

Adaptive Fractional Polynomial Modeling

George J. Knafl

School of Nursing, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Email: gknafl@unc.edu

How to cite this paper: Knafl, G.J. (2018) Adaptive Fractional Polynomial Modeling. *Open Journal of Statistics*, 8, 159-186.
<https://doi.org/10.4236/ojs.2018.81011>

Received: January 19, 2018

Accepted: February 23, 2018

Published: February 26, 2018

Copyright © 2018 by author and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Regression analyses reported in the applied research literature commonly assume that relationships are linear in predictors without assessing this assumption. Fractional polynomials provide a general approach for addressing non-linearity through power transforms of predictors using real valued powers. An adaptive approach for generating fractional polynomial models is presented based on heuristic search through alternative power transforms of predictors guided by k -fold likelihood cross-validation (LCV) scores and controlled by tolerance parameters indicating how much a reduction in the LCV score can be tolerated at given stages of the search. The search optionally can generate geometric combinations, that is, products of power transforms of multiple predictors, thereby supporting nonlinear moderation analyses. Positive valued continuous outcomes can be power transformed as well as predictors. These methods are demonstrated using data from a study of family management for mothers of children with chronic physical conditions. The example analyses demonstrate that power transformation of a predictor may be required to identify that a relationship holds between that predictor and an outcome (dependent or response) variable. Consideration of geometric combinations can identify moderation effects not identifiable using linear relationships or power transforms of interactions. Power transformation of positive valued continuous outcomes along with their primary predictors can resolve model assumption problems.

Keywords

Adaptive Regression, Childhood Chronic Conditions, Fractional Polynomials, Moderation, Nonlinearity

1. Introduction

Regression analyses reported in the applied research literature commonly assume that relationships are linear in predictors without assessing this assumption. When addressed, nonlinear relationships are often based on standard po-

lynomials, usually quadratic or cubic polynomials. However, these models cannot handle general nonlinearity. For this reason, Royston and Altman [1] proposed addressing nonlinearity with fractional polynomials based on power transforms of predictors using real valued powers. They proposed consideration of finite sets of powers. For example, they recommended the eight powers $-2, -1, -0.5, 0, 0.5, 1, 2, 3$ (with the 0 case corresponding to the natural log transform) for degree 1 fractional polynomials based on a single power transform (for more details on standard fractional polynomial modeling also see [2] [3]). Power transforms are also called Box-Tidwell transforms [4]. Box and Cox [5] proposed a related set of transforms for transforming outcome (dependent or response) variables. Carroll and Ruppert [6] used these Box-Cox transforms to transform both outcomes and predictors.

The finite set of recommended powers can effectively model nonlinearity in many situations, but not always. Adaptive modeling generalizes standard fractional polynomial modeling to address the complete set of real valued powers. Knafl *et al.* [7] first formulated adaptive modeling in the Poisson regression case. This was extended to handle generalized linear models and modeling of variance/dispersions [8] and also to handle repeated measures modeling [9]. Knafl and Ding [10] provided a general adaptive approach for fractional polynomial modeling of univariate and multivariate continuous, discrete, and count outcomes.

This paper describes adaptive fractional polynomial modeling and demonstrates the kinds of novel insights it can provide through example adaptive analyses of data from a study of family management by mothers of children with chronic physical conditions such as diabetes and Crohn's disease [11]. These methods can be used in any area of the behavioral, health, and social sciences as well as any other application area that uses regression modeling. They can provide novel insights into the data not possible with standard regression methods.

2. Methods

2.1. Fractional Polynomials

Power transforms X^p of a positive valued predictor X are well-defined for all real valued powers p . This can be extended to transforms $f(X, p)$ for real valued predictors X by defining $f(X, p)$ as X^p for $X > 0$, as 0 for $X = 0$, and as $\cos(\pi \cdot p) \cdot |X|^p$ for $X < 0$ where \cos denotes the cosine function, π is the usual constant, and $|X|$ is the absolute value of X . Note that the sign of $f(X, p)$ when $X < 0$ oscillates between ± 1 as p varies. Indicator predictor variables X with values 0 and 1 do not require transformation.

Fractional polynomials have the form

$$\beta_1 \cdot f(X_1, p_1) + \beta_2 \cdot f(X_2, p_2) + \cdots + \beta_r \cdot f(X_r, p_r)$$

where, if any of the predictors X_i and X_j are the same, then the associated powers p_i and p_j are not the same. An intercept can be included by setting

$X_1 = 1$, that is, the unit predictor, with $p_1 = 1$. When X is positive valued and $p = 0$, $f(X, 0)$ corresponds to the natural log transform $\log(X)$ if the model contains an intercept parameter (demonstrated by taking the limit as $p \rightarrow 0$); otherwise $f(X, 0)$ corresponds to the unit transform, adding in an intercept to the model. Royston and Altman [1] also consider fractional polynomials to include interactions between transforms of predictors and their natural logs, but these are not considered here.

2.2. Geometric Combinations

Power transforms of interactions can be generalized to geometric combinations consisting of products of power transforms of multiple distinct predictors using possibly different powers. These have the form

$$f(X_1, p_1) \cdot f(X_2, p_2) \cdots f(X_m, p_m)$$

where X_i are distinct for $1 \leq i \leq m$. They can replace standard terms $f(X_j, p_j)$ in fractional polynomials.

Geometric combinations can be used to address the issue called moderation [12], also called effect modification, but generalized to handle nonlinearity. An adaptive model containing geometric combinations does not necessarily imply distinct moderation. That only holds if the model with the geometric combinations substantially improves (as defined later) on the associated additive model.

2.3. Likelihood Cross-Validation

Let $S = \{s : 1 \leq s \leq n\}$ denote indexes for the observations of a data set. Partition these indexes, and hence the data, into k disjoint subsets $S(h)$ for $1 \leq h \leq k$; these subsets are called folds. Let θ denote the vector of model parameters, $L(\cdot, \theta)$ a likelihood function or a likelihood-like function such as the extended quasi-likelihood used with generalized linear models [13], and $\theta(S)$ the estimate of θ obtained by maximizing $L(S, \theta)$ over θ . The associated k -fold likelihood cross-validation score is given by

$$\text{LCV} = \prod_{1 \leq h \leq k} L(S(h), \theta(S \setminus S(h)))^{1/n}.$$

In other words, the LCV score is the product of likelihoods for the data in the folds $S(h)$ evaluated at parameter estimates computed with the data in the other folds $S \setminus S(h)$ and normalized by the sample size. Fold assignment is random, but using the same initial seed for models of the same outcome Y . In this way, the same fold assignments are used for all such models, and then their LCV scores are comparable.

Larger LCV scores indicate better models, but not necessarily substantially (or distinctly) better models. This issue can be addressed using LCV ratio tests computed using the χ^2 distribution as for standard likelihood ratio tests. These tests are expressed in terms of a threshold for a substantial (distinct) percent decrease in the LCV score, which varies with the sample size. The formula is given

in Section 4.4.2 of [10] and is computed conservatively with the 95th percentile 3.84146 of the χ^2 distribution with the smallest positive integer degrees of freedom $DF=1$.

If models M_1 and M_2 have LCV scores satisfying $LCV(M_1) < LCV(M_2)$, model M_2 provides a substantial (distinct) improvement over model M_1 if the percent decrease $100\% \cdot (LCV(M_2) - LCV(M_1)) / LCV(M_2)$ in the LCV score is larger than the threshold for the data. Otherwise model M_1 is a competitive alternative to model M_2 . If model M_1 is also simpler, then it is preferable as a parsimonious competitive alternative. LCV ratio tests are more conservative than standard tests for zero slopes in the sense that the removal from the model of a predictor with a significant slope can generate a competitive LCV score compared to the model with the predictor included (an example is provided later).

LCV scores should only be compared when computed for the same outcome variable Y . However, when Y is continuous and positive valued, it is possible to compute power-adjusted $LCV(q)$ scores for power transforms Y^q of Y (the formulation is given in Section 6.3 of [10]). The power $q=0$ corresponds to the natural log transform $\log(Y)$. These scores can be compared to choose a power transform for Y along with power transforms for its predictors. Estimated means for transformed Y can be inversed transformed to generate estimated means for untransformed Y (i.e., using the power $1/q$ for $q \neq 0$ and the exponential transform for $q=0$).

2.4. Adaptive Modeling Process

The adaptive modeling process is formulated in Chapter 20 of Knafl and Ding [10]. An overview is provided here. A base model is first expanded by systematically adding in power transforms of primary predictors. The transform added to the model next is the one generating the best LCV score for primary predictors currently under consideration for inclusion in the model. Next, the expanded model is contracted, removing power transforms from the model and adjusting the powers of the remaining transforms to improve the LCV score. When the contraction leaves the expanded model unchanged, the powers of the expanded model transforms are adjusted to improve the LCV score.

