Appendices

A Proof of Theorem 1

We consider the L_2-metric

$$d(g_1, g_2) = \|g_1 - g_2\|_{L_2(P)} = \left[P(g_1 - g_2)^2\right]^{\frac{1}{2}},$$

where Pf denotes the expectation of a measurable, data-dependent function f. Now, by Condition C1, it is easy to see that the expected objective function is

$$M(g) = Pm_g = -P(g - g_{0,i})^2, \quad i = 1, 2,$$

where $g_{0,i}$ is the true curve for the ith population. This implies the identifiability condition

$$\sup_{g: g \notin G} M(g) < M(g_{0,j})$$

for any neighborhood G_i of $g_{0,i}$.

For simplicity of presentation, we omit the subscript i in the remainder of the proof, and note that the same arguments regarding the consistency of the B-spline estimator hold for both groups. Then, following Kim et al. (2017)[1], we define a linear operator map $\mathcal{Q}$ from $\mathcal{G}$ to the sieve space $\mathcal{G}_n$, as:

$$\mathcal{Q}(\psi) = \sum_{k=1}^{K_n+m} \phi_k(\psi) B^m(x),$$

for any $\psi \in \mathcal{G}$, where $\{\phi_k\}_{k=1}^{K_n+m}$ are linear functionals in $L_\infty(\mathbb{X})$. Then we define $g_n(x) = \mathcal{Q}(g_0)$. Arguments similar to those used by Kim et al. (2017)[1] lead to the following inequality

$$\|g_n - g_0\|_{L_\infty(\mathbb{X})} \leq O(K_n^{-p}),$$

which also implies that

$$\|g_n - g_0\|_{L_2(P)} \leq O(K_n^{-p}). \tag{1}$$

It is straightforward to see that

$$d(\hat{g}_n, g_0) \leq d(\hat{g}_n, g_n) + d(g_n, g_0). \tag{2}$$

To show the convergence of the first term in the right side of (2) to 0, we let $\mathbb{M}_n(g)$ be the empirical objective function based on the data and $\mathbb{P}_n f(X) = n^{-1} \sum_{i=1}^n f(X_i)$. It follows that

$$\begin{aligned} \sup_{g \in \mathcal{G}_n} |\mathbb{M}_n(g) - M(g)| &\equiv \|\mathbb{M}_n(g) - M(g)\|_{\mathcal{G}_n} \\ &\lesssim \|\mathbb{P}_n(g - g_0)e\|_{\mathcal{G}_n} + \|(\mathbb{P}_n - P)(g - g_0)^2\|_{\mathcal{G}_n}. \end{aligned}$$

From Conditions C1, C2, and the Law of Large Numbers, we have $\|\mathbb{P}_n(g-g_0)e\|_{\mathcal{G}_n} = o_p(1)$.

For the second term we consider the class of functions $\mathcal{F}_n = \{(g-g_0)^2 : g \in \mathcal{G}_n\}$. A calculation by Shen and Wang (1994)[2] implies that $N_{[]}(\epsilon, \mathcal{G}_n, L_1(P)) \le (1/\epsilon)^{c(K_n+m)}$. Based on the set of ϵ-brackets $\{[l_j, u_j] : j = 1, \dots, (1/\epsilon)^{c(K_n+m)}\}$ in $L_1(P)$ for $\mathcal{G}_n$, we can construct a set of ϵ'-brackets $\{[l_j', u_j'] : j = 1, \dots, (1/\epsilon)^{c(K_n+m)}\}$, where

$$l_j' = \left[I(l_j \ge g_0)(l_j - g_0)^2 + I(u_j < g_0)(u_j - g_0)^2\right]\left[1 - I(l_j < g_0, u_j \ge g_0)\right]$$

and

$$\begin{aligned} u_j' &= I(l_j \ge g_0)(u_j - g_0)^2 + I(u_j < g_0)(l_j - g_0)^2 \\ &\quad + I(l_j < g_0, u_j \ge g_0)\max((u_j - g_0)^2, (l_j - g_0)^2) \end{aligned}$$

in $L_1(P)$ for $\mathcal{F}_n$. Therefore, $N_{[]}(\epsilon, \mathcal{F}_n, L_1(P)) < \infty$ for any $\epsilon > 0$. This implies that $\|(\mathbb{P} - P)(g-g_0)^2\|_{\mathcal{G}_n} \stackrel{as*}{\to} 0$, and thus $\|\mathbb{M}_n(g) - M(g)\|_{\mathcal{G}_n} = o_p(1)$. This fact along with the inequality

$$M(g) - M(g_n) \le -\frac{1}{4}P(g - g_n)^2,$$

for any g with $P(g-g_n)^2 \ge 4P(g_n - g_0)^2$ (van der Vaart and Wellner, 1996 [3]) implies that $d(\hat{g}_n, g_n) = o_p(1)$. The second term on the right side of (2) is $o(1)$ by condition C3 and inequality (1) and this leads to $d(g_n, g_0) = o(1)$ and, therefore, $d(\hat{g}_n, g_0) \stackrel{p}{\to} 0$.

For the rate of convergence we consider the key inequality

$$\begin{aligned} E^\star \sup_{P(g-g_n)^2 \le \delta^2, g \in \mathcal{G}_n} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n (g - g_n)(X_i)e_i \right| &\lesssim \\ \tilde{J}_{[]}(\delta, \mathcal{G}_n(\delta), L_2(P)) &\left[1 + \frac{\tilde{J}_{[]}(\delta, \mathcal{G}_n(\delta), L_2(P))}{\delta^2 \sqrt{n}}\right] \end{aligned}$$

holds with $\mathcal{G}_n(\delta) = \{g : g \in \mathcal{G}_n, d(g, g_n) < \delta\}$, given in p. 335 in van der Vaart and Wellner (1996) [3]. The calculation by Shen and Wang (1994)[2] implies that the ϵ-bracketing number for the class $\mathcal{G}_n(\delta)$ is bounded by $(\delta/\epsilon)^{c(K_n+m)}$. Therefore,

$$\tilde{J}_{[]}(\delta, \mathcal{G}_n(\delta), L_2(P)) = \int_0^\delta \sqrt{1 + c(K_n + m)\log\left(\frac{\delta}{\epsilon}\right)}\, d\epsilon \le c(K_n + m)^{1/2}\delta.$$

