Multivariable Fractional Polynomials in SAS

An Algorithm for Determing the Transformtion of Continous Covariates and Selection of Covariates

Carolina Meier-Hirmer, Carina Ortseifen and Willi Sauerbrei
Institute of Medical Biometry, Freiburg
Germany

June 11, 2003

Contents

1 Introduction

Fractional polynomials (FPs) were introduced by Royston and Altman (1994) for modelling the functional form of the relationship between a continuous covariate and the outcome in general types of regression models. For multivariable analysis a modified version was proposed by Sauerbrei and Royston (1999). It combines backward elimination with a systemic search for a 'suitable' transformation to represent the influence of each continuous covariate on the outcome. An application of multivariable FPs in modelling prognostic and diagnostic factors

in breast cancer is given by Sauerbrei et al. (1999), the stability of the models selected is investigated in Royston and Sauerbrei (2003).
Briefly, fractional polynomials models are especially useful when one wishes to preserve the continuous nature of the covariates in a regression model, but suspects that some or all of the relationships may be non-linear. At each step of a 'backfitting' algorithm `%mfp8` constructs a fractional polynomial transformation for each continuous covariate while fixing the current functional forms of the other covariates. The algorithm terminates when the functional forms of the covariates do not change.
The name of the SAS-Macro is `%mfp8` as it should be used with version 8.* of SAS. There is also a less performing macro `%mfp6` available for version 6.* of SAS.

2 Syntax of the %mfp8 Macro

```
%mfp8  (
       dsname= ,                          dataset
       yname= ,                           dependent variable
       xname= ,                           positive, continuous covariate(s)
       xbin= ,                            binary covariate(s)
       xinclude= ,                        covariate(s) forced in all models
       model=N,                           kind of model
       pw=-2 -1 -0.5 0 0.5 1 2 3,         global power values of the polynomial
       m=2,                               global maximum degree of the polynomial
       dp= ,                              degree of the polynomial for single covariate(s)
       power= ,                           powers of the polynomial for single covariate(s)
       censvar= ,                         censoring variable for survival data
       censval=0,                         censoring value(s) for survival data
       strata= ,                          stratifying variable(s) for survival data
       ties=efron,                        method for treating ties for survival data
       alpha=0.05,                        significance level
       mselect=ra2,                       method for choosing best model
       maxcycle=10,                       maximum number of cycles
       macpath= ,                         path for SAS-macros
       dsout=fpout,                       result dataset
       showres=r                          results in output window
       );
```

2.1 Mandatory Parameters

In order to run the macro, the following parameters have at least to be specified:
The dataset used (`dsname`), the response variable `yname`, the covariate(s) (either `xname` or `xbin` is necessary) and the path of the directory containing the SAS-macros `macpath`. The dataset `dsname` must not be stored within the SAS working library. This library is emptied at the startup of the macro.
For further parameters default values have been specified, which can be seen above.

2.2 Macro Description

`%mfp8` fits fractional polynomials in `xname` to `yname`. Covariates specified in `xbin` or `xinclude` are all modelled without considering any transformation. `model` may be `N` (Gaussian model), `L` (logistic model) or `S` (Cox model). Before the initial cycle the covariates entered in `xname` or `xbin` are sorted in ascending order based on the p-values of the full (linear) model. The aim is to investigate first possible transformations of important covariates.
To simplify notation, the term 'linear included' will also be used for binary covariates simply indicating that the binary covariate was included.
At the initial cycle the best-fitting FP function is determined for the first covariate, with all the other covariates included linear. All significance tests are carried out using an approximate p-value calculation based on a difference in deviances ($-2 \cdot \log(likelihood)$) having a χ^2 or F distribution, depending on the regression in use (cf. Royston and Altman (1994)). A cycle is defined as a pass through all covariates, whereas a step denotes processing of one covariate.
The functional form (but NOT the estimated regression coefficients) for the first covariate is kept, and the process is repeated for the following covariates. Covariates with a p-value larger than `alpha` (calculated in each step) are excluded for the rest of the cycle, but their influence is reinvestigated in the following cycle. The first iteration concludes when all covariates given in `xname` and `xbin` have been processed in this way. The next cycle starts similarly, except that the functional forms from the initial cycle are retained. The transformations used for the adjusted covariates are continuously updated with the results from the ongoing cycle. The covariate(s) determined in `xbin` are processed in a slightly different way as the macro only examines whether the covariate(s) are to be included in the model or dropped.
Updating FP functions and candidate covariates continues until the functions and covariates included in the overall model do not change (convergence). Convergence is almost always achieved within 2-5 cycles.

It is very important to specify the path for the auxiliary macros `macpath`. Also the relevant dataset `dsname` has to be specified, which must not be stored within the SAS working library. This library is emptied at the startup of the macro. The total number of covariates in `xname` and in `xbin` is limited to 99.
If at least one of the covariates specified has missing values, all observations containing missing values are ignored by the macro (complete case analysis). A warning message is given denoting the number of complete cases.
The macro uses the intrinsic SAS procedures `REG`, `PHREG` and `LOGISTIC`.

