Instantaneous Dynamics of Functional Data

by

Jeffrey Bone

B.Sc., The University of Victoria, 2014

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate Studies

(Statistics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

October 2016

Abstract

Time dynamic systems can be used in many applications to data modeling. In the case of longitudinal data, the dynamics of the underlying differential equation can often be inferred under minimal assumptions via smoothing based procedures. This is in contrast to the common technique of assuming a prespecified differential equation, and estimating it's parameters.

In many cases, one wants to learn the dynamics of a differential equation that incorporates more than just one stochastic process. In the following, we propose extensions to existing two-step smoothing methods that allow for the presence of additional functional data arising from a second stochastic process. We further introduce model comparison techniques to assess the hypothesis that there is a significant change in fit provided by this additional process. These techniques are applied to the instantaneous dynamics of mouse growth data and allow us to make comparisons between mice who have been assigned different genetic and physical conditions. Finally, to study the statistical properties of our proposed techniques, we carry out a simulation study based on the mouse growth data.

Preface

This thesis is an original and unpublished work of the author, Jeffrey Bone, under the supervision of Dr. Nancy Heckman.

The research question is an extension of the work done by Nicolas Verzelen, Wenwen Tao and Hans-Georg Müller on stochastic dynamics of functional data. Namely, we propose techniques to include additional stochastic processes in the data-driven differential equation framework and to assess their impact in describing functional data.

Table of Contents

List of Tables

List of Figures

List of Supplementary Materials

- 1D_Plots: Scatterplot and animations of estimated values of $f(t, x)$ for the eight mouse groups at each t.

- 2D_Plots: Contour plot and animations of estimated values of $f(t, x, w)$ for the eight mouse groups at each t.

- Bootstrap: Bootstrap confidence intervals for the observed value of $R^2_W(t)$ for each of the eight mouse groups.

- Contours_1D_Fit: Contour plots of the estimated value of $f(t, x)$ for each of the eight mouse groups.

- Density_Curves: The estimated densities and corresponding observed values of $\hat{S}$ for each of the eight mouse groups.

Acknowledgments

Firstly, I wish to thank NSERC for providing me with partial financial support during my studies and for the NSERC Discovery Grant 7969.

I'd like to extend my gratitude to the Department of Statistics at University of British Columbia for their admittance and support. To the many excellent professors and administrative staff, without you, the opportunities such as the ones presented to myself would not be possible.

To Patrick Carter, thank you for allowing me to use your dataset and taking time to discuss your research and its relation to my work.

To my second reader, Ruben Zamar, I appreciate the time and effort you have spent looking over and critiquing my works. Your valuable feedback has improved the quality of this thesis.

Lastly, but most importantly, I want to extend a huge thank you to my supervisor, Nancy Heckman. Without your support, encouragement, and willingness to engage in this research, it certainly would not have been possible. I feel fortunate to have worked with you over the last two years and I am very grateful for your time and effort.

Dedication

These works are dedicated to my Opa, Mr. Frank Fiederer. There is no greater gift than that of an education.

Chapter 1

Introduction

In functional data analysis (FDA), one generally has data samples consisting of points that are assumed to come from a smooth curve or other infinite dimensional object. This can be thought of as data on N different scatterplots, each corresponding to a curve. Data of this form arise in many applications such as genetic trait modeling (Kirkpatrick and Heckman, 1989), online auction dynamics (Liu and Müller, 2009) and growth studies (Gasser et al., 1984; Verzelen et al., 2012). The standard introduction to FDA is provided in an accessible manner by Ramsay and Silverman (2002, 2005).

A common approach in FDA is to use a prespecified differential equation to model each of the curves (Cao and Ramsay 2007; Ramsay et al. 2007; Liang and Wu 2008; Cao et al. 2012). The parameters of this prespecified differential equation are then estimated for each of the curves. This approach relies on the ability to identify a differential equation that is appropriate for the data before doing any fitting. The curve by curve method is a powerful one in situations when the data are densely observed or when one is interested in the exact dynamics of individual observations. On the other hand, in longitudinal studies where data are repeatedly observed for many subjects, pooling information across curves may improve estimation. Moreover, in situations where the underlying processes are stochastic, the presumed underlying dynamics may not be well understood and thus prescribing a differential equation can lead to poor fits. Examples of such processes can be seen in subject specific studies of viral levels in HIV patients (Miao et al., 2009) and in auction price trajectories (Reddy and Dass, 2006). In these situations, alternative ways of viewing and modeling the data may improve the analysis.

This alternative is to view each curve as a realization of some unknown underlying stochastic process (Yao et al. 2005; Liu and Müller 2009; Müller and Yang 2010; Müller and Yao 2010; Verzelen et al. 2012). From this viewpoint, there are no prespecified dynamics, but rather one tries to learn the underlying dynamics from the data. This approach does not require strong assumptions on the data, often just that the underlying stochastic process admits a Karhunen-Loève expansion (Ash and Gardner, 1975). Much work

has been done in developing techniques for learning the dynamics of these underlying processes. Usually these techniques borrow information across curves to better estimate the underlying process. When the data are sparse but the pooled data are sufficiently dense, Yao et al. (2005) have proposed Principal Component Analysis through Conditional Expectation (PACE). This method provides a structure for estimation of covariances relating to the underlying process, X, for instance, for estimating the covariance between $X(t)$ and $X'(s)$ or between $X(t)$ and $X(s)$. It should also be noted that the PACE method is in contrast to approaches such as functional regression (Ramsay and Dalzell, 1991), where the entire process is included in the modeling of the entire response. The applications are plentiful and include areas such as yeast cell cycles (Yao et al., 2005) and online auction dynamics (Liu and Müller, 2009).

Often, a relationship of particular interest is that between $X'(t)$ and $X(t)$. For Gaussian processes, the idea of a dynamic transfer function has been developed to understand this relationship and to estimate $E[X'(t)|X(t)]$ (Liu and Müller, 2009; Müller and Yang, 2010). This transfer function is defined as β in the equation

$$\begin{aligned} X'(t) &= \mu_{X'}(t) + \beta(t)[X(t) - \mu_X(t)] + Z(t) && (1.1) \\ E[X'(t)|X(t)] &= \mu_{X'}(t) + \beta(t)[X(t) - \mu_X(t)], && (1.2) \end{aligned}$$

where $X(t)$ and $Z(t)$ are independent. Müller and Yao (2010) show that the Gaussianity assumption guarantees the existence of such a transfer function, and of a first order linear differential equation satisfied by each observed trajectory. Verzelen et al. (2012) have extended this work to non-Gaussian processes. They show that each trajectory of a smooth stochastic process satisfies a first order nonlinear differential equation given by

$$\begin{aligned} X'(t) &= f(t, X(t)) + Z(t) && (1.3) \\ E\{X'(t)|X(t)\} &= f(t, X(t)) \end{aligned}$$

and provide a two-step smoothing procedure to estimate f. In the Gaussian case, $f(t, X(t))$ reduces to $\mu_{X'}(t) + \beta(t)[X(t) - \mu_X(t)]$ from (1.1). That being said, as Heckman (2010) has pointed out, this conditional expectation does not give us the exact underlying differential equation nor the behavior of the process, X, it only provides a way to study the relationship between $X'(t)$ and $X(s)$.

To date, the work on instantaneous dynamics has only included a single process in the conditional expectation given in (1.3). We propose an addition

to the model in (1.3) that allows for a second process, $W(\cdot)$, where $W(t)$ is thought to have an influence on $X'(t)$. We then extend the work of Müller and Yao (2010) and Verzelen et al. (2012) for determining the domains where one model explains the variation in $X'(t)$ significantly better than another. In our case, we compare the model that includes $W(t)$ to that which does not. To determine this, we formulate a hypothesis test and a permutation approach to calculating its significance level.

As mentioned previously, FDA, and in particular the approach of Yao et al. (2005) for longitudinal data, lends itself well to growth studies (Gasser et al., 1984; Verzelen et al., 2012). In continuing with this theme, we apply the two-step smoothing procedure of Verzelen et al. (2012) and our subsequent extensions of this to data comprised of observations on eight distinct groups of mice. These observations include the weekly body masses of the mice, as well as their weekly amounts of food eaten. The eight distinct groups are characterized by selective breeding (yes/no), access to an exercise wheel (yes/no) and gender (male/female). It is of biological and evolutionary interest as to how the instantaneous dynamics between body mass and growth rate differ between each of these groups. With this in mind, we first estimate the growth rates of each mouse and then use the approach of Verzelen et al. (2012) to estimate the relationship between body mass at week t and growth rate at week t. Further, we apply our new methods for including the additional stochastic process, W, corresponding to the weekly amount eaten, in model (1.3). This allows us to determine for which of the eight groups and during which weeks, the amount eaten in a week significantly effects the instantaneous relationship between body mass and growth rate.

The remainder of the thesis is organized as follows. In Chapter 2, we give a detailed description of the mouse data set, including the breeding and experimental design as well as some exploratory analysis and observations. Chapter 3 reviews foundational smoothing techniques such as smoothing splines and kernel smoothing, as these are essential for the subsequent analyses. We also address some standard techniques for choosing the smoothing parameters. The topic of FDA and in particular the contrasts between the curve by curve approach and that of learning the differential equation from the observed data is discussed in greater detail in Sections 4.1 and 4.2. In the remainder of Chapter 4 we formulate the models, model fitting procedures and model comparison techniques proposed by Müller and Yao (2010) and Verzelen et al. (2012), as well as our extensions. The analysis of the mouse data introduced in Chapter 2 is described in Chapter 5, along with the presentation and discussion of the results. Finally, Chapter 6 provides

the method and results from a simulation study to determine the statistical properties of our nonparametric testing method for determining the significance of $W(t)$ in the relationship between $X'(t)$ and $X(t)$, while Chapter 7 provides concluding remarks.

Chapter 2

The Data Set

The data set, first described in Swallow et al. (1998), was provided by Professor Patrick Carter, School of Biological Sciences, Washington State University. The data are comprised of weekly measurements taken from 320 house mice (*Mus domesticus*), who have been housed individually in a laboratory setting over a period of 80 weeks. Only 292 of these mice are included in the subsequent analysis, as the remaining 28 died early.

2.1 Breeding Design

The lines of house mice used here are from replicate lines selected for 16 generations. The breeding resulted in eight closed genetic lines. These lines were established as follows (Swallow et al., 1998). A set of 224 mice (112 of each male and female) were purchased from Harlan Sprague Dawley, Indianapolis. This initial group of mice was paired for breeding, with the exception that sibling mating was disallowed. This resulted in approximately 112 litters. From each of these litters, one male and one female were randomly selected, thus resulting in approximately 224 mice, referred to as generation -1. The mice from generation -1 were then randomly paired, again with sibling pairings disallowed. From these generation -1 pairs, eight lines were formed by randomly selecting 10 pairs for each line. The eight lines were randomly split into two groups of four (selection and control). The offspring of the chosen generation -1 pairs were designated generation 0. Figure 2.1 gives a schematic of the above description of breeding up to generation 0.

Within each line, from each of generations 0-9, 13 males and 13 females were chosen to produce the next generation. At each generation, these 26 breeder mice were selected at age 10 weeks. The 13 males and 13 females were randomly paired to breed, with the condition that no pair were siblings. The first 10 litters with two pups of each sex were used to maintain the line. The 13 pairs, rather than just 10, were used to ensure that there would be at least 10 litters with two pups of each sex.

In each generation, the 13 pairs of breeding mice in a selection line were chosen based on the average number of wheel revolutions run on days 5 and

6 on an activity wheel. From each of the 10 families, the highest running male and female were selected to breed. To make up the remaining 3 pairs of the 13 required, three additional males and females were chosen. These six mice were chosen based on being the second highest runners from the families with the highest running totals, with the condition that no two of the six additional mice were siblings. These 26 mice were randomly paired to breed the next generation as described in the preceding paragraph. Figure 2.2 provides a schematic of how (for the selected lines) generation i produces generation $i + 1$.

In each generation, the 13 pairs of breeding mice each control line were chosen randomly as follows. One male and one female mouse were randomly selected from each of the 10 families. Then an additional 3 males and 3 females were randomly chosen, with the condition that no two of the six additional mice were siblings. These 13 males and 13 females were then randomly paired for breeding, again with the condition of no pair being made up of siblings.

2.2 Experimental Design

Our analysis is based on observations made from generation 15 of the above breeding design. As described in the preceding section, of the eight genetic lines, four are control and four are selected for wheel running. From these eight lines, 5 pairs of breeding mice were selected to produce the next generation. For each of the five families within each line, two mice of each sex were assigned to be active, i.e, have access to a running wheel, and two mice of each sex were to be sedentary (Theodore J. Morgan, 2003). This divides the resulting 320 mice into two groups of 160: active and sedentary. These two groups can be further partitioned by the sex and line type (1-8) of the mice. The mice are thus divided into 32 balanced categories. In our analysis, when only 292 mice are used, these 32 categories have sizes of approximately 8-12 mice.

Weekly measurements were taken for each mouse over a period of 80 weeks. These measurements include the body mass of the mouse (grams), the amount eaten in the last week (grams) and (for the active group) the number of revolutions ran in the last week. At the end of each week, each mouse was weighed and the amount of food left in the bowl was measured. The amount eaten was calculated as the difference between the food in the bowl at the beginning of the week and the amount remaining in the bowl at the end. This could be subject to error, as some mice would bury or dispose

of food without actually eating it.

For the purpose of our analysis, we ignore the dependence within the 8 lines, thus treating the mice as independent. Therefore, we consider categories formed by sex, the presence of an exercise wheel and whether or not a mouse is selectively bred or not. This results in eight groups, each comprised of 30-40 mice, where each group has roughly comparable size.

2.3 Data Summary

Figures 2.3 and 2.4 show the body masses over the 80 weeks for the 292 mice, and for each of the eight groups, respectively. Figure 2.5 shows the point-wise standard deviations of the body masses of the 292 mice. Clearly, as the mice age there is greater variation in the body masses. This pattern in the variation is also evident within each group, as shown in Figure 2.6. There are some weeks when the body mass was not observed, typically corresponding to holidays. This can be seen by the missing slices in Figures 2.3 and 2.4. In the original data set, the researchers who collected the data had imputed these data points with the proceeding week's observations. With the permission of researcher Patrick Carter, we treat the first $k-1$ of k consecutive identical values as missing. On average, it appears that the sedentary mice are heavier for both males and females. Also, for both genders, those mice that were not selectively bred seem to be heavier.

Figures 2.7 and 2.8 show the amount eaten over the 80 weeks for the 292 mice, and for each of the eight groups, respectively. Each of the eight groups appears to show similar patterns, although the female sedentary mice exhibit greater variation. Figure 2.9 shows that the point-wise standard deviations of the amount eaten are fairly uniform, particularly after the twentieth week. This pattern is similar within each group, as seen in Figure 2.10. There are also some substantial outliers evident in Figures 2.7 and 2.8. For example, in the lower left panel of Figure 2.8, there is a single point at 100 grams, while every other measurement is below 75 grams. These outliers explain the spikes in the point-wise standard deviations, seen in Figure 2.10. After discussion with Professor Carter, this and three other similar points were determined to be inaccurate measurements or were due to food being wasted without the researcher who collected the data knowing and thus we treated them as missing values. There was also a negative measurement on an active male, in the selected group, which we also replaced by a missing value.

