



Research Article

Volume 11 Issue 1 - November 2022
DOI: 10.19080/BBOAJ.2022.11.555803

Biostat Biom Open Access J
Copyright © All rights are by Xueya Cai

Dynamic Panel Estimation in Prediction of Child Growth



Xueya Cai^{1*}, Shan Gao¹, Sarah Obudzinski² and James Sanders²

¹Department of Biostatistics and Computational Biology, University of Rochester, USA

²Department of Orthopaedics, University of North Carolina Chapel Hill, USA

Submission: November 4, 2022; **Published:** November 30, 2022

***Corresponding author:** Xueya Cai, Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY, USA

Abstract

Understanding of child growth patterns typically requires longitudinal follow-up of children with serial anthropometrics. Current statistical models used for growth prediction tend to ignore individual level dynamic changes, require intensive computational efforts, or be so complex that are hardly interpretable in a clinical meaningful way. Vector autoregression is an econometric method that has been improved to accommodate longitudinal data with a small number of repeated measures over time, and dynamic panel estimation (DPE) method further improved vector autoregression to incorporate individual variations with more flexible model assumption. Thus, the DPE method is well suited to predict child growth in which a relatively small number of repeated measures are collected over years without a definitive beginning measuring point. In this study we extended the use of the econometric dynamic panel estimation to child growth prediction, in which one step generalized method of moments was used for parameter estimations. Model goodness of fit was evaluated by residuals and cross validation strategy, and external validation was also performed using different child growth data. We found that the DPE method fits well in development and validation samples in child growth prediction, and the model was easy to interpret.

Keywords: Dynamic panel estimation; Growth models; Panel data; Child growth

Introduction

Predicting future growth is important to many areas in children's care, but its accuracy is particularly important in musculoskeletal care where growth asymmetry necessitates precise estimates to balance growth based upon differential rates. Prior models of growth have not achieved the degree of accuracy needed for precise growth modulation as needed by newer techniques such as spinal tethering for spinal deformities. Human postnatal growth follows very distinct phases [1-13] starting with rapid but decelerating infant growth. This is followed by the second phase of slower and somewhat linear childhood growth beginning at age 3 or 4 years and extending until the adolescent growth spurt which is a period of rapid growth acceleration followed by deceleration terminating in growth cessation. Most growth models assume children remain a constant percentile in height compared to their same sex peers through growth, but this is incorrect. When children enter the childhood phase, their trajectory is relatively constant, and percentile crossing is uncommon until adolescence. This consistent trajectory is termed "canalization". With illness or environmental challenges such as starvation, growth is stunted, but recovery, if sufficient time allows, results in "catch up" growth. While the pattern of growth during third phase or the adolescent

growth spurt is similar between children, the timing of the growth spurt is non-uniform. Boys undergo their growth spurt about two years later than girls, and the change in percentiles as a child ages can be very impressive as some children are growing rapidly at one time while their peers are not. This variation in timing limits the utility of cross-sectional studies to predict growth. Because the childhood phase is not as tied to skeletal maturity as the adolescent phase, knowing how long the childhood phase will persist is not known when a child is still prepubescent. However, as the growth spurt approaches, skeletal maturity becomes tightly related to growth. An ideal model to advance childhood growth predictability would join childhood growth prediction with maturity prediction to estimate adolescent growth spurt initiation when future growth becomes predictable.

Different statistical methods have been developed and applied for child growth predictions.

The Centers for Disease Control and Prevention (CDC) reported their methodology of growth charts using the locally weighted regression (LWR) procedure followed by Box transformation, in which parameter estimates were obtained from the modified

Least Mean Square (LMS) method [14,15]. The World Health Organization (WHO) applied the same methodology in their development of child growth standards [16]. Hierarchical mixed models have also been applied to provide average growth curve estimates at the population level and to control for within-subject variations [17,18]. Recently, functional data analysis has been developed and used in child growth data to relax assumptions in the mixed models and to better predict growth trajectory at the individual level. In functional data analysis, functional derivatives are constructed and estimated using nonparametric method in order to estimate outcome trajectories against changes in predictors [19]. Functional data analysis is computationally intensive, and the resulting prediction model can hardly be interpreted in a clinically meaningful way.

Autoregressive time-series data are commonly seen in econometric studies, in which a sequence of the same outcome(s) is measured over time. Vector autoregression (VAR), one of the most commonly used econometric methods in autoregressive time-series data analysis, assumes that the outcome measure is a function of the previous value(s) of the same measure at the aggregate level [20-22]. When VAR is used to analyze longitudinal (or panel) data with a few repeated measures over time such as the growth prediction, however, the homogeneity assumption is likely violated. Holtz-Eakin, et. al., developed the dynamic panel estimation (DPE) method by adding individual variations into VAR to accommodate heterogeneity [23-26]. The DPE model is shown to provide unbiased estimates at the individual level, and its interpretations are relatively easy to non-statisticians.

