# 14

# Functional responses, functional covariates and the concurrent model

## 14.1  Introduction

We now consider a model for a functional response involving one or more functional covariates. In this chapter the influence of a covariate on the response is of a particularly elementary nature: The response $y$ and each covariate $z_j$ are both functions of the same argument $t$, and the influence is *concurrent, simultaneous* or *point-wise* in the sense that $z_j$ only influences $y(t)$ through its value $z_j(t)$ at time $t$. This contrasts with the more general situation that we will defer for two chapters in which the influence of $z_j$ can involve a range of argument values $z_j(s)$.

We will see that this functional/functional model involves only minor changes at the computational level of the functional response and multivariate covariate model in the last chapter. Perhaps this is not surprising, since a scalar covariate can be viewed as a functional covariate expanded in terms of a constant basis, where the single coefficient multiplying the basis function value 1 is the value of the covariate. Therefore the functional/multivariate model is really contained within what we take up in this chapter. But of course the fact that the functional covariate is not constant does add new features that now need to be considered. We begin with a concrete problem.

## 14.2  Predicting precipitation profiles from temperature curves

### 14.2.1  The model for the daily logarithm of rainfall

Predicting temperature is relatively easy, but predicting rainfall is quite another thing. Certainly there are important precipitation effects due to climate zones, but can we get additional predictability from the behavior of temperature? It seems likely, for example, that on days when the average temperature is high, precipitation tends to be low, at least in the summer. In the winter, on the other hand, most of the snowfall comes when the temperature is only a little below freezing; when it is really cold, it seldom snows since the atmosphere is too dry.

Here is an extension of the functional ANOVA model (13.1) that we could describe as a *functional analysis of covariance* model:

$$\log[\texttt{Prec}_{mg}(t)] = \mu(t) + \alpha_g(t) + \texttt{TempRes}_{mg}(t)\beta(t) + \epsilon_{mg}(t). \qquad (14.1)$$

We consider the log of precipitation as the response since precipitation is a magnitude, and experience indicates that logging magnitudes tends to improve the fitting power of linear models. As in Chapter 13, $g$ indexes climate zones, $m$ indexes weather stations within climate zones, and climate zone effects satisfy the constraint $\sum_g \alpha_g(t) = 0$.

The variable $\texttt{TempRes}_{mg}$ is the residual temperature after removing the temperature effect of climate zone $g$ by using the techniques of Chapter 13. The motivation for removing temperature climate effects from the temperature profiles before using them in this model is that we have already allowed for these effects in the model. We don't want climate zones in the equation twice.

### 14.2.2  Preliminary steps

The average daily precipitation data for some extremely dry stations such as Resolute contain a number of zeros, and we dealt with this by replacing these with 0.05 mm since the smallest nonzero value was 0.1 mm. This permits us to smooth the logarithm of average precipitation directly. We first used 365 Fourier basis functions, and the same harmonic acceleration roughness penalty that we have been using for the weather data. The generalized cross-validation or GCV criterion was minimized for $\lambda = 10^6$, a level of smoothing that is equivalent to about 9.5 degrees of freedom. In order to speed up computation, we then opted for a simple Fourier basis expansion with eleven basis functions and no roughness penalization. For this analysis, we used an expansion of the daily average temperature residual in terms of 21 Fourier basis functions.

