

TESTING METHODS OF MINIMUM DISTANCE

[illegible]

CONTROLLED TABULAR ADJUSTMENT

Jordi Castro,
University of Catalunya
Department of Statistics and Operations Research

Sarah GIESSING,
Federal Statistical Office of Germany
Division Mathematical Statistical Methods

Controlled Tabular Adjustment

Change cell value of sensitive cells **sufficiently**,
and adjust values of non-sensitive cells **minimally**
to maintain additive table structure:

$$\min_x \|x - a\|_L$$

subject to $Ax = b$

$$\underline{a}_i \leq x_i \leq \bar{a}_i \quad i = 1, \dots, n$$

$$x_i \leq a_i - lpl_i \text{ or } x_i \geq a_i + upl_i \quad i \in P$$

Controlled Tabular Adjustment

Change cell value of sensitive cells **sufficiently**,
and adjust values of non-sensitive cells **minimally**
to maintain additive table structure:

$$\min_x \|x - a\|_L$$

$$\text{subject to } Ax = b$$

$$\underline{a}_i \leq x_i \leq \bar{a}_i \quad i = 1, \dots, n$$

$$x_i \leq a_i - lpl_i \text{ or } x_i \geq a_i + upl_i \quad i \in P$$



Controlled Tabular Adjustment

Change cell value of sensitive cells **sufficiently**,
and adjust values of non-sensitive cells **minimally**
to maintain additive table structure:

$$\min_x \|x - a\|_L \quad \leftarrow$$

subject to $Ax = b$

$$\underline{a}_i \leq x_i \leq \bar{a}_i \quad i = 1, \dots, n$$

$$x_i \leq a_i - lpl_i \text{ or } x_i \geq a_i + upl_i \quad i \in P$$

Variants tested

$$\min_{z^+, z^-} \sum_{i=1}^n w_i (z_i^+ + z_i^-)$$

L1-distance

subject to $A(z^+ - z^-) = 0$

$$0 \leq z_i^+ \leq \bar{z}_i \quad i = 1, \dots, n$$

$$0 \leq z_i^- \leq -\underline{z}_i \quad i = 1, \dots, n$$

$$\left\{ \begin{array}{l} z_i^+ \geq upl_i \\ z_i^- = 0 \end{array} \right\} \quad \text{or} \quad \left\{ \begin{array}{l} z_i^- \geq lpl_i \\ z_i^+ = 0 \end{array} \right\} \quad i \in P$$

$$\min_z \sum_{i=1}^n w_i z_i^2$$

L2-distance

subject to $Az = 0$

$$\underline{z}_i \leq z_i \leq \bar{z}_i \quad i = 1, \dots, n$$

$$z_i \leq -lpl_i \quad \text{or} \quad z_i \geq upl_i \quad i \in P$$

Controlled Tabular Adjustment

Change cell value of sensitive cells **sufficiently**,
and adjust values of non-sensitive cells **minimally**
to maintain additive table structure:

$$\min_x \|x - a\|_L$$

subject to $Ax = b$

$$\underline{a}_i \leq x_i \leq \bar{a}_i \quad i = 1, \dots, n$$

$$x_i \leq a_i - lpl_i \text{ or } x_i \geq a_i + upl_i \quad i \in P$$

$$(1 - \alpha_i) a_i \leq x_i \leq (1 + \beta_i) a_i \quad i = 1, \dots, n$$

„Restricted CTA method“

Quality Criteria

- It should be possible to combine CTA and cell suppression procedures
 - CTA must be able to provide a feasible solution, also when changes are allowed only in a given subset of cells.

Quality Criteria

- It should be possible to combine CTA and cell suppression procedures
 - CTA must be able to provide a feasible solution, also when changes are allowed only in a given subset of cells
- CTA procedure should produce as much accurate information as cell suppression, considering the data's relevance and significance
 - Overall, few cells with relative changes exceeding fixed threshold (like 5%, 10%,...)
 - On high hierarchical levels:
few relative changes $> O(1/n)$, or
 $> 1/\sqrt{X}$, or

Test Results

CTA Variants				
Weight				
Distance				
Range	Number of cells by ranges of relative deviation			
0				
0-2				
2-5				
5-10				
>10				
Number of high-level cells with large relative deviation				
> 1/N				
> 1/√X				
> 1/∛X				

