Compositional Loess modeling

Bergman, Jakob; Holmquist, Björn

Published in:
Proceedings of the 4th International Workshop on Compositional Data Analysis

2011

Proceedings of the 4th International Workshop
on Compositional Data Analysis (2011)

Compositional Loess modeling

J. BERGMAN[1] and B. HOLMQUIST[1]

[1]Department of Statistics - Lund University, Sweden. email: jakob.bergman@stat.lu.se

Abstract

Cleveland (1979) is usually credited with the introduction of the locally weighted regression, Loess. The concept was further developed by Cleveland and Devlin (1988). The general idea is that for an arbitrary number of explanatory data points x_i the value of a dependent variable is estimated $\hat{y}_i$. The $\hat{y}_i$ is the fitted value from a dth degree polynomial in x_i. (In practice often $d = 1$.) The $\hat{y}_i$ is fitted using weighted least squares, WLS, where the points x_k $(k = 1, \dots, n)$ closest to x_i are given the largest weights.

We define a weighted least squares estimation for compositional data, $\mathcal{C}$-WLS. In WLS the sum of the weighted squared Euclidean distances between the observed and the estimated values is minimized. In $\mathcal{C}$-WLS we minimize the weighted sum of the squared simplicial distances (Aitchison, 1986, p. 193) between the observed compositions and their estimates.

We then define a compositional locally weighted regression, $\mathcal{C}$-Loess. Here a composition is assumed to be explained by a real valued (multivariate) variable. For an arbitrary number of data points x_i we for each x_i fit a dth degree polynomial in x_i yielding an estimate $\hat{y}_i$ of the composition y_i. We use $\mathcal{C}$-WLS to fit the polynomial giving the largest weights to the points x_k $(k = 1, \dots, n)$ closest to x_i.

Finally the $\mathcal{C}$-Loess is applied to Swedish opinion poll data to create a poll-of-polls time series. The results are compared to previous results not acknowledging the compositional structure of the data.

1 Introduction

There exist different approaches of extracting information of time series. Smoothing of the series by weighting together 'neighboring' observations – in the belief that these consistently reflect a common, slowly varying property – comes in many different forms.

It may be a non-parametric approach in which weights are chosen according to some kernel function suitably chosen to reflect the behavior in the underlying property, or as in wavelets by letting them reflect a certain frequency resolution to be traced.

It can also be a parametric, or a locally parametric approach where the structure of the property is described as a parametric function. Cleveland (1979) is usually credited with the introduction of the locally weighted regression, Loess. The concept was further developed by Cleveland and Devlin (1988). The general idea is that for an arbitrary number of explanatory data points x_i the value of a dependent variable is estimated $\hat{y}_i$. The $\hat{y}_i$ is the fitted value from a dth degree polynomial in x_i. (In practice often $d = 1$.) The $\hat{y}_i$ is fitted using weighted least squares, WLS, where the points x_k $(k = 1, \dots, n)$ closest to x_i are given the largest weights.

Figure 1 shows the results for a large number of polls of political party preferences to nine groups of political parties during approximately four recent years. It is evident that the preferences are changing over time in a non-linear fashion, and also that there are large random deviations in the polls for the different parties. Smoothed versions of the series would facilitate a way of revealing changes in political trends.

The difficulties occurring when applying standard smoothing techniques to multivariate data of compositional type, is that the techniques do not account for the special structure inherent (among the components) in the data.

We shall consider a technique similar to Loess for smoothing multivariate data of compositional type, which take into account the special nature of the multivariate data being compositions.

Egozcue, J.J., Tolosana-Delgado, R. and Ortego, M.I. (eds.)
ISBN: 978-84-87867-76-7

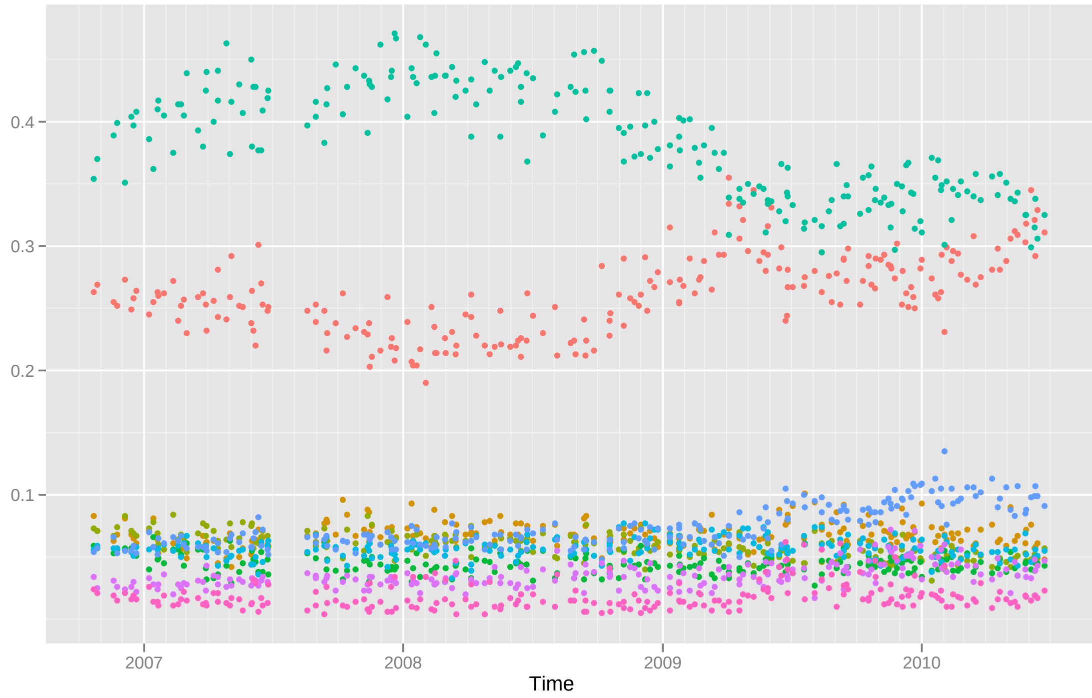

Figure 1: Polls of political party preferences to nine groups of political parties (different colors) during approximately four recent years