In this study we implemented the DPE method in child growth prediction using the Bolton-Brush study data and the Berkley longitudinal growth study data, and generalized method of moment (GMM) method was used to obtain consistent parameter estimates [27,28]. After summarizing the DPE model and its parameter estimation, we fit the grow prediction model on the Berkley study data, evaluate goodness of fit of the model using studentized residuals and cross-validation strategy, and extrapolate model prediction to the Bolton-Brush data for external validation. Finally, we conclude this study with discussions of the strengths and limitations of the DPE model when used for child growth predictions.

Methods

Dynamic Panel Estimation (DPE)

Vector autoregressions (VA) are widely applied in time series data analysis especially in econometric studies on dynamic data with repeated measures from same subjects over time 23,24. However, its application to panel data, data with fewer number of repeated measures than the number of subjects, violates the assumption required for vector autoregression of stationary individual effects. To address this issue, Holtz-Eakin, et al., improved the VA method with the approach introduced

by Chamberlain [29] to allow for individual effects and non-stationarities across time in panel data prediction. In addition, the DPE model does not require the projection to be based on all past measures [23]. The specification of the DPE model is:

$$y_{it} = \beta_0 + \sum_{m=1}^M \beta_m y_{it-1} + \sum_{m=1}^M \gamma_m x_{it-1} + \varepsilon_{it}, i = 1, 2, \dots, N, t = 1, 2, \dots, T$$

In this model, y_{it} denotes the outcome measure for subject i at time t , m is the lag length, y_{it-1} denotes the m

past outcome measures, x_{it-1} denotes the m past independent variables, ε_{it} is the error term, and β_0, β_m , and γ_m are corresponding coefficients. Consistent parameter estimates can be obtained through the two-stage least square (2SLS) strategy [23,30,31]. Holtz-Eakin et al., also indicated that the number of repeated measures (T) and the number of past outcome measures included in the model (m) should follow this rule in order for parameters to be estimable: $T \geq 3M + 2$.

Human growth studies typically recruit participants with different characteristics and collect several repeated clinical measures during follow-ups. According to the model specification, DPE is reasonable choice for such longitudinal clinical data analyses and can accommodate well the relatively small number of repeated measures in growth predictions as well as participant individual growth patterns.

Data sources

Two important longitudinal data sets of childhood growth include the Brush Foundation. Study, and the Berkeley Guidance Study. The Brush Foundation Study of Child Growth and Development is the largest and most complete longitudinally collected collection of combined anthropometrics and skeletal radiographs. The prospective study started in 1931 enrolling healthy children from 3 months to 14 years of age during each successive year until 1942. Subjects had measurements and radiographs every 3 months until 1 year of age, every 6 months until age 5 and annually thereafter. The study contains records of 4483 children with follow-up ranging from two to twelve years [1]. This study is most well-known for the Greulich and Pyle [2] hand and wrist skeletal maturity atlas. The Berkeley Guidance Studies of the Institute of Human Development [3], best known for Bayley's work on mental development [4,32] enrolled every third child born in Berkeley in 1928-1929 and followed them to growth completion with serial anthropometrics. Both the Brush and the Berkley study data included participants' sex, and age and standing height at each follow-up visit in longitudinal format.

In this study, we took advantage of the child growth stage data available from the Berkley study to fit the longitudinal prediction model using the DPE method. We fit the prediction model of child height with age for boys and girls separately given their substantial differences in start ages of growth spurt. The models were further externally validated using the Brush data.

Statistical Analysis

Several strategies were implemented in this study to identify the best cutoffs for child growth stages, to fit the childhood growth model, and to validate model prediction using external data. First, threshold analysis by child sex was performed to divide child growth into infantile growth phase, childhood growth phase, and adolescent growth spurt phase using the Berkley data. Informed by clinical knowledge we assumed the lower cut-off point for the three phases to be between 2 and 5 years of age, and the higher cut-off point to be between 7 and 12 years. Regression model of child height was then fit against age, cutoff points, and the interactions between age and cutoff points. The interactions between the second and third order age polynomials and both cutoff points were also included in the model to accommodate non-linear growth pattern within each of the three major growth phases. Alternative cutoff points for each child were considered and the ones with the smallest Aikake Information Criteria (AIC) were selected. After identification of the best cutoff points, we plotted the age and height association using 50% random samples from the original data based on the Bernoulli sampling approach.

We then fit dynamic panel estimation (DPE) models for boys and girls separately. In each sex group, we fit DPE models with different number of past observations up to the maximum number of past observations which were estimable according to Holtz-Eakin et al. criterion. Ten-fold cross validation was applied, and the DPE model with lowest root mean square error (RMSE) was chosen to be the best fitted model [33,34]. Studentized residuals obtained from the final DPE model were plotted against child age to demonstrate the goodness of fit of the model across age, in which the absolute values of studentized residuals were expected to be less than 3 under good model fitting [35].

