Abstract Title Page

Not included in page count.

Title: Modeling Longitudinal Data with Generalized Additive Models: Applications to Single-Case Designs

Authors and Affiliations: Kristynn J. Sullivan & William R. Shadish, University of California, Merced

Abstract Body

Limit 4 pages single-spaced.

Background / Context:

Description of prior research and its intellectual context.

Single case designs (SCDs) are short time series that assess intervention effects by measuring units repeatedly over time both in the presence and absence of treatment. For a variety of reasons, interest in the statistical analysis and meta-analysis of these designs has been growing in recent years (Huitema, 2011; Shadish, Rindskopf, & Hedges, 2008). One of these reasons is that SCDs have been identified by groups such as the What Works Clearinghouse as one of the acceptable empirical designs that can be included in evidence based practice reviews (Kratochwill et al., 2010). When statistically analyzing or meta-analyzing SCD data, one must take into account both (a) trend, or systematic, non-zero, change over time that is not dependent on the implementation of a treatment, and (b) autocorrelation, or the serial dependence of observations nested within the same person. Various methods have been proposed for the statistical analysis of SCDs, but all of these methods have flaws when it comes to dealing with either trend or autocorrelation, or both. For example, trend is often either ignored or assumed to be linear, and autocorrelation is also often ignored, both of which are potentially suspect. Failing to address and properly model trend and the trend-treatment interaction can result in biased parameter estimates and inflated standard errors (Allison & Gorman, 1993). In addition, large autocorrelation of errors within a dataset can affect inferences based on those data (Huitema, 2011). Properly modeling the trajectory of the data can reduce the inflation of the autocorrelation that is due to model misspecification (Shadish, Ridskopf, & Hedges, 2008).

Purpose / Objective / Research Question / Focus of Study:

Description of the focus of the research.

This paper proposes modeling SCD data with Generalized Additive Models (GAMs), a semi-parametric method from which it is possible to estimate the functional form of trend directly from the data, arguably capturing the true functional form better than ordinary least squares regression methods in which the researcher must decide which functional form to impose on the data. In addition, autocorrelation estimates can be inflated when trend is not modeled properly, and so this paper also shows how to calculate the autocorrelation from residuals extracted from GAM models, which tends to result in shrunken estimates.

Significance / Novelty of study:

Description of what is missing in previous work and the contribution the study makes.

This work applies a semi-parametric regression method, Generalized Additive Models (GAMs), to single case designs (SCDs), which, to our knowledge has not been done previously. Modeling SCDs with GAMs, one is able to address questions of non-linearity in the trend and trend-treatment interaction terms of the model by estimating the linearity of these terms directly from the data compared to imposing hypothesized functional forms on the data in parametric regression. This may allow more accurate modeling of the functional form of trend and trend-treatment interactions, more accurate parameter estimates, and shrunken autocorrelation estimates.

Statistical, Measurement, or Econometric Model:
Description of the proposed new methods or novel applications of existing methods.

Generalized Additive Models. GAMs are an expansion of generalized linear models (GLM). GAM models replace the standard linear terms of a GLM with an additive predictor that consists of a sum of smoothing functions, or scatterplot smoothers, applied to some, or all, of the model covariates (Hastie & Tibshirani, 1986; Hothorn & Everitt, 2010; Wood, 2006). The basic model is:

$$Y_{ij} = \mathbf{X}_i\boldsymbol{\theta} + s_1(x_{1ij}) + s_2(x_{2ij}) + s_3(x_{3ij}) + \ldots + \varepsilon_{ij} \qquad (1)$$

where Y_{ij} is the outcome [1]or response variable for person i at timepoint j, $\mathbf{X}_i\boldsymbol{\theta}$ is the matrix row and corresponding parameter vector of the parametric model components (not including the smoothed predictor x's) for person i, $s_1(x_{1i}), \ldots s_p(x_{pi})$ are smoothing functions for each corresponding predictor (x's), and ε_{ij} is an error term, i.i.d., $N(0,\sigma^2)$. For example, the model could include an ordinary parametric dummy-coded treatment variable in the $\mathbf{X}_i$ design matrix, but have a smooth function separately for both the trend and the interaction terms, which is not possible within fully non parametric regression.