2 The method

Let $(x_1, \mathbf{y}_1), \ldots, (x_n, \mathbf{y}_n)$ be n pairs of observations of which the $\mathbf{y}_i$'s are compositions in the simplex $\mathbb{S}^D$ and x is an explanatory variable. This explanatory variable may be univariate or multidimensional, with or without manifold restrictions. It may thus for example be another compositional variable. We shall exemplify the procedure by letting the explanatory variable be a univariate real variable representing 'time' and therefore let $(t_1, \mathbf{y}_1), \ldots, (t_n, \mathbf{y}_n)$ be the n pairs of observations of a compositional time series at time points $t_1, \ldots, t_n$.

Now let $t'_1, \ldots, t'_m$ be a second set of time points. These may constitute a larger set of time points than $t_1, \ldots, t_n$ within the convex hull of the latter and may include those latter time points as a subset (interpolation), or it could simply be equal to, or a subset of the set of the time points $t_1, \ldots, t_n$ (smoothing). It may even be a complementary set of time points for prediction purposes (extrapolation).

We use the notation $\mathbf{a} \oplus \mathbf{b}$ for the perturbation $(a_1 b_1, \ldots, a_D b_D)/\sum_{i=1}^{D} a_i b_i$ of $\mathbf{b}$ relative to $\mathbf{a}$ (or $\mathbf{a}$ relative to $\mathbf{b}$) where a_i (b_i) are the components of $\mathbf{a}$ ($\mathbf{b}$). Further $c \odot \mathbf{b}$ defines the power transformation $(b_1^c, \ldots, b_D^c)/\sum_{i=1}^{D} b_i^c$ for $c \in \mathbb{R}$.

2.1 Weighted least squares for compositions, $\mathcal{C}$-WLS

Suppose that the data is generated by

$$\mathbf{y}_i = g(t_i) \oplus \varepsilon_i$$

where g is supposed to be a smooth function of the t variable. The idea to estimate g, is to locally fit a first degree polynomial in the simplex space, by letting the closest points influence most.

At each time point t'_k we find compositions $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ that minimize

$$Q_{\mathcal{C}}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1) = \sum_{i=1}^{n} w_i(t'_k) d_S^2\{\mathbf{y}_i; \boldsymbol{\beta}_0 \oplus (t_i \odot \boldsymbol{\beta}_1)\}$$

where

$$w_i(t'_k) = W(\frac{d_E(t'_k, t_i)}{d(t'_k)})$$

for the 'tricube' weight function

$$W(u) = \begin{cases} (1-u^3)^3, & \text{if } 0 \le u < 1 \\ 0, & \text{otherwise} \end{cases}$$

and $d(t'_k) = d_E(t'_k, \text{the } q \text{ closest } t_i)$ for some given integer q, $1 \le q \le n$. Here closeness is measured in a metric of the space of t, and thus $d_E(t', t)$ is the Euclidean distance between t' and t.

The (squared) distance measure on the simplex considered here will be the (squared) Aitchison distance

$$d_S^2(\mathbf{y}; \mathbf{z}) = \frac{1}{2D} \sum_{i=1}^{D} \sum_{j=1}^{D} (\ln(\frac{y_i}{y_j}) - \ln(\frac{z_i}{z_j}))^2$$

for compositions $\mathbf{y}$ and $\mathbf{z}$, proposed by Aitchison (1992). The relation is

$$d_S^2(\mathbf{y}; \mathbf{z}) = D d_E^2(\operatorname{clr}(\mathbf{y}), \operatorname{clr}(\mathbf{z}))$$

to the Euclidean distance of centered log-ratio transformed compositions (see Aitchison (1986) for more details). The technique proposed here is however not restricted to this measure and it is possible to use other measures e.g. a measure based on divergence as suggested by Martín-Fernández et al. (1999).

If $\mathbf{z} = \boldsymbol{\beta}_0 \oplus (t \odot \boldsymbol{\beta}_1)$ then

$$d_S^2(\mathbf{y}; \mathbf{z}) = \frac{1}{2D} \sum_{i=1}^{D} \sum_{j=1}^{D} (\ln(\frac{y_i}{y_j}) - \ln(\frac{\beta_{0,i}}{\beta_{0,j}}) - t \ln(\frac{\beta_{1,i}}{\beta_{1,j}}))^2$$

where $\beta_{0,i}$ and $\beta_{1,i}$ are the components of $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$, respectively.

Here $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ are compositions in $\mathbb{S}^D$ and hence the minimization of $Q_{\mathcal{C}}$ is subject to the restrictions $\boldsymbol{\beta}_0 \ge \mathbf{0}$, $\boldsymbol{\beta}_1 \ge \mathbf{0}$ and $\boldsymbol{\beta}_0^T \mathbf{1} = \boldsymbol{\beta}_1^T \mathbf{1} = 1$.

