

Model checking the polynomial model

Once a model is selected, we should do some model checking to make sure that the residuals are normally distributed and that the random coefficients have a multivariate normal distribution. We will use the model with the fixed quartic month term (month4) and random quadratic month term (month2) to check these assumptions.

First we will check the residuals by plotting them against their predicted values and using a normal plot.

```
PROC MIXED data=a; CLASS type pat;
MODEL virus=age type month month2 month3 month4/ S DDFM=SATTERTH outp=resout
outpm=predout;
RANDOM int month month2/ SUB=pat TYPE=UN solution;
id pat month virus;
ods output solutionr=solut;
run;
```

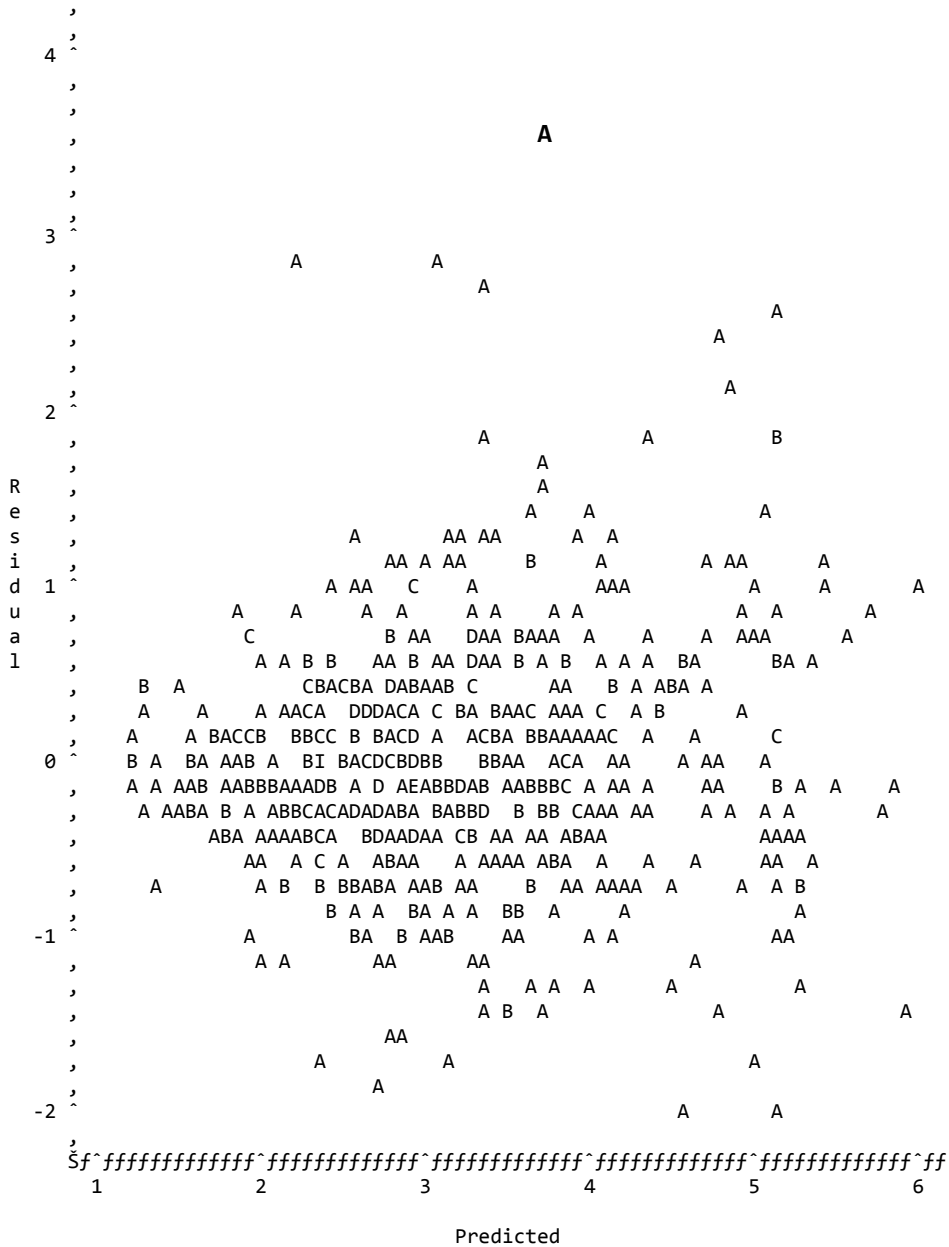
Note the additional options in the model statement, requesting that the residuals and the predicted values be output into two datasets. The ID statement carries the pat month and virus information be included in the output dataset. The ods statement outputs the solutions for the three random effects into the output dataset solut. Following are the first few records of the resout dataset. The residuals give us an indication about how far apart the predicted and actual value are. A large residual value indicates a possible outlier.

