Statistical Modelling 2017; **17(6)**: 468–493

Dynamic child growth prediction: A comparative methods approach

Andrada E Ivanescu[1], Ciprian M Crainiceanu[2] and William Checkley[3]

[1]Department of Mathematical Sciences, Montclair State University, Montclair, NJ, USA.
[2]Department of Biostatistics, Johns Hopkins University, Baltimore, MD, USA.
[3]Department of Medicine, Johns Hopkins University, Baltimore, MD, USA.

Abstract: We introduce a class of dynamic regression models designed to predict the future of growth curves based on their historical dynamics. This class of models incorporates both baseline and time-dependent covariates, start with simple regression models and build up to dynamic function-on-function regressions. We compare the performance of the dynamic prediction models in a variety of signal-to-noise scenarios and provide practical solutions for model selection. We conclude that (a) prediction performance increases substantially when using the entire growth history relative to using only the last and first observation; (b) smoothing incorporated using functional regression approaches increases prediction performance; and (c) the interpretation of model parameters is substantially improved using functional regression approaches. Because many growth curve datasets exhibit missing and noisy data, we propose a bootstrap of subjects approach to account for the variability associated with the missing data imputation and smoothing. Methods are motivated by and applied to the CONTENT dataset, a study that collected monthly child growth data on 197 children from birth until month 15. R code describing the fitting approaches is provided in a supplementary file.

Key words: functional data, function-on-function regression, functional regression, longitudinal data, height, weight

Received April 2016; **revised** January 2017; **accepted** March 2017

1 Introduction

The prediction of child-specific growth has been previously studied in the literature. Tanner et al. (1975) and Tanner et al. (1983) provided approaches for predicting the height of individuals based on measurements made during childhood. The focus of these studies was to predict adult height based on a linear regression on the height of the child, and covariates recorded during childhood. For child height, Karlberg et al. (1994) proposed an exponential function of age for infancy years (from birth to 3 years of age). Nonlinear modelling of child growth curves was proposed using linear mixed models based on linear splines regression (Howe et al., 2012, 2013; Cheung,

Address for correspondence: Ciprian M Crainiceanu, Department of Biostatistics, Johns Hopkins University, 615 N. Wolfe Street, E3636, Baltimore, MD, 21205, USA.
E-mail: ccraini1@jhu.edu

 10.1177/1471082X17707619

2014), cubic splines regression (Grajeda et al., 2016) and polynomial regression (Tilling et al., 2014). Some disadvantages of these methods include: (a) emphasis on population-level estimation instead of subject-specific dynamic prediction, where the future of each growth trajectory is estimated at every time point; (b) lack of explicit modelling of the multivariate nature of the response and predictors; and (c) lack of modelling approaches for dynamic prediction.

Quantile regression models (Carey et al., 2004; Wei and He, 2006) have also been used to describe the marginal quantile structure of growth curves. Regression models using conditional scores (Cheung, 2014; Johnson, 2015) obtained from longitudinal models have been proposed to predict a scalar response. Conditional growth charts (Zhang and Wei, 2015) provide a modelling approach that uses principal components based on conditional information in order to estimate growth curves. Quantile regression (Carey et al., 2004; Wei and He, 2006; Cheung, 2014; Zhang and Wei, 2015) models are not designed for dynamic prediction, do not provide subject-specific predictions and cannot incorporate multiple functional predictors.

Dynamic prediction has been recently proposed in the literature (Chiou, 2012; Goldberg et al., 2014; Chiou et al., 2016; Hyndman and Shang, 2016), but methods were not designed to incorporate multiple functional predictors and multiple scalar covariates. Our proposal is to consider explicit regression models that can be used in complex dynamic prediction scenarios. In particular, methods are designed to work with sparsely sampled data, multiple functional predictors and covariates, while providing inference on intuitively defined model parameters. The most important novelty of our approach is that we generalize function-on-function regression to the case of dynamic prediction. We also evaluate and compare different generalization approaches. Our approach was motivated by and applied to a child growth study conducted by one of the co-authors of this article (Dr Checkley).

1.1 Study setting and design

The methods proposed are motivated by and applied to the CONTENT child growth study, which was conducted in Pampas de San Juan Miraflores and Nuevo Paraiso, two peri-urban shanty towns with high population density, 25 km south of central Lima. These peri-urban communities (Checkley et al., 1998, 2002) are comprised of 50 000 residents, the majority of whom are immigrants from rural areas of the Peruvian Andes who settled nearly 35 years ago and later claimed unused land on the outskirts of Lima. In the last two decades, Pampas has undergone many economic and social developments. In 1989, most homes were temporary structures, constructed of wooden poles and woven thatch, without water or sewage lines. Currently, more than 85% of homes are constructed from brick or cement with in-home water and sewage lines. The study was approved by the European Union Ethics Committee, A. B. PRISMA and Universidad Peruana Cayetano Heredia, in Lima, Peru, and the Bloomberg School of Public Health, Johns Hopkins University, in Baltimore, USA. Mothers or caregivers provided written informed consent.

The study team is led by Dr Checkley, co-author of this article, and keeps an ongoing vital status and birth census database in the study areas. Based on a sampling frame of all pregnant women from May 2007 to February 2011, pregnant women in Peru were selected at random to participate in the study. Exclusion criteria included: children with severe disease that required hospitalization, severe chronic illness, child of a multiple pregnancy, birth weight less than 1 500 grams and intentions to move during the period of the study. Of 304 eligible children, 11 did not have available anthropometric data, 95 did not complete 500 days of surveillance and 11 did not have a baseline socioeconomic status (SES) data or carbon 13-labelled (C-13) urea breath test results (UBT). Our analysis included a total of 197 children who passed the inclusion criteria. At enrollment, field workers asked caretakers to complete a questionnaire on socio-economic and living conditions, including floor material, number of inhabitants, number of rooms, water source and storage, and durable assets. Field workers visited the household daily and obtained the history on breastfeeding and diarrheal symptoms, including consistency of feces, loss of appetite and intake of antibiotics. CONTENT collected growth outcomes more frequently in the first 3 months (weekly) and $3-11$ months (fortnightly) and less frequently between 12 and 24 months (monthly). Height was measured with a locally made wooden platform with sliding footboard.

1.2 Growth curve data structure

The study of growth curves of children is of fundamental importance for: (a) quantifying the development of the population; (b) comparing parameters of growth across different geographic locations and times; and (c) understanding the population and subject-specific dynamics of patterns of growth. An important component of modelling subject-specific growth dynamics is to predict the future trajectory of growth of a subject based on their known growth history. Indeed, predictions can be used for: (a) identifying children who are at increased risk of stunting; (b) quantifying and classifying early patterns of growth associated with healthy growth; and (c) quantifying the relative importance of growth predictors, which can be used for targeting and designing growth intervention studies or clinical trials.

To better understand the problem, Figure 1 displays data from the CONTENT study, which measured growth on 197 children from birth until month 15. Here we are using a slightly modified version of the CONTENT data to ensure that lagged regression models can be implemented. More precisely, if more than one observation was sampled per month, then we used the within-month average for that child. As we are not interested in small within-month fluctuations of growth curves, averaging within-month growth observations will be associated with only minimal loss of information. At this aggregated level, individual subjects exhibit different patterns of missing data at the month level. The left-top panel in Figure 1 displays the height-for-age z-score (HAZ) separated by boys (right) and girls (left). The x-axis is time in months, where zero is birth, and the y-axis is expressed in z-score units. To better understand the measurement, it is worth mentioning that HAZ data are

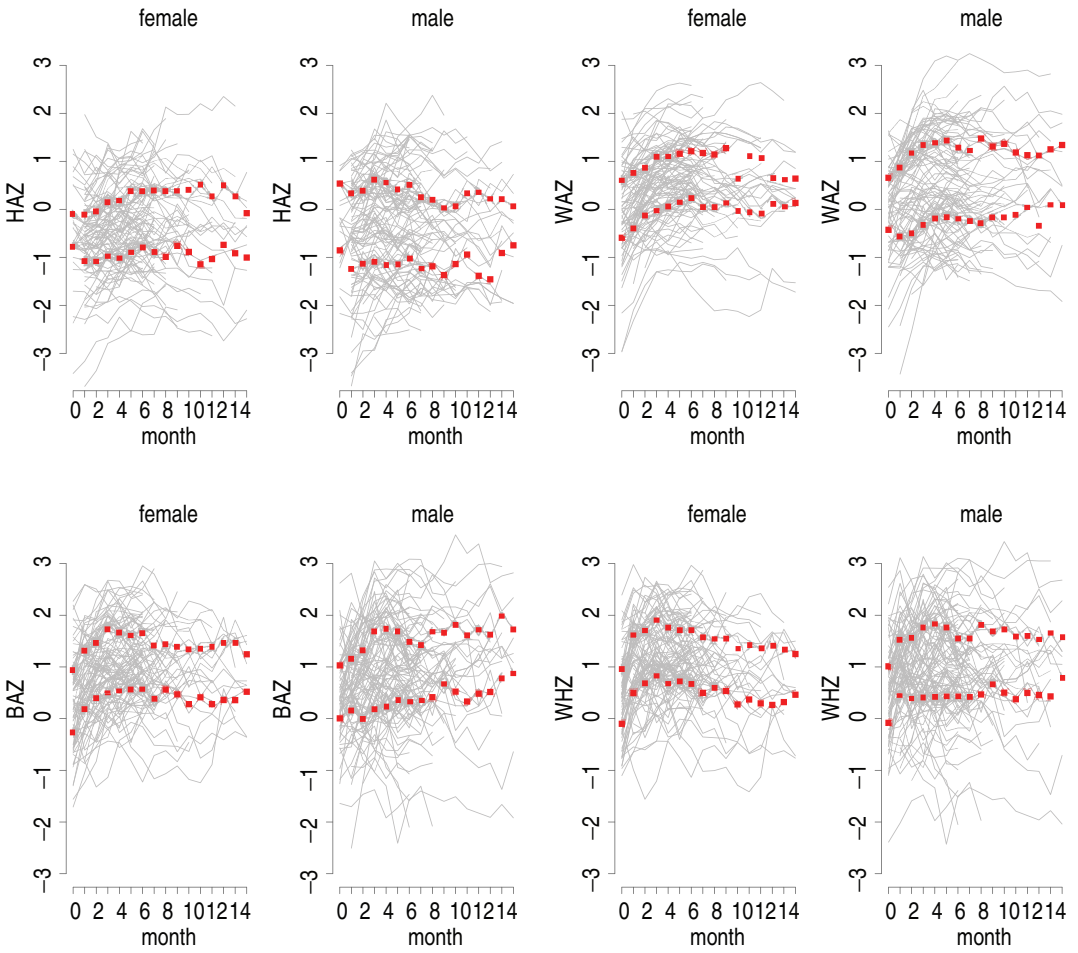

Figure 1 Displayed are HAZ, WAZ, BAZ and WHZ measurements by sex from birth to month 15. The 0.25 and 0.75 quantiles are depicted as red squares at each month.