In addition to the ten-fold cross validation, we performed external validation of the growth prediction models developed on Berkeley Guidance Study data. In the external validation, the prediction models were applied to the Bolton child phase data for boys and girls separately, and residuals and studentized residuals obtained after predictions were plotted against child age to examine goodness of fit of the models. In addition, we plotted the histogram of residuals by child age to examine potential outliers by age.

Results

Berkley boys

The Berkley data collected 2416 longitudinal age and height measures for 66 boys. Threshold analyses indicated that the lower and upper cutoffs of 3 and 12 years provided the smallest AIC value (Figure 1). panel A illustrates the growth patterns of randomly selected 33 boys from the data, and it suggests that the linear assumption of childhood growth pattern for boys at this age range was reasonable. Our prediction model of Berkley boys was then focused on those aged 3-12.

According to Holtz-Eakin's suggestion, we fit first order and second order DPE models and compared their performance using the ten-fold cross-validations. Our analyses showed that the first order DPE model provided much smaller RMSE (1.35) comparing to the second order model (RMSE=134.92).

Therefore, the first order DPE model was chosen as the best prediction model. This prediction model was estimated as

$$E(\widehat{height}) = 14.28 + 0.83 \times height(1) - 5.54 \times age(1) + 6.47 \times age$$

in which $height(1)$ and $age(1)$ are age and height of the boy, respectively, that were previously measured, and age is the boy's current age.

(Figure 1) Panel B and Panel C show the residual plot and the studentized residual plot, respectively, obtained from the above DPE model. From both plots we identified two potential outliers, one at age 4 and the other at age 12. The residual plot showed a 4.0cm difference between the observed (111.0cm) and predicted (107.0cm) height for the subject at age 4, and a 4.1cm difference between the observed (163.6cm) and predicted (159.5cm) height for another subject at age 12 (Figure 1).

Berkley girls

The Berkley data collected 2,472 longitudinal age and height measures for 70 girls. Threshold analyses indicated that the lower and upper cutoffs of 3 and 9 years provided the smallest AIC value (Figure 2). panel A illustrates the growth patterns of randomly selected 35 girls from the data, and similarly suggests that the linear assumption of childhood growth pattern for girls at this age range was reasonable. Our prediction model of Berkley girls was then focused on those aged 3-9 years.

According to Holtz-Eakin's criterion only the first order DPE model was appropriated to fit on the Berkley girls data. Our analyses further showed that the RMSE of the ten-fold cross-validated, first order DPE model was 3.89. This prediction model was estimated as

$$E(\widehat{height}) = 19.99 + 0.80 \times height(1) - 3.91 \times age(1) + 4.69 \times age$$

(Figure 2) Panel B and Panel C show the residual plot and the studentized residual plot, respectively, obtained from the above DPE model. From both plots we identified three potential outliers. The residual plot showed a 4.3cm difference between the observed (152.5cm) and predicted (148.2cm) height for a subject at age 9, a 6.3cm difference between the observed (120.5cm) and predicted (114.2cm) height for the second subject at age 6, and a 5.2cm difference between the observed (95.3cm) and predicted (100.5cm) height for the third subject at age 4. (Figure 2).

External Validation: After the DPE models were fit on the Berkley boys and girls data, we applied the exact models obtained from the Berkley data to predict boy and girl growth patterns at their child phase using the Bolton data for external validation.

The prediction models using Bolton data included 215 height measures for 24 boys and 256 height measures for 31 girls. (Figure 3) Panel (a) and Panel (b) show the residual plot and studentized residual plot, respectively, for the growth prediction model of boys aged 3 to 12 years with ± 3 months variation. It can be seen that most predicted heights based on the Berkley DPE model were close to the observed heights, except for two subject measures (outliers) at ages of 11 and 12, respectively. (Figure 3) Panel (c)

shows the residual histograms by subject approximate age with ± 3 months variation (e.g., the first set of histograms are for Bolton boys with ages of 3.5 years ± 3 months). The results suggest that younger boys had relatively unbiased height predictions with DPE prediction model, and that the predictions were slightly skewed to the left (slight over-predictions) when boys were older than 7 years old (Figure 3).

