Brazilian Journal of Biometrics (2022), **40**, 469–489
doi:10.28951/bjb.v40i4.619

ARTICLE

Increasing the flexibility of mixed models by using fractional polynomials

Edijane Paredes Garcia*[,1] and Luzia Aparecida Trinca*[,2]

[1]Department of Statistics, Federal University of Amazonas, Manaus, Brazil
[2]Biosciences Institute, São Paulo State University "Júlio de Mesquita Filho", Botucatu, Brazil
*Corresponding author. Email: edijane@ufam.edu.br; luzia.trinca@unesp.br

(Received: June 1,2022; Revised: December 14,2022; Accepted: December 14,2022; Published: December 30,2022)

Abstract

The class of regression models incorporating Fractional Polynomials (FPs), proposed by Royston and colleagues in the 1990's, has been extensively studied and shown to be fruitful in the presence of non-linearity between the response variable and continuous covariates. FP functions provide an alternative to higher-order polynomials and splines for dealing with lack-of-fit. Mixed models may also benefit from this class of curves in the presence of non-linearity. The inclusion of FP functions into the structure of linear mixed models has been previously explored, though for simple layouts, e.g. a single covariate in the random intercept model. This paper proposes a general strategy for model-building and variable selection that takes advantage of the FPs within the framework of linear mixed models. Application of the method to three data sets from the literature, known for violating the linearity assumption, illustrates that it is possible to solve the problem of lack-of-fit by using fewer terms in the model than the usual approach of fitting higher-order polynomials.

Keywords: Lack-of-fit; Longitudinal data; Random effects; Selection of variables; Transformation; Variance-covariance structure.

1. Introduction

The linear mixed models (LMM) are extensions of the linear models capable of dealing with dependent or correlated data. Correlations among observations result from the design used in observational (sampling) or experimental studies. Sampling schemes other than simple random sampling, e.g. stratified, clustering and complex random surveys, induce grouping of correlated data just as do blocking and randomization restrictions in the design of experiments. Correlations also arise when several measurements are taken in the same unit under different conditions, leading to the so-called repeated measurement and longitudinal studies. The common feature of these studies is that they deliver clustered data in one or more levels. The flexibility of the linear mixed model for clustered

continuous responses is that efficient and well-established estimation methods are available to model the mean, the variance-covariance and the serial correlation structures. Furthermore interpretations are direct for the fitted marginal and conditional models.

As in any model within the large class of generalized linear models, from which the linear is a particular case, often, there are non-linearity issues to be taken into account in the predictor function. The most common form of dealing with non-linearity is including conventional low-order polynomials. However, the second-order polynomial is symmetric around the maximum or minimum response and may not fit the data well. Higher-order polynomials may fit the data better, but they present several stationary points and may lack interpretation. Another alternative to deal with non-linearity is the use of splines. Yet, splines lack the global effect interpretation (Regier & Parker, 2015), require dense data and might present other difficulties in multiple variable cases.

Royston & Altman (1994) proposed the fractional polynomial (FP) instead and demonstrated their practical usefulness in generalized linear and proportional hazards models. By allowing power parameters assuming non-integer values, more plausible curves or surfaces can be fitted, sometimes of easier interpretation than higher-order polynomials. Fractional polynomials are, in fact, non-linear functions, and to ease the burden in their fitting, the authors proposed selecting the powers from a discrete set of rational values. Royston & Sauerbrei (2008) extended the idea to multiple linear regression and proposed a procedure for variable selection.

FPs are also useful in generalized linear mixed models and are gaining attention recently, see Aregay *et al.* (2013), Long & Ryoo (2010), and Tilling *et al.* (2014) for instance. However, applications are reasonably simple, e.g. just one explanatory variable requiring power transformation. Mixed models with multiple explanatory variables have issues beyond those of variable selection in the classic and generalized linear models since they include at least three parts that could be modeled separately. Changes in one of them may influence the other and thus, some sound strategy is required. Our aim in this paper is to propose a strategy of model-building and variable selection in the LMM when transformations in some continuous explanatory variables may be required. In Section 2 we present details of the FPs functions and algorithms involved and a very concise description of the LMM. In Section 3 we present the fractional polynomial mixed model and the model-building strategy we propose. In Section 4 we present the modeling for three data sets, all obtained from the literature and which were modeled differently. The first one is from an observational longitudinal study with only time as the explanatory. We use this example to show the flexibility of FP functions to model a rather asymmetric curve. The second one is from an experiment, also within the longitudinal framework with two interacting explanatory factors, one continuous and the other binary. The last example is an observational study with multiple continuous variables measured at two levels. To finalize, we draw our conclusions in Section 5.

2. Background

2.1 Fractional Polynomial Functions

The class of FP curves to model average relationships between a response variable and continuous covariates were introduced by Royston & Altman (1994) as an extension of the so-called Box-Tidwell transformations (Box & Tidwell, 1962). For a covariate x, $x > 0$, the FP function of degree m, $(m = 1, 2, \cdots)$, φ_m, is defined by

$$\varphi_m(x, \boldsymbol{\beta}, \mathbf{p}) = \beta_0 + \sum_{j=1}^{m} \beta_j x^{(p_j)}, \tag{1}$$

where $\mathbf{p} = (p_1, p_2, \cdots, p_m) \in \mathbb{R}^m$ with $p_1 \leq p_2 \leq \cdots \leq p_m$ and $\boldsymbol{\beta} \in \mathbb{R}^{m+1}$ the vector of regression coefficients. The $x^{(p)}$ term obeys the Box-Tidwell transformation given by

$$x^{(p)} = \begin{cases} x^p & \text{if } p \neq 0, \\ \log(x) & \text{if } p = 0. \end{cases} \tag{2}$$

For $m > 1$, φ_m follows a recursive formulae in case of powers equality (see Royston & Altman (1994) and Royston & Sauerbrei (2007)) but, usually, for most applications, first and second degrees give good approximations. Thus, for $m = 2$, with powers $\mathbf{p} = (p_1, p_2)$, $x^{(\mathbf{p})}$ is given by

$$x^{(\mathbf{p})} = x^{(p_1,p_2)} = \begin{cases} (x^{(p_1)}, x^{(p_2)}) & \text{if } p_1 \neq p_2, \\ (x^{(p)}, x^{(p)}\log(x)) & \text{if } p_1 = p_2 = p, \end{cases} \tag{3}$$

so that equation (1) reduces to

$$\varphi_2(x, \boldsymbol{\beta}, \mathbf{p}) = \beta_0 + \beta_1 x^{(p_1)} + \beta_2 x^{(p_2)}, \tag{4}$$

and, as p_2 converges to $p_1 = p$, we have the limiting case

$$\varphi_2(x, \boldsymbol{\beta}, p) = \beta_0 + \beta_1 x^{(p)} + \beta_2 x^{(p)}\log(x). \tag{5}$$

For short, $\varphi_m(x, \boldsymbol{\beta}, \mathbf{p})$ will be indicated by FPm in the text such that it will be straight to relate a function of degree one to FP1 and degree two to FP2.

Any statistical model involving a linear predictor function on continuous positive covariates may benefit from this approach. The positivity of x is intrinsic to the definition of FP. For $x \leq 0$, Royston & Sauerbrei (ibid.) provided several useful pre-transformations, from which, the simplest, for non-negative x, is adding a small $\delta > 0$ $(x + \delta)$. Scaling x is also recommended for avoiding numerical problems such as overflow or underflow. Other types of pre-transformations exist targeting, for example, to reduce the effect of influential points (see Royston & Sauerbrei (ibid.) for details).

Allowing $\mathbf{p} \in \mathbb{R}^m$ introduces a huge challenge for estimation and Royston & Altman (1994) proposed a restrict set of powers for the search. Restricting $p_j \in S$, $S = \{-3, -2, -1, -\frac{1}{2}, 0, \frac{1}{2}, 1, 2, 3\}$, simplifies the estimation burden while providing a wide class of curves. Note that such set includes the conventional polynomials, the second-order model in a transformed variable $(\varphi_2(x, \boldsymbol{\beta}, \mathbf{p} = (0, 0)) = \beta_0 + \beta_1 \log(x) + \beta_2 \log(x)^2)$, asymmetric quadratic type curves $(\varphi_2(x, \boldsymbol{\beta}, \mathbf{p} = (-2, 1)) = \beta_0 + \beta_1 x^{-2} + \beta_2 x)$ and some types of asymptotic curves. For graphical illustrations of FP curve shapes see Royston & Altman (ibid.), Regier & Parker (2015) and Garcia (2019).