obtained relative to the World Health Organization guidelines (WHO, 2016). More precisely, height is normalized using the age and sex-specific reference mean and standard deviation from the WHO standard. Thus, a value of −1 at time 6 on the HAZ scale corresponds to a child who is one standard deviation below the mean of the WHO standard for age 6 months. A closer inspection of the HAZ data indicates that the HAZ for girls seems to decrease slightly or stay about the same immediately after birth with values increasing slightly starting with month 2. The HAZ for boys seems to stay about the same or decrease immediately after birth with values staying about the same or decreasing after month 4. This observation may indicate a slight sex effect that we aim to account for in analyses. A plot of the 0.25 and 0.75 marginal

quantiles for both males and females (shown in red) seems to suggest the existence of a potential gender effect. Figure 1 also displays weight-for-age (WAZ) in the top-right panel, body-mass-index for age (BAZ) in the bottom-left panel, and weight-for-height z-score (WHZ) in the bottom-right panel. These measures are calculated for the same children using a similar normalization approach to the one for HAZ. For these data, boys tend to be heavier than girls, though the body-mass-index and weight-for-height differences are smaller. This may indicate that for the same height, girls' weight is not increasing as much as boys' weight.

Visual inspection of the panels in Figure 1 indicates that some visits are missing for some children. Figure 2 displays the histogram of percent missingness by month. Thus, data at birth are available for only 29% of children, with availability increasing to about 88% at month 1 and close to 100% at months 2 through 6, and a steady decline from month 7 (82%) through month 15 (10%). This is a problem in many growth studies and requires careful handling. Indeed, consider the case when a prediction model uses the birth size of the child to predict the future trajectory. Such a model would only use 29% of the sample to conduct the analysis. However, a lot of information about birth size is contained in the size of the child at month 1 and

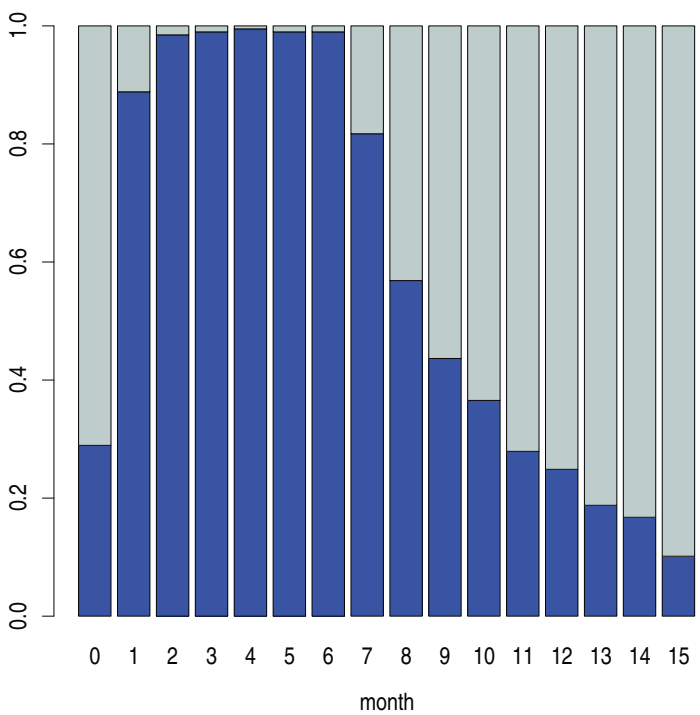

Figure 2 Histogram of percent missingness (gray bars) and percent available (blue bars) HAZ data at each month for the study time frame starting with birth until month 15.

2, which could be used to predict birth size, which, in turn, could be used to predict future growth. The idea is to use the data from the known history (e.g., HAZ at month 1 and 2) to fill in the missing history data (e.g., HAZ at birth) and predict the future growth (e.g., HAZ at months 3 through 15). Note that we do not propose to use the future to predict itself. Of course, there are many ways of filling in the missing observations and the choice of imputation mechanism may affect inferential results. Here we will focus on filling in data based on functional principal component analysis (fPCA), see Yao et al. (2005) and Goldsmith et al. (2013) for more details. For now, we focus on complete data modelling and we will revisit the missing data problem.

One of the fundamental problems in child growth curve research is to predict growth-related observations from historical growth data and other covariates. For example, one may be interested in predicting the future height trajectory of a child based on their recorded growth data up to 6 months and other covariates, such as the SES of the mother and sex of the child. Here, we are interested in prediction based on all lengths of growth history. We refer to this context as dynamic prediction, to indicate that prediction depends on and changes with the assumed length of known growth history. We approach this problem using an explicit dynamic regression approach, where the dependence between growth outcomes and growth predictors is described *explicitly* using regression equations. This type of approach is complementary to the mixed effects modelling approaches, where dependence is induced via *implicit* modelling using subject-specific random effects. Our goal is not to promote one approach versus the other, but to provide an explicit approach to modelling that may provide additional insight and intuition for data analysts.

1.3 Description of scientific and methodological questions

We turn our attention to the problem of predicting subject-specific growth curves at all times $t > t^*$ for every known growth history measured at months $0, \ldots, t^*$. We assume that there is no missing data and we will revisit the missing data approaches later. From a data-structure perspective, there are two different types of variables that could be incorporated in the growth curve prediction. Some variables are measured at baseline (i.e., before or at birth) and some are time-dependent, such as HAZ and WAZ. These variables will need to be carefully incorporated in and interpreted for each modelling approach. For example, one may be interested in quantifying how large is the effect of birth HAZ relative to mother SES on future growth. Or, one may be interested in quantifying how much additional information is provided by HAZ at month 6 in addition to HAZ at birth for predicting HAZ data at month 15.

The basic scientific questions can be described as follows:

1. Given the history of growth of a child up to month t^* (for example $t^* = 5$) and their baseline covariates, what is the predicted HAZ for the child for each one of the months $t^* + 1, \ldots, 15$?
2. Quantify the relative importance of predictors of growth as a function of available historical information (t^*) and when in the future the prediction is being made (t).

3. What are the modelling options and how can they be compared in terms of prediction and inference?
4. How can one handle the different patterns of missing data within children?
5. How can modelling complexity be reduced without affecting prediction performance?

A first step towards addressing these questions is to investigate the shape and type of possible associations between the variables of interest. Figure 3 displays

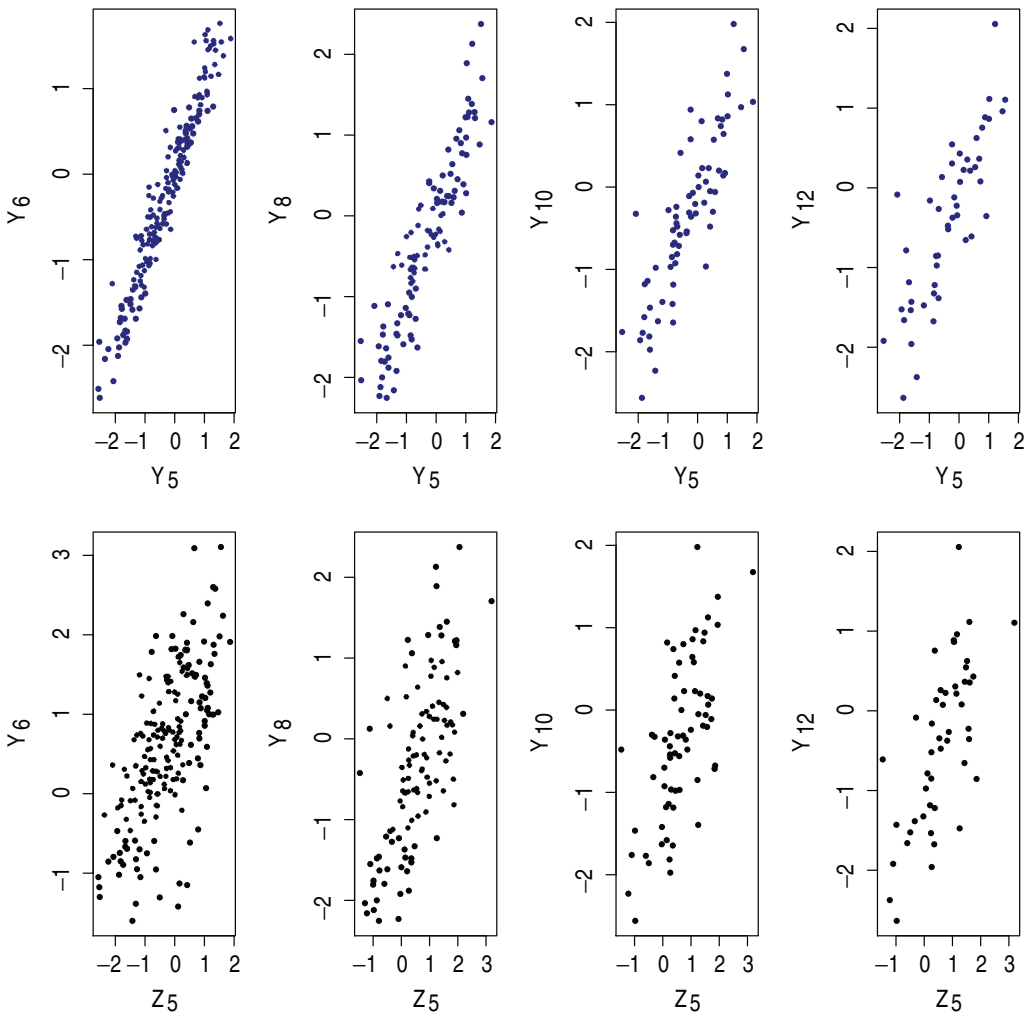

Figure 3 Association for Y_{t^*} (HAZ at $t^* = 5$) and upcoming HAZ observations at Y_{t^*+1}, Y_{t^*+3}, Y_{t^*+5} and Y_{t^*+7} are displayed on top row. The associations for the same HAZ observations beyond t^* (Y_{t^*+1}, Y_{t^*+3}, Y_{t^*+5} and Y_{t^*+7}) versus Z_{t^*} (WAZ at $t^* = 5$) are displayed in the bottom row.

some initial exploratory data analysis based on the observed data. The top panels display the scatterplot for the observed HAZ data at month 5 (x-axis) versus the observed HAZ data at months 6, 8, 10 and 12 (four top panels from left to right). These plots suggest a strong linear association between the variables, indicating that linear regression with specific predictors may be an excellent starting point for model building. As the prediction time into the future ($t - t^*$) increases, data become more spread-out around the linear trend. This is expected and reassuring, because it represents a natural decrease in correlation with increasing temporal distance between measurements. To quantify the observed degree of association between HAZ measurements, the top panel in Figure 4 displays the R^2 for the linear regression of HAZ measurements at months 6 through 12 on HAZ measurements at month 5. The R^2 tracks the same type of observed associations, indicating a decrease from 0.93 at month 6 to 0.69 at month 12. The decline in R^2 is roughy linear with the increased lag. The blue lines corresponds to the unadjusted linear model $Y_t|Y_{t*}$, while the red line corresponds to the same model adjusted for the sex of the child.