The smooth log precipitation curves for all 35 weather stations are shown in Figure 14.1. The rainiest place in Canada is unlucky Prince Rupert,
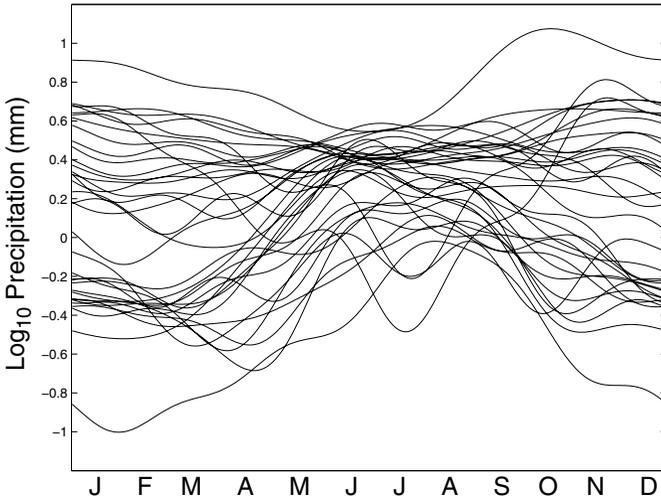
Figure 14.1. The logarithm (base 10) of average daily precipitation after smoothing for 35 Canadian weather stations.
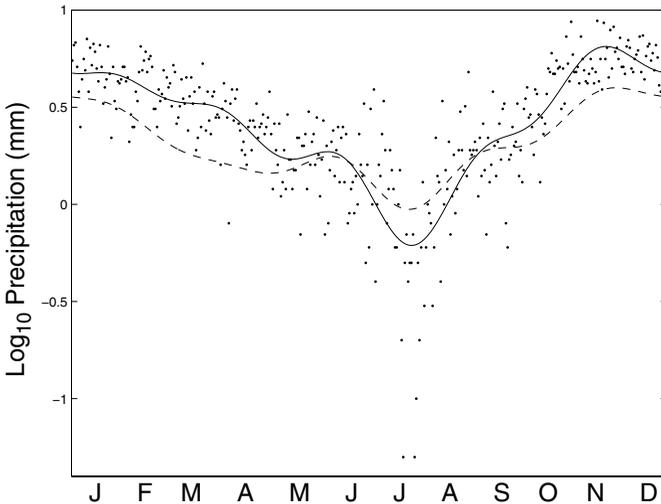


Figure 14.2. The $\log_{10}$ of average precipitation at Vancouver over 34 years is indicated by the dots, the smooth of the data using 11 Fourier basis functions by the solid curve, and the fit to the smooth curves by the point-wise linear model (14.1) by the dashed curve.
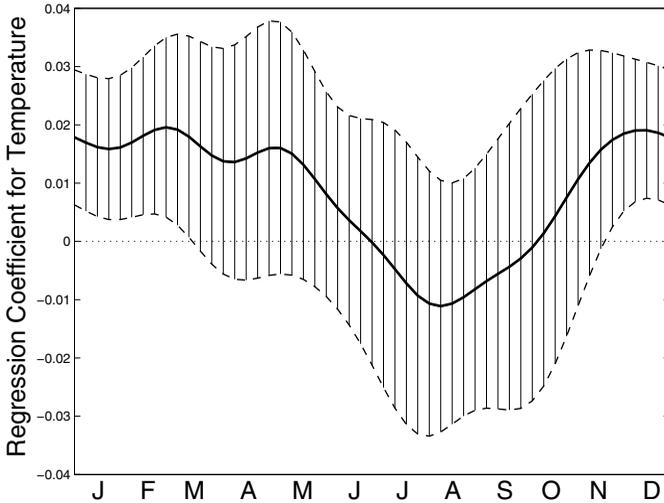
Figure 14.3. The regression coefficient for temperature with climate zone effects removed in a model predicting $\log_{10}$ rainfall. The solid line is the regression function and the cross-hatched area is the point-wise 95% confidence region for the function.

which averages nearly 12 mm of rain a day in October. The driest station is Resolute in the high arctic, where snowfall has a barely measurable rain equivalent of 0.1 mm per day in the winter. Figure 14.2 shows the resulting smooth to the precipitation data for Vancouver, a station that shows a sharp drop in rainfall during the summer months, and even records two days with no precipitation in 34 years.

