17

Derivatives and functional linear models

17.1 Introduction

This chapter is an introduction to the idea of a *differential equation*, and aims to provide for readers unfamiliar with differential equations some of the basic ideas that will carry them forward into the next chapters. We begin with an example where we see the advantages of modelling the *rate of change* of a function as the dependent variable. Of course, by term "rate of change" we mean a derivative of a function, and in this case the first derivative. Models for derivatives are often termed models for the *dynamics* of a system, or *dynamic models.*

We will see how these dynamic models, expressed as differential equations, permit us to model both the function itself and one or more of its derivatives *at the same time.* How does this differ from what we have already been doing, say, with the growth data? There, by contrast with a truly dynamic model, we begin with a model for the observed data, the height measurements. To be sure, we selected this model with an eye to looking at derivatives, but fundamentally we modelled the data and then let the derivatives emerge as by-products. Now, however, and in the next chapters, we look at linking derivatives and function values together so as to take away the privileged place of the function as the object to be estimated.

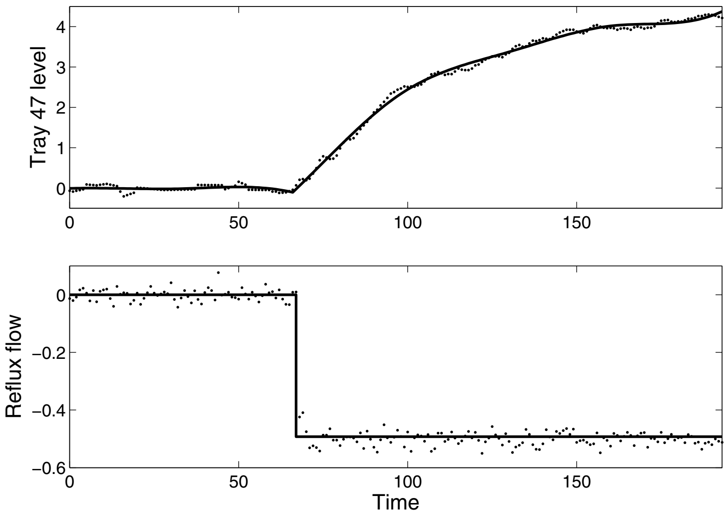

Figure 17.1. The upper panel shows the level of material in a tray of a distillation column in an oil refinery, and the lower level shows the flow of material being distilled into the tray. The points are measured values, and the solid lines are smooths of the data using regression splines. Time is in minutes.

17.2 The oil refinery data

A distillation column or cracking tower in an oil refinery converts crude oil to refined petroleum products like gasoline by boiling the crude and passing the vapor through a series of trays where, at each level, the condensate becomes more refined. Figure 17.1 shows the output from tray number 47 in the upper panel in response to the input shown in the lower panel. Both functions have been centered on their values at time 0, the time and flow units are unknown, and input flow has been measured in the downward direction.

We see that the output changes slowly in response to an abrupt change in input, although it is clearly headed toward some stable upper level between four and five units. It seems to have a fair amount of inertia, and the results are analogous to those of a person pushing a car on level ground. Otherwise there does not seem to be much to understand here; we increase the flow into a tank with an outlet, and the level rises.

The refinery data show variation on two time scales: The long-term scale involves the overall change in level from zero to near five that takes place over several hundred time units, and the the shorter time scale covers period from time 67 to where the new level is achieved, covering about one hundred units. We would like to find a way to model both the long-term change in

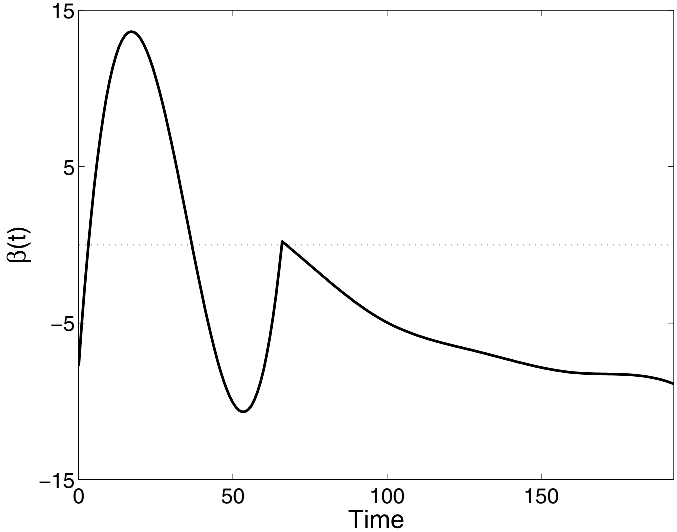

Figure 17.2. The regression coefficient function β for model (17.1) for the oil refinery data.

the output, and the rate of change over the shorter period that produces this change.