The process is controlled by tolerance parameters indicating how much of decrease in the LCV score can be tolerated at given stages of the process. For example, the contraction stopping tolerance is set using a LCV ratio test, so that the final model is parsimonious. Primary predictors are dropped from consideration for inclusion in the expansion if the inclusion of their transforms decreases the LCV score by more than an associated tolerance parameter. Transforms are similarly dropped from consideration for removal in the contraction if their removal decreases the LCV score by more than the associated tolerance parameter.

The expansion process can optionally generate geometric combinations. Each power transform is multiplied together with the prior product of power trans-

forms without adjusting the powers in those prior transforms, and so the final geometric combination is further power transformed to increase the LCV score. Transforming the whole geometric combination with a single power simplifies the process compared to adjusting each of the powers in the geometric combination separately.

Fractional polynomial models can be generated for continuous outcomes using adaptive linear regression modeling with the identity link function (as demonstrated later), for dichotomous outcomes using adaptive logistic regression modeling with the logit link function, for ordinal outcomes using adaptive ordinal regression modeling with the cumulative logit link function, for nominal outcomes using adaptive multinomial regression modeling with the generalized logit link function, and for count/rate outcomes using adaptive Poisson regression modeling with the natural log link function. These outcomes can be univariate or multivariate. Fractional polynomial models are special cases of standard models, and so have the same assumptions. Furthermore, once the predictors and powers of a fractional polynomial are specified, efficiency of estimation of associated slope parameters is the same as for standard models of the same type.

The adaptive modeling process can be applied to generate fractional polynomial models for only means (or expectations) assuming constant variances or unit dispersions. However, such assumptions need assessment. This can be conducted through adaptive modeling of both means and variances/dispersions using fractional polynomial transforms of primary predictors. The natural log link function is used for modeling the variances/dispersions. The assumption of constant variances or unit dispersions is reasonable if the associated adaptive model provides a competitive alternative to the adaptive model for both means and variances/dispersions in combination.

2.5. Computational Support

The adaptive modeling process has been implemented in the `genreg` (for general regression) SAS[®] (SAS Institute, Inc., Cary, NC) macro, which supports adaptive linear, logistic, and Poisson regression modeling of either univariate or multivariate continuous, dichotomous, ordinal polytomous, nominal polytomous, and count/rate outcomes. The `ypower` SAS macro is available for generating power transforms Y^q of positive valued continuous outcomes Y as well as power transforms of their predictors. It uses a grid search to choose the power q . These macros report in their output the threshold for a substantial percent decrease in the LCV score. Knafl & Ding [10] provide extensive coding details on the use of both macros. Section 4 describes code for generating the adaptive analyses reported here.

2.6. Example Data

Example adaptive analyses are conducted using data from a study of family management by parents of children with chronic physical conditions such as

diabetes and Crohn's disease [11]. Data are available for both mothers and fathers, but only the data for 344 partnered mothers and for 65 single mothers are used in the example analyses. The child's functional status (CFS) is measured using the Functional Status II [14] assessing the child's health status. Higher scores indicate better child functioning. The effort to manage the child's condition and the parental mutuality (PM) in managing the condition are measured using scales of the Family Management Measure [11]. Higher effort scores indicate lower levels of family management of the condition while higher PM scores indicate better levels. PM is only collected for partnered parents. There are no missing data values for all reported analyses.

Institutional Review Board approval was obtained for the original study and also for conducting secondary analyses of these data including the analyses reported in this manuscript.

3. Results

The threshold for a substantial percent decrease in the LCV score for analyses of data for the 65 single mothers is 2.91% while it is 0.56% for analyses of data for the 344 partnered mothers. LCV scores are computed with $k = 10$ folds unless otherwise indicated.

3.1. Adaptive Analyses of Data for Single Mothers

Table 1 provides results for standard polynomial models of the outcome effort for single mothers as a function of the primary predictor CFS. The linear polynomial model generates the best LCV score 0.058705 but the associated F test is nonsignificant ($p = 0.051$). Moreover, the constant model has LCV score 0.058412 with insubstantial percent decrease 0.50% (*i.e.*, less than the threshold of 2.91% for the data), and so is a parsimonious, competitive alternative. F tests for the quadratic and cubic polynomial models are also nonsignificant, and these generate inferior LCV scores. In this case, a standard polynomial assessment of possible nonlinear dependence of effort on CFS leads to the conclusion that mean effort does not depend on CFS.

Some authors feel that higher order integer power transforms should not be included in models without also including lower order integer powers [15]. To assess this issue for the single mother data, the model in only CFS^2 has LCV score 0.058467 and so improves on the full quadratic polynomial model with LCV score 0.058188 (**Table 1**). In this case, the percent decrease 0.48% is insubstantial, but the model with only the quadratic term is simpler while also providing an improvement. On the other hand, the model in only CFS^3 has LCV score 0.058269 while the full cubic polynomial model with LCV score 0.054129 (**Table 1**) generates a substantial percent decrease of 7.10%. In this case, insisting on inclusion of lower order terms makes the model highly ineffective.

An adaptive expansion generates the model based on CFS^{-1} plus an intercept.

Table 1. Standard polynomial modeling of the outcome effort as a function of the primary predictor child functional status for single mothers.

Model	Terms	10-fold LCV score	Percent Decrease	Model F Test
Constant	1	0.058412	0.50%	
Linear	1, CFS	0.058705	0.00%	$F(1, 63) = 3.95, p = 0.051$
Quadratic	1, CFS, CFS ²	0.058188	0.88%	$F(2, 62) = 2.06, p = 0.136$
Cubic	1, CFS, CFS ² , CFS ³	0.054129	7.79%	$F(3, 61) = 1.57, p = 0.207$

LCV: likelihood cross-validation; CFS: child functional status.

The associated slope is significantly ($t(1) = 2.07, p = 0.043$) nonzero. The LCV score is 0.059062, and so the constant model with insubstantial percent decrease in the LCV score of 1.10% is a parsimonious, competitive alternative. This is an example where the LCV ratio test is more conservative than the standard test for a zero slope. On the other hand, the adaptive contraction removes the intercept and adjusts the power for CFS from -1 to -0.4 . This model has LCV score 0.060490 and provides a substantial improvement over the constant model with substantial percent decrease 3.44%. This is an example where a zero intercept model is required to identify a substantial impact to a primary predictor. The percent decrease for the model linear in CFS is substantial at 2.95%, and so mean effort is distinctly nonlinear in CFS. The same adaptive model is generated with 5 and 15 folds, and so the results are robust to the choice of the number of folds.

Figure 1 contains the plot of the raw data and the predicted value curve generated by the adaptive model for mean effort as a function of CFS. Mean effort decreases nonlinearly from 19.6 at the smallest observed value for CFS of 42.9 to 13.9 at the largest observed (and possible) value for CFS of 100. The estimated constant standard deviation is 3.9. The raw data indicate that the functional status for these children is relatively high (only seven or 10.8% of the children have CFS scores less than 70). The possible values for effort range from 4 to 20, and so these mean effort values are relatively high.

3.2. Adaptive Analyses of Data for Partnered Mothers

The model for mean effort for partnered mothers as a linear function of CFS has LCV score 0.061785 and provides a substantial improvement over the constant model with LCV score 0.057434 and percent decrease 7.04% (*i.e.*, larger than the threshold of 0.56% for the data). Moreover, mean effort significantly decreases as would be expected with increased CFS (estimated slope -0.14 , $t(1) = -7.46$, $p < 0.001$). On the other hand, the adaptive model for mean effort restricted to include at most one transform of CFS is based on CFS⁶ with an intercept. The LCV score for this model is 0.063383, providing a substantial improvement over the linear model in CFS with percent decrease 2.52%. Consequently, mean effort

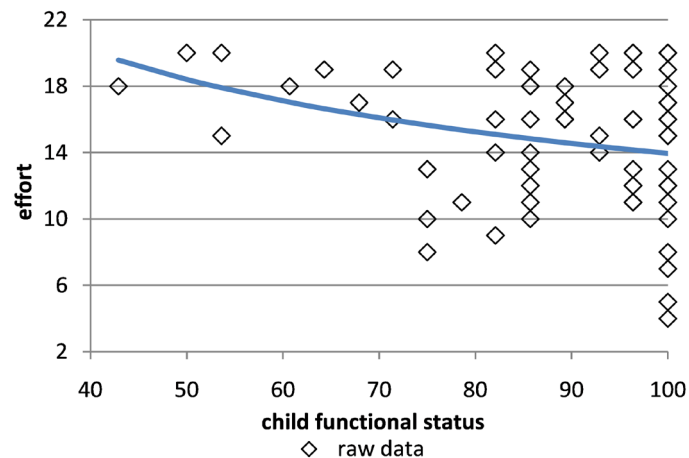


Figure 1. Raw data and predicted value curve for the adaptive model for mean effort as a function of child functional status for single mothers.

is distinctly nonlinear in CFS. Moreover, the best LCV score for the eight Royston and Altman powers for degree 1 fractional polynomials is 0.062620 achieved with CFS^3 and with substantial percent decrease 1.20%. This is an example using actual, nonsimulated data where restricting to the Royston and Altman recommended powers generates a substantially inferior model compared to using the adaptive modeling process.