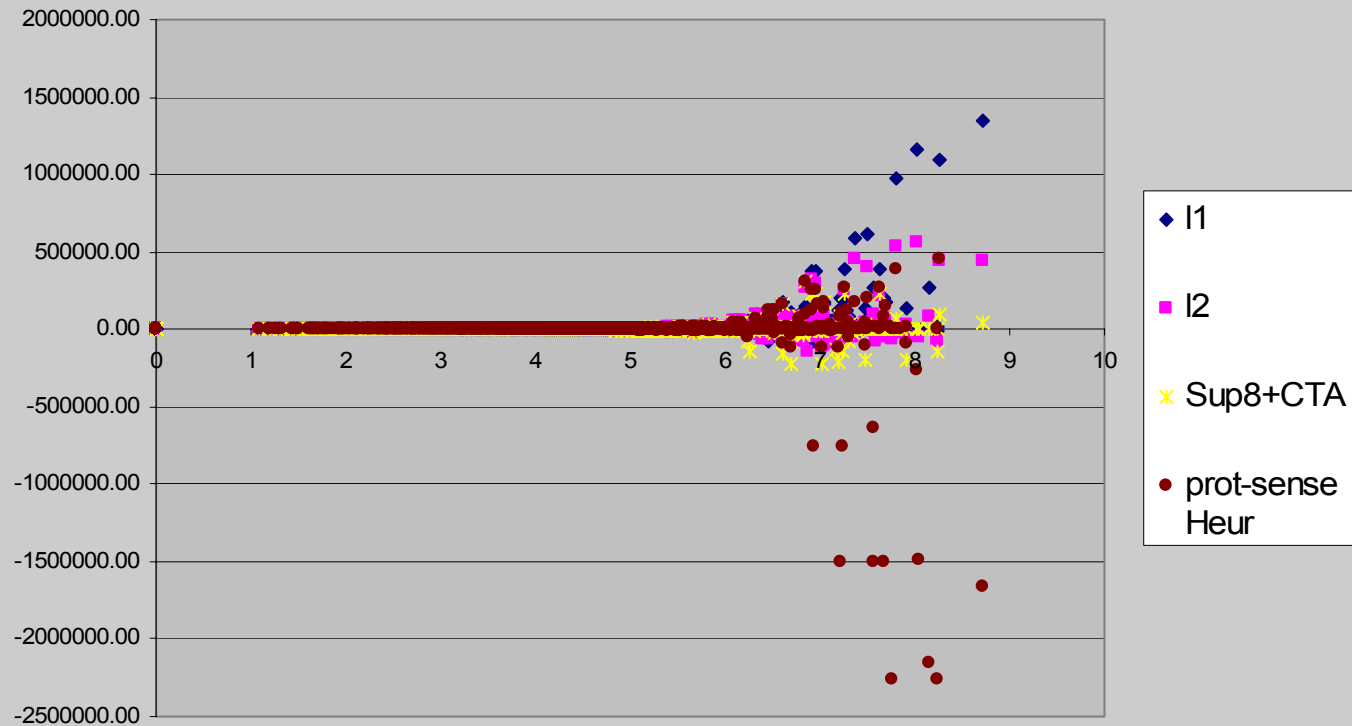
Cell suppression: 888 (20 high-level) secondary suppressions

Test Results

CTA Variants			
Weight	1/a		
Distance	L1	L2	Sup8 L1
Range	Number of cells by ranges of relative deviation		
0	2439	0	2341
0-2	655	2841	641
2-5	119	309	169
5-10	4	61	68
>10	3	9	1

Cell suppression: 888 (20 high-level) secondary suppressions

Deviations to original cell values



Test Results

CTA Variants			
Weight	1/a		
Distance	L1	L2	Sup8 L1
Range	Number of cells by ranges of relative deviation		
0	2439	0	2341
0-2	655	2841	641
2-5	119	309	169
5-10	4	61	68
>10	3	9	1
Number of high-level cells with large relative deviation			
> 1/N	85	90	42
> $1/\sqrt{X}$	83	82	40
> $1/\sqrt[3]{X}$	38	17	16

Cell suppression: 888 (20 high-level) secondary suppressions

Test Results

CTA Variants					
Weight	$1/a$		const.		
Distance	L1	L2	Sup8 L1	L1	L2
Range	Number of cells by ranges of relative deviation				
0	2439	0	2341	2164	0
0-2	655	2841	641	540	1812
2-5	119	309	169	164	396
5-10	4	61	68	78	233
>10	3	9	1	274	779
Number of high-level cells with large relative deviation					
> 1/N	85	90	42	37	90
> $1/\sqrt{X}$	83	82	40	33	65
> $1/\sqrt[3]{X}$	38	17	16	15	19