## 14.2.3    Fitting the model and assessing fit

The unweighted least squares criterion for assessing fit is

$$\texttt{LMSSE}(\mu, \alpha_g, \beta) = \sum_{g}^{4} \sum_{m}^{N_g} \int \texttt{LogPrecRes}^2_{mg}(t)\, dt, \qquad (14.2)$$

where

$$\texttt{LogPrecRes}_{mg}(t) = [\texttt{LogPrec}_{mg}(t) - \mu(t) - \alpha_g(t) - \texttt{TempRes}_{mg}(t)\beta(t)].$$

When we fit the model, using an approach that will be described in detail in the next section, we obtain a standard error of 467.9. If we drop `TempRes` from the model, this increases to 510.8, and these values are equivalent to $R^2 = 0.08$. Overall, the temperature residual functions don't seem to improve the fit by much. Figure 14.3 confirms this by showing point-wise 95%

confidence intervals for the estimated regression function for the residual temperature functions. The only part of the year where temperature seems to make a contribution is December through February.

However, it is potentially misleading to report that the regression coefficient is "significantly different from zero" at the end of January, since we are, in a sense, optimizing significance over a year's worth of results. The right way to proceed is to construct a contrast, a linear weighting of the entire year's information that focusses on the effect of interest. We can reasonably say that focussing on the effect in the winter is a test that we could propose in advance of collecting the data; we knew already that there is much more potential variation in rainfall across weather stations in the winter months and much more variability in temperature available then to predict it. As a contrast function or linear probe, we could propose

$$\xi(t) = \cos[2\pi(t - 64.5)/365],$$

where the shift value of 64.5 is defined by finding the low point in the mean precipitation profile, marking out empirically mid-winter. The inner product of the regression coefficient function with this probe,

$$\int_0^{365} \xi(t)\beta_6(t)\, dt = 2.32,$$

in effect accumulates information across the entire year about the difference between the summer and winter influence in temperature. Using the techniques described in Section 14.4, we can also work out the sampling standard error of this quantity, which in this case works out to 0.77. Taking ratio of the probe value to its standard error, we obtain $z = 3.0$. It is fairly reasonable to interpret this as a standard normal value under the null hypothesis of no difference in influence between summer and winter, and the value that we obtain appears to be inconsistent with this null hypothesis. It seems appropriate to declare that temperature has a small but statistically significant capacity to predict the log precipitation mean in mid-winter. We can conclude that, if the mean temperature residual for a weather station is high in winter, as it would be for marine stations like Prince Rupert, then precipitation will also be high for that station relative to other stations within the same climate zone.

## 14.3   Long-term and seasonal trends in the nondurable goods index

The nondurable goods manufacturing index, introduced in Chapter 1 and displayed in Figure 14.4 from 1952 to 2000, is a single long time series with a typical multiresolution structure. The global trend across these years is rather linear over large sections after logarithmic scaling. On a shorter
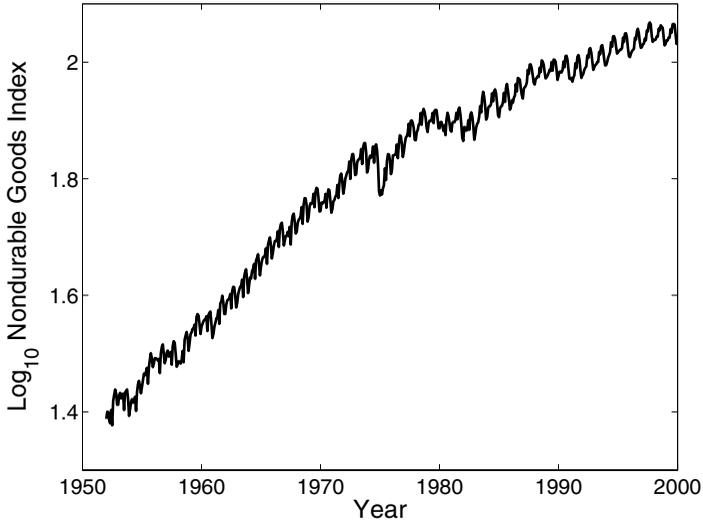
Figure 14.4. The United States nondurable goods manufacturing index plotted in logarithmic coordinates over the years 1952 to 2000.
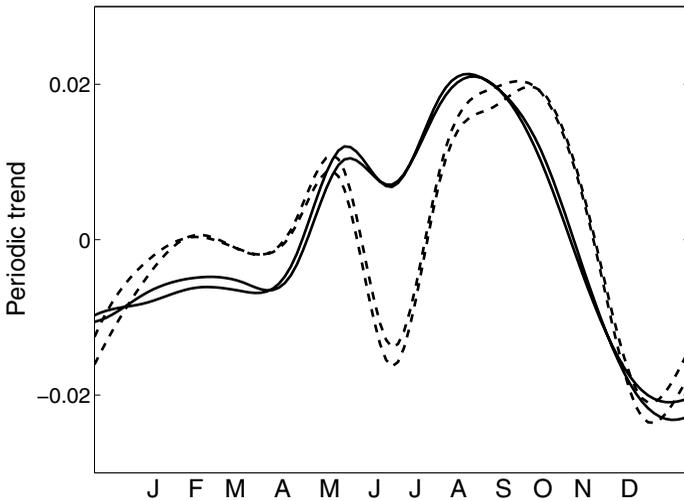