Using the adaptive modeling process applied to the means allowing for multiple transforms of CFS with constant variances generates the model for the means based on three transforms: CFS^6 , $CFS^{2.8}$, and $CFS^{-2.5}$ without an intercept and LCV score 0.064067. Under this model, mean effort (plot not provided) decreases from the lowest observed CFS value of 42.9 up to CFS 65.4, then increases up to CFS 82.1, and then decreases after that up to the largest observed value for CFS of 100.

The purpose of the parent study [11] that collected these data was to develop an instrument measuring family management of childhood chronic conditions. CFS was collected for construct validity purposes, and it was hypothesized that negative measures of family management like effort would decrease with increased CFS. Thus, a monotonically decreasing relationship had been previously hypothesized, and so the generated nonmonotonic adaptive model was counter-intuitive.

One way to assess monotonicity of mean effort in CFS is to restrict to a single power transform of CFS, but that is the model based on CFS^6 with LCV score 0.063383 and substantial percent decrease 1.07%. On the other hand, restricting the contraction not to remove the intercept generates the model based on the two transforms CFS^5 and $CFS^{6.12}$ with LCV score 0.063755 and insubstantial percent decrease 0.49%. Moreover, under this competitive model, mean effort is monotonically decreasing in CFS as would be expected.

Figure 2 contains the plot of the raw data and the predicted value curve for

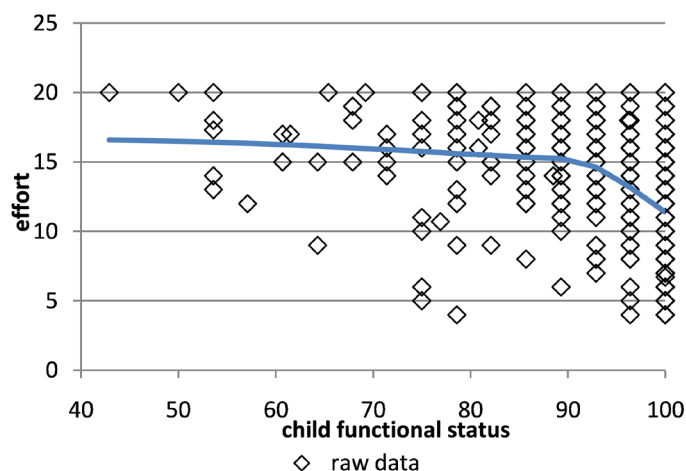


Figure 2. Raw data and predicted value curve for the adaptive model for mean effort as a function of child functional status for partnered mothers based on untransformed effort and restricted to include an intercept.

the monotonic model. Mean effort decreases slowly up to about CFS 90 and then decreases quickly up to CFS 100. The estimated constant standard deviation is 3.8. This plot suggests the possibility that mean effort is constant for $CFS \leq 90$. The adaptive model in CFS bounded to be no lower than 90, that is, $\max(CFS, 90)$ is based on the single transform $\max(CFS, 90)^6$ with an intercept and a competitive LCV score 0.063885 with insubstantial percent decrease 0.28%. The same adaptive model is generated with 5 and 15 folds, and so the results are robust to the choice of the number of folds. Under this model, mean effort (plot not provided) is constant at 15.6 up to CFS 90 and then decreases to 11.4 at CFS 100.

The adaptive model for both the means and variances in terms of $\max(CFS, 90)$ has means based on $\max(CFS, 90)^6$ with an intercept and variances based on $\max(CFS, 90)^{1.3}$ without an intercept. The LCV score is 0.064329, substantially improving on the constant variances model in $\max(CFS, 90)$ with substantial percent decrease 0.69%, indicating that the variances are nonconstant in $\max(CFS, 90)$. Mean effort is constant at 15.6 up to CFS 90 and then decreases to 11.4 at CFS 100 while the standard deviation for effort is constant at 3.4 up to CFS 90 and then increases to 4.1 at CFS 100 (plots not provided). However, this model generates two extreme outliers with standardized residuals -3.39 and -3.10 . The standardized residuals are also skewed towards the low end ranging from -3.39 to 2.11, suggesting model assumption problems.

An adaptive search over power transforms for effort along with power transforms of $\max(CFS, 90)$ for modeling mean transformed effort with constant variances identifies the power $q = 1.5$ with the best power-adjusted score $LCV(1.5) = 0.064814$. This is a substantial improvement over the model for untransformed effort with power-adjusted score $LCV(1) = 0.063885$ (the same as its non-power-adjusted LCV score) and substantial percent decrease 1.43%.

Mean effort^{1.5} is based on the one transform: $\max(\text{CFS}, 90)^{-4}$ without an intercept. The adaptive model for the means and variances in $\max(\text{CFS}, 90)$ provides an improvement with larger LCV score 0.064897 but the associated constant variances model generates an insubstantial percent decrease 0.13%. Consequently, mean effort^{1.5} is reasonably considered to have variances constant in $\max(\text{CFS}, 90)$.

Figure 3 contains the plot of the raw data and the predicted value curve for mean untransformed effort as a function of bounded CFS generated by the adaptive constant variances model for effort^{1.5} (obtained by inverse transforming estimated mean effort^{1.5} using the power $1/q = 2/3$). Mean effort is constant at 15.8 up to CFS 90 and then decreases to 11.9 at CFS 100.

Standardized residuals for transformed effort range from -2.77 to 2.43 , and so there are no extreme outliers and the skewness has been reduced. **Figure 4** contains the associated normal (probability) plot, which is reasonably close to linear. Adaptive transformation of effort has resolved model assumption problems, including outliers, skewness, and nonconstant variances.

3.3. Adaptive Moderation Analyses of Data for Partnered Mothers

Mean untransformed effort decreases significantly with untransformed $\max(\text{CFS}, 90)$ assuming constant variances (estimated slope -0.42 , $t(1) = -9.03$, $p < 0.001$; estimated standard deviation 3.8). Whether this effect is moderated by PM can be addressed with a standard linear moderation analysis by adding PM and the interaction $\max(\text{CFS}, 90) \cdot \text{PM}$ into the model. The slope for this interaction is nonsignificant ($t(1) = -0.51$, $p = 0.609$), indicating that linear moderation does not hold. The adaptive additive (*i.e.*, without considering interactions) model for the means in $\max(\text{CFS}, 90)$ and PM depends on $\max(\text{CFS}, 90)^6$ and $\text{PM}^{-0.1}$ without an intercept and with LCV score 0.064201. The adaptive model in only $\max(\text{CFS}, 90)$ is a competitive alternative with LCV score 0.063885 (as reported earlier) and insubstantial percent decrease 0.49%.

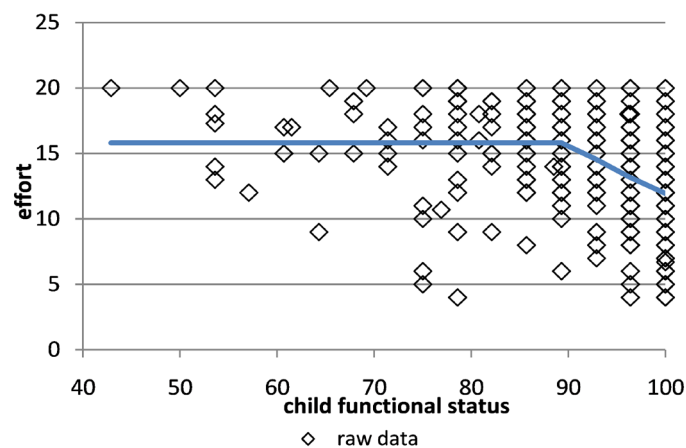


Figure 3. Raw data and predicted value curve for the adaptive model for mean effort as a function of bounded child functional status for partnered mothers based on transformed effort.