2.4 Previous Research and Research Objectives

The data collected from the mice of generation 10 from the breeding design described in Section 2.1 have been used to study a variety of genetic and evolutionary traits. Natural questions concern things such as how the body mass, amount of energy used and food consumption vary between groups (Koteja et al. 1999; Koteja et al. 2001). For example, it was found that the selected mice from generation 10 ran 70% more total revolutions per day than their control counterparts and that overall, males ran less than females (Koteja et al., 1999). Other work has addressed topics such as the variation in the amount of food wasted between the groups (Koteja et al., 2003), where it was found that there were significant differences in the amount of food wasted between replicate lines, but not between the selected and control groups. In most of these studies, it is of interest to compare the selected mice to those that are randomly bred and the active mice to the sedentary. In general, the primary focus is not on between gender comparisons as these do not provide as many conclusions about the evolutionary process of the mice.

Our objective for the data is two-fold. We treat both the body mass and the amount eaten as being governed by underlying stochastic processes. We first aim to explore the relationship between growth rate at a given age, t, and body mass at t for each of the eight groups. This is in contrast to using a more complex historical model that includes all the information about the body mass up to age t in order to study the growth rate. After estimating this relationship, we try to draw some general conclusions as well as make comparisons between the groups. The second objective is to explore how the relationship between body mass at age t and growth rate at age t is changed by including the amount eaten at t in the model. This allows us to determine in which of the groups the amount eaten at t contributes significantly to explaining the growth rate at t, providing a better understanding of the underlying biological traits.

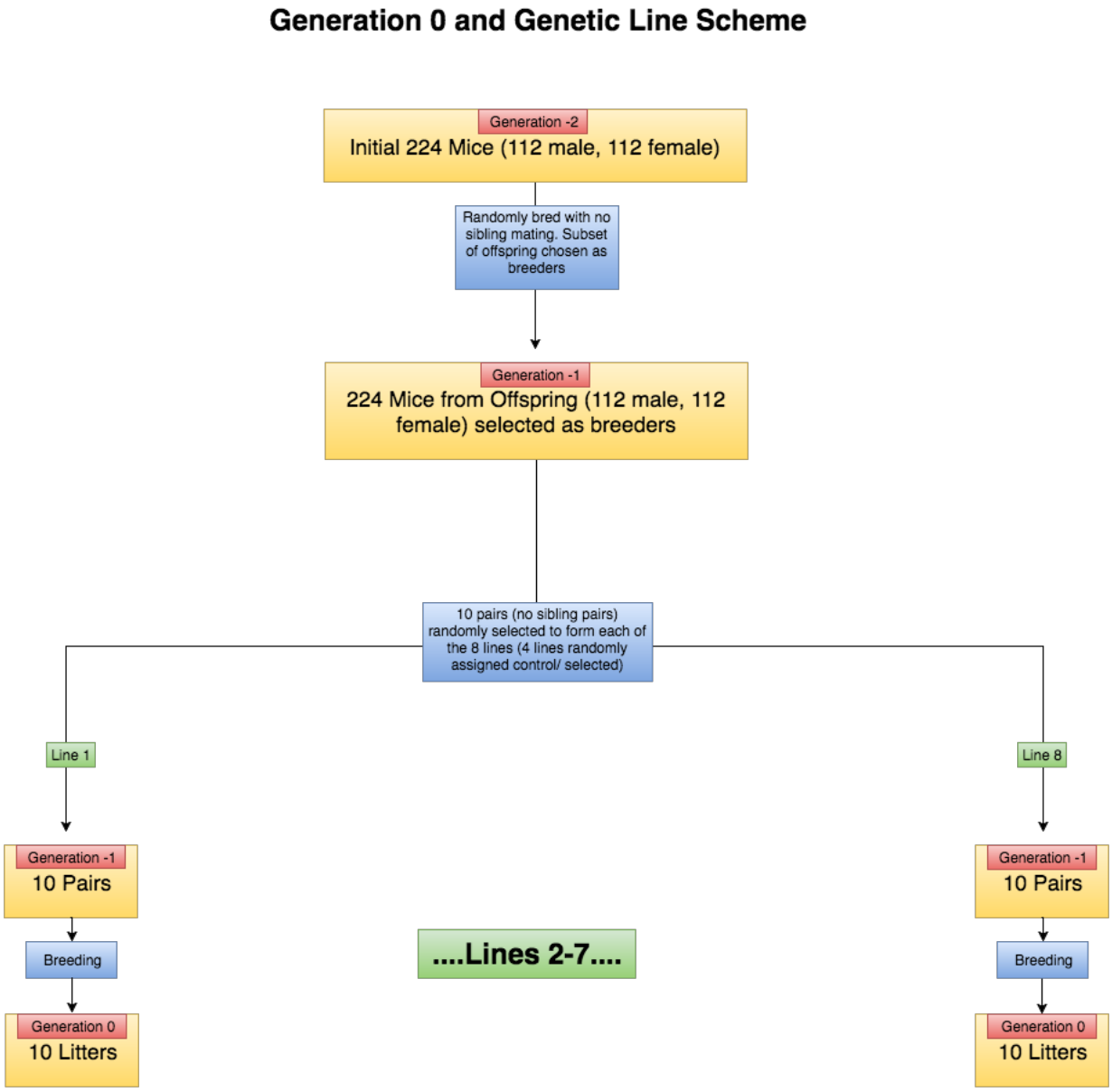


Figure 2.1: A schematic depicting the initial breeding and formation of the eight genetic lines.

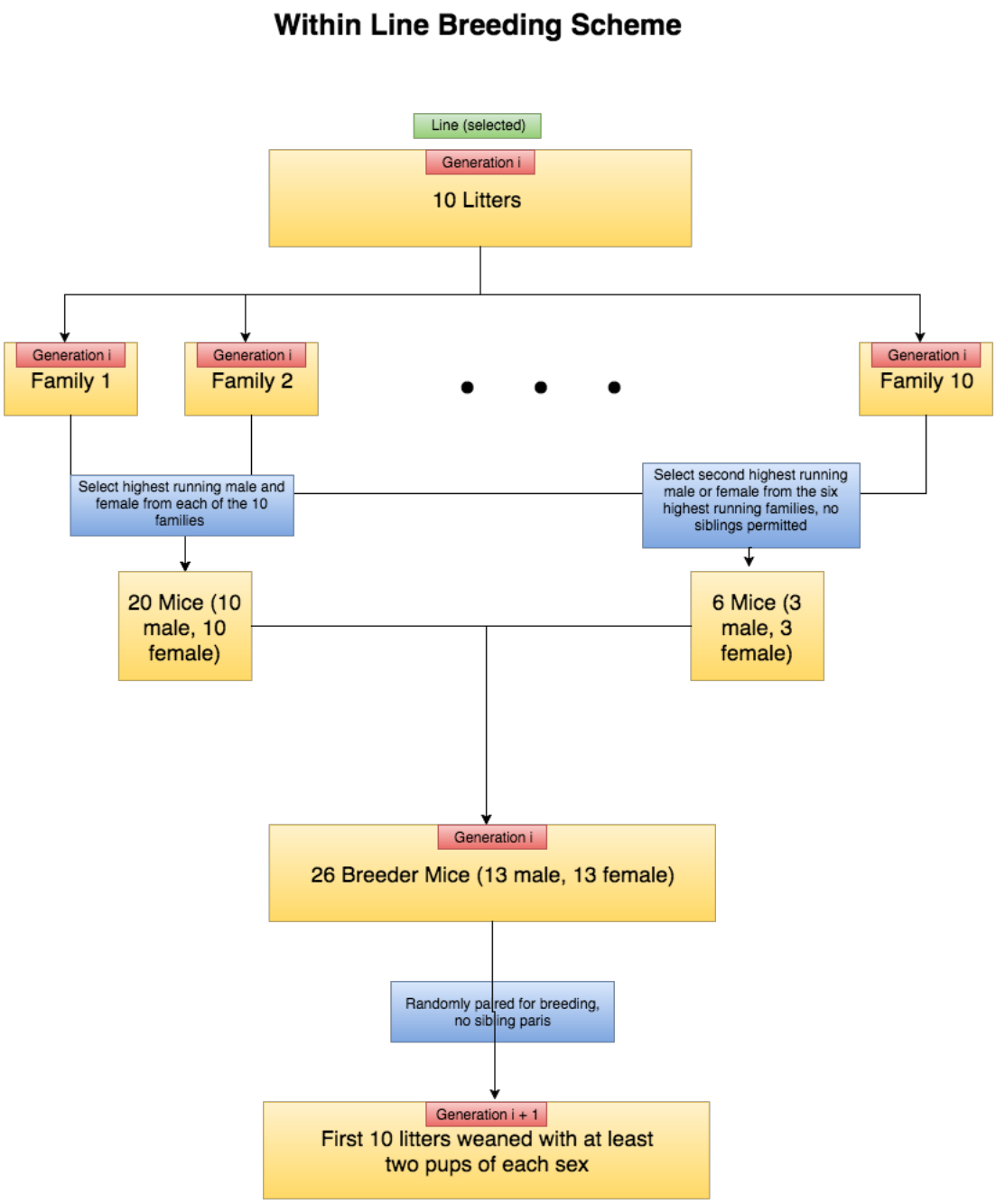


Figure 2.2: A schematic depicting the formation of generation $i + 1$ from generation i. This figure corresponds to one of the four selected lines.

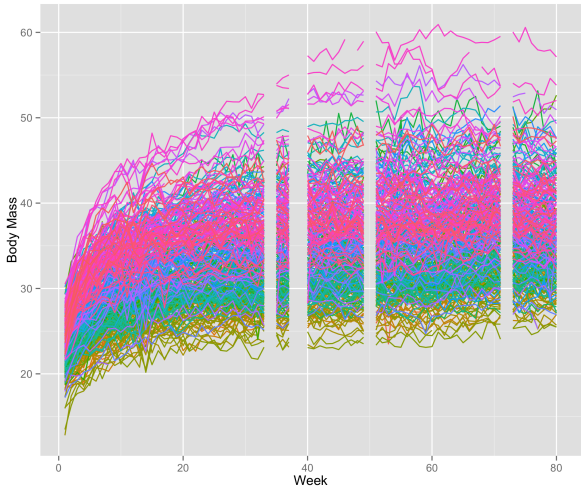


Figure 2.3: The body mass (grams) of all 292 mice as a function of age (weeks). The missing slices correspond to weeks when the body mass was not recorded.

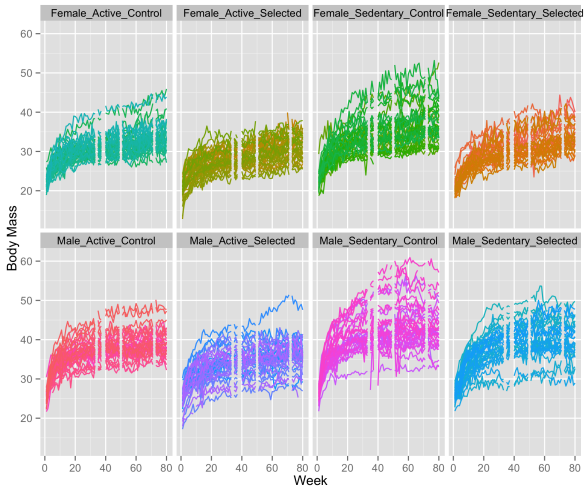


Figure 2.4: The body masses (grams) of the 292 mice as a function of age (weeks), organized into each of the eight groups. The missing slices correspond to weeks when the mass was not recorded.

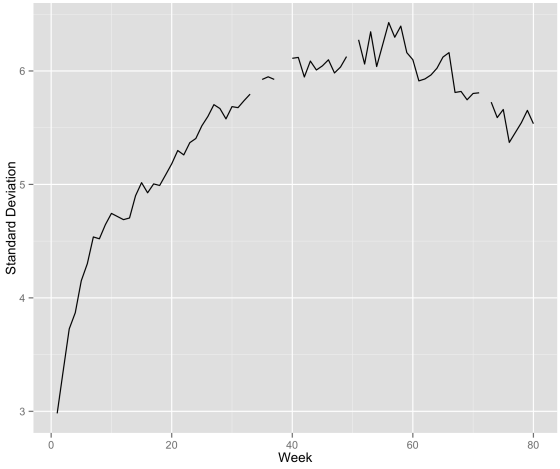


Figure 2.5: The point-wise standard deviations of the body masses (grams) as a function of age (weeks).

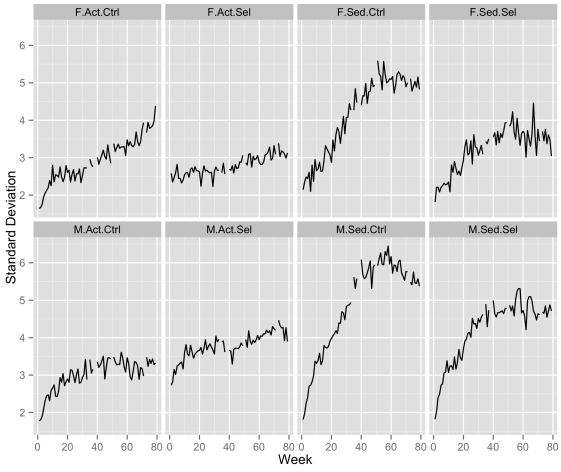


Figure 2.6: The point-wise standard deviations of the body masses (grams) for each of the eight groups as a function of age (weeks).

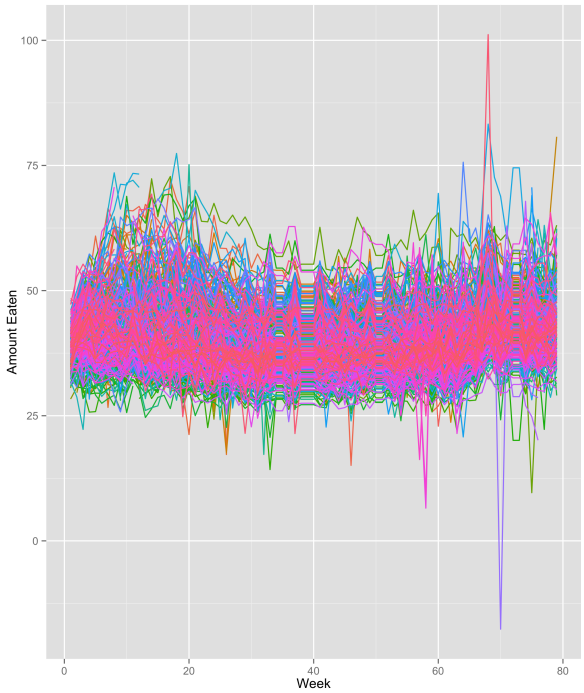


Figure 2.7: The amount eaten (grams) by the 292 mice as a function of age (weeks).