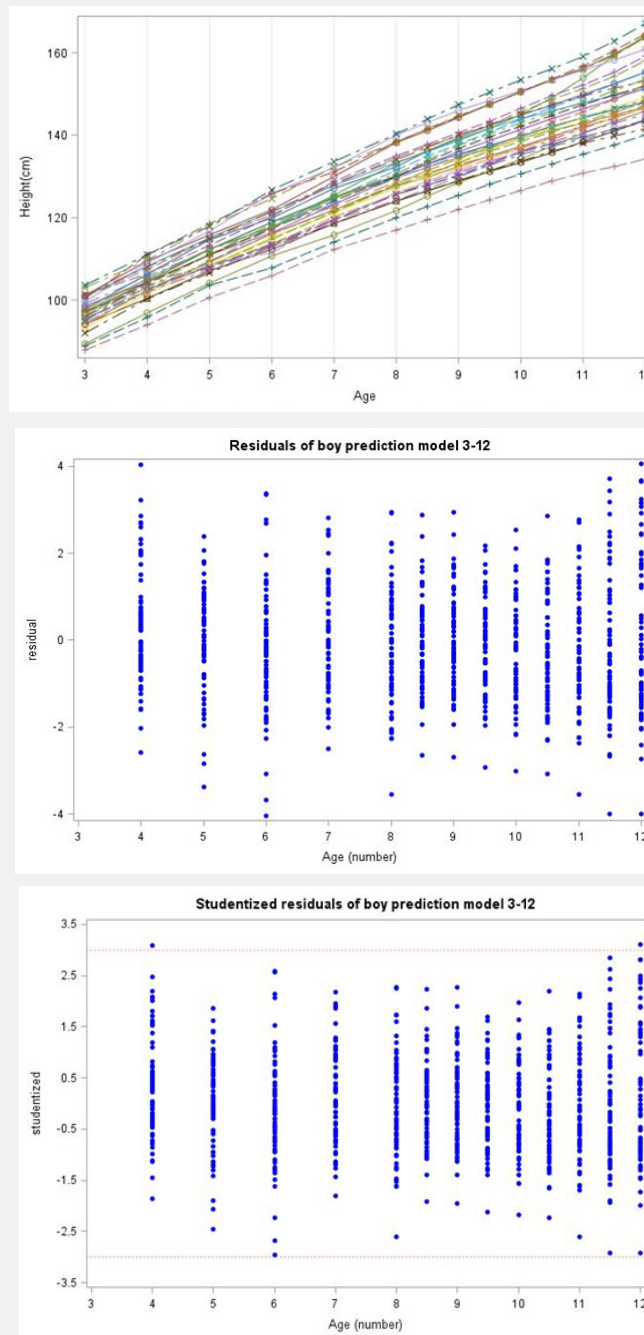


Figure 1: Height trajectory and prediction for Berkley boys aged 3-12 years. Panel (a) – height trajectory; Panel (b) – residual plot; and Panel (c) – studentized residual plot of the Dynamic Panel Estimation (DPE) model.