Thus, the key function $\phi_n(\delta)$ given in Theorem 3.4.1. in van der Vaart and Wellner (1996) [3] is

$$\phi_n(\delta) = (K_n + m)^{1/2}\delta + \frac{(K_n + m)}{\sqrt{n}}.$$

After some algebra we conclude that

$$n^{2pv}\phi_n\left(\frac{1}{n^{pv}}\right) \leq \sqrt{n}$$

if $pv \leq (1-v)/2$. Thus, if the rate $r_n = \min(pv, (1-v)/2)$ then

$$r_n^2\phi_n\left(\frac{1}{r_n}\right) \leq \sqrt{n}.$$

Also, it can be argued that $\mathbb{M}_n(\hat{g}_n) - \mathbb{M}_n(g_0) \geq -O_p(r_n^{-2})$. If $v = 1/(1+2p)$, Theorem 3.4.1 in van der Vaart and Wellener (1996) [3] and (1) and (2) imply that

$$d(\hat{g}_n, g_0) = O_p(n^{-\frac{p}{1+2p}}).$$

Next, consider the test statistic based on two independent samples of size n_1 and n_2

$$\frac{1}{N}\sum_{i=1}^{2}\sum_{j=1}^{n_i}[\hat{g}_1(\mathbf{x}_{ij}) - \hat{g}_2(\mathbf{x}_{ij})]^2 \equiv \mathbb{P}_N(\hat{g}_1 - \hat{g}_2)^2,$$

where $N = n_1 + n_2$.

Let us consider the consistency of the test statistic to the true L^2 distance between the two curves under the probability measure P underlying X_{ij}, $i = 1, 2$, $j = 1, \ldots, n_i$. It is straightforward to show that

$$\begin{aligned}\left|\mathbb{P}_N(\hat{g}_1 - \hat{g}_2)^2 - P(g_1 - g_2)^2\right| \quad &\leq \quad \left|\mathbb{P}_N[(\hat{g}_1 - \hat{g}_2)^2 - (g_1 - g_2)^2]\right| \\ &\quad + \left|(\mathbb{P}_N - P)(g_1 - g_2)^2)\right|. \qquad (3)\end{aligned}$$

The second term in (3) is $o_p(1)$, as a consequence of the condition C2 and the law of large numbers. Now see that

$$\begin{aligned}\mathbb{P}_N(\hat{g}_i - g_i)^2 \quad &= \quad \mathbb{P}_N\left\{\sum_{k=1}^{K_{n_i}+m}[\hat{\phi}_{k,i}(g_i) - \phi_{k,i}(g_i)]B_k^m + \left[\sum_{k=1}^{K_n+m}\phi_{k,i}(g_i)B_k^m - g_i\right]\right\}^2 \\ &\equiv \quad \mathbb{P}_N\left\{\sum_{k=1}^{K_{n_i}+m}[\hat{\phi}_{k,i}(g_i) - \phi_{k,i}(g_i)]B_k^m + (g_{n_i,i} - g_i)\right\}^2 \\ &\leq \quad 2\left\{\sum_{k=1}^{K_{n_i}+m}[\hat{\phi}_{k,i}(g_i) - \phi_{k,i}(g_i)]\right\}^2 \mathbb{P}_N\left(\max_{k=1,\ldots,K_{n_i}+m} B_k^m\right)^2 \\ &\quad +2P(g_{n_i,i} - g_i)^2 + o_p(1).\end{aligned}$$

By the uniform boundedness of the B-spline basis functions and the consistency of the estimator of the control points $\hat{\phi}_{k,i}(g_i)$ from the fact that $d(\hat{g}_{i,n}, g_{i,n}) = o_p(1)$ shown above, it follows that the first term in the right side of the above inequality is $o_p(1)$. Also, the second term in the right side of the above inequality is $o_p(1)$, as it was argued above. Therefore,

$$\mathbb{P}_N(\hat{g}_i - g_i)^2 = o_p(1), \qquad i = 1, 2. \tag{4}$$

(3) can be written into

$$\begin{aligned} \left|\mathbb{P}_N[(\hat{g}_1 - \hat{g}_2)^2 - (g_1 - g_2)^2]\right| &\leq \sum_{i=1}^{2} \left|\mathbb{P}_N(\hat{g}_i^2 - g_i^2)\right| + |\mathbb{P}_N \hat{g}_1(\hat{g}_2 - g_2)| \\ &\quad + |\mathbb{P}_N g_2(\hat{g}_1 - g_1)| \\ &\equiv \sum_{i=1}^{2} A_{N,i} + B_N + C_N. \end{aligned} \tag{5}$$

It is not hard to see that the first term of (5) becomes

$$\begin{aligned} \sum_{i=1}^{2} A_{N,i} &\leq \sum_{i=1}^{2} \mathbb{P}_N\,(\hat{g}_i - g_i)^2 + \sum_{i=1}^{2} \mathbb{P}_N\,|2g_i(\hat{g}_i - g_i)| \\ &\leq \sum_{i=1}^{2} \mathbb{P}_N(\hat{g}_i - g_i)^2 + K \sum_{i=1}^{2} \mathbb{P}_N\,|\hat{g}_i - g_i| \\ &\leq \sum_{i=1}^{2} \mathbb{P}_N(\hat{g}_i - g_i)^2 + K \sum_{i=1}^{2} [\mathbb{P}_N(\hat{g}_i - g_i)^2]^{1/2} \\ &= o_p(1), \end{aligned}$$

where K represents a generic constant with $K \in (0, \infty)$. By C2 and (5), $i = 1, 2$, we have that for a sufficiently large N

$$(B_N + C_N) \leq K \sum_{i=1}^{2} [\mathbb{P}_N(\hat{g}_i - g_i)^2]^{1/2} = o_p(1).$$

Therefore,

$$\mathbb{P}_N(\hat{g}_1 - \hat{g}_2)^2 \xrightarrow{p} P(g_1 - g_2)^2.$$

This result along with Lemma 14.15 of van der Vaart (2000)[4] leads to the consistency of the proposed test against every fixed alternative hypothesis with $g_1 \neq g_2$. Thus, the proof of Theorem 1 is complete.

B Simulation Results

In Part B of the Appendix, we summarize the results from the simulation studies. Details of the simulation were presented in Section 5 of the manuscript.

B.1 Comparisons of curve functions

We report the average type I error rates and power for parameter settings considered in the simulation. Simulations for curve comparisons were based on 1000 generated samples. Tests were performed at the 0.05 significance level.

