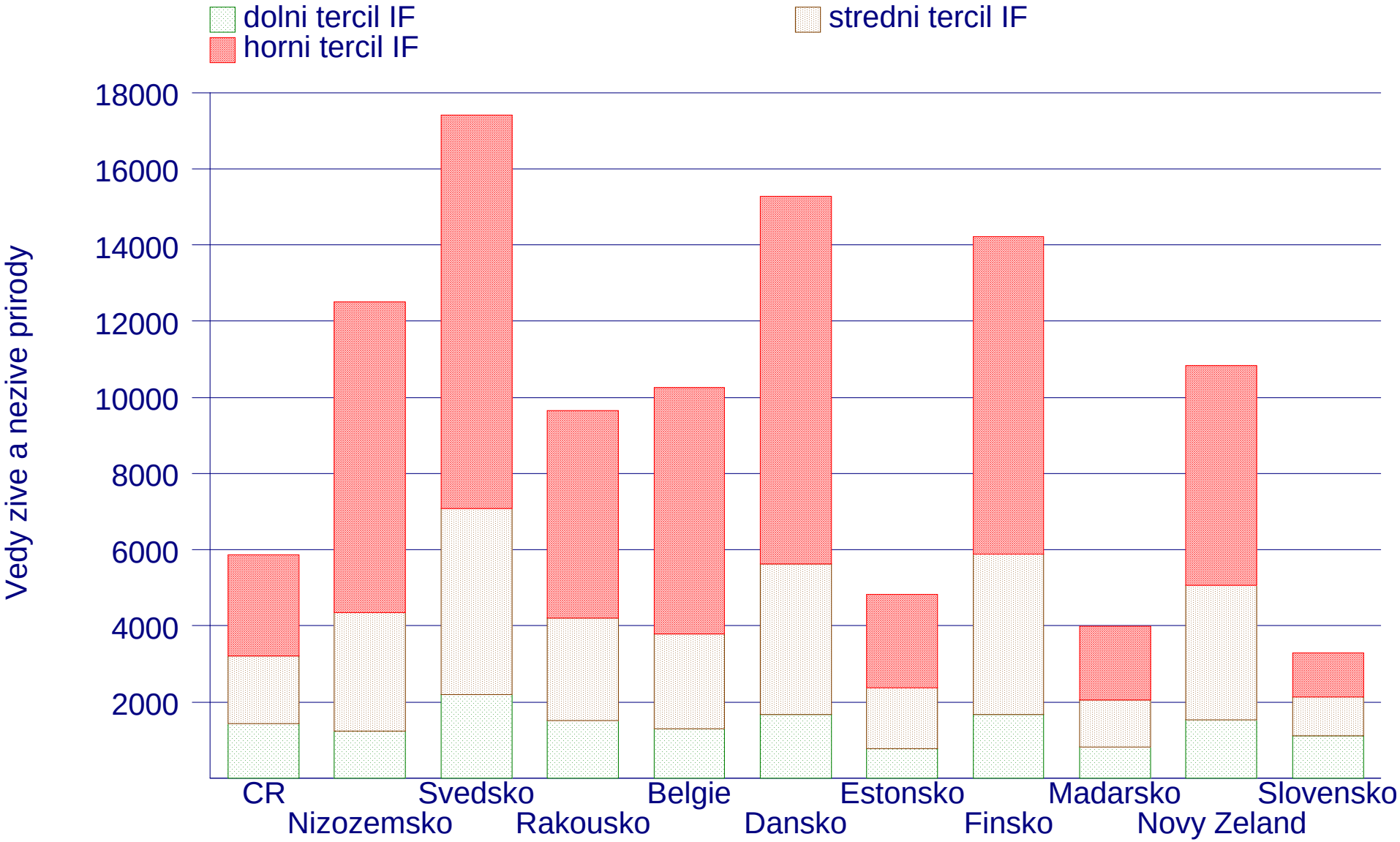
- NOBODY KNOWS (output trend to be divided by inputs)
- We do not even know which fields are more productive
- Wrong bibliometric data is used to finance Czech R&D&I

Data:

- 2011 CWTS measures of outputs, not scaled by inputs
- Other local analyses misleading and/or non-informative:
  - Analysis of R&D&I by the R&D&I Council -2010
  - Field comparison of Thomson Reuters indicators 2010
  - Field priority setting analysis, Technology Center 2011
  - R&D&I Council's Evaluation Methodology (2010-2015?)

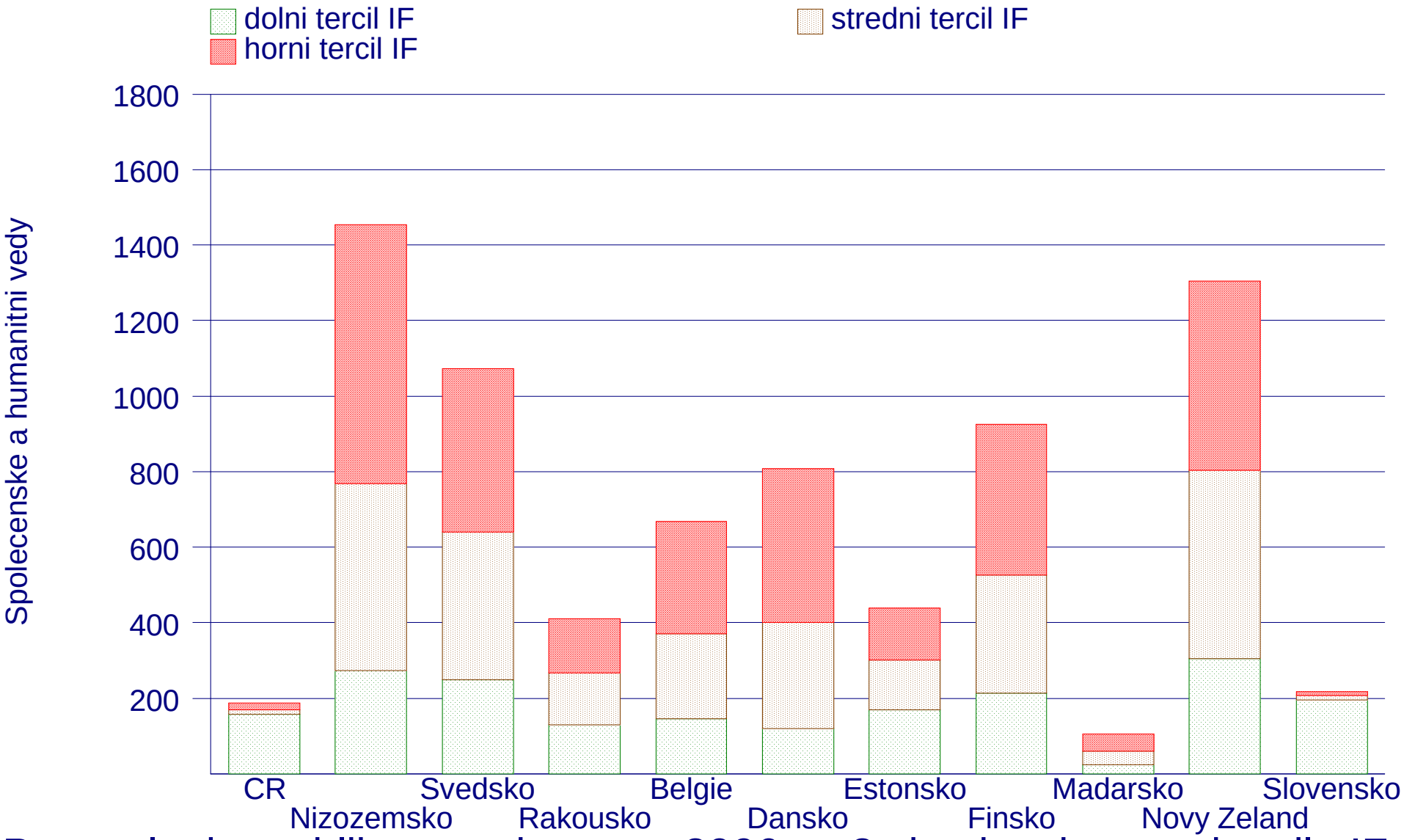
- I will substantiate these claims and offer some examples.
- But first, some own calculations based on WoS IF data (IF not ideal, but available for comparably sized countries)