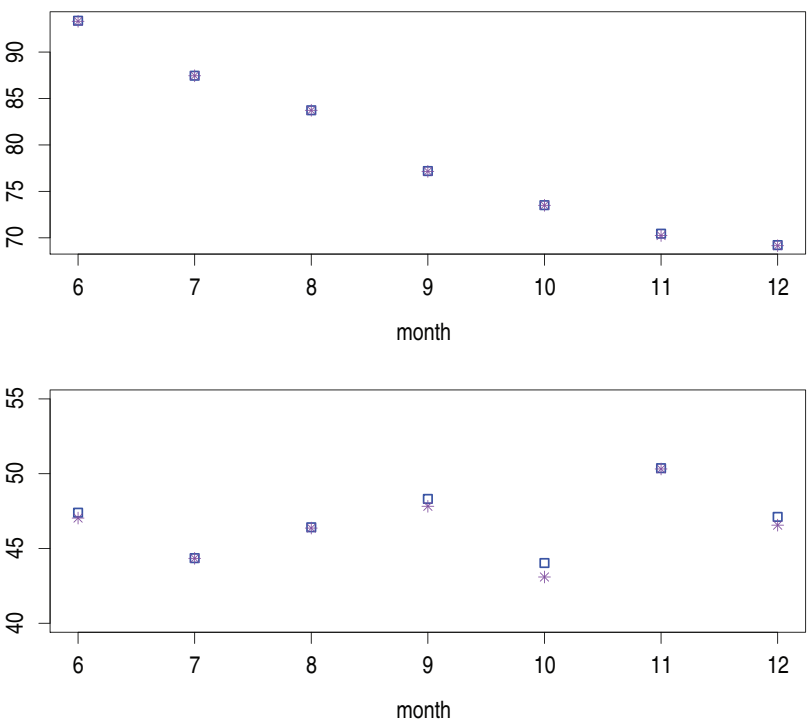

Figure 4 Change in R-squared to predict HAZ at months $\{t^* + 1, .., t^* + 7\}$ for model with predictors observed at $t^* = 5$: HAZ (blue squares in top panel), HAZ and sex (purple stars in top panel), WAZ (blue squares in bottom panel) and WAZ and sex (purple stars in bottom panel).

The combination of the top panels in Figures 3 and 4 indicates that (a) linear models using past HAZ observations are a good starting point; (b) additional covariates may be especially useful for predicting in the more distant future (there is not much to improve on a 0.93 R^2, but there is room for improvement on a 0.7 R^2); and (c) once data are corrected for past HAZ observations, the impact of sex may be relatively small for predicting future HAZ (note the small R^2 differences in the top panel of Figure 4). The bottom panels in Figures 3 and 4 are similar to the corresponding top panels, though the focus is on scatter plots of WAZ at month 5 (x-axis) versus HAZ at months 6, 8, 10 and 12. The story is similar, though with some important differences. As before, associations are strongly linear, though the strength of the association does not seem to change as dramatically with increased time differences. An inspection of the corresponding R^2 in Figure 4 reveals that R^2 varies around 48%, with some spikes that are hard to explain and may just represent random variation. These results also make sense and are quite encouraging. Indeed, it is natural to expect smaller correlation between HAZ in the future and WAZ in the past than between HAZ in the future and HAZ in the past. These exploratory data analyses suggest a framework for modelling of the data that is described in the next section.

2 Conceptualizing the problem

We now describe the general conceptual framework. We are interested in modelling at every time point, t^*, the future of the growth process (e.g., HAZ) denoted by Y_l as a function of the known history of the HAZ process $H\{Y_{t^*}\} = \{Y_l : l \in 0, \ldots, t^*\}$, time-independent covariates, W, and the history of time-dependent covariates, Z_l. For example, we denote by $H\{Z_{t^*}\} = \{Z_l : l \in 0, \ldots, t^*\}$, the known history of the WAZ process, though we could incorporate other time-dependent processes or a combination of such processes. To keep presentation simple, we focus on only one time-dependent predictor and one time-independent covariate.

The regression models we focus on are of the type

$$Y_t | W, H\{Z_{t^*}\}, H\{Y_{t^*}\}. \tag{2.1}$$

This is the most general modelling structure we consider and it contains a large number of models. At the most basic level, we start with a model that depends only on the first and last observations of the known history of both Y and Z processes, Y_0, Y_{t^*}, Z_0, Z_{t^*}, and the covariate, W. More precisely, the simplest model is

$$Y_t = W\gamma_{t^*,t} + Y_0\beta_{0,t,t^*} + Y_{t^*}\beta_{t^*,t,t^*} + Z_0\delta_{0,t,t^*} + Z_{t^*}\delta_{t^*,t,t^*} + \epsilon_{t^*,t}. \tag{2.2}$$

We refer to this as the BENchmark DYnamic (BENDY) prediction model to underscore its highly adaptive structure. Note that the model parameters depend both on t and t^*, which reflects the dynamic nature of the prediction models. Thus,

the prediction equation depends both on when in the growth process the prediction is performed (t^*) and at what time into the future ($t > t^*$). This explicit dependence ensures that predictions are tailored to the time point where they are produced.

2.1 Statistical issues and methodological challenges

The BENDY model leaves several problems unadressed. First, we are interested whether additional historical information in the Y and Z processes (Y_s, Z_s, for $s = 1, \ldots, t^* - 1$) can be used to improve fit and prediction. Second, we would like to borrow strength between prediction equations at various pairs (t^*, t). Indeed, one could expect that model coefficients do not vary dramatically with small changes in t^* or t, which could be used to smooth predictions and reduce variability. Addressing these two issues will naturally lead to penalized function-on-function regression, which we describe in Section 3. The models we consider are all particular cases of the general framework in equation (2.2), and build up in complexity naturally. The first model is BENDY, which is then extended to the dynamic linear model (DLM) that contains all available historical observations. Two functional models dynamic penalized functional regression (DPFR) and dynamic penalized function-on-function (DPFFR) are obtained from the DLM by imposing increasing amounts of smoothing on the model coefficients. Models are fit using existent software, making them more useful. An R markdown document is provided to ensure reproducibility of methods and illustrate their implementation.

The general framework can be described as follow:

1. Consider the original data, smooth it and impute missing observations.
2. Simulate datasets that mimic the structure of the original dataset with various levels of noise: low, moderate, realistic and large.
3. Compare and choose the dynamic prediction model with the smallest prediction mean squared error (MSE).

This general framework contains several innovations: (a) consider the entire prediction problem and account for its natural dynamic nature; (b) start with simple, benchmark models and assess the variability explained by the models at various levels of available historical information; (c) expand models to incorporate all available temporal information; (d) induce smoothing of the dynamic model parameters to account for correlation of predictors and reduce variability of predictions; and (e) implement methods in reproducible and stable software.

3 Methods

The main goal of our article is to introduce models that incorporate more information than just the first and last available historical observation of the time-varying

processes. This can be done by expanding model given in equation (2.2) to the following DLM

$$Y_t = W\gamma_{t^*,t} + \sum_{l=0}^{t^*} Y_l\beta_{l,t,t^*} + \sum_{l=0}^{t^*} Z_l\delta_{l,t,t^*} + \epsilon_{t^*,t}. \tag{3.1}$$

The main difference between this model and the BENDY model is that it incorporates linear dependence on Y_l and Z_l for $l \in \{1, \ldots, t^* - 1\}$. This is a useful extension of the BENDY model when additional information is contained in these data. The DLM has a functional data interpretation and can be written as

$$Y_t = W\gamma_{t^*,t} + \int_{l=0}^{t^*} Y_l\beta_{l,t,t^*}dl + \int_{l=0}^{t^*} Z_l\delta_{l,t,t^*}dl + \epsilon_{t^*,t}. \tag{3.2}$$

Indeed, data are equally spaced and, up to a constant, the Riemann sum numerical approximations of the integrals given in equation (3.2) are the corresponding sums in equation (3.1). From a practical perspective, this formula is not relevant, though it is useful to understand the next, more sophisticated, modelling approaches. Unfortunately, incorporating the entire historical information creates a series of unintended consequences. Indeed, the observations Y_l and Z_l, $l = 0, \ldots, t^*$ are highly correlated, which leads to stability problems when running DLM. Moreover, while associations tend to be well represented by linear regressions, these associations change subtly with t and t^*. This can lead to problems of interpretability, especially given the large number of parameters for each (t^*, t) pair.

To alleviate these issues, we impose smoothing priors on the model coefficients. The first approach is to treat every DLM as a scalar-on-function regression, where the scalar is Y_t and the functions are Y_l, $l = 0, \ldots, t^*$ and Z_l, $l = 0, \ldots, t^*$. There is a large literature on fitting this type of models, though here we focus on penalized functional regression (PFR) (Goldsmith et al., 2011) and use the associated penalization described therein. The method is implemented in the pfr function of the `refund` (Goldsmith et al., 2016) `R` package. Without going too deeply into technical details, pfr imposes a penalized spline smoothing assumption on the β_{l,t,t^*} and δ_{l,t,t^*} coefficients for every t and t^*. Thus, the smoothing coefficients depend both on t and t^*. We call this class of models dynamic penalized functional regression (DPFR) models to indicate that we are interested in using pfr in a dynamic prediction context.