2.3 Parameter Explanation

`dsname` dataset which contains all required variables. The macro terminates with an error message if the library is not properly assigned or if not all covariates are found in the dataset. The dataset must not be stored within the working library.

`yname` name of the response variable. The macro stops terminates with an error message if the response variable does not match the parameter given in `model`.

`xname` name(s) of the continuous covariate(s), i.e. the covariate(s) which are considered for transformation. The macro terminates with an error message if a covariate in `xname` has any negative or zero value. In this case one has to create a new variable with a preliminary transformation (e.g. $X_{1new} = X_1 + 1$).

`xbin` name(s) of binary covariate(s) or covariate(s) which generally should not be transformed. Note: If non-binary variables are used linearity is assumed. The algorithm considers only the possibility of inclusion or elimination.

`xinclude` name(s) of the covariate(s) which are always included (independent of p-values). Transformations are not investigated. In most cases these covariates are binary or categorical covariates, e.g. treatment, sex.

`model = N` specifies the used statistical model. `N` for the Gaussian model, `S` for the survival model and `L` for the logistic model. The Gaussian model is default. The Gaussian model uses the SAS procedure `PROC REG`, the survival model `PROC PHREG` (Cox-Model) and the logistic model `PROC LOGISTIC`. The latter model requires 0-1 coding of the outcome variable.

`m = 2` global definition of the maximum degree of fractional polynomials fitted for all covariates entered in `xname`. `2`: FP of second degree (default), `1`: FP of first degree, `0`: linear model.

`pw = -2 -1 -0.5 0 0.5 1 2 3` global definition of the set of fractional polynomial powers from which models are to be chosen. The default set is proposed by Royston and Altman (1994). The logarithm is indicated by 0.

`dp` possibility to limit the global degree of the fractional polynomial (`m`) for specified covariates. For example `dp=age:1 size:0`. `1` fits a first degree FP, `0` a linear model.

`power` possibility to limit the global set of powers for the FP (`pw`) for specified covariates, e.g. `power=age[0.5 1 2 3]`. Different covariates can be entered, separated by blanks. The set of powers must be sorted in an ascending order and also separated by blanks. Restrictions of the functions may be appropriate when including knowledge of the underlying structure (e.g. section 3.1.3 in Sauerbrei and Royston (1999)).

`censvar` if a survival model is calculated, a variable has to be specified which contains the censoring information.

`censval = 0` denotes the censoring value of the variable `censvar`. The default value is 0. Several values can be given separated by blanks. Only relevant in survival models.

`strata` calculating a survival model, strata variables can be specified separated by blanks. Not possible with other models.

`ties = efron` definition of the method used for handling ties in survival data. Possible entries are: `efron`, `breslow`, `discrete` or `exact`. The default value is `efron`.

`alpha = 0.05` significance level for testing between FP models of different degrees and for deciding about elimination of a covariate. It can be entered in two ways. First, a single value can be specified. In this case the value is globally used for all covariates. Second, a vector can be entered, e.g.: `alpha = 0.05 0.01 0.05`. The length of the vector must match the number of covariates in `xname` plus the number of covariates in `xbin`. The selection levels are applied in the same order you have specified the covariates in `xname` followed by the covariates in `xbin`. The default selection level is `0.05`.

`mselect = ra2` two different test algorithms are possible: `seq` and `ra2`. They are explained in section 3. The default value is `ra2`.

`maxcycle = 10` maximum number of iteration cycles permitted. If the selected model has changed after each cycle until cycle `maxcycle`, the macro stops showing a warning message. If only a univariate FP is fitted, i.e. only one covariate is specified, `maxcycle` is set to 1. The default value is `10`, but usually the program converges after less than 5 cycles.

`macpath` path of the directory containing the SAS macros used by `%mfp`. Has to be specified.

`dsout = fpout` specifies the name of the result datasets. Two datasets `dsout1` and `dsout2` are generated. The first one contains the transformations of the covariates, the second one the parameter estimates. See the last two tables in the example (5. Names within every library are possible. The default value is `fpout`. Note: when the macro starts to execute, all datasets in the working directory are deleted.

`showres = r` determines the amount of results given in the output window. '`n`' stands for no results. Only the result datasets are generated. '`r`' means that only main results are presented. '`d`' represents displaying almost all auxiliary calculations and all results. The default value is '`r`'.

3 Test Algorithms

In what follows we describe two model selection procedures for a single continuous covariate x representing one step of the iterative algorithm. In each procedure a significance level is chosen to test for inclusion of x and to decide about the FP transformation selected. For further information please refer to Sauerbrei and Royston (2002). The null distribution of the likelihood ratio statistic used in the significance tests is assumed to be F for normally distributed data, χ^2 in other cases. In the descriptions below, the most complex model allowed for x is taken to be an FP with $m = 2$, though extensions or restrictions are obvious. The parameter for choosing the test algorithm in the SAS-macro is `mselect`. By default it is set to `ra2`.

