Regression Analysis for a Functional Response

Julian J. FARAWAY

Department of Statistics
University of Michigan
Ann Arbor, MI 48109
(faraway@umich.edu)

Functional responses are encountered when units are observed over time. Although the whole function itself is not observed, a sufficiently large number of evaluations, as is common with modern recording equipment, are assumed to be available. Functional regression analysis relates the smooth functional response, $y(t)$, to known covariates, x, by a linear combination of parameter functions, $\beta(t)$, which are to be estimated. The model takes the standard form, $y(t) = x^T \beta(t) + \varepsilon(t)$. This approach provides an alternative to standard longitudinal data methods used in the biological sciences, where less and noisier data necessitate parametric modeling. The methodology is illustrated by an application in ergonomics.

KEY WORDS: Curve estimation; Ergonomics; Functional data analysis; Longitudinal data analysis; Nonparametric regression; Repeated measures.

Functional regression analysis is designed for the situation in which the ith response is a smooth real function $y_i(t), i = 1, \ldots, n, t \in \mathcal{T}$, with associated covariate vector x_i, which is constant in time throughout this article. Of course, it is only possible to observe the function $y_i(t)$ at a finite number of points, possibly with error, but it may be helpful and more natural to consider the response to be a function.

A well-known technique for analyzing data in which individuals are measured repeatedly through time is called longitudinal data analysis; see, for example, Diggle, Liang, and Zeger (1995). This methodology is mainly applied in the biological sciences, in which data are characterized by relatively few measures per individual and high variation in the measurement, possibly partly due to unmeasured time-varying covariates. This means that the individual's progress or curve can only reasonably be approximated by a simple parametric form, often linear, because there are insufficient data to try more complex models. The effect of various covariates on, say, the slope of line for an individual may be assessed, after possibly specifying a parameterized covariance matrix for the errors, using standard methods of parametric inference. Other methods avoid the necessity of assuming a parametric form for the individual curves by having a separate parameter for each time point at which observations are recorded. These methods require that the points of observation be the same for each individual and fail to take advantage of the smoothness that might be supposed to exist in the individual curves.

Imagine that we have many more accurate and well-controlled measurements for each individual or unit. In the physical and engineering sciences, real-time recording devices can easily produce many accurate measurements for a unit over time. Furthermore, suppose that examination of the data for each unit reveals no obvious parametric form and that the shape of the curves varies according to the covariates. In the absence of strong theory about what parametric form for the curves would be appropriate and with a relatively large number of observations per unit, it is not reasonable to make strong parametric assumptions. Instead, one might merely assume that the underlying curves are smooth and then ask how these curves might be modeled as a function of the covariates.

Thus functional regression analysis is not intended as a competitor to longitudinal data analysis—it augments it. It fills a gap in current methodology to better handle cases in which many observations per unit are recorded. All we need for functional regression analysis to be viable is that a sufficiently large amount of data should be available so that the function can be adequately approximated. When measurement error is small, interpolation may be adequate, but otherwise nonparametric regression or smoothing techniques can be used to estimate the function. Exactly how much data is enough depends on the amount of noise in the measurement and the smoothness of the functions. A further use of functional regression analysis is as a precursor to parametric longitudinal models. An analogy is the alternat-

ing condition expectation method of Breiman and Friedman (1985) or the generalized additive models of Hastie and Tibshirani (1990) that have been used to suggest suitable parametric forms for generalized linear models. A further advantage of functional regression analysis is that it possibly provides a more elegant theoretical basis to longitudinal methods, just as the matrix-based approach is cleaner for standard regression.

Functional data analysis is a new and relatively unexplored area. Much of the prior work concentrates on a study of variation, whereas I shall be more interested in building predictive models. Besse and Ramsay (1986) described the principal-components analysis of functional response data. Ramsay and Dalzell (1991) laid out some general ideas on functional data analysis. The functional data were separated into a structural and a residual component. Functional analysis was used to describe how this should be done. They also discussed various other types of regression models involving functional responses. Rice and Silverman (1991) concentrated on the study of functional variation after the removal of the mean structure. An important difference between the approach of Ramsay and Dalzell (1991) and Rice and Silverman (1991) is how and when the various curves should be smoothed. Ramsay, Wang, and Flanagan (1995) exemplified and extended the approach of the former, whereas Pezulli and Silverman (1993) and Silverman (1995a, b), developed the ideas of the latter. Leurgans, Moyeed, and Silverman (1993) considered canonical correlation analysis when the data are curves.

Methodology is laid out in Section 1, and its application to a problem in ergonomics is illustrated in Section 2. A discussion follows in Section 3.

1. METHOD

1.1 Estimating the Response

In practice, we do not observe $y_i(t), i = 1, \ldots, n$, only $y_i(t_{ij}), j = 1, \ldots, m_i$. The times of measurement will not always be the same for every unit, and there may possibly be noninformative missing values, so the t_{ij} do not need to be the same for each i. Our first step is to attempt to reconstruct the curves $y_i(t)$.

Notice that there are two kinds of variability here, variability specific to the particular observations, $y_i(t_{ij})$, and variability related to the whole curve, $y_i(t)$. The former might be thought to represent measurement error and the latter individual-specific variation. An analogy can be made between the error terms and the random effects in a parametric longitudinal model and these two types of error in our model. Of course, it is not easy to separate these two types of variability, but any smooth estimation is an attempt to eliminate (or at least reduce) variability of the first kind.