Another way of thinking about the problem is to consider it from the function-on-function regression perspective. The functional outcome in this context is Y_t, $t = t^* + 1, \ldots, 15$ and the functional predictors are Y_l, $l = 0, \ldots, t^*$ and Z_l, $l = 0, \ldots, t^*$. Just as in the case of DPFR, we impose smoothing using penalized spline approaches (Ivanescu et al., 2015; Scheipl et al., 2015). The main difference from the DPFR approach is that we impose smoothing simultaneously for all values of $t = t^* + 1, \ldots, 15$. Because we apply regression dynamically and use penalized function-on-function regression at every time point, we will refer to

this as dynamic penalized function-on-function (DPFFR) regression. Note that the smoothing parameters depend only on t^* and not on t, which reduces the number of smoothing parameters, for example, from $15 - t^*$ to 1 for a given model parameter and t^*. Thus, one would expect the DLM fits to be rougher than DPFR fits, which, in turn, are expected to be rougher than the DPFFR fits.

Both DPFR and DPFFR are particular cases of DLMs with penalties on the model parameters. DPFR uses a penalty for parameters on the history ($t \leq t^*$), but does not borrow strength across times where prediction is made. Thus, DPFR uses many linear penalized functional regressions and a univariate parameter space. In contrast, DPFFR uses a more aggressive penalty and borrows strength both across history and times where prediction is made. The smoothing is done on the associated bivariate parameter surface.

From a technical perspective, the DPFFR model parameter β_{l,t,t^*} is expressed as a smooth function of the lag l and time t for each t^* and has the following representation $\beta_{l,t,t^*} \approx \sum_{b=1}^{\kappa} a_b(l,t)\beta_{b,t^*}$, where $a_b(\cdot,\cdot)$ is a bivariate basis constructed from spline basis functions, and β_{b,t^*} are the associated basis coefficients. Thus, for a fixed t^*, the β_{l,t,t^*} model parameters are represented as a smooth bivariate surface. Similar modelling and interpretation is used for δ_{l,t,t^*}. The coefficients β_{b,t^*} are estimated by using quadratic penalties (Wood, 2006; Ivanescu et al., 2015) of the form $P(\beta) = \beta^t D_\beta \beta$, where D_β is a known penalty matrix corresponding to the chosen spline basis. Moreover, we obtain the best linear unbiased predictor (Ruppert et al., 2003) in a mixed model that uses variance components and smoothing parameters within a restricted maximum likelihood (REML) estimation framework. For DPFR and DPFFR implementations, we use of the R packages mgcv (Wood, 2006) and refund (Goldsmith et al., 2016); for more details on R code see the technical appendix.

4 Data analysis

4.1 Comparative prediction performance

In this section, we design realistic simulation scenarios based on the observed data and compare prediction MSE across methods. We start by smoothing the observed HAZ and WAZ data using an fPCA approach (Yao et al., 2005; Goldsmith et al., 2013). More precisely, we obtain the empirical covariance estimator from the data and smooth it using the `fpca.sc` function in the `refund` (Goldsmith et al., 2016) R package. This function uses a fast penalized spline bivariate smoother designed specifically for covariance estimation (Goldsmith et al., 2013; Xiao et al., 2016). Missing data are then imputed using BLUPs in the associated mixed effects model (Yao et al., 2005). Specifically, the method starts with the assumption that the observations are noisy realizations of a smooth underlying trajectory $Y_i(t) = X_i(t) + e_{it}$, where e_{it} are i.i.d. mean zero variance σ_e^2 random variables. If $K_Y(s,t)$ and $K_X(s,t)$ denote the (s,t) entry of the covariance operators of the $Y()$ and $X()$ processes, respectively, then it can be shown that $K_Y(s,t) = K_X(s,t) + \sigma_e^2 \delta_{s=t}$. The idea, suggested by Staniswalis

and Lee (1998), is to smooth the method of moments estimators $\widehat{K}_Y(s,t)$ using only the off-diagonal elements of $\widehat{K}_Y(s,t)$, estimate $K_X(t,s)$ for all (t,s) using smoothing, and then estimate $\widehat{\sigma}_e^2 = \int_t \widehat{K}_Y(t,t) - \widehat{K}_X(t,t)dt$. For smoothing the covariance matrix $\widehat{K}_Y(s,t)$ for $s \neq t$, we use a thin-plate spline bivariate smoother, where $\widehat{K}_Y(s,t) = \sum_{k=1}^{K} b_k \psi_k(t,s)$, where $\psi_k(,)$ is the thin-plate spline in 2-dimensions and b_k are shrunk using the low-rank radial smoothers penalty (Ruppert et al., 2003, pp. 252–253). The smoothing parameter can be estimated using REML or cross-validation (CV) and both approaches are implemented in the `mgcv` package in R (Wood, 2016). Once the covariance operator $K_X(s,t)$ is estimated, its eigenfunctions $\phi_{X,k}(t)$ can be used to model

$$Y_i(t) = \sum_{k=1}^{c} \xi_{ik}\phi_{X,k}(t) + e_{it},$$

where ξ_{ik} are mutually independent distributed $N(0,\lambda_k)$ and independent of $e_{it} \sim N(0,\sigma_e^2)$. This is a mixed effects model and the scores can be estimated using BLUPs. All these are implemented in the `refund` (Goldsmith et al., 2016) R package, which contains the combination of robust covariance estimation and BLUP calculation.

An example of HAZ and WAZ smoothed and imputed growth data are shown in Figure 5 (the two left panels correspond to one simulated HAZ and one simulated WAZ dataset, respectively). These curves are smoother than the observed data (shown in Figure 1), while capturing the overall dynamics and correlation characteristics. HAZ curves are predicted for the $n = 197$ children in the dataset. The right panels in Figure 5 display the first three estimated principal components for HAZ and WAZ, respectively. In both datasets, the first three principal components explain more than 99% of the observed variability. The first HAZ eigenfunction (black curve) indicates that a large part of the between-curve variability (91.61% of total variability) is accounted for by a vertical shift, which corresponds to sustained and constant differences in HAZ across the observational age range. In standard mixed effects analysis, this component could be captured by a random intercept. The second principal component (red curve which explains 5.32% of the variability) corresponds to the contrast between the first and the second half of the age domain. A subject with high absolute scores on this component will tend to be low/high or high/low on the HAZ scale for first/second part of the growth curve. The third principal component (which captures 3.06% of the observed variability) captures patterns of growth that have a contrasting behaviour at the beginning and end of the curve relative to the behaviour in the middle of the period. Subjects with high absolute loading on this component will tend to fluctuate either as high/low/high or low/high/low at the beginning/middle/end of the period.

Once HAZ and WAZ data are reconstructed, we compare the performance of prediction algorithms by simulating data from the following models

$$Y_i(t) = f_{Y,i}(t) + \epsilon_i(t);$$

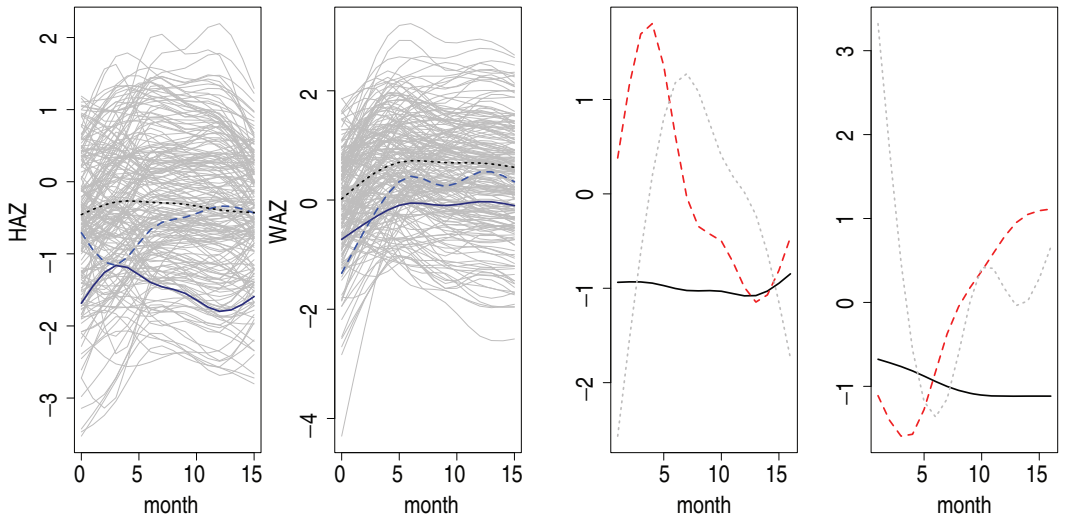

Figure 5 *Left two panels*: The simulated HAZ curves, displayed in the left panel, are generated using the fPCA method with eigenfunctions derived from the data application. The simulated WAZ data, displayed in the right panel, was obtained using the estimated eigenfunctions from the data application. In each set of simulated data three subjects are highlighted with different colours and line types. Scenario: low noise. *Right two panels*: Results from the method of functional principal components analysis for reconstructing the HAZ and WAZ trajectories are displayed in the left and right panel, respectively. The first three estimated eigenfunctions that explain 99% of the variability in the data application are shown. The estimated eigenfunctions are displayed as three distinct curves with a different colour corresponding to each eigenfunction: black solid line (first), red dashed line (second) and gray dotted line (third).

$$Z_i(t) = f_{Z,i}(t) + \eta_i(t),$$

where $f_{Y,i}(t)$ and $f_{Z,i}(t)$, $i = 1, \ldots, n$, are the smooth estimated HAZ and WAZ functions, respectively, shown in the left panels of Figure 5, and $\epsilon_i(t)$ and $\eta_i(t)$ are mutually independent normal random errors. In simulations, we consider cases when both $\epsilon_i(t)$ and $\eta_i(t)$ are homoscedastic with standard deviations σ_ϵ and σ_η, respectively. More precisely, we first estimated smooth underlying trajectories for every subject using the smoothing approach described. Figure 6 displays data for three subjects together with their estimated smooth trajectories. At every time point and for every subject, we obtain a difference between the observed and estimated curve. The standard deviation of these deviations at that time point across subjects is the month-specific standard deviation. For example, for month 6 in the left panel of Figure 6, we display the data as a square. In the right panel at month 6, we indicated by a square the standard deviation of the data deviations from the smooth curve. We repeat the process for every month and the standard deviation is estimated only over those subjects who have data at that particular point. A plot of these month-specific standard deviations are shown as red dots in the right panel of Figure 6. Note that the month-specific standard deviations vary between the minimum of 0.15 (shown