3.1 Sequential

The algorithm for the sequential model selection procedure (`mselect=seq`) works as follows:

1. Perform a 2 df test at the α level of the best-fitting second-degree FP against the best-fitting first-degree FP. If the test is significant stop, (the final model is the FP with $m = 2$), otherwise continue.

2. Perform a 1 df test at the α level of the best-fitting first-degree FP against a straight line. If the test is significant, stop (the model is the FP with $m = 1$), otherwise continue.

3. Perform a 1 df test at the α level of a straight line against the model omitting x. If the test is significant, the final model is a straight line, otherwise omit x.

Because several tests are carried out, the actual Type I error rate may exceed the nominal value of α when the true relationship is a straight line. Therefore the procedure tends to favour more complex models over simple ones. For results from a simulation study see Ambler and Royston (2001). Nevertheless the procedure is more intuitive than the following one.

3.2 RA2

`%mfp` uses the `ra2` algorithm as default. This algorithm is described in Ambler and Royston (2001) and in Sauerbrei and Royston (2002). It has the flavour of a closed test procedure (Marcus et al. 1976) which maintains approximately the correct Type I error rate for each component test. The procedure allows the complexity of candidate models to increase progressively from a prespecified minimum (a null model) to a prespecified maximum (an FP) according to an ordered sequence of test results.
The algorithm works as follows:

1. Perform a 4 df test at the α level of the best-fitting second-degree FP against the null model. If the test is not significant, drop x and stop, otherwise continue.
2. Perform a 3 df test at the α level of the best-fitting second-degree FP against a straight line. If the test is not significant, stop (the final model is a straight line), otherwise continue.
3. Perform a 2 df test at the α level of the best-fitting second-degree FP against the best-fitting first-degree FP. If the test is significant, the final model is the FP with $m = 2$, otherwise the FP with $m = 1$.

The tests in step 1, 2 and 3 are of overall association, non-linearity and between a simpler or more complex FP model, respectively.

4 Miscellaneous

4.1 Warnings and Error Messages

The macro stops and displays warnings or error messages in the output window if:

- the `dsname` dataset is stored within the working library.
- the dataset is not found. Probably a problem with the library assignment.
- at least one covariate entered has value zero or less. These covariates should be transformed by the user.
- the parameter `model` is set to another value than `N, S` or `L`.
- survival model: the survival time entered is less than zero.
- survival model: no censoring variable is specified.
- logistic model: the response variable has values other than 0 or 1.

- no response variable is specified.
- all observations have missing values. The FP macro uses only complete cases.
- not all variables given in `yname`, `xname`, `xbin`, `xinclude` and, additionally in survival models, in `censvar` or `strata` are found in the specified dataset.

4.2 Batchmode

It is possible to run the macro in batch mode. In this case you include the `%mfp8` macro in beforehand entering:

```
%include "C:\FP\mfp8.sas";
```

5 Example

We illustrate two of the analyses performed by Sauerbrei and Royston (1999). We use a dataset which contains data from a study of the German Breast Cancer Study Group for patients with node-positive breast cancer. The dataset can be downloaded in text-format from the Web site `http://www.blackwellpublishers.co.uk/rss/`. The response variable is recurrence free survival time (`rfs`) and the censoring variable is `cens`. There are 686 patients with 299 events. We use Cox regression to predict the log hazard of recurrence from prognostic factors of which 5 are continuous (`x1`, `x3`, `x5`, `x6`, `x7`) and 3 are binary (`x2`, `x4a`, `x4b`). Originally `x4` was an ordered categorical covariate with values 1, 2 and 3. Two dummy covariates were generated to present the effect (`x4a` and `x4b`). The treatment covariate (`hormon`) is forced into the model. We use `%mfp` to build a model from the initial set of 8 covariates using the backfitting model selection algorithm. We set the selection level for the algorithm to 0.05. The covariate `x1` is divided by 50 in order to avoid very high values in comparison to the other covariates. To `x6` and `x7` one is added to avoid zero values. This was done in a data step before the evocation of `%mfp`. `RA2` is chosen as selection algorithm and the ties are handled with the `breslow` method.

```
%mfp8 (  model=S,
         dsname= fp.bmft,
         yname= rfs,
         xname= x1 x3 x5 x6 x7,
         xbin= x2 x4a x4b,
         xinclude= hormon,
         alpha= 0.05,
         censvar= cens,
         mselect= ra2,
         ties= breslow,
         macpath= C:\FP,
         showres=r
         );
```

MFP8 called with following settings

```
 Dataset name          :  fp.bmft
 Dependent variable    :  rfs
 Covariates            :  x1 x3 x5 x6 x7 x2 x4a x4b
 Covar.  always incl.  :  hormon
 Model (N, S or L)     :  S
 Power values          :  -2 -1 -0.5 0 0.5 1 2 3
 Maximal degree        :  2
 Alpha                 :  0.05
 Model selection       :  RA2
 Censoring variable    :  cens
 Censoring value       :  0
 Strata variable       :
 Ties                  :  breslow
 Path to SAS macros    :  C:\FP
 Number of cycles      :  10
```

