BIOS 755: Parametric mean curves

Alexander McLain

February 4, 2025

Parametric Curves

- An alternative approach for analyzing the parallel-groups repeated measures design is to consider parametric curves for the time trends.
- We model the means as an explicit function of time
  - Linear trend
  - Quadratic Trend
  - Linear spline

Linear Trend

- If the means tend to change linearly over time, we can fit the following model:

$$E(Y_{ij}) = \beta_0 + \beta_1 t_j + \beta_2 TRT_i + \beta_3 t_j TRT_i$$

- Let Trt=1 if the subject is in group 1 and 0 otherwise. For subjects in treatment group 1,

$$E(Y_{ij}) = \beta_0 + (\beta_1 + \beta_3) t_j + \beta_2$$

- For subjects in the control group 0

$$E(Y_{ij}) = \beta_0 + \beta_1 t_j$$

- Thus, each group's mean is assumed to change linearly over time.

Quadratic Trend

- If the means tend to change over time in a quadratic manner, we can fit the following model:

$$E(Y_{ij}) = \beta_0 + \beta_1 t_j + \beta_2 t_j^2 + \beta_3 TRT_i + \beta_4 t_j TRT_i + \beta_5 t_j^2 TRT_i$$

- For subjects in treatment group 1,

$$E(Y_{ij}) = \beta_0 + (\beta_1 + \beta_4) t_j + (\beta_2 + \beta_5) t_j^2 + \beta_3$$

- For subjects in the control group 0

$$E(Y_{ij}) = \beta_0 + \beta_1 t_j + \beta_2 t_j^2$$

- Thus, each group's mean is assumed to change quadratically over time.

Quadratic Trend

- To avoid problems of collinearity in the quadratic (or in any higher-order polynomial) trend model, should always "center" t_j on its mean prior to the analysis (i.e. replace t_j by its deviation from the mean).
- For example, suppose $t_j = (1, 2, \ldots, 10)$. Then the correlation between t_j and t_j^2 is 0.975.
- However, if we create a "centered" variable, say,

$$t_j^* = t_j - \bar{t}$$

then the correlation between t_j^* and t_j^{*2} is 0.

Log Trend

- If the means are monotonically increasing (or decreasing), but there appears to be "diminishing returns" in the changes then a log transformation can be used

$$E(Y_{ij}) = \beta_0 + \beta_1 \log(t_j) + \beta_2 TRT_i + \beta_3 \log(t_j) TRT_i$$

- For subjects in treatment group 1,

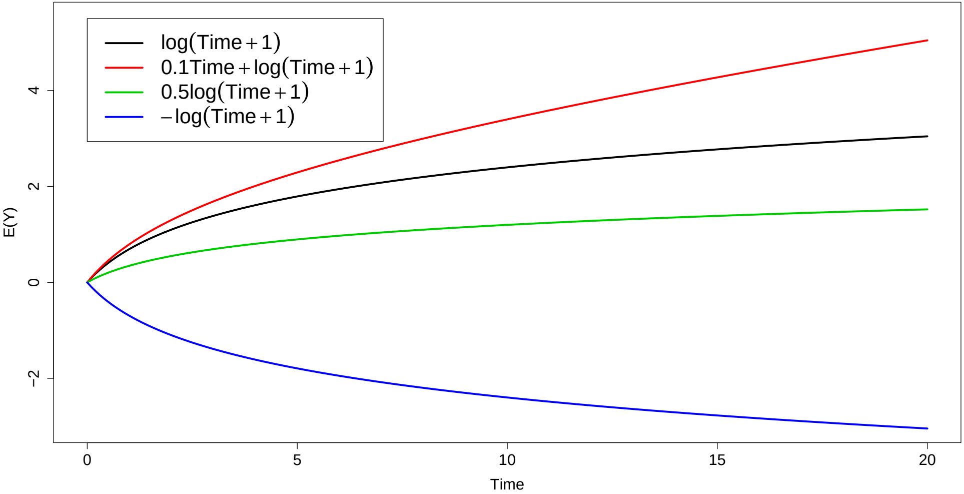
$$E(Y_{ij}) = \beta_0 + (\beta_1 + \beta_3) \log(t_j) + \beta_2$$