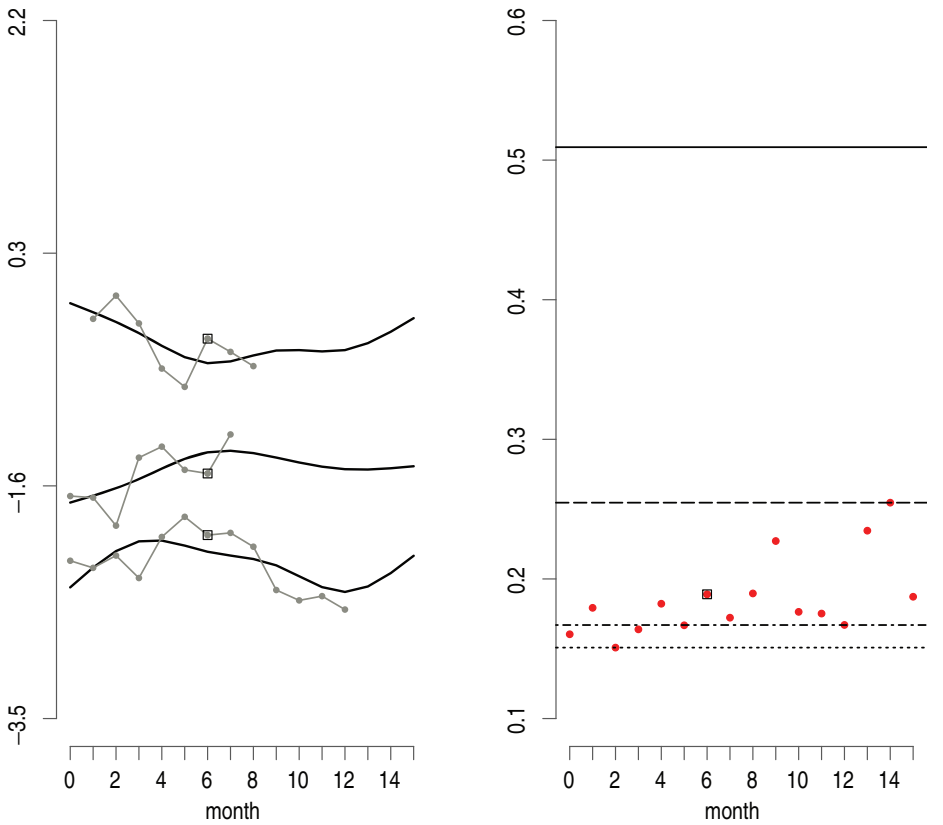

Figure 6 *Left panel*: Estimated trajectories (solid black lines) and HAZ data (gray points with gray lines) for three subjects. *Right panel*: Estimated month-specific standard deviation of the differences between the observed HAZ values and fPCA-reconstructed curves is displayed as a red solid point at each month. Across the 16 months, the minimum standard deviation was 0.15 (dotted line), the first quartile was 0.167 (dotted dashed line), the maximum standard deviation was 0.255 (dashed line). Twice the maximum standard deviation ($0.509 = 0.255 \times 2$) is shown as a solid line.

as a dotted line) and the maximum of 0.255 (shown as a dashed line) with the first quartile equal to 0.167 (shown as a dashed-dotted line).

To build realistic simulation scenarios, we have the following choices for $(\sigma_\epsilon, \sigma_\nu)$ of (0.15, 0.15), (0.167, 0.167), (0.255, 0.255) and (0.509, 0.509). We refer to these choices as low noise, mild noise, maximum noise and very high noise. Note that the very high noise corresponded to a choice of standard deviation is twice as large as the maximum month-specific standard deviation. To better visualize the very high noise scenario, a solid black line is shown at 0.509.

For quantifying the predictive performance of the proposed dynamic prediction methods, we used leave one-curve out CV. More precisely, we used $n - 1$ curves to fit

Table 1 Results for prediction MSE for dynamic regression methods to predict the future trajectories of HAZ at months $\{t^* + 1, ..., 15\}$. Results are displayed as $MSE \times 100$.

	Method	Low noise	Mild noise	Maximum noise	Very high noise
$t^* = 3$					
	BENDY	17.80	18.81	25.57	54.97
	DLM	16.85	17.96	24.66	52.99
	DPFR	16.78	17.89	24.57	52.59
	DPFFR	16.79	17.88	24.42	52.11
	dynamic_FLR	21.82	22.96	29.20	54.69
$t^* = 5$					
	BENDY	8.70	9.59	15.78	45.23
	DLM	6.56	7.75	15.05	44.12
	DPFR	6.36	7.51	14.61	42.66
	DPFFR	6.40	7.55	14.61	42.07
	dynamic_FLR	6.96	8.27	16.30	45.33
$t^* = 7$					
	BENDY	5.72	6.61	12.80	42.05
	DLM	4.42	5.35	11.54	39.07
	DPFR	4.23	5.11	10.99	37.10
	DPFFR	4.18	5.05	10.83	36.35
	dynamic_FLR	4.24	5.18	11.28	37.78
$t^* = 11$					
	BENDY	4.23	5.07	10.95	39.42
	DLM	3.63	4.45	10.17	37.88
	DPFR	3.24	3.97	9.08	33.69
	DPFFR	3.19	3.90	8.91	33.15
	dynamic_FLR	3.32	4.05	9.38	35.26

the model and the fitted model was then applied to obtain predictions corresponding to the left-out curve. This process is repeated n times for each curve in the sample. As a measure of predictive accuracy, we calculated the average of the squared deviations between the predicted and simulated values.

MSE results for the predictions are displayed in Table 1, where predictions were obtained for all months $\{t^* + 1, ..., 15\}$. Our methods have been compared with the method of Hyndman and Shang (2016) which is implemented in the ftsa R package. We label this method as dynamic_FLR and display the results in Table 1. Results indicate that BENDY performs well showing that inclusion of the baseline and last known values is important. However, all models using all historical information (DLM, DPFR and DPFFR) outperform BENDY in all scenarios and for all levels of historical information. For $t^* = 3$ (i.e., only 4 months of historical information), the more sophisticated smoothing approaches (DPFR and DPFFR) did not outperform the simple regression (DLM). However, as the noise and length of the history increases, the smoothing approaches seem to have better relative prediction properties than DLM. While this is the case, we contend that models like BENDY and DLM are extremely useful. Indeed, they are more intuitive, easier to implement and require minimum changes to standard software. In practice, one may still end up choosing a

more sophisticated, penalized model, but this would be especially useful for datasets with many observations per child and large noise levels. For $t^* = 3$ and for low, moderate and noise on real data levels, dynamic_FLR is outperformed by the rest of the methods, including BENDY, while increasing historical information produces smaller MSE for dynamic_FLR.

To get a better idea about the visual differences between the predictions produced by the various models, Figure 7 displays the predicted curve for one subject up to

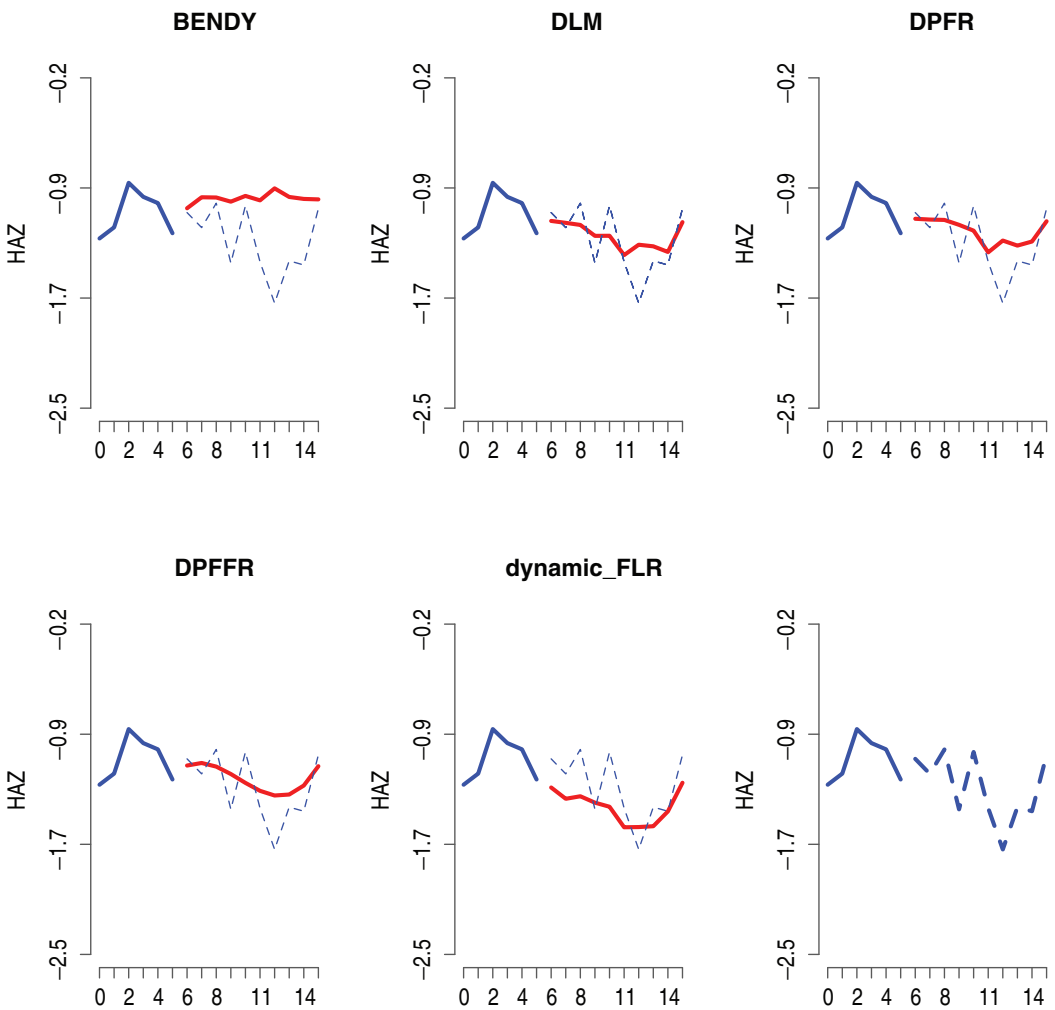

Figure 7 Displayed are predictions obtained with BENDY, DLM, DPFR, DPFFR and dynamic_FLR methods for a subject. In each case, the predicted future HAZ trajectory is displayed as solid red line. For each case, the historical monthly HAZ data for months $\{0, ..., t^*\}$ is depicted in blue solid line. The dashed line in each panel displays the observed data curve. Scenario: low noise, $t^* = 5$.

