Self-Modeling Regression with Random Effects Using Penalized Splines

Naomi S. Altman   Julio C. Villarreal

Cornell University

Abstract

In many longitudinal studies, the response can be modeled as a (discretely sampled) curve over time for each subject. Often these curves have a common shape function and individual subjects differ from the common shape by a transformation of the time and response scales. Lindstrom (1995) represented the common shape by a free-knot regression spline, and used a parametric random effects model to represent the differences between curves. We extend Lindstrom's work by representing the common shape by a penalized regression spline, and use a parametric random effects model to represent the differences between curves. The use of penalized regression splines allows for a generalization in the modeling, estimation, and testing of parameters and is easily implemented. An iterative two-step algorithm is proposed for fitting the model.

keywords: longitudinal data; semi-parametric; smoothing; penalized spline; random effects; functional data

1. Introduction

In many longitudinal studies, the response to be modeled is a continuous curve measured over time. Examples include EKG readings, growth curves, or serum glucose levels following a meal. The common factor in all of these examples is that the response curves share a similar shape e.g. the same number of extrema or inflection points located relatively near some common region. The data can be represented as Y_{ij} for $j=1, \dots ,n_i$ and $i=1, \dots ,m$; a sample of m individuals, curves, or experimental units, with the ith individual measured at n_i times, t_{ij}.

In this paper we will consider the Self- Modeling Regression (SEMOR) Model introduced by Lawton, Sylvestre and Maggio (1972). The model assumes a common underlying regression function μ_0 and transformations of both the time (t) and response (Y) axes. The SEMOR model is:

$$Y_{ij} = \phi_i\{ \mu_0 [\kappa_i(t_{ij})]\} + \varepsilon_{ij} \qquad (1)$$

where $\phi_i (x)$ is a monotone inverse link transforming the regression function and $\kappa_i (x)$ is a monotone transformation of the time axis.   One can choose to model $\phi_i (x)$ and $\kappa_i (x)$ parametrically or nonparametrically (depending on the regression problem) for specified $\mu_0 (t)$. The errors, ε_{ij} can also be modeled parametrically or nonparametrically and can be serially correlated or heteroskedastic, which is useful in the longitudinal setting. This paper will focus on nonparametric modeling of μ_0 and parametric modeling of $\phi_i (x)$ and $\kappa_i (x)$ with known correlation structure for the errors.

We give special attention to the case when $\phi_i (x)$ and $\kappa_i (x)$ are both are affine transformations: $\phi_i (x) = \alpha_{0i} + \alpha_{1i} x$ and $\kappa_i (x) == \beta_{0i} + \beta_{1i} x$. Then Equation (1) becomes the Shape Invariant Model (SIM) (Lawton et. al.1972):

$$Y_{ij} = \alpha_{0i} + A_{1i} \mu_0 (\beta_{0i} + B_{1i} t_{ij}) + \varepsilon_{ij} \qquad (2)$$

Since A_{1i} and B_{1i} should be positive, we express them as $A_{1i} = \exp(\alpha_{1i})$ and $A_{1i} = \exp(\beta_{1i})$

If one has physical or theoretical justification to pre-specify $\mu_0 (t)$ parametrically, a number of techniques are available. These include Bates and Lindstrom (1986) for the case when the parameters are fixed and Lindstrom and Bates (1990), Davidian and Giltinan (1995) and Pinheiro and Bates (1995) when parameters are random. The semi-parametric SEMOR model allows flexible modeling by estimating $\mu_0 (t)$ nonparametrically.

Several different approaches have been studied in fitting the SIM model. Lawton et. al 1972, Kneip and Gasser 1988, Kneip and Engel 1995 considered $\theta_i = (\alpha_{0i}, \alpha_{1i}, \beta_{0i}, \beta_{1i})^\top$ to be fixed effects. We will follow Lindstrom (1995) in modeling θ_i as mixed effects.

We model the regression function $\mu_0(t)$ by a penalized regression spline (p-spline) implemented as a linear mixed effects model. One major gain is that the smoothing parameter can be interpreted as a ratio of variance components that can readily be estimated along with the other parameters of the model. This allows us to use estimation methods based on maximum likelihood.

Estimation of the regression function μ_0, mixed effects θ_i, and variance components will be done via a two-step iterative algorithm. The algorithm consists of a linear mixed effects step (p-spline fitting) and a nonlinear mixed effects step (mixed effects and variance component estimation).

2. A Brief Introduction to P-Splines

Using penalized regression splines (Eilers and Marx 1992, 1996; Ruppert and Carroll, 1997) for nonparametric modeling is a compromise between smoothing splines and free-knot regression splines.

Suppose we have the following observations (Y_i, t_i) for i=1, ..., n and t_i in the interval (a,b) modeled by:

$$Y_i = \mu(t_i) + \varepsilon_i \qquad (3)$$

where μ is the unknown smooth regression function and the ε_i are independently and identically distributed random variables with zero mean and constant variance σ_ε^2. (The i.i.d. assumption can be relaxed to handle correlated or heteroskedastic errors). The regression spline model of order $p \geq 1$ is:

$$\mu_o(t;\boldsymbol{\beta}) = \beta_0 + \beta_1 t + \beta_2 t^2 + ... + \beta_p t^p + \beta_{p+1}(t-\tau_1)_+^p + ... + \beta_{p+k}(t-\tau_k)_+^p \qquad (4)$$

where the parameter to be estimated is $\beta=(\beta_0, ..., \beta_{p+k})$ and $\{\tau_1, ..., \tau_k\}$ are k fixed knots with $a \le \tau_1 < ... < \tau_k \le b$ and $(x)^p_+ = x^p I_{\{x \ge 0\}}$. The knots are pre-selected, typically with large k. To control smoothness, a roughness penalty function P is placed on the regression coefficients of the form $P(\beta_{p+1}, ..., \beta_{p+k}) \le \lambda$. The purpose of the penalty is to prevent the coefficients representing jumps in high order derivatives from becoming too large.