- For subjects in the control group 0

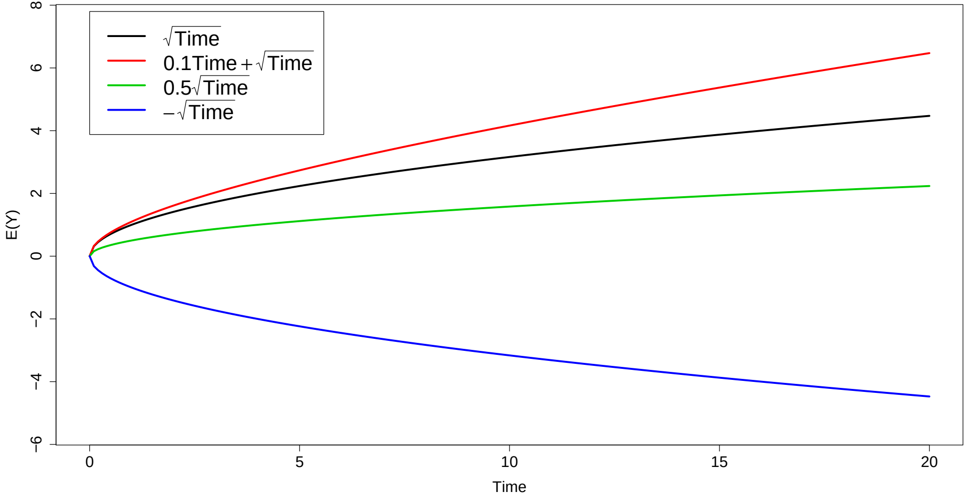
$$E(Y_{ij}) = \beta_0 + \beta_1 \log(t_j)$$

- The square root function $\sqrt{t}$ is another option for "diminishing returns".

- We can also add a linear trend to a log transformation.



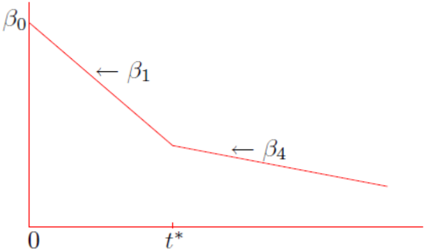
Square-root transformation



Linear Spline

- If the means change over time in a piecewise linear manner, we can fit the following linear spline model with knot at t^*:

$$
\begin{aligned}
E(Y_{ij}) &= \beta_0 + \beta_1 t_j + \beta_2 TRT_i + \beta_3 t_j TRT_i \quad t_j \le t^* \\
&= \beta_0 + \beta_1 t^* + \beta_2 TRT_i + \beta_3 t^* TRT_i + \beta_4 (t_j - t^*) + \beta_5 (t_j - t^*) TRT_i \quad t_j > t^*
\end{aligned}
$$



Linear Spline

- For subjects in group 1

$$\begin{aligned} E(Y_{ij}) &= \beta_0 + \beta_2 + (\beta_1 + \beta_3) t_j \quad t_j \le t^* \\ &= \beta_0 + \beta_2 + (\beta_1 + \beta_3) t^* + (\beta_4 + \beta_5)(t_j - t^*) \quad t_j > t^* \end{aligned}$$

- For subjects in group 2

$$\begin{aligned} E(Y_{ij}) &= \beta_0 + \beta_1 t_j \quad t_j \le t^* \\ &= \beta_0 + \beta_1 t^* + \beta_4 (t_j - t^*) \quad t_j > t^* \end{aligned}$$

Non-linear Spline

- Non-linear splines (e.g., cubic B-splines can also be used).

- These are used more with unbalanced data; we'll explore these with mixed models.

Summary of Features of Parametric Curve Models

- Allows one to model time trend and treatment effect(s) as a function of a few parameters.
- The treatment effect can be captured in one or two parameters, leading to more powerful tests when these models fit the data.
- Once you go beyond linear, quadratic, log, or square-root methods, the interpretations of the parameters are challenging (they are not easy for some of these!).
- Using profile analysis is usually preferred once things get beyond simple non-linear functions, especially for balanced data.

Summary of Features of Parametric Curve Models

- For unbalanced data, parametric curve modeling is more common. As a result, we'll return to some of these when we get to mixed models.
- Why? Since $E(Y_{ij})$ is defined as an explicit function of the time of measurement, t_j, there is no reason to require all subjects to have the same set of measurement times or the same number of measurements.
- These models can tell the difference, for example, between a kid who is 6.0 and 6.9 years old.

Parametric Curve Models

- Differentiating between a kid who is 6.0 and 6.9 years old could be important if you're trying to predict a child's risk for lung issues.
- Would it matter when determining the effect of an exposure of interest (say parents' occupation)?