Here are a few technical asides on how we smoothed these data. Both functions show a sharp break at time number 67; the upper curve has a discontinuous derivative, and the lower curve is itself discontinuous. In order to have the smooth curve for the output to have a derivative discontinuity at 67, we used order four splines and placed three coincident knot values at that time. There was also one knot positioned midway between 0 and 67, and three equally spaced knots between 67 and 193. These knot choices imply a total of eleven basis functions. The lower curve was fit with order one splines with a single interior knot placed at 67.

Suppose that we model these data using the concurrent functional linear model described in Chapter 14, so that

$$\texttt{Tray}(t) = \texttt{Reflux}(t)\beta(t) + \epsilon(t). \tag{17.1}$$

We used nearly the same basis system for the single regression coefficient $\beta(t)$ except that we dropped the interior knot in the first interval, thus using ten splines. Figure 17.2 displays the estimated regression function. After time 67, β simply mirrors the behavior of the output, and we have little interest in its behavior before time 67, where it captures some of the data's wanderings around zero. The fit to the data, not shown, is virtually the same as that shown in Figure 17.1.

This seems disappointing. We haven't learned much from the shape of the regression function that we couldn't see in the original data. In fact, a

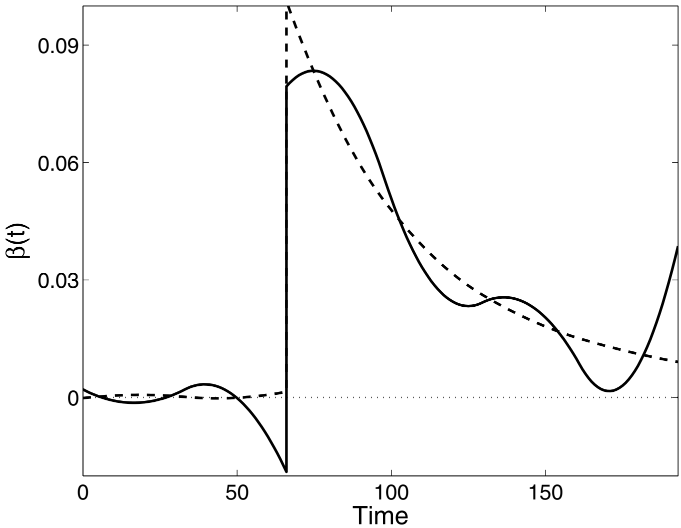

Figure 17.3. The first estimated derivative of Tray 47 level is shown as a solid line, and the fit to this derivative from model (17.2) is shown as a dashed line.

little thought convinces us that simply making β negatively proportional to `Tray` is going to work just fine.

Now let's take a different approach, involving explicit use of tray 47's first derivative as computed from the smooth in Figure 17.1, and shown in Figure 17.3. There is a fair amount of variability in this derivative estimate, but we do see something like exponential decay in the derivative after time 67, which seems consistent with what we see in Figure 17.1.

We propose to make this derivative the dependent variable, and to use two independent variables, namely `Tray` level itself and the input, Reflex flow. The model is therefore

$$\texttt{DTray}(t) = -\beta_1(t)\texttt{Tray}(t) + \beta_2(t)\texttt{Reflux}(t) + \epsilon(t). \tag{17.2}$$

It is the usual practice in formulating a linear differential equation model to place a minus sign in front of coefficient functions such as $\beta_1(t)$.

The motivation here is to model the behavior of the rate of change of the output as a function of both the output level and the input. This time we will impose extreme simplicity on both the regression functions by using a constant basis for each. The results that we obtain are $\beta_1(t) = 0.02$ and $\beta_2(t) = -0.20$. Figure 17.3 shows the fit to the first derivative offered by this model, and we have captured nicely the idea of zero derivative up to time 67 and exponential decay afterwards.