For given m, $\boldsymbol{\beta}$ is estimated by maximum likelihood (ML) in a grid of values for the powers. The combination of values of $\widehat{\mathbf{p}}$ and $\widehat{\boldsymbol{\beta}}$ resulting in the largest likelihood value are the ML estimators. Such model, for one covariate, has $2m + 1$ degrees of freedom, one for the intercept, m for the powers and m for the regression coefficients. For $m \leq 2$, a function-selection procedure, e.g. FSP algorithm from Ambler & Royston (2001), is then applied to compare the fitted models by using likelihood ratio tests (LRT). Remind that the LRT is applied for nested models and its statistics can be translated in terms of the difference between the deviance values (minus twice the log likelihood) from the two models being compared, say model 1 and model 2. That is $\Delta D(1, 2) = D(df_1) - D(df_2)$ where $D(df_\ell)$ and df_ℓ are, respectively, the deviance and the degrees of freedom of model ℓ for $\ell = 1, 2$. Asymptotically, under the simpler model $(\ell = 1)$, $\Delta D \sim \chi^2$ with $df_2 - df_1$ degrees of freedom. Basically, for one covariate, the FSP procedure involves comparing up to four models at a specified α level, the null model $(df_{null} = 1)$, the straight line on x $(df_{sl} = 2)$, the FP1 on x $(df_{FP1} = 3)$ and the FP2 on x $(df_{FP2} = 5)$. These models are compared sequentially, following the FSP algorithm:

1. x inclusion: null model vs. the best FP2;
2. linearity: straight line vs. FP2;
3. curvature complexity: FP1 vs. FP2.

In each step, evidence for the more complex model leads to the test in the next step. Otherwise, the simpler model in the current step is declared the *best* fit. The reference distribution for the test statistics in each step is χ^2 with four, three and two degrees of freedom, respectively. These are approximations since $\mathbf{p}$ is estimated in the discrete set and produces conservative P-values. However, simulation studies from Ambler & Royston (2001) indicate the approach performs reasonably well for practical use. The procedure preserves the overall type I error probability at the chosen nominal α level. It is possible to set different levels at step 1, say α_1, and steps 2 and 3, say α_2, since step 1 tests the inclusion of x in the model, and 2 and 3 test the complexity of the functional form for x. This would allow, for example, forcing x into the model ($\alpha_1 = 1$), a requirement due to study design or problem-specific.

Royston & Sauerbrei (2008) presented extensions of the above procedure for the case of multiple covariates, the MFP procedure, and for the inclusion of interaction terms, MFPI and MFPIGen procedures for interaction between binary and transformed continuous and for interaction between continuous covariates, respectively. The methods are inspired by the backward selection procedure and effects hierarchy principle, e.g. first search for main effects, then for low-order interactions if necessary, and so on.

For multiple continuous covariates the fractional polynomial model is a sum of FP functions. We briefly present the MFP procedure (restricting $m \leq 2$, that is the maximal model allowed is second-degree), that combines variable selection and function selection:

1. Set $\alpha = (\alpha_1, \alpha_2)$ and the maximal FP degree m for each variable. As for the FSP algorithm, α_1 is the significance level for inclusion of the variable ($\alpha_1 = 1$ forces it in the model, no variable selection) and α_2 is the level for function selection ($\alpha_2 = 1$ forces the most complex allowed FP function in the model, no function selection).
2. Fit the full model (linear effects) and use the partial effects' P-values to rank the variables from most to least significant.
3. Apply the FSP algorithm for the first variable in the rank list, including all original covariates in the model. The outcome is "the variable is dropped" or "no transformation is required" or "an FP transformation is chosen".
4. Apply the FSP algorithm for the next variable in the rank list, including the remaining and the already transformed variables in the model. Repeat this step until the last variable is visited. This ends one cycle of the algorithm.
5. Start a new cycle from 3 where, in the initial model, the variable being visited is set to linear, and all others enter as the FP transformed versions found in the previous cycle. Stop when two consecutive cycles result in the same MFP model.

Usually the algorithm converges with two or three cycles.

As noted, the FP transformations might be useful in any model that includes a form of linear predictors such as generalized linear models (GLM) and survival models. Algorithms are implemented in `R` (R Core Team, 2022) in the `mfp` package from Ambler & Benner (2022), and in commercial statistical software (Saurbrei *et al.,* 2006). FP transformations are likely to improve the flexibility of LMM, which is relevant in practice since estimation in non-linear mixed models is more complex. Besides, formulation of nonlinear models in more than one covariates is difficult.

2.1.1 Linear Mixed Model

A mixed model includes terms that are fixed constants and terms which are random variables. Random terms are usually present by sampling or experimental design such that observational units exhibit some pattern of dependency or correlation. The usual LMM, using standard notation, may be expressed as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon} \tag{6}$$

where $\mathbf{Y}$ is the vector of response variable, $\mathbf{X}$ is the mean model matrix, $\boldsymbol{\beta}$ is the mean model parameters (fixed), $\mathbf{Z}$ is the random effects model matrix, $\mathbf{b} \sim N\left(\mathbf{0}, \sigma^2\mathbf{D}(\theta_{\mathbf{b}})\right)$ is a vector of random coefficients, $\boldsymbol{\epsilon} \sim N\left(\mathbf{0},\ \sigma^2\mathbf{R}(\theta_{\boldsymbol{\epsilon}})\right)$ is the random error vector and $\theta = (\theta_b^\top, \theta_{\boldsymbol{\epsilon}}^\top, \sigma^2)^\top$ is the vector of the variance-covariance structure parameters. The model also assumes $\mathbf{b} \perp \boldsymbol{\epsilon}$. $\mathbf{D}(\theta_{\mathbf{b}})$ and $\mathbf{R}(\theta_{\boldsymbol{\epsilon}})$ are symmetric positive-definite matrices of dimension $q \times q$ and $N \times N$ where q and N are the dimensions of vectors $\mathbf{b}$ and $\boldsymbol{\epsilon}$, respectively. Several options are allowed for the structure of these matrices, from diagonal (for the case of uncorrelated random terms) to the so called unstructured type (the most complex correlation pattern). For useful and popular structures see, for example, Littell *et al.* (1996) and Pinheiro & Bates (2000).

The model specified in (6) is very general in the sense that several grouping patterns, as well as regression coefficients' heterogeneity (i.e. random slopes), can be accommodated thorough the specification of the $\mathbf{Z}$ matrix.

Restricted Maximum Likelihood (REML) and Maximum Likelihood (ML) methods are the most popular among the classical methods to estimate the parameters of model in (6). By substituting whatever estimate of θ in the generalized least squares equation, we obtain $\widehat{\boldsymbol{\beta}}$. REML estimator of θ has better properties than ML, and we will use it generally, except when comparing nested models for the fixed effects, in which case we will use the ML (See Demidenko (2013), Littell *et al.* (1996) and Pinheiro & Bates (2000), for example).

There is a long debate about the *correct* approach for hypothesis testing on the LMM's parameters (Bates *et al.,* 2015; Crainiceanu & Ruppert, 2004; Drikvandi *et al.,* 2013; Littell *et al.,* 1996). For fixed effects the problem is that the degrees of freedom associated to t and F type statistic tests are not known in general settings of unbalanced data. Although computationally intensive approaches exist, the likelihood ratio test is asymptotically valid and commonly used. The problem is more serious for testing the inclusion/exclusion of random effects because the parameter value at the usual null hypothesis is in the boundary of the parametric region. Then, the LRT statistics does not follow a chi-square distribution with the regular degrees of freedom. It follows, instead, a mixture of chi-squared distributions. However, the components of the mixture may not be easily obtained. To overcome the difficulty Lee & Braun (2012) proposed two alternative tests based on permutations, the "*Best Linear Unbiased Predictors (BLUP) based permutation test*" and the "*likelihood ratio based permutation test*". Both approaches rely on weighted marginal residuals. In the applications in Section 4, we will use the BLUP-based permutation test. For further details on the methods, see Lee & Braun (ibid.) and Garcia (2019).