Notation for each testing method in the Table 1 is consistent with that presented in the main manuscript. Method 1 $T_{B-spline}$: L^2 distance of pointwise B-spline regression functions with $k = \sqrt[3]{n_i}$; Method 2 $T_{P-spline}$: P-spline regression functions with the default number of knots from GCV ($k \approx 6$ in this example); Method 3 T_4: Kernel-based regression function $\tilde{g}_i(x)$, $T_4 = \sum_{i=1}^{I}\sum_{j=1}^{i-1}[\tilde{g}_i(x) - \tilde{g}_j(x)]^2$; Method 4 T_2: Variance estimating method, $T_2 = \hat{\sigma}^2 - \frac{1}{N}\sum_{i=1}^{I} n_i\hat{\sigma}_i^2$; Method 5 T_1: The scaled chi-square test, $T_1 = \frac{1}{N}\sum_{i=1}^{I}\sum_{j=1}^{n_i}[\tilde{g}(x_{ij}) - \tilde{g}_i(x_{ij})]^2$.

[Figure 1 about here.]

[Table 1 about here.]

We report the average type I error rates and power for parameter settings considered in the simulation.

Simulations for three-group curve comparisons were based on 500 generated samples. Tests were performed at the 0.05 significance level. Simulation results were presented in Table 2.

[Table 2 about here.]

B.2 Exploration of asymptotic property

To confirm the theoretical results that power increases with the sample size, we randomly chose a scenario that $d = 1$, $(\sigma_1, \sigma_2) = (0.2, 0.15)$ in two curve comparisons based on 1000 simulated samples at the significance level of 0.05. Numerical results are presented in Table 3, and power curve in Figure 2.

[Table 3 about here.]

[Figure 2 about here.]

B.3 Comparisons of surface functions

The average type I error rate and power for various parameter settings in surface comparisons based on 500 simulated samples at the significance level of 0.05 are presented in Tables 4 and 5, where "TP-Spline", "TP-Spline$_+$" and "TP-Spline$_-$" indicate tests using thin-plate splines with $\sqrt[3]{n_i}$, $\sqrt[3]{n_i} + 1$, and $\sqrt[3]{n_i} - 1$

knots. "TE" indicates testing method using tensor-product basis functions, "TP-Spline.p" and "TE-Spline.p" are tests using penalized splines. T_4(Method 3): Kernel based estimating regression function $\tilde{g}_i(\mathbf{x})$, $T_4 = \sum_{i=1}^{I} \sum_{j=1}^{i-1} [\tilde{g}_i(\mathbf{x}) - \tilde{g}_j(\mathbf{x})]^2$; T_2(Method 4): variance estimating method, $T_2 = \hat{\sigma}^2 - \frac{1}{N} \sum_{i=1}^{I} n_i \hat{\sigma}_i^2$; T_3(revised Method 5): $T_3 = \frac{1}{N} \sum_{i=1}^{I} \sum_{j=1}^{n_i} [\hat{g}(\mathbf{x}_{ij}) - \hat{g}_i(\mathbf{x}_{ij})]^2$; T_1(Method 5): matching with a scaled chi-square distribution, $T_1 = \frac{1}{N} \sum_{i=1}^{I} \sum_{j=1}^{n_i} [\tilde{g}(\mathbf{x}_{ij}) - \tilde{g}_i(\mathbf{x}_{ij})]^2$.

[Table 4 about here.]

[Table 5 about here.]

B.4 Models with correlated data

We present the average type I error rates and power for simulation settings in curve comparisons based on 1000 simulated samples with certain correlations at the significance level of 0.05; see Tables 6. We generated 200 simulated samples for comparison of the rejection probabilities between the proposed methods and Zhang's scaled χ^2 testing method. See Table 7.

[Table 6 about here.]

[Table 7 about here.]

C Package gamm4.test and its R Shiny interface

As described in the manuscript, we developed an R package, `gamm4.test`, together with an interactive interface for the implementation of the proposed tests. In Part C of the Appendix, we briefly illustrate the use of the package.

C.1 Analysis of cross-sectional data: An example

Using the pubertal growth study data as an example, we compared the weight-for-age curves between boys and girls, using only the baseline measurements. The data are cross-sectional and uncorrelated.

```
R> library("gamm4.test")
R> data("outchild")
R> child <- outchild[order(outchild$SID,outchild$age),]
R> bs <- aggregate(.~SID, child, FUN=head, 1)

R> childcur <- bs[,c("SEX","WEIGHT","age")]
R> test.grpsex1 <- gam.grptest(WEIGHT~s(age, bs="cr"),
      test=~SEX,data=childcur)
R> test.grpsex1
```

The output from the program thus far is as follows

```
Test the equality of curves based on L2 distance

Comparing 2 semiparametric regression curves
Penalized semiparametric regression is used for curve fitting.
Wide-bootstrap algorithm is applied to obtain the null distribution.

Null hypothesis: there is no difference between the 2 curves.
T =  71.92     p-value =  0.01493
```

The density function of the test statistic under the null hypothesis estimated from the bootstrap samples can be obtained by using function

```
plot(test.grpsex1, test.statistic=TRUE).
```

The function provides the flexibility of displaying either a histogram or a density curve through option `test.stat.type`. In Figure 3 we request a histogram by using

```
plot(test.grpsex1, test.statistic=TRUE, test.stat.type="hist")
```

[Figure 3 about here.]

Function gam.grptest is the main function for comparing the curves. Options bs= and k= can be used in the model formula for changing the basis functions and for specifying the number of knots. If no value for k is provided, a penalized semiparametric model estimation with the default number of knots will be used. The following are some sample commands:

```
R> test.grpsex1 <- gam.grptest(WEIGHT~s(age, bs="tp"),test=~SEX,
      data=childcur)
      #penalized thin-plate spline basis
R> test.grpsex1 <- gam.grptest(WEIGHT~s(age, bs="tp", k=5), test=~SEX,
       data=childcur)
      #thin-plate spline basis with five equally spaced knots
      #over the range of variable age
```

Option N.boot specifies the number of bootstrap samples. The default value is N.boot = 200. Option parallel=TRUE calls for parallel computing and it distributes the computational burden to all available CPU cores.

```
R> test.grpsex1 <- gam.grptest(WEIGHT~s(age, bs="cr"),test=~SEX,
      data=childcur, N.boot=300, parallel= TRUE)
R> test.grpsex1
```

The following code produces a plot of the estimated curves with a 95% pointwise confidence interval. The plot is shown in Figure 4.

```
R> plot(test.grpsex1)
R> plot(test.grpsex1, se.est=TRUE)
```

[Figure 4 about here.]

