Scientific Software International

Nonlinear curve for predicting height

Contents

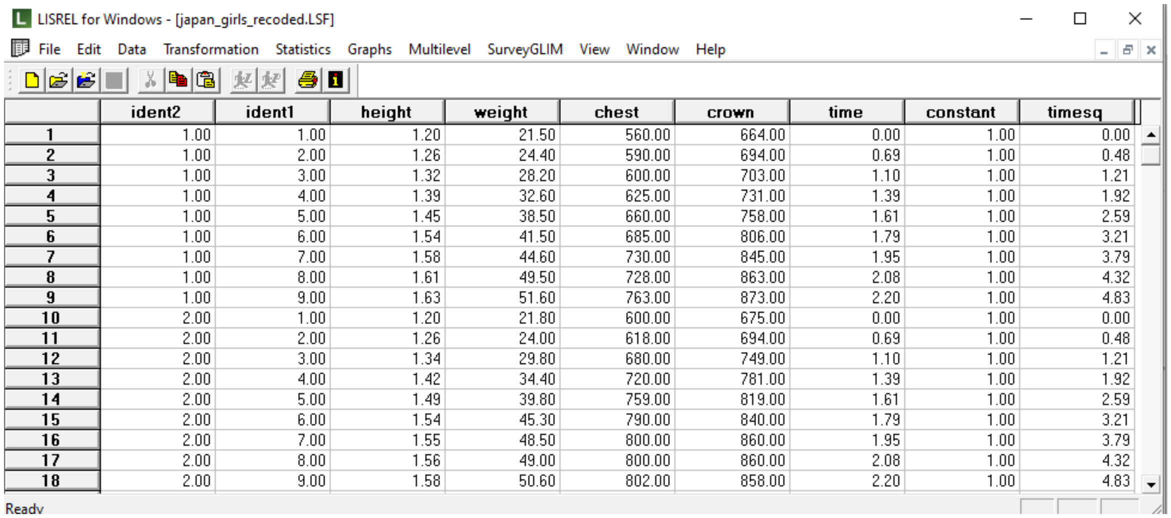
1. Introduction

In this example we consider the fitting of a nonlinear models to the height measurements of Japanese females on 9 occasions. Data are from a follow-up survey of growth of physical constitution of girls for nine years (in the prefecture of Chiba) was conducted by a lecturer Hiroaki Terao at the Faculty of Medicine, Teikyo University. Here we use only the girls' data (6 years old to 14 years old). A full description and the data are contained in: Hayashi, Chikio, & Hayashi, Fumi (1982). A new algorithm to solve PARAFAC-model. *Behaviormetrika, 11*, 49-60.

Data are given in **japan_girls_recoded.lsf** and the data for the first two females are shown below.

The variables of interest are:

- CIdent2: the identifier of each girl
- Height: the height of the individual at that point in time (in meters)
- Weight: the height of the individual at that point in time (in kilogram)
- Time: the natural logarithm of the measurement occasion (occasion = 1, 2, …, 9)
- Timesq: the squared value of Time
- Time_o: the original measurement occasion.



