European Journal of Clinical Nutrition (2016) **70**, 656–661

www.nature.com/ejcn

ORIGINAL ARTICLE

# Modelling human height and weight: a Bayesian approach towards model comparison

M Preising[1], A Suchomlinov[2], J Tutkuviene[2] and C Aßmann[3]

**BACKGROUND/OBJECTIVES:** Given the availability of large longitudinal data sets on human height and weight, different modelling approaches are at hand to access quantities of interest relating to important diagnostic aims.

**SUBJECTS/METHODS:** Statistical modelling frameworks for longitudinal data on human height and weight have to consider the issues of individual heterogeneity and time dependence to provide an accurate statistical characterisation. Further, missing values inevitably occurring within longitudinal data sets have to be addressed adequately to allow for valid inference. The Bayesian framework is illustrated to facilitate stringent comparison of available non-nested model frameworks addressing these issues using simulated and empirical data sets.

**RESULTS:** Comparing random-effects and fixed-effects modelling approaches with the Preece–Baines (PB) model reveals that, for simulated data, the Bayesian approach towards model comparison is effective in discriminating between different model specifications. With regard to analysis of 14 longitudinal data sets, the implicit trade-off between model fit, that is, description of the data, and a parsimonious parameterisation favouring prediction is often best addressed via the PB model.

**CONCLUSIONS:** The Bayesian approach is illustrated to allow for effective comparison in case model specifications for longitudinal data are not linked directly via parametric restrictions.

*European Journal of Clinical Nutrition* (2016) **70,** 656–661; doi:10.1038/ejcn.2016.23; published online 30 March 2016

## INTRODUCTION

Human growth charts serve as well-accepted indicators for measuring the population and in particular the health status of children.[1] Therefore, several statistical approaches were developed to model these curves. As a range of more traditional analysis models consist of rather parsimonious specifications, such as the Preece–Baines (PB) curve for height in puberty,[2] the Jenss–Bayley curve for weight in early life[3] or the least mean square approach of Cole–Green,[4] rather recent setups available in the literature include latent conditioning factors of growth.[5] The advantages of these latter models are their flexibilities in terms of simultaneous curve fits to the surveyed subjects, meaning that a single growth curve can be fitted to each subject by adjusting only a small number of (latent) parameters.

The aim of this paper is to point at a valid model comparison approach for non-nested statistical models incorporating at different degrees individual- and time-specific heterogeneity observable within the growth data. Each statistical criterion allowing for model comparison has to address the trade-off between model fit and model parsimony. This trade-off is operationalised in the Bayesian approach by setting up prior distributions quantifying the parameter uncertainty and thus the informational content of a model parameter. Therefore, and because of missing values that occur within available longitudinal data, which can be stringently accounted for via the device of data augmentation, we choose a Bayesian approach for our analyses and compare the (non-)nested models by calculating marginal likelihoods. To check the accuracy of the suggested approach towards model comparison, we specify fixed- and random-effects models, as well as a mixed-model approach. Furthermore, these models will be enhanced by a dynamic component. The PB model serves as a benchmark to our specifications.

## SUBJECT AND METHODS

### Statistical models

The set of considered statistical models includes the linear fixed-effects panel data model for $i = 1,\ldots, N$ individuals and $t = 1,\ldots,T$ time periods. Note that linear models, although restrictive with regard to possibilities to capture time dependencies, offer direct possibilities to incorporate external regressors—for example, parental height. Hence, the data-generating process (DGP) can be stated as follows:

$$y_{it} = c_i + a_t + X_{it}\beta + \varepsilon_{it}$$

with $y_{it}$ and $X_{it}$ denoting the dependent and independent variables of individual $i$ in period $t$. $c_i$ indicates an individual-specific and $a_t$ a time-specific intercept. The error terms $\varepsilon_{it}$ are assumed to be independently normally distributed with mean zero and time-specific variance denoted as $\sigma_t^2$. Summarising all model parameters as $\theta$, this DGP yields the likelihood function in matrix notation

$$L_{FE}(y|\theta, X) = \prod_{t=1}^{T} (2\pi)^{-\frac{N}{2}} |\Sigma_t|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(y_t - c - \iota a_t - X_t\beta)' \Sigma_t^{-1} (y_t - c - \iota a_t - X_t\beta)\right\}$$

with $y_t = (y_{1t}, y_{2t}, \ldots, y_{Nt})'$, $X_t$ as time-specific design matrix possibly including covariates of interest, $\Sigma_t$ as time-specific $(N \times N)$ covariance matrix containing $N$ time-specific disturbance variances $\sigma_t^2$ on its main diagonal, $c = (c_1,\ldots,c_N)'$ and $\iota = (1, 1, \ldots, 1)'$ as an additional $N$-dimensional vector of ones for indicating the $N$ time-specific intercepts $a_t$ for period $t$. An advantage of the fixed-effects estimation is its flexibility to consider latent heterogeneity and thereby typically an exceptional accurate fit, for example, for the purpose of child-specific height curve behaviours. As this modelling approach provides $N+T-2$ time- and individual-specific

