DISCUSSION: CONDITIONAL GROWTH CHARTS

BY RAYMOND J. CARROLL[1] AND DAVID RUPPERT[2]

Texas A&M University and Cornell University

1. Overview. Wei and He are to be congratulated on an innovative and important article. The conditional approach to growth charts described in their article is important in a practical sense, and the use of quantile regression is both natural and well motivated. We look forward to further application of their idea to actual practice, because the concept of "falling behind" in one's growth cycle has two meanings: the usual standard growth chart, and the conditional growth chart described here. We have described the Wei and He approach to pediatricians, and they all grasped the essential clever idea immediately and were enthusiastic about the idea.

Our commentary will focus on three aspects of the approach used by the authors, specifically (a) the use of *unpenalized* B-splines as described by the authors; (b) conditional versus marginal semiparametric modeling of longitudinal data; and (c) some alternative modeling approaches to "catch-up" that may get at the issue more directly and flexibly.

2. B-splines should be penalized. One purpose of discussions, of course, is to make things a little lively, and here is our contribution. In our view, one should have some skepticism of how nonparametric unpenalized B-splines and unpenalized regression splines really are in the context of *non*parametric regression. More precisely, and less inflammatory, the connection between asymptotic theory for unpenalized splines and actually attempting to be at least reasonably nonparametric is not at all clear.

There is obviously a need to balance practical behavior and ease of implementation with theory. Kernel methods (see below) are one means of doing this. In the spline context, there are four approaches: smoothing splines, penalized regression splines, unpenalized regression splines and free-knot splines; see [18] for a recent review. As authors, we have observed the following.

- Smoothing splines are basically penalized regression splines that place a knot at every value of the covariate. We have no idea how to do asymptotic theory for

smoothing splines in the Wei and He context, and we would be interested to see if such a theory is even possible, for example, along the lines of [13].

- Lower-order knots penalized regression splines are rapidly becoming a practical method of choice. However, every time we write a paper about this technique, we have been reviewed by a smoothing spline person who asks the questions: "how many knots" and "where do you place the knots." In response, Ruppert [16] did an extensive numerical study and showed that, in effect, penalized regression splines with data-adaptive penalties and 40 knots will work quite well in many practical settings. Ruppert and Carroll [17] and Ruppert, Wand and Carroll [18], Chapters 5 and 17, describe strategies for selecting the number of knots. There is little in the way of asymptotic theory for penalized regression splines, unfortunately, presumably for the same reason that penalized smoothing splines are difficult to analyze; this point is also brought up in every review. For some theory, see [3].

- Unpenalized regression splines with 40 knots are generally ridiculously non-smooth, which is of course why penalties are used. For example, in Figure 1, we generated data according to the regression model $Y = \sin(2X) + \varepsilon$, where X is equally spaced on the unit interval, $n = 400$ and $\varepsilon = \text{Normal}(0, 1)$. We used 40 knots and fit penalized and unpenalized cubic B-splines, using the Matlab software available from Brian Marx at www.stat.lsu.edu/faculty/marx/.

- There is a substantial literature on unpenalized B-splines that achieve their smoothness via knot selection, the so-called free-knot spline methodology. Much of this literature is Bayesian, being based on model averaging; see, for example, [2, 4, 19]. With such complex methods, theoretical results are naturally generally lacking, although see [12] and [20] for recent non-Bayesian approaches with impressive (asymptotic) theoretical treatment.

In summary, the nonparametric spline literature generally uses a fairly large number of knots, realizes that penalties must be imposed in one way or another to obtain smoothness, and achieves this smoothness either by direct penalties, or by some version of knot selection. Often for these methods asymptotic theory is not available.

In contrast, unpenalized regression spline methods do have a beautiful asymptotic theory, as exemplified here and in a series of important papers by Jianhua Huang and colleagues [5–9]. Plots, such as in Wei and He, are often pleasantly smooth, in contrast to Figure 1. The obvious question is: what is going on?