3. Materials and Methods

3.1 Fractional Polynomial Mixed Models

As with any model that includes a linear predictor, for continuous covariates, in the mixed model represented in (6), some non-linearity may be present, and transformations might be useful. Suppose $\mathbf{X} = (\mathbf{X}_L|\mathbf{X}_P)$ is the matrix of explanatory variables with $L + P$ columns. The $\mathbf{X}_L$ columns are dummies representing the intercept and, if available, the effects of qualitative variables (main and, possibly, interaction effects). The $\mathbf{X}_P$ columns refer to the continuous explanatory variables according to the FP degree each one requires. For example, suppose there are three continuous variables such that the first one does not require transformation, the second requires a FP1 and the

third a FP2. The i-th row of matrix $\mathbf{X}_P$ would be $(x_{i1}, x_{i2}^{(p_2)}, x_{i3}^{(p_{31})}, x_{i3}^{(p_{32})})^\top$, with relations (2) and (3) applied as appropriate, where p_2 is the transformation parameter for x_2 and p_{31} and p_{32} are the transformation parameters for x_3.

Then, the LMM in (6), including FP transformations and linear by linear interaction terms, may be written as

$$\mathbf{Y} = \mathbf{X}_L\boldsymbol{\beta}_L + \mathbf{X}_P\boldsymbol{\beta}_P + (\mathbf{X}_L \odot_c \mathbf{X}_P)\boldsymbol{\beta}_{LP} + (\mathbf{X}_L \odot_c \mathbf{Z})\mathbf{b}_L + (\mathbf{X}_P \odot_c \mathbf{Z})\mathbf{b}_P + \boldsymbol{\epsilon}, \tag{7}$$

where $(\mathbf{A} \odot_c \mathbf{B})$ is a matrix with columns referring to the required linear-by-linear interaction effects between variables belonging to $\mathbf{A}$ and $\mathbf{B}$ matrices (element-wise product of each pair of columns), $\mathbf{b}_L$ and $\mathbf{b}_P$ are possible random effect vectors associated to the effects of $\mathbf{X}_L$ and $\mathbf{X}_P$, respectively. This parametrization carries the same assumptions for model in (6) about the random terms since we can write $\mathbf{b} = (\mathbf{b}_L^\top, \mathbf{b}_P^\top)^\top$.

Fitting (6) (or (7)) involves the modeling of three components: the mean or fixed-effects structure, the random-effects structure and the serial correlation structure. Depending on the problem being dealt with, some terms in (7) may be unnecessary or, perhaps, some other added, for example, higher-order interactions. The representation in (7) has solely the aim to show that we need a model-building strategy to obtain a parsimonious fit since modeling one part may affect the modeling of the other parts.

3.1.1 Model-building and Variable Selection Strategy

One of the early questions for fitting mixed models, as well as mean and variance joint models, was from where we should start modeling since, in general, it is difficult, or perhaps, it does not make sense to model all parts simultaneously. That is particularly true for the LMM including FPs (7). The sequence of steps we propose is based on recommendations generally accepted as common sense, that is, first propose a well-specified mean model, then model the variance-covariance structure and then, with the variance-covariance model fixed, simplify the mean model. A well-specified model alludes to the maximal, complete or saturated mean profile model. However, the concept of a complete model is unattainable for observational studies with continuous covariates, or even for experiments with covariate levels not obeying the same spacing. Besides, before knowing which terms are required in $\mathbf{X}_P$ we do not know about the necessity of modeling the slope heterogeneity effects. Thus, the model-building strategy we propose is:

Step 1 Preliminary mean model: With a minimal random structure, i.e. $\mathbf{Z}$ specified for the sampling or experimental design only and $\mathbf{R} = \mathbf{I}$ (homogeneous and uncorrelated errors), find a well-specified mean model by searching for the *best* FP transformations in $\mathbf{X}_P$. To achieve this the MFP, MFPI and MFPIGen procedures are applied, as required, within the mixed model framework. As described in sub-section 2.1, the MFP procedure occurs in cycles and follows a backward strategy combined with the FSP algorithm. After application of MFP, the model building goes further by including interaction terms in case interactions are thought to contribute to the study at hand.

Step 2 Random terms model: With the mean model found in Step 1, search for the required random effects apart from those given by design. In this step we have the opportunity to model slopes' heterogeneity.

Step 3 Variance structure for $\boldsymbol{\epsilon}$: With the model built in the previous steps, search for the *best* variance-covariance matrix for $\boldsymbol{\epsilon}$. In this step, the requirement of an $\mathbf{R}$ matrix different from the identity is investigated.

Step 4 Model simplification: With the variance-covariance structures found in Step 2 and Step 3, repeat the MFP procedure, re-estimating the FP transformations and reduce the mean model if possible.

In Step 1 and Step 4, to compare the models, we use LR type tests since comparisons concern nested mean models. In Step 2 and Step 3 we use REML to estimate the parameters and Lee & Braun (2012)'s BLUP-based permutation tests for the inclusion/exclusion of random effects, one at a time. Additionally, in all steps, to guide the search for finding parsimonious models, we strongly recommend the graphical diagnostic tools proposed by Singer *et al.* (2017).

The resulting model from following these steps will be denoted as FPMM1(p) for first degree and FPMM2(p_1, p_2) for second degree fractional polynomial mixed models.

3.1.2 Data Sets

We applied the methods to three data sets from the literature. The first one is from a longitudinal study on serum bilirubin levels (μ mol/L) in healthy newborns. The study was conducted at the Escola Paulista de Medicina, UFSP and published in Rocha (2004). The sample considered here are bilirubin levels from 89 breastfeeding newborns measured every day from day 1 to day 6, and every other day until 12 days of age. The study aimed to develop a reference curve for bilirubin levels as a function of age but the data showed substantial variability among babies. The data is balanced with all subjects measured at the same time points.

The second example is from an experiment to study the relationship between the ultrafiltration rate and the transmembrane pressure that is applied to high flux membrane dialyzers. Dialyzers are used in hemodialyses to treat patients suffering from chronicle renal disease. Vonesh & Carter (1992) described the experiment and modeled the data by fitting a nonlinear mixed model. Twenty dialyzers were evaluated *in vitro* by using bovine blood. Two levels of blood flow rate were used, with ten dialyzers for each. The ultrafiltration rate in each dialyzer was measured at seven transmembrane pressure values with the set of values varying among dialyzers. This data set (`Dialyzer` available from package `nlme`) was subsequently analyzed by Littell *et al.* (1996) and Pinheiro & Bates (2000) by fitting a high-order polynomial exploring variance heterogeneity within dialyzers.

The third example is from an observational study on the relationships of willingness to pay for improvements in air quality presented by Harrison & Rubinfeld (1978). The data comprises 92 clusters, in the Boston Metropolitan Area, each subdivided into census tracts totaling 506 of them. Most of each cluster represents a town but Boston city is, itself, represented by 14 clusters. There are fourteen variables, from which eight are measured at the level of the census tract, including the outcome which is the logarithm of the median house price. The other variables are measured at the cluster level. This data set is very rich in the sense that it allows the illustration of several alternative methods to remedy the data violations on assumptions of the classical linear model. See, for example, Belsley *et al.* (1980), Longford (1993) and Singer *et al.* (2017) for some accounts. There are a few sources for the data set, with some variations. Here we use the version available from http://lib.stat.cmu.edu/datasets/.

3.1.3 Computational Resources

We wrote `R` functions for estimating the parameters of the fractional polynomial mixed models in Step 1 that wraps the `lme` function from the `nlme` package (Pinheiro *et al.,* 2018). Steps 2 and 3 use `lme` and its facilities to model variance-covariance structures and `R` code adapted from Lee & Braun (2012) for testing variance parameters when required. For model fitting diagnostic graph tools we acknowledge wide use of the function `lmmdiagnostic` freely available for download at http://www.ime.usp.br/~jmsinger/lmmdiagnostics.zip (see Singer *et al.* (2017) for guidelines for the tools). Our `R` code to run each application is available at https://github.com/edijane/mixed_models_using_fractional_polynomials.

4. Results and Discussion

In this section we apply the proposed LMM model-building strategy using FP to the three examples described in sub-section 3.1.2.

4.1 Bilirubin Concentrations in Newborns

The newborn bilirubin profiles (Figure 1) show that the relationship between concentration and the age is nonlinear with large variability among newborns justifying a model including subjects' random effects. Starting with the model with a minimal random structure we searched for the best

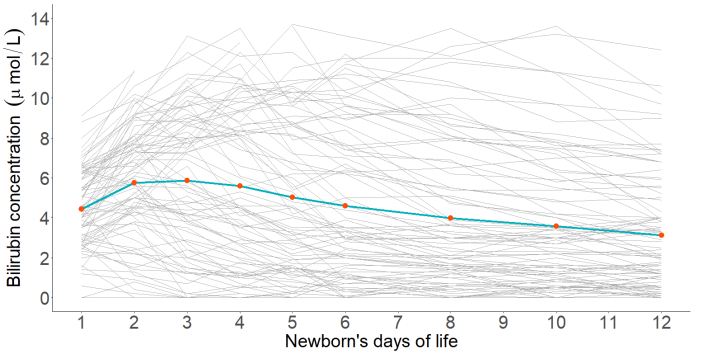

Figure 1. Bilirubin example – Bilirubin concentrations (μ mol/L) measured at 9 different ages (days) in 89 newborns.