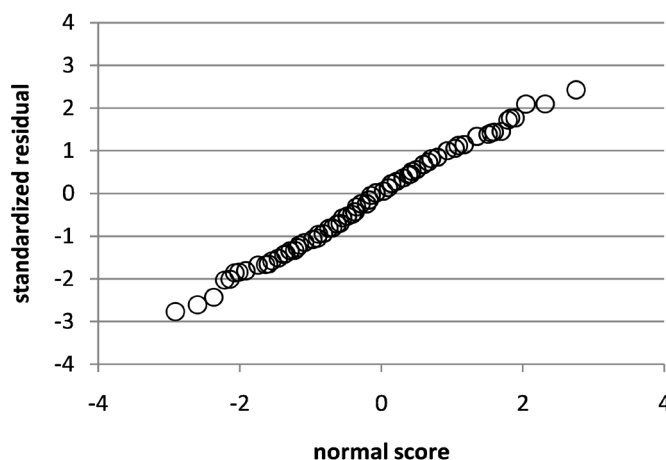


Figure 4. Normal plot generated by standardized residuals for the adaptive model for mean transformed effort as a function of bounded child functional status for partnered mothers.

One way to address the possibility of nonlinear moderation is to generate the adaptive model for mean effort based on the three primary predictors $\max(\text{CFS}, 90)$, PM, and $\max(\text{CFS}, 90) \cdot \text{PM}$, then compare its LCV score to that of the adaptive additive model in $\max(\text{CFS}, 90)$ and PM. The generated model is the same as the adaptive additive model suggesting that nonlinear moderation does not hold.

The more general way to address the possibility of nonlinear moderation is to generate the adaptive model for mean effort based on the two primary predictors $\max(\text{CFS}, 90)$ and PM along with geometric combinations in $\max(\text{CFS}, 90)$ and PM. The generated model is based on $\left(\text{PM}^{-2} \cdot \max(\text{CFS}, 90)^{-36}\right)^{-0.1599}$, $\left(\text{PM}^{-2} \cdot \max(\text{CFS}, 90)\right)^{-0.05}$ and $\left(\max(\text{CFS}, 90)^5 \cdot \text{PM}^{-2.3}\right)^{1.5}$ without an intercept and with LCV score 0.064666. This model provides a substantial improvement over the adaptive additive model with percent decrease 0.72%. Consequently, nonlinear moderation does hold, but can only be identified using geometric combinations.

Figure 5 contains predicted value curves for mean effort as a function of $\max(\text{CFS}, 90)$ for selected values of PM. For CFS up to 90, mean effort is constant, at lower levels with increased PM, and with less of a decrease with increasing levels of PM. For CFS from 90 to 100, mean effort decreases with increased CFS close to linearly, at lower levels and with steeper rates with increased PM, and with less of a change with increasing levels of PM. Assessments of moderation using transformed effort and nonconstant variances as well as of model assumptions are needed but are not addressed here for brevity.

3.4. Simulation

An example was presented earlier where the recommended set of degree 1 powers had too restrictive a range to effectively identify a power well outside of that

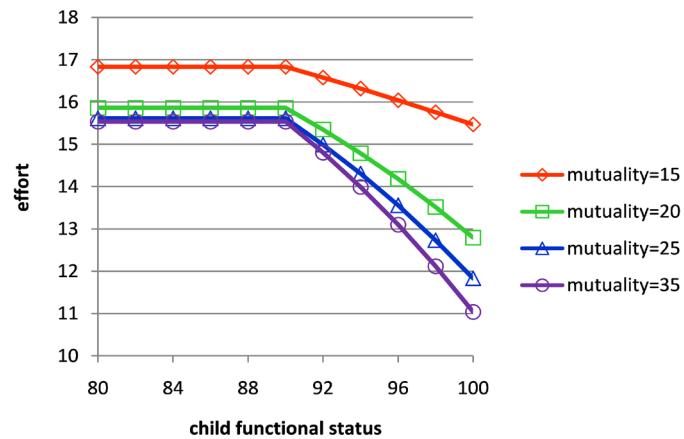


Figure 5. Predicted value curves for the adaptive model for mean untransformed effort as a function of bounded child functional status moderated by parental mutuality for partnered mothers.

range. Knafl and Ding [10] (Section 2.12) provide a simulation demonstrating that this holds more generally. The true power for the simulated data was -7 . The adaptively generated power was -6.9 , and the model based on the true power was a competitive alternative to the one based on the estimated power. The best recommended power was -2 , which generated a very substantial percent decrease of 55.8% compared to a threshold of 0.19% (due to a sample size of 1001 observations).

However, the issue of whether the recommended set of powers can be ineffective when the true power is in between the extreme values of -2 and 3 has not yet been addressed. For that reason, the following simulation was conducted. The predictor x_{sim} had 101 equally spaced values from 0 to 1, and so 0.01 units apart. The outcome y_{sim} satisfied

$$y_{sim} = (0.5 + 1 \cdot x_{sim}^{1.5} + \epsilon) / 1.5$$

where ϵ was normally distributed with mean 0 and standard deviation 0.01. The normalizing constant 1.5 was set to the maximum value for the numerator rounded to 1 decimal digit so that y_{sim} values were bounded by 1. This meant that after rounding the true intercept was 0.33, the true slope 0.67, and the true standard deviation 0.0067. These data are plotted in Figure 6. The threshold for a substantial percent decrease in the LCV score for these data was 1.86%.

Using $k = 10$ folds, the adaptive model for mean y_{sim} with constant variances was based on the true power 1.5. Rounded estimated values for the intercept, slope, and standard deviation were 0.33, 0.67, and 0.0067, respectively; all equal to the true values. The LCV score was 34.8004. The best LCV score for the recommended powers was generated by the power 2 with value 10.0569 and very substantial percent decrease of 71.1%. These results demonstrate that the set of recommended powers in between the extreme powers of -2 and 3 can also produce ineffective models when the variability in the data is small, as for these simulated data.

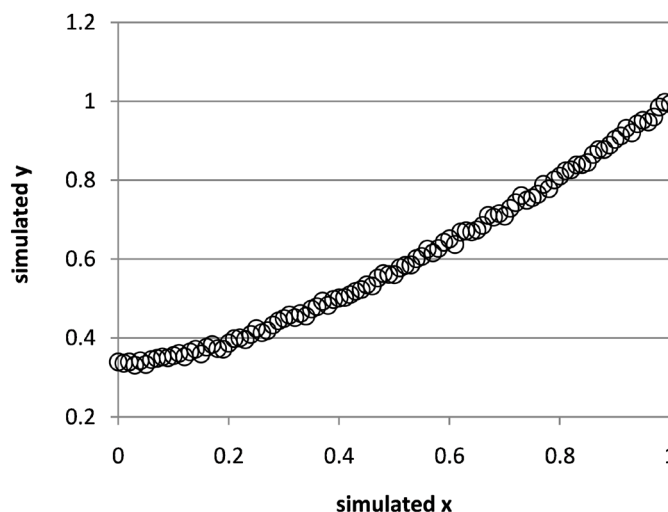


Figure 6. Simulated data with mean of the simulated y variable a function of the simulated x variable raised to the power 1.5.

4. Overview of Adaptive Fractional Polynomial Modeling in SAS

The `genreg` and `ypower` macros can be loaded into SAS using `%include` statements. The version is indicated by the date in the name of the macro. The version of `genreg` used in reported analyses was created on 3/19/2017 and the version of the `ypower` macro on 2/25/2017. These versions of the macros as well as code to generate reported analyses are available at

<http://www.unc.edu/~gknafl/AFPM2.html> (accessed February 4, 2018).

Assume that a data set with name `partnered` has been created in the SAS default library. Also assume that this data set contains the variables named `effort`, `CFS`, and `mutuality` containing values for the effort to manage the child's chronic condition, the child functional status, and parental mutuality in managing the condition, respectively, for 344 partnered mothers.

4.1. Standard Regression Modeling

A standard linear polynomial model for mean effort as a linear function of `CFS` with constant variances is generated as follows.

```
%genreg(modtype=norml,datain=partnered,yvar=effort,xvars=CFS,procmod=y);
```

The `genreg` macro is invoked by attaching a percent sign (%) to its name, followed by a list of settings for its macro parameters separated by commas and in parentheses. The `modtype` parameter determines the likelihoods used to generate parameter estimates and LCV scores. In this case, the value “norml” means that the outcome is continuous and to be analyzed using linear regression models and likelihoods based on the normal distribution. Other choices include “logis” for discrete outcomes and “poiss” for count outcomes. The `datain` parameter provides the name of the data set to use in the analysis, in this case the `partnered` data set. The `yvar` parameter names the outcome (or y) variable to be the varia-

ble effort. The `xvars` parameter provides a list of names for predictor (or `x`) variables, in this case only the single predictor CFS. The `procmod` parameter is used to generate SAS output for the appropriate SAS procedure (or PROC), in this case PROC REG. The value “`y`” is short for yes. The other possible value is “`n`” for no.

All other macro parameters take on their default values (as specified in the `genreg` code). For example, the default value for the `xintrcpt` parameter is “`y`” indicating that the model for the means should include an intercept parameter. Consequently, the model generated in the above code is the standard linear polynomial model for mean effort with an intercept. The `foldcnt` macro parameter is used to set the number of folds. In this case, its default value of “10” is used.