Because the use of the penalty function decreases the problem of overfitting, knot selection is less critical than for regression splines (Eilers and Marx 1996; Ruppert and Carroll 1997) and is usually done non-adaptively. We use a relatively large number of knots with a quadratic penalty function (Ruppert and Carroll 1997). The estimator β_λ^* is defined as the minimizer of

$$\sum_{i=1}^{n}\{Y_i - \mu_o(t;\boldsymbol{\beta})\}^2 + \lambda \sum_{j=1}^{k}\beta_{p+j}^2 \qquad (5)$$

where λ is the smoothing parameter.

We can restate the model in matrix form as follows. Let $\mathbf{Y} = (Y_1, ..., Y_n)^T$ and let X(t) be the design matrix with the ith row of X(t) equal to

$$x_i(t) = \{1, t_i, t_i^2, ..., t_i^p, (t_i - \tau_i)_+^p, ..., (t_i - \tau_k)_+^p\}$$

Let D be the diagonal matrix with the first p+1 entries zero and the remaining k entries equal to one. The estimator β_λ^* is the minimizer of the following equivalent expression in matrix notation:

$$(Y - X(t)\beta)^T (Y - X(t)\beta) + \lambda\beta^T D\beta \qquad (6)$$

with

$$\beta^*(\lambda) = (X(t)^T X(t) + \lambda D)^{-1} X(t)^T Y$$

Note that β_λ^* is a ridge regression estimator and is readily computed for any fixed choice of λ. There are several approaches to estimating λ such as cross-validation and generalized cross validation. However, it is convenient to use the generalized maximum likelihood (GML) approach (Wahba, 1985), which can be derived via a linear mixed effects (LME) model.

Consider partitioning the parameter β and the design matrix X respectively into two submatrices, $\beta = [(\gamma; u]^T$ and $X = [X^*, Z]$, with $\gamma = (\gamma_0, ..., \gamma_p)^T$, $u = (u_1, ..., u_k)^T$ and the ith rows of X^* and Z given by

$$x_i^* = \left(1, t_i, t_i^2, ... t_i^p\right) \qquad (7)$$

$$z_i = \left((t_i - \tau_1)_+^p, (t_i - \tau_2)_+^p, ... (t_i - \tau_k)_+^p\right)$$

respectively. Then consider the following LME model:

$$Y = X^*\gamma + Zu + \varepsilon \qquad (8)$$

where ε is distributed multivariate normal with mean equal to the zero vector and covariance matrix $\sigma_\varepsilon^2 I$, γ is the fixed effect and u is the random effect, independent of ε, with $u \sim$ i.i.d $N(0, \sigma_u^2 I)$. The maximum likelihood estimators of γ and u are the minimizers of

$$(Y - X^*\gamma - Zu)^T (Y - X^*\gamma - Zu) + \frac{\sigma_\varepsilon^2}{\sigma_u^2} u^T u$$

(Searle, Casella and McCulloch, 1992). This expression has the same form as in (6) where the smoothing parameter λ is replaced by the ratio of the variance components $\sigma_\varepsilon^2/\sigma_\gamma^2$ since $X\beta = X^*\gamma + Zu$ and $\beta^\tau D\beta = u^T u$. The GML method estimates the smoothing parameter by the estimated variance ratio.

The advantages of using LME models for fitting a p-spline are quite clear: the mixed model is readily understood and λ is automatically estimated. Standard statistical software can be used for

fitting LMEs, such as lme in S+ or PROC MIXED in SAS. By using the LME p-spline for SEMOR we have a relatively quick and easy method for estimating the regression function for a fixed number of knots. P-spline fitting also will allow for a generalization of the SEMOR model to include nonlinear transformation functions ϕ and κ.

3. Model Formulation

3.1 General Set-up of Model

We extend SIM model (2) to consider mixed effects for the scaling and shifting parameters.

$$Y_{ij} = \alpha_0 + a_{0i} + \exp[(\alpha_1 + a_{1i})] \mu_0 [\beta_0 + b_{0i} + \exp[(\beta_1 + b_{1i})] t_{ij}] + \varepsilon_{ij}$$

for i=1, ... ,m; j=1, ... ,n_i, t_{ij} in (a,b) where (α_0, α_1, β_0, β_1) are fixed effects and (a_{0i}, a_{1i}, b_{0i}, b_{1i}) are random effects. To impose identifiability in the SIM model, we constrain (α_0, α_1, β_0, β_1) to be (0,0,0,0), the mixed model equivalent to setting the sum of each parameter equal to zero as suggested by Kneip and Engel (1995). The random effects (a_{0i}, a_{1i}, b_{0i}, b_{1i}) are modeled as independently distributed multivariate normal. We consider the covariance matrix of the random effects to be diagonal, although they need not be.

Let $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, ..., Y_{in_i})$ for $i = 1, 2, ..., m$; $j = 1, 2, ..., n_i$. The SIM mixed model including the penalized spline is:

$$\mathbf{Y}_i = (\alpha_0 + \alpha_{0i}) + \exp(\alpha_1 + \alpha_{1i}) \{x_i^* [(\beta_0 + b_{0i}) + \exp(\beta_1 + \beta_{1i}) t] \gamma + \qquad (9)$$

$$Z_i [(\beta_0 + b_{0i}) + \exp(\beta_1 + \beta_{1i}) t] u\} + \varepsilon_i$$

$$\varepsilon_i \sim i.i.d.N\left(0_{n_i}, I_{n_i}\sigma_\varepsilon^2\right) \quad u_i \sim i.i.d.N\left(0, \sigma_u^2\right)$$

$$a_{0i} \sim i.i.d.N\left(0, \sigma_{a_0}^2\right) \qquad a_{1i} \sim i.i.d.N\left(0, \sigma_{a_1}^2\right)$$

$$b_{0i} \sim i.i.d.N\left(0, \sigma_{b_0}^2\right) \qquad b_{1i} \sim i.i.d.N\left(0, \sigma_{b_1}^2\right)$$

where $\mathbf{0}_n$ is the zero vector of length n_i and the rows of $\mathbf{X}^*(t)$ and $\mathbf{Z}(t)$ are given by   (7).

