A theory of quantitative trend analysis and its application to South African general elections

Jan M. Greben*

Trends are usually defined as progressive changes in a particular phenomenon. If the phenomenon can be characterized by one variable over time (such as the price of some item), then a trend analysis is fairly simple. Phenomena are often characterized, however, by more than one variable at particular discrete time intervals. In such cases a trend analysis becomes more complex and ambiguous. If enough data are available, however, trends between two subsequent times can be represented by so-called transition matrices. The theory of such matrices for the description of voting patterns was developed in the United Kingdom in the 1970s. South African general elections represent an ideal test case for such theories as both the number of independent results (17 000 voting districts versus about 240 three-way contested constituencies in the U.K.) and the number of parties (about 20 versus 4 in Britain) are much greater in the South African case. This paper reports such a trend analysis using extensions of previous methods, and introduces two new methods to avoid negative transition matrix elements. The applicability of the old and new methods is discussed and their results examined. Other areas of application of transition matrices to trends are briefly reviewed.

Introduction

Trends arise in many facets of society. They emphasize change: what is popular now in relation to what prevailed before. In many cases the popularity of certain items can be characterized numerically, such as in sales records. The changes can then be described in numbers. If data for a particular item are available for a continuous period of time, then one can conduct statistical analyses and establish the general behaviour by eliminating seasonal and accidental changes. The resulting outcome is called a trend-line. However, phenomena often are known or measured only at particular times and have a multi-dimensional nature. One can then define the trend as the change from one state to another. Such transitions can be represented mathematically by a transition matrix, the construction of which is discussed in this paper in the context of two consecutive South African national elections.

I first briefly review the theory of transition matrices. When there is a small number of discrete initial and final states, Markov models can profitably be used. In these models the elements of the Markov matrix represent the transition rate from one state to another (for a recent application, see Pelzer *et al.*[1]). However, if the states have many components which can assume continuous values, as is true in the case of elections, the concept of a transition matrix is more applicable. Extensive theoretical work on such techniques was conducted in the period 1969–1980 in the context of British elections. The first paper of note was written by Hawkes,[2] who discussed swings in voting outcomes using a transition matrix. The matrix was determined using a regression approach applied to a number of constituencies. However, his method does not ensure non-negativity of the transition matrix elements. Miller[3] applied a similar analysis in 1972. Like Hawkes, Miller emphasizes the need to consider constituencies with similar characteristics to justify the averaging process for the matrix. However, neither author was particularly successful in ensuring this stability. I have also considered the construction of groups of voting districts with similar characteristics,[4] but the limitation to more homogeneous samples comes at a price, as it is the diversity in voter patterns which allows one to construct unique transition matrices. McCarthy and Ryan[5] emphasized the importance of ensuring a positive transition matrix and introduced methods to do so. They also discussed the consequences of a changing electorate. Their work was subsequently criticized by Upton[6] in a study of voter surveys. This led to criticism of the mathematical techniques employed by McCarthy and Ryan. In the current paper, I discuss the merits and demerits of these various approaches and justify the use of my methods.

The application concerns the voting trends between the 1999 and 2004 South African national elections. These elections are suited to a trend analysis, because results are available for all 17 000 voting districts, which gives much better statistics than the constituency results in British polls (for a discussion of the 1999 and 2004 elections in South Africa, see refs 7 and 8, respectively). The transition matrix tells us approximately how the votes from one party in the first election are distributed among the parties in the subsequent voting. This is equivalent to comparing all the sales slips in a supermarket in one period with the purchase records at a later time. In the marketing context, a transition matrix does not just establish what was and what is popular, but also demonstrates which particular products were dropped in favour of certain new items. This transition information is not easy to deduce, but could be very valuable. For instance, retail businesses could plan their marketing strategy, anticipate future client behaviour, and optimize their earnings. In the case of elections, the products are replaced by parties or candidates, as in the British elections.[2] In voting, the trend information also can be of value. For example, political commentators could improve their analysis[9] by taking into account the mathematically determined trends, while political parties could plan their campaigns with a detailed knowledge of voter behaviour. The analysis could even contribute towards democracy, as the wishes of the electorate will be understood more accurately through such an analysis.

Central to a quantitative construction of the transition matrix is the availability of data. In the supermarket, itemized sales records are required. When these records also contain client identification, a more detailed model can be developed, which also can be used for personalized marketing strategies and could be linked to other data-mining strategies. For elections, it is ideal to have all polling outcomes in the smallest possible voting units. The only way to construct exact transition matrices is from individual voting records, which naturally are not available. The smallest political unit in South Africa is the voting district, consisting typically of a thousand voters. These aggregated results obscure voter movements; however, their large number (about 17 000) improves the statistics and allows us in principle to estimate the confidence intervals[2] for the transition matrices.

*Logistics and Quantitative Methods, CSIR, P.O. Box 395, Pretoria 0001, South Africa. E-mail: jgreben@csir.co.za

In the following section I introduce the transition matrix, for which an important issue is the elimination of negative matrix elements. I then discuss its actual application and display the results for some of the main political parties in the most recent South African elections. Other possible applications of these techniques are then discussed.

