The Stata Journal

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go "beyond the Stata manual" in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The *Stata Journal* is indexed and abstracted by *CompuMath Citation Index*, *Current Contents/Social and Behavioral Sciences*, *RePEc: Research Papers in Economics*, *Science Citation Index Expanded* (also known as *SciSearch*), *Scopus*, and *Social Sciences Citation Index*.

For more information on the *Stata Journal*, including information for authors, see the webpage

http://www.stata-journal.com

Subscriptions are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-STATA-PC, fax 979-696-4601, or online at

http://www.stata.com/bookstore/sj.html

Subscription rates listed below include both a printed and an electronic copy unless otherwise mentioned.

U.S. and Canada		Elsewhere	
Printed & electronic		**Printed & electronic**	
1-year subscription	$115	1-year subscription	$145
2-year subscription	$210	2-year subscription	$270
3-year subscription	$285	3-year subscription	$375
1-year student subscription	$ 85	1-year student subscription	$115
1-year institutional subscription	$345	1-year institutional subscription	$375
2-year institutional subscription	$625	2-year institutional subscription	$685
3-year institutional subscription	$875	3-year institutional subscription	$965
Electronic only		**Electronic only**	
1-year subscription	$ 85	1-year subscription	$ 85
2-year subscription	$155	2-year subscription	$155
3-year subscription	$215	3-year subscription	$215
1-year student subscription	$ 55	1-year student subscription	$ 55

Back issues of the *Stata Journal* may be ordered online at

http://www.stata.com/bookstore/sjj.html

Individual articles three or more years old may be accessed online without charge. More recent articles may be ordered online.

http://www.stata-journal.com/archives.html

The *Stata Journal* is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the *Stata Journal*, StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA, or emailed to sj@stata.com.


The *Stata Journal* (ISSN 1536-867X) is a publication of Stata Press. Stata, STATA, Stata Press, Mata, MATA, and NetCourse are registered trademarks of StataCorp LP.

The Stata Journal (2016)
16, Number 1, pp. 72–87

mfpa: Extension of mfp using the ACD covariate transformation for enhanced parametric multivariable modeling

Patrick Royston
MRC Clinical Trials Unit
University College London
London, UK
j.royston@ucl.ac.uk

Willi Sauerbrei
Center for Medical Biometry and Medical Informatics
Medical Center—University of Freiburg
Freiburg, Germany
wfs@imbi.uni-freiburg.de

Abstract. In a recent article, Royston (2015, *Stata Journal* 15: 275–291) introduced the approximate cumulative distribution (ACD) transformation of a continuous covariate x as a route toward modeling a sigmoid relationship between x and an outcome variable. In this article, we extend the approach to multivariable modeling by modifying the standard Stata program `mfp`. The result is a new program, `mfpa`, that has all the features of `mfp` plus the ability to fit a new model for user-selected covariates that we call FP1(p_1, p_2). The FP1(p_1, p_2) model comprises the best-fitting combination of a dimension-one fractional polynomial (FP1) function of x and an FP1 function of ACD (x). We describe a new model-selection algorithm called function-selection procedure with ACD transformation, which uses significance testing to attempt to simplify an FP1(p_1, p_2) model to a submodel, an FP1 or linear model in x or in ACD (x). The function-selection procedure with ACD transformation is related in concept to the FSP (FP function-selection procedure), which is an integral part of `mfp` and which is used to simplify a dimension-two (FP2) function. We describe the `mfpa` command and give univariable and multivariable examples with real data to demonstrate its use.

Keywords: st0425, mfpa, mfp, continuous covariates, sigmoid function, ACD transformation, multivariable fractional polynomials, regression models

1 Introduction

Over the years, fractional polynomials (FPs) have steadily gained popularity as a tool for flexible parametric modeling of regression relationships. A recent search in Google Scholar (22 February 2016) yielded 1,181 citations of the original article by Royston and Altman (1994). The multivariable fractional polynomials (MFP) method of multiple regression modeling (Sauerbrei and Royston 1999) simultaneously removes weakly influential predictors and determines a suitable functional form (FP or linear) for continuous predictors. MFP is implemented as the `mfp` command in Stata. Its appeal may lie in a combination of relative simplicity and familiarity (an extension of conventional polynomials) with added flexibility for representing nonlinear functional forms and usually a low probability of introducing uninterpretable artifacts into the fitted functions. Furthermore, unlike splines—which have only a local interpretation of the fitted function (piecewise between knots)—FPs provide a curve with a global interpretation.

MFP extends backward elimination by systematically searching for improvement in fit by modeling possible nonlinearity in the effects of continuous variables. The heart of MFP lies in modeling each continuous predictor using FP functions combined with a principled function-selection procedure (FSP) to yield a simplified functional form, if appropriate. Each predictor is modeled univariately by this method, adjusted for the other predictors, within an overarching back-fitting algorithm that visits each predictor in turn.