Cell suppression: 888 (20 high-level) secondary suppressions

Test Results

CTA Variants							
Weight	$1/a$			$const.$		$1/\log(a)$	
Distance	L1	L2	Sup8 L1	L1	L2	L1	
Range	Number of cells by ranges of relative deviation						
0	2439	0	2341	2164	0	2300	
0-2	655	2841	641	540	1812	644	
2-5	119	309	169	164	396	136	
5-10	4	61	68	78	233	42	
>10	3	9	1	274	779	98	
Number of high-level cells with large relative deviation							
> 1/N	85	90	42	37	90	51	
> $1/\sqrt{X}$	83	82	40	33	65	48	
> $1/\sqrt[3]{X}$	38	17	16	15	19	18	

Cell suppression: 888 (20 high-level) secondary suppressions

Test Results

CTA Variants							
Weight	$1/a$			<i>const.</i>		$1/\log(a)$	adapted to hierarchy
Distance	L1	L2	Sup8 L1	L1	L2	L1	L1
Range	Number of cells by ranges of relative deviation						
0	2439	0	2341	2164	0	2300	2320
0-2	655	2841	641	540	1812	644	577
2-5	119	309	169	164	396	136	124
5-10	4	61	68	78	233	42	63
>10	3	9	1	274	779	98	136
Number of high-level cells with large relative deviation							
$> 1/N$	85	90	42	37	90	51	8
$> 1/\sqrt{X}$	83	82	40	33	65	48	11
$> 1/\sqrt[3]{X}$	38	17	16	15	19	18	7

Cell suppression: 888 (20 high-level) secondary suppressions

Future Work

- Continue work on weights
- Restricted CTA: develop suitable method to determine adjustment senses for sensitive cells
- Processing of multivariate data
- CTA for Table Server



The image shows two open pages of a statistical table from the Statistisches Bundesamt. The left page is titled '1.3 Durchschnittswerte' and the right page is titled '2.1 Durchschnittswerte'. Both pages contain multiple columns of data, including years, indices, and various statistical measures. The tables are organized into sections, with the left page having a 'Basis' section and the right page having a 'Basis' and 'Index' section. The data is presented in a structured, tabular format with clear headings and sub-headings.

THANKS FOR YOUR ATTENTION

Name	n	$ \mathcal{P} $	m	N.coef
bts4	36570	2260	36310	136912
destatis	5940	621	1464	18180
five20b	34552	3662	52983	208335
five20c	34501	4022	58825	231345
hier13	2020	112	3313	11929
hier16	3564	224	5484	19996
nine12	10399	1178	11362	52624
nine5d	10733	1661	17295	58135
ninenew	6546	858	7340	32920
two5in6	5681	720	9629	34310

Table 1: Dimensions of the complex instances

Instance	$\gamma = 0$		$\gamma = 1/2$		$\gamma = 1$	
	L_1	L_2	L_1	L_2	L_1	L_2
bts4	1402	1515	1016	1184	962	933
destatis	164	396	125	416	119	309
five20b	2841	3013	2478	2815	2426	2605
five20c	3218	3477	2769	3096	2777	2822
hier13	101	103	75	82	79	68
hier16	127	145	108	124	112	95
nine12	787	889	685	787	695	709
nine5d	875	999	947	993	978	918
ninenew	613	646	521	598	531	510
two5in6	451	529	388	499	424	384

a) relative deviation between 2% and 5%

Instance	$\gamma = 0$		$\gamma = 1/2$		$\gamma = 1$	
	L_1	L_2	L_1	L_2	L_1	L_2
bts4	741	799	353	521	279	292
destatis	352	1012	11	524	7	70
five20b	1284	1434	650	1161	445	579
five20c	1352	1542	699	1202	559	706
hier13	32	32	26	27	26	24
hier16	60	69	29	46	17	112
nine12	378	427	162	310	120	149
nine5d	606	724	223	523	163	128
ninenew	298	360	154	258	107	131
two5in6	244	80	128	163	90	86

b) relative deviation greater than 5%

Table 2: Number of cells with a relative deviation between 2% and 5% (a)), and greater than 5% (b)), for $\gamma = 0, 1/2, 1$ and the complex instances