Within the GAM framework, to be able to estimate the smoothing terms (s's) using GLM methods such as iterative least squares, the smoothing terms have to be represented in such a way that the GAM becomes a linear model. To do this, one chooses a basis, or a set of linearly independent vectors, that defines a functional space. The functional space is essentially the geometric plane, drawn from and defined by a specific dataset, from which the smoothing functions are drawn. The basis vectors, when linearly combined, can represent any potential vector in the functional space. All of the potential smoothing terms of the model are an element, or basis function, of the chosen basis. So, any potential smoothing term is some linear combination of linearly independent vectors in the basis. That is, the linear combination of vectors in the basis results in a smoothing function that is directly estimated from one's data. Choosing a basis is an essential step in this process: it allows the estimation of a nonlinear term from the data, but constrains the geometrical plane (functional space) from which they can be estimated (so that the smoothing does not result in an unrealistic value, such as a 100th order polynomial smoother).

There are many potential basis options. A common basis is penalized cubic regression splines (CRS). Spline bases relate the smoothing function to the entire domain of data rather than a single point of the data. CRSs are constructed from pieces of cubic polynomial curves joined together into a continuous function. The curves are joined together at the knots of the data set; knots are the places where an inflection in the curve appears. CRSs are computationally efficient, and their results are easily interpretable. They also can be implemented on small data sets; almost always with 20 or more data points, but also sometimes with as few as six. With CRSs the researcher has to specify where to place knots, or the location of the potential bends in the functional relationship. One can choose either to manually place the knots across the data set, or to equally space these knots across the span of the data. Generally, evenly spacing the knots across the data (the computer program default) captures the shape of the data quite well, and so this is not an arduous process.

[1] Y_{ij} is only the generalized outcome, when the outcome is normally distributed. In this paper, the outcomes are either counts or percentages/proportions modeled with the Poisson and binomial distributions, respectively. Both of these functions use a log link function, and so in this paper, Y_{ij} actually is $\log(Y_{ij})$.

Introducing smoothing parameters requires estimating the degree of smoothing necessary for each covariate, for example, the trend term in the present case. Each *s* function of a GAM model contains a smoothing parameter. The smoothing parameter estimates the optimal amount of smoothing to fit the data while simultaneously adding a penalty for increased "wiggliness" of the smoothing function. Adding a penalty matrix to the least squares estimation model avoids over-fitting the smooth to the data. Within this framework, *s* approaches a straight line as the smoothing parameter approaches infinity. The optimal degree of smoothness can be estimated directly from the data. The optimal smoothing parameter is chosen by calculating a generalized cross-validation (GCV) score of each iteration. The smaller the GCV score, the better the model fit.

The model output also lists the effective degrees of freedom of the smoothing term. When using cubic regression splines, effective degrees of freedom are *very* roughly equivalent to the polynomial order of the smoother plus one (Hothorn & Everitt, 2010, Chapter 10). An example of this phenomenon is shown in the middle row of Figure 1; (please insert Figure 1 here) a quadratic term (or second degree polynomial) has just over three effective degrees of freedom. However, it is important to remember that this rule of thumb is extremely approximate, and as the effective degrees of freedom approach one (e.g., between 1 and 3), it no longer applies.

Modeling procedure. For every time series, the analyst runs four different models to assess non linearities. Each model includes four terms: 1) an intercept, or the participant's initial outcome level, 2) a continuous trend variable, in the present case measured as calendar time across sessions (e.g. two sessions conducted one day apart would be 1, 2; two sessions conducted one week apart would be 1, 8; session number can be used if time is not available), 3) a treatment variable, usually dummy coded (0 for baseline, 1 for treatment; this could be expanded either by adding another dummy code to the model for a second treatment or by changing this dummy code into a categorical variable [e.g., with levels 0, 1, 2] if the treatments could be thought of as ordinal), and 4) a trend-treatment interaction term, calculated as:

$$Int_{ij} = [X_{ij} - (t_1)]*z_{ij}, \qquad (3)$$

where X is the trend variable value for data point ij, t_1 is the trend variable value for the first data point in the first treatment phase, and z_i is the treatment dummy code for data point ij. Modeling the trend-treatment interaction in this way, rather than perhaps the traditional interaction calculation (X_i*z_i), has been recommended to capture the change in slope, as well as level treatment change, beginning at the implementation of treatment , rather than from the first time point (Huitema & McKean, 2000). The four models differ in which terms are smoothed. After running all four models, the model that best fits that particular time series data set is selected.