Royston (2015) described an extension of univariate FP modeling via the so-called approximate cumulative distribution (ACD) covariate transformation. The ACD transformation is a smooth function that maps a continuous covariate, x, to an approximation, $\text{ACD}(x)$, of its distribution function. By construction, the distribution of $\text{ACD}(x)$ in the sample is roughly uniform on $(0, 1)$. FP modeling is then performed with the transformed values $\text{ACD}(x)$ instead of x as a predictor. Royston (2015) showed that such an approach could successfully represent a sigmoid function of x, something a standard FP function cannot do (Royston and Sauerbrei 2008, sec. 5.8.1). He went on to demonstrate that useful flexibility in functional form could be achieved by considering both x and $a = \text{ACD}(x)$ simultaneously as independent predictors and applying the MFP algorithm to x and a. To limit instability and overfitting, he suggested restricting the models considered for x and a to FP1 functions. Royston (2015) also noted that models based on $\text{ACD}(x)$ may have other advantages in terms of interpretability of regression coefficients and resistance to the potential influence of extreme covariate observations.

In the present article, we take the modeling process further. We show how to select optimal FP1 functions for x and $\text{ACD}(x)$ in a univariable context. We describe a modified version of the FP FSP adapted to the x and $\text{ACD}(x)$ approach. We then modify the MFP algorithm to produce a new but closely related algorithm called MFPA, in which the FP FSP is replaced by the modified version (FSP with ACD transformation [FSPA]) just mentioned. MFPA may help with situations in which a sigmoid function is needed, which MFP cannot provide. Also, as mentioned, MFPA may reduce the influence of extreme covariate values on a selected function.

The structure of the article is as follows. Section 2 describes how to select a univariable model based on applying the FSPA to combinations of x and $\text{ACD}(x)$. Section 3 introduces MFPA as a modification of MFP. Section 4 gives examples of applying MFP and MFPA to two real datasets. Section 5 describes `mfpa`, a new command that extends the standard `mfp` command by allowing the FSPA instead of the FSP to be applied to one or more of the candidate continuous predictors. Additionally, `mfpa` supports Stata's factor variables. Section 6 contains some final remarks.

2 Choosing a suitable function

In this section, we propose a method to select a univariable model. We consider estimation with a single continuous predictor, x, combined with the preliminary transformation $a = \text{ACD}(x)$. In section 3, we describe how the selected function can be used in an iterative multivariable modeling procedure, MFPA, that is closely related to MFP. We first define the ACD transformation.

2.1 The ACD transformation

Let X be a continuous random variable to be considered as a covariate in some kind of regression model. We wish to approximate the empirical cumulative distribution function of a random sample $x_1, \ldots, x_n$ of n observations from the distribution of X. We define the ACD $(\cdot)$ transformation in several steps as follows. Let $\text{rank}(x_i)$ be the rank of x_i, with ranks 1 and n denoting the lowest and highest sample values, respectively. Define

$$\begin{aligned} z_i &= \Phi^{-1}\left\{\left(\text{rank}\left(x_i\right) - 0.5\right)/n\right\} \\ E\left(z_i\right) &= \beta_0 + \beta_1\left(x_i + s\right)^p \\ \widehat{z}_i &= \widehat{E\left(z_i\right)} = \widehat{\beta}_0 + \widehat{\beta}_1\left(x_i + s\right)^{\widehat{p}} \\ \text{ACD}\left(x_i\right) &= a_i = \Phi\left(\widehat{z}_i\right) \end{aligned}$$

where $\Phi\left(\cdot\right)$ is the standard normal cumulative distribution function (`normal()` in Stata), $\Phi^{-1}\left(\cdot\right)$ is its inverse (`invnormal()` in Stata), and $\widehat{p}$ is the best-fitting estimate of p in an FP1 regression model $E\left(z_i\right) = \beta_0 + \beta_1\left(x_i + s\right)^p$. Powers p are selected from the set $S = \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$. Ordinary least-squares regression of the z_i on the values $\left(x_i + s\right)^p$ is used to estimate the parameters β_0, β_1, and p, with $p = 0$ meaning log transformation. If any $x_i \leq 0$, then all the x_i are shifted by a constant, s, chosen to ensure that $\left(x_i + s\right) > 0$ for all i; if all $x_i > 0$, then $s = 0$. See, for example, Royston and Sauerbrei (2008, 84–85) for details of how s may be determined. In the following, we assume that $x_i > 0$ and $s = 0$ so that s can be ignored in the formulation.

An explanation of the rationale for the above approach is given in the section "The ACD transformation" in Royston (2015). Depictions of $\text{ACD}\left(x_i\right)$ when X has a normal or lognormal distribution are given in figure 1 in the section "Example 1: Simulated distributions" of Royston (2015).

2.2 The model FP1(p_1, p_2) and some submodels

In an example analysis of the prognostic importance of tumor thickness in malignant melanoma (Baade et al. 2015), Royston (2015) demonstrated that applying MFP to select FP1 functions of x = tumor thickness and of $a = \text{ACD}\left(x\right)$ simultaneously could give rise to a well-fitting function that a standard FP1 or FP2 function in x or in a could not match. The chosen function had a linear component in x and an FP1 component in a, with the latter being a sigmoid function of x. The result hinted that models comprising FP functions of x and a might be of value in particular cases as an alternative to the standard FP class.

In this section, we take the idea further and consider a four-parameter model class, $\beta_1 x^{p_1} + \beta_2 a^{p_2}$, called $\text{FP1}(p_1, p_2)$ and based on FP1 transformations of x and a. The aim is to adapt to $\text{FP1}(p_1, p_2)$ the FSP that, starting with the FP2 class, is used to determine a parsimonious FP function of x. Function selection needs to be done in a systematic and principled way. We address function selection in section 2.4.