Figure 14.5. Four seasonal cycles for the logged United States nondurable goods manufacturing index are plotted with any overall linear trend removed. Two cycles in the 60's are plotted as dashed lines, and two cycles in the 90's as solid lines.
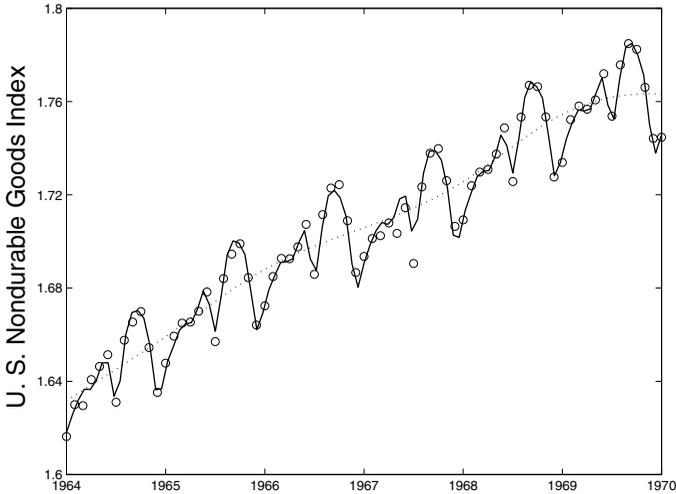
Figure 14.6. The fit to the smoothed logged United States nondurable goods manufacturing index using the point-wise linear model for six typical years. The points indicate the monthly values of the smoothed index, and the solid line is the fit based on the point-wise linear model. The dotted line indicates the estimated smooth nonseasonal trend.

scale, however, we see shocks to the system such as the end of the Vietnam War in 1974, and these seem to result in long-term changes in trend.

Moreover, like most economic indicators, there is a somewhat complex seasonal trend, and this is illustrated for four fairly representative years in Figure 14.5. There are three cycles evident in most years, separated by the Easter/Passover, summer school, and Christmas holidays, respectively.

The seasonal behavior seems to be fairly stable from one year to the next, but exhibits longer-term changes. The large autumn cycle shows a phase shift between the 60's and 90's, but there is little change in amplitude. The small winter cycle is much smaller in the 90's, but the dip due to the summer holidays is much more profound in the 60's.

We can use the point-wise linear model to separate out the smooth long-term trend from the seasonal trend, and at the same time show how the seasonal trend evolves. Our objective here is also to showcase the analysis of a single long time series rather than shorter but replicated series. This analysis used the 577 monthly values in the years 1952 to 2000. The original values were first smoothed by a smoothing spline with curvature penalized with a smoothing parameter value $\lambda = 10^{-6}$, and the smoothed version had a degrees of freedom equivalent of about 521.