We can handle this estimation in a two step approach: one is estimating the parameters for the p-spline that model μ_0, and the other is estimating the variance components, fixed effects and the Best Unbiased Predictors (BUPS) which enter the model nonlinearly. Note that Equation (9) can be placed in a general Bayesian context with non-normal priors for the parameters. We can readily extend to parametric structure for error variance.

3.2 Algorithm for SIM

From the model we have formulated, we can write the log likelihood function:

$$l(\cdot\,;\mathbf{Y}) = Log \int \{f(\mathbf{Y}\,|\,\mathbf{u},\mathbf{a}_0,\mathbf{a}_1,\mathbf{b}_0,\mathbf{b}_1)f(\mathbf{u})f(\mathbf{a}_0)f(\mathbf{a}_1)f(\mathbf{b}_0)f(\mathbf{b}_1)\}d\mathbf{u}d\mathbf{a}_0 d\mathbf{a}_1 d\mathbf{b}_0 d\mathbf{b}_1 \qquad (10)$$

This is a different approach from the earlier literature for SIM modeling since we have an explicit likelihood function. While estimation of the parameters might be done by maximizing equation (10) directly, this is computationally difficult.

Equation (2) suggests the following algorithm. Let

$$\mathbf{\theta}_i = (\theta_{1i},\theta_{2i},\theta_{3i},\theta_{4i})$$
$$= (\alpha_0 + a_{0i},\alpha_1 + a_{1i},\beta_0 + b_{0i},\beta_1 + b_{1i})$$

and

$$\mathbf{\theta} = (\mathbf{\theta}_1^T,\mathbf{\theta}_2^T,...,\mathbf{\theta}_m^T).$$

If we knew the true shape of μ_0, then we could use nonlinear mixed effects (NLME) methods to estimate $\mathbf{\theta}$. Similarly if we knew $\mathbf{\theta}$, we could estimate μ_0 through linear mixed effects (LME). Hence the following iterative algorithm is proposed.

SIM Algorithm

Step 0 Choose initial estimates of $\theta_i^{(0)} = (0,0,0,0)$. Set k=0.

Step 1 Transform data and time with

$$Y*_i^{(k)} = (Y_i - \theta_{0i}^{(k)})/\exp(\theta_{1i}^{(k)}) \text{ and } t_i*^{(k)} = \theta_{3i}^{(k)} + \exp(\theta_{4i}^{(k)}) t_i$$

Step 2 Using LME, estimate $\gamma^{(k)}$ and $u^{(k)}$ by fitting:

$$Y*^{(k)} = X(t_i*^{(k)})\gamma^{(k)} + Z(t_i*^{(k)}) u^{(k)}$$

Step 3 Using NLME, estimate $\theta^{(k+1)}$ by fitting the model

$$Y_{ij} = \alpha_0 + a_{0i} + \exp(\alpha_1 + a_{1i})\{X(\beta_0 + b_{0i} + \exp(\beta_1 + b_{1i}) t_{ij})\gamma^{(k)} + Z(\beta_0 + b_{0i} + \exp(\beta_1 + b_{1i}) t_{ij}) u^{(k)}\} + \varepsilon_{ij}$$

Check for convergence. Else normalize the parameters with the transformations (using the notation of Equation 2):

$$a_{1i}*^{(k+1)} = \exp(A_{1i}*^{(k+1)})$$

$$A_{1i}*^{(k+1)} = A_{1i}^{(k+1)}/(\Sigma_j A_{1j}^{(k+1)}/m)$$

$$\cdot a_{0i}*^{(k+1)} = a_{0i}^{(k+1)} - A_{1i}*^{(k+1)}*\Sigma_j a_{0j}^{(k+1)}/m$$

$$b_{1i}*^{(k+1)} = \exp(B_{1i}*^{(k+1)})$$

$$B_{1i}*^{(k+1)} = B_{1i}^{(k+1)}/(\Sigma_j B_{1j}^{(k+1)}/m)$$

$$b_{0i}*^{(k+1)} = b_{0i}^{(k+1)} - B_{1i}*^{(k+1)}*\Sigma_j b_{0j}^{(k+1)}/m$$

and set $\theta_i^{(k+1)} = (a_{0i}*^{(k+1)}, a_{1i}*^{(k+1)}, b_{0i}*^{(k+1)}, b_{1i}*^{(k+1)})$ Go to step 1.

Step 1 is suggested from the graphical diagnostics presented by Lindstrom (1995). The normalization in Step 3, similar to Kneip and Engel (1995), forces the fixed effects to be (0,0,0,0).

Other methods to select the initial values include choice of the median curve to estimate $\gamma^{(0)}$ and $u^{(0)}$ or use of structural points as in Kneip and Engel (1995) to choose better initial parameter estimates. Other normalizations include setting the maximum and minimum values of μ_0.

There are several convergence criteria that can be used to terminate the algorithm. Lawton et al (1972), terminated the fitting algorithm when MSE of the parameters converged. We terminated the algorithm when the change in the log-likelihood in the nonlinear mixed effects step converges.