[1]Department of Statistics and Econometrics, Otto-Friedrich-Universität, Bamberg, Germany; [2]Department of Anatomy, Histology and Anthropology, Faculty of Medicine, Vilnius University, Vilnius, Lithuania and [3]Leibniz Institute for Educational Trajectories, Bamberg, Germany. Correspondence: M Preising, Department of Statistics and Econometrics, Otto-Friedrich-Universität Bamberg, Feldkirchenstraße 21, Bamberg 96045, Germany.
E-mail: marcel.preising@uni-bamberg.de
Received 16 December 2015; accepted 6 January 2016; published online 30 March 2016

intercepts, the inherent identification problem is solved by restricting one time- and individual-specific intercept to zero each. Further note that characterising measurements via time- or individual-specific effects is data consuming in the sense that these effects increase the requirements with regard to the information set available for prediction. Nevertheless, this does not rule out these characterisations of measurements *per se*.

An alternative specification is given in terms of a random-effects approach, where the components of $c=(c_1,...,c_N)'$ and $a=(a_1,\ldots,a_T)'$ are assumed to be independently identically normally distributed with mean zero and variance $\omega_c^2$ and $\omega_a^2$, respectively. Hence, in case of the random-effects model only one time- and individual-specific variance parameter each are estimated, allowing for more restrictive but also more parsimonious forms for heterogeneity compared with the fixed-effects approach given above. The corresponding likelihood is given as:

$$L_{RE}(y|\theta,X)=\iint L_{FE}(y|\theta,X)f(c,a)dcda.$$

The fixed-effects as well as the random-effects models can be extended by a further component, in more detail an autoregressive first-order process capturing time dependencies. An additional parameter $\rho$ controls for the regression of a child's height in period $t$ on the same child's height in period $t$-1, that is,

$$y_{it}=\rho y_{it-1}+c_i+a_t+X_{it}\beta+\varepsilon_{it}.$$

The corresponding likelihood functions can be formulated as:

$$L_{AR,FE}(y|\theta,X)=\prod_{i=1}^{N}\prod_{t=2}^{T}(2\pi)^{-\frac{1}{2}}\left(\frac{1}{\sigma_t^2}\right)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma_t^2}(y_{it}-c_i-a_t-X_{it}\beta-\rho y_{it-1})^2\right\}$$

and

$$L_{AR,RE}(y|\theta,X)=\iint L_{AR,FE}(y|\theta,X)f(c,a)dcda.$$

Further, we specify a mixed model in which the time-specific intercepts $a_t$ are assumed as fixed effects, whereas the individual-specific intercepts $c_i$ are taken into account as random effects. Hence, six model specifications are available for the analysis in total, a fixed, a random and a mixed model setup, each formulated with and without a dynamic process. As a benchmark to test our specifications, we use the PB model, which is formulated as follows:

$$y_{it}=H_T-\frac{2(H_T-H_\theta)}{\exp\{s_0(t-\xi)\}+\exp\{s_1(t-\xi)\}}+\varepsilon_{it},$$

where $H_T$ indicates the adult height, $s_0$ and $s_1$ denote rate constants and $H_\theta$ and $\xi$ control for the time and age of the adolescent growth spurt. With regard to $\varepsilon_{it}$, the same properties as for the fixed- and random-effects models are assumed; however, variances of $\varepsilon_{it}$ are assumed as homoscedastic.

## Bayesian estimation and model comparison

We adapt a Bayesian approach towards model comparison for our analyses, as for one thing this allows for a non-nested model comparison, whereas for another thing it allows for accounting of the missing values occurring within longitudinal data sets. The approach is prominent in the statistical literature,[6,7] but very much less so in case missing values occur within longitudinal data sets. Calculating the marginal likelihood and simultaneously imputing the missing values using data augmentation[8] provides a solution to both challenges. Technically, the missing values are filled in at every iteration of the Gibbs sampler via draws from the corresponding full conditional distributions. It is specific to the considered setups that either the likelihood of observed measurements only is directly at hand or the full conditional distributions of missing values can be derived directly from the assumed joint distribution in case of complete observations. Therefore, the objective of a multidimensional Bayesian model framework is to calculate the joint posterior distribution of a random variable parameter vector $\theta$ augmented to include possible missing values, that is $p(\theta|y, X)$, which can, by Bayesian rule, be formulated as:

$$p(\theta|y,X)=\frac{L(y|\theta,X)\pi(\theta)}{M(y|X)},$$

and to deduce the quantities of interest—for example, the posterior moments serving as parameter estimators. In the expression, $\pi(\theta)$ denotes the prior distribution of the parameters, the likelihood function and $M(y|X)$ the marginal likelihood as a normalising constant. To assess these moments in the context of the six described model setups, we apply a Gibbs sampler, meaning that we approximate the joint posterior distribution by deriving the full conditional posterior distributions of the parameters in closed form and sample iteratively from these conditionals.[9] Hence, conjugate prior distributions are chosen for the model parameters. A complete description of the set of full conditional distributions for the different model specifications is available upon request from the corresponding author.

As neither sampling from the joint nor the full conditional posterior distributions of the PB model parameters is directly possible, a Metropolis–Hastings algorithm provides a solution. Therefore, a multivariate normal distribution serves as the proposal density with mean vector and covariance matrix obtained from the maximisation of the likelihood function corresponding to the PB model (Note that the Metropolis–Hastings algorithm may require tuning of the proposal density chosen as normal possibly not closely matching the posterior over the complete parameter space). The possibility of an acceptance of a current candidate from the proposal density is then calculated in relation to the previous draw.

As the logarithm of the marginal likelihood can be expressed by

$$\log M(y|X)=\log L(y|\theta^*,X)+\log\pi(\theta^*)-\log p(\theta^*|y,X),$$

where $\theta^*$ in our context denotes the arithmetic means of the sampled posterior draws, their densities are to be more accurate at a high-density point with more samples available. As the corresponding log-likelihood $\log L(y|\theta^*,X)$ and log prior $\log\pi(\theta^*)$ can be directly calculated, we apply the propositions by Chib,[10] in cases of Gibbs sampling, and by Chib and Jeliazkov,[11] for the Metropolis-Hastings output to approximate the logarithm of the posterior distributions $\log p(\theta^*|y,X)$. Thus, as not only the log-likelihood but also the prior and posterior distribution ordinates are taken into account for the model evaluations, the latter and therefore the marginal likelihood decreases by a higher number of model parameters by default. Hence, the log marginal likelihood penalises less parsimonious models such as the fixed-effects, even if the actual log-likelihood prefers the more extensive setup.

# RESULTS

## Design and results of simulation study

To assess the properties of the suggested approach, the simulation study analyses four data scenarios (I – IV). First, data were simulated setting a dynamic fixed-effects model as the true DGP (I). The second DGP is a dynamic random-effects model (II), whereas scenarios three and four also suppose fixed (III) and random effects (IV) but without autoregressive process. Each DGP consists of 100 individuals and 19 time periods and includes three covariates. Therefore, we simulate an intercept, three slope parameters and, in cases of dynamic models, a corresponding correlation parameter $\rho$. The error disturbances are set to be normally distributed with mean zero and time-specific variance $\sigma_t^2=\frac{1}{10}t$. In the fixed-effects context, the individual- as well as the time-specific intercepts are set as random draws from a continuous uniform distribution ranging from minus to plus five. For the random-effects model, individual- and time-specific intercepts are independently normally distributed with mean 0 and s.d. 0.03. Furthermore, 20% of the observations of the dependent variable $y_{it}$, and in the dynamic model settings thus also of the covariate $y_{it-1}$, are completely randomly set to be missing. Each DGP is applied to simulate 100 data sets, whereas all of them are analysed by all the four models performing 8000 Gibbs sampling draws after a burn-in period of 2000. To evaluate the models, we count the frequency in which the algorithm chose the actual correct model as the superior one both in terms of the likelihood $L(y|\theta,X)$ by calculating likelihood ratio tests and the marginal likelihood $M(y|X)$. Furthermore, coverages of the intercepts, slopes and, for scenarios one and two, parameters related to the dynamic process are presented. Hence, in each analysis, the 95% highest posterior density regions are checked for covering the true value of the DGP. Conjugate almost uninformative prior distributions with assumed prior independence of the parameters are set as presented in Table 1.