FP8: Model without transformation

The PHREG Procedure

Model Information

Data Set	WORK.DSNAME
Dependent Variable	rfs
Censoring Variable	cens
Censoring Value(s)	0
Ties Handling	BRESLOW

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
686	299	387	56.41

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	3576.346	3471.637
AIC	3576.346	3489.637
SBC	3576.346	3522.941

MFP8: Model without transformation

The PHREG Procedure

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	104.7094	9	<.0001
Score	120.5901	9	<.0001
Wald	114.7760	9	<.0001

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Confidence	Ratio Limits
hormon	1	-0.34624	0.12907	7.1959	0.0073	0.707	0.549	0.911
x1	1	-0.47267	0.46501	1.0332	0.3094	0.623	0.251	1.551
x3	1	0.00780	0.00394	3.9194	0.0477	1.008	1.000	1.016
x5	1	0.04878	0.00745	42.8991	<.0001	1.050	1.035	1.065
x6	1	-0.00222	0.0005735	14.9514	0.0001	0.998	0.997	0.999
x7	1	0.0001978	0.0004504	0.1929	0.6605	1.000	0.999	1.001
x2	1	0.25816	0.18348	1.9797	0.1594	1.295	0.904	1.855
x4a	1	0.63598	0.24920	6.5130	0.0107	1.889	1.159	3.078
x4b	1	0.14337	0.13626	1.1072	0.2927	1.154	0.884	1.507

MFP8: IMPORTANT: The new ordering of the covariates based on p-values is:
x5 x6 x4a x3 x2 x4b x1 x7
It is used in the following calculations and outputs.

>>> -------------------- CYCLE 1 --------------------

MFP8: Variable -x5-

Best Functions for Different Degrees m

Function	m	p1	p2	deviance	diffra2	pdiffdev
Omitted	-1	.	.	3503.610	61.366	0.00000
Linear	0	.	.	3471.637	29.393	0.00000
First Degree	1	0	.	3449.203	6.959	0.03082
Second Degree	2	0.5	3	3442.244	0.000	1.00000

MFP8: Variable -x6-

Best Functions for Different Degrees m

Function	m	p1	p2	deviance	diffra2	pdiffdev
Omitted	-1	.	.	3464.113	29.917	0.00001
Linear	0	.	.	3442.244	8.048	0.04503
First Degree	1	0.5	.	3435.550	1.354	0.50813
Second Degree	2	-2	0.5	3434.196	0.000	1.00000

MFP8: Variable -x4a-

Best Functions for Different Degrees m

Function	m	p1	p2	deviance	diffra2	pdiffdev
Omitted	-1	.	.	3440.749	5.199	0.02260
Linear	0	.	.	3435.550	0.000	1.00000

MFP8: Variable -x3-

Best Functions for Different Degrees m

Function	m	p1	p2	deviance	diffra2	pdiffdev
Omitted	-1	.	.	3436.832	3.560	0.46883
Linear	0	.	.	3435.550	2.278	0.51683
First Degree	1	-1	.	3433.677	0.405	0.81678
Second Degree	2	-2	3	3433.273	0.000	1.00000

MFP8: Variable -x2-

Best Functions for Different Degrees m

Function	m	p1	p2	deviance	diffra2	pdiffdev
Omitted	-1	.	.	3437.589	0.756	0.38444
Linear	0	.	.	3436.832	0.000	1.00000

MFP8: Variable -x4b-

Best Functions for Different Degrees m

Function	m	p1	p2	deviance	diffra2	pdiffdev
Omitted	-1	.	.	3437.848	0.259	0.61082
Linear	0	.	.	3437.589	0.000	1.00000

MFP8: Variable -x1-

Best Functions for Different Degrees m

Function	m	p1	p2	deviance	diffra2	pdiffdev
Omitted	-1	.	.	3437.893	18.085	0.00119
Linear	0	.	.	3437.848	18.040	0.00043
First Degree	1	-2	.	3433.628	13.820	0.00100
Second Degree	2	-2	-0.5	3419.808	0.000	1.00000

MFP8: Variable -x7-

Best Functions for Different Degrees m

Function	m	p1	p2	deviance	diffra2	pdiffdev
Omitted	-1	.	.	3420.805	3.715	0.44600
Linear	0	.	.	3419.808	2.717	0.43730
First Degree	1	3	.	3417.689	0.598	0.74138
Second Degree	2	-0.5	3	3417.091	0.000	1.00000

```
>>> ------------------- CYCLE 2 -------------------
```