In the above code, the `xpowers` macro parameter also has its default empty setting meaning not to transform the `xvars` variables (or equivalently, transform them with power 1). To generate a standard quadratic polynomial model in CFS use the settings “`xvars=CFS CFS`” and “`xpowers=1 2`”.

4.2. Adaptive Regression Modeling

An adaptive model for mean effort as a function of at most one transform of CFS can be generated as follows.

```
%genreg(modtype=norml,datain=partnered,yvar=effort,
        expand=y,expxvars=CFS, multtrns=n,contract=y);
```

The base model for this analysis is the constant model for mean effort due to the default settings “`xintrcpt=y`” and “`xvars=`” (*i.e.*, the empty setting for `xvars` meaning include no predictors). The setting “`expand=y`” requests that this base model be expanded by including power transforms of the variables listed in the `expxvars` parameter setting, in this case only the variable CFS. The setting “`multtrns=n`” means that multiple transforms of the `expxvars` variables are not to be included in the expanded model. The setting “`contract=y`” means contract the expanded model, adjusting powers for remaining transforms with each removal of a transform. In this case, the base constant model is expanded to include the single transform of CFS with power 6. The contraction leaves the expanded model unchanged.

To allow for multiple transforms of CFS in the model, change to “`multtrns=y`”, which is the default setting; so removing “`multtrns=n`” from the code has the same effect. In this case, the base constant model is expanded to include three transforms of CFS with powers 6, 5, and -2.5 in that order. The contraction removes the intercept, adjusts the three powers to 6, 2.8, and -2.5 , and then stops. This model can be generated directly as follows.

```
%genreg(modtype=norml,datain=partnered,yvar=effort,xintrcpt=n,
        xvars=CFS CFS CFS,xpowers=6 2.8 -2.5);
```

By default, the contraction considers removal of the intercept from the model for the means. This is controlled by the `nocnxint` (for no contraction of the `x` in-

tercept) macro parameter with default setting “n”. Add “nocnxint=y” to the above code to restrict the contraction not to remove the intercept. As before, the base constant model is expanded to include the three transforms of CFS with powers 6, 5, and -2.5 , but now the contraction removes the transform with power -2.5 rather than the intercept, adjusts the other two powers to 5 and 6.12, and then stops.

This model (see [Figure 2](#)) suggests that mean effort is constant up to the value 90 for CFS. This can be assessed by generating the adaptive model in the variable `bnddCFS`, defined to equal CFS bounded to be at least 90, assuming this variable has been added to the partnered data set (using the code “`bnddCFS=max(CFS,90);`”). The adaptively generated model in `bnddCFS` provides a competitive alternative to the one in unbounded CFS and so `bnddCFS` is used in subsequent analyses instead of CFS.

4.3. Adaptive Variance Modeling

The `genreg` macro also supports adaptive modeling of the variances (or the dispersions for logistic and Poisson regression) in terms of fractional polynomial models. The log of the variances is modeled in terms of fractional polynomials to guarantee that the variances are positive valued. At each step of the expansion, a best transform is identified for expanding the model for the means and also for the model for the variances. The choice with the better LCV score is added next to the model. Similarly, at each step of the contraction, the next term removed is the one for either the means or the variances generating the better LCV score.

An adaptive model for both means and variances of effort in `bnddCFS` is requested as follows.

```
%genreg(modtype=norml,datain=partnered,yvar=effort,expand=y,
        expxvars=bnddCFS,expvvars=bnddCFS,contract=y);
```

The `expvvars` macro parameter requests that the expansion and contraction also consider transforms of `bnddCFS` for modeling the variances. There are several parameters like `expvvars` for modeling the variances (or the *v* component) analogous to those for modeling the means (or *x* component) like `expxvars`. For example, the `vintrcpt` and `vvars` parameters control the base model for the variances with default settings “`vintrcpt=y`” and “`vvars=`”, thereby generating the constant variances model. Consequently, the above code requests a base model with constant means and variances.

The first two pages of the `genreg` output contain details on macro parameter settings, including the value for the threshold for a substantial percent decrease in the LCV score, rounding in this case to 0.56%. The third page contains results for the base model. Part of this output is in [Table 2](#). The sample size in number of measurements denoted by *m* was 344. The estimated constant mean was 13.5. The estimated log of the constant variance was 2.86 so that exponentiating and taking the square root gives an estimated standard deviation of 4.18 (this value is reported in the output, but that has not been included in [Table 2](#)). The last entry

described as the “ m th root of the likelihood using deleted predictions” is the LCV score, which equals 0.057434.

Page 4 of the output describes settings of macro parameters controlling the expansion. Page 5 describes the expanded model. Part of this output is in **Table 3**. The expansion first adds the transform of bnddCFS to the power 6 to the model for the means (order = 1), next it adds the power transform of bnddCFS to the power 3 to the model for the variances (order = 2), and then stops. The LCV score rounds to 0.064183.

Page 6 of the output describes settings of macro parameters controlling the contraction. Page 7 describes the contracted model. Part of this output is in **Table 4**. The contraction removes the intercept from the model for the variances (order = 1), adjusts the power for bnddCFS in the model for the variances to 1.3, leaves the power for bnddCFS in the model for the means unchanged at 6, and then stops. The LCV score improves to 0.064329.

4.4. Outcome Transformation

The nonconstant variances model in bnddCFS provides a substantial improvement in the LCV score over the associated constant variances model (as reported

Table 2. Part of the output describing the constant base model for mean effort with constant variances for partnered mothers.

m , the number of measurements:		344
base expectation component		
predictor	power	estimate
XINTRCPT	1	13.510756
base log variance component		
predictor	power	estimate
VINTRCPT	1	2.8628203
m th root of the likelihood using deleted predictions:		0.057434

Table 3. Part of the output describing the expanded model for both means and variances of effort as a function of bounded child functional status (bnddCFS) for partnered mothers.

expanded expectation component				
predictor	power	estimate	score	order
XINTRCPT	1	20.280044	0.057434	0
bnddCFS	6	−8.87E−12	0.0638847	1
expanded log variance component				
predictor	power	estimate	score	order
VINTRCPT	1	1.45469	0.057434	0
bnddCFS	3	1.3631E−6	0.0641826	2
m th root of the likelihood using deleted predictions:				0.0641826

Table 4. Part of the output describing the contracted model for both means and variances of effort as a function of bounded child functional status (bnddCFS) for partnered mothers.

contracted expectation component			
predictor	old power	new power	estimate
XINTRCPT	1	1	20.280639
bnddCFS	6	6	-8.87E-12
contraction component unchanged			
contracted log variance component			
predictor	old power	new power	estimate
bnddCFS	3	1.3	0.0070624
discarded	old power	score	order
		0.0641826	0
VINTRCPT	1	0.0643285	1
<i>m</i> th root of the likelihood using deleted predictions:			0.0643285

earlier), indicating that the usual assumption of constant variances is not appropriate for these data. This model also has extreme outliers and the residuals are distinctly skewed. Outcome transformation has the potential to remedy such problems. This is addressed using the `ypower` macro.

The following code requests generation of power adjusted $LCV(q)$ scores for adaptive modeling of mean transformed effort as a function of bnddCFS over a grid of powers q .

```
%ypower(datain=partnered,yvar=effort,yfst=-2.5,ycnt=11,ystp=0.5,
        expand=y, expxvars=bnddCFS,contract=y);
```

The `datain`, `yvar`, `expand`, `expxvars`, and `contract` macro parameters have the same meanings as for the `genreg` macro. The requested values for the outcome powers q range from -2.5 to 2.5 by steps of size 0.5 . The first value for q is set to -2.5 by the `yfst` macro parameter. The step size of 0.5 is set by the `ystp` macro parameter. The `ycnt` macro parameter determines how many powers to consider; 11 in this case so that the last power is 2.5 . Only constant variances were considered to reduce the computation times. The `ypower` macro invokes the `genreg` macro to generate the adaptive model for mean transformed effort in terms of bnddCFS for each requested power q , then uses `genreg` output to compute associated $LCV(q)$ scores. Part of the output for the above code is provided in **Table 5**. The best $LCV(q)$ score of 0.064814 is generated by the power 1.5 . A grid search around 1.5 over multiples of 0.1 from 1.1 to 1.9 can be generated by changing to the settings “`yfst=1.1`”, “`ystp=0.1`”, and “`ycnt=9`”. The best score is still generated by the power 1.5 .

The adaptive model for both the means and variances of effort transformed by the power 1.5 is generated with the following code.

Table 5. Part of the output for choosing an outcome transformation for effort with its means a function of bounded child functional status (bnddCFS) for partnered mothers.