First, we consider six models, M1–M6, each of which represents the best-fitting model within its respective class. They are potentially useful in deriving a more parsimonious "final" model, aiming to reduce the risk of overfitting the most complex allowed function, M1 = FP1(p_1, p_2). M2–M6 are submodels of M1. The models are listed in table 1.

Table 1. Six submodels of FP1(p_1, p_2). A dot (.) indicates that the corresponding term is omitted.

Model	Notation	Function	Comment
M1	FP1(p_1, p_2)	$\beta_1 x^{p_1} + \beta_2 a^{p_2}$	The most complex allowed function
M2	FP1$(p_1, .)$	$\beta_1 x^{p_1}$	Standard FP1 function of x
M3	FP1$(., p_2)$	$\beta_2 a^{p_2}$	Usually a singly or doubly asymptotic curve in x
M4	FP1$(1, .)$	$\beta_1 x$	Linear reduction of model M2
M5	FP1$(., 1)$	$\beta_2 a$	Linear reduction of model M3
M6	FP1$(., .)$	–	Null model; x is omitted altogether

The models have been chosen to provide two nesting hierarchies that can be applied for model reduction: M1 $\supset$ M2 $\supset$ M4 $\supset$ M6 and M1 $\supset$ M3 $\supset$ M5 $\supset$ M6. For example, M1 $\supset$ M2 means that M2 is nested in M1. These hierarchies are used to provide sets of nested models for use in function selection (see section 2.4).

Plots of some of the functional forms available with models M1, M3, and M5 may be seen in several of the figures in Royston (2015). Next, we consider estimation of the parameters of M1–M6.

2.3 Estimation

Models M2–M5 are conventional FP1 or linear models in x or in a. In univariable settings, M6 is simply a constant. Powers p_1 or p_2 in M2 and M3 are estimated in the usual way by finding the corresponding values that maximize the likelihood in the set of power transformations S.

To estimate p_1 and p_2 in M1, one might consider applying MFP (with maximum allowed complexity FP1 functions) to x and a, treating them as though they were independent variables. However, because of the high collinearity of x and a, the approach may produce a suboptimal fit; it does not always find the best values of p_1 and p_2. Instead, we systematically search all $8 \times 8 = 64$ possible pairs (p_1, p_2) for the maximum likelihood solution by fitting each of the FP1 models and finding the pair giving the highest likelihood.

When p_1 and p_2 have been determined for M1–M5, models M1, M2, and M3 are conditionally linear and β_1 and β_2 are estimated by maximum likelihood in standard fashion.

2.4 Function-selection procedure FSPA

To select a suitable model among M1–M6 above, we need a systematic model-selection procedure akin to the FSP. Full details of the FSP are given by Royston and Sauerbrei (2008, 82–84). In summary, the FSP has three steps with the following characteristics:

1. The FSP is a closed test procedure that maintains the preselected nominal significance level (α_1) for testing whether x is influential. The first test (FP2 versus null) achieves this. If FP2 is not a significantly better fit than null, then x is dropped and the procedure ends. Note that α_1 is set by `mfp`'s option `select(#)`, whose default value is 1, meaning that x is automatically selected and the procedure continues to the function-selection stage. The α_1 significance level is of course much more relevant to multivariable modeling than in the present context of function selection for a single x.

2. Assuming x is deemed influential after the first test, the FSP is also a closed test procedure that maintains a second preselected nominal significance level (α_2) for testing whether the functional form of the relation between x and the outcome is nonlinear. The second test (FP2 versus linear) achieves this. If FP2 is not a significantly better fit than linear, then a linear function of x is selected and the procedure ends. Often, the significance levels α_1 and α_2 are taken as equal. Note that α_2 is set by `mfp`'s option `alpha(#)`; the default is `alpha(0.05)`.

3. If nonlinearity is found at the second step, a final test (FP2 versus FP1), also at the α_2 level, is applied to refine the selected function further. The procedure ends, selecting either an FP1 or an FP2 function.

Allowing ACD transformation, we can reproduce the main features of the FSP starting with FP1(p_1, p_2) as the most complex permitted function. We call the modified procedure the FSPA. To enable testing, deviances ($-2 \times$ log likelihood) for each of M1–M6 are first obtained, requiring $64(\text{M1}) + 8(\text{M2}) + 8(\text{M3}) + 1(\text{M6}) = 81$ distinct model fits. (Models M4 and M5 are already fit as special cases of FP1 models M2 and M3, respectively.) The FSPA then runs as follows.

1. Step 1 is identical to step 1 of the FSP except that M1 is tested against M6 (on 4 degrees of freedom [d.f.]). This provides a closed test at the α_1 level for x being influential. If the test is nonsignificant, then drop x and end. Otherwise, continue to step 2.

2. Step 2 is identical to step 2 of the FSP except that M1 is tested against M4 (on 3 d.f.). This provides a closed test at the α_2 level for the functional form for x being nonlinear. If the test is nonsignificant, then accept a linear function for x and end. Otherwise, continue to step 3.

3. Step 3 is similar to step 3 of the FSP except that M1 is tested against M2 (on 2 d.f.) and the procedure may continue. If the test is nonsignificant at the α_2 level, then accept M2 and end. Otherwise, continue to step 4.

4. We now know that M1 is a significantly better fit than M2. However, it may be possible to simplify M1 in the direction of the ACD model M3; therefore, M1 is tested against M3 (on 2 d.f.). If the test is significant at the α_2 level, then accept M1 and end. Otherwise, continue to step 5.

