1 Fitting the Reflux Data Semi-Parametrically

In section **??** the Reflux data was discussed. An ODE model along the lines of $y' = \beta y + u_0$ as described in Equation **??** was fitted to the data. Next, `Data2LD` was used to fit a sophisticated semi-parameteric model where the ODE $y'(t) = \beta(t) + u(t)$ only holds approximately.

For the purposes of introducing some of the ideas behind `Data2LD,` an ODE model that lies between these two extremes will be briefly examined in this section. The ODE model we have in mind takes the form:

$$\begin{cases} y'(t) &= \begin{cases} \beta_1 y(t) & t < t_0 \\ \beta_1 y(t) + u(t) & t \geq t_0 \end{cases} \\ y(0) &= \beta_0 \end{cases}$$

Here, the forcing function $u(t)$ is unknown and must be estimated along with β, so this is a semi-parametric model. How would one go about fitting this model?

The general solution to this ODE given by:

$$y(t) = \beta_0 + e^{\beta_1 t} \int_{t_0}^{\min(t, t_0)} u(s) ds$$

As before, the breakpoint at t_0 will be assumed fixed in advance. Letting $U(t) = \int u(s)ds$, the general solution can be more conveniently written as:

$$y(t) = \beta_0 + e^{\beta_1 t} U(t) H(t - t_0)$$

Here $H(t)$ denotes the Heaviside step function:

$$H(t) = \begin{cases} 0 & t \leq 0 \\ 1 & t > 0 \end{cases}$$

Let I denote the set of observations for which $t_i > t_0$. If β were known, $U(t)$ could estimated by non-parameterically regressing the values $(y_i - \beta_0)/e^{\beta_1 t_i}$ against t_i where $i \in I$. The forcing function $u(t)$ could then be estimated by differentiating the estimate for $U(t)$.

If $U(t)$ were known on the other hand, β_0 and β_1 could be estimated by minimising the following non-linear least squares criterion:

$$SSE(\beta_0, \beta_1; U(t)) = \sum_{i \notin I} (y_i - \beta_0)^2 + \sum_{i \in I} [y_i - \beta_0 - e^{\beta_1 t_i} U(t_i)]^2$$

A hierarchical estimation strategy seems natural. For a given choice of β, let $U(t|\beta_0, \beta_1)$ denote the estimate of $U(t)$ produced by regressing the $(y_i - \beta_0)/e^{\beta_1 t_i}$ against t_i. Define the associated least squares error by:

$$
\begin{aligned}
H(\beta_0, \beta_1) &= SSE(\beta; U(t|\beta_0, \beta_1)) \\
&= \sum_{i \notin I} (y_i - \beta_0)^2 + \sum_{i \in I} [y_i - \beta_0 - e^{\beta_1 t_i} U(t|\beta_0, \beta_1)]^2
\end{aligned}
$$

The sketch presented here fails to address two important questions: how to to estimate $U(t)$ given β_0 and β_1, and how to optimise $H(\beta_0, \beta_1)$. Using the nomenclature of the field, there are two fitting problems, where the results of one is used as a covariate by the other. The *inner problem* entails estimating $U(t)$ given the parameters β_0 and β_1, and the *outer problem* entails minimising $H(\beta)$.

`Data2LD` employs a powerful hierarchical fitting methodology for FDA problems known as the *Parameter Cascade.*