Table 2 presents the results of the four simulation scenarios. In scenario I, that is the dynamic fixed-effects model as the DGP, the actual correct analysis model provides coverages of the intercept of 94%, whereas the estimations of the slope parameters yield even better results. In case of the parameter that controls for the dynamic process $\rho$, the estimation achieves coverages of 91%. By comparison, the dynamic random-effects modelling only shows coverages of 76% for the intercept but instead 97% for the dynamic component. Estimations of the non-dynamic models show biased intercept parameters with coverages of 13% and 3%, respectively. As presumed, the likelihood criterion shows superior results for the dynamic fixed-effects analyses in all over the 100 data sets compared with the three other specifications. The evaluation by the marginal likelihood shows nearly the same result, except that in only 94% the dynamic fixed-effects model estimates superiorly compared with the dynamic random-effects specification. This result could hint at the advantage of the more parsimonious modelling of the random-effects model and seems to be affirmed by the second block of results.

In case of the dynamic random-effects model as true DGP, that is, scenario II, the coverages are in line with those of scenario I. However, the marginal likelihoods now illustrate the superiority of the dynamic random-effects model, although the likelihood criterion, which do not attempt to the trade-off of a high number of parameters, prefers the less parsimonious dynamic fixed-effects specification.

Although the fourth DGP scenario affirms the findings of the first two concerning in relation to the superior marginal likelihood of the correct model, scenario three with its non-dynamic fixed effects differs as the actual correct analysis model shows inferior marginal likelihood in every data set compared with both the non-nested and nested random-effects models. It seems that the latter models are able to cover even the non-normally distributed individual- and time-specific intercepts sufficiently. Vice versa, the non-dynamic fixed-effect model approach loses ground when concerned in relation to the marginal likelihoods, although it is preferred in terms of the likelihood ratio tests. Thus it seems that the uninformative prior distributions with their huge variances of 1000 tend to decrease the marginal likelihood and hence the actual performance of the fixed-effects model considerably.

Summarised, in three of four scenarios the simulation study confirms the algorithm accuracies in terms of high coverages and

**Table 1.** Prior distributions

| | Parameter | Distribution | Mean | Variance |
|---|---|---|---|---|
| *Fixed-effects specification* | | | | |
| | $\beta_1$ | N | 0 | 1000 |
| | $c_i$, $i=1,\ldots, N$-1 | N | 0 | 1000 |
| | $a_t$, $t=1,\ldots, T$-1 | N | 0 | 1000 |
| | $\sigma_t^2$, $t=1,\ldots, T$ | IG | 1 | 1 |
| (dyn.) | $\rho$ | N | 0 | 1000 |
| *Random-effects specification* | | | | |
| | $\beta_1$ | N | 0 | 1000 |
| | $\sigma_\varsigma^2$ | IG | 1 | 1 |
| | $\sigma_g^2$ | IG | 1 | 1 |
| | $\sigma_t^2$, $t=1,\ldots, T$ | IG | 1 | 1 |
| (dyn.) | $\rho$ | N | 0 | 1000 |
| *Mixed model specification* | | | | |
| | $\beta_1$ | N | 0 | 1000 |
| | $\sigma_c^2$ | IG | 1 | 1 |
| | $a_t$, $t=1,\ldots, T$-1 | IG | 1 | 1 |
| | $\sigma_t^2$, $t=1,\ldots, T$ | IG | 1 | 1 |
| (dyn.) | $\rho$ | N | 0 | 1000 |
| *Preece–Baines specification* | | | | |
| | $H_T$ | N | 0 | 1000 |
| | $H_\theta$ | N | 0 | 1000 |
| | $s_0$ | N | 0 | 10 |
| | $s_1$ | N | 0 | 10 |
| | $\xi$ | N | 0 | 1000 |
| | $\sigma_t^2$ | IG | 1 | 1 |

Abbreviations: IG, inverse gamma distributed; N, normal distributed.