To get such smoothness in an unpenalized regression spline, one necessarily must insist that the number of knots be small. Indeed, Wei and He use cubic B-splines with three knots in their examples. This is no fluke: if one fixes the knots, and does no penalization or knot selection, there is no getting around phenomena such as in Figure 1. Indeed, the basic point is actually one of the conditions of their Theorem 3.1. Note that the essential condition for a function of bounded second
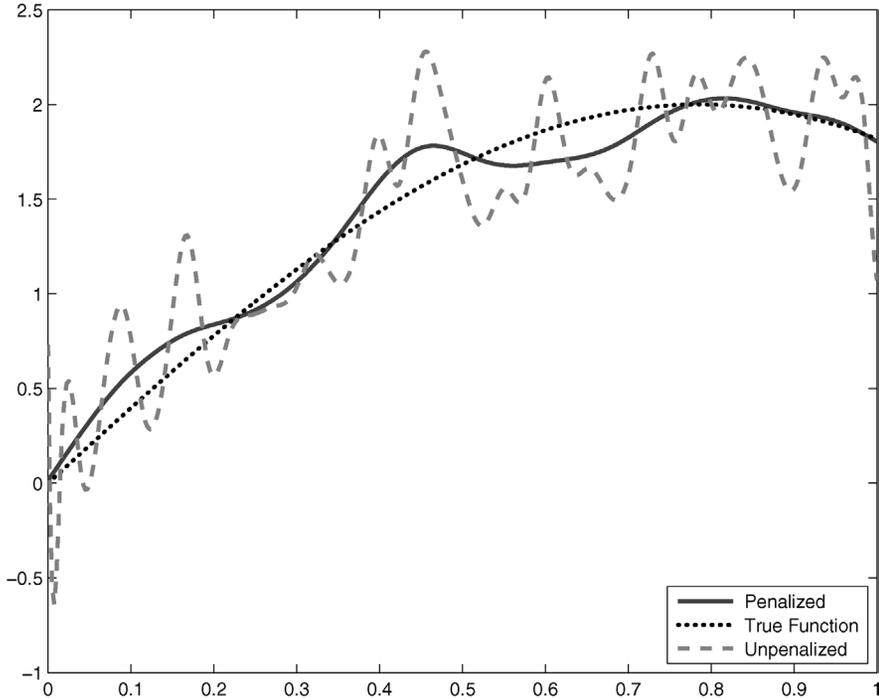
FIG. 1.   *Fits to the regression model* $Y = \sin(2X) + \varepsilon$, *where X is equally spaced on the unit interval and* $\varepsilon = \mathrm{Normal}(0, 1)$. *The dotted line is the true function. The solid line is a penalized cubic B-spline fit with* 40 *knots. The dashed line uses the same basis*, *but lacks a penalty.*

derivative is that the number of knots be proportional to $k \propto n^{1/5}$, or equivalently, that the sample size $n \propto k^5$. Since $3^5 = 243$, $4^5 = 1024$, $5^5 = 3125$ and $6^5 = 7776$, we see that the asymptotics essentially require that for any practical problem, the number of knots should be no more than six. This is hardly nonparametric regression! Of course, these calculations are deliberately shocking and totally slanted in order to add some controversy, because we have not mentioned constants of proportionality, but once one does get into estimating the number of knots and where they should be placed, then it is not clear what the theory would be.

We note that Wei and He's criterion function (2.2) could have been analyzed by kernel methods, using local-likelihood ideas. Theorem 3.1 would then have been easy to analyze using either profiling or backfitting; see, for example, [1] for the computationally far easier backfitting and [21] for the more complex profiling. We conjecture, of course, that the same limiting result as in Theorem 3.1 would have been obtained, since one way to interpret the work of Newey [15] is that different implementations of the same criterion function should not lead to different limiting results for parametric components. It would be interesting to know whether this conjecture is correct.

3. Marginal versus conditional longitudinal models. Wei and He deal with longitudinal semiparametric models, using a conditional approach that is ideally suited to their main idea.

The usual approach to this problem is via a marginal model, for example, in a slightly altered notation,

(1)
$$Y_{ij} = m(X_{ij}, \beta) + \theta(Z_{ij}) + \varepsilon_{ij};$$

$$\text{cov}(\varepsilon_{i1} \ldots) = \Sigma_i(\tau),$$

where $m(\cdot)$ is a known function. Wang, Carroll and Lin [22] describe the semiparametric efficient solution to this problem if the ε_{ij} are Gaussian using kernels, although the work of Lin et al. [11] essentially shows the same result for smoothing splines. There is considerable controversy about this problem. Early solutions (and many later ones) simply ignored the correlation structure, with a notable exception being Zeger and Diggle [23]. The semiparametric efficient solution for Gaussian data is explicit, that is, not iterative, but it still takes some work to implement.