MFP8: Variable -x5-

Best Functions for Different Degrees m

Function	m	p1	p2	deviance	diffra2	pdiffdev
Omitted	-1	.	.	3494.867	74.143	0.00000
Linear	0	.	.	3451.795	31.071	0.00000
First Degree	1	0	.	3428.023	7.299	0.02600
Second Degree	2	-2	-1	3420.724	0.000	1.00000

MFP8: Variable -x6-

Best Functions for Different Degrees m

Function	m	p1	p2	deviance	diffra2	pdiffdev
Omitted	-1	.	.	3452.093	32.704	0.00000
Linear	0	.	.	3427.703	8.313	0.03996
First Degree	1	0.5	.	3420.724	1.334	0.51314
Second Degree	2	0	0	3419.389	0.000	1.00000

MFP8: Variable -x4a-

Best Functions for Different Degrees m

Function	m	p1	p2	deviance	diffra2	pdiffdev
Omitted	-1	.	.	3425.310	4.586	0.03223
Linear	0	.	.	3420.724	0.000	1.00000

MFP8: Variable -x3-

Best Functions for Different Degrees m

Function	m	p1	p2	deviance	diffra2	pdiffdev
Omitted	-1	.	.	3420.724	5.305	0.25741
Linear	0	.	.	3418.769	3.350	0.34070
First Degree	1	-1	.	3415.746	0.327	0.84920
Second Degree	2	-0.5	0	3415.419	0.000	1.00000

MFP8: Variable -x2-

Best Functions for Different Degrees m

Function	m	p1	p2	deviance	diffra2	pdiffdev
Omitted	-1	.	.	3420.724	0.214	0.64370
Linear	0	.	.	3420.510	0.000	1.00000

MFP8: Variable -x4b-

Best Functions for Different Degrees m

Function	m	p1	p2	deviance	diffra2	pdiffdev
Omitted	-1	.	.	3420.724	0.145	0.70316
Linear	0	.	.	3420.579	0.000	1.00000

```
                  MFP8:  Variable -x1-
         Best Functions for Different Degrees m
Function        m   p1   p2   deviance  diffra2  pdiffdev

Omitted        -1   .    .    3440.057   19.333   0.00068
Linear          0   .    .    3440.038   19.314   0.00024
First Degree    1   -2   .    3436.949   16.225   0.00030
Second Degree   2   -2  -0.5  3420.724    0.000   1.00000

                  MFP8:  Variable -x7-
         Best Functions for Different Degrees m
 Function        m   p1  p2  deviance  diffra2  pdiffdev
 Omitted        -1   .   .   3420.724   2.152    0.70784
 Linear          0   .   .   3419.968   1.396    0.70643
 First Degree    1   3   .   3418.817   0.245    0.88482
 Second Degree   2  -1   3   3418.572   0.000    1.00000

>>> -------------------- CYCLE 3 --------------------

                  MFP8:  Variable -x5-
         Best Functions for Different Degrees m
 Function        m   p1  p2  deviance  diffra2  pdiffdev

 Omitted        -1   .   .   3494.867  74.143    0.00000
 Linear          0   .   .   3451.795  31.071    0.00000
 First Degree    1   0   .   3428.023   7.299    0.02600
 Second Degree   2  -2  -1   3420.724   0.000    1.00000

                  MFP8:  Variable -x6-
         Best Functions for Different Degrees m
 Function        m    p1   p2  deviance  diffra2  pdiffdev

 Omitted        -1    .    .   3452.093  32.704    0.00000
 Linear          0    .    .   3427.703   8.313    0.03996
 First Degree    1   0.5   .   3420.724   1.334    0.51314
 Second Degree   2    0    0   3419.389   0.000    1.00000

                 MFP8:  Variable -x4a-
         Best Functions for Different Degrees m
   Function  m   p1  p2  deviance  diffra2  pdiffdev

   Omitted  -1   .   .   3425.310   4.586    0.03223
   Linear    0   .   .   3420.724   0.000    1.00000

                  MFP8:  Variable -x3-
         Best Functions for Different Degrees m
Function        m    p1    p2  deviance  diffra2  pdiffdev
Omitted        -1    .     .   3420.724   5.305    0.25741
Linear          0    .     .   3418.769   3.350    0.34070
First Degree    1    -1    .   3415.746   0.327    0.84920
Second Degree   2   -0.5   0   3415.419   0.000    1.00000
```

```
                 MFP8:  Variable -x2-
         Best Functions for Different Degrees m
   Function   m   p1   p2  deviance  diffra2  pdiffdev

   Omitted   -1   .    .   3420.724   0.214    0.64370
   Linear     0   .    .   3420.510   0.000    1.00000

                 MFP8:  Variable -x4b-
         Best Functions for Different Degrees m
   Function   m   p1   p2  deviance  diffra2  pdiffdev

   Omitted   -1   .    .   3420.724   0.145    0.70316
   Linear     0   .    .   3420.579   0.000    1.00000

                 MFP8:  Variable -x1-
         Best Functions for Different Degrees m
Function       m   p1    p2   deviance  diffra2  pdiffdev

Omitted       -1   .     .    3440.057   19.333   0.00068
Linear         0   .     .    3440.038   19.314   0.00024
First Degree   1   -2    .    3436.949   16.225   0.00030
Second Degree  2   -2  -0.5   3420.724    0.000   1.00000

                 MFP8:  Variable -x7-
         Best Functions for Different Degrees m
 Function       m   p1   p2  deviance  diffra2  pdiffdev

 Omitted       -1   .    .   3420.724   2.152    0.70784
 Linear         0   .    .   3419.968   1.396    0.70643
 First Degree   1   3    .   3418.817   0.245    0.88482
 Second Degree  2   -1   3   3418.572   0.000    1.00000

>>> ---------------------------------------------------

>>> MFP8 converged after 3 cycles.
              MFP8:  Final multivariable fractional polynomial model

               Variable  Function       Alpha  Status     p1   p2

               HORMON    Linear          1.00  forced in  1    .
               x5        Second Degree   0.05  in         -2   -1
               x6        First Degree    0.05  in         0.5  .
               x4a       Linear          0.05  in         1    .
               x3        Omitted         0.05  out        .    .
               x2        Omitted         0.05  out        .    .
               x4b       Omitted         0.05  out        .    .
               x1        Second Degree   0.05  in         -2   -0.5
               x7        Omitted         0.05  out        .    .
```