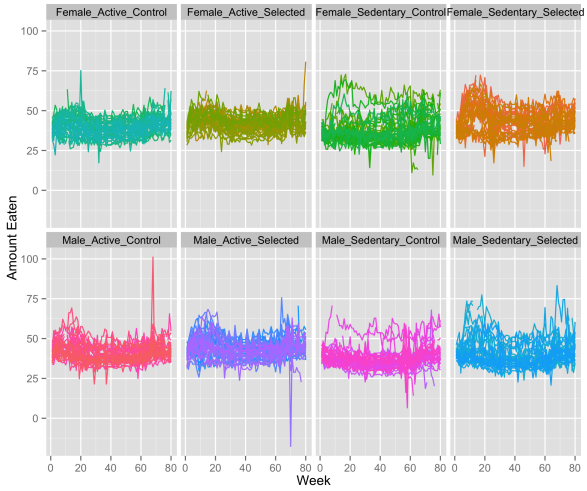


Figure 2.8: The amount eaten (grams) as a function of age (weeks), for each of the eight groups.

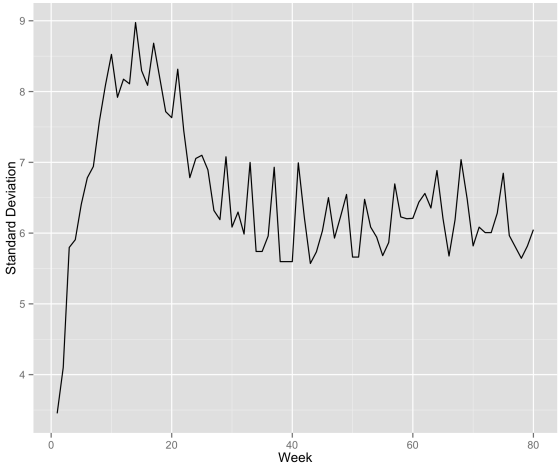


Figure 2.9: The point-wise standard deviations of the amount eaten (grams) for the 292 mice as a function of age (weeks).

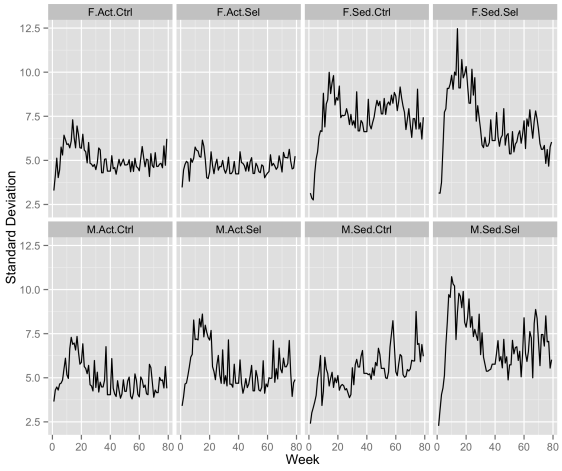


Figure 2.10: The point-wise standard deviations of the weekly amount eaten (grams) for each of the eight groups as a function of age (weeks).

Chapter 3

Smoothing Techniques

Consider a set of regression points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ such that $x_i \in [a, b]$ for all i and suppose that

$$y_i = m(x_i) + \epsilon_i, \text{where } \epsilon_i \text{ are i.i.d with } \mathrm{E}(\epsilon_i) = 0, \mathrm{Var}(\epsilon_i) = \sigma^2.$$

If instead we have regression points, $(x_1, w_1, y_1), \ldots, (x_n, w_n, y_n)$, then the previous model is adapted to

$$y_i = m(x_i, w_i) + \epsilon_i, \text{where } \epsilon_i \text{ are i.i.d with } \mathrm{E}(\epsilon_i) = 0, \mathrm{Var}(\epsilon_i) = \sigma^2.$$

This can be further generalized to include more covariates, but we restrict our focus to the univariate and bivariate cases.

Smoothing techniques attempt to capture trends in the data by providing an estimate of the function m. These estimates of m are obtained in such a way that the noise in the data is reduced. Smoothing techniques are typically used to extract information from the data, while providing flexibility and robustness.

In contrast to parametric regression, smoothing does not require any predetermined assumptions about the form of the relationship between the response and the explanatory variables. Instead, this relationship is determined completely by the information from the data. For this reason, larger sample sizes are typically needed for smoothing, when compared with those for parametric regression. In this section we will discuss two specific types of smoothing: smoothing splines and kernel smoothing.

3.1 Smoothing Splines

Definition: $\phi : [a, b] \to \mathbb{R}$ *is a spline of degree* p*, with interior knots* $t_1 < \cdots < t_n$*, where* $a < t_1$ *and* $t_n < b$*, if:*

(1) the restrictions of ϕ *to the intervals* $[a, t_1], [t_i, t_{i+1}]$ *and* $[t_n, b]$ *are polynomials of degree* p *for* $i = 1, \ldots, n$*,*

(2) $\phi(\cdot)$ *is a* $p-1$ *continuously differentiable function at the points* t_i*, for* $i = 1, \ldots, n$*.*

In the context of regression data, $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ such that $x_i \in [a, b]$ and $x_i < x_{i+1}$ for all i, we are usually interested in minimizing the mean squared error (MSE),

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - m(x_i))^2, \tag{3.1}$$

with respect to m. Now, suppose we want to penalize m based on certain characteristics. Then we can include a penalty term in (3.1) and minimize:

$$\mathcal{G}(m, \lambda) = \frac{1}{n}\sum_{i=1}^{n}(y_i - m(x_i))^2 + \lambda P(m), \tag{3.2}$$

for a fixed λ. We call λ the smoothing parameter and P the penalty function. This criterion indicates that we are willing to accept an estimate of m that increases the MSE if it also reduces the penalty. A common choice is to penalize m based on its curvature; that is, $P(m) = \int_a^b [m''(x)]^2 dx$ (Hastie and Tibshirani, 1990). In this case, it can be shown that the minimizer of $\mathcal{G}$ is a natural cubic spline with interior knots $x_1, \ldots, x_n$. These natural cubic splines are cubic splines that are subject to the condition $\int_a^{x_1} m''(x) = \int_{x_n}^b m''(x) = 0$ and are thus linear on the intervals $[a, x_1]$ and $[x_n, b]$. See Figures 3.1 and 3.2 for examples of cubic splines fit to data.

To better understand the penalized optimization problem in (3.2), we consider the extreme cases of λ. One can show that, when $\lambda \to \infty$, the minimizer of (3.2), $\hat{m}_\lambda$, approaches the least squares line. Heuristically, for increasing λ we must have $P(\hat{m}_\lambda)$ tending to 0 and the problem reduces to minimizing (3.1) with $\hat{m}_\lambda$ restricted to having $P(\hat{m}_\lambda) = 0$; that is, $\hat{m}_\lambda$ restricted to a line. On the other hand, if $\lambda = 0$, then we can choose $\hat{m}_\lambda$ so that the first term in (3.2) is 0 by setting $\hat{m}_\lambda$ to be the interpolating spline that passes through each data point. For intermediate values of λ, the minimizing function $\hat{m}_\lambda$ is a compromise between the least squares line and the interpolating function.

3.1.1 Fitting Smoothing Splines

We now focus on how a smoothing spline can be fit to regression data. Consider the case where the penalty function in (3.2) is given by the curvature of m. Then, as stated before, the minimizer in (3.2) is a natural cubic

spline. Our input data, $x_1, x_2, \ldots, x_n$, are the interior knots of the cubic spline and give us $n+1$ segments of the interval $[a, b]$. Therefore, it appears we need to determine four coefficients for each segment, for a total of $4(n+1)$ coefficients. Fortunately, our natural cubic splines are smooth and twice continuously differentiable on $[a, b]$ and linear on the intervals $[a, x_1]$ and $[x_n, b]$. This places restrictions on the coefficients and thus reduces the number of basis functions needed to n. Thus, the dimension of the space of natural cubic splines with interior knots $x_1, \ldots, x_n$ is equal to n. Denote a set of basis functions as $\{B_i, i = 1, \ldots, n\}$. We can now write our natural cubic spline as

$$m(x) = \sum_{i=1}^{n} \alpha_i B_i(x).$$

Typical choices for the B_i's include the truncated power basis or the B-spline basis. Notice that we have reduced our infinite dimensional class of m's to an n-dimensional model that is linear in the B_i's. We can thus rewrite our objective function from (3.2) in matrix form. Setting $(\boldsymbol{B})_{ij} = B_j(x_i)$, this yields:

$$\mathcal{G}(m, \lambda) = \frac{1}{n}(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\alpha})^T(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\alpha}) + \lambda \boldsymbol{\alpha}^T \boldsymbol{C} \boldsymbol{\alpha}, \tag{3.3}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)^t$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^t$ and $(\boldsymbol{C})_{ij} = \int_a^b B_i''(x) B_j''(x) dx$ contains the curvature information of the basis functions. Setting the derivative of $\mathcal{G}$ with respect to $\boldsymbol{\alpha}$ equal to zero and solving yields

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{B}^T\boldsymbol{B} + n\lambda\boldsymbol{C})^{-1}\boldsymbol{B}^T\boldsymbol{y},$$

which is the unique minimizer of $\mathcal{G}$ provided $\boldsymbol{B}^T\boldsymbol{B} + n\lambda\boldsymbol{C}$ is positive definite.

This minimizing $\hat{\boldsymbol{\alpha}}$ yields the following fitted values for y_i,

$$\hat{\boldsymbol{y}} = \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B} + n\lambda\boldsymbol{C})^{-1}\boldsymbol{B}^T\boldsymbol{y} = \boldsymbol{S}_\lambda \boldsymbol{y},$$

where $\boldsymbol{S}_\lambda$ is called the smoothing or hat matrix. There are various ways to choose the smoothing parameter, λ, that will be discussed in Section 3.3. Finally, the above discussion is restricted to univariate splines, as these are all that is required in our application. In more complex situations, one may require, for example, bivariate splines.

3.2 Kernel Smoothing

An alternative method to smoothing splines is to smooth the data via Kernel smoothing. Kernel smoothing is typically used in two different applications.

The first is for obtaining a density estimate from a sample. The second is for investigating a regression relationship, such as the one outlined at the beginning of this chapter. We will focus our proceeding discussion on the latter case.

For both density estimation and regression, the smoothing is done via a kernel function, $K(\cdot; h)$ (Wand and Jones, 1995). We restrict $K(\cdot; h)$ to be a symmetric density function with mean 0. The scale parameter, h, controls the degree of smoothing and can sometimes be thought of as the standard deviation of the random variable with density equal to the kernel function. Some typical choices for K are the Gaussian, the Epanechnikov and the box kernels.

The simplest approach to nonparametric kernel regression is to adopt a local mean approach. The estimate at an evaluation point z is given by

$$\widehat{m}_h(z) = \frac{\sum_{i=1}^n K(x_i - z; h) y_i}{\sum_{i=1}^n K(x_i - z; h)}.$$

This is referred to as the Nadaraya-Watson estimate (Nadaraya, 1964; Watson, 1964). More generally, for a fixed p, the $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p$ that minimize

$$\sum_{i=1}^n [y_i - \beta_0 + \beta_1(x_i - z) + \cdots + \beta_p(x_i - z)^p]^2 K(x_i - z; h)$$

provide the p^{th} degree polynomial estimates of m and its derivatives evaluated at z. Namely, the estimates are given by $\widehat{m^{(i)}}_h(z) = i!\hat{\beta}_i(z)$ for $i = 0, \ldots, p$. Note that when $p = 0$ the Nadaraya-Watson estimate is recovered. Setting $p = 1$ is a common choice, and is referred to as local-linear estimation. The local linear approach is good when used to estimate m, but higher order polynomials are recommended for obtaining estimates of m's derivatives (Wand and Jones, 1995). Figure 3.3 gives an example of a Nadaraya-Watson (left pane) and local linear (right pane) estimate of m based on simulated regression data.

If we have two real-valued covariates, x and w, and data (x_i, w_i) for $i = 1, \ldots, n$, then the local mean estimator of m is given by

$$\widehat{m}_{h_1,h_2}(z_1, z_2) = \frac{\sum_{i=1}^n K(x_i - z_1, w_i - z_2; h_1, h_2) y_i}{\sum_{i=1}^n K(x_i - z_1, w_i - z_2; h_1, h_2)}$$

and the local linear estimate of $m(z_1, z_2)$ is the value of β_0 gotten from the least squares criterion

$$\min_{\beta_0,\beta_1,\gamma_1} \sum_{i=1}^n [y_i - \beta_0 + \beta_1(x_i - z_1) + \gamma_1(w_i - z_2)]^2 K(x_i - z_1, w_i - z_2; h_1, h_2).$$

The multivariate kernel, K, can take many forms. Here, we restrict our multivariate kernels to those of the multiplicative form

$$K(x_i - z_1, w_i - z_2; h_1, h_2) = K_1(x_i - z_1; h_1)K_2(w_i - z_2; h_2), \tag{3.4}$$

where K_1 and K_2 are univariate kernels. See Figure 3.4 for an example of a local linear bivariate smooth.

In both the univariate and bivariate case, $\widehat{m}$ can be seen as a weighted average of the y_i's. These weights depend on the choice of the kernel function, bandwidth and the proximity of the evaluation points to the data. Furthermore, as the minimizing criterion is a least squares problem and thus quadratic, we have that $\hat{\boldsymbol{y}}$ is linear in $\boldsymbol{y} = (y_1, \ldots, y_n)^t$. Therefore, as in the case of smoothing splines, there exists a smoothing (hat) matrix, $\boldsymbol{S}_h$ such that $\hat{\boldsymbol{y}} = \boldsymbol{S}_h\boldsymbol{y}$.

3.3 Choice of Smoothing Parameter

In both of the previous sections our estimates of m depend on a value which determines the degree to which our estimating function smooths the data. In the case of smoothing splines, the smoothing parameter, λ, controls smoothing. Likewise, for kernel smoothing, the bandwidth, h, controls smoothing. Choosing these parameters is important as a poor choice can lead to over- or under-smoothing of the data.

In both smoothing splines and kernel smoothing, leave-one-out cross-validation (Hastie et al., 2001) is often used to determine λ, h or (h_1, h_2). This algorithm can be described as follows for choosing λ. Choosing h or (h_1, h_2) is similar. For each value of $i = 1, \ldots, n$, the i^{th} data point is left out of the fitting and is then predicted with the resulting natural spline. The prediction error, $e_i^*(\lambda) = y_i - \hat{y}_i^{(-i)}$, for this data point is then computed. We define the cross-validation function as the sum of the squared prediction errors, $CV(\lambda) = \sum_{i=1}^{n}[e_i^*(\lambda)]^2$. The λ with the lowest CV is chosen as the smoothing parameter.