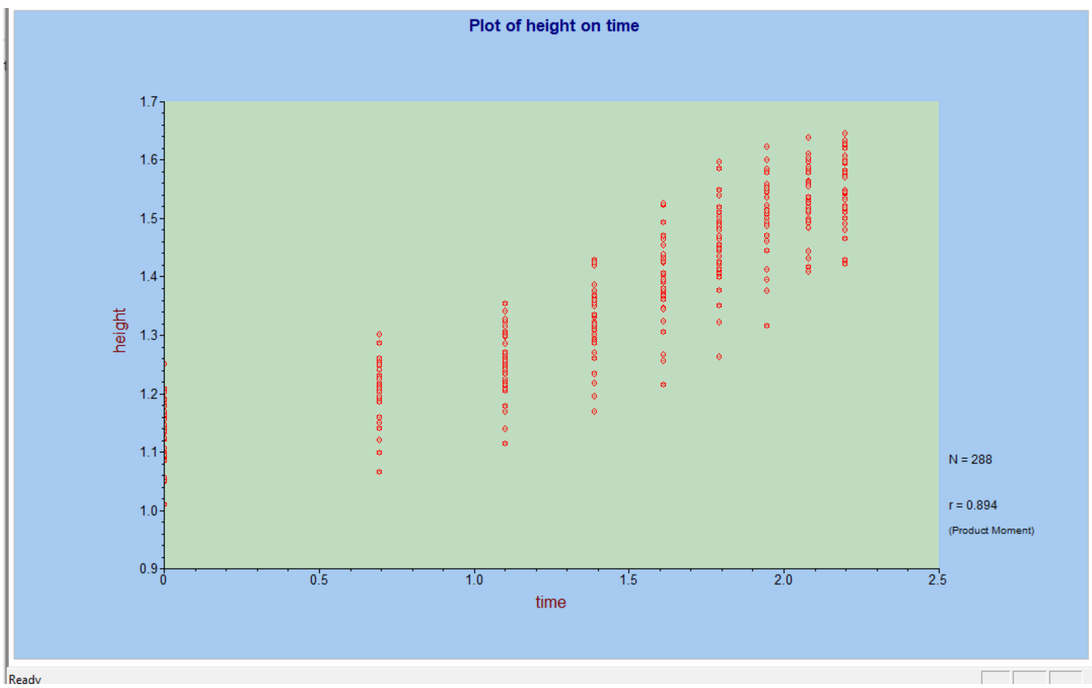
LISREL for Windows - [japan_girls_recoded.LSF]

File Edit Data Transformation Statistics Graphs Multilevel SurveyGLIM View Window Help

	ident2	ident1	height	weight	chest	crown	time	constant	timesq
1	1.00	1.00	1.20	21.50	560.00	664.00	0.00	1.00	0.00
2	1.00	2.00	1.26	24.40	590.00	694.00	0.69	1.00	0.48
3	1.00	3.00	1.32	28.20	600.00	703.00	1.10	1.00	1.21
4	1.00	4.00	1.39	32.60	625.00	731.00	1.39	1.00	1.92
5	1.00	5.00	1.45	38.50	660.00	758.00	1.61	1.00	2.59
6	1.00	6.00	1.54	41.50	685.00	806.00	1.79	1.00	3.21
7	1.00	7.00	1.58	44.60	730.00	845.00	1.95	1.00	3.79
8	1.00	8.00	1.61	49.50	728.00	863.00	2.08	1.00	4.32
9	1.00	9.00	1.63	51.60	763.00	873.00	2.20	1.00	4.83
10	2.00	1.00	1.20	21.80	600.00	675.00	0.00	1.00	0.00
11	2.00	2.00	1.26	24.00	618.00	694.00	0.69	1.00	0.48
12	2.00	3.00	1.34	29.80	680.00	749.00	1.10	1.00	1.21
13	2.00	4.00	1.42	34.40	720.00	781.00	1.39	1.00	1.92
14	2.00	5.00	1.49	39.80	759.00	819.00	1.61	1.00	2.59
15	2.00	6.00	1.54	45.30	790.00	840.00	1.79	1.00	3.21
16	2.00	7.00	1.55	48.50	800.00	860.00	1.95	1.00	3.79
17	2.00	8.00	1.56	49.00	800.00	860.00	2.08	1.00	4.32
18	2.00	9.00	1.58	50.60	802.00	858.00	2.20	1.00	4.83

Ready

A scatterplot of the height measurements over time is shown below. The relationship between height and time curve shows a nonlinear trend.



2. Logistic curve

We start by fitting a logistic curve to these data. The model can be formulated as

$$y = b_1 / (1 + s * \exp(b_2 - b_3 * time) + e$$

The level-2 model is:

$$b_1 = \beta_1 + u_1$$
$$b_2 = \beta_2 + u_2$$
$$b_3 = \beta_3 + u_3$$

The syntax file for this model is shown in the syntax file **Japan_logistic_rec.prl**. The variable ident2 is used as level-2 identifier (ID2).

Japan_logistic_rec.prl

```
!---------------------------------------------------------------
! The file japan_girls.psf contains body measurements of
! 32 girls.
!---------------------------------------------------------------
OPTIONS METHOD = ML CONVERGE = 0.00010 MAXITER =100 QUADPTS =30;
TITLE = Body measurements Japanese girls;
SY=Japan_girls_recoded.lsf;
ID1 = ident1;
ID2 = ident2;
RESPONSE = height;
FIXED = time;
MODEL = Logistic;
```

The ML solution is as follows.