**Table 2.** Simulation studies: results

| *Analysis model* | *Coverage* $\beta_1$ | *Coverage* $\beta_2$ | *Coverage* $\beta_3$ | *Coverage* $\beta_4$ | *Coverage* $\rho$ | $L_{True}$ *superior* | $M_{True}$ *superior* |
|---|---|---|---|---|---|---|---|
| *True DGP: dynamic fixed-effects specification—I* | | | | | | | |
| FE dyn. | 0.94 | 0.96 | 0.96 | 0.95 | 0.91 | — | — |
| RE dyn. | 0.76 | 0.96 | 0.95 | 0.96 | 0.97 | 1 | 0.94 |
| FE | 0.13 | 0.89 | 0.86 | 0.85 | — | 1 | 1 |
| RE | 0.03 | 0.89 | 0.88 | 0.86 | — | 1 | 1 |
| *True DGP: dynamic random-effects specification—II* | | | | | | | |
| FE dyn. | 0.91 | 0.96 | 0.97 | 0.98 | 0.92 | 0 | 1 |
| RE dyn. | 1.00 | 0.95 | 0.97 | 0.99 | 0.93 | — | — |
| FE | 0.06 | 0.91 | 0.85 | 0.84 | — | 1 | 1 |
| RE | 0.00 | 0.92 | 0.89 | 0.92 | — | 1 | 1 |
| *True DGP: fixed-effects specification—III* | | | | | | | |
| FE dyn. | 0.93 | 0.94 | 0.94 | 0.97 | — | 0 | 1 |
| RE dyn. | 0.74 | 0.93 | 0.94 | 0.97 | — | 1 | 0 |
| FE | 0.94 | 0.94 | 0.94 | 0.96 | — | — | — |
| RE | 0.83 | 0.94 | 0.94 | 0.97 | — | 1 | 0 |
| *True DGP: random-effects specification—IV* | | | | | | | |
| FE dyn. | 0.93 | 0.95 | 0.98 | 0.97 | — | 0 | 1 |
| RE dyn. | 1.00 | 0.96 | 0.98 | 0.98 | — | 0 | 1 |
| FE | 0.95 | 0.96 | 0.97 | 0.96 | — | 0 | 1 |
| RE | 1.00 | 0.96 | 0.98 | 0.98 | — | — | — |

Abbreviations: DGP, data-generating process; FE, fixed effects; FE dyn., dynamic fixed effects; RE, random effects; RE dyn., dynamic random effects.

**Table 3.** Empirical analysis: frequencies of observed and missing values in available height and weight data sets

| *Study* | *A, height* | | *A, weight* | | *B* | | *C* | | *D* | | *E* | | *F* | | *G* | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Boys* | *Girls* | *Boys* | *Girls* | *Boys* | *Girls* | *Boys* | *Girls* | *Boys* | *Girls* | *Boys* | *Girls* | *Boys* | *Girls* | *Boys* | *Girls* |
| *Height and weight* | | | | | | | | | | | | | | | | |
| $T$: | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 |
| $N_{\text{obs}}$ | 4819 | 4983 | 4884 | 5057 | 624 | 864 | 736 | 960 | 894 | 1120 | 1126 | 1262 | 1424 | 1424 | 1901 | 1769 |
| $N_{\text{mis}}$ | 3357 | 3225 | 3372 | 3231 | 0 | 0 | 0 | 0 | 2 | 0 | 10 | 18 | 0 | 0 | 19 | 23 |
| *Height and weight increments* | | | | | | | | | | | | | | | | |
| $T$: | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 |
| $N_{\text{obs}}$ | 2640 | 2541 | 2675 | 2606 | 624 | 864 | 736 | 960 | 894 | 1120 | 1126 | 1262 | 1424 | 1424 | 1901 | 1769 |
| $N_{\text{mis}}$ | 1200 | 1123 | 1197 | 1138 | 0 | 0 | 0 | 0 | 2 | 0 | 10 | 18 | 0 | 0 | 19 | 23 |

model evaluations. Nevertheless, one should keep in mind the disadvantages of the high number of parameters within the fixed-effect specifications for the empirical illustrations, which are presented in the next section.

### Empirical analysis

With regard to the empirical analysis, we use 14 available longitudinal data sets from Lithuania, Berkeley, France, Lubin, Prague, Propotec and Zurich (7 studies of boys, 7 of girls), each containing at the maximum 16 observed periods per child (for a better clarity, we use Latin letters A to G to indicate the studies in just this rank order) (For example, the Lithuanian longitudinal data were derived from the personal health records of children born in 1990 in the city and surrounding villages of Vilnius in Lithuania, including height and weight records from birth to 18 years of age (recorded every month during the first year of life, later—annually)). For comparison with the considered other studies, we use only observation up to 17 years of age. Only full-term born and healthy children were selected for the data analysis. Children with disabilities and different chronic diseases at any age were not included in the further analysis. Data were collected during 2009 and 2010 in the four largest outpatient clinics in the Vilnius city and region. The total number of personal health records analysed was 1484, of 756 boys and 728 girls. Permission to conduct the study was granted by the Lithuanian Bioethics Committee. Gaps between the measurements of body size indices were observed in almost every personal health record. The frequency of the measurements at various age groups differed widely, with three times as many measurements at some ages than others. As the children's measurements were taken during monitoring visits to the outpatient clinics, the frequency of those visits was not associated with morbidity. The children's heights are documented starting with the age of 2 years and ending with the age of 17 years. Further, as the dynamic specifications nest the non-dynamic specification and comparison between these specifications is possible via inspection of the highest density interval of the corresponding parameters, we focus on empirical analysis of the dynamic model specifications only.