Fixing the trace of the smoothing matrix is another method to specify the smoothing parameter or bandwidth. Recall, that in both smoothing splines and kernel smoothing, the fitted values could be written as $\hat{\boldsymbol{y}} = \boldsymbol{S}\boldsymbol{y}$, for some matrix $\boldsymbol{S}$ that depends on λ or h. Notice that this is a similar framework to that of linear regression, where the hat matrix, $\boldsymbol{H}$, plays the role of $\boldsymbol{S}$ in determining $\hat{\boldsymbol{y}}$. In this case, the trace of $\boldsymbol{H}$ is called the degrees of freedom and is equal to the number of parameters in the regression model. Similarly, for smoothing splines and kernel smoothing, we define the degrees

of freedom to be the trace of $\boldsymbol{S}$. Since the smoothing matrix $\boldsymbol{S}$ depends on the smoothing parameter, by fixing $trace(\boldsymbol{S}) = \sum_{i=1}^{n} S_{ii}$ to some desired value, the smoothing parameter can then be computed numerically. As $\boldsymbol{S}_\lambda$ is an $n \times n$ matrix, the degrees of freedom are equal to the sum of $\boldsymbol{S}_\lambda$'s eigenvalues. In general, the lower the degrees of freedom the greater the degree of smoothing. For an example of this, see the two splines in Figure 3.2. Finally one possible advantage to tuning the degrees of freedom as opposed to for example, cross validation, is that one may get better performance for approximating derivatives. Moreover, if one wishes to have the same amount of smoothing over many curves (for example, mice body masses), tuning via the degrees of freedom provides a way to ensure this.

There are other methods, such as ones based on the Akaike Information Criterion (Sakamoto et al., 1986), that can also be used for determining the smoothing parameters. In particular, for local polynomial kernel regression the so called "plug in" method is often used (Wand and Jones, 1995). The plug in method uses the idea of plugging in estimates of the unknown parameters in the expression of the asymptotically optimal bandwidth. This asymptotically optimal bandwidth is the one which optimizes the asymptotic mean square error. Finding estimates for the unknown parameters in the formula for the optimal bandwidth can be challenging. These estimates are often found using kernel smoothing, thus requiring their own bandwidths, which must be chosen. This hierarchal structure and its properties have been explored in detail by Hall et al. (1992).

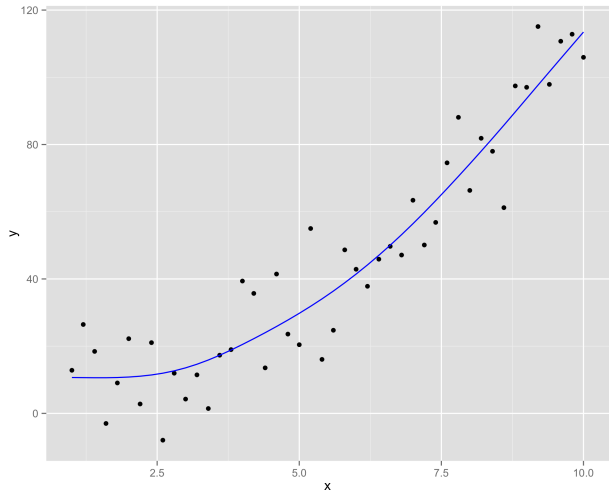


Figure 3.1: An example of a cubic smoothing spline fit to 46 data points, where the smoothing parameter has been selected by leave-one-out cross-validation.

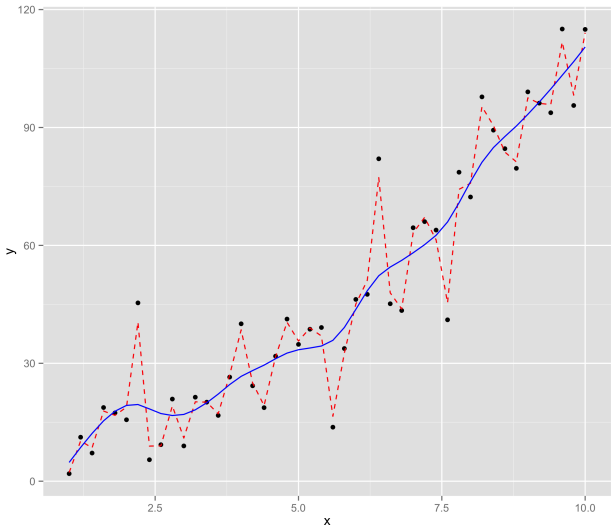


Figure 3.2: An example of two cubic smoothing splines fit to 46 data points, with different degrees of freedom (smoothing parameters). The dashed line corresponds to using 40 degrees of freedom, while the solid line corresponds to using 10.

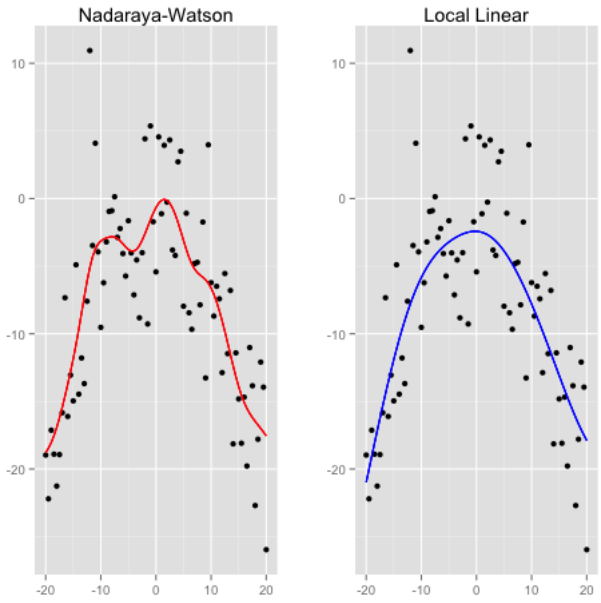
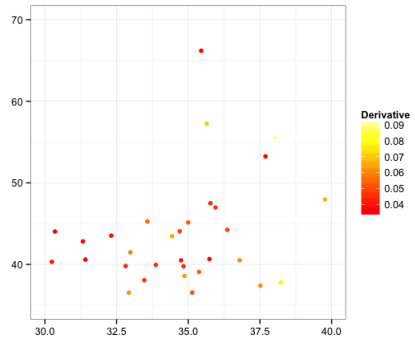
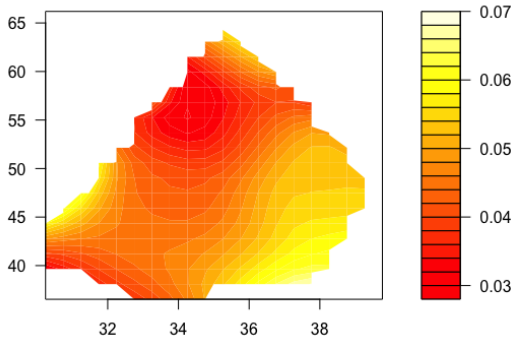


Figure 3.3: An example of Nadaraya-Watson (left) and Local Linear (right) estimates of a regression function, based on 81 data points. A standard normal kernel with bandwidth 5 is used.



(a)



(b)

Figure 3.4: An example of a contour plots of a bivariate local linear estimate of m using a multiplicative Gaussian kernel with standard deviations 1.3 and 3.75. Plot (a) shows the estimate of m evaluated at the data points. Plot (b) shows the estimate of m evaluated on a 20×20 grid of evaluation points derived from data (bottom).

Chapter 4

Differential Equation Models

Differential equations can be used to model processes encountered in a variety of disciplines. The objective is typically either to estimate the trajectory specific parameters of the differential equation or to understand the governing dynamics of the underlying process from the observed data.

4.1 Parametric Models

In many previous works, (Ramsay and Silverman 2005; Cao and Ramsay 2007; Cao et al. 2012) a previously specified nonrandom differential equation is assumed to describe the underlying process, $X(\cdot)$. The goal is then to estimate the parameters of this differential equation from the observed data, denoted (t_j, Y_j) for $j = 1, \ldots, J$. The Y_j's are assumed to be a noisy realization of the underlying process

$$Y_j = Y(t_j) = X(t_j) + \epsilon_j, \tag{4.1}$$

where ϵ_j's are mean zero independent identically distributed random variables with $\text{Var}(\epsilon_j) = \sigma^2$. Formally, the aim is to minimize the following penalized least squares problem

$$\mathcal{J}(\boldsymbol{c}|\boldsymbol{\beta}, \lambda) = \sum_{i=1}^{n} [Y(t_j) - X(t_j)]^2 + \lambda \int [L_\beta X(t)]^2 dt,$$

where $\boldsymbol{c}$ is the vector of coefficients from a basis function expansion of $X(\cdot)$ and L_β is a differential operator depending on the unknown parameter vector, $\boldsymbol{\beta}$. We define a differential operator as in Ramsay et al. (1997) to be

$$L_\beta X(t) = \sum_{j=0}^{m-1} \beta_j D^j X(t) + D^m X(t),$$

where the notation $D^j X(t)$ indicates the j^{th} derivative of $X(t)$. One reason to use a general linear operator, instead of a more specific term such as the

curvature, is to be able to penalize X in more accurately. For example, if we know that locally X satisfies some differential equation, then we may wish to use a differential operator that penalizes departure from this specific equation.

Much previous work has been done to estimate the parameters of the above minimization problem. Heckman and Ramsay (2000) jointly estimate $\boldsymbol{c}$ and $\boldsymbol{\beta}$ but find the estimators to be unsatisfactory. Cao and Ramsay (2007) modify this criterion and propose a two step estimation method. They first minimize $\mathcal{J}$, for a fixed $\boldsymbol{\beta}$, with respect to the coefficient vector, $\boldsymbol{c}$, yielding $\hat{\boldsymbol{c}}(\boldsymbol{\beta})$. An un-penalized criterion involving $\hat{\boldsymbol{c}}(\boldsymbol{\beta})$ is then used to determine the optimum $\boldsymbol{\beta}$. It is important to note that this method is not iterative.

The trajectory-wise approach described above can be generalized to situations where there is more than one observed trajectory of $X(\cdot)$. In these cases, often each trajectory is modeled via the same parametric differential equation but with a different parameter vector, $\boldsymbol{\beta}$, which is treated as a random effect. Alternatively, one can estimate the parameter vector separately for each observed trajectory, as described above.

4.2 Modeling Dynamics

As described in the preceding section, the trajectory-wise approach to fitting a differential equation model relies on the ability to specify a differential equation that describes the underlying process. If the underlying process is not well understood then this may not always be practical. Conversely, if many realizations of the underlying process are available, one can try to learn the differential equation directly from these trajectories (Liu and Müller, 2009; Müller and Yang, 2010; Verzelen et al., 2012). Specifically, given n realizations, $X_1, \ldots, X_n$, of a stochastic process X on a domain $\mathcal{T}$, we assume that we observe the noisy measurements of the process, Y_i, according to

$$Y_i(t_{ij}) = X_i(t_{ij}) + \epsilon_{ij}, j = 1, \ldots, J_i, \tag{4.2}$$

where the ϵ_{ij}'s are mean zero, independent identically distributed random variables with $\mathrm{Var}(\epsilon_{ij}) = \sigma^2$.

The method of attempting to learn the differential equation from the data uses observations from all n realizations together to estimate the underlying dynamics, rather than estimating the parameter of the aforementioned prespecified differential operator on a path by path basis. This method is particularly powerful in situations when data are sparsely observed for each realization of X. In these cases, estimation of these individual trajectories

is difficult, and more so for their derivatives. The borrowing of information from each trajectory can often aid in this estimation. For such sparse situations, Yao et al. (2005) propose the use of Functional Principle Component Analysis through Conditional Expectation (PACE). The PACE method has been applied to Gaussian processes (Müller and Yang, 2010) and to online auction dynamics (Müller and Yao, 2010).

In longitudinal studies, one may be interested in relating $X(t)$ or $X'(t)$ to $X(s)$ for a fixed $s \leq t$. Studying these relationships can provide insight into typically unknown mechanisms governing the observations. For Gaussian processes, Liu and Müller (2009) have proposed the use of dynamic transfer functions to provide an estimate of the influence that $X(t)$ has on $X^{(\nu)}(t)$ for $\nu > 0$. These transfer functions can be related to the conditional expectation of $X^{(\nu)}(t)$ given $X(t)$. Müller and Yang (2010) have extended the idea of transfer functions for Gaussian processes, to allow one to predict trends in, for example, $X'(t)$, based on previous levels of the process X at s. The instantaneous dynamics between $X(t)$ and $X'(t)$ will be the focus of Sections 4.3-4.5 but before proceeding, we reiterate the methodological differences between the trajectory wise approach described in Section 4.1 and the alternative of learning the underlying process from the data, as described at the start of this section.

The two predominant differences between these two approaches are in how the data are used. When the governing differential equation can be reasonably assumed, each of the observed trajectories can be analyzed separately (Ramsay and Silverman 2005; Cao and Ramsay 2007; Cao et al. 2012). In contrast, when the underlying dynamics are to be inferred from the data, then for each t, all observations from the n realizations of X can be used to estimate the underlying process at t (Yao et al. 2005; Liu and Müller 2009; Müller and Yang 2010; Müller and Yao 2010; Verzelen et al. 2012). Essentially, for the approach described in Section 4.1, an equation is first specified and then the unknown parameters are estimated. On the other hand, when trying to learn the differential equation from the data, one estimates the overlying equations from the observations. Furthermore, when each trajectory is analyzed individually, typically the entire trajectory is to be estimated (Ramsay and Silverman 2005; Cao and Ramsay 2007; Cao et al. 2012), while when information is borrowed from across the observed trajectories, the estimation is "local" in t. Indeed, with this local approach, the focus may be on studying specific time intervals (Müller and Yao, 2010; Verzelen et al., 2012).

4.3 Instantaneous Dynamics

In many applications, estimation of the instantaneous relationship between $X(t)$ and $X'(t)$ is of interest. More precisely, for a fixed time, determining how the value of the stochastic process is effecting its rate of change can be informative. For example, in growth studies, one may wish to infer how the current body mass of an individual at age t is related to the individual's growth rate at age t. Likewise, in finance, one may wish to infer the direction of a stock price based on its current valuation. As mentioned in Section 4.2, for Gaussian processes, the use of transfer functions to describe these relationships has been explored (Liu and Müller, 2009; Müller and Yang, 2010). Verzelen et al. (2012) extended this work to non-Gaussian processes, proposing a two step kernel estimation procedure. To our knowledge, a further generalization to include the effect of additional stochastic processes on these instantaneous dynamics has not yet been studied.

This section is comprised by first reviewing how the rate of change can be partitioned into deterministic and random components and by examining some specific cases of this decomposition, as described in Verzelen et al. (2012). We then conclude by discussing a similar decomposition to account for the influence of an additional stochastic process.