Natural Sciences, article counts by terciles of IF, scaled to Czech population



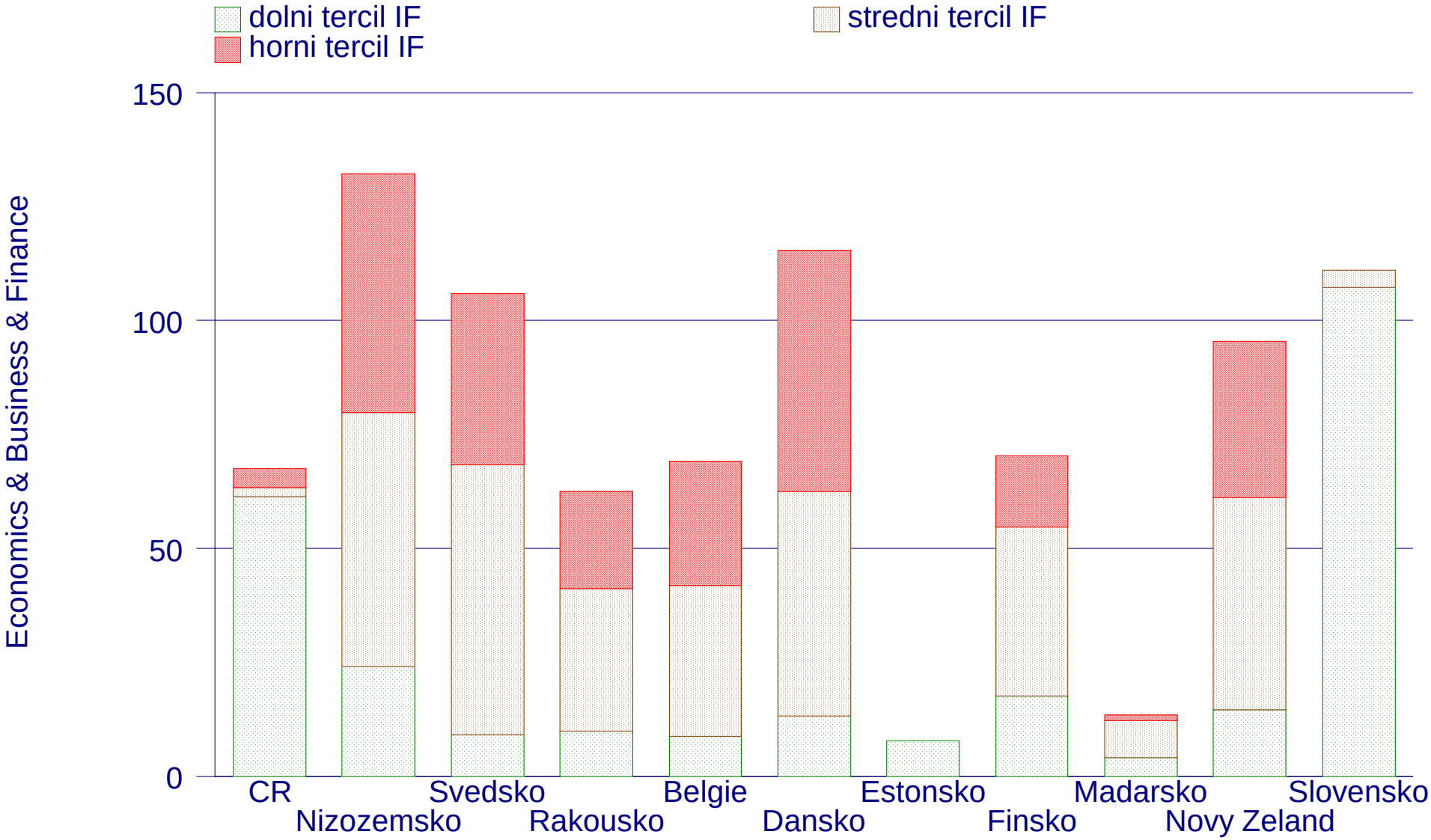
Pocet clanku publikovanych v roce 2006 ve 3 skupinach casopisu dle IF

Social sciences and humanities



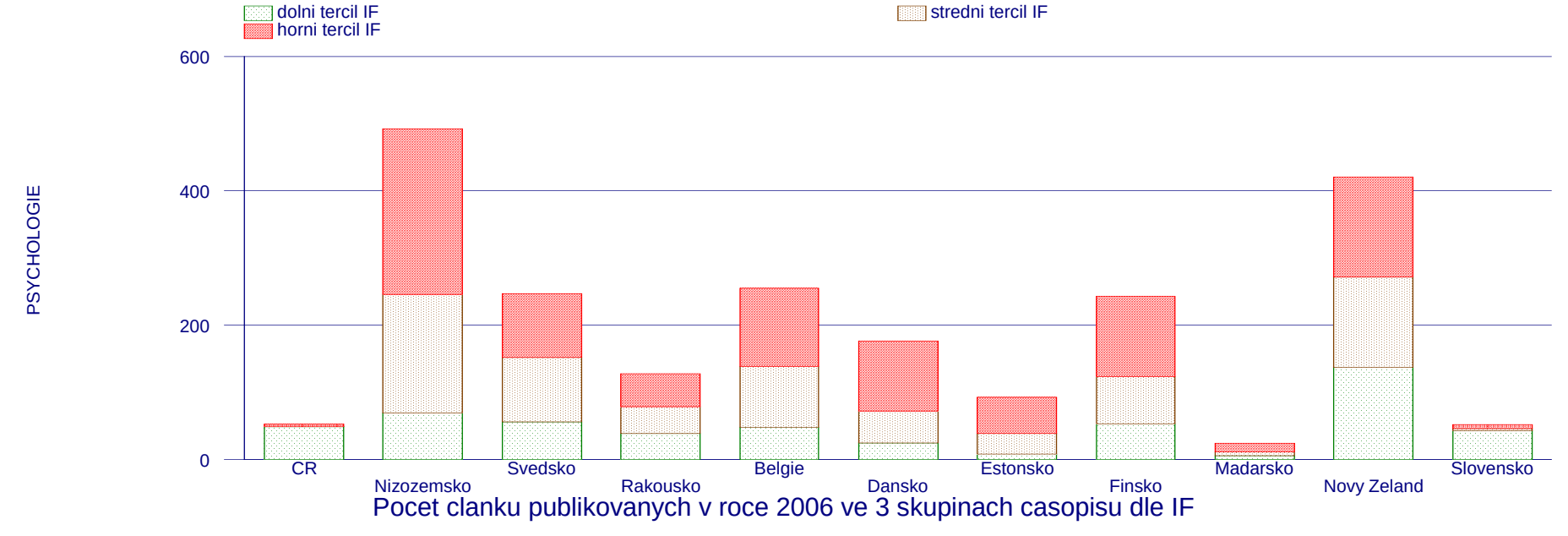
Pocet clanku publikovanych v roce 2006 ve 3 skupinach casopisu dle IF