Similarly, one could use the function gam.grptest for comparison of surface functions. In the pubertal growth example, we express the body weight as a function of age and height, i.e., $WEIGHT = f(HEIGHT, age)$. The following code produces a comparison of the function f between boys and girls.

```
R> childsurf <- bs[,c("SEX","HEIGHT","WEIGHT","age")]
R> test.grpsex2 <- gam.grptest(WEIGHT~s(HEIGHT,age),test=~SEX,
      data=childsurf)
R> test.grpsex2
R> plot(test.grpsex2)
R> plot(test.grpsex2,type="persp",theta=-35,phi=40)
R> plot(test.grpsex2,type="plotly.persp")
```

with the following output:

```
Test the equality of surfaces based on L2 distance

Comparing 2 semiparametric regression surfaces
```

```
Penalized semiparametric regression is used for surface fitting.
Wide-bootstrap algorithm is applied to obtain the null distribution.

Null hypothesis: there is no difference between the 2 surfaces.
T =  20.92     p-value =  0.4179
```

The option `type=plotly.persp` generates an interactive 3-D plot using the R package `plotly`. The generated plots are shown in Figure 5 and plotly output is uploaded on https://zhaoshi169.github.io/chap5plotlysuf.html.

[Figure 5 about here.]

The R package `fANCOVA` provides functions for kernel-based testing methods. We made a minor correction for the function `T.L2` in our package with a new function name `T.L2c`, which helps to conduct two group comparison.

```
R> n1 <- 200
R> x1 <- runif(n1,min=0, max=3)
R> sd1 <- 0.2
R> e1 <- rnorm(n1,sd=sd1)
R> y1 <- sin(2*x1) + cos(2*x1) + e1
R>
R> n2 <- 120
R> x2 <- runif(n2, min=0, max=3)
R> sd2 <- 0.25
R> e2 <- rnorm(n2, sd=sd2)
R> y2 <- sin(2*x2) + cos(2*x2) + x2 + e2
R>
R> dat <- data.frame(rbind(cbind(x1,y1,1), cbind(x2,y2,2)))
R> colnames(dat)=c('x','y','group')
R>
R> T.L2c(formula=y~x,test=~group,data=dat)
R> gam.grptest(y~s(x,bs="cr"), test=~group, data=dat,
      parallel=TRUE)
R> library(fANCOVA)
R> T.aov(dat$x, dat$y, dat$group)
R> T.var(dat$x, dat$y, dat$group)
```

In this simple example, all testing methods correctly reject the null hypothesis. But as we have shown in the simulation studies, different testing methods do have different operating characteristics.

C.2 Analysis of correlated data

We use the same example to illustrate the analysis of correlated data. Here, we included data from all the visits for the analysis. The goal is to compare the growth functions over age between boys and girls.

The R code for data preparation and testing is presented as below.

```
R> data("outchild")
R> child.rep <- outchild[(outchild$age<16 &outchild$age>10),]
R> child.reptest1 <-  gamm4.grptest(HEIGHT~s(age,bs="cr"),
      random=~(1|SID),test=~SEX,data=child.rep)
R> child.reptest1
R> plot(child.reptest1,test.statistic=TRUE)
R> plot(child.reptest1)
```

The output is as follows:

```
Test the equality of curves based on L2 distance

Comparing 2 semiparametric regression curves
Penalized semiparametric regression mixed model is used for curve fitting.
Wide-bootstrap algorithm is applied to obtain the null distribution.

Null hypothesis: there is no difference between the 2 curves.
T =  33.91     p-value =  0.004975
```

As expected, the height-for-age growth curves are significantly different between the sexes ($p<0.001$). The empirical distribution of the test statistic under the null hypothesis shows that the value of the test statistic, $T = 33.91$, is located to the far right of the plotted range. At the same time, the 95% pointwise confidence bands were relatively narrow, showing diverging growth patterns around the time of puberty. See Figure 6.

[Figure 6 about here.]

The height-for-age scatter plot is presented in Figure 6. The corresponding regression curves and the 95% pointwise confidence intervals from the semiparametric mixed model analysis were presented in Figure 6.

To implement surface comparisons, we conducted hypothesis testing on the simultaneous nonlinear effects of height and age on weight between boys and girls among blacks.

```
R> child.repw <- child.rep[(child.rep$RACE==1),]
R> child.reptest2 <- gamm4.grptest(WEIGHT~t2(age,HEIGHT),
      random=~(1|SID),test=~SEX,data=child.repw)
R> child.reptest2
R> plot(child.reptest2,type="contour")
R> plot(child.reptest2,type="persp",theta=-35,phi=40)
```

which produces the following output:

```
Test the equality of surfaces based on L2 distance

Comparing 2 semiparametric regression surfaces
Penalized semiparametric regression mixed model is used for surface fitting.
```

```
Wide-bootstrap algorithm is applied to obtain the null distribution.

Null hypothesis: there is no difference between the 2 surfaces.
T =  10.88     p-value =  0.0995
```

Plots are shown in Figure 7. The option `type=plotly.persp` generates an interactive 3-D plot using the R package `plotly`.

[Figure 7 about here.]

C.3 The R Shiny Interface

To enhance the usability of the testing methods, we created an interactive R Shiny interface for the `gamm4.test` package. This interface allows analysts that do not use R to access the testing procedure through a web link. The interface can be access at https://heather.shinyapps.io/shinygamm4/ and a youtube tutorial is available at https://youtu.be/SHqaZXSLaMw.

To illustrate, we used the `outchild` data from the "gamm4.test" package as an example.

We first put the observed data in the pre-specified format.

```
colnumes(childcur) <- c("grp", "y", "age")
```

Then click "Enter Data" to upload the dataset. To compare curves and plot the estimated regression functions, click "Test summary and plots".

For surface comparison, we first put the raw data in the specified format.

```
R> colnumes(childsurf) <- c("grp", "y", "x1", "x2")
```

We then compare surfaces by clicking "Enter Data" and "Test summary and plots".

References

[1] Kim S, Zeng D, and Taylor J. Joint Partially Linear Model for Longitudinal Data with Informative Drop-Outs. *Biometrics* 2017; 73: 72-82.

