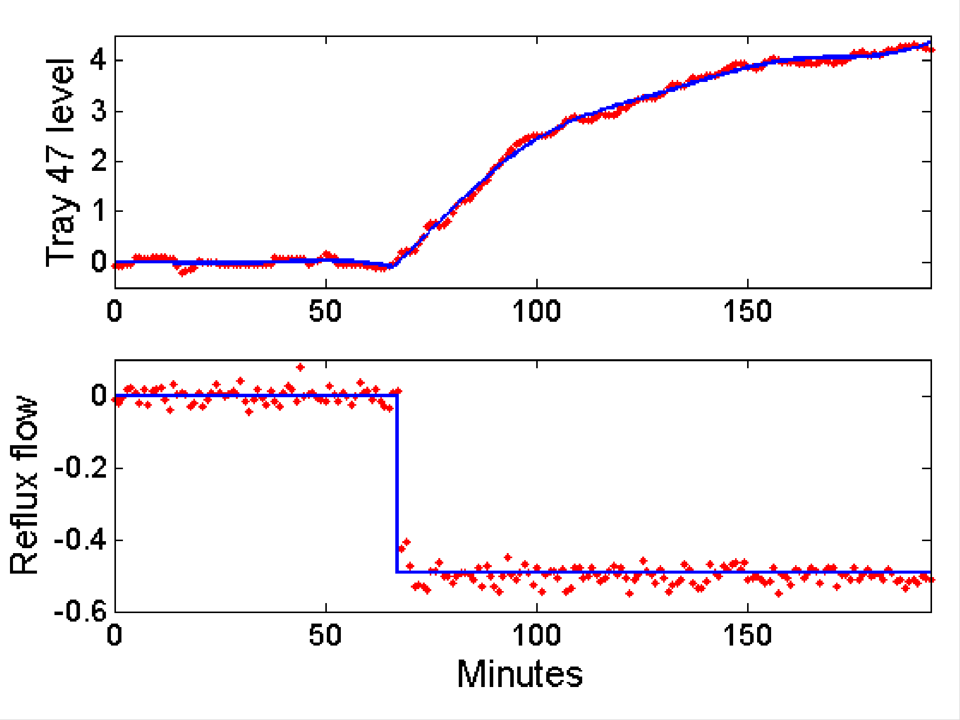
Derivatives and functional linear models

A first look at a differential equation

as a modelling tool.

1. The oil refinery data

Refinery output $x(t)$ (top panel) and input $u(t)$ (bottom panel)



- This is a simple input/output system in an oil refinery in Corpus Christi, Texas.
- A fluid, called *reflux*, flows into *tray 47* in a distillation column in an oil refinery.
- The input variable $u(t)$ is the flow rate.
- The fluid level in the tray is the output variable $x(t)$.

Variation on two time scales

- Over the longer scale, tray level changes from 0 to around 4.
- But we are also interested in how rapidly the change takes place; that is, short-scale variation.
- It looks like about 2/3 of the change takes about 50 minutes, and the final value is reached in 200 minutes or so.