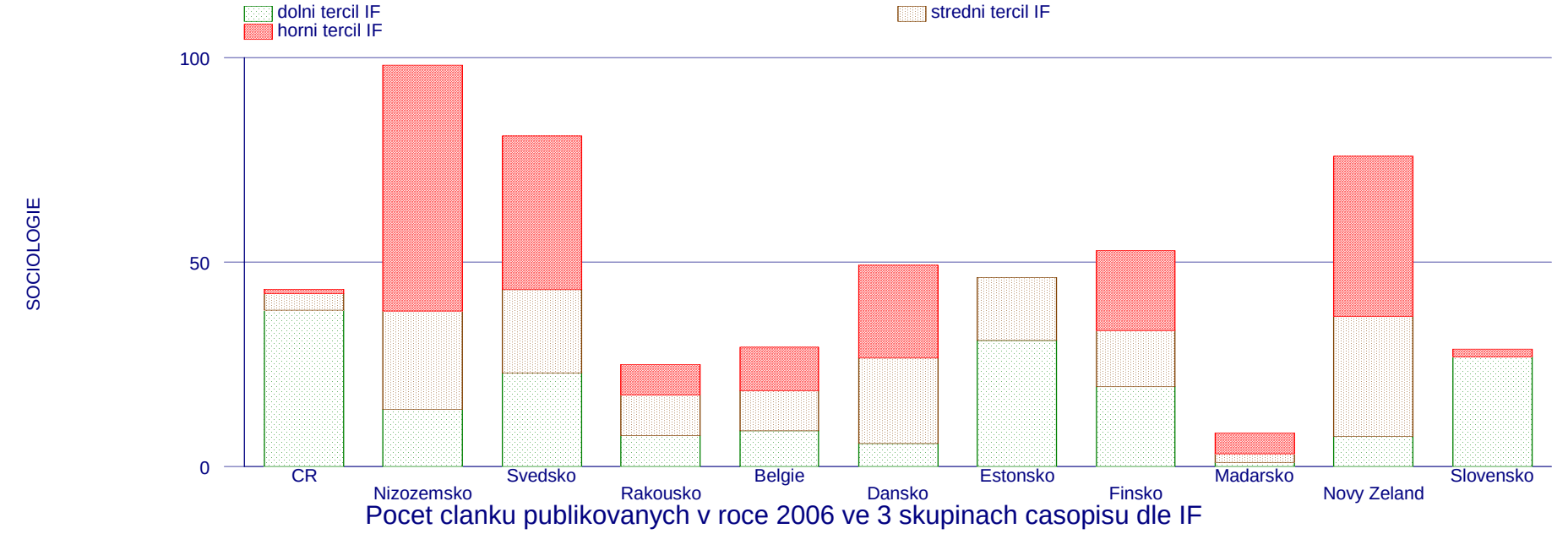
Economics and Business and Finance



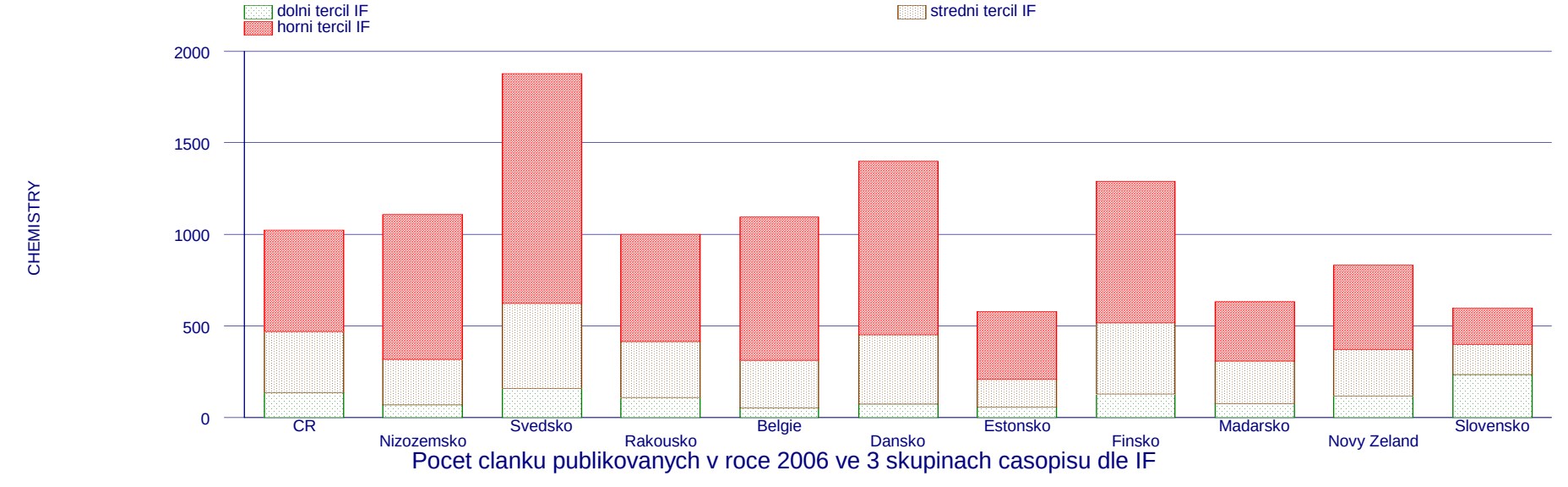
Pocet clanku publikovanych v roce 2006 ve 3 skupinach casopisu dle IF