The first model is a GAM linear model with no smoothing functions:

$$Y_{ij} = \beta_{0i} + \beta_1 X_{ij} + \beta_2 z_{ij} + \beta_3 Int_{ij} + \varepsilon_{ij}. \qquad (4)$$

This model produces equivalent parameter estimates to a GLM, as long as both are modeled with the same data distribution. However, the standard errors are calculated in a slightly different fashion for the GAM within the mgcv R package. For more information, see Wood (2010). This model is the reference model to which all other models are compared. The second model applies a smoother to the interaction term (trend by treatment) only:

$$Y_{ij} = \beta_{0i} + \beta_1 X_i + \beta_2 z_{ij} + s_3(Int_{ij}) + \varepsilon_{ij}. \qquad (5)$$

This model implies that nonlinearities in the data are a function of the implementation of treatment. That is, trend over time, if it exists, is linear, but once treatment is introduced, the data curves. Imagine a case in which a behavior has a floor effect during baseline, and when treatment is introduced, the behavior increases rapidly but then asymptotes. This situation would be captured in this model. The third model applies the smoothing function to the trend term only:

$$Y_{ij} = \beta_{0i} + s_1(X_{ij}) + \beta_2 z_{ij} + \beta_3 Int_{ij} + \varepsilon_{ij}. \quad (6)$$

This model implies that any nonlinearity is a function of time, and not the implementation of treatment. Imagine a behavior that follows a cubic order over time, and implementation of the treatment does not affect the trajectory of the behavior at all. This model would capture that relationship. Lastly, the fourth model applies a smoothing function to both the interaction term and the trend term:

$$Y_{ij} = \beta_0 + s_1(X_{ij}) + \beta_2 z_{ij} + s_3(Int_{ij}) + \varepsilon_{ij}. \quad (7)$$

This model is the most complex, and implies that there is both a nonlinear data trajectory without the implementation of treatment, as well as a (potentially different) nonlinear data trajectory after the implementation of treatment.

Usefulness / Applicability of Method:

Demonstration of the usefulness of the proposed methods using hypothetical or real data.

These model can be implemented in R using the mgcv package (Wood, 2010).
R. The first author has fully annotated code available. This procedure can be implemented in a wide range of data sets, and this paper gives several examples of how to implement and interpret these model.

Findings / Results:

Description of the main findings with specific details.

Table 1 shows the model output of all four model runs on one example case, a SCD intervention study in which the authors intervened to increase social and play skills in children with autism, using a multiple baseline design (Liber, Frea, & Symon, 2008). (Please insert Table 1 here.)The participant in the selected case is a nine-year-old boy (Wally), and the outcome is the percentage of unprompted social play skills. In the linear (baseline) model, there was not a significant treatment effect. However, in all three models with a smoother, there is a significant treatment effect, indicating that properly modeling the data trajectory of this example (the raw data is presented in Figure 2 (please insert Figure 2 here)) is essential to assess intervention effectiveness.

Conclusions:

Description of conclusions, recommendations, and limitations based on findings.

Generalized Additive Models provide a flexible way to model SCD data, allowing the data to inform the researcher both as to whether significant trend or trend treatment interaction exists, as well as which of those terms need nonlinear representations and which can remain linear. GAMs often capture the nonlinearities that appear visually in graphs of the outcome variables presented in SCD literature, and account for them in their measurement of a treatment effect. For many example cases, the model that best fit the data had either the trend or trend-treatment interaction term smoothed, indicating at least slight non-linearity for those parameters. Though this is only a small subset of available data, these findings cast doubt on the typical assumption of SCD researchers that trend, if it exists, is linear. In addition, extracting the residuals from the best fitting model resulted in a shrunken autocorrelation.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Allison, D. B., & Gorman, B. S. (1993). Calculating effect sizes for meta-analysis: The case of the single case. *Behavioral Research Therapy, 31*, 621–631.