FPMM1(p) and FPMM2(p_1, p_2) that could represent the curvature along with age. That is, we fitted the models:

$$y_{ij} = \beta_0 + \beta_1 x_{ij}^{(p)} + b_{0i} + \epsilon_{ij} \tag{8}$$

$$y_{ij} = \beta_0 + \beta_1 x_{ij}^{(p_1)} + \beta_2 x_{ij}^{(p_2)} + b_{0i} + \epsilon_{ij}, \tag{9}$$

where y_{ij} denotes the concentration of bilirubin in the i-th newborn at the j-th age x_{ij}, $b_{0i} \sim \mathcal{N}(0, \sigma_b^2)$ e $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$. The distributional parameters of the random terms of (8) and (9) are not necessarily the same, but for the sake of clean notation, this distinction will not be made explicit.

The best fittings, e.g. smallest deviance values, were given by FPMM1(2) and FPMM2(–0.5, –0.5) for the mean model in (8) and (9), respectively. Applying the FSP algorithm, as shown in Table 1, indicated that the FPMM2(–0.5, –0.5) is significantly better for simplifying the mean profile. The alternative model based on a conventional polynomial leads to the cubic model. In Figure 2 we plot the marginal fitted curves from both models from where we can see the FPMM2(–0.5, –0.5) accommodate better the observed mean. As a side check, we also found no evidence of lack-of-fit of the FPMM2(–0.5, –0.5) compared to the full mean profile model. We would expect that the curve governing bilirubin concentration holds an asymptote towards zero but there are no data available for the model to capture such behavior. Thus, we have found a well-specified model for the mean (Step 1, see parameter estimates in Table 2) and now move to model the random part in Step 2.

As anticipated by Figure 1, there seems to be a lot of variation among individual profiles. To take into account such heterogeneity, the models to fit in Step 2 are:

$$y_{ij} = \beta_0 + \beta_1 x_{ij}^{-0.5} + \beta_2 x_{ij}^{-0.5} \log(x_{ij}) + b_{0i} + \epsilon_{ij} \tag{10}$$

$$y_{ij} = \beta_0 + \beta_1 x_{ij}^{-0.5} + \beta_2 x_{ij}^{-0.5} \log(x_{ij}) + b_{0i} + b_{1i} x_{ij}^{-0.5} + \epsilon_{ij} \tag{11}$$

$$y_{ij} = \beta_0 + \beta_1 x_{ij}^{-0.5} + \beta_2 x_{ij}^{-0.5} \log(x_{ij}) + b_{0i} + b_{1i} x_{ij}^{-0.5} + b_{2i} x_{ij}^{-0.5} \log(x_{ij}) + \epsilon_{ij} \tag{12}$$

Table 1. Bilirubin example – Application of the FSP algorithm to compare the fitted models by using generalized likelihood ratio tests (GLRT)

Models	df	-2LL	Power	Comparison	GLRT	p-value
FPMM2	7	3288.088	-0.5; -0.5	FPMM2 vs Null	240.666	< 0.001
FPMM1	5	3352.211	2	FPMM2 vs LMM	84.552	< 0.001
LMM	4	3372.640	1	FPMM2 vs FPMM1	64.123	< 0.001
Null	3	3528.754	–			

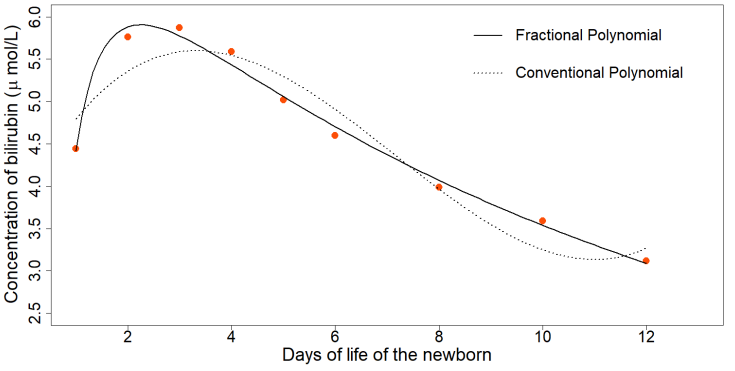

Figure 2. Bilirubin example – Predicted curves for the mixed-effects fractional polynomial and the mixed-effects conventional polynomial regression models.

The results from the BLUP-based permutation tests support that both coefficients related to terms $x^{-0.5}$ and $x^{-0.5}\log(x)$ vary between newborns, i.e. the two random effects are significant (p-value < 0.0001), and should be included in the individual profiles. Therefore, we continue the analysis with the model represented in (12), where

$$\mathbf{b}_i = \begin{bmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{bmatrix} \sim \mathcal{N}_3\left(\mathbf{0}, \sigma^2\mathbf{D}\right), \qquad \epsilon_{ij} \sim \mathcal{N}\left(0, \sigma^2\right),$$

for unstructured $\mathbf{D}$ and $\mathbf{R}_i = \sigma^2\mathbf{I}_9$. The REML estimates for all parameters are presented in Table 2, Step 2.

Residual plots (not shown) of the fitted model indicate variance heterogeneity of the measurement error at the individual level. Variances are smaller at 1, 8 and 10 days and, to model the heteroscedasticity, we grouped the age variable into two strata and fitted the model specified in (12), however, using a variance function with different variances for each stratification level of the age variable,

$$\text{Var}(\epsilon_{ij}) = \sigma^2\delta_{ij}^2, \tag{13}$$

where $\delta_{ij} = \delta_1 = 1$ for $j = 1, 7, 8$ (ages 1, 8 and 10 days) and $\delta_{ij} = \delta_2 > 0$ for $j = 2, 3, 4, 5, 6, 9$ (ages 2-6 and 12 days), for $\forall i$.

The very small p-value (< 0.0001) based on the LRT statistics indicates that the heteroscedastic model in (13) explains the data significantly better than the homoscedastic model in (12). The REML parameter estimates at this point are presented in Table 2, Step 3.

Diagnostic plots of the model in (13) are displayed in Figure 3. In the index plot of the modified Lesaffre–Verbeke measure (Figure 3a) four newborns are flagged. These are the subjects whose

bilirubin levels were still quite high at ages 10/12 days. The random effects do seem to follow a normal distribution, as depicted in Figure (3b) and the normal probability plot of the least confounded conditional residuals (Figure 3c) show no evidence against the adopted Gaussian assumption for the conditional error terms.

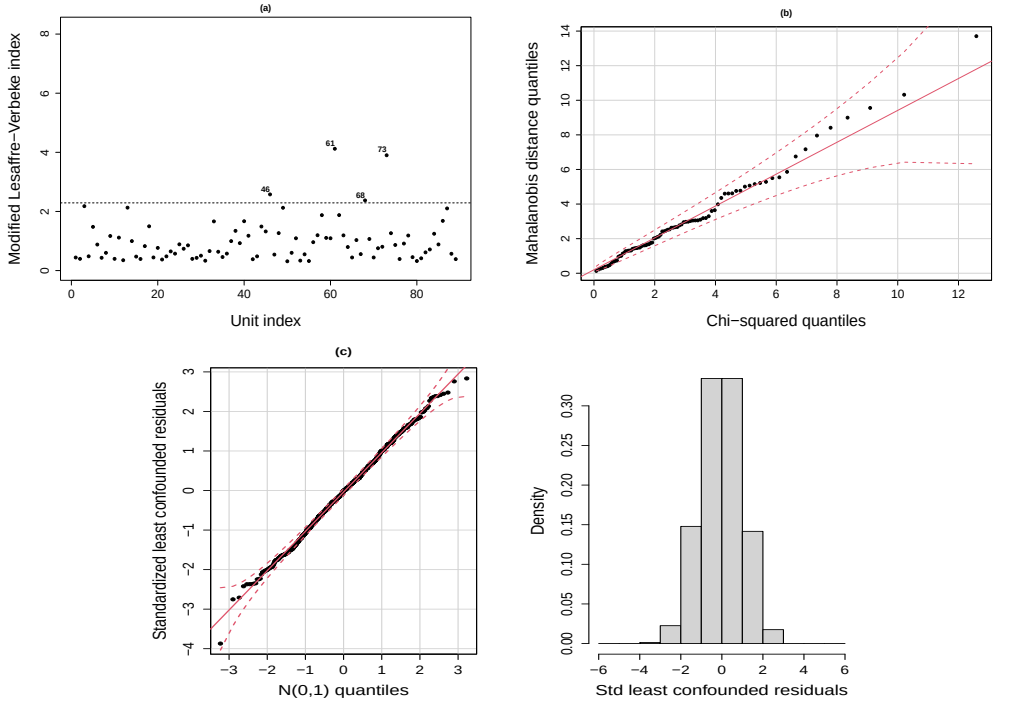

Figure 3. Bilirubin example – Diagnostic plots for the fitted model (13).