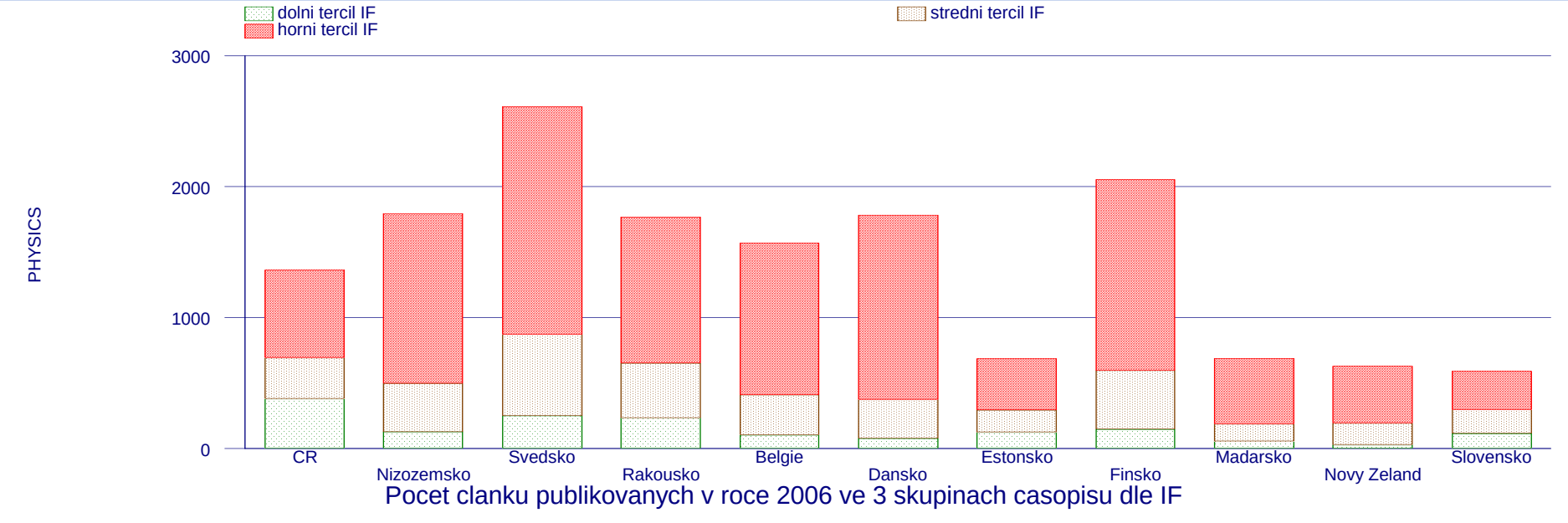
Psychology, Sociology





Chemistry and Physics



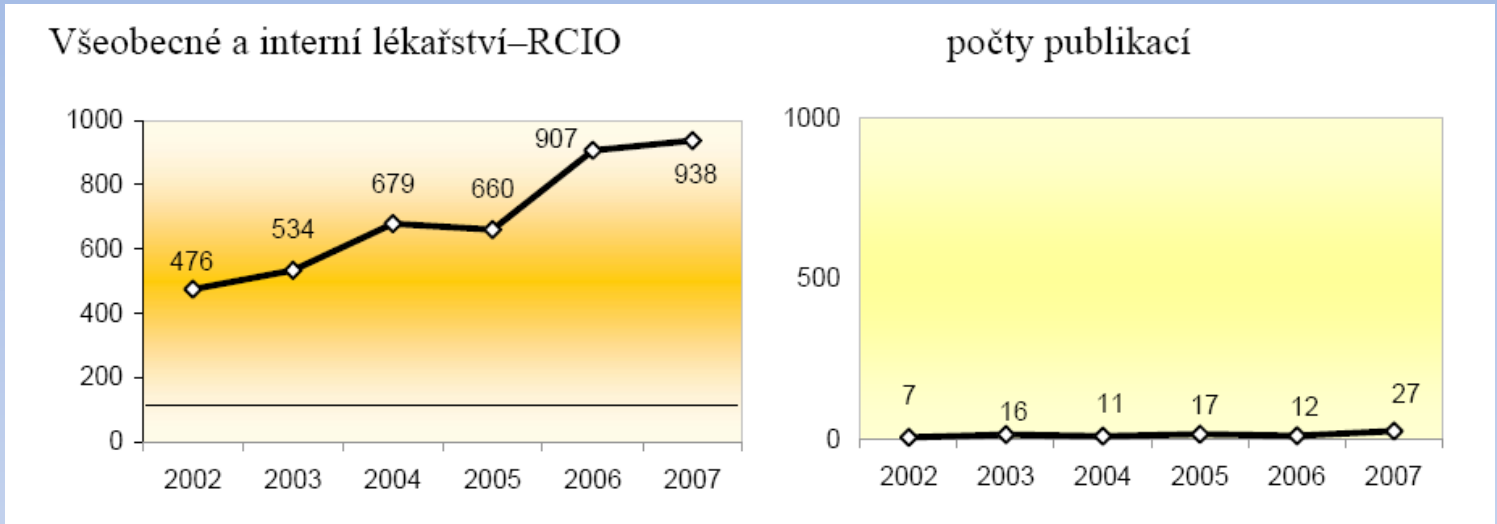


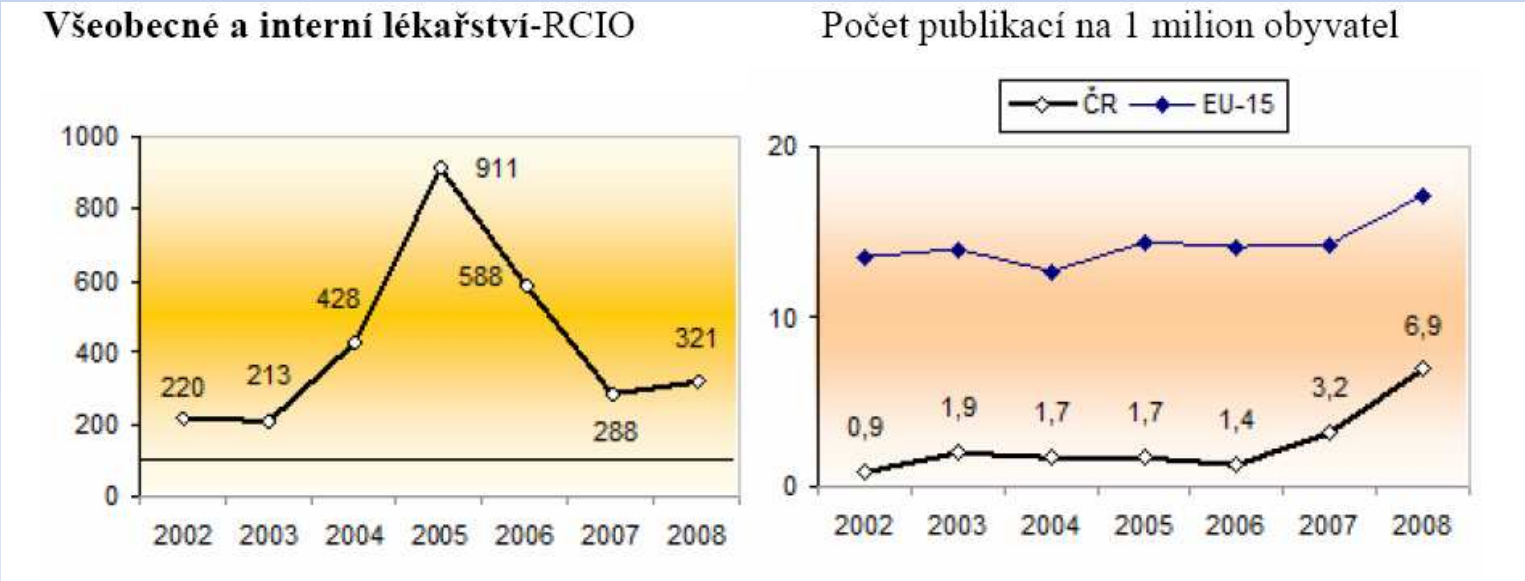
Existing Czech official analyses

1. Annual R&D&I “Analysis”: RCIO indices, no normalization by inputs or output counts
2. R&D&I ‘points’ to distribute ~10bln CZK p.a.
3. Thomson Reuters bibliometric analysis: no address harmonization, no field normalization
4. Technology Center field priority analysis
5. CWTS: WoS field-normalized citation impacts based on address harmonization, OP-funded

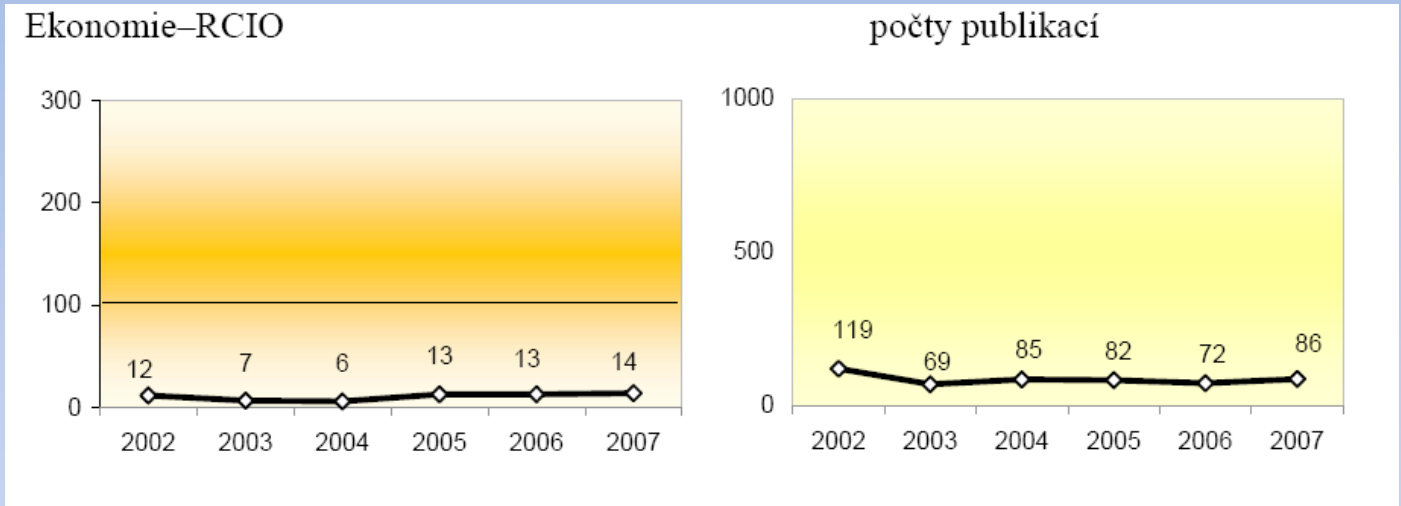
1-4 commissioned by the Czech R&D&I Council

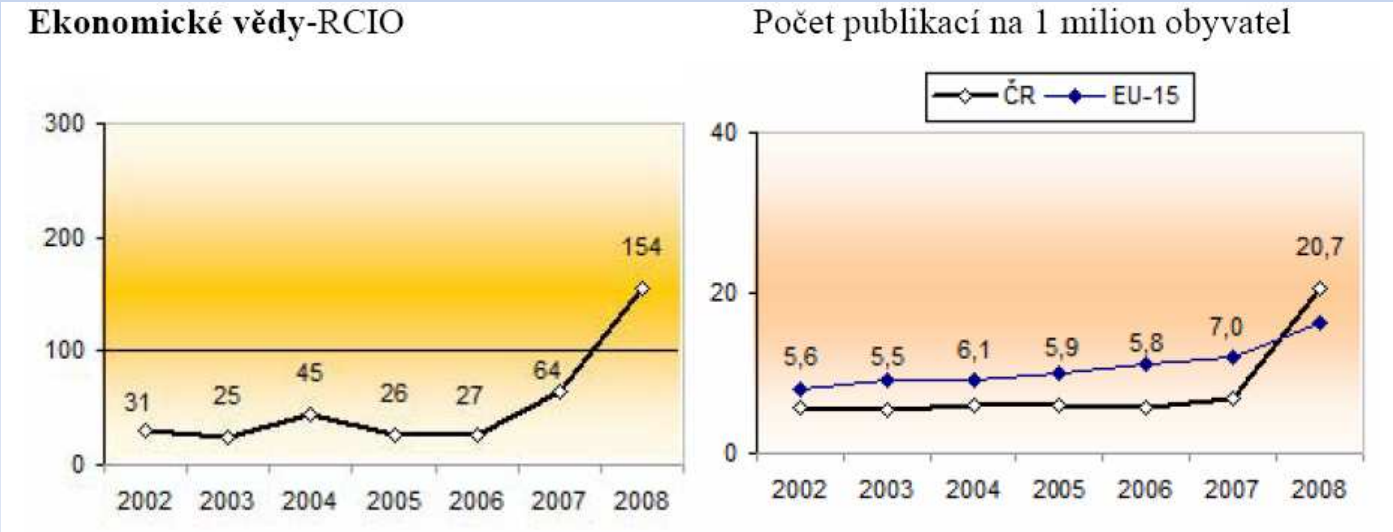
1. R&D&I Council “Analysis”: RCIO





1. R&D&I Council “Analysis”: RCIO





1. RCIO misuse, cont.

- Relative citation indices (field-normalized citations per paper) tend to be high where only few researchers from a large field publish in IF journals even though almost all should.
- In Czech lands, normalizing by number of researchers problematic since we do not know the physical number of researchers (only FTE). Also ‘national’ journals distort citations (more on this issue below).