A special case of this problem was considered by Hart and Wehrly (1986) in which there are no covariates and the task is just to estimate the common mean of a sample of curves, all observed at the same t_j. This method cannot be used directly in the regression case. For the same problem, Rice and Silverman (1991) used cross-validation to deter-

mine the amount of smoothing, leaving out the whole ith individual instead of the usual method of leaving out just one observation when cross-validating. This method could be directly applied to the regression case, but problems will arise if the form of the chosen regression model is incorrect.

An alternative approach is to smooth each y_i individually without reference to the particular model being fit. Regression analyses typically involve the consideration of several possible models, so model-free smoothing-parameter selection would be advantageous. Methods do exist for selecting the smoothing parameter automatically in the presence of correlated errors; see, for example, Diggle and Hutchison (1989), Altman (1990), or Hart (1991). These authors described ways of overcoming the problem that well-known smoothing-parameter selection methods such as cross-validation behave poorly when the errors are correlated, usually tending to undersmooth. Assumptions need to be made about the form of the covariance structure if these methods are to be successful.

Another issue is that kernel- or spline-based nonparametric regression methods tend not to be robust to outliers. This is an important practical failing of these methods, especially because the smoothing step is an attempt to filter out aberrant observations. In such cases a robust method such as LOWESS (Cleveland 1979) should be used. Because automatic methods of smoothing-parameter selection are not consistently reliable, especially when unexpected aberrations occur, I recommend that the user select the smoothing parameter by eye, using contextual knowledge of how smooth the functions are expected to be. It is possible that there might be so much data that such smoothing might become tiresome, but typically the cost of obtaining the data outweighs such concerns. It is best to err on the side of undersmoothing to avoid missing features of interest.

One final idea is not to smooth at all—just interpolate the observed $y_i(t_{ij})$ to get estimates of $y_i(t_j)$. Because the $\hat{\beta}(t)$ is a weighted average, it will tend to be smoother than the interpolated $y_i(t_{ij})$. Furthermore, there are situations, particularly in the physical sciences, in which the measurement error is extremely small so that there really is no need to smooth. Even where the measurement error is appreciable and so better results will usually be obtained by some smoothing, this interpolation approach provides a useful estimate of $\beta(t)$ because these can then be smoothed visually if an exploratory result is all that is desired.

1.2 Estimation and Prediction

Suppose that the functional responses, $y_i(t)$, arise from the model $y = X\beta + \varepsilon$, where β is a vector of functions $(\beta_1(t), \ldots, \beta_p(t))^T$ and X is the familiar $n \times p$ design matrix formed from the p-vector valued covariates $x_i, i = 1, \ldots, n$. As in scalar (i.e., scalar y) regression, the first column of X will usually all be ones, and categorical predictors can be handled by assigning appropriate dummy variables. Various transformations of the predictors can be incorporated in X as in scalar regression. Moreover, y is a vector of response functions $(y_1(t), \ldots, y_n(t))^T$ and ε is a vector of error functions $(\varepsilon_1(t), \ldots, \varepsilon_n(t))^T$. I assume that each $\varepsilon_i(t)$

is an independent realization of a stochastic process with mean 0 and covariance function $\gamma(s, t), s, t \in \mathcal{T}$. So I allow for correlated errors within individuals but not between individuals.

I will choose $\hat{\beta}$ to minimize $\sum_{i=1}^{n} \|y_i - x_i^T \beta\|^2$, where $\|\cdot\|$ is the L_2 norm on $\mathcal{T}$. Certainly other choices of norm are possible, including a weighted L_2 norm (although that would not change the following $\hat{\beta}$) or, say, an L_1 norm if a more robust estimate was required (although the need for robust estimation should be reduced by the smoothing discussed previously).

By considering each $t \in \mathcal{T}$ separately and provided that X has full rank as in the usual regression situation, it is clear that the solution is $\hat{\beta} = (X^T X)^{-1} X^T y$. Let $\hat{y} = X\hat{\beta}$ and $\hat{\varepsilon} = y - \hat{y}$.

Given a new value of the covariates x_0, we predict the response as $\hat{y}_0 = x_0^T \hat{\beta}$. Pointwise standard errors for $\hat{y}_0(t)$ and β can be computed using the usual regression formula. Weights may be incorporated in the obvious way.

The $\hat{\beta}(t)$ can now be plotted, which might help suggest a parametric form for a standard longitudinal model. The residuals could also be examined to help suggest a parametric form for the covariance. Alternatively, we can deal with the functional regression model in its own right.

Now it may be observed that if we consider the response to be the vector consisting of the values of $y_i(t)$ on the grid of timepoints, then the preceding estimation procedure is identical with multivariate multiple regression provided we only require estimates of β or predictions on that same grid of timepoints. There are two advantages to taking the functional regression approach, however. The first is conceptual—the underlying quantities are curves and not vectors. This fact is crucial in the construction of the $y_i(t)$ when the times of measurement are not coincident across individuals. It is also used in the plotting of the results and the determination of appropriate regression diagnostics. The second reason is even more compelling—the inferential techniques used in multivariate multiple regression break down when used on curve data because of the high dimension of the approximating vectors. Details are given in the next section.

1.3 Inference

Consider the comparison of two nested linear models, ω and Ω, where $\dim(\Omega) = p$ and $\dim(\omega) = q$. The model ω results from a linear restriction on the parameters of Ω.