The development of a transition matrix to describe voting trends

I start by briefly reviewing the formalism, and compare the results of earlier ('old', 1999) and more recent ('new', 2004) elections. The number of parties in each election is indicated by P_{old} and P_{new} (16 and 21, respectively). Individual parties are labelled by the index p. This label does not necessarily refer to the same party in the old and new elections, as parties (even party names) change from one election to the next. The old and new election results in voting district ν are expressed as percentages: $x_p^{(\nu)}$ and $y_p^{(\nu)}$, respectively. These percentages satisfy the equations:

$$\sum_{p=1}^{P_{\text{old}}} x_p^{(\nu)} = 100, \quad \nu = 1 \cdots V_{\text{old}}, \tag{1}$$

and

$$\sum_{p=1}^{P_{\text{new}}} y_p^{(\nu)} = 100, \quad \nu = 1 \cdots V_{\text{new}}, \tag{2}$$

where $V_{\text{old}}(V_{\text{new}})$ is the total number of voting districts in the old (new) elections. The movement of voters in voting district ν from one party to another in the subsequent elections is now characterized using a transition matrix $S_{pp'}$:

$$y_p^{(\nu)} = \sum_{p'=1}^{P_{\text{old}}} S_{pp'}\, x_{p'}^{(\nu)}, \qquad p = 1, \cdots P_{\text{new}}. \tag{3}$$

Some complications arise when comparing the old and new results. For instance, the number of voting districts generally increases with every election, in accordance with the growth of the population. The redefinition of voting districts means that the old and new results may not refer to the same area. However, since there is geographical overlap between old and new districts, it is possible to 'construct' the equivalent 1999 election results for the 2004 voting districts by a suitable mapping procedure.[8] As a result, we can limit ourselves to a comparison of the 1999 and 2004 election results for a single set of V voting districts, where V equals the number of voting districts in 2004 (17 000). This means that we replace V_{old} and V_{new} by a single number V in Equations (1) and (2).

A second complication is that the number of voters in 2004 and 1999 is not the same for individual voting districts. In the work by McCarthy and Ryan,[5] this problem was dealt with by treating the excess (or deficit) of voters in the new elections separately, leading to an extra term in Equation (3). They assumed that all 'additional' voters choose in the same way. I question this assumption and believe it to be more natural to assume that the new voters in a particular voting district would have voted in the same way in the previous election (had they been present) as the actual voters did. Another objection to the approach by McCarthy and Ryan is that the separate treatment of the excess voters is unnatural: the number of new voters is substantially larger than the tally of excess voters, as there are also voters who have left the district. If one wants to follow McCarthy and Ryan's philosophy consistently, then one should develop a model for the incident as well as for the lost voters. The latter could be old people or migrants, who may also have voted similarly before they stopped voting in the constituency. However, the assumption that either of these groups can be characterized by a single voting pattern does not appeal to me, hence these voters are lumped together with the actual voters in the respective polling districts. This approach can be realized by expressing the election results as percentages rather than actual votes, as was done in Equation (3). This also avoids the explicit treatment of voters who abstain.

Let us now continue the discussion of transition matrices. These matrices appear in various areas of science. In physics, the vectors $\vec{x}^{(\nu)}$ and $\vec{y}^{(\nu)}$ would represent initial and final states, while the S-matrix (complex in this case) would govern the transition between the two states. In statistics, it is more common to call states cross sections,[1] as they represent the system parameters at a certain point in time. The transition matrix must express the conservation of probability. In physics, this implies that the S-matrix is unity. In the present case the following condition must be applied:

$$\sum_{p=1}^{P_{\text{new}}} S_{pp'} = 1, \qquad p' = 1, \cdots P_{\text{old}}. \tag{4}$$

This condition ensures that the sum rules (1) and (2) are consistent with Equation (3). This equation suggests that S is a probability matrix. However, this is only the case if all matrix elements are non-negative, which is not necessarily true.

Equations (3) and (4) are not sufficient to determine the transition matrix $S_{pp'}$ uniquely for a single voting district. The transition matrix contains $P_{\text{old}} \times P_{\text{new}}$ elements, while there are only $P_{\text{old}} + P_{\text{new}} - 1$ independent equations (3) and (4). Even for a two-party election, there is not a unique answer. The only exact way to derive this matrix for a single voting district is to obtain the individual voting results after both elections. The secrecy of the ballot makes this impossible. A practical way to obtain direct information on voter movements is to conduct interviews with (a subset of) voters in a subset of voting districts.[6] However, such a procedure also has severe limitations, as was observed by Miller.[3] Despite these limitations, such a survey can provide an insightful test of the mathematical procedure and I will discuss the work by Upton[6] when analysing the significance of my results. Theoretically, one can limit the freedom in choosing the transition matrix by imposing additional constraints, such as the maximum entropy principle. This idea has been applied[10] successfully in marketing to a given set of market shares (comparable to a single ballot result in the elections), and might also be applicable in the current context. However, I will follow a different strategy.