A concurrent functional regression

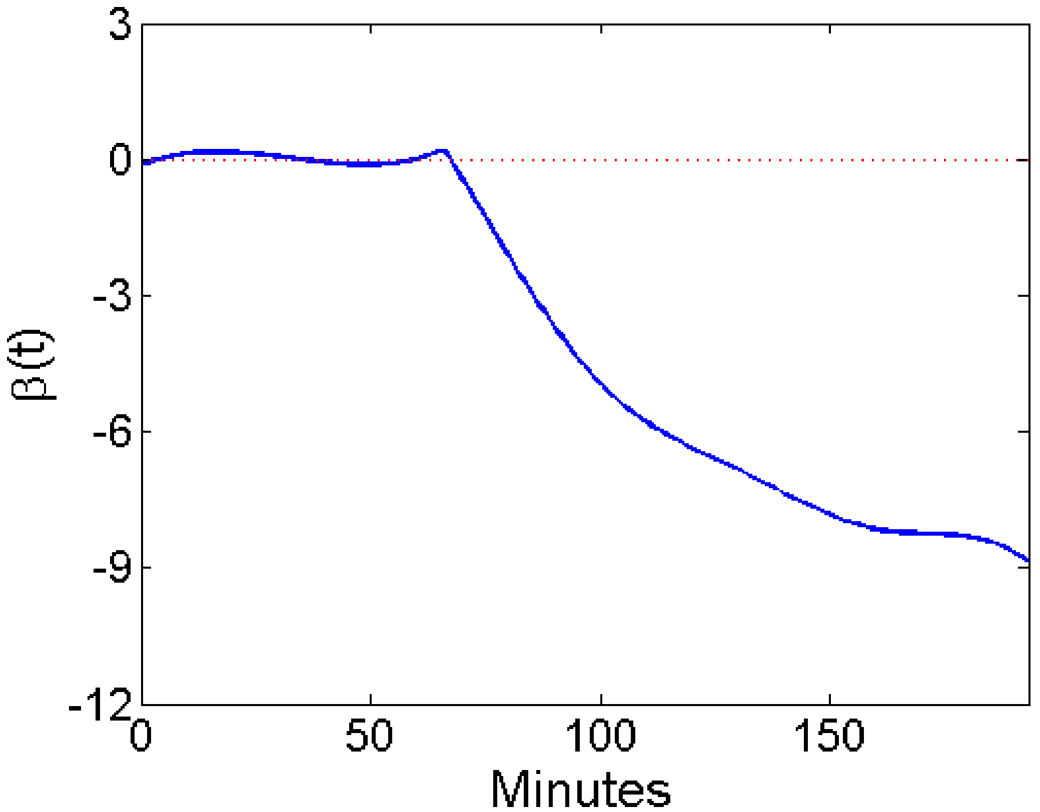
- We can model the output *state* $\texttt{Tray}(t)$ as a simple time-varying regression on input state $\texttt{Reflux}(t)$:

$$\texttt{Tray}(t) = \texttt{Reflux}(t)\beta(t) + \epsilon(t)$$

- In functional data analysis, we call this a *concurrent* regression because only the simultaneous influence of the input on the output is modelled.
- A least squares estimate $\hat{\beta}(t)$ of the regression coefficient function minimizes

$$\texttt{SSE} = \int [\texttt{Tray}(t) - \texttt{Reflux}(t)\beta(t)]^2 \, dt$$

- $\hat{\beta}(t)$ is modelled with an expansion in terms of several B-spline basis functions.



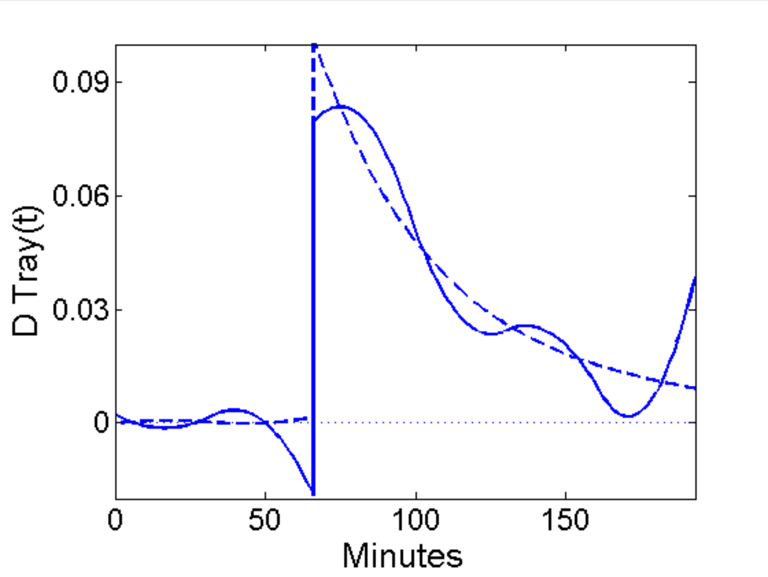
But does this tell us anything new? The regression function is just as complicated as the output.