Consider a differentiable stochastic process, X, on a domain, $\mathcal{T}$, such that X and X' have finite variance. Provided $E[X'(t)]$ exists, we can decompose $X'(t)$ as

$$X'(t) = E\{X'(t)|X(t)\} + Z(t) \text{ where } Z(t) = X'(t) - E\{X'(t)|X(t)\}.$$

Note that we always have $E\{Z(t)|X(t)\} = 0$ and $\text{Cov}(Z(t), E\{X'(t)|X(t)\}) = 0$ almost surely. Moreover, we can write $E\{X'(t)|X(t)\} = f(t, X(t))$ for some function f. This allows us to write the nonlinear differential equation

$$\begin{aligned} X'(t) &= f(t, X(t)) + Z(t) \\ f(t, X(t)) &= E\{X'(t)|X(t)\} \end{aligned} \tag{4.3}$$

We refer to f as the deterministic part of the equation and Z as the stochastic part. When f is unknown and must be estimated from the data, we refer to the first equation in (4.3) as a data-driven differential equation. In some applications, f may be time independent, in which case (4.3) is referred to as an autonomous system (Verzelen et al., 2012).

The simplest case of (4.3) is when f is just the mean function of X' and so X' and X are uncorrelated. In this case, (4.3) reduces to

$$X'(t) = \mu_{X'}(t) + Z(t). \tag{4.4}$$

If X is a Gaussian process, then it can be shown that $f(t, X(t))$ is of the form $\mu_{X'}(t) + \beta(t)[X(t) - \mu_X(t)]$ and thus only $\mu_{X'}(t)$ and $\beta(t)$ need to be estimated. In this case, β is referred to as the transfer function (Liu and Müller, 2009; Müller and Yang, 2010). Here, (4.3) becomes

$$\begin{aligned} X'(t) &= \mu_{X'}(t) + \beta(t)[X(t) - \mu_X(t)] + Z(t) \\ &\equiv \mu^*(t) + \beta(t)X(t) + Z(t) \end{aligned} \tag{4.5}$$

This linear relationship may also hold for non-Gaussian processes in some cases. However, many processes are more complex than what can be captured by the linear dynamics in (4.5). In these cases, Verzelen et al. (2012) propose estimating $f(t, X(t))$ with a two step smoothing procedure, which is described in Section 4.4. We now conclude this section by considering a natural extension to that of the system in (4.3).

While the system in (4.3) can be used to describe various types of relationships between $X(t)$ and $X'(t)$, it leaves no room for modeling the dependence of $X'(t)$ on additional processes. For example, in the case of how growth rate is effected by an individual's body mass, one could easily hypothesize that the amount eaten at age $t - \Delta$, for some $\Delta \geq 0$, also plays a role in the growth rate at t. In the preceding, we consider only $\Delta = 0$.

For applications requiring the modeling of $X'(t)$'s dependence on a process $W(\cdot)$, (4.3) can be generalized to:

$$\begin{aligned} X'(t) &= f(t, X(t), W(t)) + \zeta(t) \\ f(t, X(t), W(t)) &= E\{X'(t)|X(t), W(t)\} \end{aligned} \tag{4.6}$$

where $\zeta(t) = X'(t) - E[X'(t)|X(t), W(t)]$. The framework that includes the process $W(\cdot)$ can be used when one suspects that a significant amount of the variation in $X'(t)$ can be explained by $W(t)$.

4.4 Two Step Estimation Procedure

Recall that in Section 4.3, we formulated three possible models to describe the instantaneous relationship between $X'(t)$ and $X(t)$ (and possibly $W(t)$). These were given in (4.3), (4.5), and (4.6), where the notation of f and Z is reused. The linear model, (4.5), is a simple, specific case of (4.3) where $f(t, X(t)) = \mu^*(t) + \beta(t)X(t)$. Thus, these models can be seen as ascending in complexity from (4.5), to (4.3), up to the nonlinear model, (4.6), where $X'(t)$ depends on $X(t)$ and $W(t)$ as opposed to just $X(t)$. We now describe how each of these models can be fit, before proceeding to comparing fits in Section 4.5.

For each of the three models, the first step is to estimate the trajectories X_i and X_i' from the raw observations, Y_{ij}, $j = 1, \ldots, J_i$, as modeled in (4.2). For $i = 1, \ldots, n$, we estimate the trajectory, X_i, and subsequently, its derivative, X_i', by the method of smoothing splines, as discussed in Section 3.1. We call these estimates $\hat{X}_i$ and $\hat{X}_i'$, respectively. This is in slight contrast to previous work, where convolution kernel smoothing estimates are used (Verzelen et al., 2012).

For the linear model in (4.5), one can view $(\hat{X}_i(t), \hat{X}_i'(t))$ as regression data for each fixed t. Thus, a natural estimate of $\beta(t)$ is the slope estimate from least squares linear regression at fixed t. Alternatively, one can use information from the whole process to estimate the best linear unbiased predictor (BLUP) of $X'(t)$ given $X(t)$ of the form $\alpha(t) + \beta(t)X(t)$. This BLUP is given by

$$\beta(t) = \frac{\text{Cov}[X'(t), X(t)]}{\text{Var}[X(t)]}.$$

If $X(t)$ and $X'(t)$ are jointly bivariate normal, then the above is the exact solution to (4.5) (Müller and Yao, 2010). Therefore to obtain an estimate, $\hat{\beta}(t)$, for $\beta(t)$, one can estimate $\text{Cov}[X'(t), X(t)]$ and $\text{Var}[X(t)]$. One way to do this is by estimating the covariance function of $X(\cdot)$ in order to estimate the eigenfunctions from X's Karhunen-Loéve expansion (Rice and Silverman, 1991). These estimated eigenfunctions then allow for the estimation of $\text{Cov}[X'(t), X(t)]$ and $\text{Var}[X(t)]$ (Liu and Müller, 2009).

For the nonlinear models (4.3) and (4.6), we use a two step smoothing procedure to estimate f. As mentioned above, the first step is to smooth the trajectories. The second step is to obtain an estimate of f. This two step procedure proceeds from the same idea used by Ellner et al. (2002) for autonomous systems and Verzelen et al. (2012) for systems as in (4.3).

In the univariate case (4.3), at each fixed time point, t, our data are given by $(\hat{X}_i(t), \hat{X}_i'(t))$ for $i = 1, \ldots, n$. We obtain the estimate of $f(t, \cdot)$, denoted $\hat{f}(t, \cdot)$, by using a local linear kernel smooth (Wand, 2015) of $\hat{X}_i'(t)$ on $\hat{X}_i(t)$, as described in Section 3.2. The corresponding bandwidth is chosen by cross-validation, as outlined in Section 3.3. Doing this for each t gives the estimate of the smooth function f. This approach is in small contrast to the work Verzelen et al. (2012), where a Nadaraya-Watson estimate is used instead of a local linear estimate.

To estimate $f(t, X(t), W(t))$ as in (4.6), we simply extend the above method for (4.3) to a bivariate local linear smooth, as described in Section 3.2. For a fixed t, our data are given by $(\hat{X}_i(t), W_i(t), \hat{X}_i'(t))$ for $i = 1, \ldots, n$. To obtain $\hat{f}$ we smooth $\hat{X}_i'(t)$ on $(\hat{X}_i(t), W_i(t))$. Again, cross validation is

used to select the bandwidths. It is important to note that we do not always need to preprocess the trajectories of W by smoothing. Rather, whether or not to smooth these trajectories relies on the assumed nature of the underlying process W. In Chapter 5, we do not smooth the amount eaten in our application to the mouse data set described in Chapter 2.

4.5 Model Comparisons

Given observations from n realizations, $X_1, \ldots, X_n$, it is natural to want to determine which of models (4.3) and (4.5) best describes the instantaneous relationship between $X(t)$ and $X'(t)$. This section will first discuss the techniques for assessing and comparing the fits of these two models, as developed by Verzelen et al. (2012). Then, assuming the presence of observations from n realizations, $W_1, \ldots, W_n$, of W, we propose a statistic to test if including both $W(t)$ and $X(t)$ leads to a significant increase in explaining the variation in $X'(t)$ when compared to $X(t)$ alone. In other words, we propose a test to assess whether the model given in (4.6) provides a superior fit to the data than that in (4.3).

We begin our discussion by reviewing the method of Verzelen et al. (2012) for assessing the fit of the model given in (4.3) as well as the specific case in (4.5). When the fit is good, $X(\cdot)$ may be close to the solution of the equation $X'(t) = f(t, X(t))$. To assess whether $X(\cdot)$ can be viewed this way, we determine the relative size of $Z(\cdot)$ using the decomposition of variance $\mathrm{Var}[X'(t)] = \mathrm{Var}[f(t, X(t))] + \mathrm{Var}[Z(t)]$. This decomposition allows us to assess the fraction of variance of $X'(t)$ explained by $f(t, X(t))$ using the coefficient of determination (Müller and Yao, 2010; Verzelen et al., 2012):

$$R^2(t) = \frac{\mathrm{Var}[f(t, X(t))]}{\mathrm{Var}[X'(t)]} = 1 - \frac{\mathrm{Var}[X'(t) - f(t, X(t))]}{\mathrm{Var}[X'(t)]}. \tag{4.7}$$

On sub-domains of $\mathcal{T}$ when $R^2(t)$ is close to 1 then $X(\cdot)$ may be close to the solution of the equation $X'(t) = f(t, X(t))$. Given n realizations, $X_1, \ldots, X_n$, a natural estimate of $R^2(t)$ is

$$\hat{R}^2(t) = 1 - \frac{\sum_{i=1}^{n} [\hat{X}_i'(t) - \hat{f}(t, \hat{X}_i(t))]^2}{\sum_{i=1}^{n} [\hat{X}_i'(t) - \bar{\hat{X}}'(t)]^2}. \tag{4.8}$$

The numerator of this estimate is the sum of the squared residuals from the estimate of f, while the denominator is $(n-1)$ times the sample variance of the $\hat{X}_i'(t)$'s.

When one wishes to assess the fit of the simpler, linear model in (4.5) then (4.8) reduces to

$$\hat{R}_L^2(t) = 1 - \frac{\sum_{i=1}^n [\hat{X}_i'(t) - \hat{\mu}^*(t) - \hat{\beta}(t)\hat{X}_i(t)]^2}{\sum_{i=1}^n [\hat{X}_i'(t) - \bar{\hat{X}}'(t)]^2}. \tag{4.9}$$

Since for a given set of trajectories of X, one does not know whether or not the linear form of $f(t, X(t))$ suffices, one may wish to compare the fit of this linear form in (4.5) to the fit of the nonlinear in (4.3). For this, one can compare the two corresponding R^2 values. When the ratio between $\hat{R}_L^2(t)$ in (4.9) to $\hat{R}^2(t)$ in (4.8) is small, then the nonlinear model provides a significantly better fit to the data. Conversely, when this ratio is large, the simpler linear model may be acceptable.

The above comparison is valid when one is interested in what type of effect $X(t)$ has on $X'(t)$. To study the improvement in the fit given by including $W(t)$, as in model (4.6) of Section 4.3, we propose modifying (4.7) to the coefficient of determination given by

$$R_W^2(t) = \frac{\mathrm{Var}[f(t, X(t), W(t))]}{\mathrm{Var}[X'(t)]} = 1 - \frac{\mathrm{Var}[X'(t) - f(t, X(t), W(t))]}{\mathrm{Var}[X'(t)]} \tag{4.10}$$

and estimated by

$$\hat{R}_W^2(t) = 1 - \frac{\sum_{i=1}^n [\hat{X}_i'(t) - \hat{f}(t, \hat{X}_i(t), W_i(t))]^2}{\sum_{i=1}^n [\hat{X}_i'(t) - \bar{\hat{X}}'(t)]^2}. \tag{4.11}$$

To determine whether or not $W(t)$ provides a substantially better fit to the data than the model only including $X(t)$, we compare $R_W^2(t)$ with $R^2(t)$ via

$$r^2(t) \equiv \frac{R_W^2(t)}{R^2(t)} = \frac{\mathrm{Var}[X'(t) - f(t, X(t), W(t))]}{\mathrm{Var}[X'(t) - f(t, X(t)]}. \tag{4.12}$$

When $r^2(t)$ is large, we have a comparable amount of variation of $X'(t)$ explained by models (4.3) and (4.6), indicating that including $W(t)$ does not provide a significantly better fit. On the other hand, when this ratio is small, it indicates that the fit with $W(t)$ included is better to that without $W(t)$ and that model (4.3) is more appropriate. Following from (4.8) and (4.11), we estimate $r^2(t)$ by

$$\hat{r}^2(t) \equiv \frac{\hat{R}_W^2(t)}{\hat{R}^2(t)} = \frac{\sum_{i=1}^n [\hat{X}_i'(t) - \hat{f}(t, \hat{X}_i(t), W_i(t))]^2}{\sum_{i=1}^n [\hat{X}_i'(t) - \hat{f}(t, \hat{X}_i(t))]^2}. \tag{4.13}$$

To test the null hypothesis that, for all t, the overall instantaneous relationship between X and X' at t is unaffected by $W(t)$ we propose the test statistic

$$S = \int_{t \in \mathcal{T}} \hat{r}^2(t) dt. \tag{4.14}$$

Since in practice, one does not have data at all $t \in \mathcal{T}$, we replace S by

$$\hat{S} = \sum_{k=1}^{J} \hat{r}^2(t_k^*), \tag{4.15}$$

for some evenly spaced grid of points, $t_1^* < \cdots < t_k^*$. When $\hat{S}$ falls below a given quantile of its null distribution, we reject the null hypothesis and conclude that (4.6) provides a better overall fit then model (4.3).

The null distribution of S is unknown and therefore must be approximated. To estimate this null distribution and thus the corresponding p-value of the test statistic, a permutation test with the W_i's is used. That is, from our derived observations $\{(\hat{X}_i, W_i), i = 1, \ldots, n\}$ we obtain N new data sets. The k^{th} data set is $\{(X_i, W_{\gamma(i)}), i = 1, \ldots, n\}$, where $\gamma(1), \ldots, \gamma(n)$ is a random permutation of the ordered set, $\{1, \ldots, n\}$. Via a permutation, each of the observed trajectories W_j is randomly assigned to a new "partner" $\hat{X}_i$. Heuristically, in the new data set, there will be no dependence between X' and W as the trajectories X_i and W_j are paired together at random. Note that this permutation induces a distribution that is a special case of the null distribution. The null distribution does not require independence of X and W. Rather, the assumption of the null distribution is that $E[X'(t)|X(t), W(t)]$ does not depend on $W(t)$, a weaker assumption.

For each of these N data sets, the smoothing described in Section 4.4 is then carried out for each of the two models (4.3) and (4.6) and the sample ratios, $\hat{r}_1^2(t), \hat{r}_2^2(t), \ldots, \hat{r}_N^2(t)$, for $t \in \{t_1^*, \cdots, t_k\}$ along with the test statistics $\hat{S}_1, \hat{S}_2, \ldots, \hat{S}_N$ are computed. We then use the empirical distribution of $\hat{S}_1, \hat{S}_2, \ldots, \hat{S}_N$ to provide an approximation the null distribution of S. When the observed ratio falls below a given quantile of this approximate null distribution, we conclude that, in terms of instantaneous dynamics, the two processes, X and W, account for a significantly higher portion of the variation in X', when compared to X alone.