Mathematically, we can remove the ambiguity by applying the same transition matrix to a number of voting districts simultaneously. Naturally, Equation (3) is then no longer satisfied exactly for each voting district considered, but is valid in some average way. This approach requires the assumption that the voting districts considered are governed approximately by the same transition matrix, but at the same time display a wide range of different voting results (otherwise, the non-uniqueness is not resolved). In the current study, the second point is emphasized and the whole set of 17 000 voting districts is included, so as to minimize the non-uniqueness. In ref. 4, I consider the effect of using subsets of voting districts, selected for similarity in voting patterns, thereby constructing transition matrices for particular segments of the population. The subsets of voting districts (denoted as clusters) are described in detail in ref. 11. Miller[3] also has studied the question of the variation of transition matrices with different types of constituencies.

To find the optimal transition matrix $S_{pp'}$, we minimize the objective function:

$$J = \frac{1}{2}\sum_{v=1}^{V} N_v \sum_{p=1}^{P_{\text{new}}} (y_p^{(v)} - \sum_{p'=1}^{P_{\text{old}}} S_{pp'} x_{p'}^{(v)})^2 - \sum_{p'=1}^{P_{\text{old}}} \varepsilon_{p'} (\sum_{p=1}^{P_{\text{new}}} S_{pp'} - 1), \quad (5)$$

where the presence of N_v (the number of votes cast in district v in 2004) ensures that large voting districts are weighted more heavily than small districts. By demanding that J is minimal with respect to variations in $S_{pp'}$, we deduce a set of equations for $S_{pp'}$. Notice that we have added a Lagrange multiplier term on the right to impose condition (4) on the transition matrix. After optimization, we obtain:

$$\sum_{p''=1}^{P_{old}} S_{pp''} \sum_{v=1}^{V} N_v x_{p''}^{(v)} x_{p'}^{(v)} = \sum_{v=1}^{V} N_v y_p^{(v)} x_{p'}^{(v)} + \varepsilon_{p'} , \quad (6)$$

which must be valid for all values $p = 1,...P_{\text{new}}$ and $p' = 1,...P_{\text{old}}$. Introducing the matrices:

$$(\underline{\underline{A}})_{pp'} \equiv A_{pp'} = \sum_{v=1}^{V} N_v x_p^{(v)} x_{p'}^{(v)} , \quad (7)$$

$$(\underline{\underline{X}})_{pp'} \equiv X_{pp'} = \sum_{v=1}^{V} N_v y_p^{(v)} x_{p'}^{(v)} , \quad (8)$$

and:

$$(\underline{\underline{\varepsilon}})_{pp'} = \varepsilon_{p'} , \quad (9)$$

we can write Equation (6) as follows:

$$\underline{\underline{S}}\,\underline{\underline{A}} = \underline{\underline{X}} + \underline{\underline{\varepsilon}} . \quad (10)$$

The solution of this equation is:

$$\underline{\underline{S}} = \underline{\underline{X}}\,\underline{\underline{A}}^{-1} + \underline{\underline{\varepsilon}}\,\underline{\underline{A}}^{-1} , \quad (11)$$

where we assumed that the inverse $\underline{\underline{A}}^{-1}$ exists, which is the case if enough independent voting results are represented in the set V. We can show that the Lagrangian multipliers can be set to zero, as

$$\underline{\underline{S}} = \underline{\underline{X}}\,\underline{\underline{A}}^{-1} \quad (12)$$

already satisfies Equation (4).

This average transition matrix has an interesting property. If we define the overall results of the old and new elections by:

$$x_p = \sum_{v=1}^{V} N_v x_p^{(v)} \Big/ \sum_{v=1}^{V} N_v , \quad p = 1, \cdots P_{\text{old}} , \quad (13)$$

and

$$y_p = \sum_{v=1}^{V} N_v y_p^{(v)} \Big/ \sum_{v=1}^{V} N_v , \quad p = 1, \cdots P_{\text{new}} , \quad (14)$$

then we can show that:

$$y_p = \sum_{p'=1}^{P_{\text{old}}} S_{pp'} x_{p'}, \qquad p = 1, \cdots P_{\text{new}} . \quad (15)$$

Hence, the transition matrix gives an exact mapping of the old overall results onto the corresponding new ones. Clearly, this is not generally true for individual voting districts.

As it stands, the transition matrix, Equation (12), is not guaranteed to be non-negative. Indeed, the application to the South African elections (see following section) reveals that about 35% of the matrix elements are negative. To interpret the matrix elements $S_{pp'}$ as the fraction of the voters switching from one party p' in the old election to party p in the new poll, they would have to be non-negative. We can expect similar problems in other settings, such as the marketing problem (see Application in other contexts below). In principle there is nothing wrong with the transition matrix having negative elements. It simply expresses the correlations between the distribution of the old and new parties. One can interpret a negative element as expressing the fact that the presence of a particular (old) party reduces the support for a new party (with the negative entry) amongst all old parties. For example, we could imagine that the members of party p discourage the voting for party p' in their electoral district, so that even the support from other parties for p' in that district is less than normal. The influence of p therefore goes beyond its own voting potential. It is difficult to distinguish between these direct and indirect influences of voters on the voting pattern, and hence there seems to be no unambiguous way to recover the direct transition matrix elements (which are naturally non-negative) from the full matrix element.