Coefficients	Beta	Std.Err.	Z-value	P > |z|
b1	45.78778	0.58570	78.17670	0.00000
b2	3.71584	0.01596	232.81652	0.00000
b3	0.16466	0.00194	84.91557	0.00000

Variance estimate	Level 1	Std.Err.	Z-value	P > |z|
Sigma**2	0.00106	0.00009	12.08830	0.00000

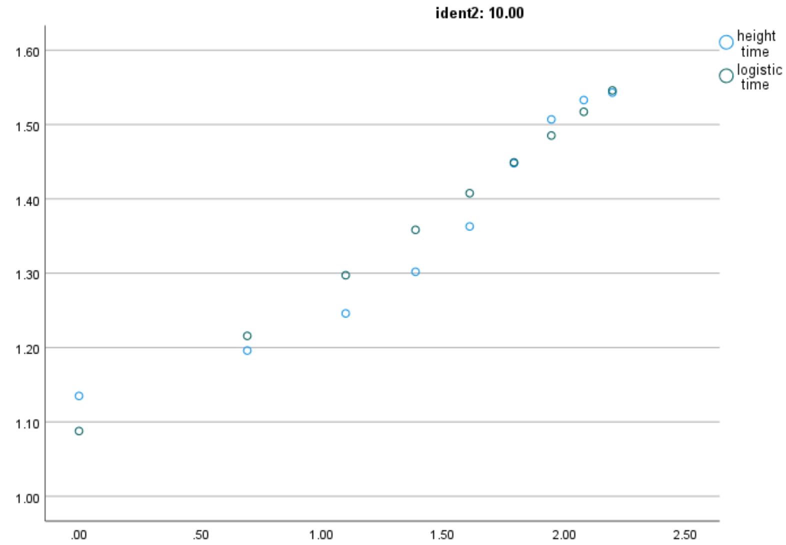
Covariances	Level 2	Std.Err.	Z-value	P > |z|
u1,u1	0.65036	2.86899	0.22668	0.82067
u2,u1	0.02668	0.08484	0.31441	0.75321
u2,u2	20.85103	0.00262	7946.27437	0.00000
u3,u1	-0.00033	0.00929	-0.03568	0.97153
u3,u2	0.91502	0.00025	3681.65742	0.00000
u3,u3	0.04016	0.00004	945.24514	0.00000

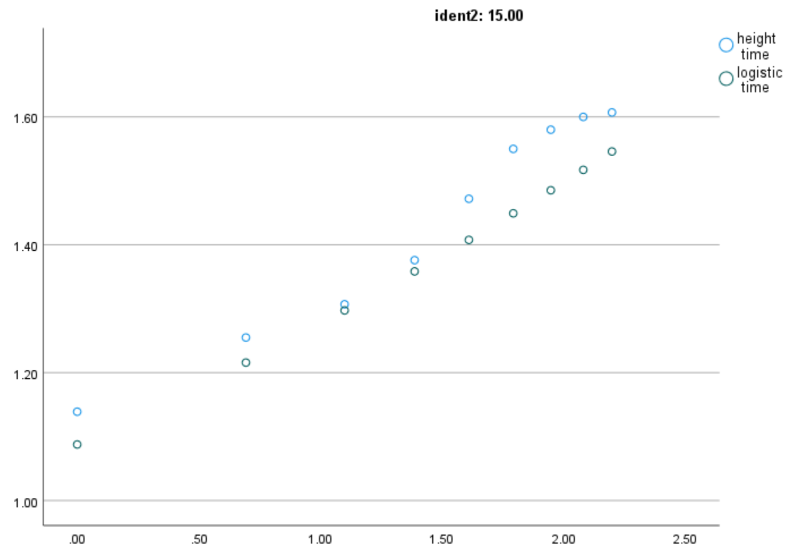
The estimated beta coefficients are all highly significant. We note that there the estimated $\mathrm{var}(u_1, u_1)$ is not significant but that both $\mathrm{var}(u_2, u_2)$ and $\mathrm{var}(u_3, u_3)$ are. Only one of the three covariances is significant. In this case, where function values increase montonically over tie, *s* has a value of 1.