5. Finally, M3 is tested against M5 (on 1 d.f.). If the test is significant at the α_2 level, then accept M3 and end. Otherwise, accept M5 and end.

With the FSPA, depending on the choices of α_1 and α_2, we may obtain any of models M1–M6 as "final". The ordered sequence of steps comprising the FSPA is designed to select a linear or FP1 model if the fit of one of them is sufficient. Only if M1 is better than both M4 and M2 do M3 and M5 (ACD-based models) come into play. Thus the FSPA favors FP1 or linear functions in the sense that it will consider an ACD-based model only if a standard FP1 or linear model fails to fit as well as M1 does. The approach follows the philosophy of MFP that an explanatory model should be as simple as possible and that increased complexity should be adequately supported by an improved fit to the data.

3 The MFP and MFPA algorithms

At each step of the MFP algorithm, the FSP is applied to each continuous covariate in turn to decide whether it is sufficiently influential (that is, significant at the α_1 level) to remain in the model, and if so, to estimate its functional form (usually an FP2, FP1, or linear function). Categorical variables are also tested for inclusion in standard fashion. The models fit at each step are adjusted for all other currently selected candidate variables, whether continuous or categorical, retaining any FP or linear functions if those have been selected so far. A cycle is defined as a complete tour, in a specified order, of all the candidate variables. The algorithm terminates when the selected functions or categorical variables do not change from one cycle to the next. Typically, MFP converges in about 2–4 cycles. Theoretically, MFP can oscillate between two different solutions, but in practice such behavior is extremely rare. In section 6.3.2 of Royston and Sauerbrei (2008), we illustrate further details of the algorithm in an example.

The MFPA algorithm is identical to MFP except that the FSP is replaced with the FSPA for any continuous variable(s) that the user wishes to assess using the ACD approach. It is possible to specify ACD and hence the FSPA for any subset of the continuous predictors. In the `mfpa` program (described below in section 5), specifying which variables are to be modeled with FP1(p_1, p_2) as the most complex permitted function of an x and the corresponding a is done through the `acd()` option.

4 Examples

4.1 Example 1: A function with an asymptote

We use the well-known German breast cancer dataset (Schumacher et al. 1994), which can be loaded into Stata via the command `webuse brcancer`. The data are prepared for survival analysis using the command `stset rectime, failure(censrec)`.

We compare five functions selected for the effect of the strongest predictor (`x5` = number of positive lymph nodes) in univariate Cox regression models, all adjusted for hormonal therapy (`hormon`). The models we consider for `x5` are as follows:

1. FP2(p_1, p_2) for which the FSP selects $(p_1, p_2) = (-2, -1)$ (that is, a quadratic function in $\texttt{x5}^{-1}$).
2. A negative exponential model, that is, a linear function of $\exp(-0.12 \times \texttt{x5})$, as suggested by Sauerbrei and Royston (1999).
3. FP1(p_1, p_2), that is, model M1 without simplification, for which the maximum likelihood estimate is $(p_1, p_2) = (-0.5, -2)$.
4. FP1(p_1, p_2) with model simplification with the FSPA using $\alpha_1 = \alpha_2 = 0.05$, for which the selected powers are $(p_1, p_2) = (., 3)$ (an instance of model M3; see table 1).
5. A restricted cubic regression spline with 4 d.f. (Royston and Sauerbrei 2007b).

The fitted curves, depicting log relative-hazards fit by the Cox model, are shown in figure 1.

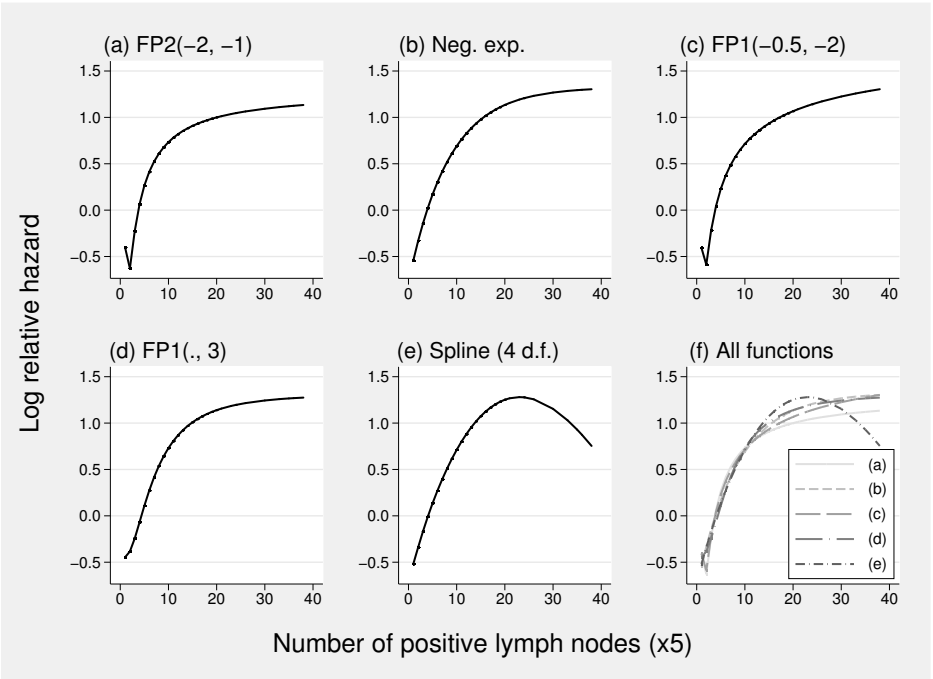

Figure 1. Five fitted functions for `x5` in the German breast cancer dataset. Graph (f) compares the functions shown individually in graphs (a) through (e).