If one wishes to evaluate the actual redistribution of the votes from the old to the new parties—and this is what most political analysts desire—then one has to construct a transition matrix without negative elements. It is natural to do this by defining the same minimization problem, Equation (5), with the additional condition that the matrix elements should be non-negative. Mathematically, this is a well-defined problem, which has a unique solution, as McCarthy and Ryan[5] first noticed. However, as noted before, one should be aware that the minimization tends to emphasize correlations, and these do not necessarily correspond accurately with the redistribution of votes because of the indirect influence of voters. For this reason, I will not only consider exact methods to solve this minimization problem, but will also examine alternative methods that display certain other desirable properties. McCarthy and Ryan[5] applied quadratic programming techniques to impose the non-negativity conditions. Instead, the Kuhn and Tucker approach is used here.[12,13] Since both methods claim to be unique, the results should be identical, as they solve the same minimization problem. A version of the Kuhn–Tucker method[14] suitable for the current context is discussed in ref. 4. In that document two other methods are introduced, a heuristic one, which preserves the validity of Equation (15), and a simple renormalization method, which does not necessarily preserve (15). All of the methods ensure the validity of the basic property, Equation (4). I now give without derivation the final form of the Kuhn–Tucker procedure, together with the explicit form of the two other ones.

The Kuhn–Tucker approach leads to the following equation for the modified transition matrix $\hat{\underline{\underline{S}}}$:

$$\hat{\underline{\underline{S}}} = \underline{\underline{X}}\,\underline{\underline{A}}^{-1} + \underline{\underline{\psi}}\,\underline{\underline{A}}^{-1} + \underline{\underline{\varepsilon}}\,\underline{\underline{A}}^{-1} + \underline{\underline{\eta}}\,\underline{\underline{A}}^{-1}, \quad (16)$$

where two additional matrices, $\underline{\underline{\psi}}$ and $\underline{\underline{\eta}}$, have been introduced. The latter matrix represents a new set of Lagrangian multipliers, which has the form:

$$(\underline{\underline{\eta}})_{pp'} = \eta_p\, x_{p'}, \quad p = 1 \cdots P_{\text{new}} \quad p' = 1 \cdots P_{\text{old}} , \quad (17)$$

and is introduced to ensure the validity of Equation (15) after the correction. The matrix $\underline{\underline{\psi}}$ is introduced to eliminate negative matrix elements. First, one defines the set:

$$\Lambda = \left\{ (\,p, p'\,) \,\middle|\, S_{pp'} < 0 \right\} . \quad (18)$$

For $(p,p') \notin \Lambda$ one sets $\psi_{pp'}$ equal to zero. For $(p,p') \in \Lambda$ one fixes the $\psi_{pp'}$ such that $\hat{S}_{pp'} = 0$. One then obtains the following solutions:[1]

$$\varepsilon_p = -\frac{1}{P_{\text{new}}}\sum_q \psi_{qp} \qquad p = 1\cdots P_{\text{old}} \,, \tag{19}$$

and

$$\eta_p = \Big(\frac{1}{P_{\text{new}}}\sum_{q,q',q''}\psi_{qq'}A^{-1}_{q'q''}x_{q''} - \sum_{q,q'}\psi_{pq}A^{-1}_{qq'}\,x_{q'}\Big) / \sum_{q,q'} x_q\,A^{-1}_{qq'}\,x_{q'} \,, \tag{20}$$

$$p = 1\cdots P_{\text{new}} \,,$$

where the non-zero matrix elements, $\psi_{pp'}$ are determined by:

$$-S_{pp'} = \sum_q \psi_{pq}A^{-1}_{qp'} - \frac{1}{P_{\text{new}}}\sum_{q,q'}\psi_{qq'}A^{-1}_{q'p'} + \Big[\frac{1}{P_{\text{new}}}\sum_{q,q',q''}\psi_{qq'}A^{-1}_{q'q''}x_{q''} - \sum_{q,q'}\psi_{pq'}A^{-1}_{q'q}\,x_q\Big] \times \sum_q x_q A^{-1}_{qp'} \Big/ \Big[\sum_{q,q'} x_{q'}A^{-1}_{q'q}\,x_q\Big] \tag{21}$$

$$\text{for } \forall (p,p') \in \Lambda \,.$$

If after solving Equation (21), $\underline{\underline{\hat{S}}}$ has new negative elements, then one has to extend Λ with these additional pairs (p, p'), and recalculate the modified matrix.