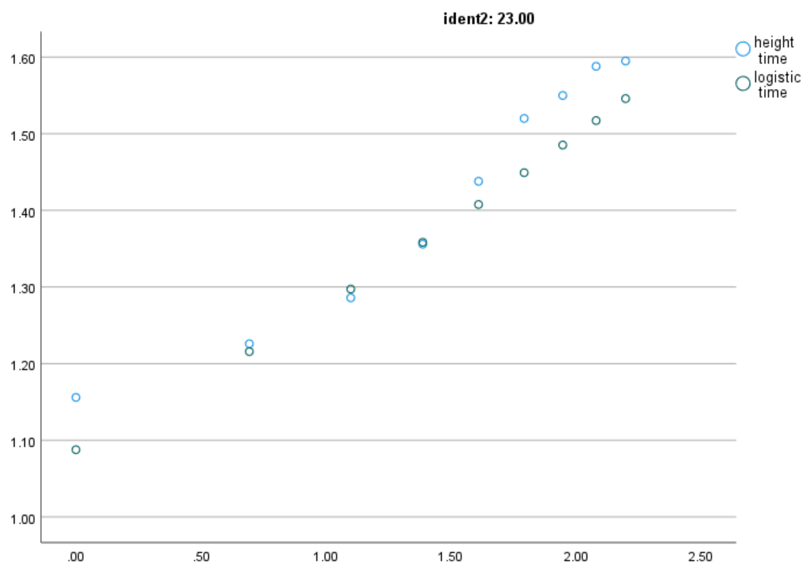
The average expected height can be estimated using the formula

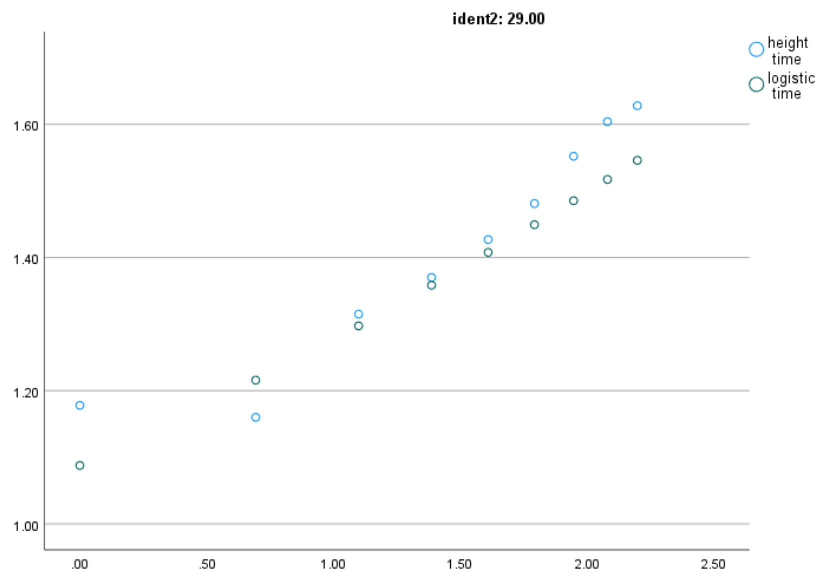
$$\text{Predicted}(y) = 45.78778 / (1 + \exp(3.71584 - 0.16466 * time)$$

When we look at scatterplots of observed and predicted height for the same selection, we note that the fitted curve does not follow the general shape of the observed height curve.









We conclude that the model does not describe the data well and continue to explore other potential models.