Model (17.2) is an example of a first order linear differential equation with constant coefficients. This is to say that the equation links the first derivative to the function value and the input function, and that the linking equation is linear with coefficients that are constant. In order to see how well the result fits the data, we need to solve equation (17.2) for `Tray`. Fortunately, any basic text on differential equations will give us the solution, which is, using $y(t)$ to stand for `Tray`(t) and $u(t)$ to stand for `Reflux`(t),

$$y(t) = e^{-\beta_1 t}[y(0) - (\beta_2/\beta_1) \int_0^t e^{\beta_1 s} u(s)\, ds]. \tag{17.3}$$

We can simplify this further by specifying that $y(0) = 0$, $u(t) = 0, t \leq 67$, and $u(t) = -0.4924, t > 67$ to get

$$y(t) = 0.4924 \frac{\beta_2}{\beta_1}[1 - e^{-\beta_1 (t-67)}],\ t \geq 67,\ \text{and 0 otherwise.} \tag{17.4}$$

The fit to the data offered by this equation is shown in Figure 17.4. The two parameter values define a model that fits the data beautifully, and predicts that the new level that `Tray` is approaching is 4.7.

Here's a summary of what we learn from the model by studying equations (17.3) and (17.4):

- When there is no input, `Tray` level will decay exponentially with a rate constant of -0.02 from whatever its level is at time 0.
- When `Reflux` increases by one unit, the level of Tray 47 will increase at an exponentially declining rate (rate constant again -0.02) to a new level $0.2/0.02 = 10$ units higher. This is the long-term change in the output.
- The time from increase in `Reflux` to the time `Tray` achieves its new level is about $4/0.02 = 200$ time units, and this is the shorter term period in which the actual change takes place.
- β_1 is the rate constant, and therefore controls the rate of change of `Tray` level. It models the dynamic behavior of `Tray`.
- β_2, along with β_1, controls the ultimate change; the long-term *gain* per unit increase in Reflux flow is β_2/β_1.

17.3 The melanoma data

Figure 17.5 presents age-adjusted melanoma incidences for 37 years from the Connecticut Tumor Registry (Houghton et al. 1980). The solid line is a smoothing spline fit by penalizing the size of the fourth derivative $D^4 x$ and choosing the penalty parameter by minimizing generalized cross-validation or GCV. Two types of trends are obvious: a steady linear increase and a periodic component. The latter is related to sunspot activity and the accompanying fluctuations in solar radiation. If we look closely, though, we can also see that there are some changes in the periodic trend; the peaks

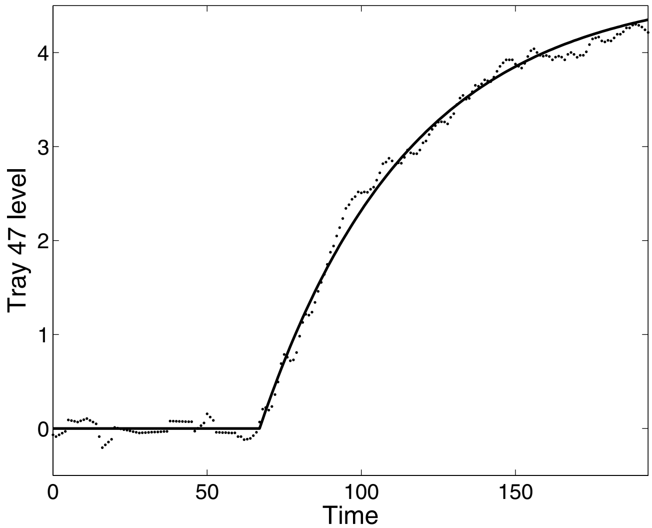

Figure 17.4. The fit to the data defined by model (17.2) is shown as a solid line, and the data as points.