We apply the suggested modelling approaches directly to height and weight and also on height and weight increments. Table 3 presents the corresponding frequencies of observed and missing values in each data set (Note that, in the non-dynamic context, there is one less period utilised, and hence no further imputation for the very first explaining period is necessary). Note that because of the dynamic model structure the model is not directly suited for the first period, as there is no appropriate covariate (the former period) available. Hence, for the first period, an unconditional modelling approach is implemented. Furthermore, we only considered individuals in our analyses who were observed at least once in each age ranging from 1 to 5 years, from 6 to 11 years and from 12 to 17 years. This seems reasonable owing to the fact that the data augmentation step for a specific child relies on the child's own observations at its available periods. Hence, it takes enough information from a couple of periods of an individual to impute its corresponding missing values.

Table 1 presents the uninformative prior distributions applied to the different introduced models assuming prior independence of the parameters, including the PB formulation as the benchmark model. Table 4 presents block-wise the empirical results with regard to height and weight of the dynamic fixed, random and mixed model setups, as well as of the PB formulation for the boys and girls studies, respectively. Each block includes the log marginal likelihoods and log-likelihoods, as well as the number of estimated parameters in the specific model context. Summarising, comparisons of the models show that the PB model provides best fits in terms of the log marginal likelihood in 9 out of the 14 data sets. Vice versa, in the five remaining studies, the dynamic random-effects specification offers the best results. On closer consideration, the results show at least two remarkable features. First, as presumed above, the dynamic fixed-effects formulation provides best fits in the meaning of the log-likelihood in almost every survey but the two Lithuanian data sets. However, this result holds merely if the prior and posterior evaluations are not taken into account. Hence, as the log marginal likelihood also evaluates the relatively high number of estimates of the dynamic fixed-effects formulation, the specification loses ground against the more parsimonious dynamic random-effects model. Thus, and taking the results of the simulation study into account, it seems that the random effects are able to fit both the individual- and time-specific intercepts by their normal distribution assumption. This result still holds if only the time-specific intercepts are estimated as fixed effects. Even then, the dynamic random-effects model provides the better fit in terms of the log marginal likelihood. Similar result findings are present with regard to the model for height and weight increments (see Table 5). Here, in the absence of the PB model, the dynamic random-effects models are identified as the best model according to the marginal likelihood criterion.

## DISCUSSION

We consider a range of statistical models incorporating latent heterogeneity at different degrees in order to model human growth curves. We chose a Bayesian framework in the context of fixed, random and mixed model setups. After their formal derivation and algorithm description, we assessed the algorithm accuracies by a simulation study. The PB formulation as the benchmark model provided the best fit not in terms of the likelihood but the marginal likelihood in the majority of the

**Table 4.** Empirical analysis: model comparison: log marginal likelihoods for height and weight studies of boys and girls