Now, with the variance-covariance structure settled, in Step 4 we perform a new search for the power in the set S, in the hope of reaching some simplification for the FP function. A FPMM with powers $p_1 = p_2 = 0$ was the one that presented the smallest deviance value ($-2LL = 2777.96$). The change in the transformation from FPMM2(–0.5, –0.5) to FPMM2(0, 0) results in a more parsimonious model, that is, a quadratic model in the log scale of age. Thus, the final model for the bilirubin concentration is the FPMM2(0; 0), described in equation form as follows:

$$\begin{aligned} y_{ij} = \beta_0 + \beta_1 \log(x_{ij}) + \beta_2 \log(x_{ij})^2 + b_{0i} + b_{1i} \log(x_{ij}) + b_{2i} \log(x_{ij})^2 + \epsilon_{ij}, \\ i = 1, \cdots, 89;\ j = 1, \cdots, 9, \end{aligned} \tag{14}$$

where

$$\mathbf{b}_i = \begin{bmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{bmatrix} \sim \mathcal{N}_3\left(\mathbf{0}, \sigma^2 \mathbf{D}\right), \qquad \epsilon_{ij} \sim \mathcal{N}\left(0, \sigma^2 \delta_{ij}^2\right),$$

with unstructured $\mathbf{D}$ and $\delta_{ij} = \delta_1 = 1$ for $j = 1, 7, 8$ (ages 1, 8 and 10 days) and $\delta_{ij} = \delta_2 > 0$ for $j = 2, 3, 4, 5, 6, 9$ (ages 2-6 and 12 days), for $\forall i$. The REML estimates for the parameters of model in (14) is displayed in Table 2.

Diagnostic plots for model in (14) show no relevant changes in relation to the diagnostic plots of the model in Step 3 (Figure 3). Furthermore, there is no evidence either of influential observations

or units in the fitting and the modeling using FP provided a reasonably precise curve for the mean model. It should be noted, however, that the variance component estimates, reflecting subjects variability, were very high (shown in Table 2) indicating, perhaps, that the data fail to allow the modeling of a reference curve of bilirubin concentrations for health babies.

Table 2. Bilirubin example – Parameter estimate and standard errors (SE) for the model at each step of the strategy. The mean model in all steps is FPMM2(p_1, p_2)

	Step 1		Step 2		Step 3		Step 4	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Power								
$(p_1; p_2)$	$(-0.5, -0.5)$		$(-0.5, -0.5)$		$(-0.5, -0.5)$		$(0, 0)$	
Fixed effects								
β_0	–7.86	0.85	–7.86	1.00	–7.77	1.02	4.49	0.21
β_1	12.28	0.73	12.28	1.08	12.20	1.10	2.38	0.42
β_2	10.32	0.69	10.32	1.03	10.23	1.02	–1.22	0.15
Var-Cov Struct								
Random effects								
σ_0^2	9.82		68.83		75.53		3.71	
σ_1^2			86.10		91.82		14.42	
σ_2^2			78.73		78.97		1.71	
Corr(b_0, b_1)			–0.99		–0.98		0.30	
Corr(b_0, b_2)			–0.87		–0.87		–0.43	
Corr(b_1, b_2)			0.89		0.89		–0.95	
Error								
σ^2	2.38		0.89		0.31		0.34	
δ_2					1.88		1.75	
Model fitting info								
AIC	3298.23		2860.04		2808.94		2805.77	
BIC	3321.64		2906.86		2860.44		2857.28	
Log-REML	–1644.12		–1420.02		–1393.46		–1391.89	

4.2 High-Flux Hemodialyzer Ultrafiltration Rates

The design in this experiment is a balanced, completely randomized scheme at the level of experimental units, the dialyzers, with the treatment, the blood flow rate, at two levels, "low" and "high" set to 200 and 300 dl/min, respectively. The response was the ultrafiltration rate (mL/hr) measured at each of seven sequentially increasing transmembrane pressure (dmHg) levels applied to the dialyzer, characterizing a longitudinal layout concerning pressure. The pressure levels varied slightly among dialyzers, as depicted in Figure 4. From the graphs, we also anticipate non-linearity and the indication of the flow rate effect interacting with pressure. As mentioned in subsection 3.1.2, Littell *et al.* (1996) and Pinheiro & Bates (2000) modeled the data using a fourth-order mixed model polynomial with variance heterogeneity. Here we investigate if the FP approach provides some simplification.

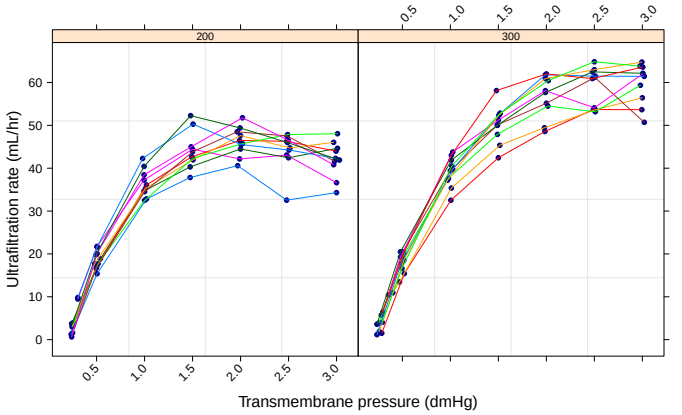

Figure 4. Ultrafiltration Rates example – Individuals plots.

Starting with the minimal random structure and restricting to second-degree FP functions only since we can already be quite sure no first-order FP would be able to account for the non-linearity present, in Step 1, we fitted the models:

$$\begin{gathered} y_{ij} = \beta_0 + \beta_1 x_{ij}^{(p_1)} + \beta_2 x_{ij}^{(p_2)} + \gamma_0 w_i + b_{0i} + \epsilon_{ij}, \qquad b_{0i} \sim \mathcal{N}\left(\mathbf{0}, \sigma_b^2\right), \quad \epsilon_{ij} \sim \mathcal{N}\left(0, \sigma^2\right) \\ i = 1, \cdots, 20; \; j = 1, \cdots, 7, \end{gathered} \tag{15}$$

where $w_i = 0$ if flow rate is low and $w_i = 1$ if flow rate is high and $p_1, p_2 \in S$. The set of powers $(1, 1)$ gave the best model ($-2LL = 837.28$). Once knowing the transformations, we included the interaction terms with the binary treatment in the mean model, e.g. we fitted

$$y_{ij} = \beta_0 + (\beta_1 + \gamma_1 w_i) x_{ij} + (\beta_2 + \gamma_2 w_i) x_{ij} \log(x_{ij}) + \gamma_0 w_i + b_{0i} + \epsilon_{ij}, \tag{16}$$

which resulted in a highly significant improvement ($\Delta D = 126.98$ on 2 df). This is the most complete mean model based on FP2 we can fit with these data. Its parameter estimates are presented in Table 3, Step 1.