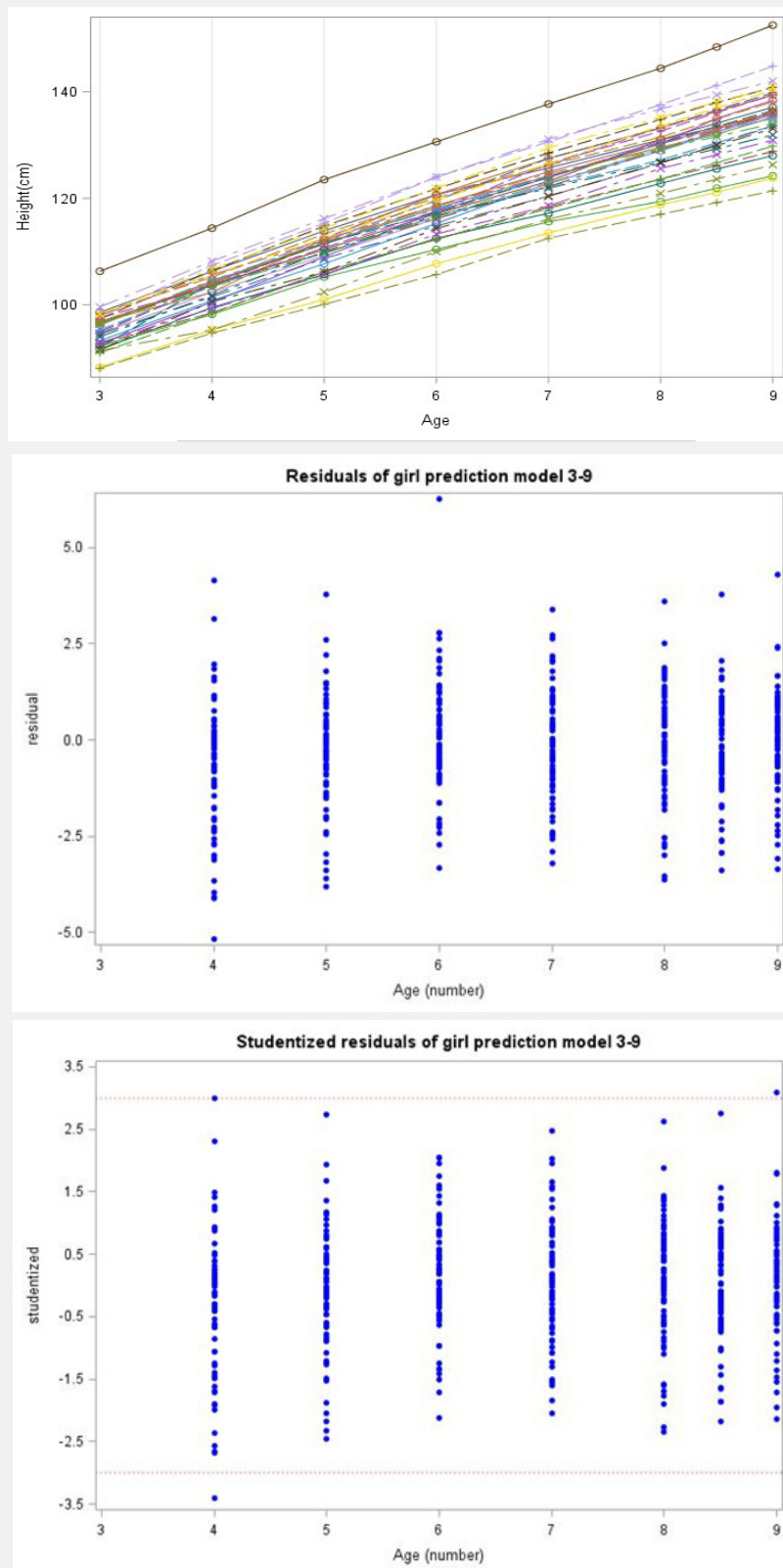


Figure 2: Height trajectory and prediction for Berkley girls aged 3-9 years. Panel (a) – height trajectory; Panel (b) – residual plot; and Panel (c) – studentized residual plot of the Dynamic Panel Estimation (DPE) model.