pat	month	Resid	Pred	virus
3	0.00	1.46533	4.02467	5.49000
3	0.75	-0.37431	3.83943	3.46512
3	2.00	-0.22146	3.57805	3.35659
3	4.00	-0.33715	3.26738	2.93023
3	7.00	-0.36411	2.99977	2.63566
3	9.00	-0.08344	2.92065	2.83721
3	10.00	-0.44611	2.90347	2.45736

The following code produces the plot of the residuals against their predicted values.

```
proc plot data=resout; plot resid*pred;
title 'residuals against their predicted values';
run;
```

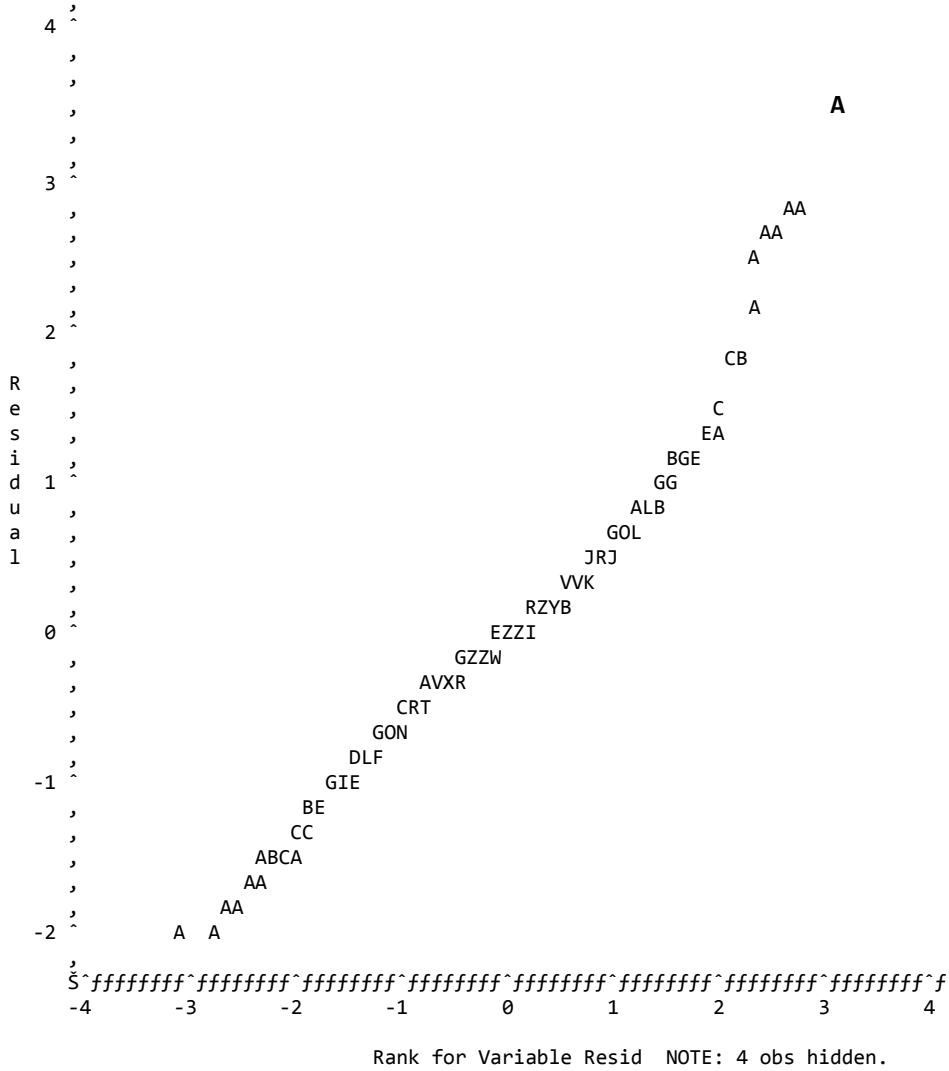
Plot of Resid*Pred. Legend: A = 1 obs, B = 2 obs, etc.