The FP2$(-2,-1)$, FP1$(-0.5,-2)$, and spline curves are all nonmonotonic, with the spline curve exhibiting a maximum log relative-hazard at about 25 positive lymph nodes. Such nonmonotonicity is implausible for biologic reasons, because more positive nodes should mean a higher risk of cancer recurrence. The negative exponential and FP1$(.,3)$ curves are closely similar and are by construction both monotonic. Thus, the FSPA provides a "good" model for `x5` within the ACD-extended FP class without resorting to special nonlinear functions such as the negative exponential transformation in figure 1(b). The FP2 function fits the data best, but the local minimum at two nodes conflicts with medical knowledge and is probably a result of overfitting the data. Sauerbrei and Royston (1999) therefore introduced the negative exponential transformation as a possible pretransformation to provide a monotonic function.

As an illustration of the workings of the FSPA, table 2 shows the results of the various tests on the deviances (minus twice the log partial likelihoods) for the six models M1–M6 for `x5`, adjusted for `hormon`.

Table 2. Models (table 2a) and accompanying tests (table 2b) comprising the FSPA when applied to `x5` (number of positive lymph nodes) in the German breast cancer data. All models are Cox regression, adjusted for `hormon`.

Table 2a.

Model		
Code	Description	Deviance
M1	FP1$(-0.5,-2)$	3483.88
M2	FP1$(0.5,.)$	3493.35
M3	FP1$(.,3)$	3486.13
M4	FP1$(1,.)$	3517.88
M5	FP1$(,.1)$	3494.06
M6	FP1$(.,.)$	3567.53

Table 2b.

FSPA model comparisons			
Step	Comparison	Dev. diff.	p-value
1	M1 versus M6	83.65	<0.001
2	M1 versus M4	34.00	<0.001
3	M1 versus M2	9.47	0.009
4	M1 versus M3	2.25	0.3
5	M3 versus M5	7.93	0.005

We see that M1 fits significantly better than all of M6 ($P < 0.001$), M4 ($P < 0.001$), and M2 ($P = 0.009$). At step 4 of the FSPA, the fit of M3 is not significantly worse than that of M1 ($P = 0.3$), leading to provisional acceptance of M3 and to the final comparison at step 5 (M3 versus M5). Because M3 fits significantly better than M5 ($P = 0.005$), M3 is finally selected.

Below, we show the output from `mfpa`, summarized in table 2, when fitting `x5` and `hormon`:

```
. webuse brcancer, clear
(German breast cancer data)
. stset rectime censrec
  (output omitted)
. mfpa, select(0.05) acd(x5): stcox x5 hormon
Deviance for model with all terms untransformed =  3517.881, 686 observations
Variable     Model (vs.)   Deviance  Dev diff.   P      Powers   (vs.)
-------------------------------------------------------------------------
(A)x5        M6     M1     3567.530    83.650  0.000*   .         -.5,-2
             M4            3517.881    34.002  0.000*   1
             M2            3493.355     9.475  0.009*   0
             M3            3486.128     2.248  0.325    3
             M5     M3     3494.056     7.928  0.005*   1         3
             Final (M3)    3486.128                     . 3
hormon       null   lin.   3496.724    10.596  0.001*   .         1
             Final         3486.128                     1
```

```
Fractional polynomial fitting algorithm converged after 1 cycle.
Transformations of covariates:
-> gen double IAx5__1 = Ax5^3-.125 if e(sample)
Final multivariable fractional polynomial model for _t
```

Variable	Initial			Final		
	df	Select	Alpha	Status	df	Powers
(A)x5	4	0.0500	0.0500	in	2	. 3
hormon	1	0.0500	0.0500	in	1	1

(*output omitted*)

The `Deviance` column shows the deviance of each model in the `Model` column. The deviance difference between it and its comparator in the `(vs.)` column is shown in the `Dev diff.` column, with the p-value in the `P` column. An asterisk indicates significance at the `alpha()` level; here, the default setting `alpha(0.05)` was applied. The selected FP powers in the FP1(p_1, p_2) models are shown in the `Powers` and corresponding `(vs.)` columns.

The tests are applied from the top down, as described in section 2.4. As noted, the tests of M6, M4, and M2 versus M1 are all significant. M3 is provisionally selected and then confirmed as the final model by the result of the fifth test. Model M3 has powers $(., 3)$, that is, no term in `x5` and one term comprising the cube of `acd(x5)`.

4.2 Example 2: A multivariable model

We consider the so-called Boston housing dataset, in which the log median house price in the Boston area is to be predicted from 13 housing- or environment-related variables, 12 of which are continuous, in a dataset of size 506. Some of the continuous variables are strongly correlated and some have a rather strange distribution. Difficulties in finding a suitable model have made it a dataset often used for comparing various modeling approaches.

The data were analyzed in some detail by Royston and Sauerbrei (2008, 207–213). The selected MFP model is described in table 9.1 of that work. Ten of the 13 variables were selected as significant at the 5% level; three of these (`crim`, `rm`, and `dis`) required FP2 functions and one (`lstat`) required an FP1 function. The remaining five continuous functions were selected as linear. The only categorical variable (`chas`) was selected. The explained variation (R^2_a), adjusted for model dimension, was 0.827.