Although this approach is unique,[12] it is not necessarily true that the given formulation of the optimization problem gives the best characterization of the redistribution of votes, as noted above. In particular, the results of McCarthy and Ryan[5] have been criticized by Upton[6] on the basis of several voter surveys. The transition matrix also reflects the influence of voters of one party on supporters of other parties, and therefore its interpretation as a redistribution model, after imposing the non-negativity conditions, is not completely satisfactory. This is reflected, for example, in the presence of many zero matrix elements in the Kuhn–Tucker transition matrix (see next section). In reality, it is extremely unlikely that so many elements are exactly zero, so this illustrates a defect of the formulation of the optimization problem with positivity conditions [this problem is probably widespread in any linear programming problem, although in some cases the high frequency of zero (or boundary) elements may be acceptable]. Because of the limitations of this 'exact' approach, it is of interest to consider other methods for converting the transition matrix into a non-negative matrix, especially if such methods do not display the same high frequency of zero elements. Below, I discuss two alternative methods, each of which is much simpler to apply than the Kuhn–Tucker or quadratic programming approach.

In the first alternative approach, one tries to remove the influence part of the transition matrix, which is responsible for the negative elements, so as to obtain non-negative redistribution elements. Let us assume that the matrix element $S_{pp'}$ is negative. This is interpreted to mean that the voters belonging to the old party p' encourage voters from other (old) parties $p'' \neq p'$ not to vote for the new party p. If the transition matrix is to represent a redistribution of votes, then this would lead to a reduction in the matrix elements $S_{pp''}$ for $p'' \neq p'$. However, since a large number of voting districts are represented with the same S-matrix, in the overall fit it is more effective to represent this effect with a negative matrix element $S_{pp'}$, because in those districts where p' is not well represented the effect is not present. I now want to reverse this effect and restore the meaning of the transition matrix as a redistribution matrix. Hence, I make $S_{pp'}$ zero and reduce the matrix elements $S_{pp''}$ for $p'' \neq p'$. Naturally, this will worsen the overall fit; however, it will allow us to interpret the modified matrix as the real redistribution of votes (if we want to use the transition matrix for predicting the results in a particular voting district then we should keep the original one). The first step then is to set the modified $\hat{S}_{pp} = 0$. The reduction of $\hat{S}_{pp''}$ for $p'' \neq p'$ can be formally imposed by demanding condition Equation (15) for the modified transition matrix. Subsequently, the matrix elements $\hat{S}_{\alpha p''}$ for $\alpha \neq p$ also have to be reduced to maintain the validity of Equation (4). Finally, the remaining matrix elements also have to be adjusted. This leads to the following algorithm ($S_{pp'} < 0$)

- $\hat{S}_{pp'} = 0$
- $\hat{S}_{p\beta} = \gamma\, S_{p\beta}; \quad \gamma = \dfrac{y_p}{y_p - S_{pp'}x_{p'}}, \qquad \beta \neq p'$ (22)
- $\hat{S}_{\alpha p'} = \gamma'\, S_{\alpha p'}; \quad \gamma' = \dfrac{1}{1 - S_{pp'}}, \qquad \alpha \neq p$
- $\hat{S}_{\alpha\beta} = S_{\alpha\beta} + (1-\gamma)\,\gamma'\, S_{\alpha p'} S_{p\beta}, \qquad \alpha \neq p \quad \beta \neq p'.$

We can easily verify that these steps preserve the identities, Equations (4) and (15). We can apply Equation (22) sequentially (treating the most negative entries first) until all negative elements have been removed. During this process, elements, which at some stage were set to zero, will turn into positive terms, making them more realistic. Other elements, which started out negative, will become positive even before their turn to be treated has arrived. The opposite effect (that positive elements become negative) is extremely rare. In contrast to the Kuhn–Tucker approach, this method does not refer back to the original objective function and its minimization once the correction process has started. Hence, it will no longer provide an optimal solution to the original mathematical problem, and therefore will be called a heuristic approach. However, it does maintain the link to the original optimal transition matrix (which the quadratic programming approach does not) and implements the necessary corrections to this matrix so that it can be reinterpreted as a redistribution matrix. So it is a valid alternative to the exact approach, with the added advantage that it is much simpler to apply than the Kuhn–Tucker or quadratic programming approaches.

The second alternative approach is a simple renormalization of the original transition matrix that does not enforce the property, Equation (15):

$$\hat{S}_{pp'} = \frac{S_{pp'} + |S_{pp'}|}{1 + \sum_{p''}|S_{p''p'}|} \,. \tag{23}$$

This new transition matrix also satisfies Equation (4) and is clearly non-negative. It turns all negative elements into zero elements, and therefore will also lead to many zero entries. However, in contrast to the Kuhn–Tucker approach, it will not lead to additional zero elements, and therefore it reduces the frequency of zero elements somewhat compared to the Kuhn–Tucker approach, which may be considered an advantage. Its big advantage is its extreme simplicity, although this comes at the expense of Equation (15), which is no longer satisfied.

To decide which procedure gives the most accurate description of the redistribution of votes, one can compare the results with a more direct determination of the transition matrix via voter surveys. Although such surveys have not been conducted in South Africa, they have been performed in the U.K.,[6] and have been compared to the quadratic programming approach by McCarthy and Ryan.[5] In the following section, possible consequences of this study for the adequacy of our three methods are considered.