We also want to look at the normal plot, which should be fairly linear. The following SAS code first ranks the residuals and then standardizes them based on their rank. We then plot the residuals against their standardized rank. There appears to be one possible outlier value with a residual greater than 3. Now we look at the normal plot.

```
proc rank data=resout out=norm normal=tukey;
var resid; ranks s_resid; run;
proc plot data=norm;
plot resid*s_resid;
title 'residuals - normal plot'; run;
```

Plot of Resid*s_resid. Legend: A = 1 obs, B = 2 obs, etc.



Both plots indicate that there may be one potential outlier with a residual greater than 3. Following are the records for this patient (67)

Obs	pat	age	agecat	month	month2	month3	month4	virus	type	month5
1	67	8	2	0.00	0.00	0.00	0.00	7.220	AL	0.00
2	67	8	2	1.30	1.69	2.20	2.86	1.472	AL	3.72
3	67	8	2	2.75	7.56	20.80	57.19	1.760	AL	157.27
4	67	8	2	3.00	9.00	27.00	81.00	1.924	AL	243.00
5	67	8	2	4.00	16.00	64.00	256.00	1.704	AL	1024.00
6	67	8	2	5.00	25.00	125.00	625.00	1.572	AL	3125.00
7	67	8	2	6.00	36.00	216.00	1296.00	1.420	AL	7776.00
8	67	8	2	7.50	56.25	421.88	3164.06	1.436	AL	23730.45
9	67	8	2	8.00	64.00	512.00	4096.00	1.700	AL	32768.00

It looks like there may have been a possible recording error at baseline for this patient. If we remove that patient and rerun our selected model we get the following results.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
UN(1,1)	pat	0.5618
UN(2,1)	pat	-0.02430
UN(2,2)	pat	0.01530
UN(3,1)	pat	0.001062
UN(3,2)	pat	-0.00045
UN(3,3)	pat	0.000015
Residual		0.5024