| *Study* | *A, height* | *A, weight* | *B* | *C* | *D* | *E* | *F* | *G* |
|---|---|---|---|---|---|---|---|---|
| *Boys* | | | | | | | | |
| FE dyn. | | | | | | | | |
| $\log M(y\|X)$ | − 15 190.82 | − 16 339.41 | − 1374.99 | − 1606.54 | − 2061.76 | − 2568.78 | − 3216.90 | − 3946.88 |
| $\log L(y\|\theta,X)$ | − 11 354.07 | − 12 157.06 | − 837.34 | − 1009.38 | − 1390.17 | − 1762.02 | − 2243.86 | − 2666.30 |
| Parameters | 541 | 546 | 69 | 76 | 86 | 101 | 119 | 150 |
| RE dyn. | | | | | | | | |
| $\log M(y\|X)$ | − 14 307.20 | − 17 053.98 | **− 1064.13** | **− 1232.70** | − 1615.76 | − 2000.45 | − 2500.71 | **− 3221.77** |
| $\log L(y\|\theta,X)$ | − 14 230.57 | − 16 981.42 | − 1019.52 | − 1186.65 | − 1568.62 | − 1950.26 | − 2448.82 | − 3168.99 |
| Parameters | 19 | 19 | 19 | 19 | 19 | 19 | 19 | 19 |
| Mixed model dyn. | | | | | | | | |
| $\log M(y\|X)$ | − 13 515.33 | − 15 993.00 | − 1165.19 | − 1420.38 | − 1769.12 | − 2116.04 | − 2586.78 | − 3239.59 |
| $\log L(y\|\theta,X)$ | − 13 309.45 | − 15 786.74 | − 986.69 | − 1243.74 | − 1589.74 | − 1927.77 | − 2387.84 | − 3034.77 |
| Parameters | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 |
| PB | | | | | | | | |
| $\log M(y\|X)$ | **− 8324.13** | **− 13 652.16** | − 1136.01 | − 1314.77 | **− 1541.02** | **− 1985.50** | **− 2425.11** | − 3659.28 |
| $\log L(y\|\theta,X)$ | − 8292.75 | − 13 650.19 | − 1109.26 | − 1289.83 | − 1515.56 | − 1958.38 | − 2400.08 | − 3641.44 |
| Parameters | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 |
| *Girls* | | | | | | | | |
| FE dyn. | | | | | | | | |
| $\log M(y\|X)$ | − 15 217.44 | − 16 225.17 | − 1860.35 | − 2035.47 | − 2391.77 | − 2892.77 | − 3161.81 | − 3665.83 |
| $\log L(y\|\theta,X)$ | − 11 246.11 | − 12 015.21 | − 1168.55 | − 1303.13 | − 1539.30 | − 1978.34 | − 2148.83 | − 2423.27 |
| Parameters | 543 | 548 | 84 | 90 | 100 | 110 | 119 | 142 |
| RE dyn. | | | | | | | | |
| $\log M(y\|X)$ | − 13 869.76 | − 15 627.70 | − 1360.14 | **− 1516.54** | **− 1764.42** | − 2233.35 | − 2386.41 | − 2805.17 |
| $\log L(y\|\theta,X)$ | − 13 795.13 | − 15 557.07 | − 1310.81 | − 1469.83 | − 1710.72 | − 2181.80 | − 2334.32 | − 2746.95 |
| Parameters | 19 | 19 | 19 | 19 | 19 | 19 | 19 | 19 |
| Mixed model dyn. | | | | | | | | |
| $\log M(y\|X)$ | − 13 587.73 | − 15 204.57 | − 1410.77 | − 1568.76 | − 1803.80 | − 2296.30 | − 2445.83 | − 2853.02 |
| $\log L(y\|\theta,X)$ | − 13 382.60 | − 14 999.11 | − 1211.52 | − 1373.51 | − 1597.28 | − 2095.48 | − 2241.39 | − 2639.89 |
| Parameters | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 |
| PB | | | | | | | | |
| $\log M(y\|X)$ | **− 8552.33** | **− 13 497.49** | **− 1228.29** | − 1572.06 | − 1874.88 | **− 2000.77** | **− 2179.80** | **− 2712.54** |
| $\log L(y\|\theta,X)$ | − 8523.37 | − 13 496.70 | − 1208.52 | − 1550.41 | − 1854.57 | − 1978.38 | − 2154.79 | − 2688.89 |
| Parameters | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 |

Abbreviations: FE dyn., dynamic fixed effects; Mixed model dyn., dynamic mixed model; PB, Preece–Baines model; RE dyn., dynamic random effects. Values in bold highlight that the corresponding model is the study-specific superior one in terms of the log marginal likelihood.

**Table 5.** Empirical analysis: model comparison for height and weight increments: log marginal likelihoods for studies of boys and girls