1. RCIO misuse, cont.

- Lesson: present relative citation measures only jointly with measures of total output per unit of input (at least per researcher).
- This will stop the confusing comparison between productive fields with large output at average RCIO from unproductive fields with tiny highly selected output with high RCIO.

2. The 2009 ‘point’ table

Result types			I- NERR specializations	II – other specializations
J_{imp}	article in impacted magazine [1]		10 to 305 [2]	
	article in prestigious impacted magazine (Nature, Science, Proc.Natl. Acad. Sci.) [3]		500	
J_{neimp}	article in non-impacted magazine	world-renowned database [4]	12	8
		list of critiqued periodicals [4]	10	4
B	scholarly book	world language [5]	40	40
		other languages		20
D	article in proceedings [6]		8	
P	patent	European or international patent (EPO, WIPO), patent of USA and Japan	500	
		Czech or national patent used on the basis of a valid license contract	200	
		other patents [7]	40	
Z	pilot plant, confirmed technology, species, breed		100	
F	usable sample		40	
	industrial sample		40	
G	prototype, functional sample		40	
H	Applied results		40	
N, L	certified methodologies and procedures, specialized maps with scholarly content		40	
R	software		40	
V	research report containing classified information		50	

1) NERR includes specializations (according to R&D IS codebook: AA – Philosophy and Religion, AB – History, AC – Archaeology, Anthropology, and Ethnology, AD – Politology and Political Science, AE – Administration – AG, Legal Science, AI – Linguistic Science, AJ – Literature, Mass media, and Audiovisuals, AL – Art, Architecture, and Cultural Heritage, AM – Education and Schools.)

2) publications indicated in the following database Web of Science of the company Thomson Reuters: Science Citation Index Expanded (SCI-EXPANDED) – 1945 – present; Social Science Citation Index (SSCI) – 1980 – present; Arts & Humanities Citation Index (A&HCI) – 1980 – present; Index Chemicus (IC) – 1993 – present; Current Chemical Reactions (CCR-EXPANDED) – 1986 – present

3) evaluation Impact 10 x 305 x Factor, where

2. Fatal flaws of Czech “point” system

A. Arbitrary allocations across fields, between basic science and applied/innovation work
B. Use of IF in all fields, with little normalization (to costs, salaries, frequency of publications)
C. Automatic financing of easy-to-produce output = powerful incentives against quality research, esp. in fields with low frequency of IF or no IF; incentives applied to micro units

OP-funded International Audit of Czech R&D:

“Discontinue the extreme system immediately”

2.A Arbitrary allocations

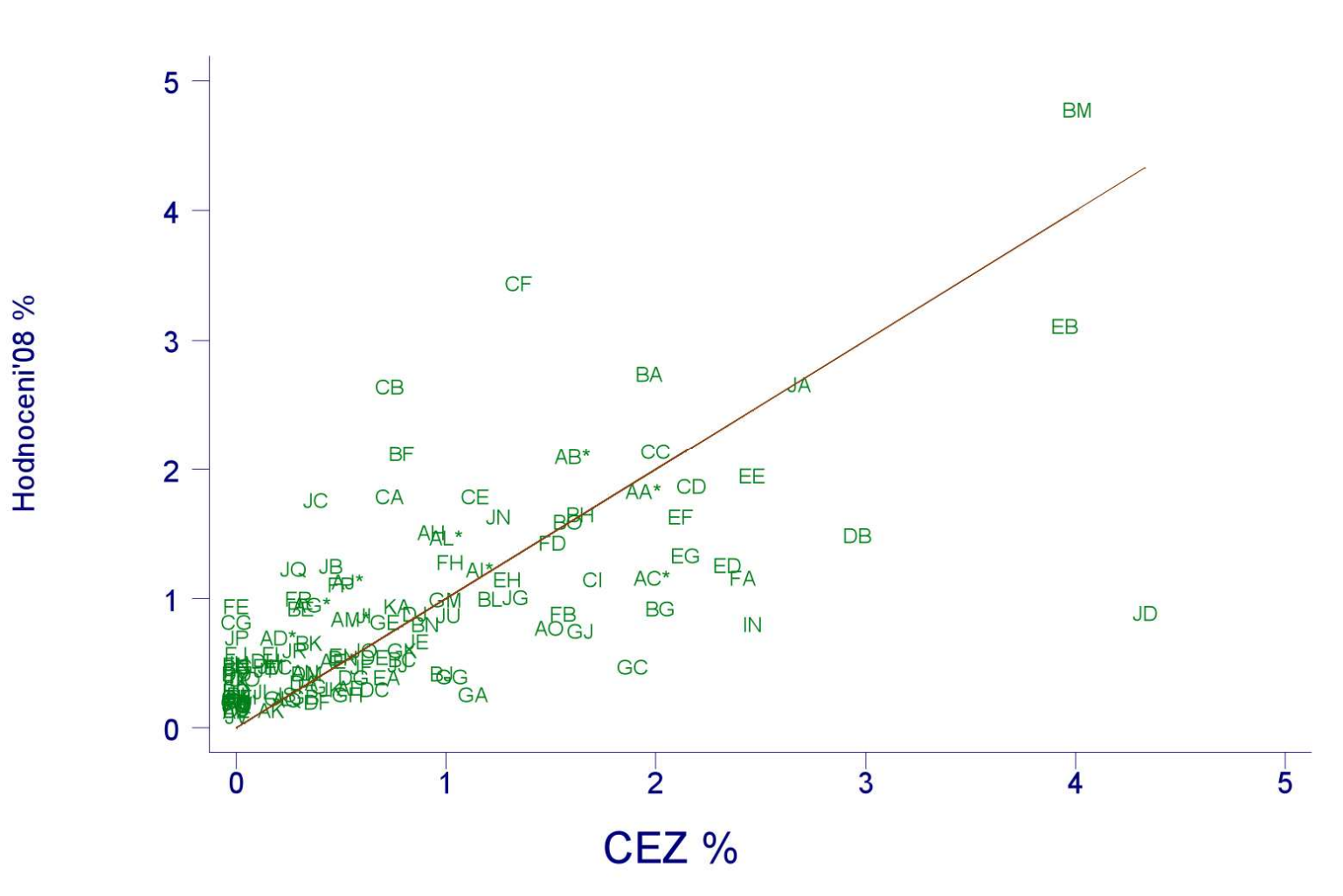
- The only R&D financing system in the world based solely on counting (not reading) outputs
- The point parameters per type of output are set and adjusted arbitrarily, resulting in arbitrary division of funding between