Variable	Coefficient	Standard Error	Pr > ChiSq	Hazard Ratio	95% Confidence	Limits
HORMON	-0.40242	0.12808	0.0017	0.6687	0.5202	0.8595
x5^(-2)	3.87904	0.76972	0.0000	48.3777	10.7015	218.6983
x5^(-1)	-5.49065	0.86426	0.0000	0.0041	0.0008	0.0224
x6^(0.5)	-0.05714	0.01109	0.0000	0.9445	0.9242	0.9652
x4a	0.50070	0.24963	0.0449	1.6499	1.0115	2.6912
x1^(-2)	1.78935	0.33027	0.0000	5.9856	3.1331	11.4350
x1^(-0.5)	-8.01542	1.74843	0.0000	0.0003	0.0000	0.0102

```
Log(Likelihood) of the resulting model:       -1710.36

>>> MFP8 has finished!
```

Some explanation of the output from the model selection algorithm is desirable. The output consists of three parts. The first one contains the results of the model assuming linear effects for all covariates. Afterwards the cycle-tables follow which show results of the selection of transformations within the subsequent cycles. After convergence of the macro the transformations for each covariate and the parameter estimates are given for the final model. We call that the result-table.

In our case first the results of the Cox model (as `model=S` was chosen) are shown. The covariate list is ordered by increasing p-values from this initial model. Afterwards the iteration starts. Consider the first few lines of output of the cycle-tables:

```
 1.                  MFP8:  Variable -x5-

              Best Functions for Different Degrees m
     Function        m    p1   p2  deviance  diffra2  pdiffdev

 2.  Omitted        -1    .    .   3503.610   61.366   0.00000
 3.  Linear          0    .    .   3471.637   29.393   0.00000
 4.  First Degree    1    0    .   3449.203    6.959   0.03082
 5.  Second Degree   2   0.5   3   3442.244    0.000   1.00000
```

Note that all p-values are calculated although the chosen model selection algorithm may stop before all transformations are tested. The model is modified variable-by-variable in subsequent steps adjusting for the other covariates. The most significant linear term turns out to be `x5` which is therefore processed first. Line 2 compares the best-fitting second degree FP `x5` with a model omitting `x5`. The deviance of the best second degree FP, which has powers (0.5, 3), can be seen in line 5 (`3442.244`). The reported deviance of 3503.610 in line 2 is the deviance of the model with all covariates included (linear relationship) and `x5` omitted. The deviance of the model for the linear relationship of `x5` can be seen in line 3 (`3471.637`), and the deviance of the $m = 1$ model (`3449.203`) in line 4. Here the power '0' is selected, which means $\log(x_5)$. The `RA2` procedure first compares the best-fitting second degree FP with the model excluding `x5`. The test for inclusion of `x5` is highly significant (line 2, `pdiffdev`) so the procedure continues. Line 3 shows that the $m = 2$ model is also significantly better than a straight line and line 4 that it is also somewhat better than an FP with $m = 1$ (p=0.03082). Thus at this stage in the model selection procedure the 'final' functional form for `x5` is an FP with powers (0.5,

3). The overall model with $m = 2$ for x5 and all other terms linear has deviance 3442.244. The value of pdiffdev for the second degree model is only of technical interest as the value 1.00000 stops the model selection algorithm.
Next the macro determines the functional form for x6, adjusting for x5 as it was just selected with an FP2 (0.5, 3), and for the other covariates assuming a linear relationship.

```
                  MFP8:  Variable -x4a-
          Best Functions for Different Degrees m
Function    m   p1   p2   deviance   diffra2   pdiffdev

Omitted    -1    .    .   3440.749    5.199    0.02260
Linear      0    .    .   3435.550    0.000    1.00000
```