3.3 Convergence of Algorithm

If we knew θ, Step 1 would align the transformed response and transformed time (Y^* and t^*) to the common shape or regression function. The transformation removes subject differences leaving us a data set with only one source of variation, σ_ε^2. We then need only fit a p-spline. Although the theoretical rate of convergence of the p-spline as a sieve estimator has not to our knowledge been established, it should be similar to the rate for B-splines, which was determined by Shen and Wong (1994). Simulation results in Villarreal (2001) support the convergence of the p-spline with GML selection of the smoothing parameter if the number of knots increases with the sample size.

In general, by transforming the data and pooling across curves to estimate μ_0, we effectively increase the number of design points to the total number of observations. Hence, we have a large sample size for estimating the regression shape. As long as the NLME routine converges the algorithm appears to converge.

4. Simulation Results

To determine the efficacy of the computational method, we performed a small simulation study, using 3 underlying curves for the common shape. These were selected to resemble shapes commonly found in growth and pharmacokinetic models.
The three functions are:

$$\mu_{0(1)}(t) = te^{-t}$$

$$\mu_{0(2)}(t) = \frac{1}{\sqrt{2\pi}(.25)}e^{-\frac{1}{2(.25)}(t-1.5)^2} + \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(t-3.5)^2}$$

$$\mu_{0(3)}(t) = \frac{e^{2(t-2.5)}}{e^{2(t-2.5)} + e^{-2(t-2.5)}}$$
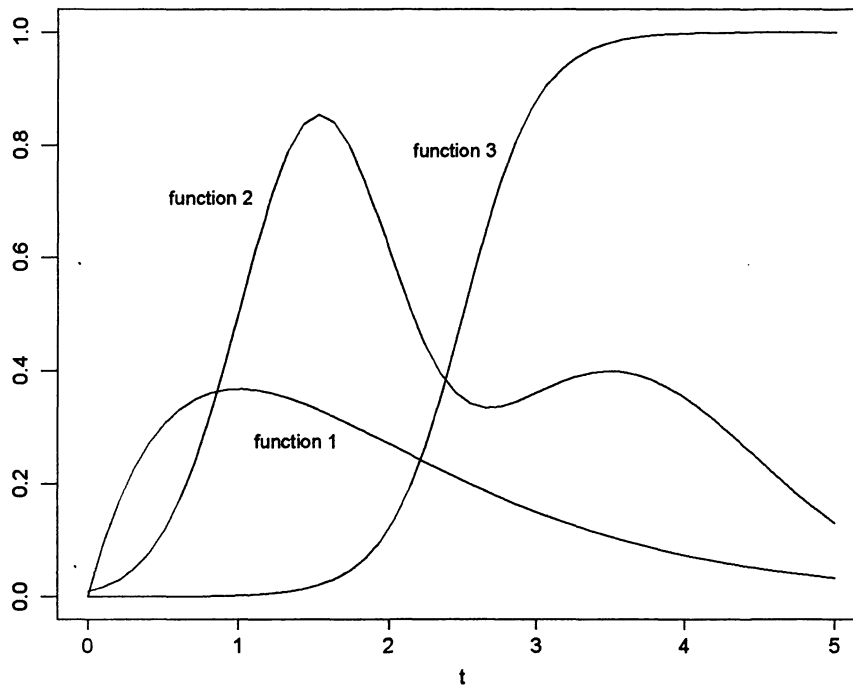
These are plotted in Figure 1.

Figure 1: Regression shapes used in the simulation study.

The two sources of error in the model are the between curve variation modeled by the random effects, and the within curve variation modeled by the error. We use two levels of variation: large and small. These are summarized in Table 1. We also consider two levels of observations per curve (n=20, 30) and 3 levels of numbers of curves (m=10, 30, 50). For each combination of factors, we generated 50 data sets.

	Case I: Small Error	Case II: Large Error

σ_{a0}	0.10	0.30
σ_{a1}	0.05	0.10
σ_{b0}	0.15	0.20
σ_{b1}	0.0375	0.1875
σ_{ϵ}	0.02*range(f)	0.10*range(f)

Table 1: Simulated Variance Components

The tolerance set for convergence was set at a change in the log-likelihood of the nonlinear step of 10^{-4} with a maximum of 50 iteration steps. For each sample, we record the estimates of the random effects and the mean squared error (MSE) in estimating the curves (at the design points).

Although the p-spline estimator is not sensitive to knot placement, in practice, we need to determine the number of knots required for fitting the curves. It is clear that for the case of a single curve, with the number of design points increasing to a dense set, the number of knots must increase with the sample size to achieve consistency. A small simulation study showed that, for sample size 100 on a single curve, increasing beyond 10 knots did not improve the fit. However, we need to adapt to multiple curves and sample sizes. In the first iteration, the number of transformed time points is either 20 or 30, and we use 4 and 6 knots respectively. After the first iteration, the number of transformed time points is the product of number of points per curve and the number of curves. For 20 points per curve we used 8,12 or 16 knots for 10, 30 or 50 curves respectively. For 30 points per curve we used 12, 18 or 24 knots for 10, 30 or 50 curves respectively.

Fitting was done using the lme and nlme procedures in Splus. Convergence of the algorithm depends on convergence of all steps: lme to fit the shape, nlme to fit the variance components, and the iterations between lme and nlme. The lme step always converged. The nlme step was more problematic. There were frequent failures, even when the iterative procedure appeared to be close to convergence. This appears to be a problem with nlme rather than with the iterative procedure. When the lme and nlme both converged, the algorithm usually converged quickly. Curves 1 and 3 always converged. Curve 2 did not converge (in 50 iterations) for 5 of the 300 samples.