1.3.1 Multivariate Analysis-Based Methods. Because $y(t)$ is approximated by a vector, the function evaluated on a grid of points, it is natural to look to multivariate multiple regression analysis (see Anderson 1984; Johnson and Wichern 1992) for ideas on how to test the null hypothesis that ω holds against the alternative of Ω. Although making the grid finer does not increase the amount of information because the data size is fixed, we can make the grid as fine as we like, so any test statistics we use should make sense as the mesh size decreases.

Suppose we evaluate $y_i(t)$ on a grid of points t_j, where $j = 1, \ldots, m$. Typically we would want the t_j to be equally spaced in $\mathcal{T}$. Now we can apply standard multivariate multiple regression with response vector y_i' and covariates x_i. The multivariate linear model is

$$y'_{(n \times m)} = X_{(n \times p)} \beta'_{(p \times m)} + \varepsilon_{(n \times m)},$$

where $E\varepsilon_i = 0$ and $\text{cov}(\varepsilon_i) = \Sigma$. Now $\hat{\beta}' = (X^T X)^{-1} X^T y'$, and estimates of $\hat{\beta}(t)$ can then be made by interpolating the rows of $\hat{\beta}'$. We can make m as large as we like because the estimates and their standard errors are computed pointwise. The validity of this method of estimating $\hat{\beta}$ follows from Ramsay and Dalzell (1991). Difficulties arise, however, with the choice of m when we want to make inference about the whole curves, $\hat{\beta}_j(t)$, or test hypotheses comparing nested regression models.

The likelihood ratio test statistic of the null H_ω versus the alternative H_Ω is

$$U \propto \log \frac{|\hat{\Sigma}_\Omega|}{|\hat{\Sigma}_\omega|} = \sum_{j=1}^{m} \log \frac{\lambda_j^\Omega}{\lambda_j^\omega},$$

where λ_i^Ω and λ_i^ω are the ordered eigenvalues of $\hat{\Sigma}_\Omega$ and $\hat{\Sigma}_\omega$, respectively.

Now let's see how the λ_i's relate to the covariance function λ. Suppose the eigenfunction decomposition of underlying covariance function $\gamma(s, t)$ is

$$\gamma(s, t) = \sum_{i=1}^{\infty} \gamma_i \phi_i(s) \phi_i(t)$$

so that a randomly selected curve from the population can be written as

$$Y(t) = X\beta(t) + \sum_{i=1}^{\infty} \xi_i \phi_i(t), \qquad (1)$$

where the ξ_i are uncorrelated random variables with zero means and variances γ_i. Now, assuming that the model is correct, the eigenvalues, λ_i, and eigenvectors of $\hat{\Sigma}$ can be used to estimate γ_i and $\phi_i(t)$. Consistency as $m \to \infty$ follows from Dauxois, Pousse, and Romain (1982), although in their result m represents the number of observations per unit, not the grid size. Typically, only the first few eigenvalues are of any size so that the residual variation can be represented by just a few terms of (1). The remaining eigenfunctions, which are approximated by the eigenvectors, represent directions in which there is much less variation. The likelihood ratio statistic, however, depends on terms $\log(\lambda_j^\Omega/\lambda_j^\omega)$, which do not become small as j becomes large. This means that the test statistic will become dominated by terms that represent unimportant directions of variation as $m \to \infty$. Although the likelihood ratio statistic may be fine for multivariate multiple regression for small m, it is not a suitable basis for comparing functional regression-analysis models.

Other tests such as the Lawley–Hotelling trace criterion, the Bartlett–Nanda–Pillai trace criterion, and Roy's maximum root are subject to the same objection in that they become dominated by variation in unimportant directions.

Now these multivariate multiple regression tests could be used if we kept m rather small, but that requires believing that a rather coarse grid (maybe just three or four points) could represent the function. This does not seem reasonable. One might consider the possibility of tests based on just the leading eigenvalues, but then the question of how many leading values and how to weight them will arise. Various shrinkage or reduced-rank-type methods might also be attempted.

1.3.2 Bootstrap-Based Testing Methods. An alternative is to consider the sum of the eigenvalues of $\hat{\Sigma}_\omega - \hat{\Sigma}_\Omega$ because this will be dominated by the directions of greatest difference. This test statistic is proportional to

$$T = \frac{1}{m} \sum_{i=1}^{n} \sum_{j=1}^{m} (\hat{y}_{ij}^\omega - \hat{y}_{ij}^\Omega)^2$$

$$\rightarrow \sum_{i=1}^{n} \int_T (\hat{y}_i^\omega(t) - \hat{y}_i^\Omega(t))^2 \, dt \quad \text{as} \quad m \rightarrow \infty.$$

This test statistic makes sense as m becomes large in contrast to the preceding statistics. Some weighting might be appropriate when the variance of $y(t)$ varies considerably over T. By making some assumptions about the error structure, it would be possible to derive the null distribution of this test statistic, but instead I propose the following bootstrap testing procedure:

1. Compute the test statistic T and the residuals under the null hypothesis: $\hat{\varepsilon}_i^\omega(t)$.
2. Resample with replacement from the residual curves, $\hat{\varepsilon}_i^\omega(t)$, to obtain $\hat{\varepsilon}_i^{\omega*}(t)$ and form the resampled response $y^* = \hat{y}^\omega + \hat{\varepsilon}^{\omega*}$.
3. Refit the null and alternative models using the response y^*, and compute the resampled test statistic T^*.
4. Repeat B times and compute the P value for the test as $\#\{T^* > T\}/B$.