Power for Transforming effort	Power-Adjusted LCV Score
-2.5	0.0000002
-2.0	0.0000277
-1.5	0.0020992
-1.0	0.0162368
-0.5	0.0397838
0.0	0.0548231
0.5	0.0605673
1.0	0.0638847
1.5	0.0648138
2.0	0.0636068
2.5	0.0607855

```
%ypower(datain=partnered,yvar=effort,yfst=1.5,ycnt=1,expand=y,
        expxvars=bnddCFS,expvvars=bnddCFS,contract=y,nogprint=n);
```

The settings “*yfst=1.5*” and “*ycnt=1*” guarantee that only the power 1.5 is considered. The *nogprint* macro parameter requests that *genreg* output be generated along with *ypower* output. The default setting is “*nogprint=y*”, indicating not to generate any *genreg* output.

4.5. Adaptive Moderation

Assume the variable *interact* has been added to the partnered data set (using code “*interact=bnddCFS*mutuality;*”). A standard linear moderation model can be generated as follows.

```
%genreg(modtype=norml,datain=partnered,yvar=effort,
        xvars=bnddCFS mutuality interact,procmod=y);
```

The setting “*procmod=y*” requests standard PROC REG output including the *p* value for the test of a significant interaction term, which in this case is 0.609.

The adaptive additive model in *bnddCFS* and *mutuality* is generated as follows.

```
%genreg(modtype=norml,datain=partnered,yvar=effort,expand=y,
        expxvars=bnddCFS mutuality,contract=y);
```

The generated model for the means is based on *bnddCFS* transformed by the power 6 and *mutuality* transformed by the power -0.1 without an intercept. The adaptive model in *bnddCFS*, *mutuality*, and *interact* is generated as follows, but it is the same as the adaptive additive model.

```
%genreg(modtype=norml,datain=partnered,yvar=effort,expand=y,
        expxvars=bnddCFS mutuality interact,contract=y);
```

Geometric combinations in bnddCFS and mutuality are requested as follows.

```
%genreg(modtype=norml,datain=partnered,yvar=effort,expand=y,  
geomcmbn=y, expxvars=bnddCFS mutuality,contract=y);
```

The setting “geomcmbn=y” requests that the expansion consider geometric combinations. The default setting is “geomcmbn=n”, which restricts to additive models. The generated model is described earlier and can be generated directly as follows.

```
%genreg(modtype=norml,datain=partnered,yvar=effort,xintrcpt=n,  
xgcs=mutuality -2 bnddCFS -36 : mutuality -2 bnddCFS 1:  
bnddCFS 5 mutuality -2.3, xgcpowrs=-0.1599 -0.05 1.5);
```

The xgcs macro parameter describes geometric combinations to include in the model for the means (or x component). Geometric combinations are determined by lists of variables and powers separated by colons(:). For example, “mutuality -2 bnddCFS -36” means raise mutuality to the power -2, raise bnddCFS to the power -36, and multiply these two transforms together. The xgcpowrs macro parameter provides associated powers for transforming the geometric combinations of the xgcs macro parameter in the same order; for example, the geometric combination associated with “mutuality -2 bnddCFS -36” is transformed by the power -0.1599.

5. Discussion

Consideration of general power transforms of a primary predictor may be required to identify that a relationship holds between that predictor and an outcome. For example, standard linear, quadratic, and cubic polynomial models for mean effort for single mothers as a function of child functional status (CFS) suggested that mean effort was constant in CFS. However, the adaptive model identified a distinctly nonlinear relationship (**Figure 1**) between mean effort and CFS. Moreover, this was only possible with consideration of zero intercept models.

Consideration of general powers can identify distinctly better models than restricting to the Royston and Altman recommended set of powers. For example, mean effort for partnered mothers depended on CFS⁶ which provided a substantial improvement on the best recommended degree 1 power transform CFS³ for this case.

Adaptive fractional polynomial modeling can identify distinctly nonlinear relationships even when there is a significant linear relationship. For example, mean effort for partnered mothers was significantly related to untransformed CFS, but adaptively transforming CFS generated a substantially better model based on a nonlinear relationship. This model was counter-intuitively not monotonically decreasing in CFS. However, the model generated by restricting the adaptive search not to remove the intercept was competitive monotonic alternative (**Figure 2**), indicating that monotonicity was a reasonable conclusion for these data.

5.1. Bounding Predictors

The model of [Figure 2](#) suggested that mean effort was constant in CFS up to a value of 90 and then decreased after that, and the associated bounded CFS model provided a competitive alternative. Bounded CFS is a special case of splines as often used in nonparametric modeling, for example, in multiple adaptive regression splines (MARS) [\[16\]](#). Its applicability to these data provides the useful insight into the relationship between mean effort and CFS that effort can only be improved (lowered) by increasing CFS to relatively high levels.

CFS might not lend itself well to intervention, but suppose mean effort was constant in low values of a measure amenable to intervention like family functioning. An intervention that produces only small to moderate improvements in family functioning would only be beneficial for families with relatively high levels of family functioning to start with and so would have too low of a scientific premise to justify an efficacy study. On the other hand, suppose mean effort decreased distinctly for improvements in family functioning at low to moderate levels but was reasonably considered to be constant for high levels of family functioning. If a study to improve family functioning included substantial numbers of families with high levels of family functioning, the intervention would have little effect on mean effort for these families, and the efficacy of the intervention would likely not be supported. However, if having a low to moderate level of family functioning was an inclusion criterion, the intervention would be likely to produce distinct reductions in effort supporting its efficacy.

5.2. Zero Intercept Models

Models commonly include an intercept even when it is nonsignificant, and so zero intercept models may seem inappropriate to consider. However, these models are quite simple to understand in the continuous outcome context; they mean that the mean outcome is zero when the predictors all have value zero. Moreover, they can generate competitive alternatives to nonzero intercept models more parsimoniously, and substantially better models in some cases (as demonstrated in the example analyses). However, the adaptive modeling process can be constrained not to remove the intercept, and the associated adaptive models can provide useful insights into the data (as in the above example supporting an expected monotonic relationship).

5.3. LCV Ratio Tests

LCV ratio tests can be more conservative than standard tests for zero coefficients. For example, mean effort depended significantly on CFS transformed with the power -1 , but the associated LCV score was not substantially better than the score for the constant model. This is not an isolated case. For example, for the 21 significant cases considered in [\[17\]](#), the constant model was a competitive alternative using LCV scores in 5 (23.8%) of these cases; this also held for 10 (83.3%) of the 12 significant cases considered in [\[18\]](#) and for 12 (56.1%) of 21 significant

cases considered in [19]. This is an important property of LCV scores, indicating that adaptive modeling usually generates more parsimonious models than model selection procedures based on standard p values, thereby reducing the chance of overfitting. Knafl and Ding [10] provided the following partial justification for why this would hold in general. Model selection based on LCV scores is asymptotically equivalent to model selection based on Akaike information criterion (AIC) scores [20] as long as these are appropriately transformed into larger is better scores. Associated AIC ratio tests use more conservative thresholds for significance than standard likelihood ratio tests. Since the recommended approach for multiple fractional polynomial modeling [2] [3] is based on standard likelihood ratio tests, it would be more likely to produce more complex models and with more of a potential to overfit the data.

5.4. Adaptive Moderation

Adaptive moderation modeling can identify moderation effects not identifiable using linear moderation modeling. Moreover, consideration of geometric combinations may be needed to identify nonlinear moderation compared to only considering power transforms of interactions. For example, mean untransformed effort for partnered mothers depended significantly on untransformed bounded CFS, but including parental mutuality (PM) in the model along with its interaction with bounded CFS generated a nonsignificant slope for the interaction. An adaptive analysis considering transforms of the interaction (and so with bounded CFS and PM raised to the same power) suggested that nonlinear moderation did not hold. However, consideration of geometric combinations (and so with different powers for bounded CFS and PM) identified nonlinear moderation of the dependence of mean effort on bounded CFS by PM (Figure 5).

Linear moderation of an effect to a predictor X on a continuous outcome Y by a third variable Z can be addressed by also including Z as a covariate and its interaction $X \cdot Z$ in the model for the mean of Y and then testing for a zero slope for the interaction. An alternative approach is to test for a significant change in R^2 for the model with the interaction compared to the covariate model including X and Z without the interaction using an F test. In the adaptive nonlinear setting, the first approach based on a test for a zero coefficient cannot be extended, but the other approach can be. First, generate the adaptive model allowing for separate transforms of X and Z along with geometric combinations in X and Z . If the generated model contains no geometric combinations, then adaptive moderation does not hold. However, if the generated model does contain geometric combinations, adaptive moderation might still not hold. In that case, also generate the adaptive additive model in X and Z , that is, only considering separate transforms of X and Z without geometric combinations. Adaptive moderation holds if this adaptive additive model generates a substantial percent decrease in the LCV score compared to the model allowing for geometric combinations (as was the case in the example analyses).