Table 1. Basic transition 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

*ANC, African National Congress; DA, Democratic Alliance; DP, Democratic Party; IFP, Inkatha Freedom Party; UDM, United Democratic Movement; ID, Independent Democrats; NNP, New National Party; ACDP, African Christian Democratic Party; VFP, Freedom Front Plus.

Table 2. Renormalized transition 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	10.9	3.1	42.1	41.4	43.4
DA	12.4	0.3	88.8	1.6	36.6	0	0	5.4
IFP	7.0	0	0.9	85.4	0.7	0	0	3.3
UDM	2.3	0.7	0	0.1	0	56.3	13.6	2.0
ID	1.7	0.1	5.4	0	18.5	0.6	14.2	0
NNP	1.7	0.3	0	0.0	29.9	0	0	0
ACDP	1.6	0.4	3.8	0.7	5.6	0	26.3	1.3
VFP	0.9	0.0	0.2	0.1	1.8	0	0	40.9

*See Table 1 for abbreviations.

Calculation of quantitative transition matrices to describe trends between the 1999 and 2004 South African elections

I have applied the formulation by comparing the South African elections in 1999 and 2004. In Table 1 I display the transition matrix corresponding to Equation (12). I have limited the display to 7 'old' parties (the actual number in 1999 was 16) and 8 'new' parties (the actual number in 2004 was 21). The DA (Democratic Alliance) is in many respects a continuation of the DP (Democratic Party). The same is true of the VFP (Freedom Front Plus) and the VF (Freedom Front). Hence, one expects the corresponding diagonal matrix elements to be large.

The top row in the table contains the overall results for the 1999 elections, while the column on the left shows the 2004 results. The transition matrix is contained in the box with 56 elements (the original total was 336). Since the matrix is expressed in terms of percentages, the columns add up to 100 instead of 1 (the sums are not exact as the smaller parties are not featured). The zeroes in this matrix are a consequence of the limitation to one decimal place, and they are not exact. Let us explain these numbers for a particular case. The entry with column index DP and row index ID is 5.8. The trend interpretation is that 5.8% of the people who voted for the DP in 1999 voted for the ID in 2004. Since the original percentage of the DP was 9.4%, the absolute contribution to the ID is 0.6%. The total percentage obtained by the ID in 2004 is 1.7%, so that about one third of this total comes from the DP.

To justify this interpretation as trend matrix, the negative elements must be removed. As much as 18 of the 56 elements are negative. In relative terms this is not very different from the total matrix, which has 336 elements out of which 118 are negative.[4] Three methods to remove negative elements were introduced above. The results for these three methods are listed in Tables 2–4.

A comparison of Tables 2–4 shows that the three methods lead to considerable differences in the corrected transition matrix. We first notice that the Kuhn–Tucker matrix does indeed contain many zero elements. While the original transition matrix contained 19 zeroes (none of them exactly zero), this 8 by 7 Kuhn–Tucker matrix contains many (as much as 28) elements that are (exactly) zero. As can be expected, the calculations of Mc Carthy and Ryan feature the same phenomenon, since they are obtained with an equivalent method. This unrealistic result is an artefact of the mathematical method, also common in linear programming problems. The renormalization results also

Table 3. Heuristic transition 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

*See Table 1 for abbreviations.

Table 4. Kuhn–Tucker transition 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	11.5	4.6	42.9	48.9	23.5
DA	12.4	0.2	94.0	1.4	38.2	0	0	17.3
IFP	7.0	0	0	84.7	0	0	0	0
UDM	2.3	0.3	0	0	0	56.3	0	0
ID	1.7	0	1.2	0	18.5	0.9	19.4	0
NNP	1.7	0	0	0	24.5	0	0	0
ACDP	1.6	0.4	3.6	0.7	6.7	0	31.7	0
VFP	0.9	0	1.2	0.4	3.5	0	0	59.2

*See Table 1 for abbreviations.

Table 5. Average transition matrix with standard deviations in small font.

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	0.4	0.1	0.1	12.2	1.7	4.8	1.8	43.0	1.0	44.4	4.0	36.9	11.6
DA	12.4	0.5	0.3	90.2	3.3	1.5	0.2	38.3	1.7	2.0	3.4	0.7	1.1	9.5	6.8
IFP	7.0	0.0	0.0	0.7	0.6	83.9	1.9	0.6	0.5	0.0	0.0	0.1	0.1	2.2	1.9
UDM	2.3	0.5	0.2	0.0	0.0	0.0	0.0	0.0	0.0	52.8	6.0	8.3	7.3	1.2	1.0
ID	1.7	0.0	0.1	3.7	2.2	0.0	0.0	17.4	1.9	0.9	0.3	15.1	4.0	0.0	0.0
NNP	1.7	0.2	0.1	0.0	0.0	0.0	0.0	25.5	4.0	0.1	0.2	0.1	0.1	0.0	0.0
ACDP	1.6	0.4	0.0	3.8	0.2	0.8	0.2	6.2	0.6	0.1	0.3	28.3	3.0	0.9	0.8
VFP	0.9	0.0	0.1	0.7	0.5	0.2	0.2	3.2	1.3	0.0	0.1	0.2	0.4	46.8	10.8

*See Table 1 for abbreviations.

contain a number of exact zeroes; however, their total is considerably less (17). On the other hand, the heuristic result contains only one exact zero (not present in Table 3 as it is one of the elements not displayed). Another distinction between the exact Kuhn–Tucker method and the other methods is the magnitude of the diagonal elements (elements relating the same party). Note that in virtually all cases the Kuhn–Tucker method gives larger diagonal elements than the other methods.