Hastie, T. & Tibshairani, R. (1986). Generalized additive models. *Statistical Science*, *1*(3), 297-310.

Hothorn, T. & Everitt, B. S. (2010). A handbook of statistical analyses using R, second edition. Boca Raton, Florida: Chapman and Hall/CRC.

Huitema, B. E. (2011). The analysis of covariance and alternatives: Statistical methods for experiments, quasi-experiments, and single-case studies. Hoboken, New Jersey: John Wiley and Sons, Inc.

Huitema, B. E. & McKean, J. W. (2000). Design specification issues in time-series intervention models. *Educational and Psychological Measurement*, *60*(1), 38-58. DOI: 10.1177/00131640021970358

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M & Shadish, W. R. (2010). *Single-case designs technical documentation*. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

Liber, D. B., Frea, W. D., & Symon, J. B. G. (2008). Using time-delay to improve social play skills with peers for children with autism. *Journal of Autism and Developmental Disorders*, *38*(2), 312-323. doi: 10.1007/s10803-007-0395-z

Shadish, W.R., Rindskopf, D.M., & Hedges, L.V. (2008). The state of the science in the meta-analysis of single-case experimental designs. *Evidence-Based Communication Assessment and Intervention,* 3, 188–196.

Shadish, W.R., & Sullivan, K.J. (2011). Characteristics of single case designs used to assess Intervention effects in 2008. *Behavior Research Methods*, *43*, 971-980.

Wood, S. N. (2006). Generalized additive models: An introduction with R. Boca Raton, Florida: Chapman and Hall/CRC.

Wood, S. (2010). mgcv: GAMs with GCV/AIC/REML smoothness estimation and GAMMs by PQL. Available at http://cran.r-project.org/package=mgcv.

Appendix B. Tables and Figures

Not included in page count.

Table 1

GAM and GLM Results for Example 3: proportion of social play skills demonstrated during play activities, results in logits and (proportions)

GAM Model 1 - all linear terms

Linear Terms

	Estimate	SE	*T*-statistic	*p*-value
Intercept	-2.66(0.07)	0.47	-5.67	<.001
Trend	0.05(0.51)	0.04	1.21	0.23
Treatment	0.21(0.55)	0.43	0.49	0.63
Interaction	-0.03(0.49)	0.05	-0.61	0.55

Smoother Terms - N/A

Adjusted R^2 = 0.62 Deviance Explained = 76.4% GCV Score = 0.75

GAM Model 2 - interaction smoother

Linear Terms

	Estimate	SE	*T*-statistic	*p*-value
Intercept	-2.64(.07)	0.54	-4.88	<0.001
Trend	0.05(.51)	0.02	2.99	0.005
Treatment	-0.07(.48)	0.21	-2.97	0.005

Smoother Terms

	Estimated DF	*F*-statistic	*p*-value
Interaction	3.57	44.13	<.001

Adjusted R^2 = 0.97 Deviance Explained = 96.3% GCV Score = 0.13

GAM Model 3 - trend smoother

Linear Terms

	Estimate	SE	*T*-statistic	*p*-value
Intercept	-0.99(.27)	0.60	-1.64	0.109
Treatment	-0.64(.35)	0.22	-2.97	0.005
Interaction	0.01(.50)	0.03	0.41	0.684

Smoother Terms

	Estimated DF	*F*-statistic	*p*-value
Trend	3.55	45.99	<.001

Adjusted R^2 = 0.97 Deviance Explained = 96.4% GCV Score = 0.13

GAM Model 4 - interaction and trend smoother

Linear Terms

	Estimate	SE	T-statistic	p-value
Intercept	-0.76(.31)	0.16	-4.72	<.001
Treatment	-0.64(.34)	0.22	-2.97	0.005

Smoother Terms

	Estimated DF	F-statistic	p-value
Trend	3.55	47.40	<.001
Interaction	1.00	0.17	0.684