The local compositional weighted least squares ($\mathcal{C}$-WLS) estimates corresponding to t'_k are

$$(\hat{\boldsymbol{\beta}}_{0k}, \hat{\boldsymbol{\beta}}_{1k}) = \operatorname*{arg\,min}_{\boldsymbol{\beta}_0, \boldsymbol{\beta}_1} Q_{\mathcal{C}}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1) \,.$$

One way to accomplish this minimization is by adding Lagrange multiplicators of the restrictions to $Q_{\mathcal{C}}$. Other possibilities include using available minimization procedures that permit restrictions on the parameters to be specified.

It is also possible to find explicit expressions for the minimizing $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ when using the distance measure d_S^2 (see Appendix). For the implemented version of the procedure we have also used (as an alternative) a minimization algorithm where positive restrictions on the parameters were specified, and where the last component in the composition were set to one minus the sum of the other parts of the compositions. This gives a flexibility for using other distance measures as mentioned above.

2.2 Loess for compositions, $\mathcal{C}$-Loess

Repeating this procedure for each t'_k in the set $\{t'_1, \ldots, t'_m\}$ we obtain a set of estimated compositions. The locally fitted first degree polynomial composition at t'_k is denoted $\hat{\mathbf{y}}'_k = \hat{\mathbf{y}}(t'_k)$ and is equal to $\hat{\boldsymbol{\beta}}_{0k} \oplus (t'_k \odot \hat{\boldsymbol{\beta}}_{1k})$ for $k = 1, \ldots, m$, where $\hat{\boldsymbol{\beta}}_{0k}$ and $\hat{\boldsymbol{\beta}}_{1k}$ are the locally fitted estimates corresponding to time t'_k. This is then a compositional locally weighted regression, which we name $\mathcal{C}$-Loess.

Let $d_i^2 = d_S^2(\mathbf{y}_i; \hat{\mathbf{y}}_i)$ where $\hat{\mathbf{y}}_i = \hat{\mathbf{y}}(t_i) = \hat{\boldsymbol{\beta}}_{0i} \oplus (t_i \odot \hat{\boldsymbol{\beta}}_{1i})$ is the fitted (estimated) composition at t_i. As a measure of lack-of-fit we may use

$$s_{\text{LOF}}^2 = \sum_{i=1}^{n} d_S^2(\mathbf{y}_i; \hat{\mathbf{y}}_i)/n$$

Figure 2: Smoothed Swedish opinion poll series with $q = 40$ with original observations. The solid black line represents the 4% election threshold to win seats in the Swedish parliament.

ISBN: 978-84-87867-76-7

based on the 'deviations' between the fitted and observed compositions at $t_1, \ldots, t_n$. It may however also sometimes be useful to plot the individual deviations d_i^2 or d_i against t_i to have a summary of the scale over the range of observations. This plot may reveal time points where the smoothed result deviate very much from what is observed. Also a smoothed version of such a scatter plot may give information on the local average deviations in the mean level. Here of course s^2_{LOF} is the 'totally smoothed' average value of the d_i^2's.

The larger q, the smoother result will be obtained from the $\mathcal{C}$-Loess procedure and the larger value of s^2_{LOF}. An 'optimum' smoothing parameter q can be obtained by minimizing a function of s^2_{LOF} penalized by sample size n and q.

The $\mathcal{C}$-Loess estimate $\hat{\mathbf{y}}(t'_k)$ is not a linear combination of the $\mathbf{y}_i$'s as it would be for unrestricted spaces. However it can be shown that it can be written as

$$\bigoplus_{i=1}^{n} (\ell_i(t'_k) \odot \mathbf{y}_i)$$

(i.e. a compositional linear combination) where the $\ell_i(t'_k)$ depend on $t_1, \ldots, t_n, W, d_E$, and q (and t'_k) but not on the $\mathbf{y}_i$'s, (see Appendix).

3 Swedish opinion poll data

The data set we consider consists of $n = 218$ number of polls of political party preferences in Sweden. The polls extend over the time period from October 2006 to May 2010 and were performed by a number of different polling institutes including Statistics Sweden (SCB). The polls included in the data set are all similar in their design. They were all essentially telephone interviews of a number of individuals each of which were given the question: "If it were general election today, what political party would you vote for?" The given definite answers were used to calculate fractions of the different party supporters (party preference compositions). The number of individuals taking part in each poll may vary from poll to poll but in general it is quite stable around 1000-2500 individuals, but there also exist polls in which as many as 7000 individuals were interviewed. There are nine parts in the compositions: the four liberal/conservative parties currently in office (M, FP, C, KD), the three environmentalist/socialist parties (S, V, MP), the nationalistic party (SD), and all other parties (Other).

Figure 2 shows the smoothed series using $q = 40$. For the largest of party in office (M) we see a clear increase around the start of the economic crisis in the second half of 2008, and at the same time a large decrease for the largest opposition party (S). A close-up of the smaller parties is shown in Figure 3. We here see a steady decline for C and an almost doubling in size for MP. Two parties are constantly close to the election threshold; KD is just above 4 % and SD is approaching from 4 % from below.

We also modeled the data using a number of different values of the smoothing parameter q, ranging from 10 to 150. In Figure 4 the smoothed series are shown with $q = 30$, $q = 40$ and $q = 50$. As is seen in the figure, $q = 30$ yields a more volatile estimate than $q = 40$ or $q = 50$ which are more smooth as expected. We choose to use $q = 40$ as it seems to give a good balance between capturing the changes in trend while not being too sensitive to individual polls.