The unit profiles (Figure 4) show that the modeling might benefit from random slopes' components, given their specific curves. With this in view, in Step 2 we start including components, first b_{1i} associated to the term of lowest order, fitting the model

$$\begin{gathered} y_{ij} = \beta_0 + (\beta_1 + \gamma_1 w_i + b_{1i}) x_{ij} + (\beta_2 + \gamma_2 w_i) x_{ij} \log(x_{ij}) + \gamma_0 w_i + b_{0i} + \epsilon_{ij} \\ \mathbf{b}_i = (b_{0i}, b_{1i})^\top \sim \mathcal{N}_2\left(\mathbf{0}, \sigma^2 \mathbf{D}\right), \qquad \boldsymbol{\epsilon}_i \sim \mathcal{N}_7\left(\mathbf{0}, \sigma^2 \mathbf{I}_7\right), \end{gathered} \tag{17}$$

for unstructured $\mathbf{D}$ and $\mathbf{b}_i \perp \boldsymbol{\epsilon}_i$. The BLUP-based permutation test showed strong evidence in favour of this term in the model. In the next step we fitted

$$\begin{gathered} y_{ij} = \beta_0 + (\beta_1 + \gamma_1 w_i + b_{1i})x_{ij} + (\beta_2 + \gamma_2 w_i + b_{2i})x_{ij}\log(x_{ij}) + \gamma_0 w_i + b_{0i} + \epsilon_{ij} \\ \mathbf{b}_i = (b_{0i}, b_{1i}, b_{2i})^\top \sim \mathcal{N}_3\left(\mathbf{0}, \sigma^2\mathbf{D}\right), \qquad \boldsymbol{\epsilon}_i \sim \mathcal{N}_7\left(\mathbf{0}, \sigma^2\mathbf{I}_7\right), \end{gathered} \tag{18}$$

for unstructured $\mathbf{D}$ and $\mathbf{b}_i \perp \boldsymbol{\epsilon}_i$, and tested the significance of variance component associated to b_{2i}, which was highly significant. Parameter estimates of the model are presented in Table 3, Step 2. Residual diagnostic for the fit in this step showed evidence of a relationship between conditional error variance and the explanatory variable pressure. Thus, in Step 3 we modeled the structure in $Var(\boldsymbol{\epsilon}_i) = \sigma^2\mathbf{R}_i$. The best variance function found was for $\mathbf{R}_i$ diagonal and variances governed by the power function given by

$$Var(\epsilon_{ij}) = \sigma^2 x_{ij}^{2\delta}. \tag{19}$$

The inclusion of error variance heterogeneity in the model resulted in good improvement of the fit. The parameter estimates are presented in Table 3, Step 3. Diagnostic inspection of the fitted model, in Figure 5, did not show any evidence of possible violations of assumptions.

We reached the last step of our approach, first checking if, given the variance-covariance structure well modeled, some simplification of the mean model would be recommended. Firstly, we re-estimated the power parameters of the FP functions. In this example, no modification of the powers was required. Then we tested the significance of the interaction terms by using the LRT statistics which did not show evidence in favour of the inclusion of γ_2 in the model. Therefore, this term was dropped from the model.

The parameter estimates of the final model are presented in Table 3, Step 4. It is noteworthy that the fitted mean curve using FP captures better the non-monotonic behavior present in the data of the ultrafiltration rate for high pressures, a deficiency of both, the nonlinear and the quartic polynomial models, recognized in Vonesh & Carter (1992) and Pinheiro & Bates (2000) (see Figure 6).

4.3 Boston Housing Values

The variables in the data set are described in the `Appendix`. In our analysis we decided not to use the variable `b` $(100 \times (\texttt{B} - 0.63)^2$, where `B` is the proportion of black population in the town). There are two sound justifications to exclude this variable from the modeling, the first is that the FP approach searches for the transformation of one variable given that the remaining variables are in the model and, it is not possible to recover back the original values of variable `B`, to properly account for its relationship with the response, the logarithm of the median house price. The second is that the background for the transformation proposed and distributed in the data set is not transparent and objective. The issue is addressed at https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_boston.html.

The fit of a multiple linear regression model to the response showed several violations of the classical linear model assumptions, residuals with long-tail distribution, lack-of-fit for several covariates (`CRIM`, `RM`, `DIS` and `LSTAT`), asymmetry and heterogeneity of residuals when comparing their patterns within-cluster units (graphs not shown).

A mixed model incorporating the grouping scheme of the data might remedy the problem of extra variability and correlations for observations in the same cluster. Then, we start, in Step 1, with the model, using standard notation, given by

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i b_{0i} + \boldsymbol{\epsilon}_i, \qquad b_{0i} \sim N(0, \sigma_0^2), \quad \boldsymbol{\epsilon}_i \sim N_{n_i}(0, \sigma^2\mathbf{I}_{n_i}), \tag{20}$$

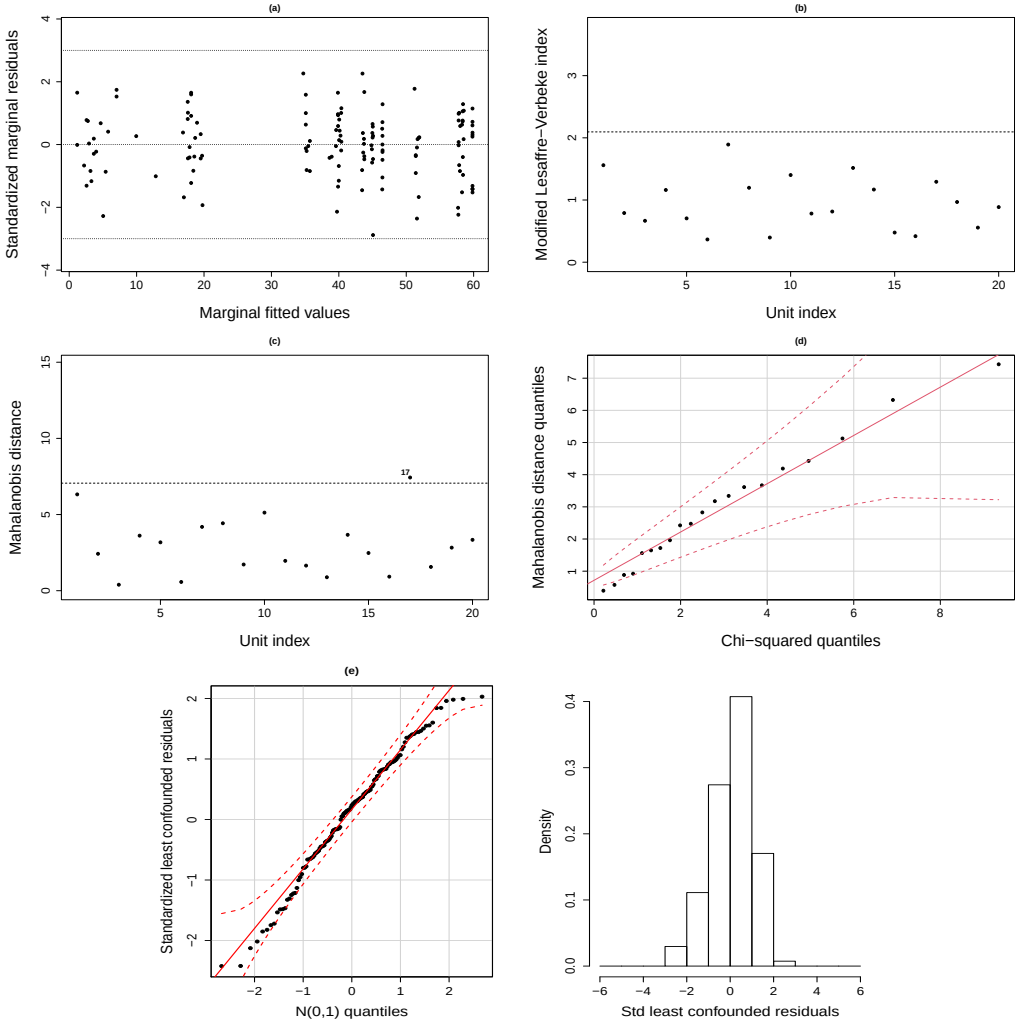

Figure 5. Ultrafiltration Rates example – Diagnostic plots for the fitted model (18) with within-group variance model given in (19).

for the i-th cluster, where n_i is the number of observations in cluster i, $i = 1, \cdots, 92$ and $\mathbf{Z}_i = \mathbf{1}_{n_i}$. The fitting of this model produces the variable visiting order, in step 2 of the MFP algorithm, for the search of FP transformations, that, hopefully will account for the non-linearity issues. Before starting, three variables were pre-transformed. `TAX` was just scaled down by dividing by 100. For `ZN` and `CRIM` that assume zero values we used expression in (21) as advised in Royston & Sauerbrei (2008)

$$\omega(x) = 0.2 + (1 - 0.2) \times \frac{x - \min(x)}{\max(x) - \min(x)}. \tag{21}$$

The visiting order list at the start is presented in the first column of Table 4. Thus, the first cycle begins with variable `LSTAT`, then moves to `RM` and so on. The MFP algorithm converged in two cycles and the final estimates are presented in Table 4, Step 1, where we see four variables were transformed, `LSTAT`, `RM`, `CRIM` and `DIS` and other four were dropped (`AGE`, `ZN`, `INDUS` and `CHAS`) and are not show in the Table.