A parametric bootstrap alternative is to assume that the errors are from a Gaussian stochastic process and generate the $\hat{\varepsilon}^{\omega*}$ using (1) by generating the ξ_i as independent normal variates with variance λ_i^ω.

Quite apart from the usual caveats about bootstrap methods, these tests can only be regarded as approximate due to the subjective nature of the smoothing carried out earlier.

If the null hypothesis is rejected, it may be interesting to find out why. Large values in $\sum_{j=1}^{m} (\hat{y}_{ij}^\omega - \hat{y}_{ij}^\Omega)^2$, $i = 1, \ldots, n$, will indicate individual curves that contribute substantially to the rejection, whereas $\sum_{i=1}^{n} (\hat{y}_{ij}^\omega - \hat{y}_{ij}^\Omega)^2$, $j = 1, \ldots, m$, will identify points in time at which the two models may differ.

Simultaneous confidence bands for $\hat{\beta}_j(t)$ and $\hat{y}_0(t) = x_0^T \hat{\beta}(t)$ may also be derived using this bootstrap procedure. Suppose that the function being estimated/predicted is $f(t)$—either one of the preceding—and that the standard error is $s(t)$; then we wish to find a c such that

$$P(\hat{f}(t) - cs(t) \leq f(t) \leq \hat{f}(t) + cs(t), \forall t \in \mathcal{T}) = 1 - \alpha$$

for a $(1 - \alpha) \times 100$ simultaneous confidence interval for $f(t)$. We can construct resampled $f^*(t)$ using the method described previously and let c be the $(1 - \alpha)$th quantile of $\sup_t |f^*(t) - \hat{f}(t)|/s(t)$. This is essentially the method used by Faraway (1990) and compared favorably to other bootstrap confidence-band construction methods by Loader (1993).

1.4 Residual Analysis

After a model has been selected, perhaps with the aid of the testing methods described previously, it is a good idea to examine the residuals to check the adequacy of the chosen model. Three kinds of plots may be useful:

1. Plots of the estimated eigenfunctions and their associated eigenvalues. The eigenfunctions are estimated using the eigenvectors of the estimated residual covariance matrix. These plots show the nature of the unexplained variation in the model, which can sometimes be given a physical interpretation as we will see in our example. The eigenfunctions could also be used to suggest a suitable parametric form for random effects in view of the form of (1).
2. Quantile-quantile (Q-Q) plots of the estimated scores $\hat{\xi}_i$. There will be a plot of n scores for each eigenfunction. Typically most of the variation is represented by the first few eigenfunctions, so only a few of these Q-Q plots will be worthy of examination. The main subject of attention is outliers—these indicate unusual curves rather than aberrant point observations. The same considerations about how to handle outliers apply as in scalar regression. Approximate normality will provide some evidence of a Gaussian stochastic error process should this be an assumption of any inferential technique.
3. For each time point t_j, plot $\hat{y}_i(t_j)$ versus $\hat{\varepsilon}_i(t_j)$. As in scalar regression, we expect to see no relationship if the assumptions of the model hold. For checking constancy of variance (with respect to $\hat{y}$ and not t, where we allow for nonconstant variance), plotting $\hat{y}_i(t_j)$ versus $|\hat{\varepsilon}_i(t_j)|$ is more effective.

Clearly, there is more scope for other plots of the residuals, some of which would be direct analogs of the scalar regression case and others that would be new, using the functional nature of the response and error.

2. EXAMPLE

Ergonomically correct design of equipment is vital to the health and comfort of users. Devising a good design for such equipment is often costly because it involves the production of prototypes. These must be tested by users of different anthropometric dimensions. The prototypes are recursively improved, which can be expensive and time consuming due to the cost of construction and testing. An accurate model of human body motion for various anthropometric dimensions would ease computer-aided design of equipment, avoiding much of the need for the production and testing of prototypes.

As part of a study of the body motions of automobile drivers, researchers at the Center for Ergonomics at the University of Michigan attached markers to various body joints

of a single individual. Four cameras tracked the marker locations as the individual reached to various locations in the car and pressed a button. The coordinates of the markers were sampled at 25 hertz and then converted to angular data. There were a few missing values caused by markers being obscured during some parts of the motion. In this article, I will analyze only the right-elbow-included angle (the angle formed by the shoulder, elbow, and wrist markers) for a single individual. In the full study, about 10 angles are investigated across 36 individuals. There were 3 reaches to each of 20 different targets situated in the glove compartment, center console gear shifter, central instrument panel (where the radio is), and the headliner (above the windshield). The order of the reaches was randomized, and there was an adequate rest time between motions. Only the portion of motion between when the hand leaves the steering wheel and when it reaches the target will be modeled.

The data recorded for each motion vary in length both because of the targets being at different distances from the driver and also because the driver may make the reach at different speeds. The objective now, however, is only to model the shape of the motion and not the speed at which it occurred. Let $y(t)$ represent the angle, and t is rescaled to vary over $[0, 1]$. Thus, t does not represent time, merely the proportion of the motion between the start and the end. For a given motion, $y(t)$ is observed on an equally spaced grid of points, but the number of such points varies from observation to observation. There were a few missing values, but this poses no problem because nonparametric regression methods can estimate the curve at these points.