Admittedly, the test statistic S combines information across t. In many applications, $W(t)$ might only be significant to the relationship between $X(t)$ and $X'(t)$ on certain sub-domains of $\mathcal{T}$, while not significant overall. The nonparametric method, described above, also allows us to explore this possibility by approximating the null distribution of $\hat{r}^2(t)$ for each fixed t.

For a fixed t, the observed value of $\hat{r}^2(t)$ can be compared to the distribution of $\hat{r}_1^2(t), \hat{r}_2^2(t), \ldots, \hat{r}_N^2(t)$. When $\hat{r}^2(t)$ is below a given quantile, we conclude that $W(t)$ together with $X(t)$ provides a significantly better fit to $X'(t)$, when compared to the fit provided by $X(t)$ alone. Of course, if this comparison of the observed value, $\hat{r}^2(t)$ is done for many t's, one should be aware of multiple testing issues and a correction factor to the rejection level could be applied.

Chapter 5

Dynamics of Mouse Growth Data

The methods described in Chapter 4 can be used to estimate the dynamics of the mouse growth data outlined in Chapter 2. We first briefly describe the smoothing of the data from each of the eight groups of mice. We then use the model fitting procedure outlined in Section 4.4, both including and not including the amount eaten. These models provide insight into not only the individual groups, but also the differences between them. Specifically, we attempt to understand the growth processes of the mice in the eight groups and how these processes vary. Finally, for each group, we compare the fits of these models and study if and when the amount eaten plays a significant role in the relationship between body mass and growth rate.

5.1 Model Fitting

5.1.1 Growth Rate Depending on Body Mass

As outlined in Section 4.4, before fitting a nonparametric regression model, the raw data are smoothed to approximate the underlying process and its derivative. For each mouse, we use smoothing splines using the statistical programming language R (R Core Team, 2015) to obtain the estimated body mass curve and the subsequent growth rate as a function of age. We set the smoothing parameter by setting the degrees of freedom, as outlined in Section 3.3. For the smoothing of the body masses we set the degrees of freedom to 10, while for the estimated growth rates they are set to 5. Here we fix the degrees of freedom, rather than using cross-validation to ensure that each mouse has a comparable amount of smoothing. Moreover, in the case of derivative estimation, cross-validation does not perform reliably. Figure 5.1 shows the smoothed body masses, $\hat{X}_i(t)$, of mice in each of the eight groups. In general, the male mice are heavier than their female counterparts. Likewise, the sedentary mice appear heavier than the active. For both males and females, the heaviest individual mice belong to the sedentary control

groups.

Figure 5.2 shows the estimated growth rates, $\hat{X}_i'(t)$, of mice in each of the eight groups. As one would expect, the growth rates at early weeks are much higher than those in the middle or later weeks. For some mice estimated growth rates are negative during later weeks. This characteristic is most prevalent in the male sedentary groups. Furthermore, the flattening out of the curves in many of the panes in Figure 5.2 indicates that, for most mice, the growth rate is relatively constant after the younger ages.

As outlined in Section 4.4, to estimate the deterministic part, $f(t, \cdot)$, of the nonlinear differential equation in equation (4.3), we use a local linear estimate (Bowman and Azzalini, 2014) where for each t, the estimated growth rate, $\hat{X}_i'(t)$, is regressed on the smoothed body mass, $\hat{X}_i(t)$. The kernel is chosen to be Gaussian and the smoothing parameter is selected by cross validation. Figure 5.3 shows this estimate, $\hat{f}(t, x)$, at weeks $t = 5$, 30, 55 and 80 from the active male mice from the control group. As we can see, the nonparametric approach picks up on some of the nonlinear trends in the data. At $t = 5$ the relationship between body mass and growth rate is monotone increasing. At $t = 30, 55$, and 80 the relationship appears to be relatively constant with noise. However, for $t = 30$ and 80, there may be slight increases in growth rate for increase in body mass for mice within the middle weights. For each of the eight groups and each age, plots such as those in Figure 5.3 can be found in the digital appendix in the subdirectory */1D_Plots*.

We can further examine the active male mice from the control group through the contour plot in Figure 5.4, which displays the entire estimate $\hat{f}(t, x)$. The values of $\hat{f}$ range from around 0.05 to 0.1 at the early ages, while at the older ages they are nearly all below 0.025. Examining the relationship between body mass and expected growth rate given in Figure 5.4 we see that all young mice have high growth rates, regardless of body mass. From ages 10 to 35, growth rate is high for both light and heavy mice. Further, from ages 35 to 60, growth rate decreases with increasing body mass. Lastly, for ages beyond 60 weeks, the pattern is unclear. Similar plots for the other 7 groups show this pattern as well and can be found in the digital appendix in the subdirectory */Contour_1D_Fits*.

Figure 5.5 provides a comparison of the estimated relationship between body mass and growth rate for each of the eight groups. Within each panel of Figure 5.5 we see 80 curves, each representing $\hat{f}(t, \cdot)$ for a fixed age t. The darker the color of the curve, the larger the value of t. Although one can attempt to draw some general conclusions from Figure 5.5, the best way to determine how the relationship between body mass and growth rate changes

over time is to examine the data and its corresponding curve at each t. Table 5.1 provides a description of this relationship for each of the eight groups. As previously mentioned, the individual scatter plots can be found in the subdirectory */1D_Plots*.

Mouse Group	**Description of $\hat{f}(t, x)$**
F.Act.Ctrl	• Relatively flat between weeks 1 and 15, with a slight uptrend and one exceptionally heavy mouse. • Monotone increasing between weeks 16 and 50, with the exception of two heavy mice. • For weeks 50 and 80, there are several heavy mice whose growth rates don't fit the pattern. Other then these mice, the estimate still appears to be monotone increasing.
F.Act.Sel	• Relatively flat between weeks 1 and 15 except for the lightest/heaviest mice. • Monotone increasing between weeks 16 and 45, with the exception of the heaviest mice, which have low growth rates. • Relatively flat with a slight uptrend from weeks 45 and 80.
F.Sed.Ctrl	• Monotone increasing between weeks 1 and 9, with the exception of the heaviest mice, which have low growth rates and the lightest mice, which have high growth rates. • Monotone increasing between weeks 10 and 40. • For weeks 41 and 49 the fits are poor due to a number of heavy mice who have very different growth rates. • Flat between weeks 50 and 60. • For weeks 60 and 80, the fit is again poor due to the heavy mice with different trends.

F.Sed.Sel	• Noisy upward trend from weeks 1 and 15, with the exception of two heavy mice. • Monotone increasing from weeks 16 and 40. • Flat from weeks 40 and 59, with the exception of one heavy mouse whose growth rate is much lower than others. • Flat from weeks 60 and 80, with the exception of a slight uptrend from weeks 75 and 80.
M.Act.Ctrl	• Monotone increasing from weeks 1 and 30. • Flat with a lot of noise from weeks 30 and 80.
M.Act.Sel	• Monotone increasing from weeks 1 and 30, with the exception of two light mice. • Flat with noise from weeks 30 and 80. From weeks 30 to 75 there is one exceptionally heavy mouse with a high growth rate. From weeks 65 t 80 there are two light mice with very low growth rates.
M.Sed.Ctrl	• Increasing overall trend (with lots of noise) from weeks 1 and 50. • Relatively flat with noise from weeks 50 and 80. There are several mice with negative growth rates that lead to poor fits at some weeks.
M.Sed.Sel	• Monotone increasing between weeks 1 and 40. • Relatively flat (with noise) from weeks 40 and 80.

Table 5.1: The general trends in the estimated instantaneous relationship between body mass and growth rate for each of the eight groups as seen in the subdirectory */1D_Plots* of the digital appendix.

Table 5.1 allows us to make some general conclusions about the groups. In both the female active groups there appears to be an increase in estimated

growth rate with an increase in body mass, with the exception of the early weeks. On the other hand, the female sedentary mice show this increase at the young and middle ages but not at the older ages. In all four of the male groups we see a similar trend of an increase in estimated growth rate with an increase in body mass up to week 30-50.

5.1.2 Including Amount Eaten

The above analysis focuses strictly on the instantaneous relationship between the body mass and growth rate in each of the eight groups. We now fit the model given in equation (4.6), where $W(t)$ is the amount eaten in week t. The weekly amount eaten is displayed for each group in Figure 5.6. As mentioned at the end of Section 4.4, we choose not to smooth W, as biologically there is no reason to assume the process is intrinsically smooth.

The model is fit via local linear bivariate kernel smoothing (Bowman and Azzalini, 2014), as outlined in Section 3.2, using the multiplicative kernel in equation (3.4), where $K_1 = K_2$ are Gaussian. The bandwidths, h_1 and h_2, are chosen by cross validation. For each t, the evaluation points for the estimate, $\hat{f}(t, x, w)$, form an equally spaced grid on $[\min_i \hat{X}_i(t), \max_i \hat{X}_i(t)] \times [\min_i W_i(t), \max_i W_i(t)]$.

Including the process W in the model makes visual representations analogous to Figures 5.4 and 5.5 difficult. Rather, the best we can do is to make contour plots for the expected growth rate as a function of body mass and amount eaten at each t. Figure 5.7 gives an example of these contour plots at $t = 5, 30, 55$ and 80 for the active males from the control group. These plots are at the same ages and in the same group as the plots in Figure 5.3, which do not include W. In Figure 5.3, at $t = 5$, expected growth rate increases with body mass. In Figure 5.7, we see a similar trend: at $t = 5$, for a fixed amount eaten there appears to be an increase in expected growth rate with an increase in body mass, with the exception of when the amount eaten is high. At $t = 30$ the expected growth rate is relatively constant as in Figure 5.3, with the exception of one heavy mouse, who can be seen in the upper right pane of Figure 5.3. For $t = 55$, for lighter mice there is an increase in expected growth rate as the amount eaten increases. For moderate and heavy body masses, the relationship is constant. Finally, at $t = 80$, the growth rate is constant with the exception of the heavier mice. For these mice, it appears that an increase in the amount eaten increases the expected growth rate.

Similar plots from 80 times points and 8 different groups, are included in the subdirectory */2D_Plots* in the digital appendix. This subdirectory

also contains animations of the 80 contour plots for each of the 8 groups in .gif files. These allow for some interpretation of how the amount eaten and body mass effect the expected growth rate. These interpretations are discussed at the end of the proceeding section.

5.2 Model Comparisons

For each of the eight groups, we now compare the models in equations (4.3) and (4.6) using the techniques outlined in Section 4.5. Recall that we quantify the extent to which the deterministic part of models (4.3) and (4.6) explains the variation in $X'(t)$ with the ratios $R^2(t)$ and $R^2_W(t)$, given in equations (4.8) and (4.11), respectively. Figure 5.8 shows a comparison between $R^2(t)$ and $R^2_W(t)$ for each of the eight groups, while Figure 5.9 shows a 95% bootstrap confidence interval for $R^2_W(t)$ for the male active control group. These point-wise confidence intervals were obtained by computing $R^2_W(t)$ for each of 200 samples (with replacement) of the data. The standard errors of the resulting 200 $R^2_W(t)$'s were then computed at each t and used with standard normal quantiles to construct the confidence bands. Similar plots for each of the eight groups can be found in the digital appendix in the */Bootstrap* subdirectory. We now make some observations about the model fits and some comparisons between groups. These observations are strictly exploratory as there are no standard values of $R^2_W(t)$ and $R^2(t)$ to be compared to.

Firstly, for all eight groups and all time points, we have $R^2_W(t) \geq R^2(t)$, thus indicating that the amount of variation in the growth rate explained by model (4.6) is uniformly higher than model (4.3). This is expected, as model (4.6) is richer than (4.3). In general, the values of R^2_W and R^2 are fairly high for ages 10-30, while for many of the groups, we see a drop in the values of R^2_W and R^2 between weeks 40 and 50. This drop is generally followed by an uptrend for the later ages, especially for model (4.6) .

Secondly, we examine the differences between groups. In both the male and female sedentary control groups, at young ages, the values of R^2_W and R^2 are large. In contrast, for older ages, only R^2_W is large. The sedentary selected males and females are similar to the sedentary control groups, with the exception that R^2_W is small for the later ages in the male group. In the active selected groups, R^2 has uniformly low values for both the male and females. Model (4.6) is slightly better, although R^2_W is still relatively low, especially in the females. In general, both models seem to fit the active control groups better, in particular at the younger and older ages.

We now formally test whether the fit from model (4.6) is significantly better than that from model (4.3). For this test, we carry out the hypothesis testing procedure based on the test statistic $\hat{S}$, outlined in Section 4.5. For each of the eight groups we use 500 permutations of the amounts eaten.

Figure 5.10 gives the approximated null density of $\hat{S}$ for the sedentary female group as well as the observed value of $\hat{S}$. Recall that when $\hat{S}$ is small, we reject the null hypothesis and conclude that the model in (4.6) provides a significantly better fit to that in (4.3). In Figure 5.10 the observed value of $\hat{S}$ is less than the 5^{th} percentile of the approximated null density and therefore the model in (4.6) provides a significantly better fit. This indicates that the amount eaten in week t significantly effects the relationship between body mass and growth rate at week t, at least for some values of t. The densities for the other seven groups are similar and can be seen in the digital appendix under */Density_Curves*. The resulting p-values are displayed in Table 5.2. At a rejection level of 5%, only the null hypothesis for the sedentary female groups (both control and selected) is rejected. That being said, the p-values are all lower for the sedentary groups versus their active counterparts. This indicates that there is more evidence of an effect of the weekly amount eaten on the instantaneous relationship between body mass and growth rate for the sedentary mice.

Figure 5.11 gives the ratios $\hat{r}^2(\cdot)$ in equation (4.13) of Section 4.5 for each of the eight groups. The dashed lines correspond to the point-wise 5^{th} and 95^{th} percentiles resulting from the permutation testing method given in Section 4.5. When $\hat{r}^2(t)$ falls below the 5^{th} percentile, this indicates that the value is unusually low, providing evidence that the amount eaten is significant for explaining the growth rate at time t, and thus in favor of the alternative hypothesis given in model (4.6). Conversely when the value lies significantly above the 5^{th} percentile, there is no evidence that the amount eaten is significant in the instantaneous relationship between body mass and growth rate.

Given the p-values in Table 5.2 it is natural to want to determine for which weeks and in what way the weekly amount eaten is significant in the instantaneous relationship between body mass and growth rate. To answer these questions, we examine Figure 5.11, the point-wise p-values obtained from the permutation method, provided in Figure 5.12, as well as the individual contour plots in */2D_/Plots* of the digital appendix. The contour plots are challenging to interpret, particularly since it is difficult to tell if results are driven by a few unusual mice. Note that a contour plot indicates that W is not important in determining expected growth rate if each vertical line on the contour plot is of one colour/growth rate. We focus

	Active		Sedentary	
	Control	Selected	Control	Selected
Female	0.752	0.330	0.002	0.000
Male	0.330	0.114	0.150	0.060

Table 5.2: The p-values for each of the eight groups, resulting from the permutation test from Section 4.5 to test the null hypothesis that the amount eaten at week t does not significantly effect the instantaneous relationship between body mass and growth rate.

our discussion on the sedentary females groups as they were the only two groups with overall p-values less than 0.05 in Table 5.2.