a.basic/applied research,

b.old/new field funding shares,

c.new/even newer field funding shares…

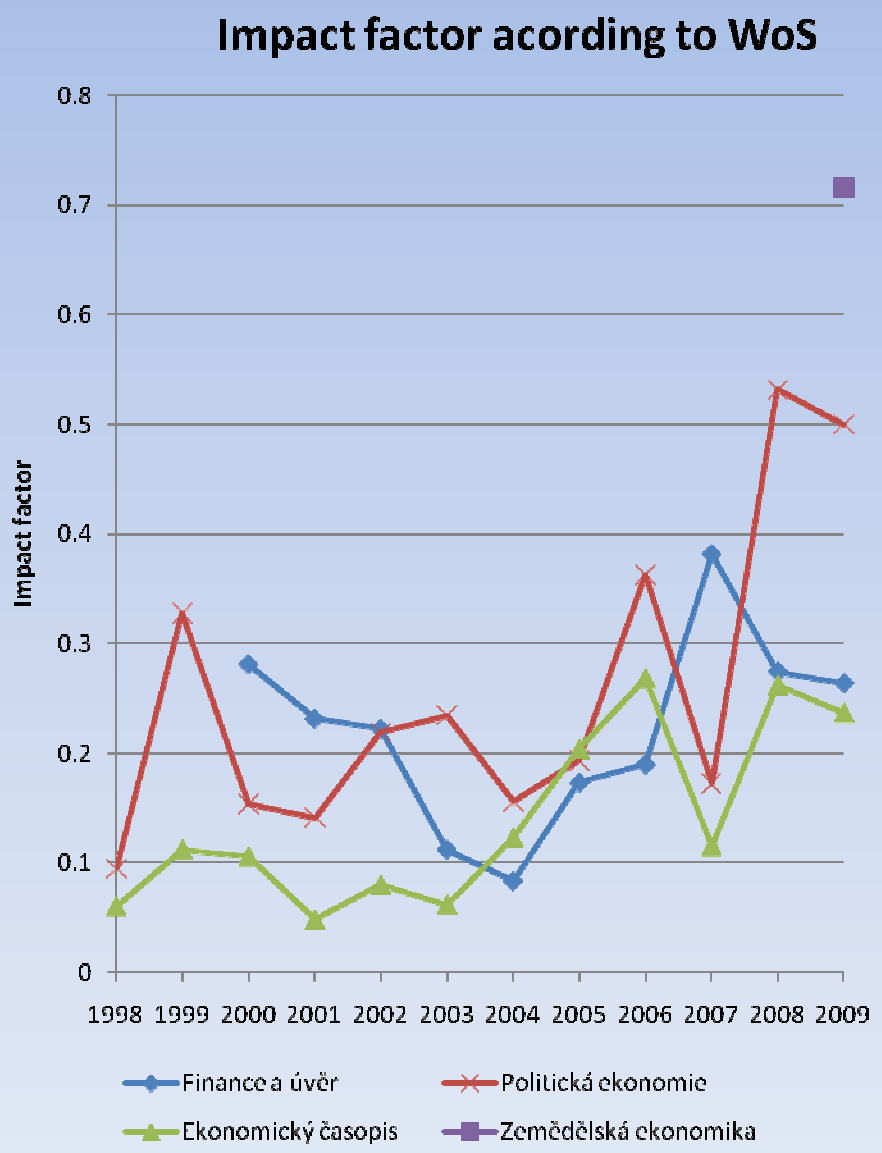
Old/new shares of science fields on total institutional funding budget

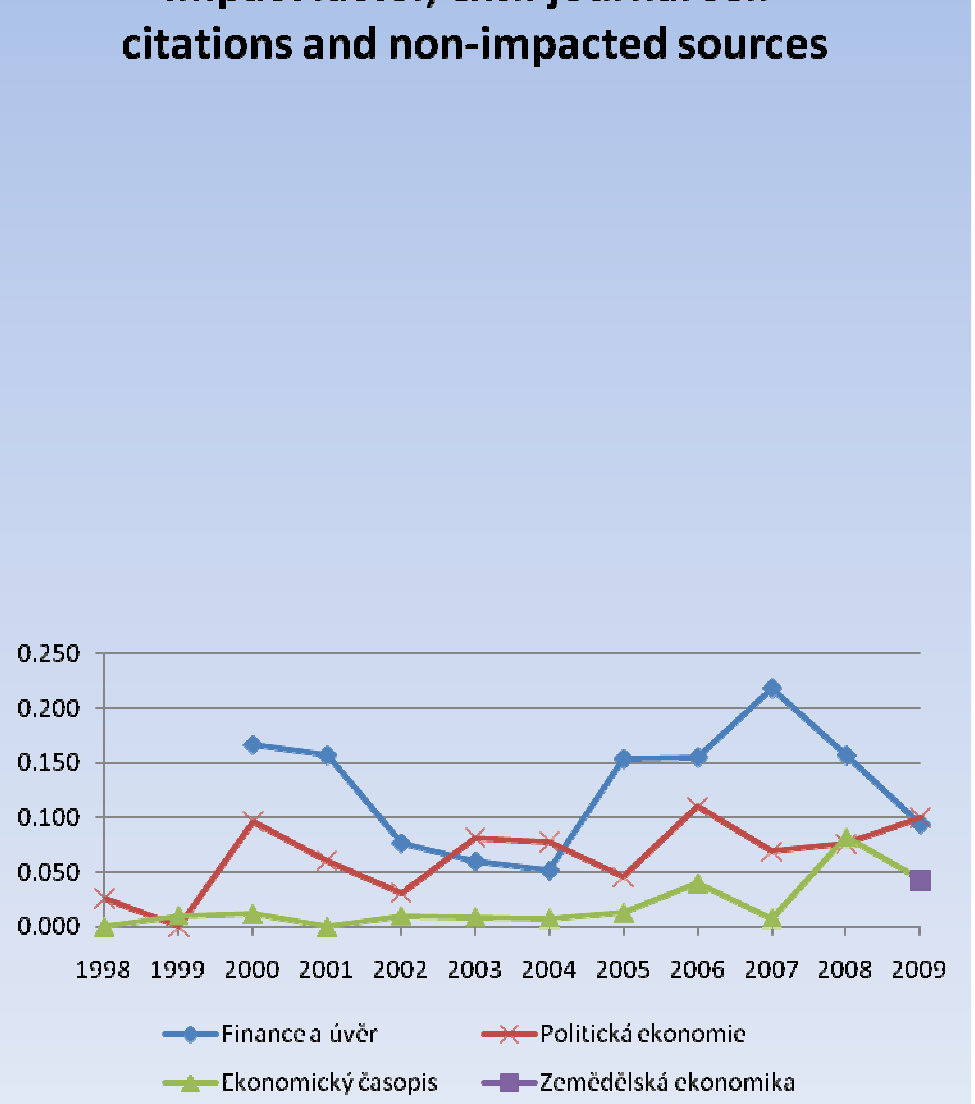


2.B IF inflation & 'national' journals

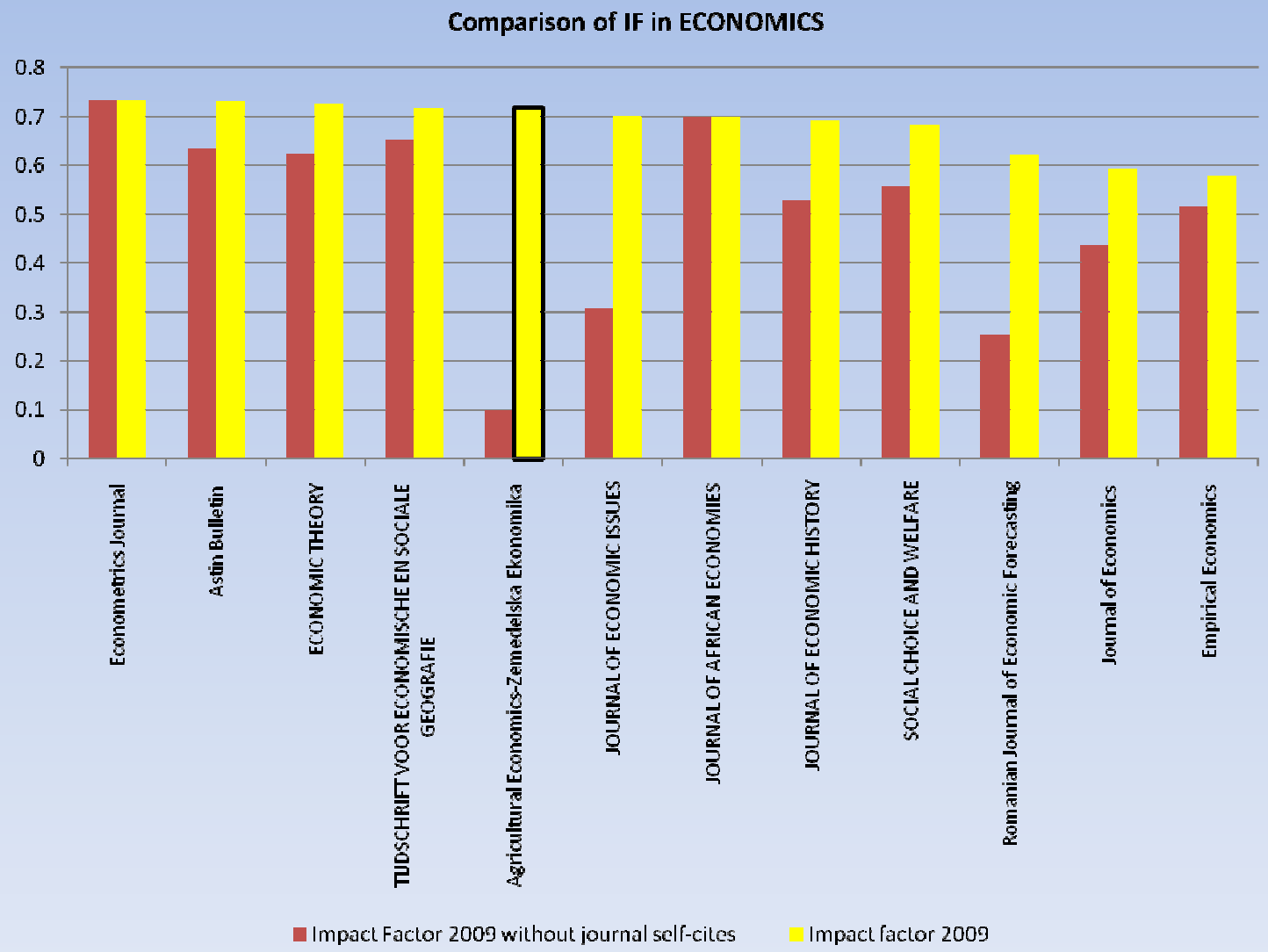
- Formulaic use of IF leads to bizarre behavior
- New WoS IF journals mushrooming with high IF based on within-journal / within-country cites
- The ability to control/manipulate an IF this way a key feature of 'national' journals, esp. in SS:
  - In 2010 *Transformations in Business and Economics* had 90% journal self-citation rate and 2nd decile IF
  - The 3rd highest IF journal in Economics on the planet in 2011 is another new Lithuanian journal…