5.5. Outcome Transformation

Consideration of power transforms of positive valued continuous outcomes as well as of predictors can generate distinctly better models and can resolve model assumption problems. For example, variances for untransformed effort for partnered mothers were distinctly nonconstant, and the associated adaptive model generated skewed standardized residuals with two extreme outliers. On the other hand, modeling mean effort^{1.5} as a function of bounded CFS generated a model (Figure 3) that provided a substantial improvement over not transforming effort with variances reasonably treated as constant, without outliers, with reduced skew, and with a reasonably linear normal plot (Figure 4). However, this might not always be the case. The need for nonconstant variances is not a problem because variances can be appropriately modeled using adaptive methods. If there are still outliers after outcome transformation, sensitivity analyses can be conducted removing outliers systematically to assess whether the outliers are highly influential on what models for the means and/or variances are generated.

High levels of skewness may not be resolved through power transforms of a positive continuous outcome. In such cases, the outcome can be categorized into ordinal levels. Conventional categorizations of the outcome can be used if available. If not, percentile splits could be used. For example, the outcome could be categorized based on a median split and then modeled using adaptive logistic regression. Alternatively, the outcome could be categorized based on a tertile or quartile split and modeled using adaptive ordinal regression.

5.6. Testing for a Bivariate Relationship

Suppose a relationship has been hypothesized between an outcome Y and a predictor X . The conventional approach for assessing this involves testing for a significantly nonzero slope for X using a bivariate regression model (or equivalently for a zero correlation). If this test is nonsignificant, a nonlinear relationship still might hold between Y and X . This can be assessed adaptively using an expansion of a constant base model restricted to at most one transform of X . If the expansion leaves the constant model unchanged, then a nonlinear relationship does not hold between Y and X . However, if the generated model does contain a transform of X , the heuristics of the expansion do not guarantee that this relationship would be significant. This could be assessed with a standard test for zero slope for transformed X . Even if this test is significant, the relationship might not be substantial, which could be addressed using a LCV ratio test comparing the model with transformed X to the constant model. Equivalently, the expanded model could be subjected to a contraction holding the intercept fixed in the model. The relationship is substantial if the contraction does not remove transformed X from the model.

5.7. Applicability

Linear relationships are typically assumed or implied in theories used in the be-

havioral, health, and social sciences. However, a variety of exceptions within the behavioral science context are discussed by Hayes and Preacher [21]. In any case, nonlinearity can hold even when theories hypothesize exact linear relationships.

The reported analyses used data from the family psychology literature, but adaptive methods apply more generally to all areas providing deeper insights into relationships between arbitrary outcomes and their continuous predictors. Adaptive methods could be used to conduct purely exploratory analyses similar to those reported here to identify nonlinear dependence of outcomes on individual transformed predictors and/or on multiple transformed predictors in combination and possibly interacting (through geometric combinations).

However, they can also be used to supplement standard analyses based on theoretical considerations. For example, suppose a pretest/posttest experiment was conducted with participants randomized to either an intervention to reduce depressive symptoms (or to affect any other appropriate outcome) compared to a control condition. Suppose that the pre-specified approach to assess the efficacy of the intervention was to model mean post-baseline depressive symptoms in terms of the indicator for being in the intervention group while controlling for baseline depressive symptoms. If the intervention group indicator has a significant effect in this theory-based model supporting the efficacy of the intervention, the next natural issue to address is how that conclusion is affected by controlling for covariates. If one includes all available covariates in the model and the intervention group effect becomes nonsignificant, this could be a meaningful conclusion or alternatively a consequence of overfitting the model with too many nonsignificant covariate terms.

This issue could be addressed adaptively as follows. Conduct an adaptive analysis starting from the model for the means based on the intervention group indicator by itself, expand the model considering baseline depressive symptoms and all available covariates as primary predictors, then contract the expanded model adjusting powers of remaining transforms but restricting the contraction not to remove the intervention group indicator (or also the intercept). Because that indicator was included in the model based on theory, it would be reasonable to use a standard test for zero slope to assess whether the intervention effect still held after accounting adaptively for transforms of the baseline value and the covariates. This kind of modeling could be considered semi-adaptive.

If a quasi-experimental or comparative design was used instead with two nonequivalent groups, the issue of covariates becomes even more important. One possibility would be to conduct an adaptive logistic regression analysis modeling treatment group membership in terms of available covariates and use the generated adaptive model to compute propensity score weights or scores for models assessing the efficacy of the intervention.

Suppose that depressive symptoms (or any other appropriate outcome) were collected at baseline and also longitudinally at multiple post-baseline times and that the efficacy of the intervention was supported by a significant group-by-

time interaction using a standard repeated measures analysis. A parsimonious model of how mean depressive symptoms changes over time separately within each of the two groups can be addressed using adaptive linear mixed modeling allowing for temporal correlation. Conduct an adaptive moderation analysis with time and the intervention group indicator as primary predictors while allowing for geometric combinations in these two predictors to account for different nonlinear trajectories for the two groups. In this case, geometric combinations would be the same as power transforms of the interaction since indicator variables are unaffected by power transformation.

Any standard model used to establish the effect of an intervention on a continuous outcome might have nonconstant variances or the outcomes may be skewed, making the conclusion that the intervention was efficacious questionable. This could be assessed through adaptive modeling allowing for nonconstant variances and/or outcome transformation. The model for the mean could be left unchanged so it would be theory-based or it could be adjusted as above for covariates. For discrete or count outcomes, the need for nonunit dispersions could be assessed using adaptive modeling comparing the model allowing for nonunit dispersion to the associated unit dispersions model, once again holding a theory-based model for the means fixed or adjusting it for covariates.

5.8. Comparison to Standard Fractional Polynomial Modeling

The example analyses demonstrate that the range of the recommended powers can be too restrictive when the appropriate power is outside that range. This is further supported by a simulation provided by Knafl and Ding [10] (Section 2.12). A simulation was also provided here that demonstrates this can also happen when the true power is strictly within the range of recommended powers. On the other hand, the recommended set of powers will often generate a competitive choice for actual data sets compared to the associated adaptive model. When the selected recommended power has an extreme value of -2 or 3 , it would be prudent to expand the search through more extreme powers. In any case, when the data have little variability, it would also be prudent to expand the search over powers nearby the selected recommended power.

The advantage of standard fractional polynomial modeling [2] [3] is that it is supported by other statistical software, not just SAS as is adaptive modeling. The advantages of adaptive modeling include use of LCV ratio tests rather than standard likelihood ratio tests, consideration of general real valued powers not just fixed sets, allowing for the more general case of both nonzero and zero intercept models, ability to model variances/dispersions as well as means, support for adaptive moderation analyses including geometric combinations generalizing power transforms of interactions, and support for multivariate outcomes not just univariate outcomes. On the other hand, standard fractional polynomial modeling has been extended to handle time-to-event outcomes while adaptive modeling of such outcomes is currently under development.

5.9. Limitations

Adaptive modeling is not directly supported in standard statistical software. However, SAS macros have been developed for conducting adaptive modeling as described in this paper. Code for conducting the analyses reported here including these macros are available at <http://www.unc.edu/~gknafl/AFPM2.html> (accessed February 4, 2018). The data used in these analyses are also available at this URL. Example code for conducting analyses like those reported here is provided above. More extensive analyses of this kind and associated code are provided in Chapters 2 - 7 of [10]. Data and code for those analyses as well as for analyses reported in other chapters are at <http://www.unc.edu/~gknafl/AdaptReg.html> (accessed February 4, 2018). Mediation [12] has not been addressed here. That requires a more complex approach based on adaptive path modeling [22].

Models based on arbitrary power transforms can be difficult to interpret, especially geometric combinations based on multiple powers (as in the example analyses). Selected powers have little meaning since competitive models could be generated by replacing those selected powers by many alternative, nearby powers. Adaptive models can only be fully understood through visualization of estimated relationships. **Figures 1-3** give examples of visualizing models based on single predictor transforms. **Figure 5** provides an example of visualizing a model based on two predictor transforms (*i.e.*, plot the mean outcome as a function of one of the predictors for selected values of the second predictor). Three predictor transforms could be handled by several plots like **Figure 5**, visualizing the relationship in two of the transforms at selected values of the third transform. Models based on more than three transforms would be difficult to visualize. Adaptive modeling can be constrained to include at most three transforms in order to guarantee that the generated model could be visualized.