| *Study* | *A, height* | *A, weight* | *B* | *C* | *D* | *E* | *F* | *G* |
|---|---|---|---|---|---|---|---|---|
| *Boys* | | | | | | | | |
| FE dyn. | | | | | | | | |
| $\log M(y\|X)$ | − 5707.38 | − 5623.96 | − 1315.75 | − 1562.71 | − 2013.05 | − 2471.06 | − 3114.97 | − 3745.36 |
| $\log L(y\|\theta,X)$ | − 3953.99 | − 3765.48 | − 787.77 | − 987.26 | − 1369.08 | − 1684.90 | − 2173.33 | − 2498.67 |
| Parameters | 268 | 270 | 67 | 74 | 84 | 99 | 117 | 148 |
| RE dyn. | | | | | | | | |
| $\log M(y\|X)$ | **− 4113.87** | **− 3929.44** | **− 890.28** | **− 1097.69** | **− 1483.23** | **− 1814.32** | **− 2323.14** | **− 2699.55** |
| $\log L(y\|\theta,X)$ | − 4047.10 | − 3869.35 | − 845.55 | − 1051.75 | − 1435.92 | − 1765.04 | − 2271.06 | − 2645.28 |
| Parameters | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 |
| Mixed model dyn. | | | | | | | | |
| $\log M(y\|X)$ | − 4201.90 | − 4020.13 | − 987.99 | − 1195.71 | − 1581.89 | − 1911.88 | − 2421.15 | − 2799.22 |
| $\log L(y\|\theta,X)$ | − 4019.93 | − 3841.62 | − 810.04 | − 1017.70 | − 1402.74 | − 1727.46 | − 2232.32 | − 2604.46 |
| Parameters | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| *Girls* | | | | | | | | |
| FE dyn. | | | | | | | | |
| $\log M(y\|X)$ | − 5486.52 | − 5617.04 | − 1722.63 | − 1944.01 | − 2295.96 | − 2698.53 | − 3031.80 | − 3382.89 |
| $\log L(y\|\theta,X)$ | − 3681.65 | − 3828.33 | − 1021.06 | − 1220.46 | − 1455.05 | − 1799.15 | − 2034.39 | − 2121.04 |
| Parameters | 257 | 262 | 82 | 88 | 98 | 108 | 117 | 140 |
| RE dyn. | | | | | | | | |
| $\log M(y\|X)$ | **− 3902.90** | **− 4002.07** | **− 1138.74** | **− 1340.70** | **− 1583.31** | **− 1932.42** | **− 2176.25** | **− 2288.08** |
| $\log L(y\|\theta,X)$ | − 3839.67 | − 3940.28 | − 1083.18 | − 1291.13 | − 1528.36 | − 1879.64 | − 2121.74 | − 2226.03 |
| Parameters | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 |
| Mixed model dyn. | | | | | | | | |
| $\log M(y\|X)$ | − 3986.78 | − 4094.45 | − 1232.41 | − 1434.79 | − 1678.77 | − 2025.98 | − 2272.02 | − 2383.64 |
| $\log L(y\|\theta,X)$ | − 3807.16 | − 3914.05 | − 1036.25 | − 1247.27 | − 1482.52 | − 1835.60 | − 2076.26 | −2176.68 |
| Parameters | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |

Abbreviations: FE dyn., dynamic fixed effects; Mixed model dyn., dynamic mixed model; RE dyn., dynamic random effects. Values in bold highlight that the corresponding model is the study-specific superior one in terms the log marginal likelihood.

empirical data. Taken together with the two major results that the dynamic fixed-effects model best fits the data in terms of the likelihood, whereas it simultaneously suffers for its high number of parameters, and that the dynamic component provides an impressive model accuracy, a next analysis step could be to formulate dynamic models that are not penalised that much in terms of their prior and posterior parameter evaluations. Thus empirical Bayes methods could provide a solution by selecting more informative prior distributions, for example, by deriving the prior moments from a previous likelihood estimation. This approach might take both the advantages of a latent dynamic model and not too lavish prior information into account.

Further research could also focus on the actual prediction of children's growth. For example, as the data augmentation steps within the Gibbs sampling imputed missing values in each iteration by utilising information of all the remaining observations in the data sets, this technique might serve as an accurate prediction procedure as well. Simulation studies could provide coverages of erased and then re-predicted observations, whereas forecast errors allow additional evaluations of the actual prediction. These approaches could be able to increase the analysis accuracies not only on the aggregate but even on the individual level.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## REFERENCES

1 WHO. *WHO Child Growth Standards: Length/Height-For-Age, Weight-For-Age, Weight-For-Length, Weight-For-Height and Body Mass Index-For-Age: Methods and Development*. World Health Organization: Geneva, Switzerland, 2006.
2 Preece MA, Baines MJ. A new family of mathematical models describing the human growth curve. *Ann Hum Biol* 1978; **5**: 1–24.
3 Jenss RM, Bayley N. A mathematical method for studying growth in children. *Hum Biol* 1937; **9**: 556–563.
4 Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med* 1992; **11**: 1305–1319.
5 Cole TJ, Donaldson, Malcolm DC, Ben-Shlomo Y. SITAR--a useful instrument for growth curve analysis. *Int J Epidemiol* 2010; **39**: 1558–1566.
6 Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc* 1995; **90**: 773–795.
7 Aßmann C, Hermanussen M. Modeling determinants of growth: evidence for a community-based target in height? *Pediatr Res* 2013; **74**: 88–95.
8 Tanner MA, Wong WH. The calculation of posterior distributions by data augmentation. *J Am Stat Assoc* 1987; **82**: 528–540.
9 Casella G, George EI. Explaining the Gibbs sampler. *Am Stat* 1992; **46**: 167–174.
10 Chib S. Marginal likelihood from the Gibbs output. *J Am Stat Assoc* 1995; **90**: 1313–1321.
11 Chib S, Jeliazkov I. Marginal likelihood from the Metropolis–Hastings output. *J Am Stat Assoc* 2001; **96**: 270–281.