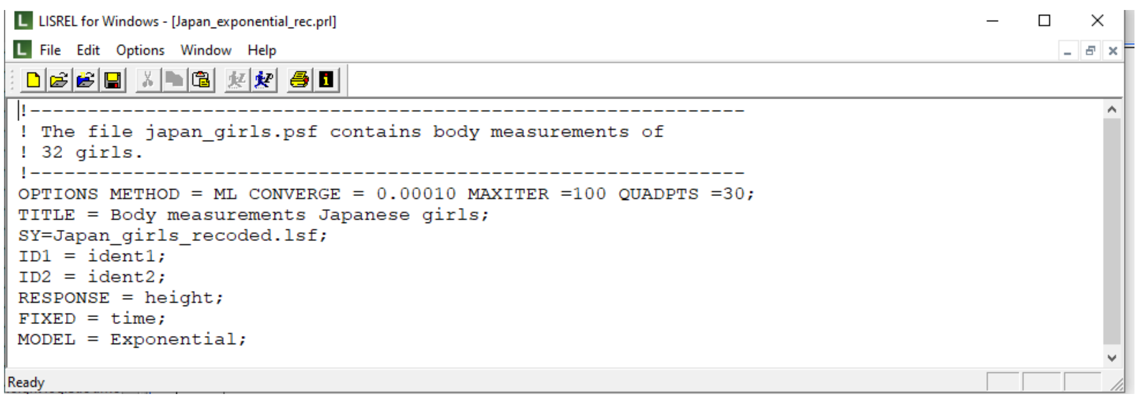
3. Exponential curve

Another potential model is the exponential model. The model can be formulated as

$$y = b_1 * \exp(-b_2 * time) + e$$

and is more parsimonious as it utilizes fewer parameters: one fewer fixed effect and 3 fewer variance-covariance elements need to be estimated under this model.

The model is given in **Japan_exponential_rec.prl**:



The output obtained for this model is:

```
--------------------------------------------------------------------------
  Coefficients          Beta          Std.Err.       Z-value        P > |z|
```

b1	1.08829	0.00689	157.86714	0.00000
b2	-0.15982	0.00189	-84.61839	0.00000

Variance estimate	Level 1	Std.Err.	Z-value	P > |z|
Sigma**2	0.00132	0.00008	17.03199	0.00000

Covariances	Level 2	Std.Err.	Z-value	P > |z|
u1,u1	0.00235	0.00054	4.35461	0.00001
u2,u1	0.00010	0.00012	0.82853	0.40737
u2,u2	0.00001	0.00004	0.25495	0.79876

The estimated fixed coefficients are again highly significant, but only one random effect ($\mathrm{var}(u_1, u_1)$) is statistically significant.

The average expected height under the current model can be estimated using the formula

$$\text{Predicted}(y) = 1.08829 * (\exp(0.15892 * time)$$

When we evaluate the results graphically and by examining the descriptive statistics of the observed and predicted height for the same set of girls, we find no change and thus no improvement. As a result, the best thing that can be said about the exponential curve fitted here is that it delivers the same fit as the logistic curve but with fewer parameters.

4. Double exponential curve

Another potential model is the exponential model. The model can be formulated as

$$y = b_1 * \exp(-b_2 * time) + c_1 * \exp(-c_2 * time) + e$$

and is more parsimonious as it utilizes fewer parameters: one fewer fixed effect and 3 fewer variance-covariance elements need to be estimated under this model.

The model is given in **Japan_d_exponential_rec.prl**:

```
Japan_d_exponential_rec.prl
!---------------------------------------------------------------
! The file japan_girls.psf contains body measurements of
! 32 girls.
!---------------------------------------------------------------
OPTIONS METHOD = ML CONVERGE = 0.00010 MAXITER =100 QUADPTS =30;
TITLE = Body measurements Japanese girls;
SY=Japan_girls_recoded.lsf;
ID1 = ident1;
ID2 = ident2;
RESPONSE = height;
FIXED = time ;
MODEL = Exponential + Exponential;
```