The covariate x4a is the first binary covariate considered by the procedure. It can be seen that the selection algorithm is reduced to decide between (linear) inclusion or exclusion. Transformations are not calculated as this is not sensible.

```
                      MFP8:  Variable -x3-
             Best Functions for Different Degrees m
Function        m   p1   p2   deviance   diffra2   pdiffdev

Omitted        -1    .    .   3436.832    3.560    0.46883
Linear          0    .    .   3435.550    2.278    0.51683
First Degree    1   -1    .   3433.677    0.405    0.81678
Second Degree   2   -2    3   3433.273    0.000    1.00000
```

x3 is the first covariate to be excluded. The difference between the second degree FP and the model excluding the covariate ist not significant (p-value 0.46883). So the covariate is eliminated. However, in the second cycle the covariate x3 is reinvestigated for influence after further transformation of other covariates.
The next covariate in turn is x2. At this time the other covariates are x5 second degree FP with exponents (0.5, 3), x6 linear, x4a linear, x3 eliminated and x4b, x1, x7 all linear as these have not been processed so far. In the following steps the macro determines an FP2 for x1 and eliminates the other covariates.
After the first cycle the obtained model is: $\beta_1 x_5^{0.5} + \beta_2 x_5^3 + \beta_3 x_6 + \beta_4 x_{4a} + \beta_5 x_1^{-2} + \beta_6 x_1^{-0.5}$. This is also the 'starting' model of cycle 2.
The line `>>> -------------------- CYCLE --------------------` delineates the beginning of a new cycle.
After reprocessing in the same way, cycle 3 gives the same results as cycle 2, i.e. convergence is achieved. The model finally chosen is "Model II" as given in Tables 3 and 4 of Sauerbrei and Royston (1999). It includes x1 with powers (-2,-0.5), x4a, x5 with powers (-2, -1) and x6 with power 0.5. The differences in deviance for comparison with a straight line after convergence (cycle 3) are 3440.038-3420.724 = 19.314 and 3451.795-3420.724 = 31.071 respectively provide that there is strong evidence of non-linearity for x1 and for x5. Covariates x2, x3, x4b and x7 are eliminated, as can be seen from their status out in the first result-tables. x6 is included as an FP1 with power 0.5. The functional relationship of x6 has changed in the second cycle, mainly because the correlated covariate x7 was eliminated and the FP2 for covariate x1 was determined at the end of the first cycle, after processing x6.
The second result-table shows the parameter estimates. If covariates were included linearly

the parameter estimate is denoted by the plain covariate name (e.g. x4a). If the covariate was transformed with an first degree FP, the transformation of the covariate is given instead of the sole variable name (e.g. x6^(0.5)). For the second degree FP there are two parameter estimates which are denoted by the transformed covariates. In our example x5 is included as second degree FP with parameter estimates x5^(-2) = 3.87904 and x5^(-1) = -5.49065.

According to Sauerbrei and Royston (1999), medical knowledge implies a monotonic risk function for x5 (number of positive nodes) in contrast to the results of the first example. Sauerbrei and Royston (1999) modified Model II by estimating a preliminary exponential transformation x5e, thus obtaining a monotonic risk function. The value of -0.12 was estimated in a preliminary step (cf. appendix A in Sauerbrei and Royston (1999)). Their model III may be estimated using the following command:

```
%mfp8 (  model=S,
         dsname= fp.bmft,
         yname= rfs,
         xname= x1 x3 x5e x6 x7,
         xbin= x2 x4a x4b,
         xinclude= hormon,
         power= x5e[0.5 1 2 3],
         dp= x5e:1,
         alpha= 0.05,
         censvar= cens,
         mselect= ra2,
         ties= breslow,
         macpath= C:\FP,
         showres=r
         );
```

In comparison to the first call of the macro, we have three changes in the second call which all relate to x5. $x_{5e} = \exp(-0.12 \cdot x_5)$ is used instead of x_5, the set of powers for x5e is restricted to $\{0.5, 1, 2, 3\}$ and the degree of the FP is restricted to 1. The resulting model is as reported in Table 4 of Sauerbrei and Royston (1999):

MFP8: Final multivariable fractional polynomial model

Variable	Function	Alpha	Status	p1	p2
HORMON	Linear	1.00	forced in	1	.
x5e	Linear	0.05	in	1	.
x6	First Degree	0.05	in	0.5	.
x4a	Linear	0.05	in	1	.
x3	Omitted	0.05	out	.	.
x2	Omitted	0.05	out	.	.
x4b	Omitted	0.05	out	.	.
x1	Second Degree	0.05	in	-2	-0.5
x7	Omitted	0.05	out	.	.

```
                          Standard                                         95%
   Variable   Coefficient    Error    Pr > ChiSq  Hazard Ratio  Confidence   Limits

   HORMON      -0.39450     0.12810     0.0021        0.6740       0.5244     0.8664
   x5e         -1.98121     0.22689     0.0000        0.1379       0.0884     0.2151
   x6^(0.5)    -0.05819     0.01109     0.0000        0.9435       0.9232     0.9642
   x4a          0.51744     0.24937     0.0380        1.6777       1.0291     2.7352
   x1^(-2)      1.74216     0.33014     0.0000        5.7096       2.9895    10.9050
   x1^(-1)     -7.81792     1.74944     0.0000        0.0004       0.0000     0.0124

Log(Likelihood) of the resulting model:       -1711.62

>>> MFP8 has finished!
```