15 months based on data available for all months 0 through 5. The first four panels provide BENDY, DLM, DPFR and DPFFR predictions. The last panel displays the observed simulated data. The most striking characteristic of these predictions is that they fundamentally agree and that they are different in terms of detail, not substance. Indeed, on this plot it is hard to tell visually the difference between BENDY and DPFR, while DPFFR seems to estimate a slightly more clear pattern towards the end of the period. Interestingly, DPFFR provides very smooth predictions that are more in line with the expected pattern of growth of children. Thus, as intended, more aggressive smoothing in DPFFR results in smoother patterns with minimum bias. This is the reason why DPFFR outperforms other approaches in several scenarios. This type of behaviour was observed for multiple subject-specific data.

4.2 Comparative inference

Each dynamic prediction model contains model parameters for the first and last observed HAZ and WAZ values. In addition, the DLM, DPFR and DPFFR models account for all available historical data by including additional model components. Even though models are quite different, it makes sense to conduct inference on all models and compare results on the shared parameters, even if they have slightly different interpretations. For example, we would like to see how the sex parameter ($\gamma_{t^*,t}$) is estimated across models and check whether there are changes in the interpretation of results. Also, we can compare the parameters β_{0,t,t^*}, β_{t^*,t,t^*}, δ_{0,t,t^*} and δ_{t^*,t,t^*} across models. The model parameters corresponding to the month-specific historical effects of HAZ and WAZ data from the DPFFR model are bivariate form and are displayed in Figure 8. The shape of the estimated β parameter suggests that the recent HAZ history during months 0–5 has a large influence on the HAZ trajectories at months 6–15. The effect of WAZ (quantified by the δ parameter) is lower in magnitude compared to the effect of HAZ. This comparison can be made

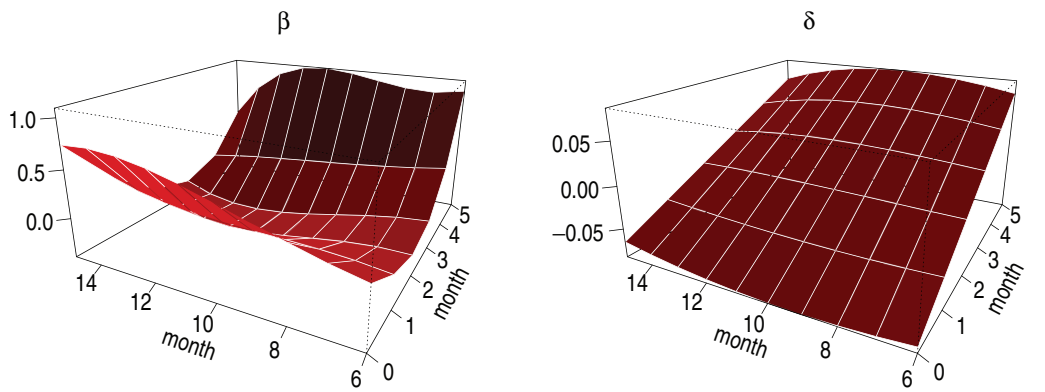

Figure 8 Displayed are the bivariate estimates for the β and δ dynamic prediction model parameters for a fit of the DPFFR method when $t^* = 5$. Scenario: low noise.

Table 2 Results for the point estimator, and, in subscript, the lower and upper bound (in round parentheses) of the confidence intervals for the dynamic regression model parameters. The dynamic regression methods predict the HAZ future trajectory at month $t = 12$ and were applied to the non-bootstrap or bootstrap data.

	Method	Low noise	Low noise	Maximum noise	Maximum noise
		Non-bootstrap	Bootstrap	Non-bootstrap	Bootstrap
$t^* = 3$					
$\gamma_{t^*,t}$ (sex)					
	BENDY	$-0.10_{(-0.24,\ 0.04)}$	$-0.01_{(-0.15,\ 0.11)}$	$-0.08_{(-0.25,\ 0.07)}$	$-0.02_{(-0.17,\ 0.13)}$
	DLM	$-0.06_{(-0.20,\ 0.07)}$	$-0.01_{(-0.15,\ 0.11)}$	$-0.08_{(-0.25,\ 0.08)}$	$-0.01_{(-0.17,\ 0.13)}$
	DPFR	$-0.07_{(-0.21,\ 0.06)}$	$-0.01_{(-0.15,\ 0.11)}$	$-0.09_{(-0.25,\ 0.06)}$	$-0.01_{(-0.17,\ 0.13)}$
	DPFFR	$-0.06_{(-0.10,\ -0.01)}$	$-0.01_{(-0.04,\ 0.02)}$	$-0.08_{(-0.12,\ -0.03)}$	$-0.01_{(-0.05,\ 0.03)}$
β_{0,t,t^*} (HAZ_0)					
	BENDY	$0.73_{(0.59,\ 0.88)}$	$0.30_{(0.14,\ 0.45)}$	$0.66_{(0.50,\ 0.81)}$	$0.30_{(0.16,\ 0.45)}$
	DLM	$0.79_{(0.52,\ 1.05)}$	$0.36_{(0.10,\ 0.62)}$	$0.57_{(0.36,\ 0.79)}$	$0.22_{(0.01,\ 0.43)}$
	DPFR	$0.84_{(0.61,\ 1.07)}$	$0.34_{(0.13,\ 0.56)}$	$0.58_{(0.40,\ 0.77)}$	$0.20_{(0.04,\ 0.35)}$
	DPFFR	$0.74_{(0.65,\ 0.83)}$	$0.36_{(0.27,\ 0.45)}$	$0.55_{(0.48,\ 0.61)}$	$0.21_{(0.14,\ 0.28)}$
β_{t^*,t,t^*} (HAZ_{t^*})					
	BENDY	$0.27_{(0.11,\ 0.43)}$	$0.64_{(0.48,\ 0.80)}$	$0.27_{(0.11,\ 0.43)}$	$0.56_{(0.40,\ 0.72)}$
	DLM	$0.54_{(0.27,\ 0.82)}$	$0.73_{(0.43,\ 1.03)}$	$0.27_{(0.06,\ 0.49)}$	$0.46_{(0.23,\ 0.68)}$
	DPFR	$0.48_{(0.22,\ 0.74)}$	$0.67_{(0.42,\ 0.92)}$	$0.21_{(0.02,\ 0.41)}$	$0.41_{(0.24,\ 0.58)}$
	DPFFR	$0.52_{(0.43,\ 0.61)}$	$0.75_{(0.64,\ 0.85)}$	$0.29_{(0.22,\ 0.35)}$	$0.46_{(0.38,\ 0.53)}$
δ_{0,t,t^*} (WAZ_0)					
	BENDY	$-0.45_{(-0.57,\ -0.32)}$	$-0.36_{(-0.49,\ -0.24)}$	$-0.36_{(-0.49,\ -0.22)}$	$-0.29_{(-0.42,\ -0.15)}$
	DLM	$-0.06_{(-0.30,\ 0.18)}$	$-0.21_{(-0.46,\ 0.03)}$	$-0.16_{(-0.35,\ 0.02)}$	$-0.24_{(-0.43,\ -0.04)}$
	DPFR	$-0.08_{(-0.30,\ 0.12)}$	$-0.24_{(-0.39,\ -0.09)}$	$-0.19_{(-0.36,\ -0.02)}$	$-0.26_{(-0.38,\ -0.13)}$
	DPFFR	$-0.09_{(-0.16,\ -0.02)}$	$-0.24_{(-0.31,\ -0.17)}$	$-0.17_{(-0.24,\ -0.11)}$	$-0.26_{(-0.32,\ -0.20)}$
δ_{t^*,t,t^*} (WAZ_{t^*})					
	BENDY	$0.34_{(0.23,\ 0.46)}$	$0.29_{(0.17,\ 0.41)}$	$0.32_{(0.19,\ 0.45)}$	$0.26_{(0.14,\ 0.39)}$
	DLM	$0.70_{(0.43,\ 0.97)}$	$0.41_{(0.13,\ 0.69)}$	$0.51_{(0.30,\ 0.71)}$	$0.29_{(0.08,\ 0.50)}$
	DPFR	$0.67_{(0.43,\ 0.91)}$	$0.36_{(0.20,\ 0.52)}$	$0.48_{(0.30,\ 0.66)}$	$0.27_{(0.15,\ 0.39)}$
	DPFFR	$0.59_{(0.50,\ 0.67)}$	$0.40_{(0.33,\ 0.48)}$	$0.42_{(0.35,\ 0.48)}$	$0.29_{(0.23,\ 0.35)}$

because both HAZ and WAZ are expressed as z-scores. The information from WAZ at month t^* (month 5) has a positive association with HAZ in the future, while WAZ at birth has a negative association with HAZ in the future.

Approximate 95% pointwise confidence intervals are reported for all model parameters. For DPFFR, we constructed non-bootstrap confidence intervals that take into account the bivariate nature of DPFFR model parameters. For example, an approximate 95% DPFFR pointwise confidence interval for β_{0,t,t^*} can be constructed as $\widehat{\beta}_{0,t,t^*} \pm 1.96\ \widehat{\text{sd}}\{\widehat{\beta}_{0,t,t^*}\}$. As $\widehat{\beta}_{0,t,t^*} = \sum_{b=1}^{\kappa} a_{b,0,t}\widehat{\beta}_{b,t^*}$, for every $(0, t)$ for a fixed t^*, $\widehat{\text{sd}}\{\widehat{\beta}_{0,t,t^*}\} = \sqrt{\boldsymbol{a}_{\boldsymbol{b},\mathbf{0},\boldsymbol{t}}\widehat{\boldsymbol{\Sigma}}\boldsymbol{a}_{\boldsymbol{b},\mathbf{0},\boldsymbol{t}}^T}$, where $\widehat{\boldsymbol{\Sigma}}$ is the estimated covariance matrix of $\widehat{\boldsymbol{\beta}}_{t^*}$, and $\boldsymbol{a}_{\boldsymbol{b},\mathbf{0},\boldsymbol{t}} = \{a_b(0, t)\}_{b=1}^{\kappa}$ is a collection of κ bivariate B-splines basis functions. We use the Bayesian posterior covariance matrix (see Ruppert et al., 2003 for more details).

Tables 2 and 3 display the average point estimators and the confidence intervals across n simulated datasets for two values of t^* (3 and 5 months) for the prediction

Table 3 Results for the point estimator, and, in subscript, the lower and upper bound (in round parentheses) of the confidence intervals for the dynamic regression model parameters. The dynamic regression methods predict the HAZ future trajectory at month $t = 12$ and were applied to the non-bootstrap or bootstrap data.