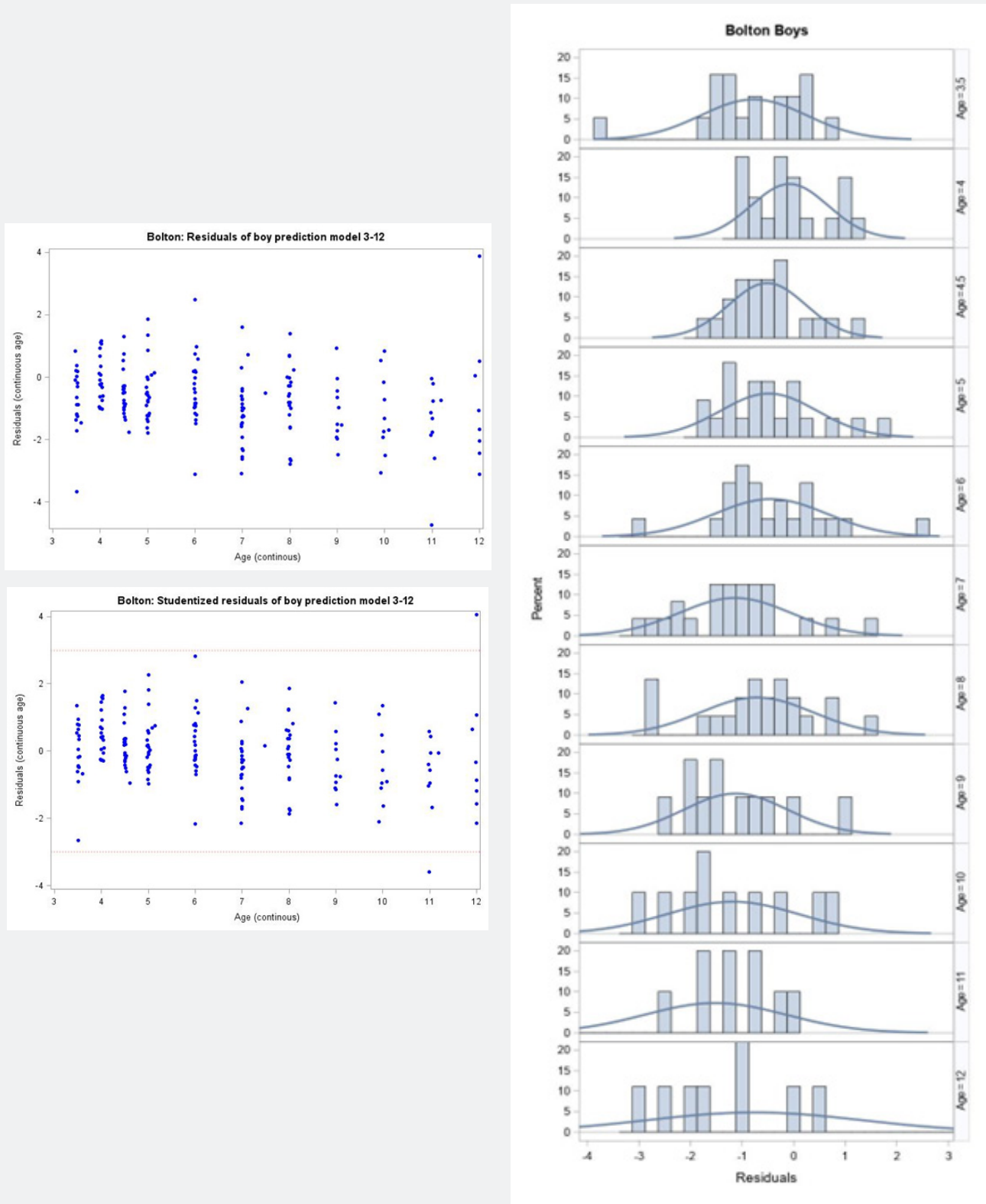


Figure 3: External validation using the Bolton boys data. (a) Residual plot; (b) Studentized residual plot; and (c) Histogram of residuals by age.

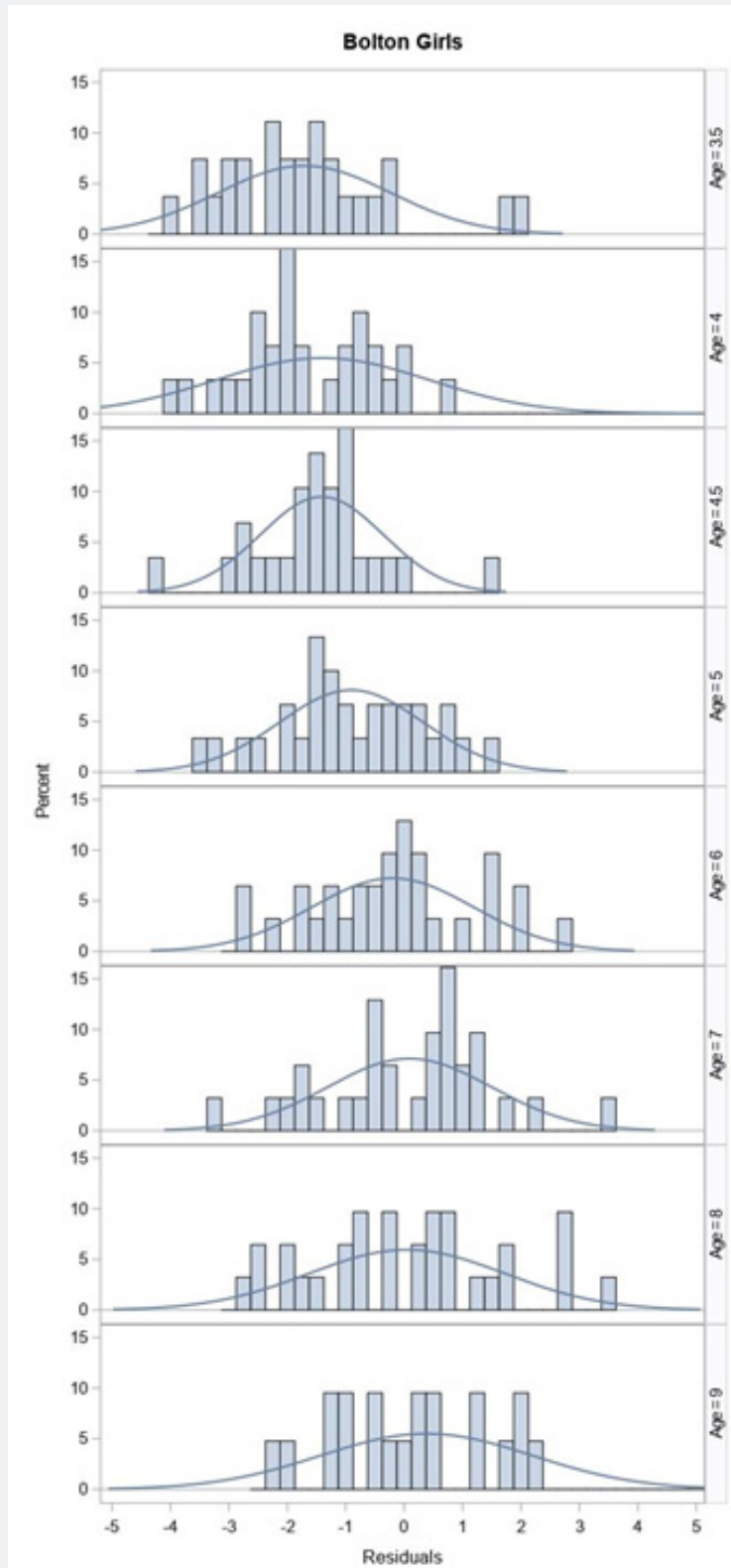
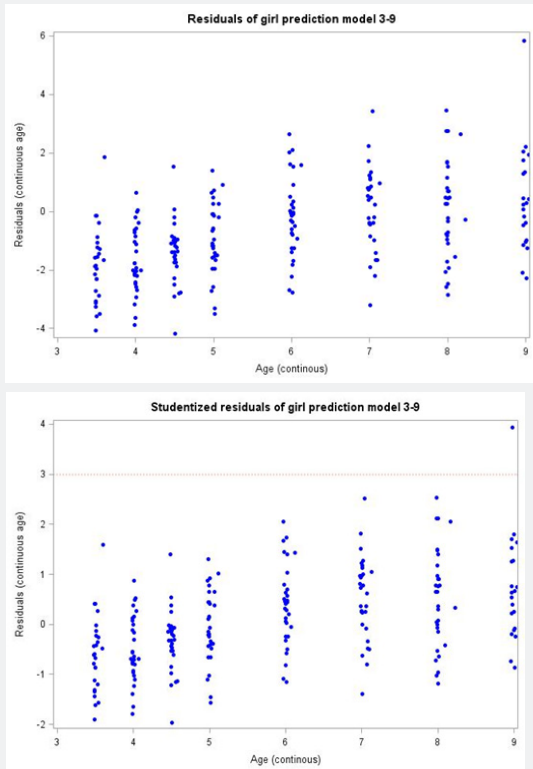


