Theory of Quantitative Trend Analysis and its Application to the South African Elections

Presented at the CSIR R&I Conference

Dr Jan Greben

Logistics and Quantitative Methods

CSIR Built Environment

Pretoria

28 February 2006

our future through science

Outline

- 1. Introduction
- 2. CSIR Election Night Forecasting
- 3. Trends in Elections
- 4. Mathematical Construction of Trend Matrices
- 5. Analysis of Some Results
- 6. Other areas of applications

1. Introduction

- Different uses of Trends
- When can we apply quantitative trend analysis?
- Application to elections
- Less data-intensive methods based on “common sense”
- Possible uses in marketing

2. CSIR Election Night Forecasting

- Applied Election Night Forecasting Model in 1999, 2000 and 2004
- Based on cluster decomposition of the Electorate
- Will again be applied to 2006 Municipal Elections
- On 1 March we will also apply trend predictions

3. Trends in Elections

Example of matrix describing trends

from 1999 to 2004 elections

Table 4. Kuhn-Tucker correction to trend matrix

KT Party 2004	Party 1999 Results	ANC 66.7	DP 9.4	IFP 8.2	NNP 6.7	UDM 3.6	ACDP 1.4	VF 0.8
ANC	69.7	97.4	0.0	11.5	4.6	42.9	48.9	23.5
DA	12.4	0.2	94.0	1.4	38.2	0.0	0.0	17.3
IFP	7.0	0.0	0.0	84.7	0.0	0.0	0.0	0.0
UDM	2.3	0.3	0.0	0.0	0.0	56.3	0.0	0.0
ID	1.7	0.0	1.2	0.0	18.5	0.9	19.4	0.0
NNP	1.7	0.0	0.0	0.0	24.5	0.0	0.0	0.0
ACDP	1.6	0.4	3.6	0.7	6.7	0.0	31.7	0.0
VFP	0.9	0.0	1.2	0.4	3.5	0.0	0.0	59.2

4. Mathematical Construction of Trend Matrices

1999 Election Results, v = voting district V = 15 000 P_{old} parties

$$\sum_{p=1}^{P_{old}} x_p^{(v)} = 100 , \quad v = 1 \cdots V$$

2004 Election Results

$$\sum_{p=1}^{P_{new}} y_p^{(v)} = 100 , \quad v = 1 \cdots V ,$$

2004 Election Results, v = voting district V = 17 000 P_{new} parties

Relate the two results:

$$y_p^{(v)} = \sum_{p'=1}^{P_{old}} S_{pp'} \, x_{p'}^{(v)} , \qquad p = 1, \cdots P_{new}$$

CSIR our future through science

Problems with the Mathematical Construction of the Trend Matrix

- Matrix can not be constructed for a single result!

- Matrix has P_{new} * P_{old} elements, while there are P_{new} + P_{old} Election input results
- Hence, we need to construct average matrix from many results
- Resulting Matrix is optimal, but not necessarily positive

our future through science

Optimization

Define Objective Function:

$$J = \frac{1}{2}\sum_{v=1}^{V} N_v \sum_{p=1}^{P_{new}} (y_p^{(v)} - \sum_{p'=1}^{P_{old}} S_{pp'} x_{p'}^{(v)})^2 - \sum_{p'=1}^{P_{old}} \varepsilon_{p'} (\sum_{p=1}^{P_{new}} S_{pp'} - 1),$$

Minimize this subject to variations in the trend matrix

Resulting Matrix:

$$\underline{\underline{S}} = \underline{\underline{X}}\ \underline{\underline{A}}^{-1}$$

$$(\underline{\underline{A}})_{pp'} \equiv A_{pp'} = \sum_{v=1}^{V} N_v x_p^{(v)} x_{p'}^{(v)}$$

$$(\underline{\underline{X}})_{pp'} \equiv X_{pp'} = \sum_{v=1}^{V} N_v y_p^{(v)} x_{p'}^{(v)}$$

Predictions

The following formula can be used for prediction

$$y_p^{(v)} = \sum_{p'=1}^{P_{old}} S_{pp'}(t)\, x_{p'}^{(v)} \qquad\qquad p = 1, \cdots P_{new}.$$

where:

$$\underline{\underline{X}}(t)_{pp'} = \sum_{v \in \Omega(t)}^{V} N_v\, y_p^{(v)} x_{p'}^{(v)} \qquad \Omega(t) = counted \quad results$$

$$(\underline{\underline{A}})_{pp'} = \sum_{v=1}^{V} N_v\, x_p^{(v)} x_{p'}^{(v)}$$

$$\underline{\underline{S}} = \underline{\underline{X}}(t)\, \underline{\underline{A}}^{-1}$$