	Method	Low noise Non-bootstrap	Low noise Bootstrap	Maximum noise Non-bootstrap	Maximum noise Bootstrap
$t^* = 5$					
$\gamma_{t^*,t}$ (sex)					
	BENDY	$-0.06_{(-0.17,\,0.03)}$	$-0.00_{(-0.11,\,0.09)}$	$-0.04_{(-0.17,\,0.08)}$	$-0.00_{(-0.14,\,0.12)}$
	DLM	$0.02_{(-0.06,\,0.10)}$	$-0.00_{(-0.09,\,0.09)}$	$0.00_{(-0.12,\,0.12)}$	$-0.00_{(-0.13,\,0.12)}$
	DPFR	$0.02_{(-0.06,\,0.10)}$	$-0.00_{(-0.09,\,0.08)}$	$0.00_{(-0.12,\,0.12)}$	$-0.00_{(-0.13,\,0.12)}$
	DPFFR	$0.00_{(-0.01,\,0.03)}$	$-0.00_{(-0.03,\,0.03)}$	$-0.01_{(-0.05,\,0.02)}$	$-0.00_{(-0.04,\,0.03)}$
β_{0,t,t^*} (HAZ_0)					
	BENDY	$0.45_{(0.35,\,0.55)}$	$0.29_{(0.18,\,0.39)}$	$0.40_{(0.28,\,0.51)}$	$0.29_{(0.17,\,0.40)}$
	DLM	$0.64_{(0.47,\,0.80)}$	$0.50_{(0.31,\,0.69)}$	$0.45_{(0.28,\,0.62)}$	$0.32_{(0.15,\,0.49)}$
	DPFR	$0.70_{(0.57,\,0.83)}$	$0.52_{(0.37,\,0.67)}$	$0.51_{(0.37,\,0.64)}$	$0.33_{(0.19,\,0.46)}$
	DPFFR	$0.69_{(0.63,\,0.75)}$	$0.50_{(0.43,\,0.57)}$	$0.50_{(0.43,\,0.56)}$	$0.31_{(0.25,\,0.37)}$
β_{t^*,t,t^*} (HAZ_{t^*})					
	BENDY	$0.63_{(0.53,\,0.72)}$	$0.68_{(0.58,\,0.78)}$	$0.61_{(0.50,\,0.72)}$	$0.62_{(0.51,\,0.74)}$
	DLM	$1.03_{(0.86,\,1.20)}$	$0.93_{(0.73,\,1.13)}$	$0.73_{(0.56,\,0.91)}$	$0.63_{(0.44,\,0.81)}$
	DPFR	$1.06_{(0.92,\,1.20)}$	$0.93_{(0.77,\,1.09)}$	$0.75_{(0.60,\,0.89)}$	$0.61_{(0.46,\,0.77)}$
	DPFFR	$1.05_{(0.99,\,1.12)}$	$0.91_{(0.84,\,0.99)}$	$0.75_{(0.68,\,0.82)}$	$0.60_{(0.54,\,0.67)}$
δ_{0,t,t^*} (WAZ_0)					
	BENDY	$-0.27_{(-0.35,\,-0.20)}$	$-0.22_{(-0.31,\,-0.13)}$	$-0.22_{(-0.31,\,-0.13)}$	$-0.18_{(-0.28,\,-0.08)}$
	DLM	$-0.02_{(-0.17,\,0.12)}$	$0.00_{(-0.18,\,0.18)}$	$-0.05_{(-0.21,\,0.09)}$	$-0.05_{(-0.22,\,0.11)}$
	DPFR	$-0.07_{(-0.13,\,-0.01)}$	$-0.01_{(-0.10,\,0.08)}$	$-0.11_{(-0.19,\,-0.04)}$	$-0.07_{(-0.16,\,0.02)}$
	DPFFR	$-0.07_{(-0.09,\,-0.05)}$	$0.00_{(-0.04,\,0.05)}$	$-0.10_{(-0.13,\,-0.07)}$	$-0.04_{(-0.09,\,-0.00)}$
δ_{t^*,t,t^*} (WAZ_{t^*})					
	BENDY	$0.19_{(0.12,\,0.27)}$	$0.17_{(0.08,\,0.25)}$	$0.20_{(0.12,\,0.29)}$	$0.18_{(0.08,\,0.27)}$
	DLM	$0.21_{(0.03,\,0.39)}$	$0.22_{(0.02,\,0.41)}$	$0.23_{(0.06,\,0.41)}$	$0.21_{(0.03,\,0.39)}$
	DPFR	$0.11_{(0.05,\,0.18)}$	$0.20_{(0.09,\,0.30)}$	$0.15_{(0.08,\,0.22)}$	$0.19_{(0.09,\,0.29)}$
	DPFFR	$0.08_{(0.06,\,0.11)}$	$0.19_{(0.14,\,0.24)}$	$0.12_{(0.09,\,0.15)}$	$0.19_{(0.14,\,0.24)}$

of HAZ at $t = 12$ months. Results are shown across various models and only for the shared parameters $\gamma_{t^*,t}$ (sex), β_{0,t,t^*} (HAZ at birth) β_{t^*,t,t^*} (last known HAZ), δ_{0,t,t^*} (WAZ at birth) δ_{t^*,t,t^*} (last known WAZ). Point estimators are shown in larger font and confidence intervals are shown as indices in smaller font. While, in general, point estimators do not have the same interpretation across models, it is reassuring that there are no major differences between them for a fixed pair (t^*, t). The sex effect $(\gamma_{t^*,t})$ tends to be statistically insignificant across a wide range of models. We attribute the lack of significance to the fact that the marginal sex effect is relatively weak in the dataset and is probably accounted for in the conditional models by using the last known HAZ value. All models agree that the effects of baseline and last known value of HAZ are highly significant. Similarly, the last value of WAZ is statistically significant in all models indicating that the last observed WAZ contains predictive information above and beyond the baseline and last known HAZ measurements. Models also tend to agree that once these covariates are accounted for, the baseline

Table 4 Results of inference (AIC, adjusted R^2) for predictive modelling.The dynamic regression methods predict the future curve of HAZ at months $t \in \{ t^* + 1, ..., 15\}$.The AIC label refers to AIC/10^4.

	Method	Low noise	Mild noise	Maximum noise	Very high noise
$t^* = 3$					
AIC					
	BENDY	47.98	51.03	66.89	103.67
	DLM	45.60	48.97	65.17	101.81
	DPFR	45.06	48.40	64.51	100.98
	DPFFR	46.26	49.22	63.98	99.52
Adjusted R^2					
	BENDY (%)	81.99	81.05	75.09	54.79
	DLM (%)	83.25	82.23	76.42	57.42
	DPFR (%)	83.29	82.27	76.48	57.45
	DPFFR (%)	83.33	82.33	76.62	57.88
$t^* = 5$					
AIC					
	BENDY	13.59	17.64	37.82	79.08
	DLM	3.89	10.36	36.12	77.93
	DPFR	2.54	9.02	34.81	76.55
	DPFFR	1.68	8.11	33.75	75.29
Adjusted R^2					
	BENDY (%)	91.17	90.31	84.57	62.89
	DLM (%)	93.55	92.42	85.85	65.35
	DPFR (%)	93.58	92.45	85.88	65.42
	DPFFR (%)	93.61	92.50	86.01	65.81
$t^* = 7$					
AIC					
	BENDY	−1.27	3.38	24.17	61.05
	DLM	−9.17	−3.29	20.49	58.36
	DPFR	−10.60	−4.72	18.99	56.69
	DPFFR	−11.21	−5.37	18.26	55.86
Adjusted R^2					
	BENDY (%)	94.05	93.15	87.24	65.11
	DLM (%)	95.67	94.79	89.22	69.68
	DPFR (%)	95.66	94.79	89.24	69.81
	DPFFR (%)	95.72	94.86	89.40	70.27

WAZ seems to significantly impact prediction for datasets with large noise levels and where $t - t^*$ is large.

Results also indicate that, for all models, the longer the length of the historic data, the narrower the confidence intervals for the model parameters. The average width of the confidence interval is shortest for DPFFR, though DPFR has an improved performance for longer available histories. Both DPFFR and DPFR provide shortened confidence intervals on all parameters, especially when the length of the available history increases. The length of the confidence intervals for DPFR is comparable to DLM in the case of the sex parameter. For the other parameters, the performance of DPFR improves substantially and the lengths of confidence intervals become comparable to those obtained using DPFFR. In the non-bootstrap analysis results

of Table 4, the AIC and adjusted R^2 metrics were obtained from the mixed model fitting software (Wood, 2006), as provided by the `gam` model fit obtained using the `mgcv` (Wood, 2016) R package. These metrics are described in Wood (2006). We report cross-validated AIC, adjusted R^2 metrics by running the models n times on datasets of size $n - 1$. The results reported in Table 4 are the averages of the model fits. AIC results indicate that there is not much difference between the various models when $t^* = 3$ irrespective of the size of the measurement error. This makes sense, because when $t^* = 3$, there are only two additional observations that are not being used by BENDY, $l = 1, 2$. When $t^* = 5$ and 7, the advantage of models incorporating the entire history becomes much larger, indicating that the subtle changes in the subject-specific history contain substantial additional information. A similar story is told by the adjusted R^2 measurements, though these results also indicate that BENDY provides reasonable predictions. This should not be surprising, given our exploratory data analyses in Section 1.2.

4.3 Imputation and bootstrap for dynamic prediction

Using the original data, we generate a bootstrap sample by randomly drawing n subjects with replacement. For each bootstrap sample, we apply a data imputation mechanism based on fPCA by using the `fpca.sc` (Goldsmith et al., 2013) function in the refund (Goldsmith et al., 2016) R package. The dynamic prediction methods described (BENDY, DLM, DPFR and DPFFR) are then applied to each bootstrap sample data. A bootstrap estimator is obtained for each parameter. For example, $\hat{\gamma}^b_{t^*,t}$, is the estimator for $\gamma_{t^*,t}$ in the b-th bootstrap sample, $1 \leq b \leq B$, whereas the final bootstrap-based estimator is the average across the corresponding B estimators. We took $B = 100$ bootstrap samples. Results for the parameters are presented in Tables 2 and 3 for two choices of t^* for predictions at $t = 12$. For the $\gamma_{t^*,t}$ parameter, the bootstrap-estimated effect is only slightly changed in magnitude, and unchanged in significance. The bootstrap analyses yield a magnified effect for the point estimate for the β_{t^*,t,t^*} parameter, corresponding to the last historical HAZ data available, irrespective of the noise setting for the short history case, while the significance of the findings remains unchanged. For the parameter corresponding to the HAZ data at birth, the bootstrap analyses show a reduction in magnitude for the β_{0,t,t^*} parameter, irrespective of noise level and length of history, while the significance of this term is unchanged. For the δ_{0,t,t^*} parameter, the bootstrap-estimated effect is increased in magnitude for the case of short history, irrespective of noise level. For short history, the bootstrap results for DPFR and DPFFR yield a significant effect for δ_{0,t,t^*}, irrespective of noise level, while for longer history, the significance of δ_{0,t,t^*} changes to become insignificant for low noise, and then changes to significant with DPFFR at higher noise level. The δ_{t^*,t,t^*} parameter bootstrap estimator is smaller at $t^* = 3$ and larger for $t^* = 5$ for most of the methods, compared to the analysis on the original data. Results using the analysis on the original data are overall in agreement with bootstrap-based methods, though some differences exist. We have found this process to be particularly useful in cases when parameters are close to being statistically significant.