The sequence of residual deviations d_i for the smoothed series in Figures 2 and 3 are shown in Figure 5. There does not seem to be any variation due to time in the deviations. The sample size of an opinion poll is usually varying between polling institutes but seldom between polls performed by the same institute. If it is believed that the sample size is the primary source of the uncertainty in the result, then there should be possible to find some smooth relation between sample size and deviation. If this is the case we could build in this into the weighting of different polls in the distance measure, letting more uncertain points weight less than more accurate polls. However, in Figure 6 we see no clear such relation; the deviations are smaller for larger sample sizes but there seems to be other sources of variability too.

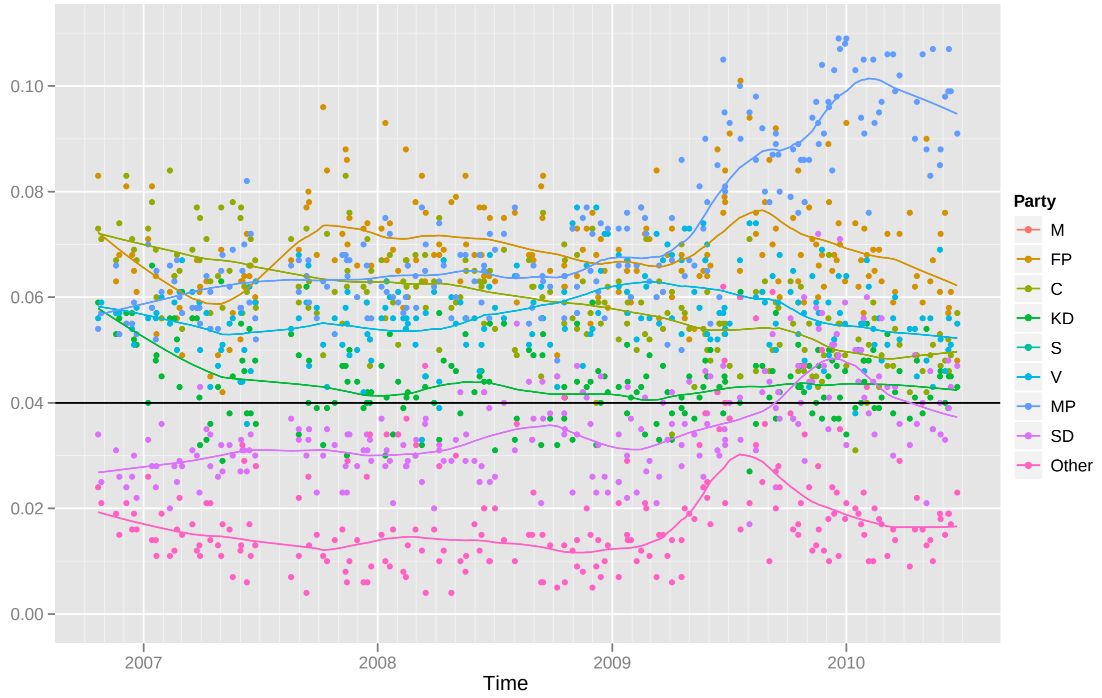

Figure 3: Close-up of the smoothed Swedish opinion poll series with $q = 40$ with original observations. The solid black line represents the 4 % election threshold for winning seats in the Swedish parliament.

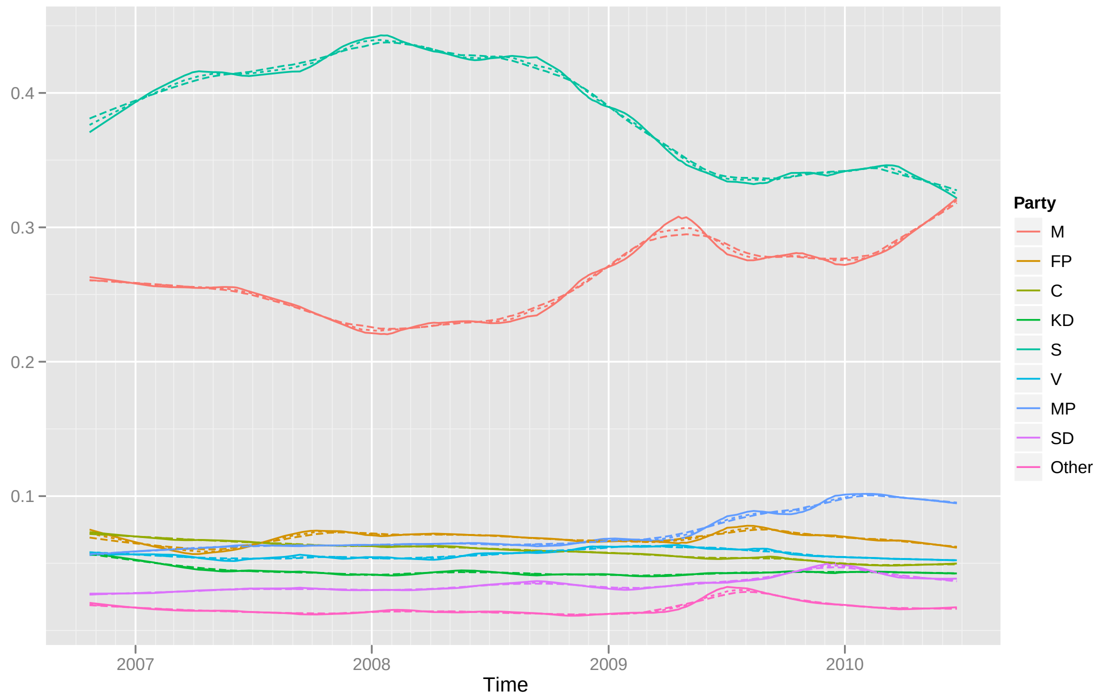

Figure 4: Smoothed series with $q = 30$ (solid line), $q = 40$ (dotted line), and $q = 50$ (dashed line). We note that $q = 30$ yields a much more volatile smooth than $q = 40$ or $q = 50$ which are more similar.