[2] Shen X and Wong WH. Convergence rate of sieve estimates. *The Annals of Statistics* 1994; 22: 580-615.

[3] Van der Vaart AW and Wellner JA. Weak convergence. *In Weak convergence and empirical processes.* Springer, New York, NY, 1996. 16-28p.

[4] Van der Vaart AW. *Asymptotic statistics (Vol. 3).* Cambridge university press, 2000.

Table 1: Two group curve comparison: Power and Type I error rates.

d	(n_1, n_2)	(σ_1, σ_2)	$T_{B-spline}$	$T_{P-spline}$	T_4	T_2	T_1
0	(125, 125)	(0.2, 0.15)	0.051	0.049	0.070	0.043	0.060
	(216, 216)	(0.2, 0.15)	0.048	0.057	0.071	0.046	0.067
	(512, 512)	(0.2, 0.15)	0.055	0.060	0.066	0.049	0.063
	(125, 125)	(0.25, 0.2)	0.049	0.051	0.065	0.043	0.060
	(216, 216)	(0.25, 0.2)	0.051	0.052	0.071	0.058	0.072
	(512, 512)	(0.25, 0.2)	0.060	0.047	0.065	0.057	0.066
1	(125, 125)	(0.2, 0.15)	0.416	0.349	0.406	0.309	0.415
	(216, 216)	(0.2, 0.15)	0.688	0.621	0.655	0.529	0.666
	(512, 512)	(0.2, 0.15)	0.969	0.967	0.973	0.926	0.972
	(125, 125)	(0.25, 0.2)	0.274	0.226	0.321	0.208	0.302
	(216, 216)	(0.25, 0.2)	0.434	0.379	0.446	0.353	0.469
	(512, 512)	(0.25, 0.2)	0.824	0.802	0.850	0.735	0.848
2	(125, 125)	(0.2, 0.15)	0.974	0.941	0.956	0.920	0.958
	(216, 216)	(0.2, 0.15)	1.000	1.000	1.000	0.995	1.000
	(512, 512)	(0.2, 0.15)	1.000	1.000	1.000	1.000	1.000
	(125, 125)	(0.25, 0.2)	0.844	0.764	0.830	0.725	0.821
	(216, 216)	(0.25, 0.2)	0.985	0.970	0.982	0.952	0.983
	(512, 512)	(0.25, 0.2)	1.000	1.000	1.000	1.000	1.000
3	(125, 125)	(0.2, 0.15)	1.000	1.000	0.999	1.000	0.999
	(216, 216)	(0.2, 0.15)	1.000	1.000	1.000	1.000	1.000
	(512, 512)	(0.2, 0.15)	1.000	1.000	1.000	1.000	1.000
	(125, 125)	(0.25, 0.2)	0.995	0.990	0.992	0.981	0.995
	(216, 216)	(0.25, 0.2)	1.000	1.000	1.000	1.000	1.000
	(512, 512)	(0.25, 0.2)	1.000	1.000	1.000	1.000	1.000

Table 2: Three-group curve comparison: Power and Type I error rates.

(d_1, d_2)	(n_1, n_2, n_3)	$T_{B-spline}$
(0, 0)	(100, 50, 75)	0.044
(0, 1)	(100, 50, 75)	0.074
(0, 2)	(100, 50, 75)	0.166
(0, 3)	(100, 50, 75)	0.356
(1, 1)	(100, 50, 75)	0.134
(1, 2)	(100, 50, 75)	0.296
(1, 3)	(100, 50, 75)	0.512
(2, 2)	(100, 50, 75)	0.482
(2, 3)	(100, 50, 75)	0.698
(3, 3)	(100, 50, 75)	0.864
(0, 0)	(200, 100, 150)	0.054
(0, 1)	(200, 100, 150)	0.110
(0, 2)	(200, 100, 150)	0.346
(0, 3)	(200, 100, 150)	0.760
(1, 1)	(200, 100, 150)	0.256
(1, 2)	(200, 100, 150)	0.618
(1, 3)	(200, 100, 150)	0.906
(2, 2)	(200, 100, 150)	0.888
(2, 3)	(200, 100, 150)	0.986
(3, 3)	(200, 100, 150)	1.000
(0, 0)	(300, 150, 225)	0.064
(0, 1)	(300, 150, 225)	0.148
(0, 2)	(300, 150, 225)	0.494
(0, 3)	(300, 150, 225)	0.894
(1, 1)	(300, 150, 225)	0.402
(1, 2)	(300, 150, 225)	0.790
(1, 3)	(300, 150, 225)	0.970
(2, 2)	(300, 150, 225)	0.960
(2, 3)	(300, 150, 225)	1.000
(3, 3)	(300, 150, 225)	1.000

Table 3: Changes of power with sample size.

(n_1, n_2)	(σ_1, σ_2)	$T_{B-spline}$	$T_{P-spline}$
(1000, 1000)	(0.2, 0.15)	1.000	1.000
(729, 729)	(0.2, 0.15)	0.995	0.993
(512, 512)	(0.2, 0.15)	0.978	0.968
(343, 343)	(0.2, 0.15)	0.871	0.837
(216, 216)	(0.2, 0.15)	0.667	0.613
(125, 125)	(0.2, 0.15)	0.408	0.345

Table 4: Surface comparison: Type I error rates.