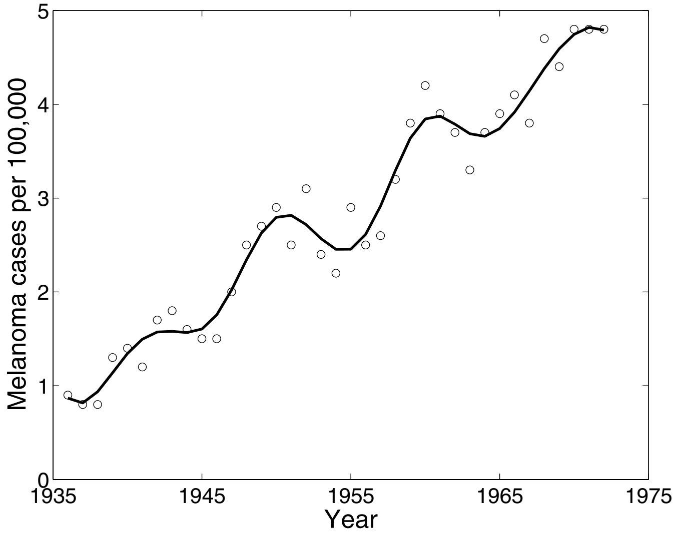

Figure 17.5. Age-adjusted incidences of melanoma for the years 1936 to 1972. The solid curve is the polynomial smoothing spline fit to the data penalizing the norm of the fourth derivative, with the smoothing parameter chosen by minimizing the GCV criterion.

around 1950 and 1960 seem stronger than those near 1940 and 1970, and perhaps the length of each cycle changes a little, too.

In short, there are three time scales here: the unlimited time over which linear trend is maintained, the short-term sunspot cycle of about ten years, and the medium term covering the range of years in the data in which the cycles themselves change.

We again want to find a simple model that will capture changes on these three time spans, and that will also tell us something about the dynamics of the cyclical variation. We already know that a straight line solves the differential equation $D^2x = 0$ and that $\sin(\omega t)$ and $\cos(\omega t)$ solve the equation $D^2x = -\omega^2 x$ for some period $2\pi/\omega$. We can put these two ideas together working with the fourth order equation $D^4x = -\omega^2 D^2 x$. Let's add one more parameter to define the differential equation

$$D^4x = -\beta_1 D^2 x - \beta_2 D^3 x, \tag{17.5}$$

where $\beta_1 = -\omega^2$ and β_2, called the *damping coefficient*, allows for an exponential decay in the oscillations by multiplying $\sin(\omega t)$ and $\cos(\omega t)$ by the factor $\exp(-\beta_2 t/2)$ where $t = \text{year} - 1935$.

Here's an algorithm for estimating the unknown coefficients β_1 and β_2:

1. Start by smoothing the data, as we have already done, using smoothing splines penalized by using D^4 with the smoothing parameter λ that minimizes GCV.

2. Compute the derivatives of the smooth up to order four.

3. Carry out a regression of the fourth derivative values, taken at each year, on the corresponding values for the second and third derivatives. The regression coefficients are estimates of β_1 and β_2.

4. Define the linear differential operator L as

$$Lx = \beta_1 D^2 x + \beta_2 D^3 x + D^4 x. \tag{17.6}$$

   Operator L is just a re-arrangement of differential equation (17.5); x satisfies the equation if and only if $Lx = 0$.

5. Now smooth the data using the roughness penalty defined by this linear differential operator, and again choose λ to minimize GCV. Hopefully, because this operator will annihilate more of the variation in the data than D^4 would, the smooth will be better and the estimates of the derivatives will also improve.

6. Check for convergence in the regression coefficients, or in the value of GCV. If convergence occurs, continue on to the last step; otherwise, return to step 2.

7. As we did for the refinery data, see how well the smooth fits the data, and also how well the data are fit by a solution to the differential equation (17.5).

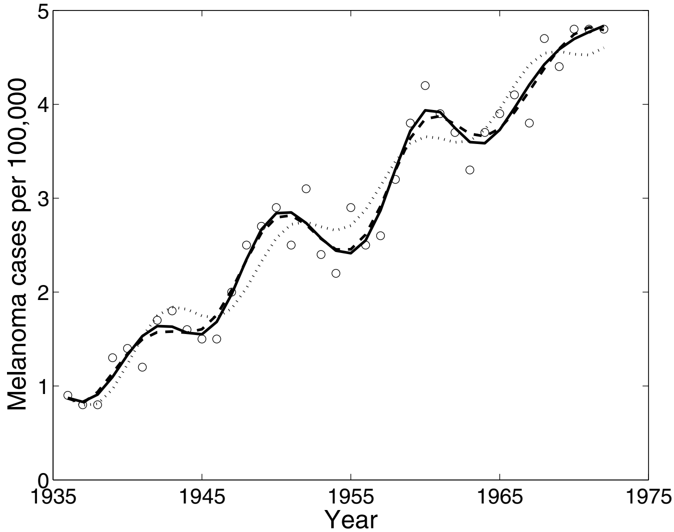

Figure 17.6. The light dashed line is the minimum–GCV fit to the data using a roughness penalty on D^4. The heavy solid line is the fit using a roughness penalty on Lx, where L is defined by (17.6). The dotted line is the best fit for functions satisfying differential equation (17.5).