For the purposes of the simulation study, we generated the required number of curves until we had fitted 50 samples without a failure of nlme. Simulations in which the algorithm failed to converge at the iterative step are included in the summaries. Simulations in which nlme failed are not included.
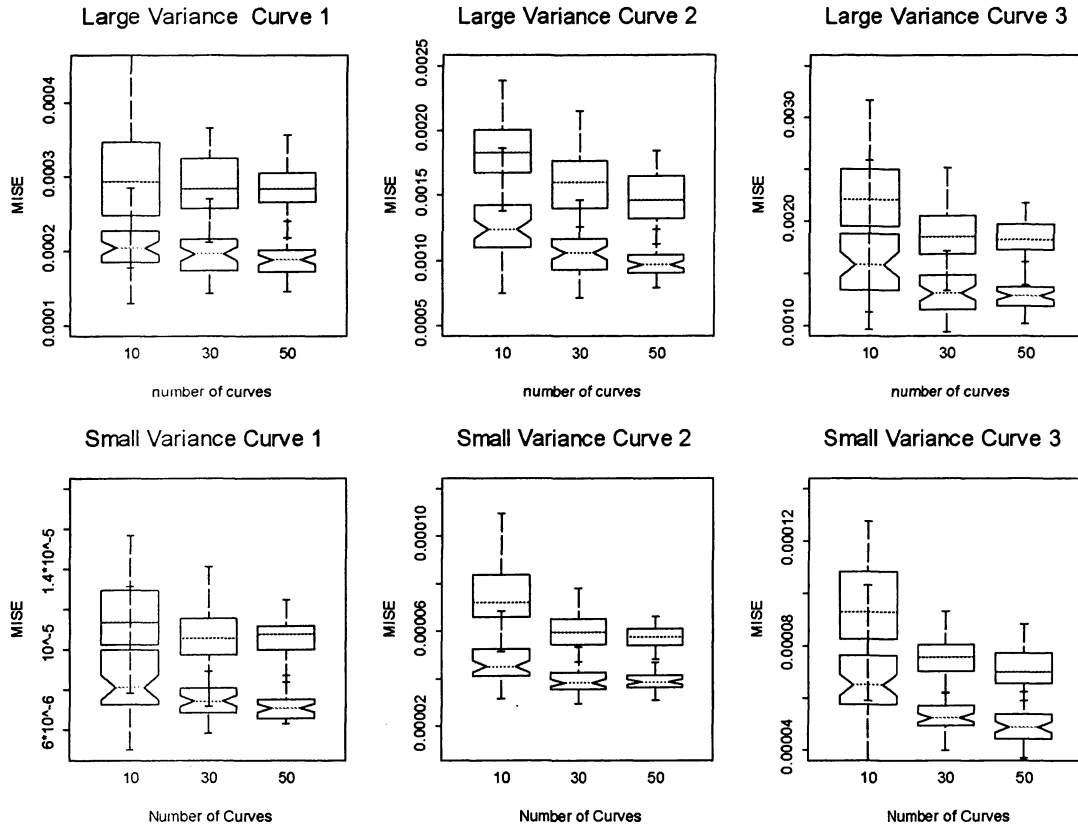

Figure 2: *MISE of fit for each simulation. Notched boxes indicate 30 data points per curve. Plain boxes indicate 20 data points per curve. Notice that the MISE improves with both the number of data points per curve and the number of curves per set.*

Figure 2 displays boxplots of the MISE under all conditions of sample size and variance components. The MISE decreases as a function of both the number of data points per curve and the number of curves per set.

To assess the accuracy of the estimates for the variance components, we look at box plots for each variance estimate as a function of the number of curves. These are displayed as the ratio SD(estimated)/SD(realized) where the SD(estimated) is the SD of the predicted parameter value and the SD(realized) is the empirical SD of the parameters generated by the model.
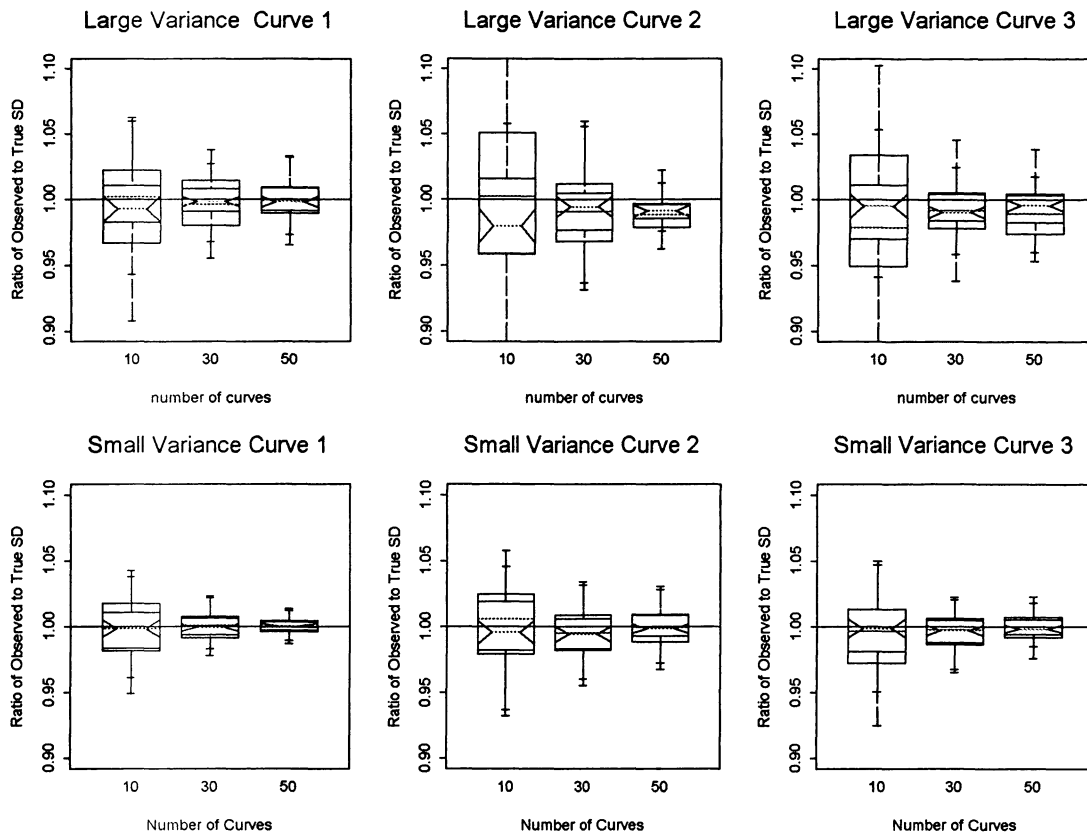
Figure 3: SD ratio for a_0. The numerator is the empirical SD of the predicted values of a_0 The denominator is the empirical SD of the generated values.