Func	(n_1, n_2)	(σ_1, σ_2)	**TP-Spline**	**TP-Spline$_+$**	**TP-Spline$_-$**	**TE-Spline**	**TP-Spline.p**	**TE-Spline.p**	T_4	T_1	T_3	T_2
a	(125, 125)	(0.5, 0.3)	0.088	0.10	0.088	0.10	0.106	0.094	0.09	0.038	0.15	0.002
	(216, 216)	(0.5, 0.3)	0.076	0.080	0.094	0.07	0.07	0.068	0.088	0.02	0.110	0.012
	(512, 512)	(0.5, 0.3)	0.074	0.068	0.058	0.080	0.046	0.06	0.064	0.026	0.064	0.012
	(125, 125)	(0.6, 0.4)	0.098	0.086	0.09	0.070	0.080	0.080	0.08	0.02	0.114	0.002
	(216, 216)	(0.6, 0.4)	0.078	0.080	0.100	0.07	0.078	0.068	0.078	0.02	0.120	0.010
	(512, 512)	(0.6, 0.4)	0.05	0.06	0.07	0.060	0.06	0.056	0.056	0.016	0.08	0.012
b	(125, 125)	(0.6, 0.4)	0.06	0.05	0.044	0.068	0.048	0.060	0.058	0.066	0.060	0.052
	(216, 216)	(0.6, 0.4)	0.044	0.038	0.038	0.034	0.044	0.038	0.070	0.080	0.088	0.026
	(512, 512)	(0.6, 0.4)	0.054	0.048	0.05	0.040	0.05	0.05	0.080	0.07	0.076	0.038
	(125, 125)	(0.8, 0.6)	0.060	0.058	0.06	0.060	0.05	0.064	0.074	0.080	0.078	0.068
	(216, 216)	(0.8, 0.6)	0.066	0.07	0.064	0.066	0.05	0.05	0.074	0.076	0.084	0.038
	(512, 512)	(0.8, 0.6)	0.064	0.058	0.06	0.050	0.058	0.050	0.060	0.060	0.06	0.026
c	(125, 125)	(0.8, 0.6)	0.038	0.040	0.04	0.056	0.048	0.040	0.040	0.046	0.04	0.056
	(216, 216)	(0.8, 0.6)	0.05	0.05	0.054	0.048	0.046	0.03	0.046	0.050	0.048	0.056
	(512, 512)	(0.8, 0.6)	0.046	0.050	0.04	0.024	0.04	0.040	0.05	0.044	0.046	0.054
	(125, 125)	(1, 0.8)	0.046	0.044	0.04	0.040	0.046	0.044	0.048	0.05	0.060	0.056
	(216, 216)	(1, 0.8)	0.066	0.05	0.054	0.050	0.058	0.038	0.050	0.046	0.050	0.052
	(512, 512)	(1, 0.8)	0.04	0.044	0.044	0.034	0.038	0.040	0.054	0.064	0.054	0.068

Table 5: Surface comparison (Cont.): Power.

Func	(n_1, n_2)	(σ_1, σ_2)	TP-Spline	TP-Spline$_+$	TP-Spline$_-$	TE-Spline	TP-Spline.p	TE-Spline.p	T_4	T_1	T_3	T_2
d	(125, 125)	(0.5, 0.3)	0.940	0.944	0.952	0.806	0.944	0.826	0.870	0.784	0.940	0.592
	(216, 216)	(0.5, 0.3)	0.998	0.998	1.000	0.984	0.998	0.990	0.988	0.976	0.994	0.960
	(512, 512)	(0.5, 0.3)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	(125, 125)	(0.6, 0.4)	0.792	0.780	0.786	0.656	0.804	0.660	0.742	0.604	0.834	0.400
	(216, 216)	(0.6, 0.4)	0.968	0.958	0.968	0.930	0.972	0.942	0.942	0.902	0.964	0.800
	(512, 512)	(0.6, 0.4)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
e	(125, 125)	(0.6, 0.4)	0.960	0.962	0.956	0.932	0.968	0.928	0.894	0.918	0.932	0.934
	(216, 216)	(0.6, 0.4)	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.998	0.998	0.998
	(512, 512)	(0.6, 0.4)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	(125, 125)	(0.8, 0.6)	0.730	0.724	0.716	0.618	0.730	0.640	0.650	0.646	0.662	0.706
	(216, 216)	(0.8, 0.6)	0.950	0.946	0.944	0.908	0.944	0.936	0.874	0.880	0.890	0.906
	(512, 512)	(0.8, 0.6)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	1.000	1.000
f	(125, 125)	(0.8, 0.6)	0.742	0.742	0.734	0.632	0.736	0.650	0.696	0.684	0.698	0.734
	(216, 216)	(0.8, 0.6)	0.960	0.958	0.962	0.908	0.960	0.926	0.860	0.882	0.894	0.960
	(512, 512)	(0.8, 0.6)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	(125, 125)	(1, 0.8)	0.478	0.500	0.490	0.386	0.482	0.388	0.484	0.502	0.500	0.490
	(216, 216)	(1, 0.8)	0.772	0.784	0.784	0.680	0.776	0.736	0.680	0.714	0.722	0.796
	(512, 512)	(1, 0.8)	0.998	0.998	0.998	0.992	0.998	0.996	0.972	0.992	0.990	1.000

Table 6: Type I error rates and power of comparisons with correlated data.

d	(n_1, n_2)	σ_1'	σ_2'	σ_1	σ_2	$T_{splinecorr}$
0	(50,60)	0.20	0.15	0.04	0.05	0.057
	(50,60)	0.20	0.15	0.10	0.12	0.058
	(50,60)	0.25	0.20	0.10	0.12	0.062
	(100,120)	0.20	0.15	0.04	0.05	0.066
	(100,120)	0.20	0.15	0.10	0.12	0.072
	(100,120)	0.25	0.20	0.10	0.12	0.068
	(150,160)	0.20	0.15	0.04	0.05	0.048
	(150,160)	0.20	0.15	0.10	0.12	0.041
	(150,160)	0.25	0.20	0.10	0.12	0.053
1	(50,60)	0.20	0.15	0.04	0.05	0.426
	(50,60)	0.20	0.15	0.10	0.12	0.383
	(50,60)	0.25	0.20	0.10	0.12	0.212
	(100,120)	0.20	0.15	0.04	0.05	0.968
	(100,120)	0.20	0.15	0.10	0.12	0.784
	(100,120)	0.25	0.20	0.10	0.12	0.516
	(150,160)	0.20	0.15	0.04	0.05	1.000
	(150,160)	0.20	0.15	0.10	0.12	0.941
	(150,160)	0.25	0.20	0.10	0.12	0.804
2	(50,60)	0.20	0.15	0.04	0.05	1.000
	(50,60)	0.20	0.15	0.10	0.12	0.994
	(50,60)	0.25	0.20	0.10	0.12	0.956
	(100,120)	0.20	0.15	0.04	0.05	1.000
	(100,120)	0.20	0.15	0.10	0.12	1.000
	(100,120)	0.25	0.20	0.10	0.12	1.000
	(150,160)	0.20	0.15	0.04	0.05	1.000
	(150,160)	0.20	0.15	0.10	0.12	1.000
	(150,160)	0.25	0.20	0.10	0.12	1.000

Table 7: Comparison of the rejection probabilities between the proposed method and Zhang et al's scaled χ^2 testing method