The final model states a linear function for `x5e`, which means that the functional relationship for `x5` - adjusted for the other covariates is $-1.98 \cdot \exp(-0.12 \cdot x_5)$. This time the effect of the treatment `HORMON` was not of primary interest, so it may also be used as a strata variable.
To plot functions for the Cox-model, standardisation of the covariates is necessary as the risk is estimated relative to an unspecified baseline hazard function. For more details see section 3.3 in Royston and Sauerbrei (2003). In figure 1 one example for such a plot is given.

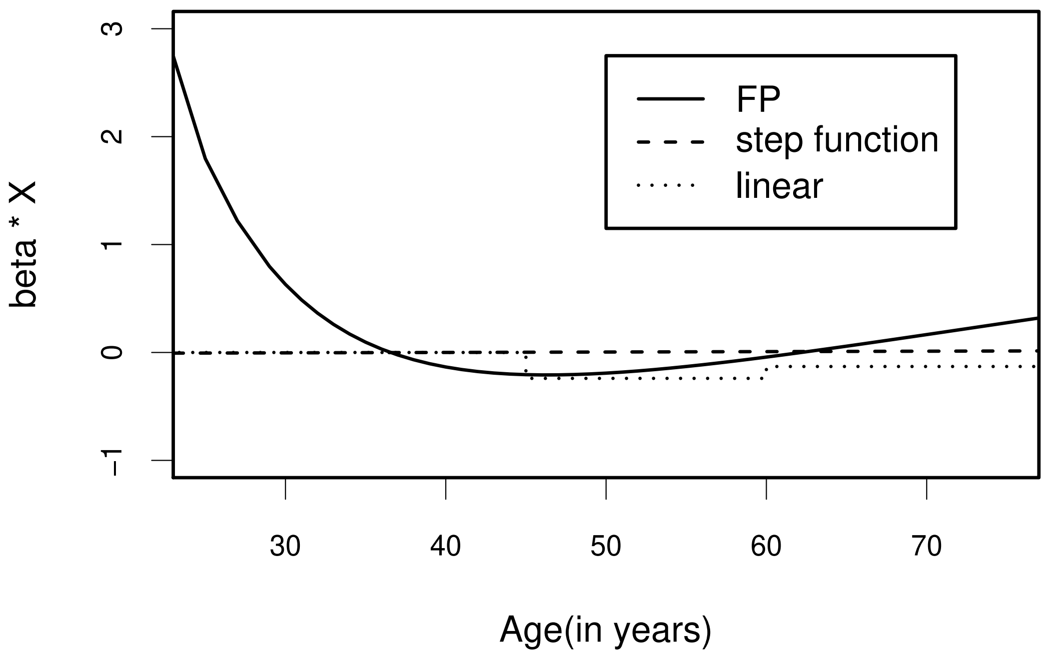

Figure 1: Postulated functional influence of age on the log relative risk of recurrence-free survival.

The linear approach does not show any effect. The fitted step function shows a reduction of the risk for woman between 45 and 65 years but this effect is not significant. “The FP approach indicates that younger patients, up to an age of about 40, have a highly increased risk, that after a fairly constant period between 40 and 55 years the risk increases again.” (Sauerbrei et al. 1999).

References

Ambler, G. and P. Royston (2001). Fractional polynomial model selection procedures: Investigation of type I error rate. *Journal of Statistical Simulation and Computation 69*, 89–108.

Marcus, R., E. Peritz, and K. Gabriel (1976). On closed test procedures wit special reference to ordered analysis of variace. *Biometrika 76*, 655–660.

Royston, P. and D. G. Altman (1994). Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling (with discussion). *Applied Statistics 43*, 429–467.

Royston, P. and W. Sauerbrei (2003). Stability of multivariable fractional polynomial models with selection of variables and transformations: a boostrap approach. *Statistics in Medicine 22*, 639–659.

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

Sauerbrei, W. and P. Royston (2002). Corrigendum: Building multivariable prognostic and diagnostic models: Transformations of the predictors by using fractional polynomials. *Journal of the Royal Statistical Society A 165*, 399–400.

Sauerbrei, W., P. Royston, H. Bojar, C. Schmoor, and M. Schumacher (1999). Modelling the effect of standard prognostic factors in node-positive breast cancer. *British Journal of Cancer 79*, 1752–1760.