Making the derivative $D\texttt{Tray}(t)$ the output

- We now model the *rate of change* $D\texttt{Tray}(t)$, using the output state $\texttt{Tray}(t)$ and the input $\texttt{Reflux}(t)$ as covariates.
- We'll use constants for the two regression functions:

$$D\texttt{Tray}(t) = -\beta_1 \texttt{Tray}(t) + \beta_2 \texttt{Reflux}(t) + \epsilon(t)$$

- This is an example of a *first order differential equation with constant coefficients*.
- We see that it is just another form of concurrent functional linear model, now with two covariates, but with regression functions β_1 and β_2 that are constant.

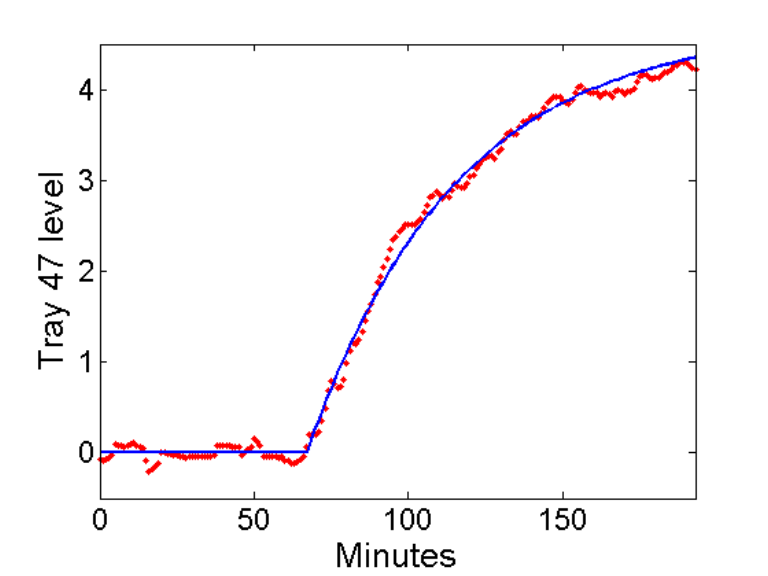


The solid line is the derivative estimated from the data, and the dashed line is the model's fit to this derivative. $\hat{\beta}_1 = 0.02$ and $\hat{\beta}_2 = 0.19$

- A differential equation that is this simple has an explicit solution.

$$x(t) = e^{-\beta_1 t}[x(0) - (\beta_2/\beta_1)\int_0^t e^{\beta_1 s} u(s)\, ds].$$

- $\beta_1 \approx 0.02$ is the rate constant, and therefore controls the rate of change of `Tray` level. About 2/3 of the change takes $1/\beta_1$ time units, and the final level is nearly reached in $4/\beta_1$ time units. β_1 models the *dynamic behavior* of `Tray`.
- $\beta_2 \approx 0.19$, along with β_1, defines the ultimate change; the long-term *gain* per unit increase in Reflux flow is $\beta_2/\beta_1 \approx 9.5$.