A plot of the data revealed no unusual single $y_i(t_{ij})$ and apparently very little noise in the observed angle curves. This is to be expected due to the accuracy of the measurement equipment and the smoothness of human motion. Thus, very little smoothing was indicated and the smoothing splines used to estimate $y_i(t)$ were close to being interpolants. The $y(t)$ were computed on an evenly spaced grid of 20 points, which was adequate to represent the curves. The curves are shown in Figure 1.
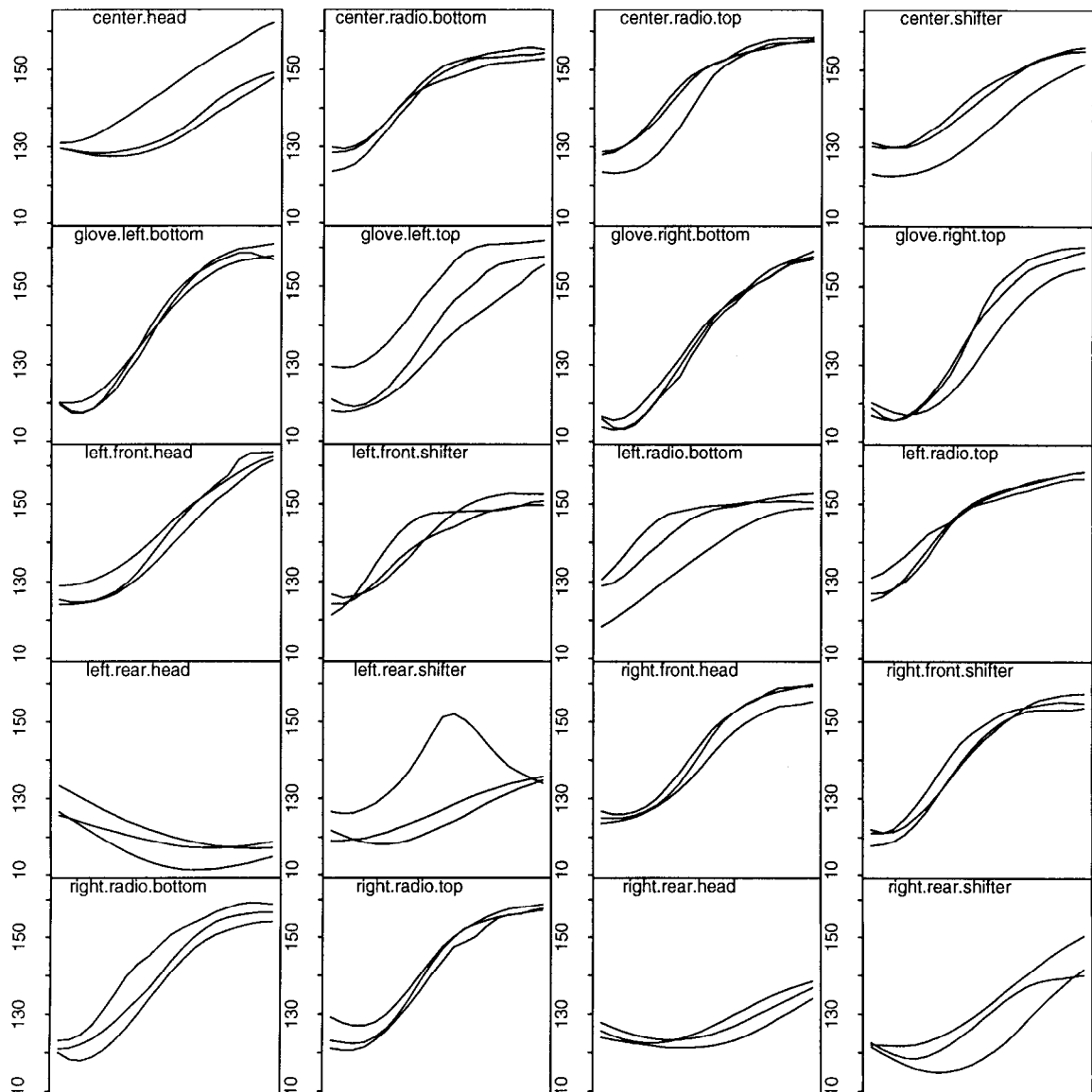

Figure 1. Estimates of the Right-Elbow-Included Angle Curve for 20 Different Locations With 3 Replications per Location. The vertical axis is in degrees.

One of the motions to the left rear shifter location is clearly wrong. Closer investigation of the data for this observation revealed that the subject changed his mind about the target in mid-reach. This observation was discarded from the rest of the analysis.

The purpose of this experiment was to find a model for predicting the motion given the coordinates (c_x, c_y, c_z) of the target. The coordinates of each of the 20 targets in the experiment are known. A linear model

$$y(t) = \beta_0(t) + c_x\beta_x(t) + c_y\beta_y(t) + c_z\beta_z(t) \qquad (2)$$

was fit to the data. To determine whether this represents an adequate fit, I compared it to the model in which we fit a different curve for each target:

$$y(t) = \beta_k(t), \qquad (3)$$

where $k = 1, \ldots, 20$ indicates the target. This model is analogous to standard one-way analysis of variance. Comparing these two models represents a lack-of-fit test for the lin-

ear model. The estimated p value for the test using either bootstrap test described in Section 1.3 with 200 resamples was 0. Thus, the linear model appears not to fit sufficiently well. An examination of the components of the test statistic reveals that the lack of fit is associated with reaches to locations that are extreme in the target space—the left rear headliner in particular. The fit in general also becomes worse approaching the end of the reach.