The first covariate function $z_1$ is simply the constant function, and it is multiplied by a regression coefficient function $\beta_1$ that was expanded in
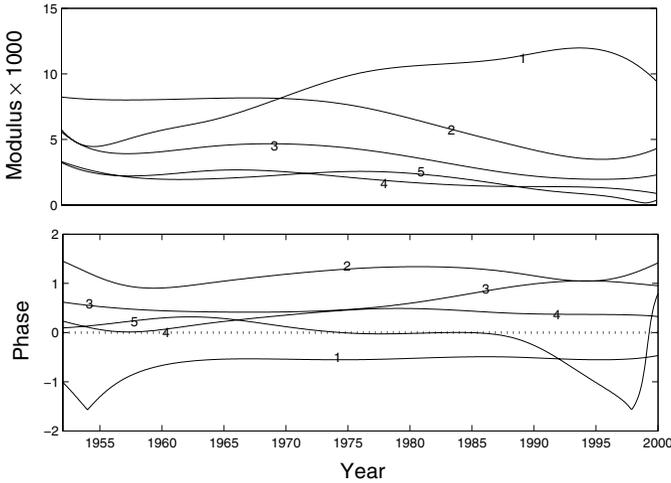
Figure 14.7. The evolution of seasonal trend in the logged United States non-durable goods manufacturing index. The top panel shows the modulus of the five sine/cosine pairs with the frequencies indicated in years. The bottom panel shows the phase for each pair, indicated as an angle in radians between $-\pi$ and $\pi$.

terms of cubic B-splines with knots placed at each year. This knot spacing is designed to allow $\beta_1$ to show smooth trend, but is too coarse to accommodate any seasonality. To further ensure that $\beta_1$ is sufficiently smooth, we penalized curvature with a smoothing parameter $\lambda = 0.01$.

An additional 10 covariate functions $z_2, \ldots, z_{11}$ were set up as a series of sine/cosine pairs with periods 1, 1/2, 1/3, 1/4 and 1/5 years, respectively. These are intended to model periodic seasonal effects. The corresponding $\beta_j$'s were expanded in terms of seven B-spline basis functions with equal knot spacing, and these coefficient functions permit us to see any smooth changes in the structure of this seasonal tend.

Figure 14.6 shows the fit to the smoothed logged goods index by this model for the years 1964 to 1970 along with the smooth nonseasonal trend estimated by $\beta_1$. We see that the fit, based on 121 parameters and some smoothing, is quite good, and certainly captures the seasonal trend in a reasonable way. The turbulent few years in the mid-seventies are not shown, but the fit was not so good there, naturally, since we only allowed for rather smooth seasonal evolution.

How does the seasonal trend evolve? The top panel of Figure 14.7 shows the amplitude or modulus

$$\texttt{Mod}_j(t) = \sqrt{\beta_j^2(t) + \beta_{j+1}^2(t)}$$

of the sine/cosine pairs corresponding to $j = 2, 4, \ldots, 10$ of increasing frequency. The conclusion seems fairly clear: In later years, more of the seasonality is represented by the lowest harmonic with frequency of one year, and the energies in the higher frequency components tend to decline. Seasonal variation is tending to smooth out with time, perhaps due to the effects of automation of production, and the shifting of manufacturing with large seasonality to off-shore locations. The bottom panel shows the phase angle, measured in radians,

$$\texttt{Phase}_j(t) = \arcsin[\beta_j(t)/\texttt{Mod}_j(t)] \ .$$

Here we see little evolution, as we would expect since the timing of the cycles is tied to holidays, in the case of summer and Christmas at least, whose timing is fixed. We no doubt could have done better if we had allowed for the variable timing of the Easter/Passover holiday.