Wei and He's conditional model approach, in contrast, neatly avoids all these issues, because their model is

(2)
$$Y_{ij} = m(Y_{i,k}, X_{ij}, \beta^*) + \theta(Z_{ij}) + \varepsilon_{ij}^*,$$

where $k < j$ and the ε_{ij}^* are independent. Here even the kernel approach is simple to implement.

We would be interested in Wei and He's thoughts on how to use model (2) to help understand marginal models such as (1), *in the context of their innovative quantile regression modeling*. Since they (quite properly) do not work in a mean-based model, the marginal interpretation of their model, if any, is unclear to us. If there were a marginal interpretation, then perhaps this could be used in the more standard unconditional growth chart arena.

On a theoretical note, if one starts from a conditional model and then turns it into a marginal model, it is not clear whether the methods of estimation are semiparametric efficient *in the marginal model*. Recent work by Lin and Carroll [10] can be used to answer this question.

4. Other literature. Much work on nonparametric estimation of marginal curves has been done in the past using kernel methods; see [14] and the references therein. We are uncertain how extensively these methods have been adopted by practitioners, but since they are not mentioned in this paper it seems that they are at least somewhat neglected. Perhaps Professors Wei and He can comment. In Müller's book there is an emphasis on estimating the first two derivatives, that is, growth velocity and acceleration. Müller mentions growth spurts during adolescence and perhaps these might be missed by a three-knot cubic spline. Therefore, we wonder whether the methods in this paper can be applied to long time spans

that include adolescence or, instead, whether a methodology with a smoothing parameter would be needed.

In (5.1), it seems more appropriate to use $H_{i,j}^3$ instead of $H_{i,j}$ as a covariate, since weight should be roughly proportional to the cube of a linear dimension. Perhaps the increase in c_τ with τ seen in Table 2 and mentioned by the authors is due to using $H_{i,j}$ instead of its cube. Similarly, the authors discuss a boy who jumped from the 0.25th quantile to the median level at age 0.61 year and concluded "that he might be overweight given his prior path and current height." This is an interesting and potentially important conclusion, but one must be certain that the effect of height is being modeled correctly before making it.

5. Alternative models with catch-up. In their Introduction, Professors Wei and He mention "catch-up" growth where subjects on the upper (or lower) centiles move toward the median at a faster rate than others. Model (2.1) is very general and should include the possibility of catch-up growth, but it would be helpful to practitioners if the mechanism in the model for catch-up growth was explicit. Consider (5.1) of the paper with $k = 1$, that is,

$$W_{i,j} = g_\tau(t_{i,j}) + (a_\tau + b_\tau D_{i,j})W_{i,j-1} + c_\tau H_{i,j} + e_{i,j}.$$

There is no apparent mechanism for catch-up growth here, since, with a_τ and b_τ both positive as in Table 2, $W_{i,j}$ is an increasing function of $W_{i,j-1}$. Of course, it might be that $H_{i,j}$ catches up and forces $W_{i,j}$ to catch up as well. This raises the question of whether $H_{i,j}$ should be a covariate or a second response. Perhaps height and BMI should be modeled jointly as a bivariate dynamic process.

We have thought about other models where catch-up growth might be more explicit. One model, which we realize may be too simple but could be a good starting point, is

$$
\begin{aligned}
W_{i,j} = {} & W_{i,j-1} + \{g_\tau(t_{i,j}) - g_\tau(t_{i,j-1})\} \\
& + b D_{i,j}\{W_{i,j-1} - g_\tau(t_{i,j-1})\} + D_{i,j}e_{i,j} \\
= {} & W_{i,j-1} + \text{average change in the population} + \text{subject-specific change}.
\end{aligned}
$$
(3)

If $b < 0$, then there is catch-up growth, because subjects with $W_{i,j-1} - g_\tau(t_{i,j-1})$ positive (negative) tend to grow less (more) than average. One feature of (3) is that $W_{i,j} \to W_{i,j-1}$ as $t_{i,j} \to t_{i,j-1}$ (so that $D_{i,j} \to 0$). Anything else, of course, would not be realistic. Besides the parameters in g_τ, (3) has only a single parameter b. However, b probably should depend on time since Professors Wei and He mention that catch-up tends to be time-specific. One might replace b by $b\{(t_{i,j} + t_{i,j-1})/2\}$ where $b(\cdot)$ is a spline.