A quadratic model,

$$\begin{aligned}
y(t) &= \beta_0(t) + c_x\beta_x(t) + c_y\beta_y(t) + c_z\beta_z(t) \\
&\quad + c_xc_y\beta_{xy}(t) + c_yc_z\beta_{yz}(t) + c_zc_x\beta_{zx}(t) \\
&\quad + c_x^2\beta_{x^2}(t) + c_y^2\beta_{y^2}(t) + c_z^2\beta_{z^2}(t),
\end{aligned} \qquad (4)$$

was fit to the data and compared to the model in (3). The bootstrap test described previously gave an estimated p value of .26, whereas the version assuming a Gaussian stochastic process gave .3. This indicates that the quadratic model cannot be rejected as an adequate model for the data.
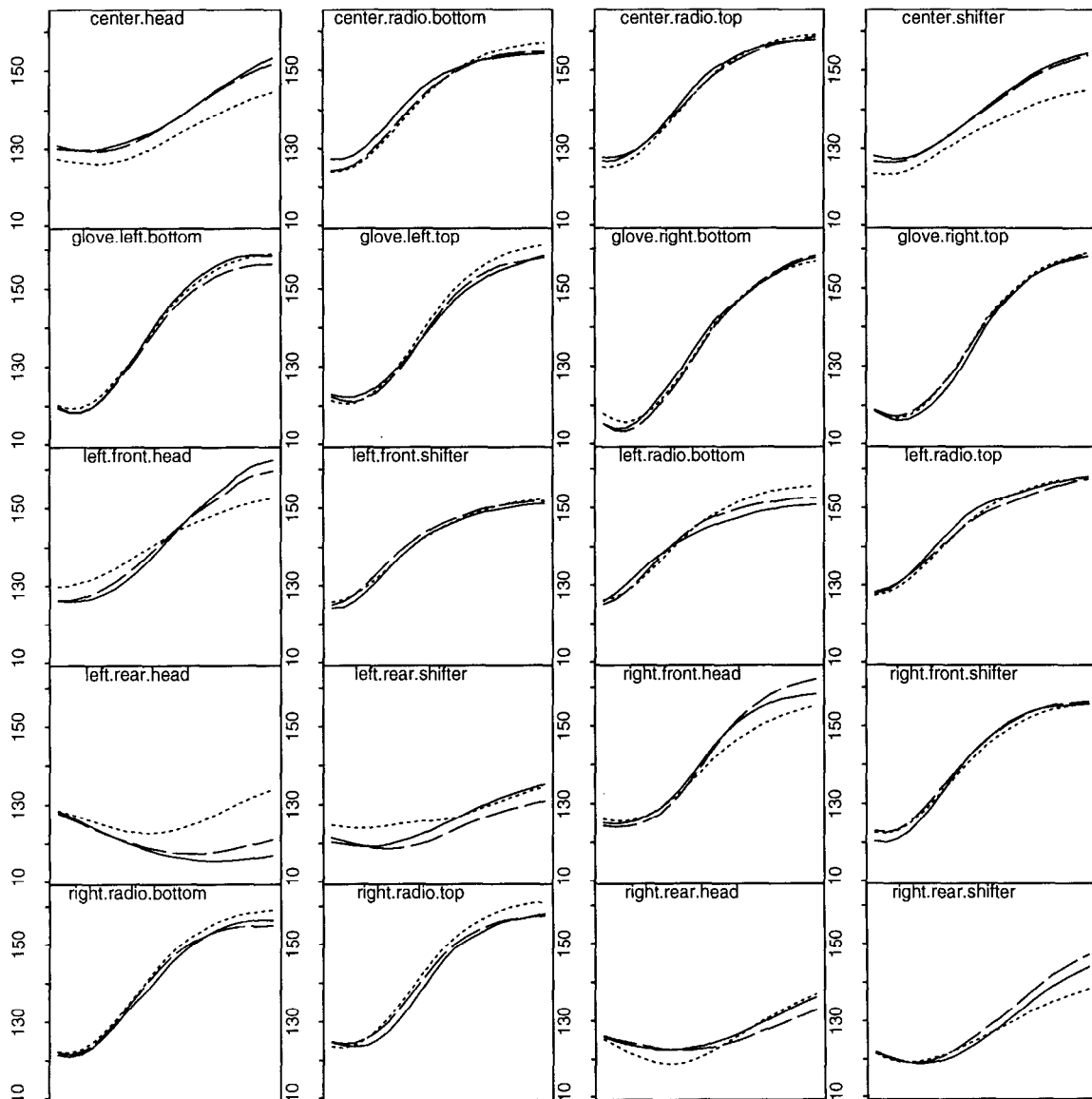