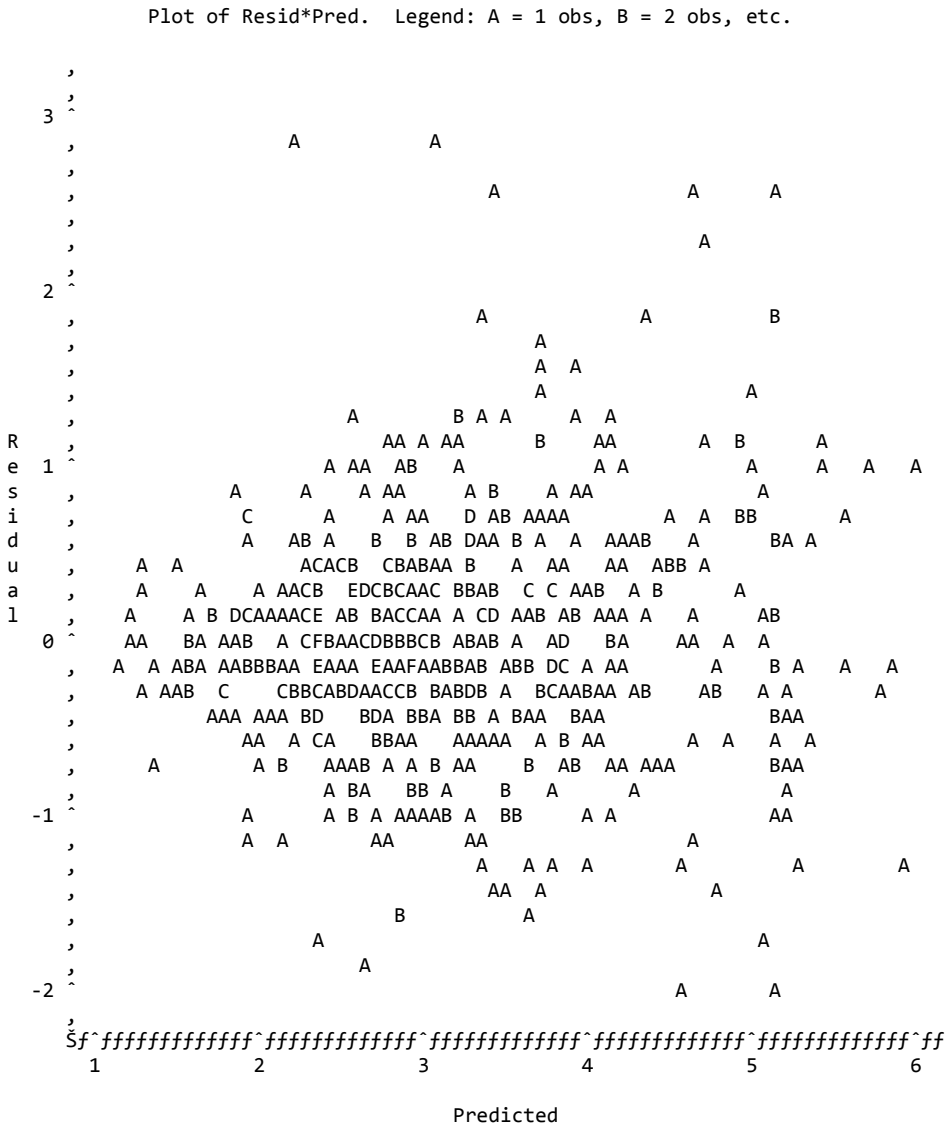
Solution for Fixed Effects

Effect	type	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		3.7615	0.2439	49	15.42	<.0001
age		-0.04355	0.03765	42.9	-1.16	0.2538
type	AL	-0.06574	0.2512	41.6	-0.26	0.7948
type	ST	0
month		-0.1616	0.04550	176	-3.55	0.0005
month2		0.01526	0.005369	497	2.84	0.0047
month3		-0.00055	0.000235	489	-2.34	0.0197
month4		6.429E-6	3.251E-6	549	1.98	0.0485