On applying `mfpa` to this dataset, we obtain eight predictors significant at the 5% level, all of them continuous. Of these, five have two FP1 powers and three are linear. The adjusted explained variation (R^2_a) is 0.853.

Table 3 describes the selected models. It is interesting that the MFPA model has two fewer predictors, one additional parameter, and a higher explained variation than the MFP model.

Table 3. Boston housing data. Comparison of predictors and functions selected at the 5% nominal significance level by the MFP and MFPA algorithms.

Covariate	MFP	MFPA	Covariate	MFP	MFPA
`crim`	$1, 2$	$0, 0.5$	`dis`	$-2, 1$	$1, -2$
`zn`	out	out	`rad`	linear	linear
`indus`	out	out	`tax`	linear	$0.5, 2$
`chas`*	in	out	`ptratio`	linear	linear
`nox`	linear	linear	`bk`	linear	out
`rm`	$0.5, 0.5$	$-1, 3$	`lstat`	0.5	$0.5, 1$
`age`	out	out	R^2_a	0.827	0.853

*Binary predictor

Figure 2 compares the partial predictors for the nine continuous predictors selected by MFP and MFPA.

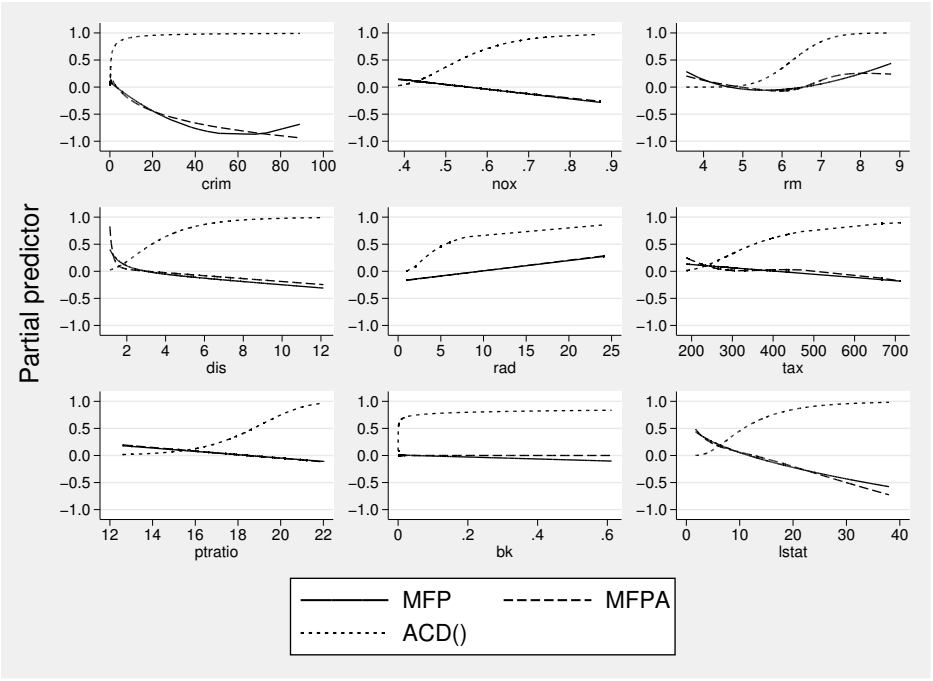

Figure 2. Boston housing data. Fitted partial predictors for the MFP (solid lines) and MFPA (long dashes) models as well as ACD (short dashes) approximations to the cumulative distribution functions of the predictors.

Note that the ACD transformations of two predictors (`crim` and `bk`) are notably skewed in distribution and that the remainder are more symmetrical.

At first glance, the differences between the fitted functions appear rather minor. However, the FP2$(1, 2)$ function for `crim` (level of criminality in the local area) seems inappropriate because it is nonmonotonic, whereas the FP1$(0, 0.5)$ function is nearly monotonic. The two functions for `rm` are both nonmonotonic but are subtly different. MFP selects `bk`, which evidently has a (very) weak effect, whereas MFPA omits it.

In terms of fit, figure 3 shows smoothed residuals for the MFP model.