Table 3. Ultrafiltration Rates example – Parameter estimates and standard errors (SE) for the model at each step of the strategy. The mean model in all steps is FPMM2(1,1)

	Step 1		Step 2		Step 3		Step 4	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Power								
$(p_1; p_2)$	$(1, 1)$		$(1, 1)$		$(1, 1)$		$(1, 1)$	
Fixed effects								
β_0	–21.90	1.19	–22.01	1.26	–23.74	0.96	–23.01	0.79
β_1	57.36	1.22	57.47	2.06	59.80	1.78	58.23	1.35
β_2	–33.18	0.88	–33.26	1.51	–35.15	1.36	–33.89	0.98
γ_0	–2.07	1.19	–1.67	1.81	–1.17	1.38	–2.67	0.79
γ_1	5.03	1.22	4.59	2.94	3.85	2.54	7.03	0.80
γ_2	1.61	0.88	1.91	2.14	2.56	1.94		
Var-Cov Struct								
Random effects								
σ_0^2	5.58		4.49		7.30		7.56	
σ_1^2			27.53		27.48		28.47	
σ_2^2			14.78		14.93		15.57	
Corr(b_0, b_1)			–0.99		–0.87		–0.87	
Corr(b_0, b_2)			0.94		0.79		0.80	
Corr(b_1, b_2)			–0.95		–0.95		–0.95	
Error								
σ^2	7.16		3.94		1.59		1.59	
δ					0.86		0.86	
Model fitting info								
AIC	726.29		685.93		647.40		650.29	
BIC	749.82		723.60		687.97		688.05	
Log-REML	–355.14		–329.97		–309.70		–312.14	

Step 2 of our modeling strategy is not relevant for this example since there are no other specific cluster effects that could be explored in the modeling and the random cluster effect is intrinsic to the data and should remain. Nonetheless, there is strong evidence in favour of the benefit of cluster effects in the model showed by the BLUP-based permutation test. However, diagnostic analysis for the fitted model indicated the random model for the error specified in (20) was not adequate, the main concern is the same $\text{Var}(\boldsymbol{\epsilon}_i) = \sigma^2\mathbf{R}_i = \sigma^2\mathbf{I}_{n_i}$ for all clusters. Such violation was expected since clusters vary a lot concerning size, the smallest clusters have just one observation while the largest has 30 observations. Intra-cluster correlations might also be cluster-specific. Thus, by inspecting the patterns of the modified Lesaffre-Verbeke Index, as advised in Singer *et al.* (2017), we proposed a stratification of clusters and incorporated the following variance-covariance structure for the errors

$$\boldsymbol{\epsilon}_i \sim N_{n_i}(\mathbf{0}, \sigma^2\delta_{si}\mathbf{R}(\rho)) \tag{22}$$

for $\delta_{1i} = 1$ and $\delta_{si} > 0$ for $s = 2, 3, 4, 5, 6$, $i = 1, 2, \cdots, 92$, that is, clusters were grouped into six strata for variance modeling. The model also includes the extra parameter $\rho \in (-1, 1)$ which is

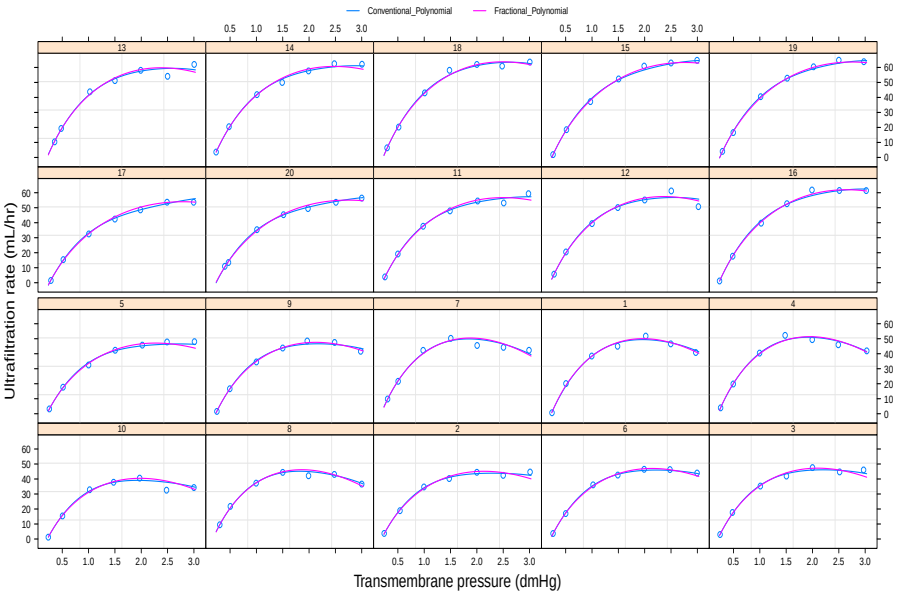

Figure 6. Ultrafiltration Rates example – Predicted ultrafiltration rates versus transmembrane pressure, by subject, from the final FPMM(1, 1) model with within-group variance model (19) and from the fourth-order polynomial with variance heterogeneity (Pinheiro & Bates, 2000).

the intra-cluster correlation parameter. Clusters in the same stratum share the same error variance-covariance parameters. To form the stratification we fitted several models sequentially, each time checking clusters flagged with non-adequate structure in the diagnostic graphs. The final model parameter estimates are presented in Table 4, Step 3. Once found a parsimonious structure, in Step 4, we updated the mean model by re-estimating the FPs. The algorithm converged in three cycles. The resulting model and parameter estimates are presented in Table 4, Step 4, with the variable visiting order in the second part of the Table. The diagnostic graphs for the fit of the final model is showed in Figure 7, from where we can be assured the data is parsimoniously modelled.

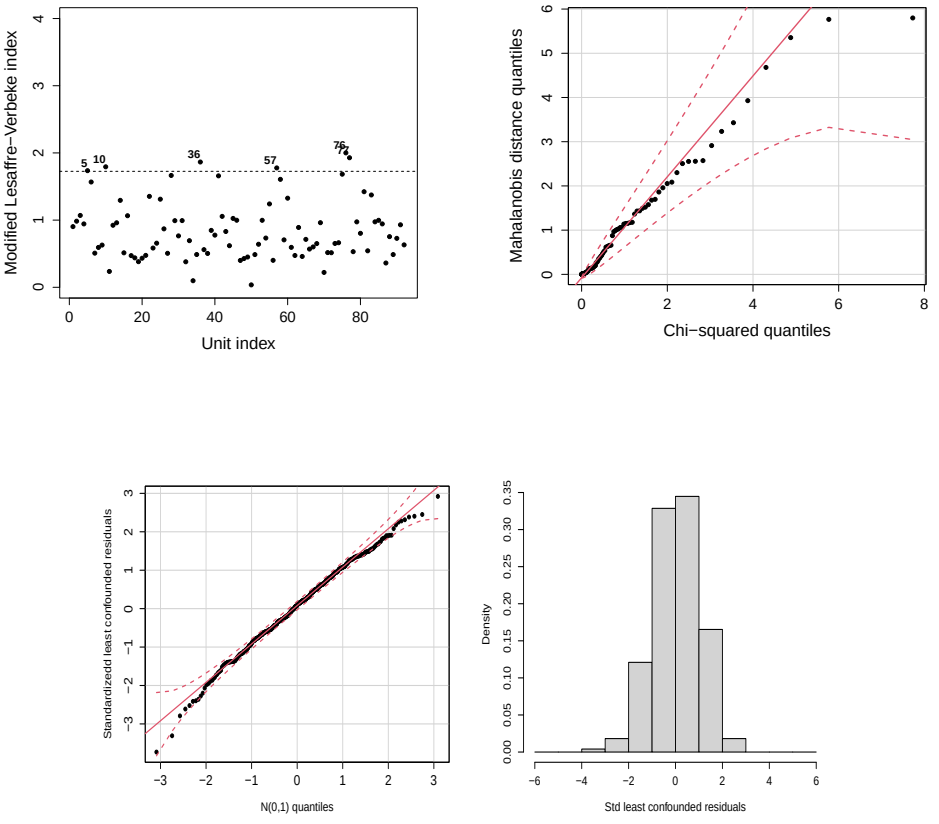

Figure 7. Housing Prices example – Diagnostic plots for the fitted model in Step 4 with within-cluster variance-covariance given in (22).