Is there an objective way to judge these results and decide which one is the most realistic? Upton[6] has checked the results of McCarthy and Ryan using voter surveys (called panel results). This type of survey had been criticized earlier by Miller,[3] who claimed that the results of such surveys tended to be biased. In particular, abstentions are under-reported, while the voters also change their memory of previous votes to favour the newly supported party, so that up to 50% of respondents do not remember that they changed allegiances.[15] Despite this criticism, these surveys provide a valuable independent means of checking the mathematical models, especially when the conclusions show a large degree of consistency over a wide number of cases. Upton finds that in all cases the estimated results (that is, the results from McCarthy and Ryan) feature diagonal elements which are much higher than the observed ones (that is, the results from the surveys). The ratios of the estimated and observed diagonal elements are 1.05 and 1.09 for the British Conservative and Labour parties, respectively, while for the smaller Liberal party the ratio grows to 1.36. In our case the ratios between the Kuhn–Tucker result and the heuristic method are 1.01 for the ANC, 1.07 for the DA/DP, 1.04 for the IFP and 1.23 for the UDM. For a small party like the VF/VFP, the ratio grows to 1.47. These numbers are similar to the ones found by Upton, hence, these results suggest that the heuristic results may be a better representation of the voter movements than the exact Kuhn–Tucker method. The situation with the small/zero elements leads to a similar conclusion. In Upton's study,[6] zero elements do not occur at all, with the smallest element in his table 2 being 2%. In the heuristic approach there are also very few small elements (see Table 3), while in the Kuhn–Tucker case there are as many as 28 elements identical to zero. Hence, Upton's findings suggest that the heuristic results are closer to the truth than the Kuhn–Tucker results. I have not commented so far on the renormalization approach, since this method is much simpler and lacks a clear justification. Its results are somewhat intermediate between those of the Kuhn–Tucker and the heuristic approach and therefore are not a bad alternative if one wants quick results.

Given the different results of the three methods, and the arguments in favour of the heuristic approach, one might wonder how best to present the results to the public and to political analysts. As argued above, I prefer the heuristic approach because it seems to lead to more realistic results. However, one still would like an idea of the errors involved. By comparing all three methods, one can derive a standard deviation and use it as measure of its uncertainty. This is not a rigorous statistical procedure; however, considering the criticism of the exact approach, a formal determination of the confidence intervals[2] of the original transition matrix has even less practical value, as the main uncertainty is probably more due to model uncertainty than to statistical error. One could also use the three results to derive an average transition matrix, as an alternative to the heuristic result. Table 5 displays this average matrix with its standard deviation.

Consider now the 'uncertainty' in the resulting matrix. The smallest (relative) uncertainty is in the diagonal elements. According to this matrix, as much as 97% of the original 1999 ANC voters, again supported the ANC in 2004. The uncertainty is only 0.4%, both in absolute and relative terms. For the DP (DA) the diagonal element is 90.2%, with an uncertainty of 3.3%. For the smaller parties the uncertainty can be bigger. For example, for the UDM only 52.8% voted again for the UDM in 2004, but the uncertainty is 6%, which is 11% in relative terms. An unexpected (at least by me) result is that nearly 37% of 1999 Freedom Front (VF) voters, voted for the ANC in 2004, although the uncertainty is quite large (11.6%). It would be surprising if any political analysts had expected this. This result shows that the current approach might well lead to new political insights. It is also of interest that the appeal of the ANC does not extend to the NNP, despite the fall in the latter's popularity in the 2004 elections and the merger it made with the ANC. Interesting as these political questions may be, further discussion of them lies outside the scope of this paper.

Application in other contexts

Transition matrices have also been used in other contexts. Pelzer *et al.*[1] give a survey of applications to socio-economic problems. Most of these applications use Markov procedures, which also have been used extensively in marketing.[17] Another method that has been used in marketing is the maximum entropy method,[10,16] especially in the study of brand switching. As long as the number of brands is not too large, this method is useful for the construction of transition matrices. Entropy calculations do not display the problem of zero elements; in fact, their absence in this method can be used as a further argument against their presence in the Kuhn–Tucker approach. The entropy approach can possibly also be used in the election problem, and then would be of value if few data are available. Whether it would give a reliable description is, however, not clear. Bass[16] has criticized some of the limitations of the entropy approach. The applicability of the methods discussed in this paper to marketing problems depends strongly on the availability of detailed data. As mentioned before, detailed sales information could be converted into transition matrices providing insight into shopping habits such as brand switching and churning. Both the original transition matrix with its direct and indirect components and the modified matrix without negative elements could be useful. In the latter case the new heuristic method introduced here could be beneficial.