Figure 3 displays boxplots of the SD ratio for parameter a_0. Plots for b_0 are similar. The variance components are estimated quite well. There appears to be a small downwards bias. The ratios are more spread out when there are 20 points per curve (rectangular boxes) than when there are 30 (notched boxes). The spread of the ratios decreases as the number of curves increases.

Figure 4 displays boxplots of the SD ratio for parameter a_1. Plots for b_1 are similar. There is some downwards bias in the estimated SDs, especially when the true variance components are large. The ratios are more spread out when there are 20 points per curve (rectangular boxes) than when there are 30 (notched boxes) and when the variance components are large. The spread of the ratios decreases as the number of curves increases.
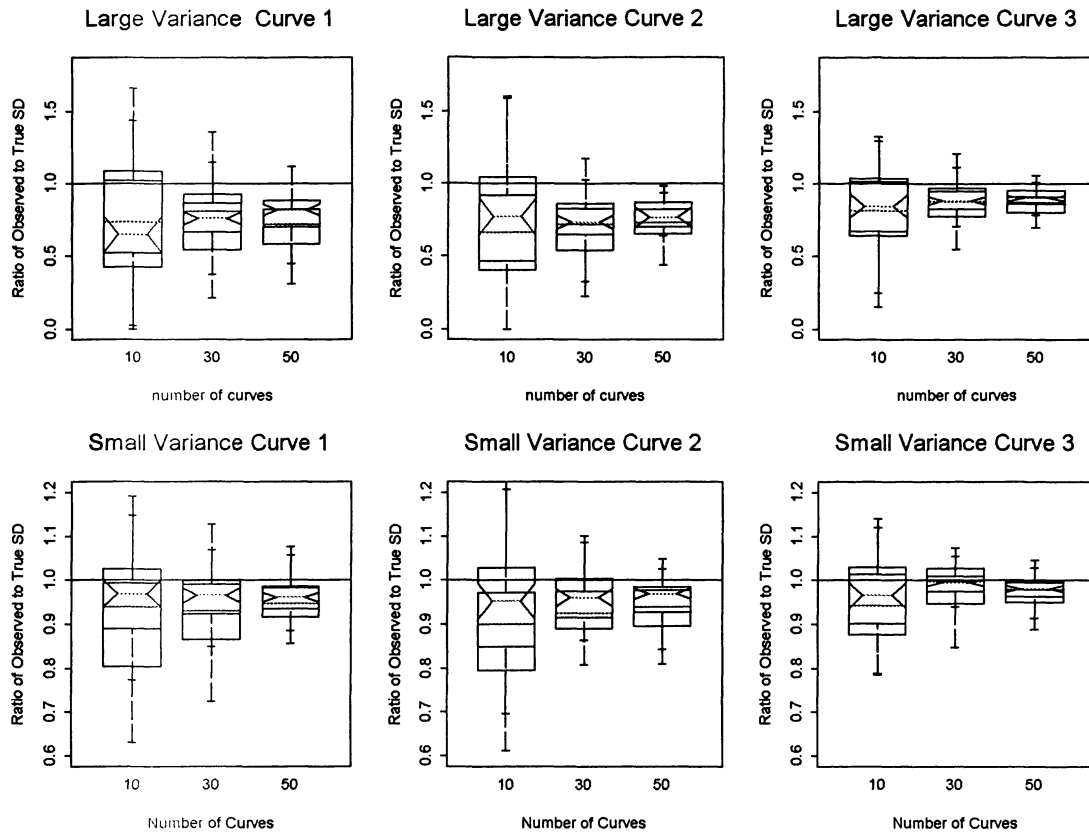
Figure 4: *SD ratio for a_1. The numerator is the empirical SD of the predicted values of b_1 The denominator is the empirical SD of the generated values.*

5. Application to Spirometer Data

For a real application, we considered the spirometer data used by Lawton et. al. (1972) and Lindstrom (1995). We fitted the SIM model, using 10 equally spaced knots. The data are displayed in Figure 5 with a fit to each curve (computed individually). The rescaled data are displayed in Figure 6, with the fit of the common shape. The lower right panel is a plot of all the rescaled data on the common time and response scale.

It is evident from both figures that the error variance is small for these data. The individual curves display a sigmoid shape. Figure 6 shows the remarkably good fit of the SIM model. The rescaled data (using the predicted parameters) have very little spread. The common shape is clear.