Table 5 Comparison of features provided by each dynamic prediction method. Features are labelled A1 through A8 and are described as follows: (A1) Produces full k-steps ahead prediction, (A2) Smooth predictions, (A3) Produces predictions with sparse designs, (A4) Prediction performance (low noise and short history), (A5) Prediction performance (high noise and short history), (A6) Prediction performance (high noise and long history), (A7) Intuitive approach and (A8) Simplicity of implementation. For features A4–A6, we list the two top performers based on the MSE results reported in Table 1.

Method	A1	A2	A3	A4	A5	A6	A7	A8
BENDY			✓				✓	✓
DLM			✓				✓	
DPFR			✓	✓	✓	✓		
DPFFR	✓	✓	✓	✓	✓	✓		

5 Discussion

Modelling and predicting child growth is a fundamental problem that needs to be addressed. Indeed, understanding the relationship between growth and baseline and time-varying covariates can provide the infrastructure for planning future studies, calculating sample size for intervention trials, and designing effective interventions and studies. Here we have proposed a simple-to-complex class of models, where the simplest approach (BENDY) could be implemented by any data analyst with knowledge of linear regression. The most sophisticated approaches come with their own software implemented in R and can easily be deployed. Thus, together with the accompanying software tools, we have significantly expanded the inferential capabilities for child growth prediction and inference. Table 5 provides a comparison of BENDY, DLM, DPFR and DPFFR across eight characteristics for predictive modelling. In addition to providing and comparing prediction models, we have introduced an inferential framework for cases when data are observed at irregular time intervals and with seizable noise. In this case, we propose to use functional principal component approaches to smooth and impute the data. To account for the variability introduced by this step, we propose to use a bootstrap of subjects followed by a subsequent fusion of results on these datasets. This approach increased some of the confidence intervals that were obtained using only a one-step procedure, though they incorporate more of the observed variability.

We conclude that more sophisticated dynamic smoothing approaches (such as DPFR and DPFFR) can substantially improve prediction, though there is a penalty on interpretability of coefficients and results. Moreover, these methods work better on studies where data are collected quite densely, as in the case of our study. It is likely that data from more traditional studies will simply not contain the level of detail necessary for sophisticated algorithms to detect. For these reasons, we consider that BENDY should be the benchmark prediction model: it is simple and provides a reasonable 'floor' for the performance of more sophisticated approaches. We also conclude that inferential results remain quite robust across models. Having a range of models provides an excellent platform for sensitivity analysis. Running such sensitivity analyses should become routine. One surprising finding was that

the DLM can be very unstable in cases of low noise and dense sampling. This problem is likely due to the large correlation of predictors and can be addressed either by simplifying the model to BENDY or by using penalized dynamic models, as introduced here.

Compared to existing methods on child growth studies, we have significantly expanded the analysis of prediction of child growth to the case of dynamic prediction of future HAZ. While some existing studies, such as Howe et al. (2013) and Grajeda et al. (2016), study the modelling and prediction of child height using mixed effects and splines, and we have highlighted in this article a different framework for modelling longitudinal data for child growth and dynamic prediction. With this new approach, we introduced a novel method for identifying the importance of child growth predictors and their contribution to future HAZ. Mainly, the study we conducted showed that HAZ history is very relevant, and the largest influence was observed for the first and last observation from the HAZ history. HAZ history as well as WAZ history were considered as predictors of growth for future HAZ. While WAZ history had relative smaller impact, our study showed that the last observation from the WAZ history had a positive impact on the future HAZ.

There are many remaining open problems that we have not addressed here. For example, one could be interested in the effect of choice of data imputation and smoothing method. There is also a need for formal hypothesis testing combined with the study of validity of methods for fitting complex DPFR and DPFFR models. Finally, while we have used fast, off-the-shelf smoother options, improved prediction may be achieved with dedicated priors and smoothers that remain to be proposed.

Acknowledgements

Ciprian M Crainiceanu and William Checkley were supported by a Bill & Melinda Gates Foundation Grant (OPP1114097).

References

Carey VJ, Yong FH, Frenkel LM and McKinney RM (2004) Growth velocity assessment in pediatric aids: Smoothing, penalized quantile regression and the definition on growth failure. *Statistics in Medicine*, **23**, 509–26.

Checkley W, Epstein L, Gilman R, Black R, Cabrera L and Sterling C (1998) Effects of cryptosporidium parvum infection in peruvian children: Growth faltering and subsequent catch-up growth. *American Journal of Epidemiology*, **148**, 497–506.

Checkley W, Gilman R, Black R et al. (2002) Effects of nutritional status on diarrhea in peruvian children. *The Journal of Pediatrics*, **140**, 210–18.

Cheung YB (2014) *Statistical Analysis of Human Growth and Development*. Boca Raton, FL: Chapman & Hall and CRC Press.

Chiou JM (2012) Dynamical functional prediction and classification, with application to traffic flow prediction. *The Annals of Applied Statistics*, **6**, 1588–614. DOI: 10.1214/12–AOAS595

Chiou J-M, Yang Y-F and Chen Y-T (2016) Multivariate functional linear regression and prediction. *Journal of Multivariate Analysis*, **146**, 301–12. DOI: 10.1016/j.jmva.2015.10.003

Goldberg Y, Ritov Y and Mandelbaum A (2014) Predicting the continuation of a function with applications to call center data. *Journal of Statistical Planning and Inference*, **147**, 53–65. DOI: 10.1016/j.jspi.2013.11.006

Goldsmith J, Bobb J, Crainiceanu CM, Caffo BS and Reich D (2011) Penalized functional regression. *Journal of Computational and Graphical Statistics*, **20**, 830–51. DOI: 10.1198/jcgs.2010.10007

Goldsmith J, Greven S and Crainiceanu CM (2013) Corrected confidence bands for functional data. *Biometrics*, **69**, 41–51. DOI: 10.1111/j.1541–0420.2012.01808.x

Goldsmith J, Scheipl F, Huang L et al. (2016) *Refund: Regression with functional data*. URL http://CRAN.R–project.org/package=refund

Grajeda LM, Ivanescu AE, Saito M et al. (2016) Modelling subject-specific childhood growth using linear mixed-effect models with cubic regression splines. *Emerging Themes in Epidemiology*, **13**. DOI 10.1186/s12982–015–0038–3

Howe LD, Tilling K, Galobardes B, Smith GD, Gunnell D and Lawlor DA (2012) Socioeconomic differences in childhood growth trajectories: At what age do height inequalities emerge? *Journal of Epidemiology and Community Health*, **66**, 143–48. DOI: 10.1136/jech.2010.113068

Howe LD, Tilling K, Matijasevich A et al. (2013) Linear spline multilevel models for summarizing childhood growth trajectories: A guide to their application using examples from five birth cohorts. *Statistical Methods in Medical Research*, 1–21. DOI: 10.1177/0962280213503925

Hyndman RJ and Shang HL (2016) *ftsa: Functional time series analysis*. URL http://CRAN.R–project.org/package=ftsa

Ivanescu AE, Staicu A-M, Scheipl F and Greven S (2015) Penalized function-on-function regression. *Computational Statistics*, **30**, 539–68. DOI: 10.1007/s00180–014–0548–4

Johnson W (2015) Analytical strategies in human growth research. *American Journal of Human Biology*, **27**, 69–83.

Karlberg J, Jalil F, Lam B, Low L and Yeung CY (1994) Linear growth retardation in relation to the three phases of growth. *European Journal of Clinical Nutrition*, **48**, 25–43.

Ruppert D, Wand MP and Caroll RJ (2003) *Semiparametric regression*. Cambridge, UK: Cambridge University Press.

Scheipl F, Staicu A-M and Greven S (2015) Functional additive mixed models. *Journal of Computational and Graphical Statistics*, **24**, 477–501. DOI: 10.1080/10618600.2014.901914

Staniswalis J and Lee J (1998) Nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association*, **93**, 1403–18.

Tanner JM, Whitehouse RH, Marshall WA and Carter BS (1975) Prediction of adult height from height, bone age, and occurrence of menarche, at ages 4 to 16 with allowance for mid parent height. *Archives of Disease in Childhood*, **50**, 14–26.

Tanner JM, Landt KW, Cameron N, Carter BS and Patel J (1983) Prediction of adult height from height and bone age in childhood. Archives of *Disease in Childhood*, **58**, 767–76.

Tilling K, Macdonald-Wallis C, Lawlor DA, Hughes RA and Howe LD (2014) Modelling childhood growth using fractional polynomials and linear splines. *Annals Nutrition Metabolism*, **65**, 129–38.

Wei Y and He X (2006) Conditional growth charts. *The Annals of Statistics*, **34**, 2069–97. DOI: 10.1214/009053606000000623

WHO (2016) *Global database on child growth and malnutrition*. URL http://www.who.int/nutgrowthdb/about/introduction/en/index2.html

Wood SN (2006) *Generalized Additive Models: An Introduction with R*. New York, NY: Chapman & Hall/CRC Press.

Wood SN (2016) *mgcv: Mixed gam computation vehicle with gcv/aic/reml smoothness*

estimation. URL http://CRAN.R–project.org/package=mgcv

Xiao L, Ruppert D, Zipunnikov V and Crainiceanu C (2016) Fast covariance estimation for high-dimensional functional data. *Statistics and Computing*, **26**, 409–21. DOI: 10.1007/s11222–014–9485–x

Yao F, Müller H-G and Wang J-L (2005) Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, **100**, 577–90. DOI 10.1198/016214504000001745

Zhang W and Wei Y (2015) Regression based principal component analysis for sparse functional data with applications to creening growth paths. *The Annals of Applied Statistics*, **9**, 597–620. DOI: 10.1214/15–AOAS811