

Comparison Programs

Presented by

Robert M^cDougall

Center for Global Trade Analysis

Purdue University

quality control

- **check types**
 - directed checks
 - undirected checks
- **control points**
 - data inputs
 - intermediate data sets
 - final output
- **ignorance**
- **large data sets**

strategy

- **comparison**
- **feature selection**
- **generic approach**

comparison

- **new to old**
- **contributor to contributor**
- **data set to data set**
 - **energy to I-O**
 - **trade to I-O**
- **stage to stage**

generic approach

- **compare single pair of arrays**
 - bilateral trade data
 - use table
- **non-negative numbers**
- **inherent scale**
- **features:**
 - elements
 - marginal totals

feature selection

- large changes in large values
- function of absolute and relative difference
- $(V_B - V_A) \cdot (\log V_B - \log V_A)$
- symmetric
- non-negative data
- matching non-zeros

sign agreement

- **smearing**

- $V(i,j) = (1 - \varepsilon) \cdot V(i,j) + \varepsilon \cdot [V(i,\bullet) \cdot V(\bullet,j) / V(\bullet, \bullet)]$

- **mutual contamination**

- $V_A = (1 - \varepsilon) \cdot V_A + \varepsilon \cdot V_B$

- $V_B = (1 - \varepsilon) \cdot V_B + \varepsilon \cdot V_A$

first try

- compare array elements
- repetitious
- favors clumpy slices

1	3	5
2	4	6

1	3	5
4	8	12

0	4	0
2	2	2
3	3	3

0	8	0
2	2	2
6	6	6

second try

- **Compare array totals.**
- **Rescale.**
- **Compare rank 1 subtotals.**
- **Rescale.**
- **Compare rank 2 subtotals.**
- *et cetera*

example

1	1	1
1	1	1
2	2	2

2	2	2
4	4	4
2	2	2

array scaling factor: 2

2	2	2
2	2	2
4	4	4

2	2	2
4	4	4
2	2	2

defects

Sometimes, the main action is not in the higher level totals:

1	1	1
1	1	1
1	1	1

1	1	1
1	1	1
4	4	4

third try

- **Compare all features: array total, marginal totals, individual elements.**
- **Identify most divergent feature.**
- **Rescale to eliminate that divergence.**
- **Compare all features.**
- **Identify greatest remaining divergence.**
- **Rescale.**
- *et cetera*

example: merchandise trade

Comm.	Source	Destn	Actual		Adj.	
*	*	*	5781272	8399882	5781272	8399881
oil	nga	ind	0	7997	2	7995
*	*	chn	235977	503906	342575	503866
otn	jpn	pan	4528	0	6440	2
lea	*	hkg	10904	489	15520	493
ele	chn	*	67462	179094	96245	179074

example: GTAP data

input	use	region	Actual		Adj.	
*	*	xee	318005	121845	317957	121893
osg_d	osg	can	77	30555	84	30548
dwe_d	mvh	mex	13603	0	13600	4
*	wtr	usa	27265	107037	27285	107018
dwe_d	crp	mex	11784	0	11781	3
ome_m	govt	xws	14526	23	14522	27

state of play

- **from rank 3 arrays to rank 2 and 4**
- **from money value comparisons to:**
 - **price data comparisons: Compare money values with different prices and common quantities.**
 - **tax rate comparisons: Compare tax-paid values with common quantities and tax-free prices.**