5. Conclusions

When modeling continuous variables we often encounter nonlinear relationships that should be dealt with empirically mainly in the case of multiple explanatory variables because of the unknown underlying mechanistic relations that generate the data. That is also, usually, the case of a unique explanatory when not much governing theory exists. FP models are regression-based powerful techniques to solve complex problems. Alternative empirical techniques to overcome non-linearity issues are conventional polynomial expansions and splines. However, conventional polynomials may produce fitted curves or surface hard to explain in practice and splines have their inherent issues of interpretability and the requirement of reasonably *dense* data (many points with small spaced intervals).

In this research, we explored the FP functions' usefulness in the context of linear mixed models. In the mixed-effects models, we have to deal with the modeling of different structures like the mean model, the variance-covariance model and the error measurement model. We have outlined a strategy for the application of FP in such a context. We have shown the modeling for three examples that benefited from the FP functions and produced parsimonious fittings. As in the case of fixed-effects models, in which FP functions extend easily for generalized linear models, our strategy is expected to extend to generalized linear mixed models to obtain a more parsimonious model for the

Table 4. Housing Prices example – Parameter estimates and standard errors (SE) for the model at each step of the strategy

	Step 1		Step 3		Step 4	
Parameters	Estimate	SE	Estimate	SE	Estimate	SE
Fixed effects						
Intercept	15.229	2.313	19.110	2.400	29.557	2.230
$LSTAT^{0.5}$	–0.200	0.012	–0.138	0.009		
$RM^{0.5}$	–9.738	1.790	–13.188	1.823		
$RM^{0.5}$ log(RM)	2.665	0.447	3.641	0.468		
$CRIM^{-0.5}$	0.628	0.067	0.629	0.085		
NOX	–0.929	0.170	–0.562	0.155		
RAD	0.022	0.004	0.017	0.003		
$DIS^{-0.5}$	0.698	0.137	0.278	0.108		
TAX	–0.070	0.019	–0.054	0.015		
PTRATIO	–0.029	0.008	–0.025	0.005		
Visiting order for Step 4						
$RM^{-0.5}$					–10.192	0.538
$RM^{-0.5}$ log(RM)					–30.370	2.856
$LSTAT^{0.5}$					–0.135	0.009
$CRIM^{-1}$					0.196	0.024
RAD					0.014	0.003
PTRATIO					–0.025	0.005
TAX^{-2}					0.787	0.177
NOX					–0.561	0.136
DIS					–0.018	0.005
Var-Cov Struct						
Random effects						
σ_0^2	0.013		6.283×10^{-5}		2.391×10^{-5}	
Error						
σ^2	0.017		0.010		0.009	
Variance function						
δ_2			1.697		1.672	
δ_3			1.803		1.845	
δ_4			1.865		1.977	
δ_5			2.809		2.799	
δ_6			4.307		4.720	
Correlation						
ρ			0.451		0.414	
Model fitting info						
AIC	–430.358		–652.123		–673.776	
BIC	–379.880		–576.404		–598.057	
Log-REML	227.179		344.061		354.888	

marginal link function. We argue the approach we propose is of vast application to solving practical problems of data analysis.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Ambler, G. & Royston, P. Fractional polynomial model selection procedures: investigation of Type I error rate. *Journal of Statistical Computation and Simulation* **69,** 89–108 (2001).
2. Ambler, G. & Benner, A. *mfp: Multivariable Fractional Polynomials.* R package version 1.5.2.2 (2022). https://CRAN.R-project.org/package=mfp.
3. Aregay, M., Shkedy, Z., Molenberghs, G., David, M.-P. & Tibaldi, F. Model-based estimates of long-term persistence of induced HPV antibodies: a flexible subject-specific approach. *Journal of Biopharmaceutical Statistics* **23,** 1228–1248 (2013).
4. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **67,** 1–48 (2015).
5. Belsley, D. A., Kuh, E. & Welsch, R. E. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity* (John Wiley & Sons, New York, 1980).
6. Box, G. E. P. & Tidwell, P. W. Transformation of the independent variables. *Technometrics* **4,** 531–550 (1962).
7. Crainiceanu, C. M. & Ruppert, D. Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66,** 165–185 (2004).
8. Demidenko, E. *Mixed Models: Theory and Applications with R* (John Wiley & Sons, Hoboken, New Jersey, 2013).
9. Drikvandi, R., Verbeke, G., Khodadadi, A. & Nia, V. P. Testing multiple variance components in linear mixed-effects models. *Biostatistics* **14,** 144–159 (2013).
10. Garcia, E. P. *Uso de polinômios fracionários nos modelos mistos.* Doutorado em Biometria (Instituto de Biociências/UNESP, Botucatu, 2019).
11. Harrison, D. & Rubinfeld, D. L. Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management* **5,** 81–102 (1978).
12. Lee, O. E. & Braun, T. M. Permutation tests for random effects in linear mixed models. *Biometrics* **68,** 486–493 (2012).
13. Littell, R. C., Milliken, G. A., Stroup, W. W. & Wolfinger, R. D. *SAS System for Mixed Models* (SAS Institute Inc, Cary, North Carolina, 1996).
14. Long, J. & Ryoo, J. Using fractional polynomials to model non-linear trends in longitudinal data. *British Journal of Mathematical and Statistical Psychology* **63,** 177–203 (2010).
15. Longford, N. T. *Random Coefficient Models* (Oxford University Press, Oxford, 1993).
16. Pinheiro, J. C. & Bates, D. M. *Mixed-Effects Models in S and S-PLUS* (Springer-Verlag, New York, 2000).
17. Pinheiro, J. C., Bates, D. M., Debroy, S., Sarkar, D. & Team, R. C. *nlme: Linear and Nonlinear Mixed Effects Models* R package version 3.1-137 (2018). https://CRAN.R-project.org/package=nlme.

18. R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing (Vienna, Austria, 2022). https://www.R-project.org/.

19. Regier, M. D. & Parker, R. D. Smoothing using fractional polynomials: an alternative to polynomials and splines in applied research. *Wiley Interdisciplinary Reviews: Computational Statistics* **7,** 275–283 (2015).

20. Rocha, F. M. M. *Seleção de estruturas de covariância para dados com medidas repetidas.* Mestrado em Estatística (IME/USP, São Paulo, 2004).

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

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

23. Royston, P. & Sauerbrei, W. *Multivariable model-building: a pragmatic approach to regression anaylsis based on fractional polynomials for modelling continuous variables* (John Wiley & Sons, Chichester, England, 2008).

24. Saurbrei, W., Meier-Hirmer, C., Benner, A. & Royston, P. Multivariable regression model building by using fractional polynomials: Description of SAS, STATA and R programs. *Computational Statistics & Data Analysis* **50,** 3464–3485 (2006).

25. Singer, J. M., Rocha, F. M. M. & Nobre, J. S. Graphical tools for detecting departures from linear mixed model assumptions and some remedial measures. *International Statistical Review* **85,** 290–324 (2017).

26. Tilling, K., MacDonald-Wallis, C., Lawlor, D. A., Hughes, R. A. & Howe, L. D. Modelling childhood growth using fractional polynomials and linear splines. *Annals of Nutrition and Metabolism* **65,** 129–138 (2014).

27. Vonesh, E. F. & Carter, R. L. Mixed-effects nonlinear regression for unbalanced repeated measures. *Biometrics* **48,** 1–17 (1992).

Appendix

Variable labels used in Subsection 4.3 and descriptions of the variables in the Boston Housing data set as retrieved from package `mlbench`: `cmedv`: median value of owner-occupied homes in USD 1000's, the *response* variable, we used the transformation (log(*cmedv*)) for the modeling; `CRIM`: per capita crime rate by town; `ZN`: proportion of residential land zoned for lots over 25,000 sq.ft; `INDUS`: proportion of non-retail business acres per town; `CHAS`: Charles River dummy variable (= 1 if tract bounds river; 0 otherwise); `NOX`: nitric oxides concentration (parts per 10 million); `RM`: average number of rooms per dwelling; `AGE`: proportion of owner-occupied units built prior to 1940; `DIS`: weighted distances to five Boston employment centres; `RAD`: index of accessibility to radial highways; `TAX`: tax full-value property-tax rate per USD 10,000; `PTRATIO`: pupil-teacher ratio by town; `b`= $1000(B - 0.63)^2$ where `B` is the proportion of blacks by town; `LSTAT`: percentage of lower status of the population.