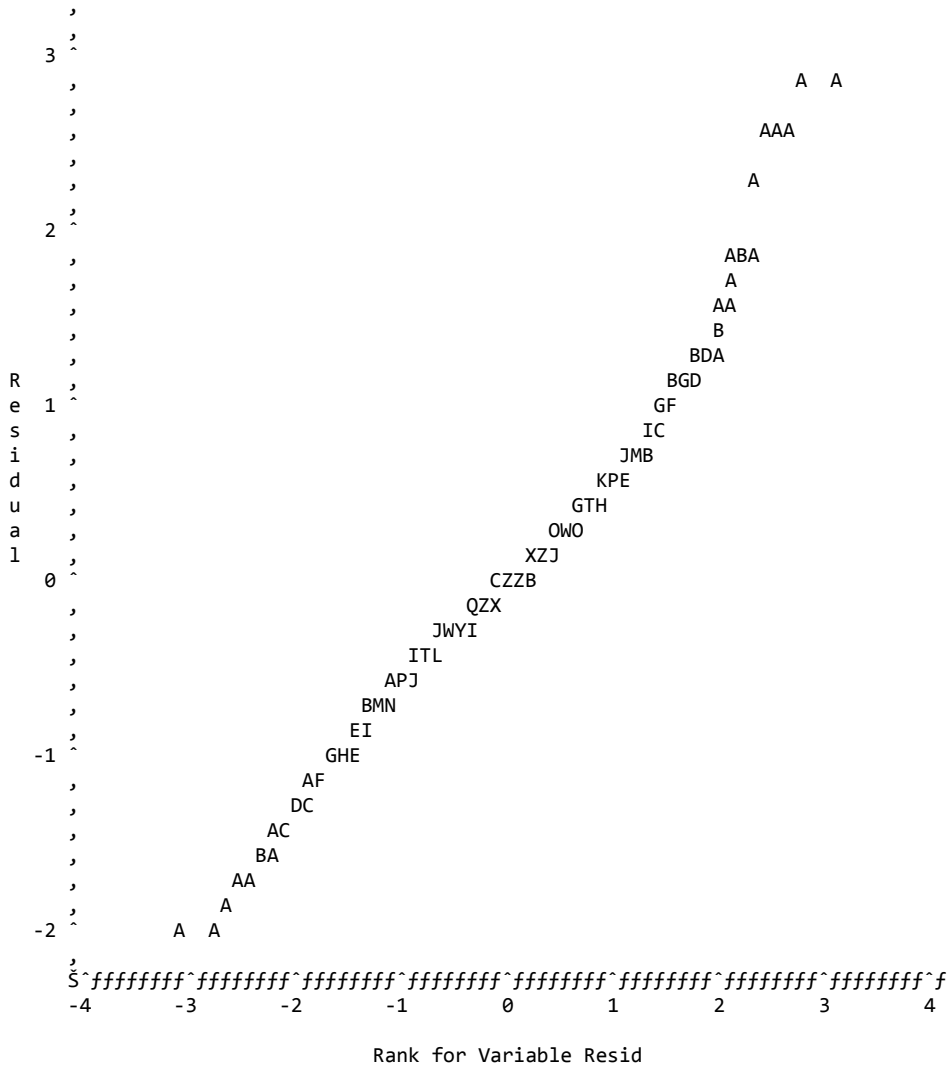
Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
age	1	42.9	1.34	0.2538
type	1	41.6	0.07	0.7948
month	1	176	12.61	0.0005
month2	1	497	8.07	0.0047
month3	1	489	5.47	0.0197
month4	1	549	3.91	0.0485

When we compare the fixed effect results from this analysis without the outlier with the original analysis with the outlier, the changes are noticeable enough to base conclusions on the analysis without the outlier. We look at the



Plot of Resid*s_resid. Legend: A = 1 obs, B = 2 obs, etc.



NOTE: 3 obs hidden.

With the outlier removed, the above plots of the residuals vs. predicted and the normal plot do not indicate any marked deviation from normality. In order to check the joint distributions of the three sets of random coefficients (patient, patient.time and patient.time2) we will plot them against each other with the following SAS code.