The final estimates for the regression mean function fit the data very closely. Lindstrom (1995) also analyzed these data. She removed the two initial points due to the assumption that they were observed before exhalation began. We use the two initial points and are able to fit the data very well. The time shift constant b_0 captures the possibility of observations prior to exhalation. We also see from the variance component estimates that the variability in the curve for the y- direction is largely due to the multiplicative factor a_1.
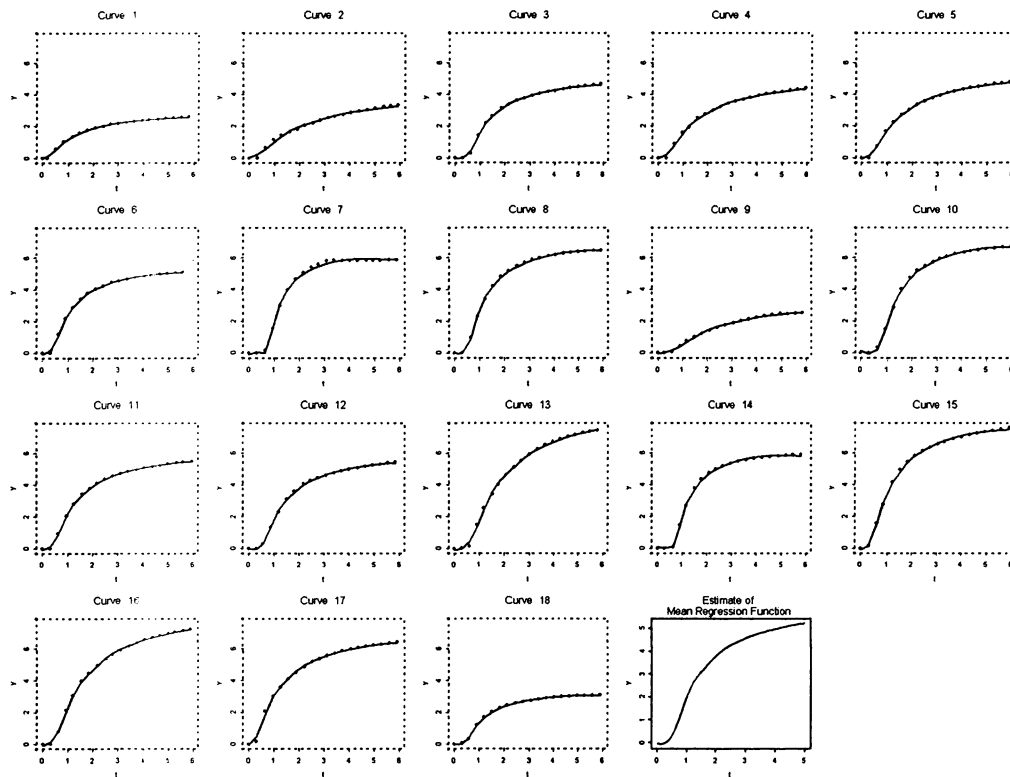

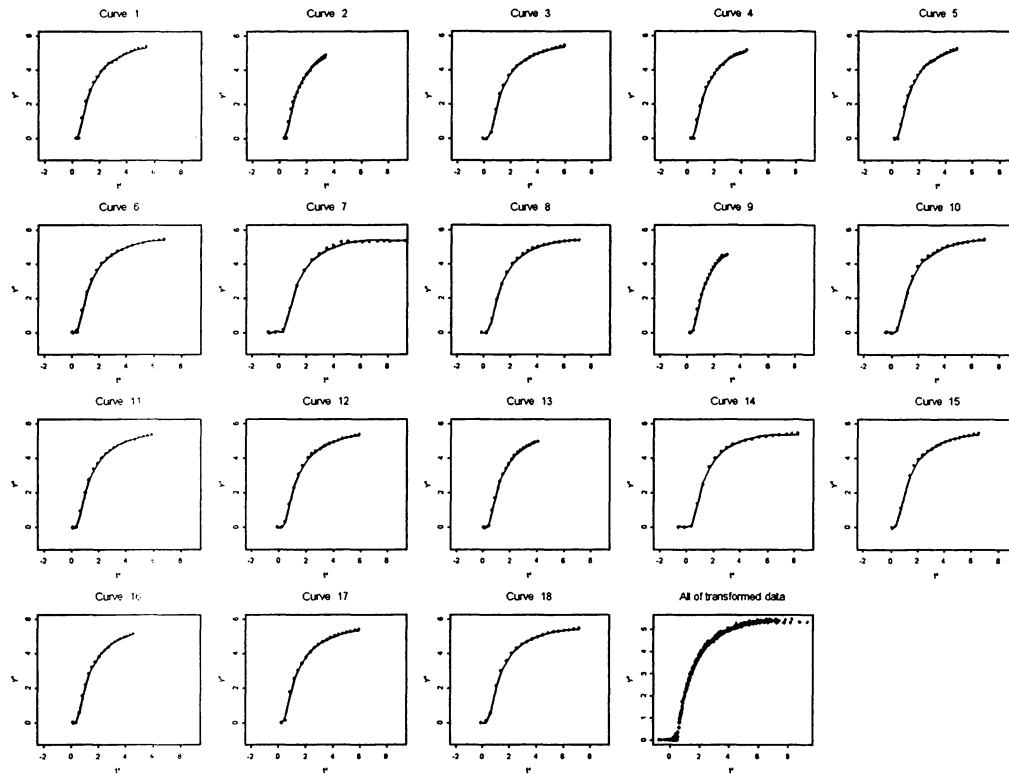***Figure 5:*** *Spirometer data (...) with individual estimated curves (-).*

Figure 6: *Scaled and shifted spirometer data (...) with individual estimated shifted curve (-)).*

6. Discussion

The paper demonstrates the efficacy of using p-splines and random effects for fitting the SIM model to a family of curves. The SEMOR model has a number of advantages over fully parametric or fully nonparametric models. It allows very flexible modeling of the common shape function, while allowing differences among curves to be expressed as parametrically. Commonly used summaries, such as the maximum, time to maximum and so on can be expressed as a function of the parameters and a functional of the shape function.

