

# Method

February 22, 2019

# 1 The Shape Invariant Model

## 1.1 Model

In shape invariant model, individual curves can be obtained by shifting up-down, left -right, stretching and shirking a single mean curve. The mathematic form of the model can be expressed as:

$$f(\alpha_{0i}, \alpha_{1i}, \alpha_{2i}; t) = \alpha_{0i} + g(e^{\alpha_{1i}t} - \alpha_{2i}) \quad (1.1)$$

where  $g(\cdot)$  determines the shape of the response curve and  $\alpha_{0i}, \alpha_{1i}, \alpha_{2i}$  are the location and scale parameters for the curve.  $\alpha_{0i}, \alpha_{1i}, \alpha_{2i}$  are individual random effects.  $\alpha_{0i}$  is the up-down shift parameter for the  $y$ -axis.  $\alpha_{2i}$  is the left-right shift parameter for the  $x$ -axis.  $\alpha_{1i}$  is the shrink-stretch scale parameter which adjusts the speed of outcome progression. Exponentiation ensures scaled time to be greater than zero. Figure 1 illustrates the idea. The solid black line is the mean curve for population with  $\alpha_{0i}, \alpha_{1i}$  and  $\alpha_{2i}$  equal to 0. The red dashed line indicates a vertical shift in  $y$ -axis corresponding to the value of  $\alpha_{0i}$ . The blue dotted line represents a horizontal shift in  $x$ -axis corresponding to the value of  $\alpha_{2i}$ . The green dot-dashed line represents a shrinkage in time scale corresponding to the value of  $\alpha_{1i}$ .

The mean curve function  $g(\cdot)$  can be either parametric or non-parametric. We propose to use splines. Splines are piecewise polynomials joint together (Smith, 1979). The joint points are called knots. A spline of degree  $n$  is continuous up to its first  $n - 1$  derivatives. There are different representations of splines. B-splines are computationally efficient. However, they lack easy statistical interpretation (Smith, 1979). Another down side of B-spline is that a B-spline is only defined over a specific range (e.g. the range of the data) but the shifting and scaling may produce data beyond the original bounds (Beath, 2007). Another approach is to use truncated polynomials to construct splines. A spline of degree  $d$  and  $k$  knots  $\tau_1, \tau_2, \dots, \tau_k$  can be expressed as:

$$g(x) = \sum_{i=0}^d \gamma_i x^i + \sum_{j=1}^k \gamma_{j+d} (x - \tau_j)_+^d \quad (1.2)$$

where the "+" function is defined as:  $x_+ = x$  if  $x > 0$  and  $x_+ = 0$  otherwise.  $\gamma_0, \gamma_1, \dots, \gamma_{d+k}$  are considered to be fixed effects which are common for every individual since the spline function represents the common curve. Only internal knots are needed to define a spline using the truncated polynomial approach. One downside of this approach is that the polynomial terms may be correlated with each other and cause multicollinearity. The coefficient estimates in the spline function may not be valid but the predictive power of the model as a whole will not be reduced by multicollinearity. The interpretations of the coefficient estimates of the polynomial terms have little use in practice. We are more concerned about the outcome prediction. Thus, we proposed to use the representation of truncated

polynomials.

Overall, the model can be expressed as

$$y_{ij} = \alpha_{0i} + g(e^{\alpha_{1i}t_{ij}} - \alpha_{2i}) + \epsilon_{ij} \quad (1.3)$$

where  $g(\cdot)$  is the spline function as in equation 1.2,  $\epsilon_{ij} \sim N(0, \sigma^2)$  and

$$\begin{pmatrix} \alpha_{0i} \\ \alpha_{1i} \\ \alpha_{2i} \end{pmatrix} \sim N(0, \Psi) \quad (1.4)$$

The predicted curves for individuals and population can be obtained from the above non-linear mixed effects model. Individual curves can be transformed back using their three random effects. If the model performs well, all curves should overlay upon the mean curve.

## 1.2 Knots and degree selection

Initial knots are placed at quartiles or tertiles of visits. Number of knots can then be selected by comparing AIC or BIC.

## 1.3 Covariates

A covariate can be incorporated into the model either by modifying the common shape function or by modifying the SIM parameters. We proposed to modify the scale parameter to be the sum of random effect and a linear combination of the covariates. The covariates are considered to be fixed effects.

$$y_{ij} = \alpha_{0i} + g(e^{(\alpha_{1i} + \beta^T \mathbf{x}_i)t_{ij}} - \alpha_{2i}) + \epsilon_{ij} \quad (1.5)$$

(Why only add covariates into speed parameter?)

## 1.4 Algorithm to choose starting value for fixed effects

Step 1 Fit a linear model without covariate. Obtain spline coefficients and  $\hat{y}_0(t)$ .

Step 2  $y_\lambda \approx f((1 + \lambda)t)$ .

First order Taylor expansion:  $f((1 + \lambda)t) \approx \hat{y}_0(t) + \lambda t f'(t)$ .