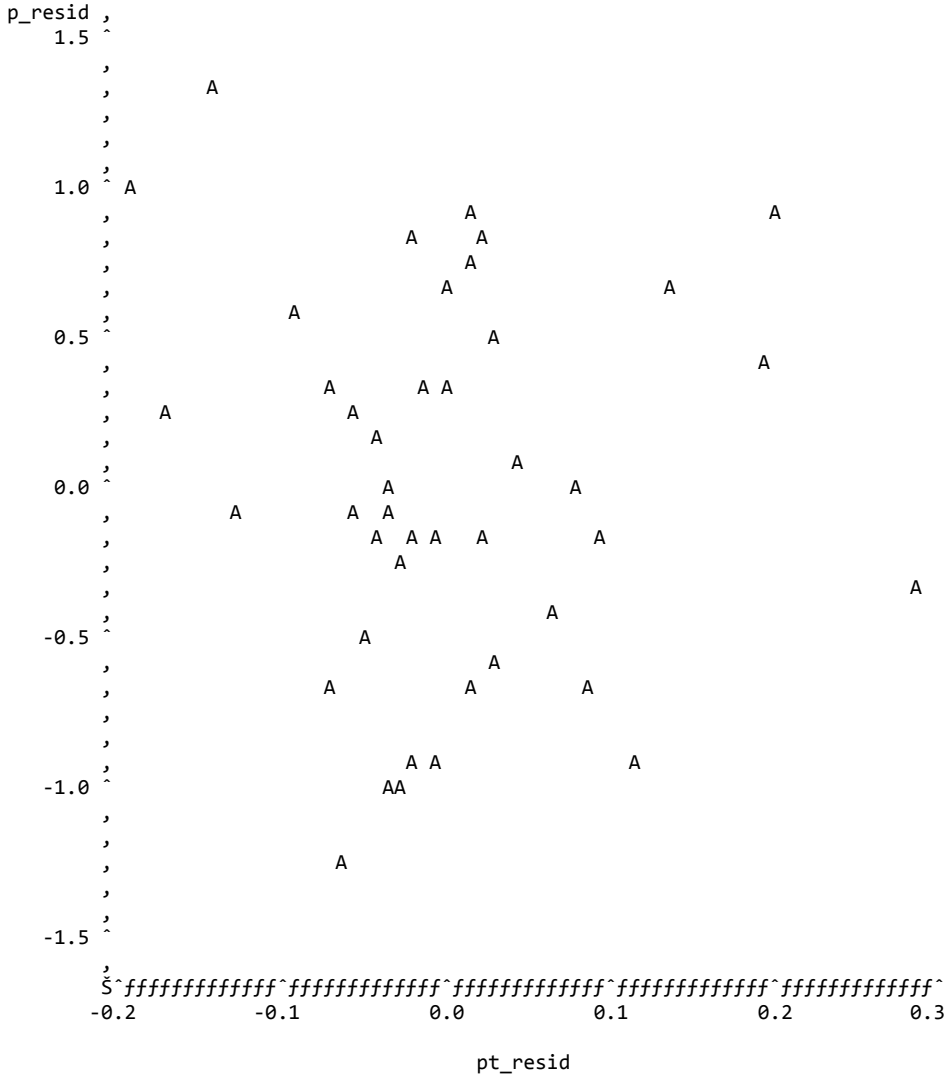
```
data solut;
set solut;
patx=pat*1;
drop pat;
run;

data p_resid(keep=pat p_resid) pt_resid(keep=pat pt_resid)
pt2_resid(keep=pat pt2_resid); set solut;
pat=patx;
if effect='Intercept' then do;
    p_resid=estimate;
    output p_resid;
end;
else if effect='month' then do;
    pt_resid=estimate;
    output pt_resid;
end;
else do;
    pt2_resid=estimate;
    output pt2_resid;
end;

proc sort data=predout;
by pat;
proc means noprint data=predout; by pat;
var pred;
output out=predm mean=p_pred n=freq;

data all;
merge p_resid pt_resid pt2_resid predm; by pat;
proc plot;
plot p_resid*pt_resid;
title 'patient coefficients vs patient.time coefficients';
run;
proc plot;
plot p_resid*pt2_resid;
title 'patient coefficients vs patient.time2 coefficients';
run;
proc plot;
plot pt_resid*pt2_resid;
title 'patient.time coefficients vs patient.time2 coefficients';
run;
```

Plot of p_resid*pt_resid. Legend: A = 1 obs, B = 2 obs, etc.



NOTE: 43 obs had missing values.

Plot of p_resid*pt2_resid. Legend: A = 1 obs, B = 2 obs, etc.