ISBN: 978-84-87867-76-7

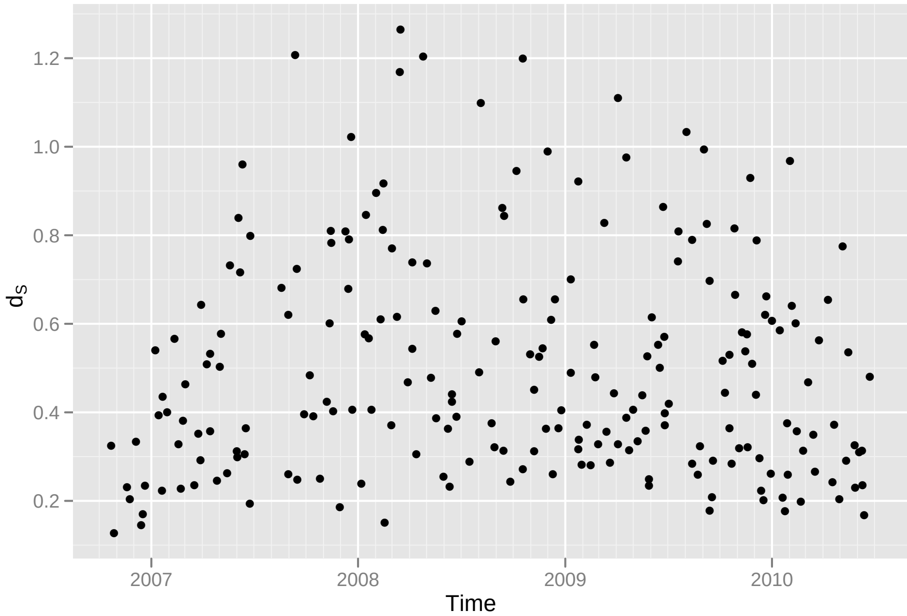

Figure 5: The deviations d_i from the smoothed series for $q = 40$ plotted versus time.

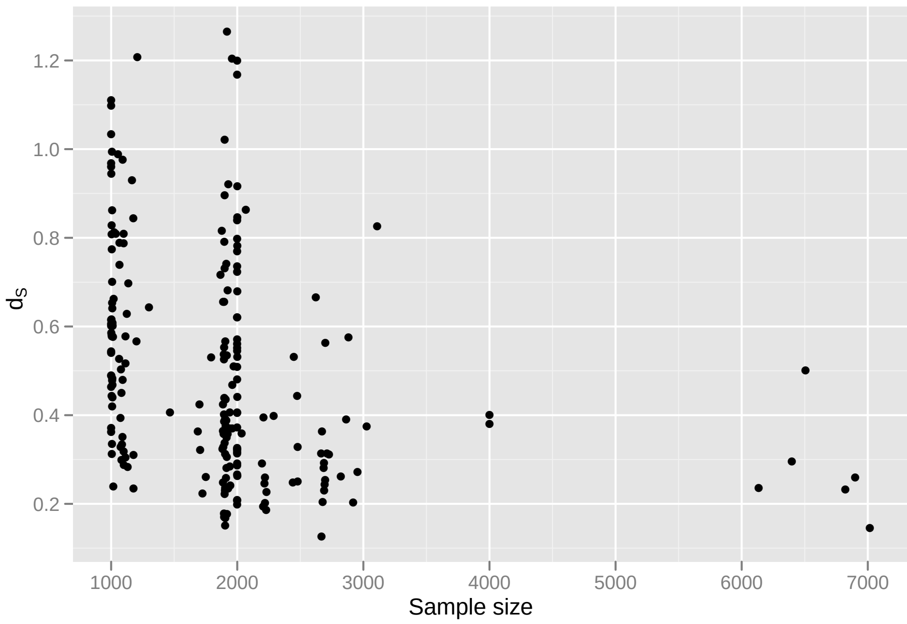

Figure 6: The deviations d_i from the smoothed series for $q = 40$ plotted versus sample size.