2.B IF inflation, cont.: Czech journals

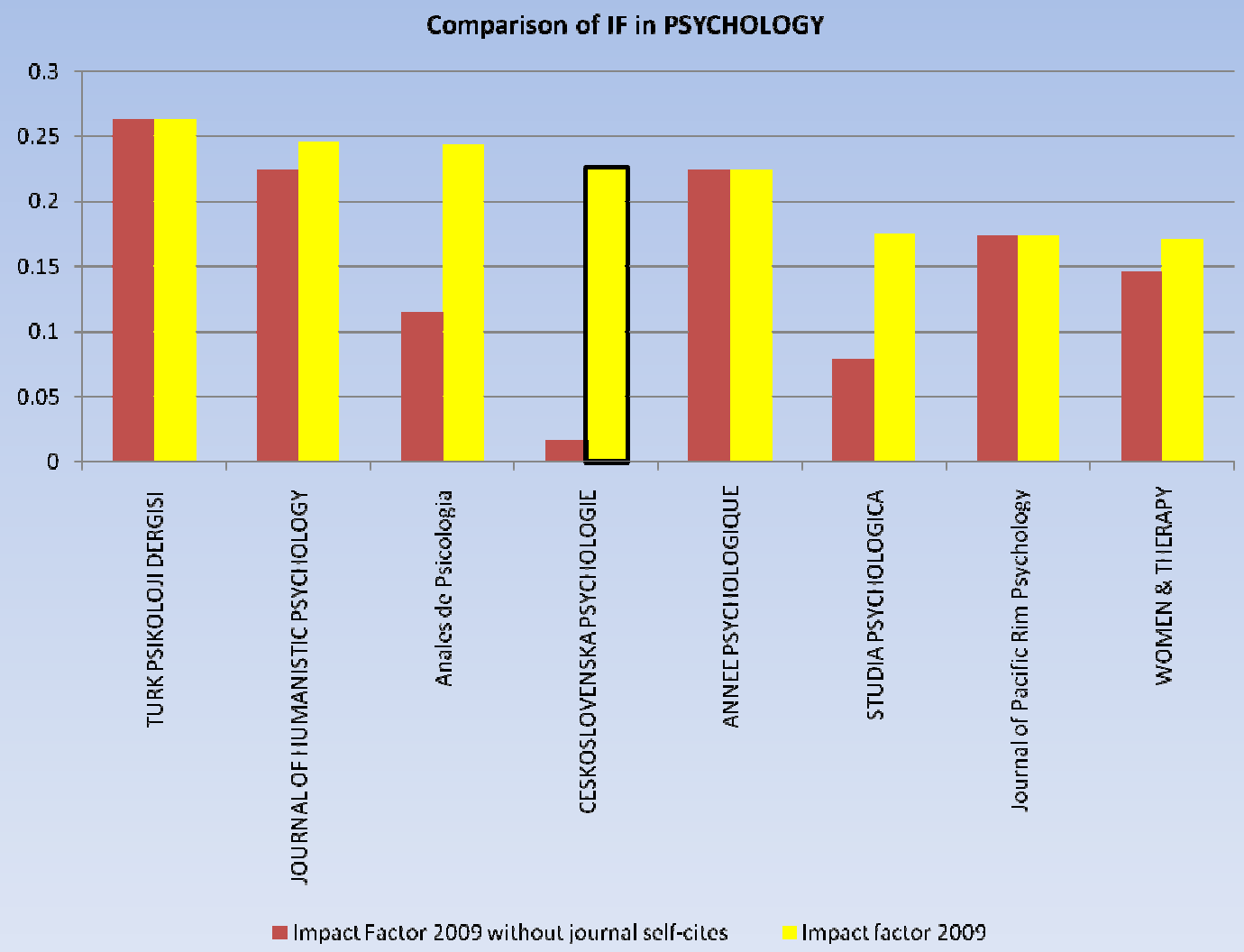




2.B 'national' IF, continued



2.B ‘national’ IF, continued



2.B ‘national’ IF, continued

- There are 4 Czech IF journals in Economics now.

  Politicka Ekonomie: 60% of citation impact from itself, over 90% from the Czech/Slovak Republics. *Macroeconomic Dynamics:* identical IF, but only 11% of citations from within the journal and the rest is international, including citations from high IF.
- Share of local IF output in SS:

  Sweden 1%, Czech Republic: 71%.
- National journals are necessary, but how many and what to make of their IF, which they control directly

2.B Use of IF in all fields with little cost/frequency normalization

- Evaluations in other countries account for field differentials in costs of research production
- IF is not an optimal citation measure (it does not account for the quality of the citing source, Scopus indexes (JCR a SNIP) do)
- The British REF exercise: out of 36 sub-panels only 12 use citation data (including economics) and only as input into peer review (that can easily eliminate rid of bogus journals)

2.B Differences in natural IF frequency

TABLE 3
Journal Publication Rates and Related Information for Eight Disciplines

Discipline	Annual journal publications per faculty (USA)	Acceptance rates in top five journals (%)[a]	Number of authors per per article[b]	Equivalent pages per per article[c]	Equivalent pages per author per year[d]
Economics	0.54[e]	9	1.6	8.8	2.97
Finance	0.26[f]	11	1.9	11.4	1.56
Geology	1.54[g]	51	2.6	13.3	7.87
Psychology	1.80[f]	22	2.2	10.6	8.67
Physics	2.10[f]	69	4.7	6.6	2.95
Oceanography	2.11[g]	72	5.0	8.0	3.38
Chemistry	2.86[h]	55	4.3	7.2	4.79
Geophysics	3.65[g]	69	2.5	8.7	12.70

Notes and Sources:

[a]Our survey of the top 5 journals in each field except oceanography, where only 3 journals provided information.
The top five economics journals we selected were the *American Economic Review, Journal of Political Economy, Southern Economic Journal, E*
the top 5 journals in the other disciplines, we asked department chairs to rank the "best" journals in their respective disciplines. In finance, the
Financial Economics, Journal of Financial and Quantitative Analysis, Review of Financial Studies, and the *Journal of Banking and Finance*. The pe
nal of Experimental Psychology, Psychological Bulletin, Psychological Review, and the *Journal of Applied Psychology*. The physics journals we
Review A, B, C, and D. The geology journals were Bulletin of the Geological Society of America, *Geology*, Bulletin of the American Association
and *American Journal of Science*. The oceanography journals were *Journal of Physical Oceanography, Journal of Geophysical Research (green)*,
nals were *Journal of Geophysical Research (JGR-Red Solid Earth), Geophysics, Tectonophysics, Geophysical Journal International*, and *Geophy*
were *Journal of the American Chemical Society, Journal of Biological Chemistry, Biochemistry, Journal of Medicinal Chemistry*, and the *Journal*
[b]Based on 1992–93 data for the top 5 journals in each discipline, except for economics, where data on the top 24 journals were used.
[c]Equivalent page per article was estimated as average number of words per other discipline journal page divided by average number of words per *A*
discipline" journals and top 24 economics journals were included in the sample. Estimates were adjusted for number of co-authors in Table 1.