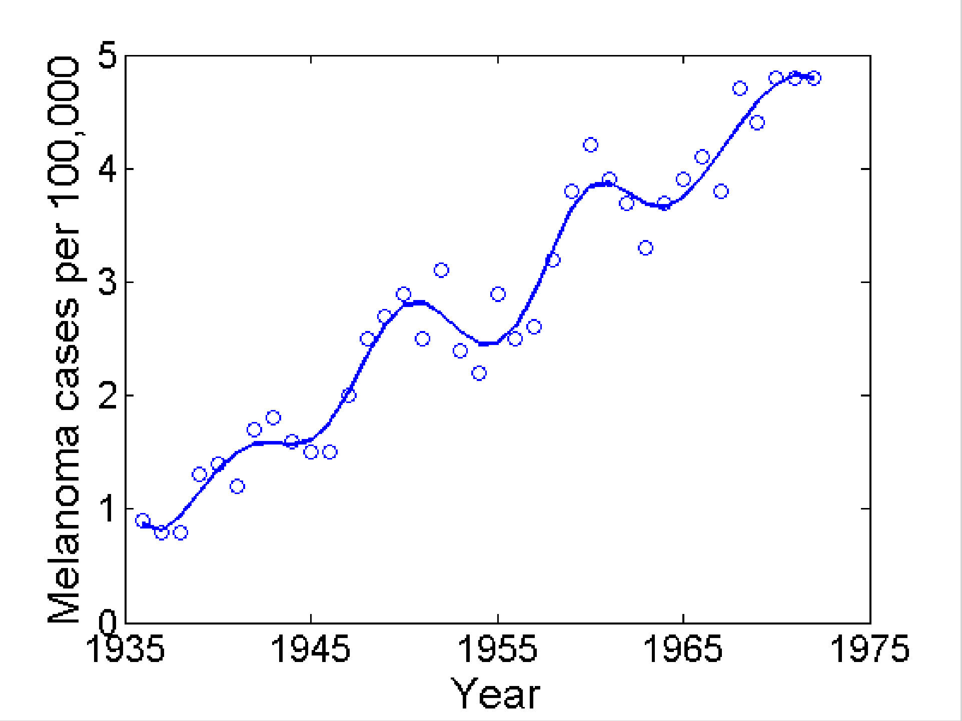


Modelling the rate of change $D\texttt{Tray}(t)$ directly produces a fine fit to the data with only two parameters.

2. The melanoma data

Age-adjusted melanoma incidences for Connecticut.

The solid line is a spline smooth with penalty on D^4x.



Estimating a differential operator L

- A differential operator L is just a re-arranged differential equation.
- Can we smooth the data and estimate the operator at the same time?
- Will this give us a better fit to the data with fewer degrees of freedom used up?
- We will try the operator