The possibility of overfitting is of concern for all regression models. The use of LCV scores reduces the possibility of overfitting by an adaptively selected model. Overfitting results when a model relies too heavily on small exceptional subsets of the data (e.g., outliers), but such subsets are likely to fall mostly, if not entirely, in one fold. While such a model will fit that fold very well, it is unlikely to fit the other folds well, thereby generating an inferior LCV score compared to models not overfitting the data. However, the possibility of overfitting should still be assessed. For example, **Figure 3** is the plot for the final selected model for mean effort as a function of CFS for partnered mothers. The smallest observed value for CFS is 42.9, more than 7 units less than the next smallest observed value of 50. However, **Figure 3** indicates that neither of these two data points affects estimated mean effort values over the associated range of CFS values very much. These estimates are controlled more by the next smallest observed CFS value of 53.6 with five associated observations. Had the estimated curve at CFS = 42.9 been very close to the associated observed effort value of 20, the model could have overfit the relationship at the low end of CFS values, especially if its slope changed quite a bit soon after that.

Other methods that are directly supported in standard statistical software could have been used instead, including classification and regression trees (CART) [23], generalized additive models (GAMs) [24], and MARS models. Fractional polynomial modeling is likely to generate smoother curves than these other methods due to the differentiability of power transforms. However, these other methods do not address nonconstant variances or outcome transformation. Knafl and Ding [10] provided comparisons of adaptive modeling to the use of GAMs and MARS, demonstrating that adaptive modeling can generate improved LCV scores using more parsimonious models and sometimes distinctly better models. In none of those example analyses did GAMs outperform adaptive modeling. However, in one exceptional example analysis, MARS modeling outperformed adaptive modeling, but that MARS model could be improved through adaptive adjustments (*i.e.*, by power transforming MARS generated splines).

6. Summary

Adaptive fractional polynomial modeling has been described and demonstrated. Reported analyses demonstrate the kinds of novel insights into the data that are possible with adaptive modeling. Specifically, fractional polynomials can outperform standard polynomials and consideration of only standard polynomials can lead to the incorrect conclusion that a relationship does not hold when in fact a nonlinear relationship exists. Zero intercept models are important to consider because the associated models are simpler and relationships in some cases can only be supported allowing for a zero intercept. While the recommended Royston and Altman powers will often generate effective models, there are cases where the recommended powers are inadequate to effectively model the data. This had only been demonstrated for powers outside the range of the recommended powers before using simulated data, not actual data as demonstrated in the example analyses. Furthermore, the example simulation analyses demonstrate for the first time that this also holds when the true power is strictly within the range of the recommended powers. Allowing for nonconstant variances can provide distinct improvements over making the usual constant variances assumption and provides an objective way to assess that assumption as opposed to subjective inspection of residual plots. Power transforms of positive valued continuous outcomes can provide distinct improvements over modeling untransformed outcomes. Basing the choice on power-adjusted LCV scores provides an objective way to choose an appropriate outcome power as opposed to subjectively inspecting scatter plots for selected power-transformed outcomes. Nonlinear moderation relationships can hold when linear moderation does not hold. Moreover, this can sometimes only be identified using general geometric combinations rather than standard interactions. The reported analyses provide the first known example of this.

In summary, adaptive modeling can provide novel insights into data com-

pared to standard regression modeling. While only analyses of univariate continuous outcomes have been addressed here, there can be similar benefits to adaptive modeling of univariate discrete and count outcomes (for examples see [17] [18] [19]) as well as multivariate continuous, discrete, and count outcomes (for examples see [10]).

References

- [1] Royston, P. and Altman, D.G. (1994) Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modeling. *Applied Statistics*, **43**, 429-467. <https://doi.org/10.2307/2986270>
- [2] Royston, P. and Sauerbrei, W. (2008) *Multivariable Model-Building: A Practical Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables*. John Wiley & Sons, Hoboken. <https://doi.org/10.1002/9780470770771>
- [3] Zhang, Z. (2016) Multivariable Fractional Polynomial Method for Regression Model. *Annals of Translational Medicine*, **4**, 174. <https://doi.org/10.21037/atm.2016.05.01>
- [4] Box, G.E.P. and Tidwell, P.W. (1962) Transformation of Independent Variables. *Technometrics*, **4**, 531-550. <https://doi.org/10.1080/00401706.1962.10490038>
- [5] Box, G.E.P. and Cox, D.R. (1964) An Analysis of Transformations. *Journal of the Royal Statistical Society, Series B*, **26**, 211-252.
- [6] Carroll, R.J. and Ruppert, D. (1984) Power Transformations When Fitting Theoretical Models to Data. *Journal of the American Statistical Association*, **79**, 321-328. <https://doi.org/10.1080/01621459.1984.10478052>
- [7] Knafl, G.J., Fennie, K.P., Bova, C., Dieckhaus, K. and Williams, A.B. (2004) Electronic Monitoring Device Event Modeling on an Individual-Subject Basis Using Adaptive Poisson Regression. *Statistics in Medicine*, **23**, 783-801. <https://doi.org/10.1002/sim.1624>
- [8] Knafl, G.J., Delucchi, K.L., Bova, C.A., Fennie, K.P., Ding, K. and Williams, A.B. (2010) A Systematic Approach for Analyzing Electronically Monitored Adherence Data. In: Ekwall, B. and Cronquist, M., Eds., *Micro Electro Mechanical Systems (MEMS) Technology, Fabrication Processes and Applications*, Nova Science Publishers, Hauppauge, NY, 1-66.
- [9] Knafl, G.J., Fennie, K.P., and O'Malley, J.P. (2006) Adaptive Repeated Measures Modeling Using Likelihood Cross-Validation. In: Boveruchuk, B., Ed., *Proceedings Second IASTED International Conference on Computational Intelligence*, ACTA Press, Anaheim, CA, 422-427.
- [10] Knafl, G.J. and Ding, K. (2016) *Adaptive Regression for Modeling Nonlinear Relationships*. Springer International Publishing, Switzerland. <https://doi.org/10.1007/978-3-319-33946-7>
- [11] Knafl, K., Deatrick, J.A., Gallo, A., Dixon, J.K., Grey, M., Knafl, G.J., *et al.* (2011) Assessment of the Psychometric Properties of the Family Management Measure. *Journal of Pediatric Psychology*, **36**, 494-505. <https://doi.org/10.1093/jpepsy/jsp034>
- [12] Baron, R.M. and Kenny, D.A. (1986) The Moderator-Mediator Variable Distinction in Social Psychology Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality & Social Psychology*, **51**, 1173-1182. <https://doi.org/10.1037/0022-3514.51.6.1173>
- [13] McCullagh, P. and Nelder, J.A. (1999) *Generalized Linear Models*, 2nd Edition,

Chapman & Hall/CRC, Boca Raton.

- [14] Stein, R.E. and Jessop, D.J. (1990) Functional Status II(R): A Measure of Child Health Status. *Medical Care*, **28**, 1041-1055.
<https://doi.org/10.1097/00005650-199011000-00006>
- [15] Aiken, L. S. and West, S. G. (1991) Multiple Regression: Testing and Interpreting Interactions. Sage Publications, Inc, Thousand Oaks.
- [16] Friedman, J.H. (1991) Multivariate Adaptive Regression Splines. *Annals of Statistics*, **19**, 1-67. <https://doi.org/10.1214/aos/1176347963>
- [17] Riegel, B. and Knafl, G.J. (2014) Electronically Monitored Medication Adherence Predicts Hospitalization in Heart Failure Patients. *Patient Preference and Adherence*, **8**, 1-13. <https://doi.org/10.2147/PPA.S54520>
- [18] Knafl, G.J. and Riegel, B. (2014) What Puts Heart Failure Patients at Risk for Poor Medication Adherence? *Patient Preference and Adherence*, **8**, 1007-1018.
<https://doi.org/10.2147/PPA.S64593>
- [19] Meghani, S.H. and Knafl, G.J. (2016) Patterns of Analgesic Adherence Predict Health Care Utilization among Outpatients with Cancer Pain. *Patient Preference & Adherence*, **10**, 81-98. <https://doi.org/10.2147/PPA.S93726>
- [20] Claeskens, G. and Hjort, N.L. (2009) Model Selection and Model Averaging. Cambridge University Press, Cambridge.
- [21] Hayes, A.F. and Preacher, K.J. (2010) Quantifying and Testing Indirect Effects in Simple Mediation Models When the Constituent Paths Are Nonlinear. *Multivariate Behavioral Research*, **45**, 627-660. <https://doi.org/10.1080/00273171.2010.498290>
- [22] Knafl, G.J., Knafl, K.A., Grey, M., Dixon, J., Deatrick, J.A. and Gallo, A. (2017) Incorporating Nonlinearity into Mediation Analyses. *BMC Medical Research Methodology*, **17**, 45. <https://doi.org/10.1186/s12874-017-0296-6>
- [23] Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1998) Classification and Regression Trees. CRC Press, Boca Raton.
- [24] Hastie, T.J. and Tibshirani, R.J. (1999) Generalized Additive Models. Chapman & Hall/CRC, Boca Raton.