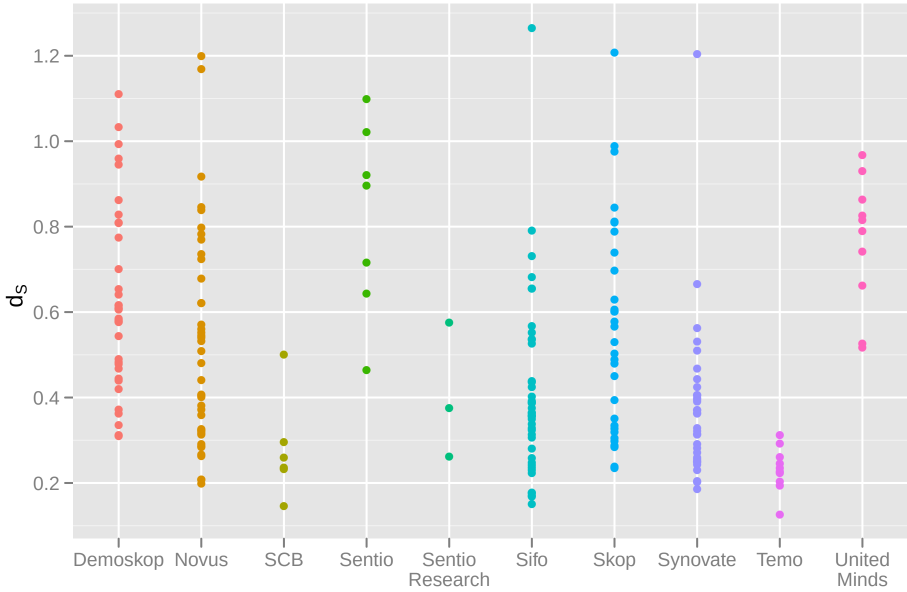

Figure 7: The deviations d_i from the smoothed series for $q = 40$ plotted versus the different polling institutes.

One other possible source of variation is shown in Figure 7, where we see the deviations plotted versus the polling institutes. There are apparently large differences between the various institutes. The small deviations for Statistics Sweden (SCB) are explained by the fact SCB uses much larger sample sizes: approximately 6000-7000 compared to the others institutes' sample sizes of 1000-2000. The large deviations of United Minds are probably explained by the fact that they use a web panel and not a telephone interviewing scheme. We believe that the other differences are probably due different degrees of weighting of the results using previous election results and various demographic statistics.

Finally, we also estimated a traditional Loess smooth for each of the nine parts, treating them each as a real univariate series. The smooth is for comparability done with $q = 40$. The smoothed series are plotted in Figure 8 together with the compositional Loess smoothed series. In the figure we see that both methods yield similar results, however in this case the traditional Loess tends to underestimate the larger parties and overestimate the smaller parties compared to the compositional Loess model.

4 Summary and conclusions

In this paper we perform smoothing of time series of compositional data, taking into account their compositional properties. We introduce a compositional weighted least squares estimation ($\mathcal{C}$-WLS) and a compositional locally weighted regression model ($\mathcal{C}$-Loess). The technique is flexible and can be performed very efficiently for certain distance measures. Is has been shown that the technique makes a difference as compared to traditional smoothing techniques where the inherent compositional structure is respected.

Acknowledgment

Henrik Oscarsson, Göteborg University, is acknowledged for making the data set used in the example easily accessible.

References

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Monographs on Statistics and Applied Probability. London: Chapman and Hall. (Reprinted in 2003 with additional material by

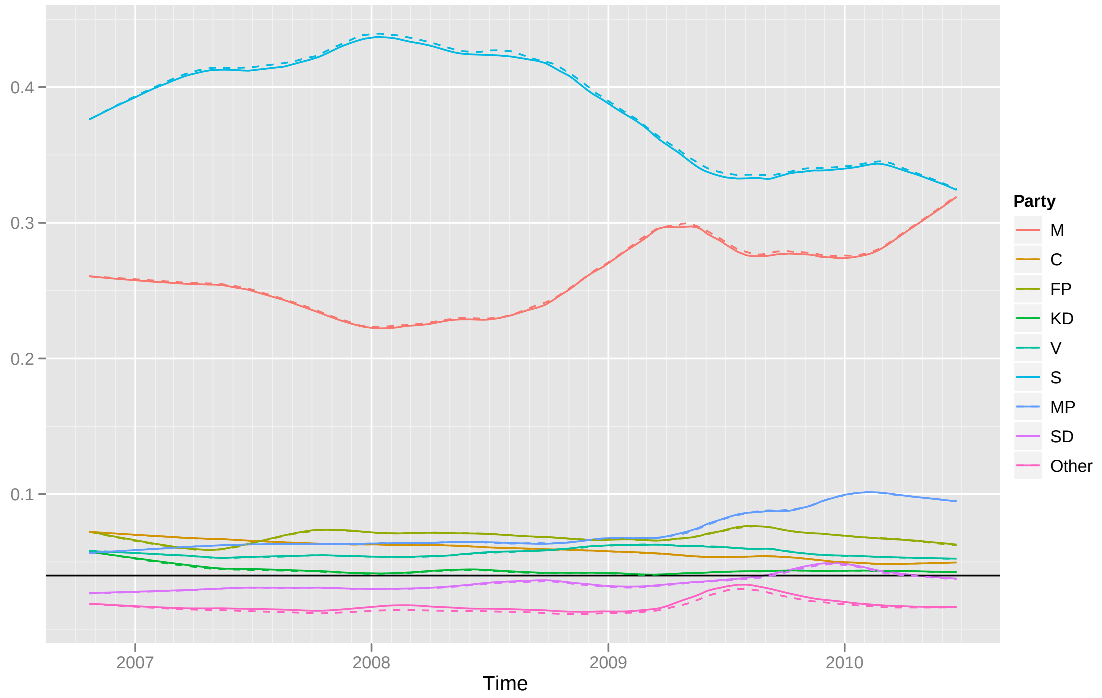

Figure 8: A comparison of the compositional Loess (dashed lines) and the traditional Loess models (solid lines).

The Blackburn Press).