This process effectively converged in five iterations, at which point β_1 and β_2 are 0.56 and 0.018, respectively. We can work out that $\omega^2 = 3\beta_2^2/4 + \beta_1$, and this corresponds to a period of 8.39 years. The period was estimated in the first iteration as 11.22 years.

Now we're in a position to compare the various fits to the data:

- The same fit as in Figure 17.5 using the D^4 operator.
- The smoothing fit using the converged value of operator L.
- The fit Lx satisfying $Lx = 0$, that is, satisfying the differential equation (17.5).

Each of these fits are shown in Figure 17.6. The final smooth tracks the data a bit better, especially between 1960 and 1965. But now we have a good estimate of the trend that can be fit with an exponentially decaying sinusoid plus linear trend, and we see that there are indeed phase differences between the smooth and the strictly periodic fit. Actually, the exponential decay is small, and scarcely visible in the plot.

The changes in the cycles resulting from iteratively updating the smoothing function and its derivatives are more visible in the phase-plane plot in Figure 17.7. In the right panel, showing the results for the estimated roughness penalty, the amplitudes of the cycles are stronger and the behavior of

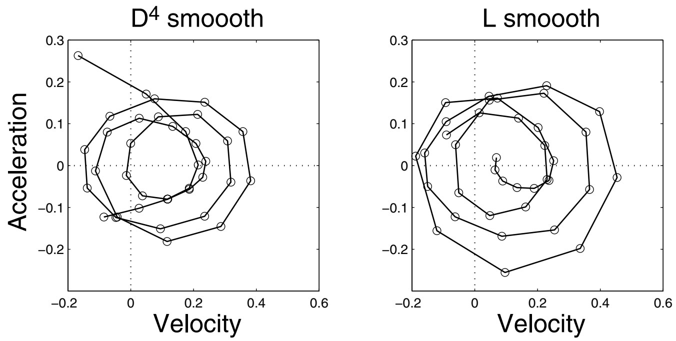

Figure 17.7. Two phase–plane plots for the fit to the melanoma data. The left panel is for the initial roughness penalty defined by the differential operator D^4, and the right panel is for the estimated operator L defined in (17.6).

the fit at the beginning and end of the curve is consistent with its behavior elsewhere.

17.4 Some comparisons of the refinery and melanoma analyses

Why was the differential equation (17.2) for the refinery data of order one and (17.5) for the melanoma model of order four? The reason is that we could express the shape of `Tray` in terms of only a single function, whereas we required four component or basis functions to express the essential structure of the melanoma data. Of course, the melanoma model required only two constants to be estimated, but that was because we could assume that the multipliers of x and Dx were zero. They are there, after all, but are just not estimated from the data.

On the other hand, the refinery data involved both an input and an output. Hence, we needed a parameter to model the impact of a change in the input, as well as a parameter to model the internal dynamics of the output. In the melanoma data, there was no input (although we could well have used sunspot activity records as an input), but the internal dynamics were, in effect, four dimensional.

In both problems we used two levels of fitting. The low-dimensional fit was defined by the solution of a differential equation, and the higher-dimensional fit was achieved by keeping smoothing parameter λ low enough that the roughness penalty did not overwhelm the data fit. This means that we *partitioned the functional variance* into two parts: the low-dimensional part captured by the differential equation, and the balance which is the difference between the low- and high-dimensional fits. The differential equations played key roles in this process.

For both models we estimated some parameters defining the differential equation from the data. In effect, the process that we used for the refinery data was simply a one-step version of the more sophisticated algorithm that we used for the melanoma data.

Perhaps this is the most important conclusion to take away from this chapter: We can use noisy data to estimate a differential equation that expresses at least a substantial part of the variation in the data. This problem is taken up in Chapter 19. First, though, you may want to read the next chapter, which offers a review of a number of results about linear differential equations and linear differential operators.