The use of p-splines to model the common shape offers further advantages. P-splines are sieve estimators - that is, for a fixed set of knots a p-spline is a parametric function, and convergence to a nonparametric class of functions is achieved by increasing the number of knots as the sample size increases. For any fixed set of knots, the p-spline can be expressed as a linear mixed model. Hence, generalization to data with correlated errors, generalized linear mixed models, Cox

regression and other generalizations of the linear mixed model is conceptually simple (although the computational cost may be high.) Varying degrees of curvature in μ_0 can be accommodated by use of a general diagonal variance matrix for σ_u in (8). The quadratic penalty on the jumps in derivatives of μ_0 at the knots can be replaced by any function corresponding to an exponential family log-likelihood.

Similarly, the use of the iterative algorithm allows us to generalize seamlessly to the general SEMOR model - the affine transformation of the response and time axes can be replaced by any monotone parametric function. In general, let both ϕ and κ be dependent on θ_i, the vector of q random effects where

$\theta_i = (\theta_{1i}, ..., \theta_{ni})^T$. $E[\theta_i] = \mu_\theta$ and $Var[\theta_i] = \Sigma_\theta$. The general set up is as follows:

$$Y_i = \phi\{X[\kappa(t,\theta_i)]\gamma + Z[\kappa(t,\theta_i)]u,\theta_i\} + \varepsilon_i$$

$$\varepsilon_i \sim i.i.d.N(0,I\sigma_\varepsilon^2) \qquad u_i \sim N(0,\sigma_u^2)$$

$$\theta_i \sim i.i.d.N(\mu_\theta,\Sigma_\theta)$$

Estimation for this general SEMOR form can be carried out similarly to SIM, with suitable normalization of the parameters to achieve identifiability. The random effects need not be mutually independent.

SEMOR Algorithm

Step 0 Choose initial estimates of $\theta_i^{(0)}$. Set k=0.

Step 1 Transform data and time with

$$Y*_i^{(k)} = \phi(Y_i; \theta_i^{(k)})^{-1} \text{ and } t_i*^{(k)} = \kappa(t_i; \theta_i^{(k)})$$

Step 2 Using LME, estimate $\gamma^{(k)}$ and $u^{(k)}$ by fitting:

$$Y^{*(k)} = X(\ t_i^{*(k)})\ \gamma^{(k)} + Z(t_i^{*(k)})\ u^{(k)}$$

Step 3 Using NLME, estimate $\theta^{(k+1)}$ by fitting the model

$$Y_{ij} = \phi\{X[\kappa(t_i;\ \theta_i^{(k)})]\ \gamma^{(k)} + Z[\kappa(t_i;\ \theta_i^{(k)})]\ u^{(k)}\} + \varepsilon_{ij}$$

Check for convergence. Else normalize $\theta^{(k+1)}$ and go to Step 1.

The iterative algorithm also suggests conditional tests when there are time-invariant covariates that might affect either the regression function or the axis transformations. The inverse axis transformations (Algorithm Step 1) aligns the data making it possible to determine if the shape of the regression on the aligned data varies as a function of the covariate. If the regression function does not vary with the covariate, then conditional on the fitted function, we have a nonlinear mixed effects model which can readily accommodate covariate effects.

Computation is still problematic for these models. Either a high dimensional integration must be performed, or computationally intensive methods are needed.

The flexibility of the SEMOR model, coupled with modern computational methods, make it a valuable addition to the statistical modeling tool kit.

7. References

Bates, D.M. and Lindstrom, M.J. (1986) 'Nonlinear least squares with conditional linear parameters,' *Proceedings of the American Statistical Association, Statistical Computing Section*, 152-15..

Davidian, M. and Giltinan, D.M.(1995) *Nonlinear Models for Repeated Measurement Data*, New York: Chapman & Hall.

Eilers, P.H.C. and Marx, B.D. (1999), "Generalized Linear Models on Sampled Signals and Curves: A P-Spline Approach", *Technometrics*, **41**, 1-13.

Kneip, A. and Gasser, T., (1988) "Convergence and consistency results for self-modeling non-linear regression", *The Annals of Statistics*, **16**, 82-112.

Kneip, A. and Engel, J., (1995), "Model estimation in non-linear regression under shape invariance", *The Annals of Statistics*, **25**, 551-532.

Lawton, W.H., Sylvestre, E.A. and Maggio, M.S., (1972), "Self modeling non-linear regression", *Technometrics*, **14**, 513-532.

Lindstrom, M.J. and Bates, D.M., (1990), "Non-linear mixed effects models for repeated measures data", *Biometrics*, **46**, 673-687.

Lindstrom, M.J., (1995), "Self-Modeling with random shift and scale parameters and a free-knot spline shape function", *Statistics in Medicine*, **14**, 2009-2021.

Pinheiro, J. C. and Bates, D. M., (1995) "Approximations to the Log-likelihood Function in the Nonlinear Mixed-effects Model", *Journal of Computational and Graphical Statistics*, **4**, 12--35.

Ruppert, D., Carroll, R.J., (1997) *Penalized Regression Splines*, tech. report, Operations Research, Cornell University.

Searle, S.R., Casella, G. and McCulloch, C.E., (1992) *Variance Components*, John Wiley & Sons, Inc.: New York.

Shen, X. and Wong, W. H., (1994), "Convergence Rate of Sieve Estimates", *The Annals of Statistics*, **22**, 580--615.

Wahba, G. (1985) "A Comparison of GCV and GML for Choosing the Smoothing Parameters in the Generalized Spline Smoothing Problem", *The Annals of Statistics*, **4**, 1378-1402.

Villarreal, J. (2001) *Self-modeling Nonlinear Regression with Random Shifts and a Penalized Regression Spline Shape Function* M.S. Thesis, Cornell University.