Let $W_{i,j}^* = W_{i,j} - g_\tau(t_{i,j})$. Then (3) with b a spline can be written as

$$W_{i,j}^* = [1 - b\{(t_{i,j} + t_{i,j-1})/2\}D_{i,j}]W_{i,j-1}^* + D_{i,j}e_{i,j},$$

which is similar to an AR(1) process.

6. Concluding comments. Notwithstanding our comments about general statistical methodology, and our comments about practical details, Wei and He have written an important paper, and we look forward to reading about further developments of their innovative ideas. We have rarely read such a thought-provoking paper that has the potential to be extremely important in practice.

REFERENCES

[1] CHEN, X., LINTON, O. and VAN KEILEGOM, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* **71** 1591–1608. MR2000259

[2] DIMATTEO, I., GENOVESE, C. R. and KASS, R. E. (2001). Bayesian curve fitting with free-knot splines. *Biometrika* **88** 1055–1071. MR1872219

[3] HALL, P. and OPSOMER, J. D. (2005). Theory for penalised spline regression. *Biometrika* **92** 105–118. MR2158613

[4] HOLMES, C. C. and MALLICK, B. K. (2001). Bayesian regression with multivariate linear splines. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **63** 3–17. MR1811987

[5] HUANG, J. Z. (2001). Concave extended linear modeling: A theoretical synthesis. *Statist. Sinica* **11** 173–197. MR1820005

[6] HUANG, J. Z. (2003). Local asymptotics for polynomial spline regression. *Ann. Statist.* **31** 1600–1635. MR2012827

[7] HUANG, J. Z., KOOPERBERG, C., STONE, C. J. and TRUONG, Y. K. (2000). Functional ANOVA modeling for proportional hazards regression. *Ann. Statist.* **28** 961–999. MR1810916

[8] HUANG, J. Z., WU, C. O. and ZHOU, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* **89** 111–128. MR1888349

[9] HUANG, J. Z., WU, C. O. and ZHOU, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statist. Sinica* **14** 763–788. MR2087972

[10] LIN, X. and CARROLL, R. J. (2006). Semiparametric estimation in general repeated measures problems. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 69–88. MR2212575

[11] LIN, X., WANG, N., WELSH, A. H. and CARROLL, R. J. (2004). Equivalent kernels of smoothing splines in nonparametric regression for clustered/longitudinal data. *Biometrika* **91** 177–193. MR2050468

[12] MAMMEN, E. and VAN DE GEER, S. (1997). Locally adaptive regression splines. *Ann. Statist.* **25** 387–413. MR1429931

[13] MAMMEN, E. and VAN DE GEER, S. (1997). Penalized quasi-likelihood estimation in partial linear models. *Ann. Statist.* **25** 1014–1035. MR1447739

[14] MÜLLER, H.-G. (1988). *Nonparametric Regression Analysis of Longitudinal Data. Lecture Notes in Statist.* **46**. Springer, Berlin. MR0960887

[15] NEWEY, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica* **62** 1349–1382. MR1303237

[16] RUPPERT, D. (2002). Selecting the number of knots for penalized splines. *J. Comput. Graph. Statist.* **11** 735–757. MR1944261

[17] RUPPERT, D. and CARROLL, R. J. (2000). Spatially-adaptive penalties for spline fitting. *Aust. N. Z. J. Stat.* **42** 205–223.

[18] RUPPERT, D., WAND, M. P. and CARROLL, R. J. (2003). *Semiparametric Regression*. Cambridge Univ. Press. MR1998720

[19] SMITH, M. and KOHN, R. (1996). Nonparametric regression using Bayesian variable selection. *J. Econometrics* **75** 317–343.

[20] STONE, C. J. and HUANG, J. Z. (2002). Free knot splines in concave extended linear modeling. *J. Statist. Plann. Inference* **108** 219–253. MR1947401

[21] VAN KEILEGOM, I. and CARROLL, R. J. (2006). Backfitting versus profiling in general criterion functions. *Statist. Sinica*. To appear.

[22] WANG, N., CARROLL, R. J. and LIN, X. (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data. *J. Amer. Statist. Assoc.* **100** 147–157. MR2156825

[23] ZEGER, S. L. and DIGGLE, P. J. (1994). Semi-parametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* **50** 689–699.

DEPARTMENT OF STATISTICS
TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS 77843-3143
USA
E-MAIL: carroll@stat.tamu.edu

SCHOOL OF OPERATIONS RESEARCH
  AND INDUSTRIAL ENGINEERING
225 RHODES HALL
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853
USA
E-MAIL: dr24@cornell.edu