Thus,  $\lambda_i = \frac{y_i - \hat{y}_0(t)}{t f'(t)}$

Step 3  $1 + \lambda_i = \exp(\beta x_i)$ . Thus,  $\log(1 + \lambda_i) = \beta x_i$ . Fit model to obtain  $\beta$ .

## References

- Beath, K. J. (2007). Infant growth modelling using a shape invariant model with random effects. *Statistics in medicine*, 26(12), 2547–2564.
- Smith, P. L. (1979). Splines as a useful and convenient statistical tool. *The American Statistician*, 33(2), 57–62.

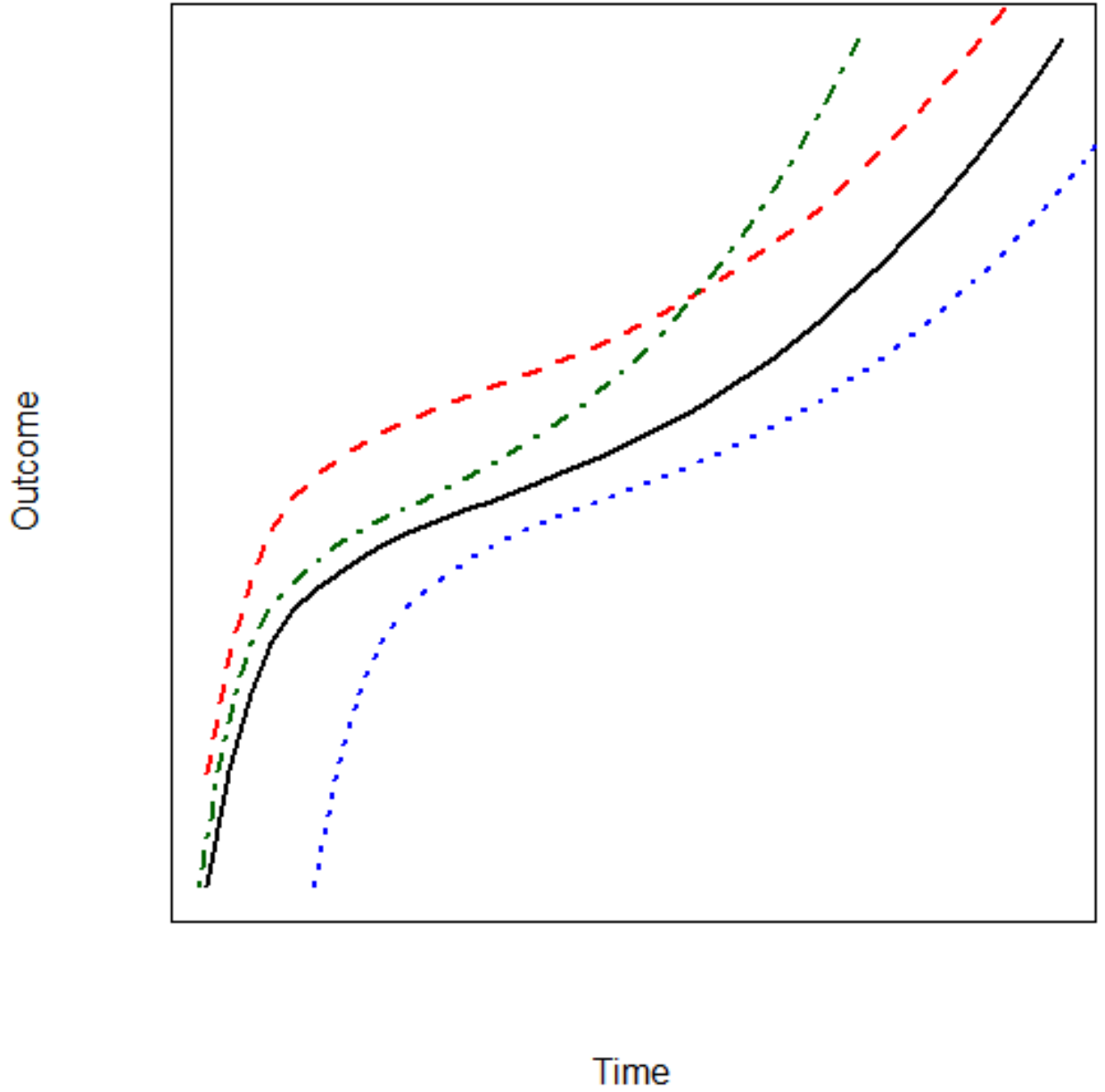


Figure 1: Illustration of the shape invariant model. The black solid line represents the mean curve for population. The red dashed line represents a shift in y-axis corresponding to  $\alpha_{0i}$ . The blue dotted line represents a shift in x-axis corresponding to  $\alpha_{2i}$ . The green dot-dashed line represents a shrinkage in time scale corresponding to  $\alpha_{1i}$ .