			$\mathbf{x_2} = \mathbf{x_1}$		$\mathbf{x_2} = \mathbf{x_1} + U(0, 0.05)$		Random $\mathbf{x_1}, \mathbf{x_2} \sim U(0,1)$	
d	(n_1, n_2)	$(\sigma'_1, \sigma'_2, \sigma_1, \sigma_2)$	Scaled χ^2	$T_{splinecorr}$	Scaled χ^2	$T_{splinecorr}$	Scaled χ^2	$T_{splinecorr}$
0	(50,50)	(0.20, 0.15, 0.04, 0.05)	0.000	0.045	0.010	0.075	0	0.070
	(50,50)	(0.25, 0.20, 0.10, 0.12)	0.005	0.070	0.005	0.070	0	0.065
	(100,100)	(0.20, 0.15, 0.04, 0.05)	0.005	0.035	0.025	0.070	0	0.065
	(100,100)	(0.25, 0.20, 0.10, 0.12)	0.005	0.050	0.010	0.035	0	0.030
1	(50,50)	(0.20, 0.15, 0.04, 0.05)	0.035	0.360	0.005	0.330	0	0.405
	(50,50)	(0.25, 0.20, 0.10, 0.12)	0.005	0.180	0.000	0.165	0	0.185
	(100,100)	(0.20, 0.15, 0.04, 0.05)	0.250	0.955	0.000	0.900	0	0.965
	(100,100)	(0.25, 0.20, 0.10, 0.12)	0.070	0.535	0.000	0.400	0	0.430
2	(50,50)	(0.20, 0.15, 0.04, 0.05)	0.765	1.000	0.020	1.000	0	1.000
	(50,50)	(0.25, 0.20, 0.10, 0.12)	0.270	0.890	0.010	0.885	0	0.935
	(100,100)	(0.20, 0.15, 0.04, 0.05)	1.000	1.000	0.025	1.000	0	1.000
	(100,100)	(0.25, 0.20, 0.10, 0.12)	0.840	1.000	0.005	1.000	0	1.000
3	(50,50)	(0.20, 0.15, 0.04, 0.05)	1.000	1.000	0.460	1.000	0	1.000
	(50,50)	(0.25, 0.20, 0.10, 0.12)	0.915	1.000	0.185	1.000	0	1.000
	(100,100)	(0.20, 0.15, 0.04, 0.05)	1.000	1.000	0.905	1.000	0	1.000
	(100,100)	(0.25, 0.20, 0.10, 0.12)	1.000	1.000	0.565	1.000	0	1.000

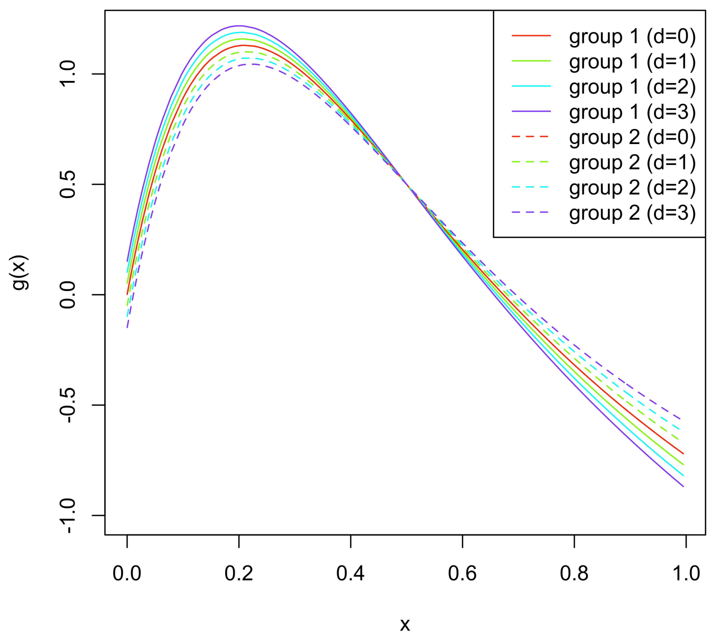

Figure 1: Functions $g_{1d}(x)$ and $g_{2d}(x)$ with $d = 0, 1, 2, 3$ used in the simulation studies

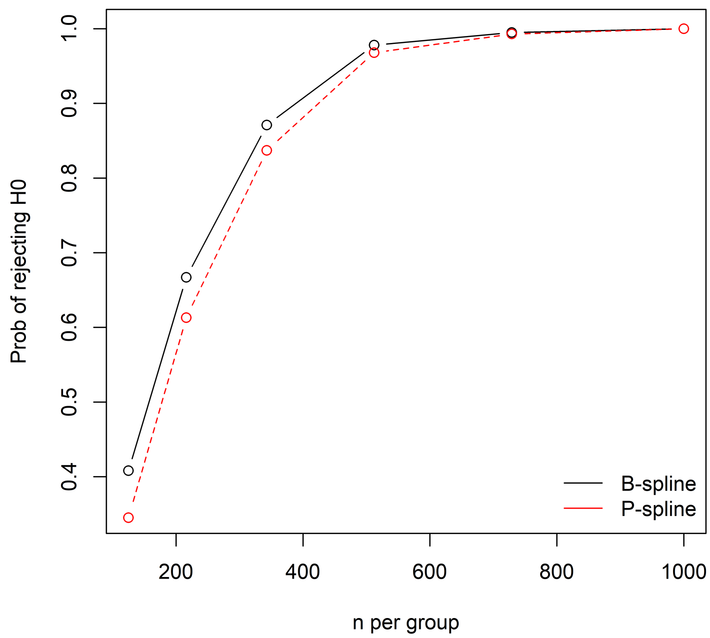

Figure 2: Probability of rejecting H_0 with $d = 1$, $(\sigma_1, \sigma_2) = (0.2, 0.15)$