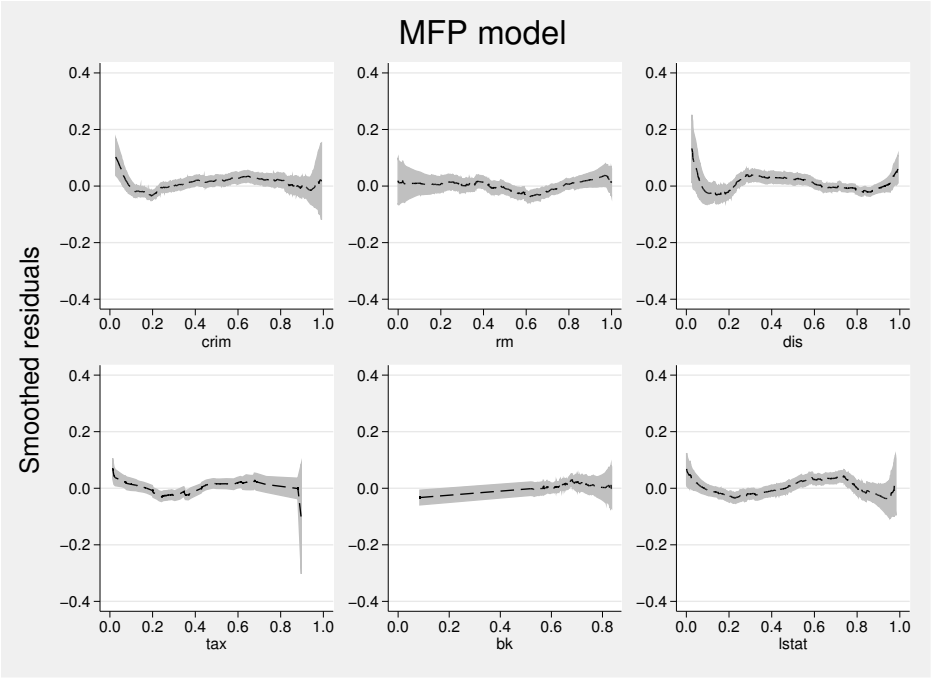

Figure 3. Boston housing data. Smoothed residuals for partial predictors in the MFP model.

Subjectively, some lack of fit is evident for `crim`, `dis`, and perhaps `lstat`.

Figure 4 shows smoothed residuals for the same predictors in the MFPA model.

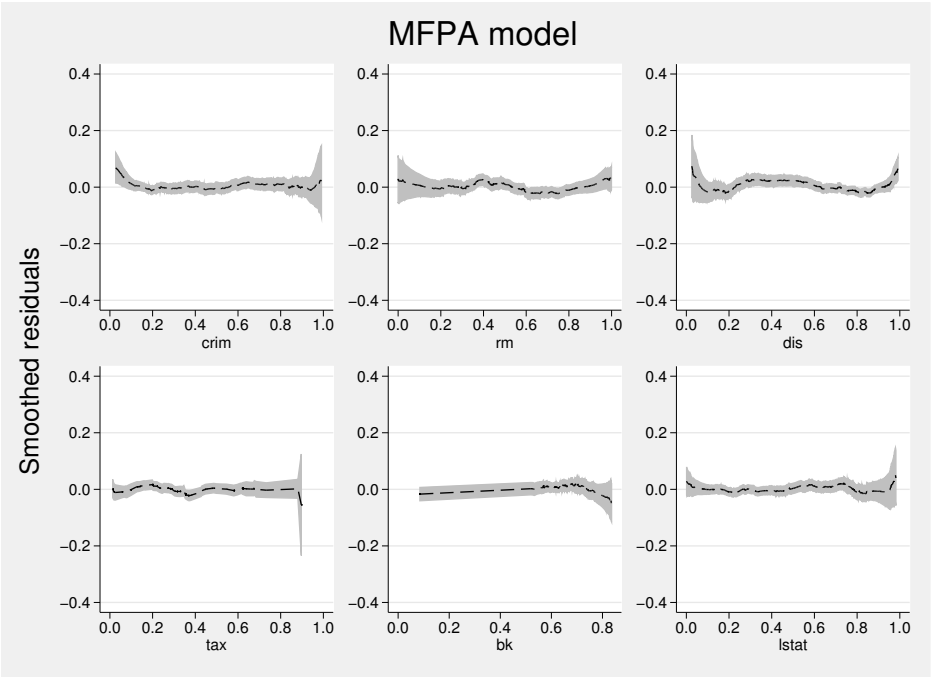

Figure 4. Boston housing data. Smoothed residuals from the MFPA model for the partial predictors in the MFP model.

Altogether the fit seems a little better, and the only predictor still exhibiting lack of fit is `dis`.

This example suggests that MFPA may uncover subtle nonlinearity missed by MFP in difficult situations with unusual distributions and a potential influence of extreme values. The overall predictive ability of the model may not be too different, but the interpretation of the effects of individual predictors may change.

5 The mfpa command

5.1 Syntax

The syntax of `mfpa` is as follows:

`mfpa`[, `acd(`*varlist*`)` `linadj(`*varlist*`)` *mfp_options*]: *regression_cmd*
[*yvar1* [*yvar2*]] *xvarlist* [*if*] [*in*] [*weight*] [, *regression_cmd_options*]

`mfpa` is identical to `mfp` except that it accepts factor variables in *xvarlist* and has two additional options, which are described below.

The standard postestimation commands `fracpred` and `fracplot` have been replaced with `xfracpred` and `xfracplot`, respectively.

Note that the `acd` program must be installed before using `mfpa`. To install `acd`, type `net install st0339, from(http://www.stata-journal.com/software/sj14-2)`.

5.2 Description

`mfpa` selects the MFP model that best predicts the outcome variable from the right-hand-side variables in *xvarlist*.

`mfpa` provides some extensions to Stata's `mfp` command:

1. `mfpa` supports factor variables, and
2. `mfpa` has two new options: `linadj(`*varlist*`)` to adjust linearly for variables in *varlist*, and `acd(`*varlist*`)` to optimize the fit for each *xvar* in *varlist* and its ACD transformation.

As mentioned above, the `mfp` postestimation commands `fracpred` and `fracplot` are replaced with `xfracpred` and `xfracplot`, respectively. The syntax is unchanged except that `xfracplot` has the additional option `nopts`, which suppresses plotting of partial residuals. Also provided with the software package for this article is `xfracpoly`, which extends the `fracpoly` command (which is no longer part of official Stata) by supporting the use of factor variables in its *xvarlist*. The three `xfrac`* commands are briefly documented in the `mfpa` help file under the heading *Related commands*.