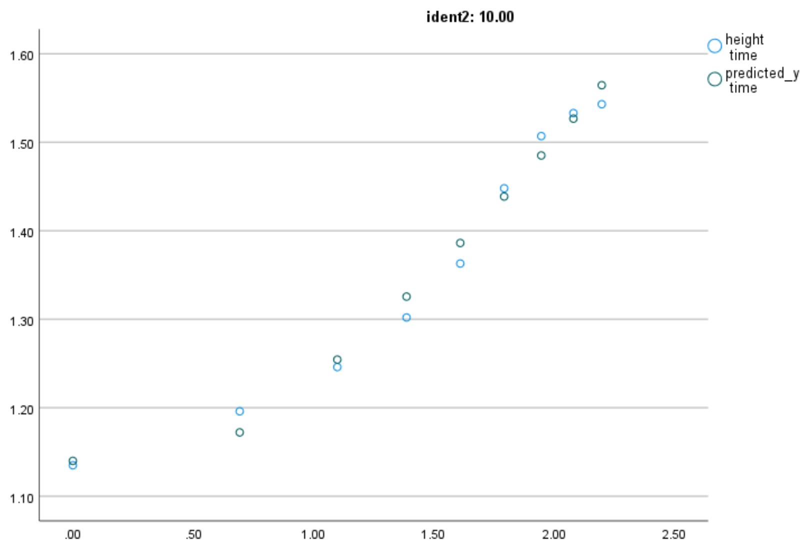
Coefficients	Beta	Std.Err.	Z-value	P > |z|

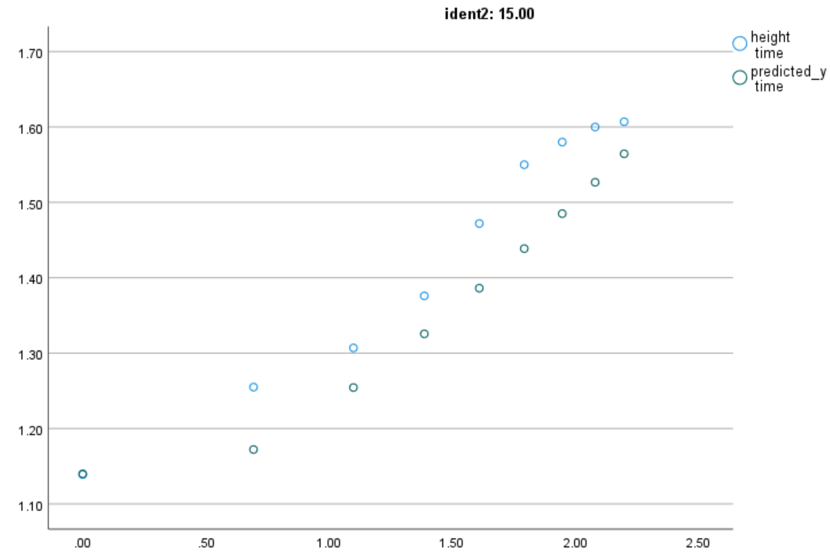
b1	0.15318	0.01143	13.39960	0.00000
b2	2.29554	0.26784	8.57062	0.00000
c1	0.98680	0.01250	78.93836	0.00000
c2	-0.20945	0.00509	-41.16160	0.00000

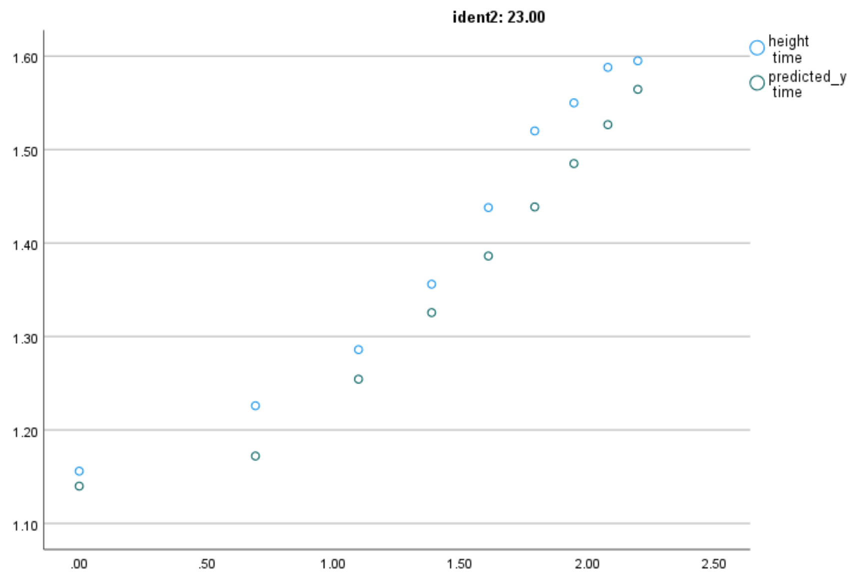
Variance estimate	Level 1	Std.Err.	Z-value	P > |z|
Sigma**2	0.00047	0.00003	17.39985	0.00000