Aitchison, J. (1992). On criteria for measures of compositional difference. *Math. Geol. 24*(4), 365–379.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc. 74*(368), 829–837.

Cleveland, W. S. and S. J. Devlin (1988). Locally weighted regression: An approach to regression analysis by local fitting. *J. Amer. Statist. Assoc. 83*(403), 596–610.

Martín-Fernández, J. A., C. Barceló-Vidal, and V. Pawlowsky-Glahn (1999). A measure of difference for compositional data based on measures of divergence. In S. J. Lippard, A. Naess, and R. Sinding-Larsen (Eds.), *Proceedings of the Fifth Annual Conference of the International Association for Mathematical Geology*, Volume 1, Trondheim (Norway), pp. 211–215.

Appendix

The centered log-ratio transform

For a composition $\boldsymbol{\alpha}$ in the interior of $\mathbb{S}^D$, the centered log-ratio transform clr and it's inverse iclr are defined by

$$\mathbf{y} = \mathrm{clr}(\boldsymbol{\alpha}) = (\ln \frac{\alpha_1}{g(\boldsymbol{\alpha})}, \ldots, \ln \frac{\alpha_D}{g(\boldsymbol{\alpha})})$$

where $g(\boldsymbol{\alpha}) = (\alpha_1 \cdots \alpha_D)^{1/D}$, and

$$\boldsymbol{\alpha} = \mathrm{iclr}(\mathbf{y}) = (\frac{\exp y_1}{\sum_1^D \exp y_i}, \ldots, \frac{\exp y_D}{\sum_1^D \exp y_i})$$

Hence

$$\mathrm{clr}(\boldsymbol{\alpha}) = (\ln \alpha_1, \ldots, \ln \alpha_D) - \mathbf{1} \ln((\alpha_1 \cdots \alpha_D)^{1/D}) = \mathbf{a}_C$$

where $\mathbf{a} = (\ln \alpha_1, \ldots \ln \alpha_D)'$ and $\mathbf{a}_C = (I_D - \frac{1}{D}\mathbf{1}_D \mathbf{1}_D^T)\mathbf{a}$.

For a perturbation $(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}) = (\alpha_1\beta_1, \ldots, \alpha_D\beta_D)/(\sum_1^D \alpha_i\beta_i)$, the clr is additive, i.e.

$$\mathrm{clr}(\boldsymbol{\alpha} \oplus \boldsymbol{\beta}) = (\ln(\alpha_1\beta_1/\sum_1^D \alpha_i\beta_i), \ldots, \ln(\alpha_D\beta_D/\sum_1^D \alpha_i\beta_i)) - \mathbf{1} \ln(\frac{\alpha_1\beta_1}{\sum_1^D \alpha_i\beta_i} \cdots \frac{\alpha_D\beta_D}{\sum_1^D \alpha_i\beta_i})^{1/D})$$

$$= (\ln(\alpha_1\beta_1), \ldots, \ln(\alpha_D\beta_D)) - \mathbf{1} \ln(\alpha_1\beta_1 \cdots \alpha_D\beta_D)^{1/D}) = \mathrm{clr}(\boldsymbol{\alpha}) + \mathrm{clr}(\boldsymbol{\beta})$$

Minimization of $Q_{\mathcal{C}}$

For each t'_k we minimize

$$Q_{\mathcal{C}} = \sum_{i=1}^n w_i(t'_k) d_S^2\{\mathbf{y}_i; \boldsymbol{\alpha} \oplus (t_i \odot \boldsymbol{\beta})\}$$

$$= \frac{1}{2D} \sum_{i=1}^n w_i \sum_{k=1}^D \sum_{j=1}^D (\ln(\frac{y_{ik}}{y_{ij}}) - \ln(\frac{\alpha_k}{\alpha_j}) - t_i \ln(\frac{\beta_k}{\beta_j}))^2$$

where y_{ij}, α_j and β_j are the components of $\mathbf{y}_i$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively, and where we in the notation have suppressed the dependence of w_i from t'_k.

Let $\eta_{ij} = \ln y_{ij}$ be the components of $\boldsymbol{\eta}_i$, $a_j = \ln \alpha_j$ and $b_j = \ln \beta_j$ then

$$Q_{\mathcal{C}} = \frac{1}{2D} \sum_{i=1}^n w_i \sum_{k=1}^D \sum_{j=1}^D (\eta_{ik} - \eta_{ij} - a_k + a_j - t_i(b_k - b_j))^2$$

With $\mathbf{c}_i = \boldsymbol{\eta}_i - \mathbf{a} - t_i\mathbf{b}$ and $\mathbf{C}_i = \mathbf{c}_i\mathbf{1}^T - \mathbf{1}\mathbf{c}_i^T$ we may write

$$Q_{\mathcal{C}} = \frac{1}{2D} \sum_{i=1}^n w_i \,\mathrm{trace}(\mathbf{C}_i\mathbf{C}_i^T) = \sum_{i=1}^n w_i \mathbf{c}_i^T (I_D - \frac{1}{D}\mathbf{1}_D\mathbf{1}_D^T)\mathbf{c}_i$$

$$Q_{\mathcal{C}} = \sum_{i=1}^n w_i(\boldsymbol{\eta}_i - \mathbf{a} - t_i\mathbf{b})^T (I_D - \frac{1}{D}\mathbf{1}_D\mathbf{1}_D^T)(\boldsymbol{\eta}_i - \mathbf{a} - t_i\mathbf{b})$$

or

$$Q_{\mathcal{C}} = \sum_{i=1}^n w_i\boldsymbol{\eta}_i^T (I_D - \frac{1}{D}\mathbf{1}_D\mathbf{1}_D^T)\boldsymbol{\eta}_i - 2\sum_{i=1}^n w_i\boldsymbol{\eta}_i^T (I_D - \frac{1}{D}\mathbf{1}_D\mathbf{1}_D^T)(\mathbf{a} + t_i\mathbf{b}) +$$

$$+ \sum_{i=1}^n w_i(\mathbf{a} + t_i\mathbf{b})^T (I_D - \frac{1}{D}\mathbf{1}_D\mathbf{1}_D^T)(\mathbf{a} + t_i\mathbf{b})$$