In the selected females who were sedentary, we have an overall p-value of 0.000. The top right panel of Figure 5.12 indicates for this group that the weekly amount eaten is significant during weeks 3 and 4, 15 and 16, 21 to 33, 38, 44 and 45, 58 to 65 and 71 to 74. To determine how the amount eaten effects the expected growth rate, we examine the contours in */2D_Plots_/F.Sed.Sel* of the digital appendix. During weeks 3, 4, 15 and 16 the growth rate is mostly constant, with the exception of the heavy mice where the expected growth rate is decreasing as the amount eaten increases. For weeks 21 to 33 the growth rate is constant for the lighter mice. For the heavier mice, expected growth rate initially decreases with an increase in amount eaten and then increases. Weeks 38, 44 and 45 are similar, but the pattern is less clear. During weeks 58 to 65 and 71 to 74, the heaviest mice have by far the fastest growth rate, making it difficult to see other patterns. That being said, in some of these weeks it appears that an increase in food consumption leads to an increase in expected growth rate.

For the sedentary females from the control group, whose overall p-value is 0.002, it appears that for many ages past week 40, the weekly amount eaten is significant in explaining the growth rate. Specifically, weeks 40 to 42, 47 to 52, 60 to 62 and 68 to 75 all have p-values less than 0.05. During weeks 40 to 42 and 47 to 52, for the heavier mice, the growth rate is largest for an intermediate value of amount eaten. In addition, during weeks 47 to 52, for the lighter mice, the expected growth rate increases as the amount eaten increases. For weeks 60 to 62 there appears to be an increase in expected growth rate for an increase in amount eaten, although one fast growing and one shrinking mouse at weeks 61 and 62 make these observations slightly more difficult to see. We see a similar, but small, increase for moderate and

heavy mice during weeks 68 to 75. During these weeks there are some light mice with negative growth rates that do not show this pattern.

We conclude this section by pointing out some important facts. Firstly, since for each group we are testing 80 different hypotheses (one for each week), we must be aware of multiple comparison issues and therefore exercise caution over significant results. Secondly, as will be discussed in Chapter 6, the point-wise power of the above test is often low. Therefore, it is possible that some of the p-values are higher than they would be under a superior testing method. As discussed, individual analysis of the contour plots can provide insight into whether this is the case. That being said, complex nonlinear relationships are not always obvious from such plots and therefore the possibility exists that some of our conclusions could be further refined. Finally, we point out that each group is made up of 30-40 mice and that this is a somewhat small number for fitting nonparametric models. An increase in sample size may provide greater clarity in the individual contour plots as currently a small number of mice who exhibit different trends from the true underlying relationship may cloud the figures.

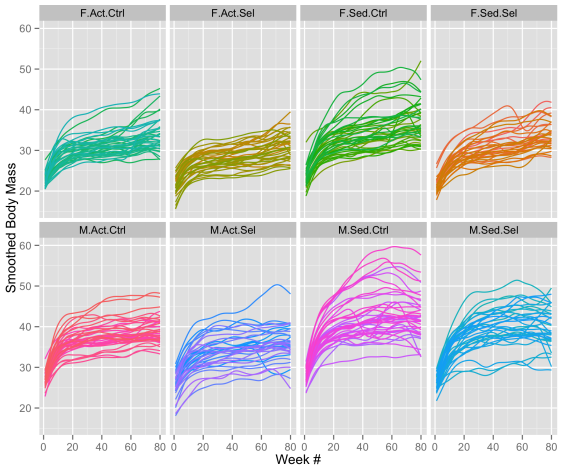


Figure 5.1: The smoothed body masses (grams) for each of the eight groups of mice as a function of age (weeks).

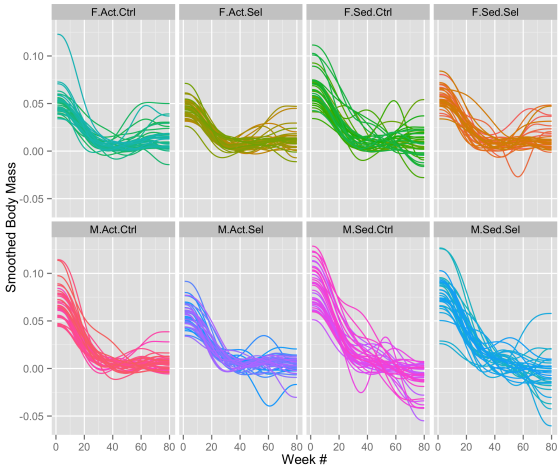


Figure 5.2: The estimated growth rates (grams/week) for each of the eight groups of mice as a function of age (weeks).

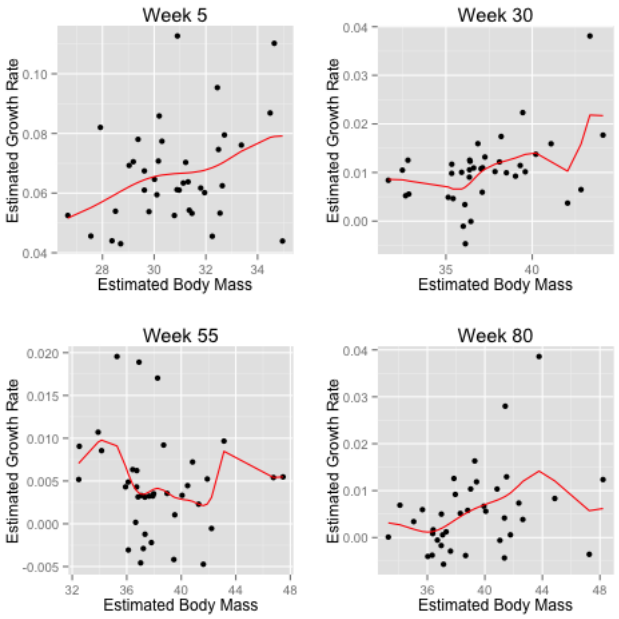


Figure 5.3: An example of the estimated deterministic component, $\hat{f}(t, \cdot)$, as a function of body mass (grams) for the active males from the control group at weeks 5, 30, 55, and 80.

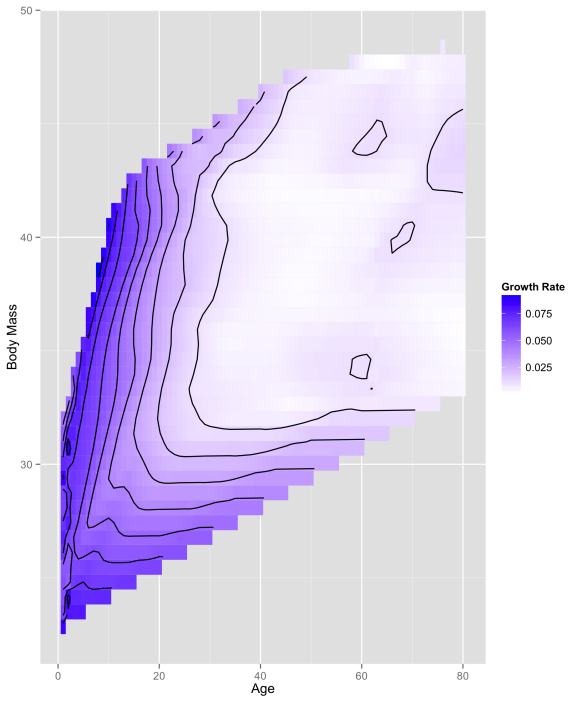


Figure 5.4: Contour plot of the nonparametric estimate, $\hat{f}(t, x)$, for the active male mice from the control group. The x-axis corresponds to age (weeks), while the y-axis is body mass (grams).

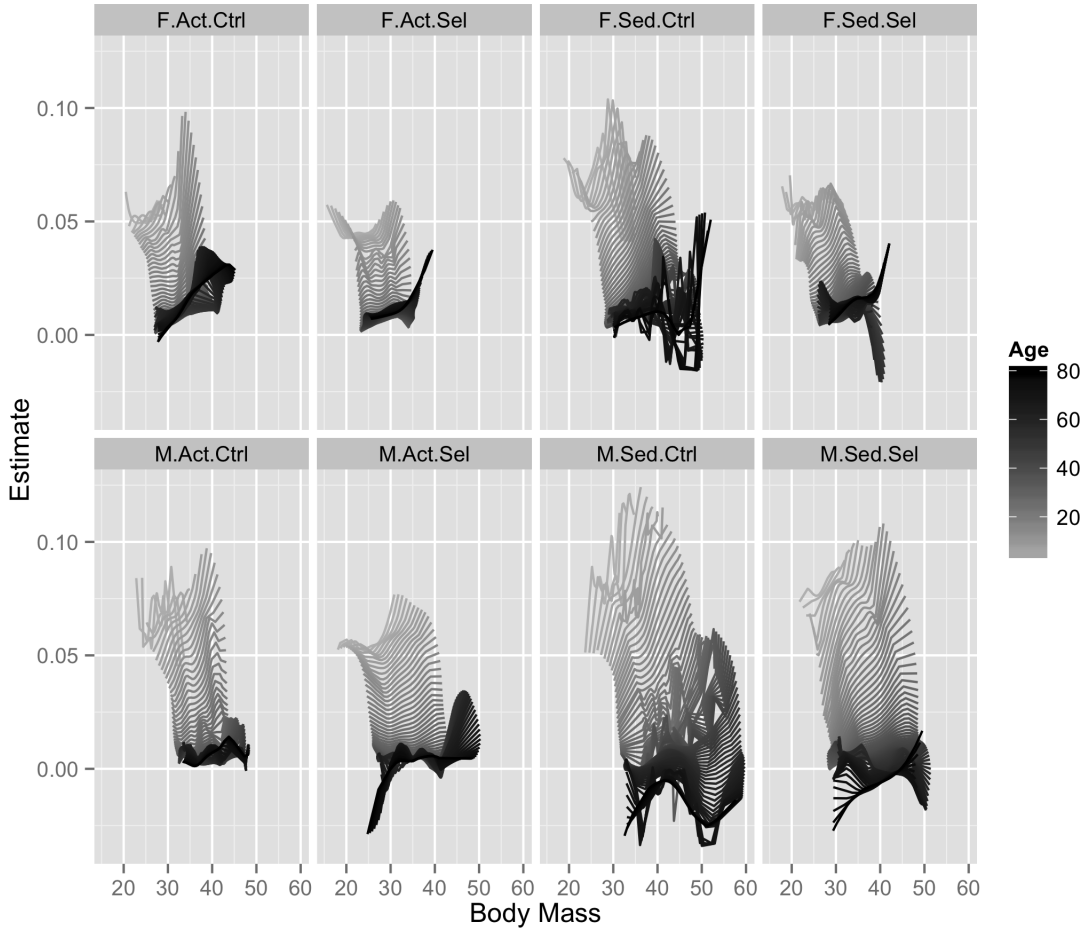


Figure 5.5: The estimated relationship, given by $\hat{f}(t, x)$, between body mass (grams) and growth rate (grams/week) for each of the eight groups. The increase in darkness of the curves indicates an increase in age.

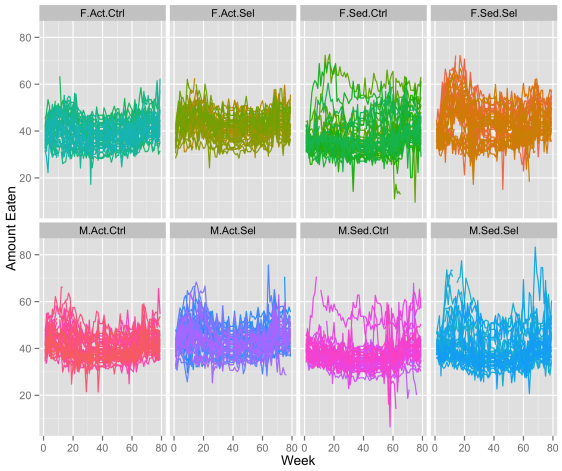


Figure 5.6: The amount eaten (grams) as a function of age (weeks), organized into each of the eight groups. This is the same as Figure 2.8, with the exception that here the outliers have been removed.

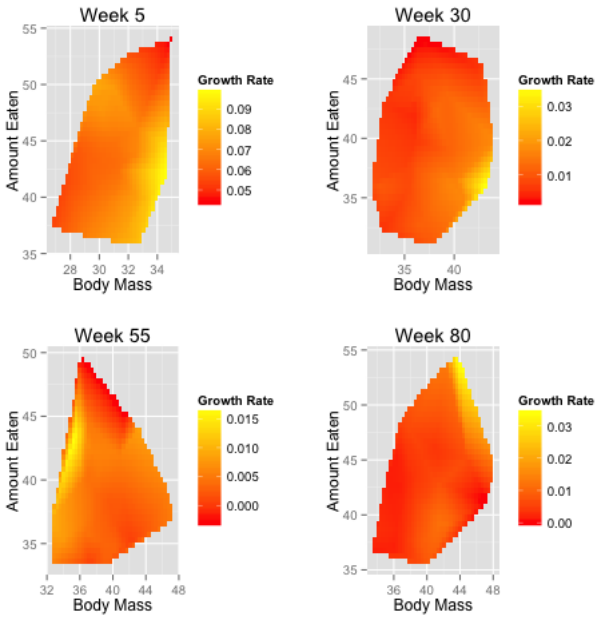


Figure 5.7: A sample of four contour plots, at weeks 5, 30, 55 and 80, from the control males who were active. The x-axis indicates body mass (grams), while the y-axis corresponds to the weekly amount eaten (grams). The coloring is based on the value of $\hat{f}(t, x, w)$, the conditional expected growth rate.

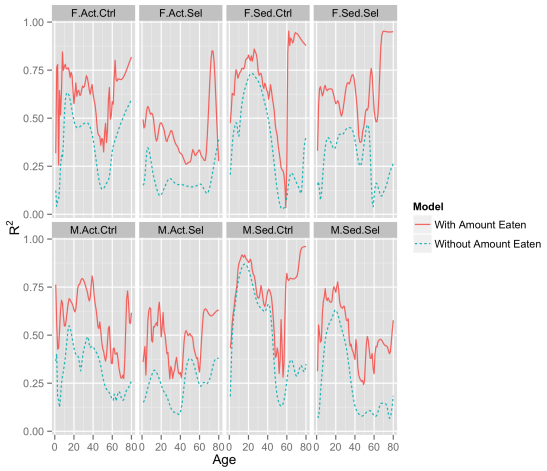


Figure 5.8: A comparison of $R^2(t)$ (dashed) and $R^2_W(t)$ (solid) as a function of age (weeks) for each of the eight groups.