## 14.4   Computational issues

We have $q$ covariate functions $z_{ij}$, each multiplied by its regression coefficient function $\beta_j$. Our concurrent multiple regression model is

$$y_i(t) = \sum_{j=1}^{q} z_{ij}(t)\beta_j(t) + \epsilon_i(t) \ . \tag{14.3}$$

Let the $N$ by $q$ functional matrix $\mathbf{Z}$ contain these $z_{ij}$'s, and let the vector coefficient function $\boldsymbol{\beta}$ of length $q$ contain each of the regression functions. The concurrent functional linear model in matrix notation is then

$$\mathbf{y}(t) = \mathbf{Z}(t)\boldsymbol{\beta}(t) + \boldsymbol{\epsilon}(t) \ , \tag{14.4}$$

where $\mathbf{y}$ is a functional vector of length $N$ containing the response functions.

   We estimate a basis function expansion for each regression function $\beta_j, j = 1, \ldots, q$ along with roughness penalties to control the smoothness of the estimates for the $\beta_j$'s. We must allow for both the basis and the roughness penalty to vary from one $\beta_j$ to another; some regression functions may be assumed to only pick up very smooth effects requiring only a few basis functions, while others may be required to model high-frequency variability in the data. This means that we will have to possibly define a roughness penalty

$$\texttt{PEN}_j(\beta_j) = \lambda_j \int [L_j\beta_j(t)]^2 \, dt$$

separately for each regression coefficient function. Each penalty is defined by choosing a linear differential operator $L_j$ that is appropriate for that functional parameter, such as the curvature operator $L_j = D^2$ or the harmonic acceleration operator $L_j = (2\pi/365)^2 D + D^3$.

The weighted regularized fitting criterion is

$$\mathrm{LMSSE}(\boldsymbol{\beta}) = \int \mathbf{r}(t)' \mathbf{r}(t)\, dt + \sum_{j}^{p} \lambda_j \int [L_j \beta_j(t)]^2\, dt, \qquad (14.5)$$

where

$$\mathbf{r}(t) = \mathbf{y}(t) - \mathbf{Z}(t)\boldsymbol{\beta}(t).$$

Let regression function $\beta_j$ have the expansion

$$\beta_j(t) = \sum_{k}^{K_j} b_{kj}\theta_{kj}(t) = \boldsymbol{\theta}_j(t)'\mathbf{b}_j(t)$$

in terms of $K_j$ basis functions $\theta_{kj}$. In order to express (14.4) and (14.5) in matrix notation referring explicitly to these expansions, we need to construct some composite or super matrices.

Defining

$$K_\beta = \sum_{j}^{q} K_j,$$

we first construct vector $\mathbf{b}$ of length $K_\beta$ by stacking the vectors vertically, that is,

$$\mathbf{b} = (\mathbf{b}_1', \mathbf{b}_2', \ldots, \mathbf{b}_q')'.$$

Now assemble $q$ by $K_\beta$ matrix function $\boldsymbol{\Theta}$ as follows:

$$\boldsymbol{\Theta} = \begin{bmatrix} \boldsymbol{\theta}_1' & 0 & \cdots & 0 \\ 0 & \boldsymbol{\theta}_2' & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\theta}_q' \end{bmatrix}. \qquad (14.6)$$

We can now express model (14.4) as

$$\mathbf{y}(t) = \mathbf{Z}(t)\boldsymbol{\Theta}(t)\mathbf{b} + \boldsymbol{\epsilon}(t). \qquad (14.7)$$

Note that the model can be formally transformed to a *constant coefficient linear model* by defining $N$ by $K_\beta$ functional matrix $\mathbf{Z}^*(t)$ as

$$\mathbf{Z}^*(t) = \mathbf{Z}(t)\boldsymbol{\Theta}(t)$$

so that

$$\mathbf{y}(t) = \mathbf{Z}^*(t)\mathbf{b} + \boldsymbol{\epsilon}(t). \qquad (14.8)$$

This doesn't really gain anything computationally since we achieve constant coefficients at the price of going from $q$ covariates to the greatly expanded number of $K_\beta$ covariates.