5. Analysis of Some Results:

Original Result contains negative elements

Table 1. Basic trend matrix characterizing the trends between the 1999 and 2004 elections in South Africa

Party 2004	Party 1999 / Results	ANC 66.7	DP 9.4	IFP 8.2	NNP 6.7	UDM 3.6	ACDP 1.4	VF 0.8
ANC	69.7	97.0	-0.9	11.0	3.2	43.2	48.0	68.4
DA	12.4	0.3	95.1	1.6	37.8	-1.7	-1.0	8.5
IFP	7.0	0.0	0.9	86.1	0.7	-0.2	-6.9	5.3
UDM	2.3	0.7	-3.4	0.1	-2.9	57.8	15.8	3.1
ID	1.7	0.1	5.8	-0.5	19.1	0.6	16.5	-28.7
NNP	1.7	0.3	-2.0	0.0	30.8	-0.3	-5.9	-23.3
ACDP	1.6	0.4	4.1	0.7	5.8	-0.1	30.5	2.1
VFP	0.9	0.0	0.3	0.1	1.9	-0.1	-0.2	64.6

our future through science

Various Ways to make Trend Matrix Positive

Table 2. Renormalized trend matrix

Party 2004	Party 1999 Results	ANC 66.7	DP 9.4	IFP 8.2	NNP 6.7	UDM 3.6	ACDP 1.4	VF 0.8
ANC	69.7	97.0	0.0	10.9	3.1	42.1	41.4	43.4
DA	12.4	0.3	88.8	1.6	36.6	0.0	0.0	5.4
IFP	7.0	0.0	0.9	85.4	0.7	0.0	0.0	3.3
UDM	2.3	0.7	0.0	0.1	0.0	56.3	13.6	2.0
ID	1.7	0.1	5.4	0.0	18.5	0.6	14.2	0.0
NNP	1.7	0.3	0.0	0.0	29.9	0.0	0.0	0.0
ACDP	1.6	0.4	3.8	0.7	5.6	0.0	26.3	1.3
VFP	0.9	0.0	0.2	0.1	1.8	0.0	0.0	40.9

Table 3. Heuristic correction to trend matrix

Party 2004	Party 1999 Results	ANC 66.7	DP 9.4	IFP 8.2	NNP 6.7	UDM 3.6	ACDP 1.4	VF 0.8
ANC	69.7	96.7	0.2	14.1	6.7	44.1	43.0	43.9
DA	12.4	0.9	87.9	1.3	40.1	6.0	2.0	5.8
IFP	7.0	0.0	1.1	81.7	1.1	0.1	0.2	3.3
UDM	2.3	0.5	0.0	0.0	0.0	45.9	11.1	1.6
ID	1.7	0.0	4.6	0.0	15.2	1.3	11.6	0.0
NNP	1.7	0.2	0.0	0.0	22.2	0.3	0.2	0.0
ACDP	1.6	0.4	3.9	1.0	6.3	0.4	26.8	1.4
VFP	0.9	0.1	0.5	0.1	4.4	0.1	0.7	40.2

Less data-intensive methods based on "common sense"

- Base exclusively on overall results
- Simpler objective function with additional criteria
- Examples of criteria:

1. People like to stay with same party (party loyalty)
2. Some people like to switch
3. New parties have strong appeal for short while
4. Parties that loose contribute mostly to winners

Example of a result of the simple approach

Table 7. Trend matrix using Method 1.

Method 1 Party 2004	Party 1999 Results	ANC 66.7	DP 9.4	IFP 8.2	NNP 6.7	UDM 3.6	ACDP 1.4	VF 0.8
ANC	69.7	99.5	2.4	4.8	13.0	11.3	12.6	18.7
DA	12.4	0.4	96.6	5.0	13.2	11.6	13.2	19.5
IFP	7.0	0.0	0.0	85.3	7.6	1.3	0.0	0.0
UDM	2.3	0.0	0.0	0.4	7.6	63.8	0.0	0.0
ID	1.7	0.1	0.9	3.0	10.8	7.2	4.9	7.3
NNP	1.7	0.0	0.0	0.4	32.1	1.3	0.0	0.0
ACDP	1.6	0.0	0.0	0.7	7.9	1.9	69.2	0.0
VFP	0.9	0.0	0.0	0.5	7.7	1.5	0.0	54.6

We see that diagonal elements are much larger than previously

6. Other Areas of Application

Marketing

- Usually only popularity of each product is evaluated
- Now we can link the reduction in purchases in one product to the increase of purchases in other products
- Switching Between Products

6. Other Areas of Application-Continuation

Whenever two cross sections of a process are known:

Given: Electricity is given by sector and by area

Problem: We would like to know the sector demand per area without additional data

Solution: Construct Correlation Matrix with simple assumptions