Figure 4: External validation using the Bolton girls data. (a) Residual plot; (b) Studentized residual plot; and (c) Histogram of residuals by age.

(Figure 4) Panel (a) and Panel (b) show the residual plot and studentized residual plot respectively, for girls growth prediction from age 3 to 9 with ± 3 months variation. The residual plots show only one outlier at age 9, and the studentized residual plot shows a slightly increasing trend over age. This slight increasing pattern can also be found in (Figure 4) Panel (c), which suggests that girl heights in the Bolton sample tended to be slightly over-predicted at younger ages (between age 3 and 5), but predictions remained overall unbiased at older ages (between age 6 and 9) (Figure 4).

Discussion

This study introduced the dynamic panel estimation (DPE) method for prediction of childhood growth, and analyses in both the development and the external validation samples suggest overall excellent predictive abilities for growth patterns among boys and girls separately. The DPE method has been widely used in econometric analyses for the prediction of various micro- and macro-economic behaviors. The DPE method allows for dynamic individual growth patterns in model prediction and theoretically is a promising method for predicting childhood growth. Our results of goodness of fit of the DPE models in child growth predictions and small root means square errors in alternative samples provide empirical evidence supporting this expectation.

As shown in the model specifications as well as model fitting in separate samples of boys and girls, the DPE method is able to effectively take into account children's variations in dynamic growth patterns at different ages, which is similar to functional data analyses method that has been applied in Berkley child growth prediction [19]. While both functional data analysis and the DPE method rely on differential equations, the specification of the DPE model is much more straightforward than that of functional data analysis and coefficient estimates of the DPE model can be interpreted in a clinically meaningful way. Another advantage of the DPE method is that it is computationally much less intensive comparing to the functional data analysis method without loss of accuracy.

The two data sets used in this study are some of the most important longitudinal studies of childhood growth in existence. The Berkeley study has complete longitudinal data for height and sitting height though growth but does not have the granularity of the Brush study which has very detailed anthropometrics but with children only followed for various periods through their growth. While the model caused some overestimation in younger children in the Brush collection, it was very accurate during the more important later childhood phase where accuracy is most crucial.