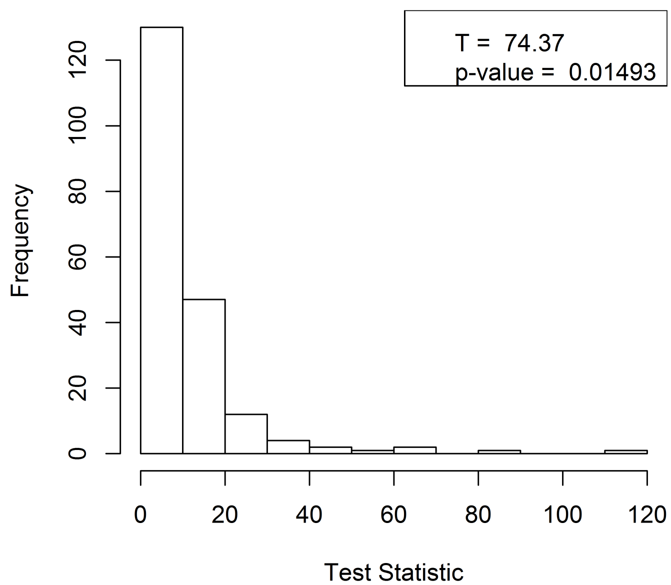

Figure 3: Empirical distribution of the test statistic under the null hypothesis

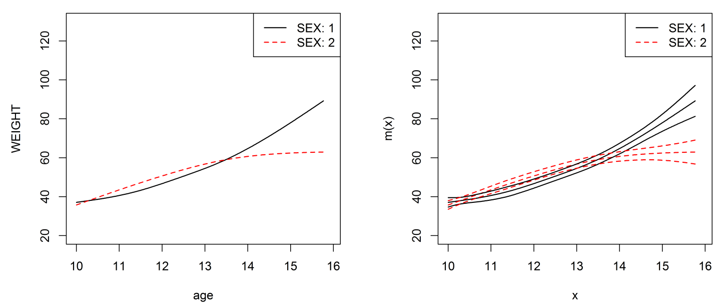

Figure 4: Estimated age effects on weight and associated pointwise 95% CIs in boys and girls

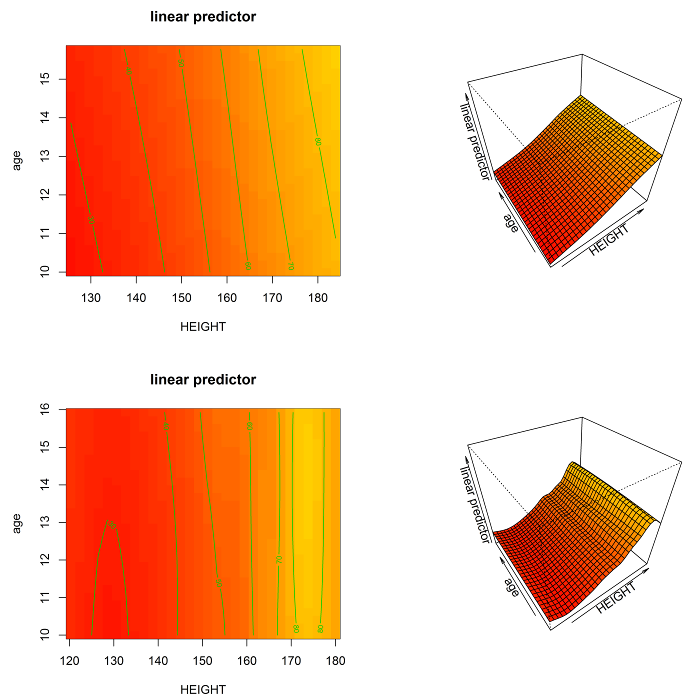

Figure 5: Estimated contour and 3D plots of height and age effects on weight by gender

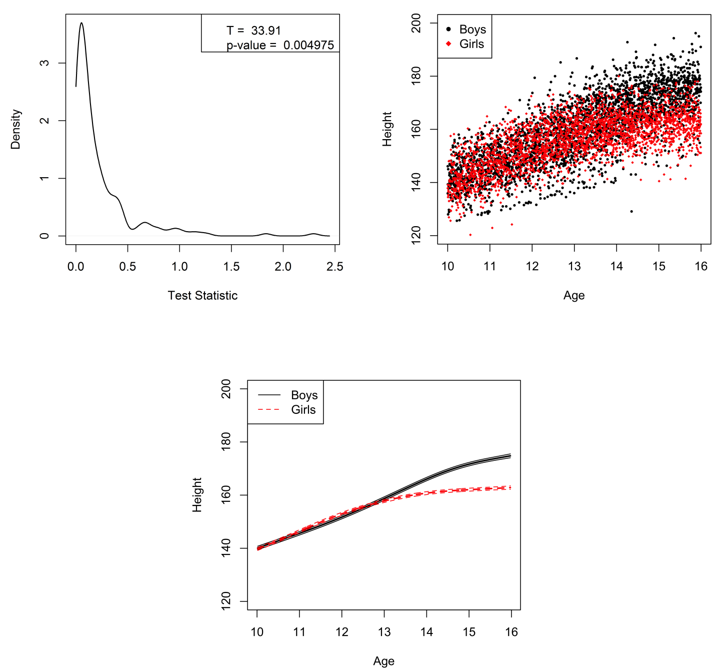

Figure 6: Empirical distribution of the test statistic under the null hypothesis and height over age by gender; height-for-age scatter plot; pointwise 95% CI by gender

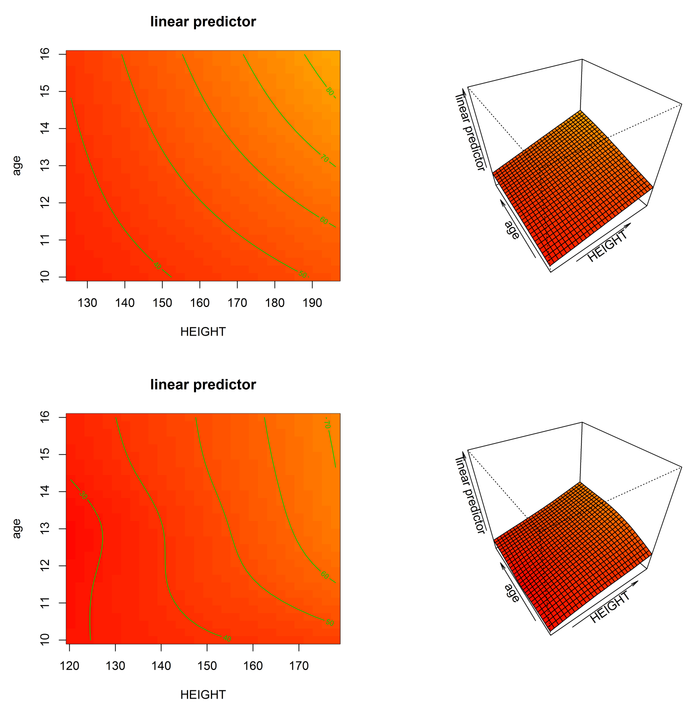

Figure 7: Estimated contour and 3D plots of weight on height and age by sex using correlated data