Figure 2. The Fitted Curves for One-Way ANOVA Model (2) (solid line), the Linear Model (1) (dotted line), and the Quadratic Model (3) (dashed line). The vertical axis is in degrees. Note the superiority of the quadratic fit and that the largest discrepancies occur in reaches to extreme locations.
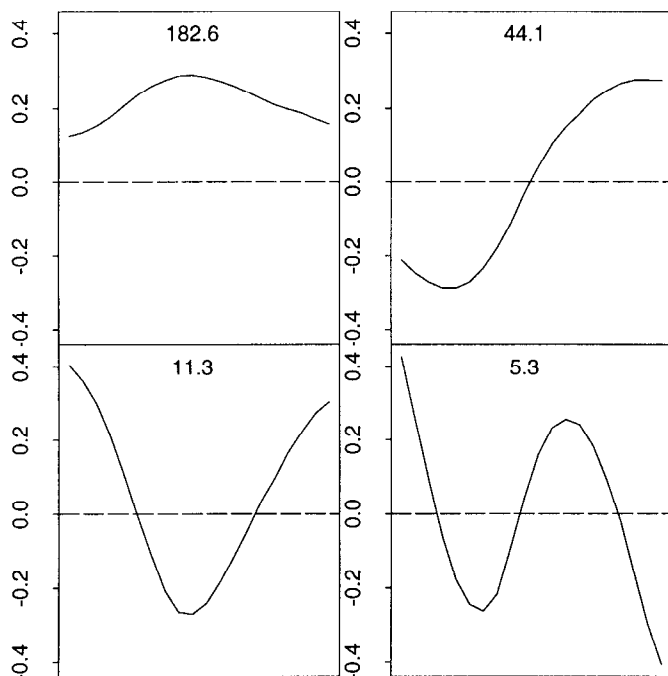
Figure 3. An Eigendecomposition of the Residuals: Estimates of the First Four Eigenfunctions With the Associated Eigenvalues Marked on the Plot.

Of course, with more data and targets, it is likely that the inadequacy of the quadratic model would be revealed because it seems unlikely that the true relationship is exactly quadratic. This model, however, appears to give a good approximation and forms a suitable basis for prediction for reaches within the general range of targets in the experiments. See the fitted curves in Figure 2.

It is possible to obtain a measure of the fit in each of these models. The coefficient of determination, R^2, may be computed pointwise and then integrated across $\mathcal{T}$. The values obtained are 84.0% for the one-way ANOVA model, 68.6% for the linear model, and 80.6% for the quadratic model. The fit was better at the end of the motion, which is not surprising because the targets, which mark the end of the motion, were the covariates.

Having found a prospective regression model, we now examine the residuals. The first four estimated eigenfunctions are shown in Figure 3.

The first eigenfunction shows that variation about the mean is mostly in the form of a vertical shift but with greater variation in the middle of the motion. This can be explained by realizing that the elbow angle will start in roughly the same position for a reach to a given target, and likewise for the endpoint, but may progress at an uneven rate. The marking of the beginning and end of the motion was determined from an examination of the coordinates of the hand. This was difficult to do conclusively, leaving varying amounts of stationary motion at the ends of curve, so variation in this could also explain the first eigenfunction. The second eigenfunction can be interpreted as variation in the start and finish angles of the elbow. The hand is not always in exactly the same position on the wheel or the arm has begun to move before the hand or the attitude of

the arm varies when the target has been reached. The other residuals checks described in Section 1.4 were carried out and revealed nothing abnormal.

The selected quadratic model can now be used for prediction. The new target, x_0, for which a predicted angle curve is required is situated between the radio and the glove compartment. The predicted curve is $\hat{y}_0(t) = x_0^T \hat{\beta}(t)$, and 95% simultaneous confidence bands for the mean response (the average predicted motion to this target) and for a new response are shown in Figure 4. The bands were constructed using the bootstrap method described in Section 1.3. The width of the bands happens to be roughly constant. The critical value was 2.29 based on 999 bootstrap samples, which compares to the 2.01 that would be used for the t-based pointwise confidence bands. Further insight into the nature of the variability in the prediction might be had by adding random errors to $\hat{y}_0(t)$ in the manner of (1).

In this example, I have not shown the $\hat{\beta}(t)$ because they would be difficult to interpret. In other data collected in this study, covariates such as the heights of individuals making the reaches were recorded, and the $\hat{\beta}(t)$ associated with such covariates would be of direct interest.

Two alternative approaches might be considered for modeling this data. Parametric longitudinal analysis could be applied if we can propose suitable parametric models for the mean and covariance of the response. For this particular example, there was no theory to suggest any particular parametric model, so one would have to examine Figure 1 and attempt to divine one. Because the curves do not seem to have any common shape, a flexible model with many parameters would be appropriate. One would then have the difficult task of relating these parameters to the covariates. [An examination of the $\hat{\beta}(t)$ from the functional regression might be helpful in suggesting a suitable parametric model.] This approach is fine for data in which there are few measurements per curve with relatively large variance
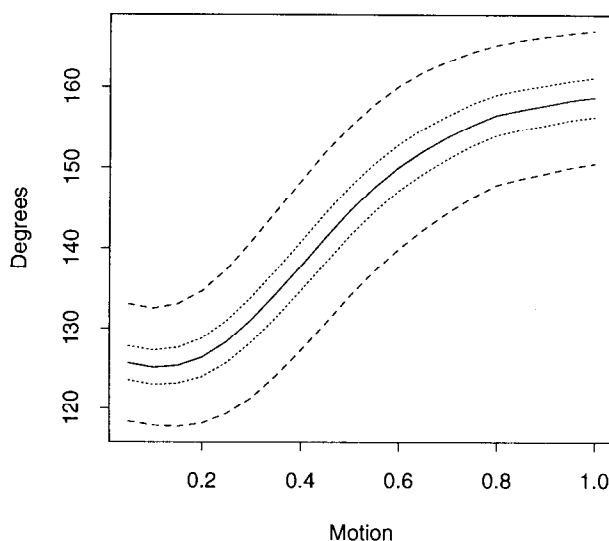

Figure 4. The Predicted Response (solid line) for a New Target Situated Between the Radio and the Glove Compartment. 95% simultaneous confidence bands for the mean response (inner pair) and for a new observation (outer pair) are also shown.

so that simpler parametric forms will work, but it seems to be unwieldy for data of the kind presented here.