But this formalism (14.8) makes clear that the functional linear model has $K_\beta$ parameters. If each of the $Y_i$ response functions is expanded in

terms of $K_y$ basis functions, then the total number of degrees of freedom for error $df_e$ in the model becomes

$$df_e = NK_y - K_\beta \ .$$

Keeping these numbers in mind helps us to avoid over-fitting the data, an ever-present hazard in the world of functional data analysis. We will show in a couple of chapters that all of the functional linear models considered in this book can be re-expressed in this constant coefficient form (14.8).

In order to take care of the roughness penalties, we also need to arrange the order $K_j$ roughness penalty matrices multiplied by their respective smoothing parameters,

$$\mathbf{R}_j = \lambda_j \int \boldsymbol{\theta}_j(t) \boldsymbol{\theta}'_j(t) \, dt \ ,$$

into the symmetric block diagonal matrix $\mathbf{R}$ of order $K_\beta$:

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{R}_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \mathbf{R}_q \end{bmatrix} . \tag{14.9}$$

We can now write down the normal equations weighted least squares solution for the composite coefficient vector $b$:

$$[\int \boldsymbol{\Theta}'(t)\mathbf{Z}'(t)\mathbf{Z}(t)\boldsymbol{\Theta}(t) \, dt + \mathbf{R}]\mathbf{b} = [\int \boldsymbol{\Theta}'(t)\mathbf{Z}'(t)\mathbf{y}(t) \, dt] \ . \tag{14.10}$$

The amount of numerical integration involved in these expressions is really quite manageable. The scalar functions

$$\omega_{j\ell}(t) = \sum_i^N z_{ij}(t) z_{i\ell}(t)$$

play the role of *weighting functions* for the functional inner products

$$\int \boldsymbol{\theta}_j(t) \boldsymbol{\theta}'_\ell(t) \omega_{j\ell}(t) \, dt, j, \ell = 1, \dots, q \ .$$

Similarly, on the right side, we have a set of inner products of the basis functions $\boldsymbol{\theta}_j$ with the unit function $\mathbf{1}$ weighted by the scalar functions $\sum_i^N z_{ij}(t) y_i(t)$. Computing these inner products by numerical integration is a fairly routine procedure.

## 14.5 Confidence intervals

In order to compute confidence intervals, we also have to explicate the role of the coefficient matrix $\mathbf{C}$ in the basis function expansions of the response

functions, expressed as $y = \mathbf{C}\boldsymbol{\phi}$, where the basis function vector $\boldsymbol{\phi}$ is of length $K_y$. This results in

$$
\begin{aligned}
\hat{b} &= \left[\int \boldsymbol{\Theta}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\Theta} + \mathbf{R}\right]^{-1}\left[\int \boldsymbol{\Theta}'\mathbf{Z}'\mathbf{C}\boldsymbol{\phi}\right] \\
&= \left[\int \boldsymbol{\Theta}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\Theta} + \mathbf{R}\right]^{-1}\left[\int \boldsymbol{\phi}' \otimes (\boldsymbol{\Theta}'\mathbf{Z}')\right]\text{vec}(\mathbf{C}) . \qquad (14.11)
\end{aligned}
$$

Here the $K_\beta$ by $K_y N$ composite matrix $\boldsymbol{\phi}' \otimes (\boldsymbol{\Theta}'\mathbf{Z}')$ has the structure

$$
\begin{bmatrix}
\phi_1 \boldsymbol{\theta}_1 \mathbf{Z}_1' & \cdots & \phi_{K_y} \boldsymbol{\theta}_1 \mathbf{Z}_1' \\
\vdots & \cdots & \vdots \\
\phi_1 \boldsymbol{\theta}_q \mathbf{Z}_q' & \cdots & \phi_{K_y} \boldsymbol{\theta}_q \mathbf{Z}_q'
\end{bmatrix},
$$

where the vector function $\mathbf{Z}_j$ is the $j$th column of $\mathbf{Z}$. Recall that in this expression $\phi_k$ is a scalar basis function, whereas $\boldsymbol{\theta}_j$ is a basis function vector of length $K_j$.