2.C 'Soft' output distorts everything

	Evaluation year	2010	Growth rate (%)	2009	2008	2007	2006
	Years counted	**2005-9**	2010/09	**2004-8**	**2003-7**	**2002-6**	**2001-5**
Jimp	**Article in WOS journal**	35617	8	33056		29773	25478
	Article in SCOPUS or ERIH journal	14113	14	12352			
	Article in Czech journal-reviewed	19263	**30**	14824			
Jneimp	**Article in non-WOS journal- Total**	33376	23	27176		47445	46581
J	**Article in journal-Total**	68992	15	60232	40124	77218	72059
B,C	**Book, chapter**	21096	**61**	13094	13111	17756	18740
B	**Book**					7164	6468
C	**Chapter**					10592	12272
D	**Proceedings**	7481	**66**	4501	2730	104340	83713
P	**Patent**	229	-38	371	276	562	363
F	**Utility model, industrial design**	566	**210**	183			
G	**Prototype, functional model**	2225	**143**	915			
H	**Results implemented into legislation or standards**	183	**215**	58			
N	**Certified method**	1325	**393**	269			
R	**Software**	1692	**192**	580			
V	**Secret report**	8	-98	400	2		
S	**Prototype, applied method**	3065	-7	3284	3133	1077	
Z	**Trial operation, variety, breed**	902	52	593			
T	**Prototype, trial operation**	352	-36	551			
Z*	**Trial operation, verified technology, variety, breed**	1253	10	1144	887	1676	1471
L	**Specialized maps**			105			
	Total number of items	*108116*	28	*84744*	*60263*	*202630*	*176350*

Numbers in black are taken from the webpages of RVVI, numbers in red are Technology Centre calculations, based on the above data

*This category was named Technologies (T) in 2006 and may include also some other types of results.

2.C ‘Points’ in Social Sciences, 2010

Shares of output types	IF	non-IF journals	Books	Proceedings	Total
AH - Economics	**23**	25	48	4	100
AN – Psychology	43	18	34	5	100
AO - Sociology, demography	27	19	52	1	100

Summing Up How 'Points' Work: Example of a Social Science: Economics

Table 7.3. ISI coverage indicators per discipline

Discipline	*1a Importance of journals (%)*	*1b ISI coverage of journal literature (%)*	*1a*1b Overall ISI coverage (%)*
Molecular biology & biochemistry	96	97	92
Biological sciences related to humans	95	95	90
Chemistry	90	93	84
Clinical medicine	93	90	84
Physics & astronomy	89	94	83
* Total ISI *	84	90	75
Applied physics & chemistry	83	89	73
Biological sciences ~ animals and plants	81	84	69
Psychology & psychiatry	75	88	66
Geosciences	77	81	62
Other social sciences ~ medicine & health	75	80	60
Mathematics	71	74	53
Economics	59	80	47
Engineering	60	77	46
Other social sciences	41	72	29
Humanities & arts	34	50	17

Disciplines are ranked by descending overall ISI coverage (last column).
Definition of the indicators:
Importance of journals as communication media: % References to documents published in journals relative to total references.
ISI cov ... SI source journals
Overall ... elative to total references. The latter indicator is the numerical product of the first two indicators.

Internationally relevant basic research in economics is published predominantly in impact factor journals.

Applied physics & chemistry includes amongst others the journal categories applied physics,

Example of ‘points’ in Economics (2004-08)

% shares of articles by IF tercile, institutions A, B and C

Institution	IF tercile Top	Middle	Bottom
A - share on economics	70	54	18
A - share on economics, political sc., sociol., psychol.	28	25	8
B - share on economics, political sc., sociol., psychol.	23	13	24
C - share on economics, political sc., sociol., psychol.	19	24	29

B is Academy of Sciences, C is Charles University (both excluding A)

% share of Economics ‘points’ by type of output

	Type of output IF	non-IF	Books	Total %
A	46	15	4	16

3. Thomson Reuters Analysis

- Publ./citation aggregates by institution, field
- No documentation, vague variable description
- No systematic check on address assignment
- Inappropriate comparisons (for example: h-index for institutions of different size or field)
- No information on inputs, so no productivity
- Results not published yet, after two years

4. Technology Center Analysis

- To provide a bibliometric evaluation of field productivity for priority setting exercises
- Not released yet
- In addition to RCIO gives some output size info (shares of Czech publ. on world field output)
- No input data, so no productivity assessment

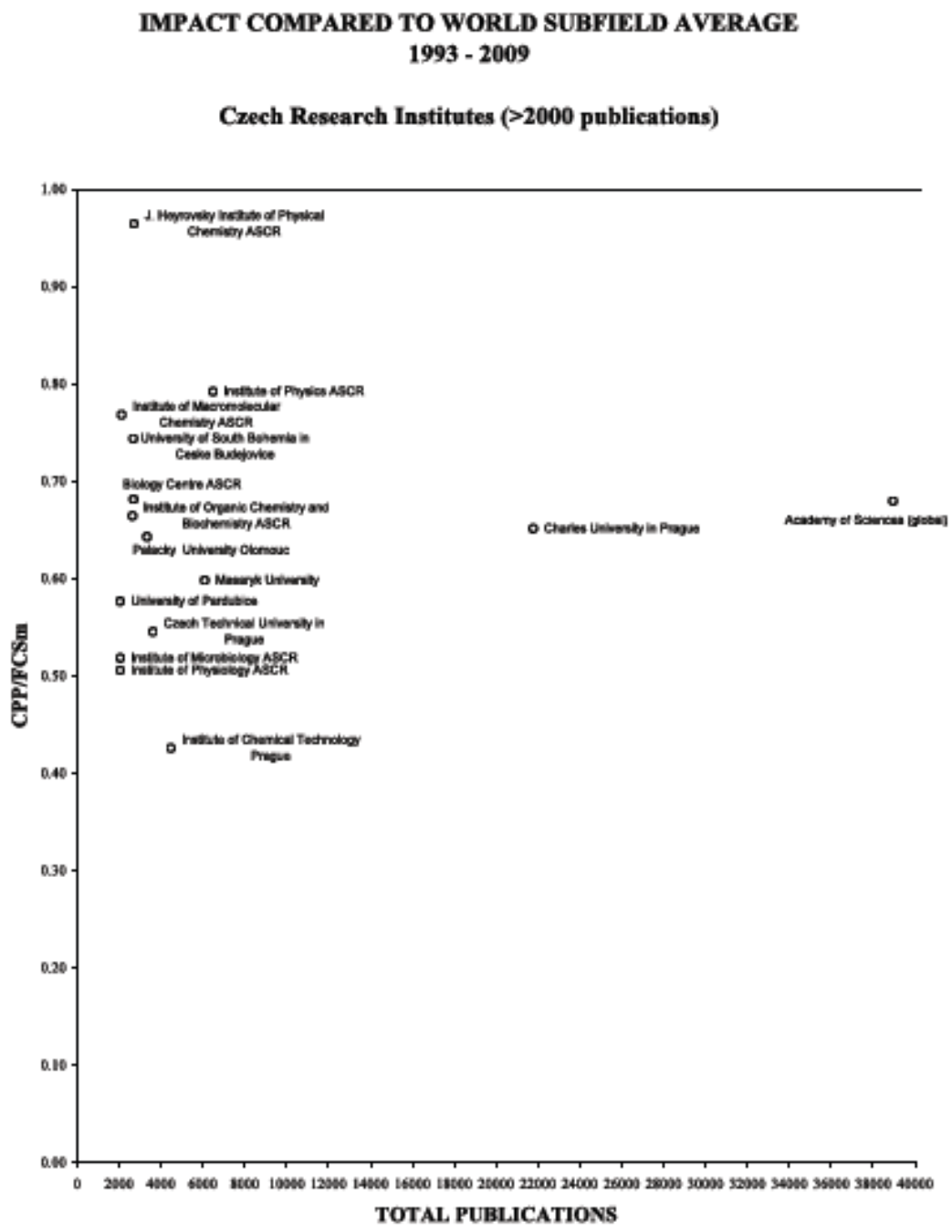
  (the share of an institution on Czech output or its number of RCIO>1 publications says little about productivity without scaling by inputs)

5. CWTS analysis

+ Normalized citation impact against aggregate output of each unit

- So far no input data, productivity?

- Citation analysis fully counts 'national' journals



Bottom line

- Incentives provided in evaluation/financing methodology are key to scientists' productivity
- Bibliometric data a vital part of accountable evaluations of research in <u>some</u> fields
- No useful Czech bibl. data 22 years after 1989
- Naïve use of bibliometric data = natural disaster

=> Urgent systemic need for transfer of know how and build up of local human capital / culture in evaluation techniques and bibliometric analysis