Introducing $\boldsymbol{\eta}_{Ci} = (I_D - \frac{1}{D}\mathbf{1}_D\mathbf{1}_D^T)\boldsymbol{\eta}_i$, $\mathbf{a}_C = (I_D - \frac{1}{D}\mathbf{1}_D\mathbf{1}_D^T)\mathbf{a}$ and $\mathbf{b}_C = (I_D - \frac{1}{D}\mathbf{1}_D\mathbf{1}_D^T)\mathbf{b}$ we have

$$Q_{\mathcal{C}} = \sum_{i=1}^{n} w_i \boldsymbol{\eta}_{Ci}^T \boldsymbol{\eta}_{Ci} - 2\sum_{i=1}^{n} w_i \boldsymbol{\eta}_{Ci}^T(\mathbf{a}_C + t_i\mathbf{b}_C) + \sum_{i=1}^{n} w_i(\mathbf{a}_C + t_i\mathbf{b}_C)^T(\mathbf{a}_C + t_i\mathbf{b}_C)$$

or

$$Q_{\mathcal{C}} = \sum_{i=1}^{n} w_i \boldsymbol{\eta}_{Ci}^T \boldsymbol{\eta}_{Ci} - 2\sum_{i=1}^{n} w_i \boldsymbol{\eta}_{Ci}^T \mathbf{a}_C - 2\sum_{i=1}^{n} w_i t_i \boldsymbol{\eta}_{Ci}^T \mathbf{b}_C +$$
$$+ \sum_{i=1}^{n} w_i \mathbf{a}_C^T \mathbf{a}_C + 2\sum_{i=1}^{n} w_i t_i \mathbf{a}_C^T \mathbf{b}_C + \sum_{i=1}^{n} w_i t_i^2 \mathbf{b}_C^T \mathbf{b}_C$$

By differentiating $Q_{\mathcal{C}}$ wrt $\mathbf{a}_C$ and $\mathbf{b}_C$ and setting these to zero we obtain, with $d_{11} = \sum_1^n w_i, d_{12} = \sum_1^n w_i t_i$, and $d_{22} = \sum_1^n w_i t_i^2$ and further $\mathbf{g}_1 = \sum_1^n w_i \boldsymbol{\eta}_{Ci}$ and $\mathbf{g}_2 = \sum_1^n w_i t_i \boldsymbol{\eta}_{Ci}$, the equations

$$d_{11}\mathbf{a}_C + d_{12}\mathbf{b}_C = \mathbf{g}_1$$

and

$$d_{12}\mathbf{a}_C + d_{22}\mathbf{b}_C = \mathbf{g}_2.$$

These have the solutions

$$\mathbf{a}_C = (d_{22}\mathbf{g}_1 - d_{12}\mathbf{g}_2)/(d_{11}d_{22} - d_{12}^2) = \sum_1^n u_i \boldsymbol{\eta}_{Ci}$$

and

$$\mathbf{b}_C = (d_{11}\mathbf{g}_2 - d_{12}\mathbf{g}_1)/(d_{11}d_{22} - d_{12}^2) = \sum_1^n v_i \boldsymbol{\eta}_{Ci}$$

where

$$u_i = (d_{22}w_i - d_{12}w_i t_i)/(d_{11}d_{22} - d_{12}^2)$$

and

$$v_i = (d_{11}w_i t_i - d_{12}w_i)/(d_{11}d_{22} - d_{12}^2)$$

which both depend on t'_k. Finally

$$\boldsymbol{\alpha} = \mathrm{iclr}(\mathbf{a}_C)$$

and

$$\boldsymbol{\beta} = \mathrm{iclr}(\mathbf{b}_C)$$

Compositional linear combination

For $\boldsymbol{\alpha} \oplus (t'_k \odot \boldsymbol{\beta})$ we have

$$\mathrm{clr}(\boldsymbol{\alpha} \oplus (t'_k \odot \boldsymbol{\beta})) = \mathrm{clr}(\boldsymbol{\alpha}) + t'_k\,\mathrm{clr}(\boldsymbol{\beta}) = \mathbf{a}_C + t'_k\mathbf{b}_C = \sum_1^n (u_i + t'_k v_i)\boldsymbol{\eta}_{Ci}$$

which can be written

$$\mathrm{clr}(\bigoplus_{i=1}^{n}(\ell_i(t'_k) \odot \mathbf{y}_i)) = \sum_1^n \ell_i(t'_k)\,\mathrm{clr}(\mathbf{y}_i) = \sum_1^n \ell_i(t'_k)\boldsymbol{\eta}_{Ci}$$

where $\ell_i(t'_k) = u_i + t'_k v_i$ where u_i, v_i of course also depend on t'_k as well as on W, q, d_E, and $t_1, \ldots, t_n$ but not $\mathbf{y}_i$.