Adjusted R^2 = 0.97 Deviance Explained = 96.4% GCV Score = 0.13

Figure 1. Graphical example of curvature of data as estimated degrees of freedom increases

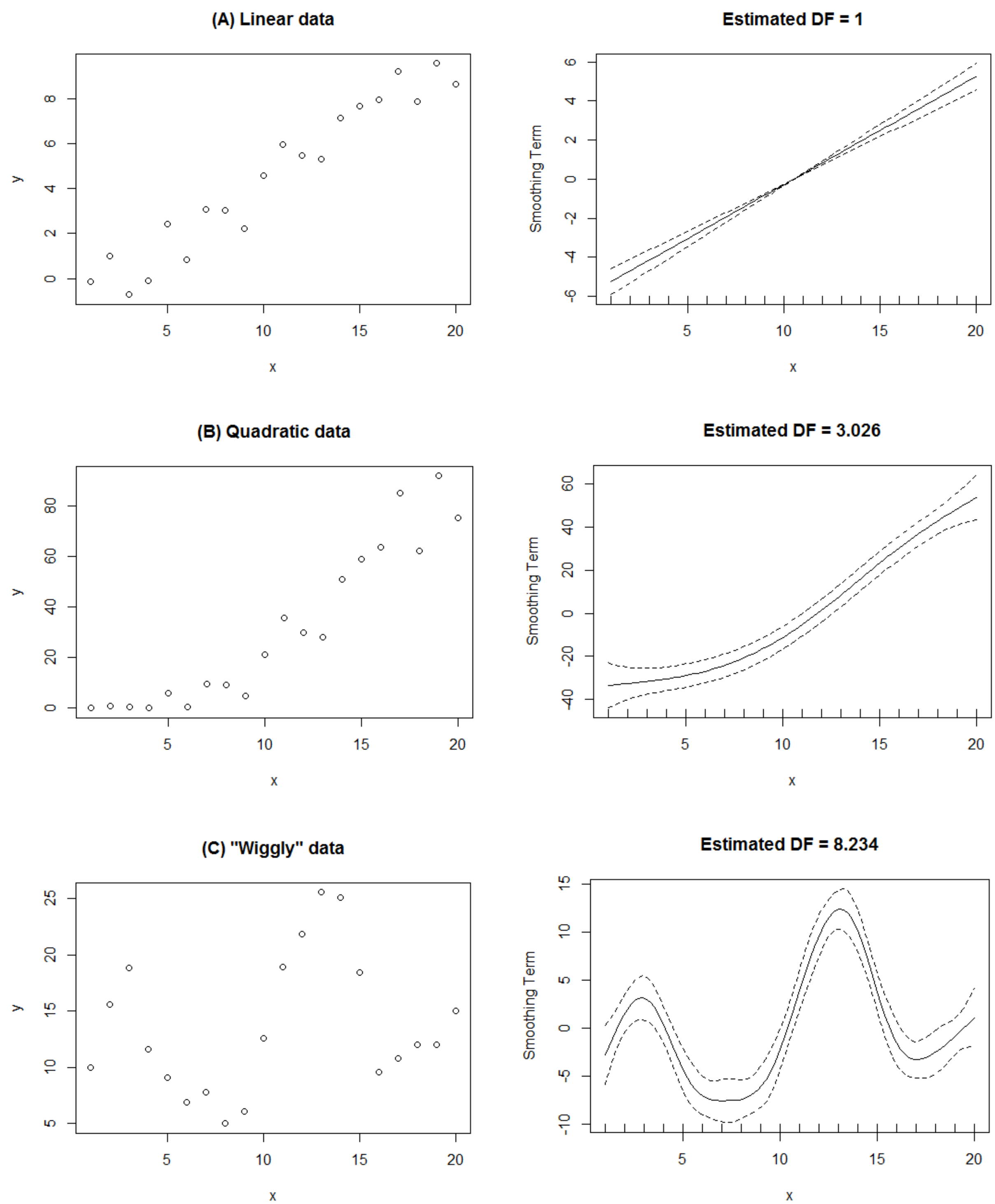

Figure 8. Example three, time series graph of percentage of social play skills demonstrated during play activities with Wally's peers.