Here again, the numerical integration can be reduced considerably when the $j$th covariate has the expansion $\mathbf{Z}_j = \mathbf{D}_j \boldsymbol{\psi}_j$. In this event, $\boldsymbol{\phi}' \otimes (\boldsymbol{\Theta}'\mathbf{Z}')$ is

$$
\begin{bmatrix}
\phi_1 \boldsymbol{\theta}_1 \boldsymbol{\psi}_1' \mathbf{D}_1' & \cdots & \phi_{K_y} \boldsymbol{\theta}_1 \boldsymbol{\psi}_1' \mathbf{D}_1' \\
\vdots & \cdots & \vdots \\
\phi_1 \boldsymbol{\theta}_q \boldsymbol{\psi}_q' \mathbf{D}_q' & \cdots & \phi_{K_y} \boldsymbol{\theta}_q \boldsymbol{\psi}_q' \mathbf{D}_q'
\end{bmatrix} .
$$

We see in this expression that we need inner products $\langle \boldsymbol{\theta}_j, \boldsymbol{\psi}_\ell' \rangle$ with weighting functions $\phi_k$.

Finally, the matrix representing the mapping `C2BMap` that we need to put together the mapping `Y2RMap` to construct confidence intervals is

$$
\texttt{C2BMap} = \left[\int \boldsymbol{\Theta}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\Theta} + \mathbf{R}\right]^{-1}\left[\int \boldsymbol{\phi}' \otimes (\boldsymbol{\Theta}'\mathbf{Z}')\right] . \qquad (14.12)
$$

## 14.6   Further reading and notes

Models that are closely related to the point-wise linear model have been considered by a number of authors. West, Harrison and Migon (1985) investigated what was essentially model (14.3), but with the restriction that the regression coefficient functions $\beta_j(t)$ have a simple autoregressive time series structure. They referred to this structure as a *dynamic generalized linear model,* and went on to consider various extensions in West and Harrison (1989).

Hastie and Tibshirani (1993) looked at a version of this model within what they called *varying coefficient* models of the form

$$
y_i = \sum_j \beta_j(R_{ij}) z_{ij} + \epsilon_i. \qquad (14.13)
$$

They explored various strategies for obtaining flexible estimates of the functions $\beta_j$s, including the use of spline basis expansions with roughness penalties. This paper, as well as the work of West, et al (1985, 1989), contain many interesting examples and illustrate the principle that a number of estimation strategies can be developed for models like these. The discussions associated with the two journal articles cited here also contain many useful alternative perspectives.

The varying coefficient model has subsequently received a lot of attention, with much of this devoted to estimation of smooth regression functions by kernel smoothing (Wu, Chiang and Hoover (1998)), local polynomial smoothing (Fan, Yao, and Cai (2003); Neilsen, Nielsen and Joensen, Madsen and Holst (2000); Zhang and Lee (2000); Zhang, Lee and Song (2002)) and local maximum likelihood estimation (Cai, Fan and Li (2000); Cai, Fan and Yao (2000); Dreesman and Tutz (2001)). Gelfand, Kim, Sirmans and Banerjee (2003) used a Bayesian model for spatial variation in regression coefficients.

While the varying coefficient model certainly involves one or more functional parameters, the data involved are more typically multivariate rather than functional. In many applications, the argument variable $r_j$ for $\beta_j$ is a spatial dimension, and the corresponding covariate $\mathbf{z}_j$ is fixed rather than varying over some argument. From this perspective, the varying coefficient model is closer to the *generalized additive model* (Hastie and Tibshirani, 1990).

It is likely, though, that the techniques associated with varying coefficient problems will prove useful in functional data settings as well. This is especially evident in Eubank, Muñoz Maldonado, Wang and Wang (2004), where the model being investigated is essentially the concurrent functional linear model.