An alternative approach is based on multivariate multiple regression. Each response is represented by a vector whose elements lie along a grid on the underlying curve. Because the vectors have to be the same length, the raw data has to be processed using the methods discussed in Section 1.1. The estimates of the parameters will be then be the same as those used in the functional regression on that grid. The prediction of future curves could be made. All inference based on this approach will be suspect, however, for the reasons discussed in Section 1.3.1. Furthermore, consideration of appropriate residual diagnostics derives from the functional representation of the response and not its vector-based approximation.

3. DISCUSSION

There is still a great amount of fundamental development to be done in the area of functional regression analysis. The inferential procedures discussed previously are tentative, and some theoretical investigation would be valuable. In some situations, the covariates x_i will change over time also. An example of this kind was discussed by Ramsay and Dalzell (1991), although, because in that case there are only 12 measurements per unit, some parametric structure needs to be imposed. In other cases, in which there are more observations per unit, a more fully nonparametric approach would be worthwhile. A related problem is the multivariate functional response. The data discussed previously are a subset of the whole experiment in which angles related to other body joints were observed. Because these are clearly related to each other, a multivariate functional regression analysis would be useful.

As mentioned previously, functional regression analysis could be used as an exploratory preliminary to parametric longitudinal data analysis. This would be valuable in smaller datasets in which the observations are unequally spaced, although it remains to be seen how few observations are sufficient for functional regression analysis to be effective.

ACKNOWLEDGMENTS

REFERENCES

Altman, N. (1990), "Kernel Regression Estimation of Data With Autocorrelated Errors," Journal of the American Statistical Association, 85, 749–759.

Anderson, T. (1984), An Introduction to Multivariate Statistical Analysis (2nd ed.), New York: Wiley.

Besse, P., and Ramsay, J. (1986), "Principal Components Analysis of Sampled Functions," Psychometrika, 51, 285–311.

Breiman, L., and Friedman, J. (1985), "Estimating Optimal Transformations for Multiple Regression and Correlation" (with discussion), Journal of the American Statistical Association, 80, 580–619.

Cleveland, W. (1979), "Robust Locally Weighted Regression and Smoothing Scatterplots," Journal of the American Statistical Association, 74, 829–836.

Dauxois, J., Pousse, A., and Romain, Y. (1982), "Asymptotic Theory for the Principal Component Analysis of a Vector Random Function: Some Applications to Statistical Inference," Journal of Multivariate Analysis, 12, 136–154.

Diggle, P., and Hutchison, M. (1989), "On Spline Smoothing With Autocorrelated Errors," Australian Journal of Statistics, 31, 166–182.

Diggle, P., Liang, K., and Zeger, S. (1995), Analysis of Longitudinal Data, Oxford, U.K.: Oxford University Press.

Faraway, J. (1990), "Bootstrap Selection of Bandwidth and Confidence Bands for Nonparametric Regression," Journal of Statistical Computing and Simulation, 37, 37–44.

Hart, J. (1991), "Kernel Regression With Time Series Errors," Journal of the Royal Statistical Society, Ser. B, 53, 173–187.

Hart, J., and Wehrly, T. (1986), "Kernel Regression Estimation Using Repeated Measurements Data," Journal of the American Statistical Association, 81, 1080–1088.

Hastie, T., and Tibshirani, R. (1990), Generalized Additive Models, London: Chapman and Hall.

Johnson, R., and Wichern, D. (1992), Applied Multivariate Statistical Analysis (3rd ed.), Englewood Cliffs, NJ: Prentice-Hall.

Leurgans, S., Moyeed, R., and Silverman, B. (1993), "Canonical Correlation Analysis When the Data Are Curves," Journal of the Royal Statistical Society, Ser. B, 55, 725–740.

Loader, C. (1993), "Nonparametric Regression, Confidence Bands and Bias Correction," in Proceedings of the 25th Symposium on the Interface Between Computer Science and Statistics, pp. 131–136.

Pezulli, S., and Silverman, B. (1993), "Some Properties of Smoothed Principal Components," Computational Statistics, 8, 1–16.

Ramsay, J., and Dalzell, C. (1991), "Some Tools for Functional Data Analysis," Journal of the Royal Statistical Society, Ser. B, 53, 539–572.

Ramsay, J., Wang, X., and Flanagan, R. (1995), "A Functional Data Analysis of the Pinch Force of Human Fingers," Applied Statistics, 44, 17–30.

Rice, J., and Silverman, B. (1991), "Estimating the Mean and Covariance Structures Nonparametrically When the Data Are Curves," Journal of the Royal Statistical Society, Ser. B, 57, 673–690.

Silverman, B. (1995a), "Incorporating Parametric Effects Into Functional Principal Components Analysis," Journal of the Royal Statistical Society, Ser. B, 57, 673–690.

——— (1995b), "Smoothed Principal Components by Choice of Norm," technical report, University of Bristol, Dept. of Mathematics.