References

- 1 Bayley N, Schaefer ES (1964) Correlations of Maternal and Child Behaviors with the Development of Mental Abilities. Data from the Berkeley Growth Study 29(6): 1-80.
- 2 Greulich WW, Pyle SI (1959) Radiographic atlas of skeletal development of the hand and wrist. Stanford University Press 91(1): 53.
- 3 Hauspie RC, (1989) Mathematical models for the study of individual growth patterns. Rev Epidemiol Sante Publique 37(5-6): 461-76.
- 4 Hoerr NL (1962) Radiographical Atlas of Skeletal Development of the Foot and Ankle. A Standard of Reference. Arch Dis Child 37(195): 563.
- 5 Bayley N (1943) Skeletal Maturing in Adolescence as a Basis for Determining Percentage of Completed Growth. Child Development 14(1): 1-46.
- 6 Bayley N, (1946) Tables for Predicting Adult Height from Skeletal Age and Present Height. Journal of Pediatrics 28(1): 49-64.
- 7 Bayley N, Schaefer ES (1960) Maternal Behavior and Personality development data from the Berkley Growth Study. Psychiatr Res Rep Am Psychiatr Assoc 13: 155-181.
- 8 Boynton B (1936) The Physical Growth of Girls: a Study of the Rhythm of Physical Growth from Anthropometric Measurements on Girls Between Birth and Eighteen Years. University of Iowa 12(5): 1-105.
- 9 Brush Foundation (1930) Physical and Mental Adolescent Growth: The Proceedings of the Conference on Adolescence. Cleveland. Ohio, USA.
- 10 Buckler JMH, Wild J (1987) Longitudinal study of height and weight at adolescence. Archives of Disease in Childhood 62: 1224-1232.
- 11 Dearborn WF, Rothney JWM, Shuttleworth FK (1938) Data on the Growth of Public School Children (From the Materials of the Harvard Growth Study). Monographs of the Society for Research in Child Development 3(1): 1-136.
- 12 Duren DL, Nahhas RW, Sherwood RJ (2015) Do Secular Trends in Skeletal Maturity Occur Equally in Both Sexes? Clinical Orthopaedics and Related Research 473(8): 2559-2567.
- 13 Buckler JMH (1990) A Longitudinal Study of Adolescent Growth.
- 14 Centers for Disease Control (2010).
- 15 Cole TJ, Green PJ (1992) Smoothing reference centile curves: the LMS method and penalized likelihood. Statistics in Medicine 11(10): 1305-19.
- 16 Borghi E, M de Onis, S Van Burren, R Martorell, C Garza et al. (2006) Construction of the World Health Organization child growth standards: selection of methods for attained growth curves. Statistics in Medicine 25(2): 247-265.
- 17 Chirwa ED, Griffiths PL, Maleta K, Norris SA, Cameron N (2014) Multi-level modelling of longitudinal child growth data from the Birth-to-Twenty Cohort: a comparison of growth models. Annals of Human Biology 41(2): 168-179.
- 18 Johnson W, Balakrishna N, Griffiths PL (2013) Modeling physical growth using mixed effects models. American Journal of Physical Anthropology 150(1): 58-67.
- 19 Hall P, Muller HG, Yao F (2009) Estimation of functional derivatives. The Annals of Statistics 37(6): 3307-3329.
- 20 Runkle DE (1987) Vector Auto regressions and Reality. Journal of Business & Economic Statistics 5(4): 437-442.
- 21 Stock JH, Watson MW (2001) Vector Auto regressions. Journal of Economic Perspectives 15(4): 101-115.
- 22 Zha T (1998) Dynamic Multivariate Model for Use in Formulating policy. Federal Reserve Bank of Atlanta 83(1):16-29(1998).
- 23 Holtz-Eakin D, Newey W, Rosen HS (1998) Estimating Vector Auto regressions with Panel Data. The Econometric Society 56(6): 1371-1395.
- 24 Arellano M, Bond S (1991) Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. The Review of Economic Studies 58(2): 277-297.
- 25 Arellano M, Bover O (1995) Another look at the instrumental variable estimation of error-components models. Journal of Econometrics 68(1): 29-51.

- 26 Maurice BJJ, Vasilis S (2013) Dynamic Panel Data Models. Universiteit van Amsterdam, Dept of Econometrics. Working papers 13-01.
- 27 Tuddenham RD, Snyder MM (1954) Physical growth of California boys and girls from birth to eighteen years. *Publications in Child Development University of California* 1(2): 183-364.
- 28 Nelson S, Cakirer B, Lai YY (2003) Longitudinal changes in craniofacial factors among snoring and nonsnoring Bolton-Brush study participants. *American Journal of Orthodontics and Dentofacial Orthopedics* 123(3): 338-344.
- 29 Griliches Z, Intriligator M (1983) *The Handbook of Econometrics* 1: 3-771.
- 30 Angrist JD, Imbens GW (1995) Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity. *Journal of the American Statistical Association* 90(430): 431-442.
- 31 Scott AJ, Holt D (1982) The Effect of Two-Stage Sampling on Ordinary Least Squares Methods. *Journal of the American Statistical Association* 77(380): 848-854.
- 32 Marion MM (1943) Growth of major long bones in healthy children: a preliminary report on successive roentgenograms of the extremities from early infancy to twelve years of age. *American Journal of Diseases of Children* 66(3): 227-257.
- 33 Kuhn M, Johnson K (2013) *Applied predictive modeling*. Springer.
- 34 James G, Witten D, Hastie T, Tibshirani R (2013) *An introduction to statistical learning: with applications in R*. Springer.
- 35 Kutner MH, Nachtsheim C, Neter J (2004) *Applied linear regression models*.



This work is licensed under Creative Commons Attribution 4.0 License
DOI: [10.19080/BBOAJ.2022.10.555803](https://doi.org/10.19080/BBOAJ.2022.10.555803)

Your next submission with Juniper Publishers will reach you the below assets

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
(Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission

<https://juniperpublishers.com/online-submission.php>