5.3 Options

`acd(`*varlist*`)` creates the ACD transformation of each member of *varlist*. It also invokes the FSPA to determine the best-fitting FP1(p_1, p_2) model, as described in section 2.4. For a given continuous predictor *xvar*, depending on the values of `select(#)` and `alpha(#)`, `mfpa` simplifies the FP1(p_1, p_2) model to select one of the six submodels described in section 2.2. The variable representing the ACD transformation of *xvar* is named `A`*xvar* and is left behind in the workspace, together with FP transformation(s) of `A`*xvar* as appropriate.

`linadj(`*varlist*`)` adjusts linearly for members of *varlist*; that is, the members are included in every model fit. This avoids the need for the more complicated and less efficient `df()` and `select()` options to achieve the same result.

mfp_options are any options appropriate to `mfp`.

regression_command_options are any options appropriate to the regression command specified in *regression_command*.

5.4 Examples

```
webuse brcancer, clear
stset rectime, failure(censrec) scale(365.24)
mfpa, acd(x5): stcox x5
mfpa, select(0.05): stcox x1 x2 x3 x4a x4b x5 x6 x7 hormon
xfracplot x5
mfpa, select(0.05) acd(x5 x6 x7): stcox x1 x2 x3 x4a x4b x5 x6 x7 hormon
xfracplot x5
```

6 Comments

In this article, we introduced MFPA and the `mfpa` command, an extension of MFP and `mfp` that supports the ACD transformation in the range of possible predictor transformations. If sigmoid relationships are relevant or expected, MFPA can be used instead of MFP. Our impression is that replacement of the FP2(p_1, p_2) family with the FP1(p_1, p_2) family does not sacrifice flexibility in functional form. The mathematical details of how this happens merit further investigation. With the possibility of modeling singly or doubly asymptotic relationships, the FP1(p_1, p_2) family offers an attractive alternative to the FP2(p_1, p_2) family in some cases. However, its interpretability and transportability are less straightforward than those of MFP, and its properties remain to be explored in greater detail and in more datasets.

The ACD transformation may provide a solution to the problem of influential covariate observations. In the MFP context, we previously proposed the $g_\delta(.)$ pretransformation (Royston and Sauerbrei 2007a), which works quite differently from ACD. For any continuous x, the distribution of ACD(x) is by construction approximately uniform $(0, 1)$. The extreme values of the uniform distribution are generally much less influential in regression models than those of the original distribution of x. In the selected functions of `x5` in the German breast cancer dataset, the FSP selects a nonmonotonic FP2 function, which contradicts medical knowledge, whereas the FSPA chooses FP1$(., 3)$, which fits the data well and, being guaranteed monotonic, makes more biologic sense.

In summary, publication of `mfpa` makes the command widely available to other researchers. We hope this will stimulate further research in this important topic area.

7 References

Baade, P. D., P. Royston, P. H. Youl, M. A. Weinstock, A. Geller, and J. F. Aitken. 2015. Prognostic survival model for people diagnosed with invasive cutaneous melanoma. *BMC Cancer* 15: 27.

Royston, P. 2015. Tools for checking calibration of a Cox model in external validation: Prediction of population-averaged survival curves based on risk groups. *Stata Journal* 15: 275–291.

Royston, P., and D. G. Altman. 1994. Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling (with discussion). *Journal of the Royal Statistical Society, Series C* 43: 429–467.

Royston, P., and W. Sauerbrei. 2007a. Improving the robustness of fractional polynomial models by preliminary covariate transformation: A pragmatic approach. *Computational Statistics and Data Analysis* 51: 4240–4253.

———. 2007b. Multivariable modeling with cubic regression splines: A principled approach. *Stata Journal* 7: 45–70.

———. 2008. *Multivariable Model-building: A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables*. Chichester, UK: Wiley.

Sauerbrei, W., and P. Royston. 1999. Building multivariable prognostic and diagnostic models: Transformation of the predictors by using fractional polynomials. *Journal of the Royal Statistical Society, Series A* 162: 71–94.

Schumacher, M., G. Bastert, H. Bojar, K. Hübner, M. Olschweski, W. Sauerbrei, C. Schmoor, C. Beyerle, R. L. A. Neumann, and H. F. Rauschecker. 1994. Randomized 2×2 trial evaluating hormonal treatment and the duration of chemotherapy in node-positive breast cancer patients. *Journal of Clinical Oncology* 12: 2086–2093.

About the authors

Patrick Royston is a medical statistician with more than 30 years of experience, with a strong interest in biostatistical methods and in statistical computing and algorithms. He works largely in methodological issues in the design and analysis of clinical trials and observational studies. He is currently focusing on alternative outcome measures in trials with a time-to-event outcome; on problems of model building and validation with survival data, including prognostic factor studies and treatment-covariate interactions; on parametric modeling of survival data; and on novel clinical trial designs.

Willi Sauerbrei has worked for more than 30 years as an academic biostatistician. He has extensive experience of cancer research and a long-standing interest in modeling observational data. Topics of interest include variable and function selection, model stability, treatment-covariate interactions, time-dependent effects in survival analysis, meta-analysis, and reporting of research findings.

Royston and Sauerbrei have collaborated on regression methods using continuous predictors for more than two decades and have written a book (Royston and Sauerbrei 2008) on multivariable modeling.