If few data are available, one might have to resort to simpler models with additional assumptions. Entropy models have already been mentioned. Other approaches have also been discussed in the literature. For example, Upton described a memory model[18] for voting transitions, while he and Sarvlik[19] have discussed a loyalty–distance model for voting change. If one can justify the presence of certain regularities in the transition matrix, then models expressing these mechanisms may well give a valid alternative to the advanced mathematical methods. As we have seen in this paper, heuristic methods with the right justification can be superior to advanced mathematical methods that solve a given problem exactly, but fail to represent reality perfectly. Further research on transition matrices, which relies less on extensive databases, is currently under way.

Conclusions

I have shown how transition matrices that relate old and new election data can lead to interesting new insights. Although it is possible to interpret such a matrix when it has negative elements, it is usually preferable to ensure non-negativity of the matrix elements. This paper presents three ways to accomplish this. To analyse the adequacy of each method, I refer to a voter study conducted in the U.K. in 1978. The Kuhn–Tucker approach, which is equivalent to the quadratic programming method used in earlier work, yields an exact solution of the minimization problem. However, it has some undesirable features, such as a high frequency of zero elements and an overestimation of the diagonal elements. The so-called heuristic method provides a more intuitive way to calculate the non-negative transition matrix. It is not plagued as strongly by the aforementioned problems, and is much easier to use than the Kuhn–Tucker approach. The trend predictions in real-time have recently been used in election-night forecasting for the 2006 South African municipal elections. Since such an application can use only partial results (those for the subset of voting districts available in real-time), there are some extra challenges and uncertainties. However, it is another example of the potential benefits of transition matrices, namely as a prediction tool.

I thank the South African Broadcasting Corporation for supporting the 2004 election prediction work by the Logistics and Quantitative Methods unit in the CSIR, and by giving us access to the data underlying the current analysis. I also thank a referee for useful comments and for drawing my attention to the paper by McCarthy and Ryan.[5]

1. Pelzer B., Eisinga R. and Franses P.H. (2001). Estimating transition probabilities from a time series of independent cross sections. *Statistica Neerlandica* **55**, 249–262.
2. Hawkes A. (1969). An approach to the analysis of electoral swing. *J. R. Statist. Soc. A* **132**, 68–79.
3. Miller W.L. (1972). Measures of electoral change using aggregate data. *J. R. Statist. Soc. A* **135**, 122–142.
4. Greben J.M. (2005). Several methods of trend analysis applied to the South African elections. CSIR Tech. Rep. CSIR/BE/LQM/IR/2007/0100B, Pretoria.
5. McCarthy C. and Ryan T.M. (1977). Estimates of voter transition probabilities from the British general elections of 1974. *J. R. Statist. Soc.* A **140**, 78–85.
6. Upton G.J.G (1978). A note on the estimation of voter transition probabilities. *J. R. Statist. Soc. A* **141**, 507–512.
7. Lemon A. (2001). The general election in South Africa, June 1999. *Electoral Studies* **20**, 305–339.
8. Greben J.M., Elphinstone C.E., Holloway J., de Villiers R., Ittmann H. and Schmitz P. (2005) Prediction of the 2004 national elections in South Africa. *S. Afr. J. Sci.* **101**, 157–161.
9. Lahiri A.K. and Roy P. (1984). Assessing swings in multi-party systems: the Indian experience. *Electoral Studies* **3**, 171–189.
10. Herniter J.D. (1973). An entropy model of brand purchase behaviour. *J. Marketing Res.* **10**, 361–375.
11. Greben J.M., Elphinstone C.E. and Holloway J. (2006). A model for election night forecasting applied to the 2004 South African elections. *J. Operations Res. Soc. S. Afr. (ORiON)* **22**, 89–103.
12. Kuhn H.W. and Tucker A.W. (1951). Non-linear programming. *Proc. 2nd Berkeley Symposium on Mathematical Statistics and Probability,* ed. J. Neyman. University of California Press, Berkeley.
13. Bernholtz B. (1963). A new derivation of the Kuhn–Tucker conditions. *Opns Res.* **12**, 295–299.
14. Kaufman L. and Rousseeuw P.J. (1990). Finding groups in data. In *An Introduction to Cluster Analysis,* pp. 182–185. Wiley-Interscience, New York.
15. Benewick R.J., Birch, A.H., Blumler J.G. and Ewbank A.(1969). The floating voter and the liberal view of representation. *Political Studies* **17**, 177–195.
16. Bass F.M. (1974). The theory of stochastic preference and brand switching. *J. Marketing Res.* **11**, 1–20.
17. Hauser J.R. and Wisniewski K.J. (1982). Dynamic analysis of consumer response to marketing strategies. *Management Sci.* **28**, 455–486.
18. Upton G.J.G (1978). A memory model for voting transitions in British elections. *J. R. Statist. Soc. A* **140**, 86–94.
19. Upton G.J.G and Sarlvik B. (1981). A loyalty–distance model for voting change. *J. R. Statist. Soc. A* **144**, 247–259.