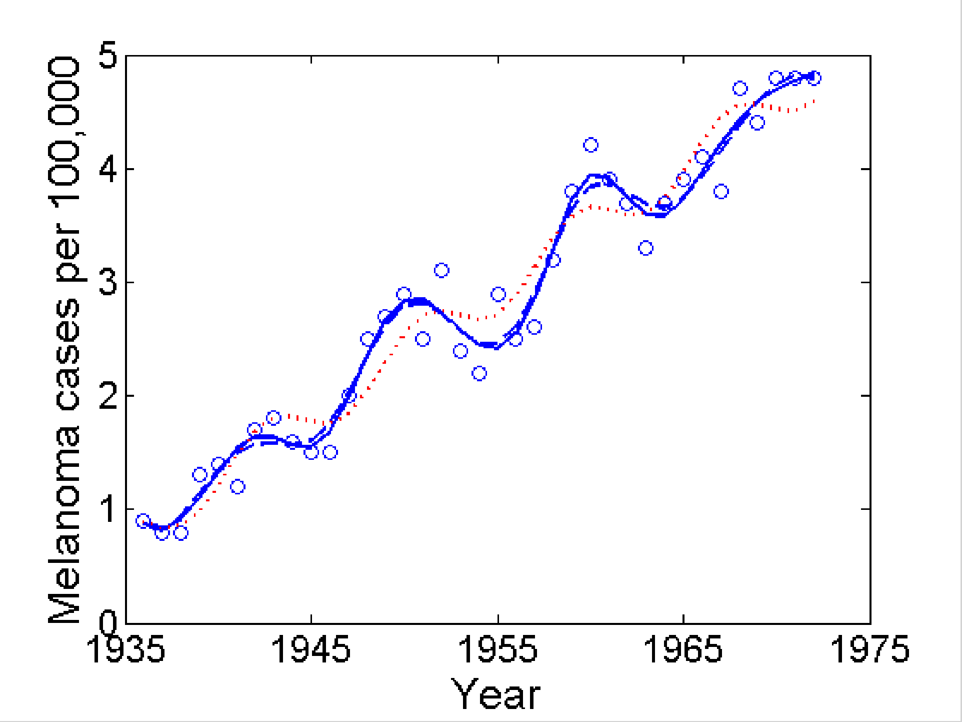
$$Lx = \beta_1 D^2 x + D^4 x$$

- This operator a tilted line plus sinusoidal trend with the period to be estimated from the data.

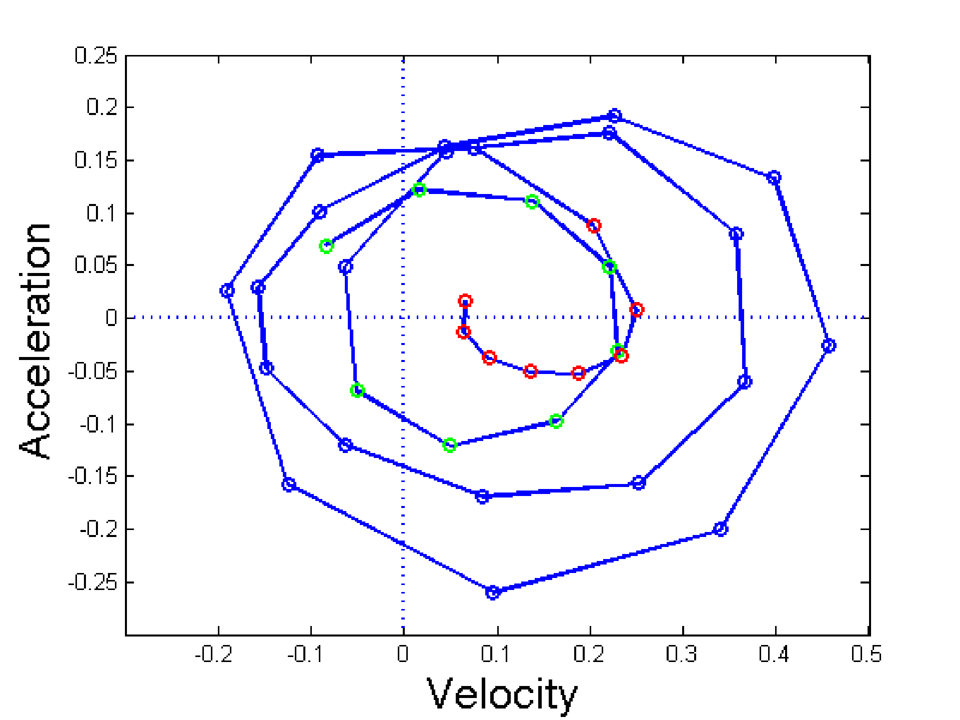
The algorithm

- Start with $\beta_1 = 0$ and $Lx = D^4x$, and estimate derivatives up to order 4, choosing λ to minimize the GCV criterion.
- Carry out concurrent functional regression to estimate β_1.
- Re-smooth using Lx, again re-computing derivatives and minimizing GCV.
- Continue until the parameter estimates converge.

Solid blue line is smooth using converged Lx penalty. Dashed line is smoothing using D^4 penalty. Dotted line is a solution to differential equation $Lx = 0$.



Phase-plane plot



3. Summary

- By using a derivative as the dependent variable, we can model the rate of change of an output variable.
- The resulting differential equation can be solved to provide a model for the output variable itself.
- We effectively get two or more models for the price of one.
- We are simultaneously modelling the *level* or *state* of a system and its *dynamics*.