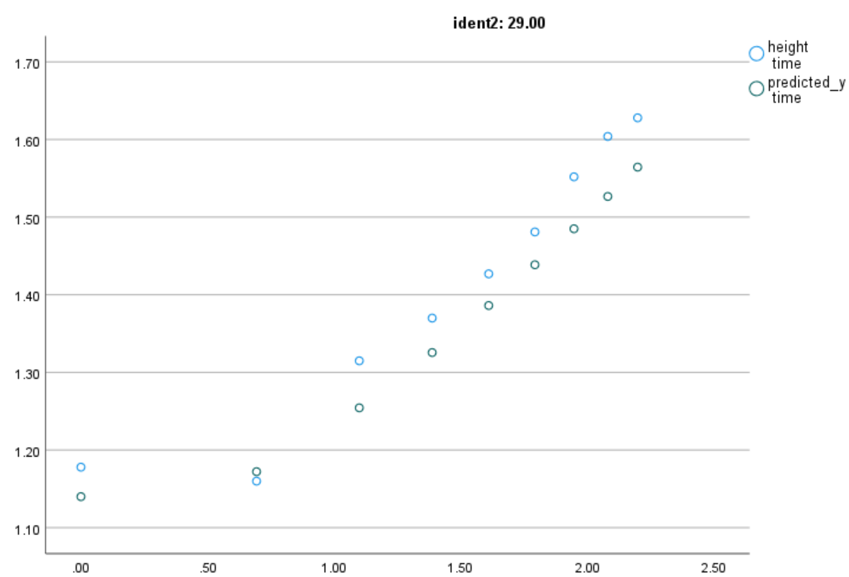
Covariances	Level 2	Std.Err.	Z-value	P > |z|
u1,u1	0.00283	0.00137	2.05748	0.03964
u2,u1	-0.06541	0.02988	-2.18877	0.02861
u2,u2	1.54409	0.76317	2.02326	0.04305
u3,u1	-0.00318	0.00141	-2.25635	0.02405
u3,u2	-0.09861	0.03151	-3.12929	0.00175
u3,u3	0.96664	0.00169	571.28645	0.00000
u4,u1	-0.00128	0.00060	-2.12889	0.03326
u4,u2	0.02721	0.01303	2.08897	0.03671
u4,u3	0.01490	0.00066	22.61477	0.00000
u4,u4	0.00079	0.00029	2.73542	0.00623

The observed and fitted values for a few girls are shown below. We see that this model describes the observed data better than either of the previous two.




ident2: 15.00
height
time
predicted_y
time
1.70
1.60
1.50
1.40
1.30
1.20
1.10
.00
.50
1.00
1.50
2.00
2.50


ident2: 23.00
height
time
predicted_y
time
1.60
1.50
1.40
1.30
1.20
1.10
.00
.50
1.00
1.50
2.00
2.50


ident2: 29.00
height
time
predicted_y
time
1.70
1.60
1.50
1.40
1.30
1.20
1.10
.00
.50
1.00
1.50
2.00
2.50

While the fit of female 10 has improved judging by the scatterplot above, the same is not true for the other three females. The double exponential model seems to be the best of the three models considered, though it comes at the cost of estimating a total of 14 parameters, compared to 10 parameters for the logistic model and 7 for the exponential model.

5. Conclusion and suggestions for further research

When we calculate the sum of squared residuals for the three models, we obtain the results shown below. It supports the previous conclusion that, of the three models fitted to these data in this example, the double exponential model describes the data best of the three models considered here.

Descriptive Statistics

	N	Minimum	Maximum	Sum	Mean	Std. Deviation
sq_resid (logistic)	288	.00	.04	1.29	.0045	.00647
sq_resid (exponential)	288	.00	.04	1.29	.0045	.00646
sq_resid (double exponential)	288	.00	.03	1.13	.0039	.00553

The models considered here used the natural logarithm of the measurement occasion as predictor. The data file also contains the original measurement occasions. It is left as an exercise to the reader to fit these models to this variable and compare the results. In addition, it might be useful to include the weight of the respondents as a covariate in the model.