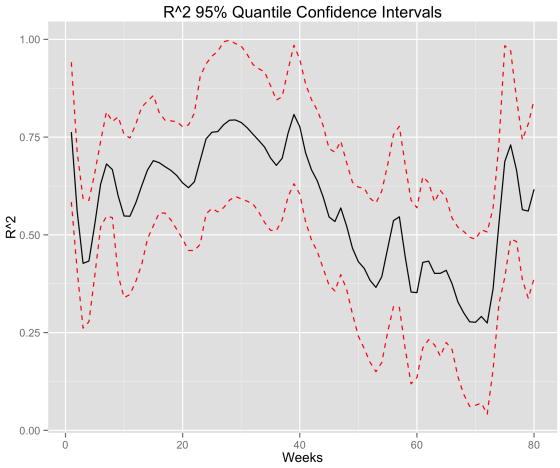


Figure 5.9: A 95% bootstrap confidence interval (dashed,red) for $R^2_W(t)$ for the active males in the control group.

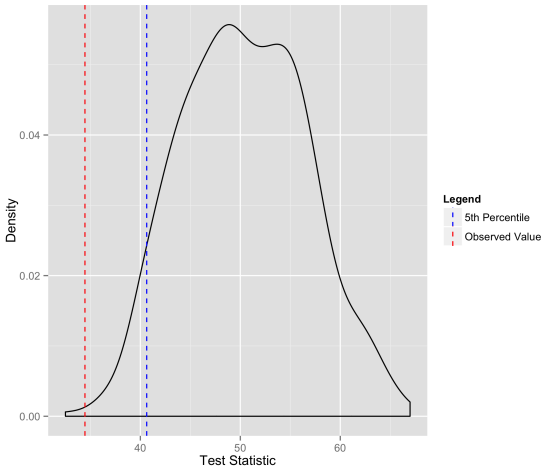


Figure 5.10: The approximated density of $\hat{S}$ for the sedentary females from the control group. The red line corresponds to the observed value from the data, while the blue line indicates the 5^{th} percentile of the approximated null density of $\hat{S}$.

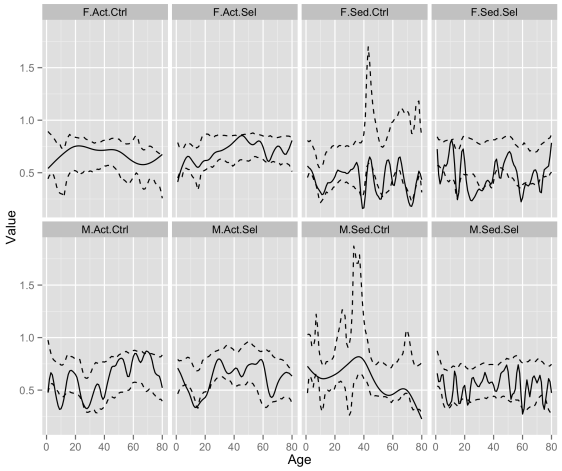


Figure 5.11: The ratio $\hat{r}^2$ as a function of age (weeks) for each of the 8 groups. The dashed lines represent the 5th and 95th percentiles resulting from the permutation method described in Section 4.5.

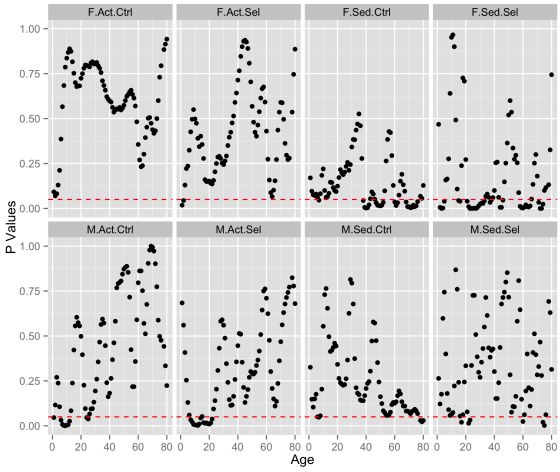


Figure 5.12: The point-wise p-values for each of the eight groups resulting from the permutation test in Section 4.5 of the null hypothesis that the amount eaten at week t does not significantly effect the instantaneous relationship between body mass and growth rate.

Chapter 6

Simulation Study

To assess the statistical properties of our methods for the model comparisons proposed in Section 4.5, we carry out a series of simulations based on the data collected on mice, as described in Chapter 2.

Recall that we are interested in how body mass, X, and the amount eaten, W, at a given age, t, effect the growth rate in the eight different groups of mice. Our findings for each of the eight groups were outlined in Chapter 5. Specifically, we tested the hypothesis that the amount eaten at age t has a significant effect on the relationship between body mass and growth rate at t. To assess the statistical properties of our methods for this hypothesis test, we simulate data sets which are similar to those in Chapter 2 and carry out the analysis described in Chapter 5. This is repeated for various levels of correlation between the body mass and the amount eaten. For simplicity, we only simulate data sets based on one of the eight groups.

6.1 Models for Simulated Data

We generate data, $\tilde{\boldsymbol{Y}}_{\boldsymbol{i}}, \tilde{\boldsymbol{W}}_{\boldsymbol{i}} \in \mathbb{R}^J$, for $i = 1, \ldots, n$, that are independent identically distributed as $(\tilde{\boldsymbol{Y}}, \tilde{\boldsymbol{W}})$, where $\tilde{\boldsymbol{Y}} = (Y_1, \ldots, Y_J), \tilde{\boldsymbol{W}} = (W_1, \ldots, W_J)$, with

$$\tilde{Y}_j = X(t_j) + \epsilon_j \tag{6.1}$$

$$\tilde{W}_j = a + \gamma e(t_j) \tag{6.2}$$

for $j = 1, \ldots, J$, where

$$\begin{aligned}
&X(t_j) = \mu_X(t_j) + \sum_{k=1}^{K} \alpha_k \varphi_k(t_j), \\
&\mu_X(t_j) = \mathrm{E}[X(t_j)], \\
&\alpha_k \sim \mathcal{N}(0, \lambda_k) \text{ and } \mathrm{Cov}[\alpha_k, \alpha_l] = 0 \text{ for all } k \neq l, \\
&\epsilon_j \sim \mathcal{N}(0, \sigma_j^2), \text{ and } \mathrm{Cov}[\epsilon_k, \epsilon_l] = 0 \text{ for all } k \neq l, \\
&\mathrm{Cov}[\alpha_k, \epsilon_l] = 0 \text{ for all } k \text{ and } l
\end{aligned}$$

and

$$
\begin{aligned}
&a \sim \mathcal{N}(\mu_a, \sigma_a^2),\\
&\mathrm{E}(e(t)) = 0, \mathrm{Cov}(e(s), e(t)) = \mathcal{I}(s = t),\\
&\mathrm{Cov}[a, e(t)] = 0 \text{ for all } t,\\
&\mathrm{Cov}[\alpha_k, a] = c_k \text{ for all } k,\\
&\mathrm{Cov}[\epsilon_j, a] = 0 \text{ for all } j,\\
&\mathrm{Cov}[\epsilon_j, e(t)] = 0 \text{ for all } j \text{ and } t.
\end{aligned}
$$

One can think of the $\tilde{\boldsymbol{Y}}_{\boldsymbol{i}}$ and $\tilde{\boldsymbol{W}}_{\boldsymbol{i}}$ as the simulated noisy body masses and amounts eaten, respectively, of the i^{th} mouse.

We now discuss how the values of the above parameters are chosen in order to generate $\tilde{\boldsymbol{Y}}_{\boldsymbol{i}}$ and $\tilde{\boldsymbol{W}}_{\boldsymbol{i}}$. To choose these parameters we use estimates based on the data from the group of male mice who were active on wheels and were from the control breeding group. We denote these data as (t_j, y_{ij}, w_{ij}) for $i = 1, \ldots, n = 38$ and $j = 1, \ldots, J = 79$, where y_{ij} denotes the body mass and w_{ij} denotes the amount eaten for mouse i at time t_j. In addition, we denote the smoothed body mass of mouse i at time t_j as x_{ij}.

To generate the X_i's, that is, to determine values for $K, \lambda_1, \ldots, \lambda_K$ and $\varphi_k(t_1), \ldots, \varphi_k(t_J)$, we perform a principal component analysis on the data vectors $(x_{i1}, \ldots, x_{iJ})^t$ for $i = 1, \ldots, n$. From this PCA, we choose $K = 2$, as the first two principal components explain 95.56% of the variation in the data vectors. We then set $(\varphi_1(t_1), \ldots, \varphi_1(t_J))$ and $(\varphi_2(t_1), \ldots, \varphi_2(t_J))$ to be the eigenvectors from the PCA, shown in Figure 6.1, and $(\lambda_1, \lambda_2) = (639.89, 24.71)$ to be the corresponding eigenvalues. Further, we set

$$\mu_X(t_j) = \frac{1}{n} \sum_{i=1}^{n} x_{ij}.$$

This function is displayed in Figure 6.2. For the error variances of the ϵ_j's we set

$$\sigma_j^2 = \sum_{i=1}^{n} (y_{ij} - x_{ij})^2 / n.$$

For the simulated amounts eaten, we set γ^2, μ_a and σ_a^2 as

$$\gamma^2 = \frac{1}{n(J-1)} \sum_{i=1}^{n} \sum_{j=1}^{J} (W_{ij} - \bar{W}_{i\cdot})^2 = 11.64 \text{ gm}^2,$$

where $\bar{W}_{i\cdot} = \sum_{j=1}^{J} W_{ij}/J$ and

$$\mu_a = \bar{W}_{\cdot\cdot} = \sum_{i,j} W_{ij}/(nJ) = 41.18 \text{ gm}$$

$$\sigma_a^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\bar{W}_{i\cdot} - \bar{W}_{\cdot\cdot})^2 - \frac{\gamma^2}{J} = 17.09 \text{ gm}^2.$$

To simplify the model we choose $\text{Cov}(\alpha_2, a) = 0$. This assumption is not unreasonable as the variance of α_2 is much smaller than that of α_1 and therefore φ_2 could possibly be omitted all together from modeling the dependence between X and W. Thus, as shown in Appendix (2), to have a proper covariance structure between $\tilde{\boldsymbol{Y}}_{\boldsymbol{i}}$ and $\tilde{\boldsymbol{W}}_{\boldsymbol{i}}$ it suffices that:

$$\text{Corr}^2(\alpha_1, a) < 1.$$

We simulate data using varying levels of correlation between α_1 and a. These levels are 0, 0.2, 0.4, 0.6, and 0.8. For each of these correlations, 500 data sets are generated and analyzed with the method of Chapter 4. An example of one of these data sets can be found in Figure 6.3. The proceeding section outlines the corresponding results.

6.2 Simulation Results

In this section we present the results from the simulation study described in Section 6.1. All simulations were carried out in R (R Core Team, 2015) and the results are displayed using ggplot2 (Wickham, 2009). As mentioned in Section 6.1, we simulate 500 data sets for each of $\text{Corr}(\alpha_1, a) = 0, 0.2, 0.4, 0.6, 0.8$. For each of these data sets, a p-value based on the test statistic

$$\hat{S} = \sum_{j=1}^{J} \hat{r}^2(t_j) \tag{6.3}$$

is calculated via the permutation method described in Section 4.5 to test the hypothesis that, for all t, the overall instantaneous relationship between X and X' at t is unaffected by $W(t)$. In the following, we have used 200 permutations for each data set.

Figure 6.4 shows the histograms of the 500 p-values obtained from the permutation method, for each of the five correlation levels. We first note that when the correlation between α_1 and a is 0, the distribution of the

p-values is approximately uniform. Moreover, as expected, when the correlation between α_1 and a and hence X and W is high, we observe that the resulting p-values are much lower than when the correlation is low. This indicates that when the correlation between the two processes is high, $\hat{S}$ can be reasonably expected to indicate whether or not the additional process, W, effects the overall instantaneous relationship between X and X'. When the correlation between the two is lower, we may not expect the test to adequately inform of the relationship. This is further illustrated in Figure 6.5 where the power curves of the test statistic, $\hat{S}$, are shown. For low correlation levels the test has little power to detect the significance of W on the instantaneous relationship between X and X', while for higher levels of correlation (and higher rejection levels), we observe substantially higher power. Tables 6.1 and 6.2 provide more detailed information about the values plotted in Figure 6.1, in particular, about the variability of the values displayed.

In Table 6.1 we display 95% confidence intervals for the achieved significance level of our test, that is, for the proportion of times the null hypothesis is rejected at various rejection levels, when the correlation between a and α_1 is 0. We expect that these confidence intervals should contain the nominal rejection level, given in the first column of Table 6.1. This is the case for each of the first four rejection levels (as well as nearly true for $\alpha = 0.2$), thus illustrating the viability of the test.

Further standard errors for the power of our 5% level test at various correlation levels can be seen in Table 6.2. The standard errors for the other rejection levels shown in Figure 6.1 are similar and are thus not shown here.

Alpha	Proportion Rejected	95% Confidence Interval
0.01	0.01	(0.00,0.02)
0.05	0.05	(0.03,0.07)
0.1	0.09	(0.07,0.12)
0.15	0.13	(0.10,0.16)
0.2	0.17	(0.13,0.20)

Table 6.1: The proportion of the null hypotheses rejected based on $\hat{S}$ when the correlation between a and α_1 is set to 0 as well as a 95% confidence interval of the expected proportion.

As mentioned in Section 4.5, as the test statistic $\hat{S}$ is a sum over all t_j, it cannot identify at which time points $W(t)$ and $X(t)$ better explain $X'(t)$

Correlation	Proportion Rejected	SE	95% Confidence Interval
0.0	0.050	0.010	(0.030,0.070)
0.2	0.070	0.011	(0.048,0.092)
0.4	0.080	0.012	(0.565,0.104)
0.6	0.140	0.015	(0.112,0.169)
0.8	0.390	0.022	(0.345,0.433)

Table 6.2: The proportion of null hypotheses rejected based on $\hat{S}$, at significance level of 0.05, their standard errors and a 95% confidence interval for the expected proportion rejected.

when compared to just $X(t)$. To determine this, for each fixed t_j, we test the hypothesis

$$\begin{aligned} H_0 &: E[X'(t_j)|X(t_j), W(t_j)] = f_0(t_j, X(t_j)) \\ H_1 &: E[X'(t_j)|X(t_j), W(t_j)] = f_1(t_j, X(t_j), W(t_j)) \end{aligned} \tag{6.4}$$

For this test, we obtain a p-value by using the empirical distribution of the 200 values of $\hat{r}^2(t_j)$ which result from the permutations of the data. Thereby, for each fixed t_j, we have 500 p-values (one for each simulated data set). To determine the point-wise power of our test, these 500 p-values are compared to a chosen rejection level to obtain the proportion of times the null hypothesis is rejected.

To give ourselves a measure for comparison, we fit for each fixed t_j, $j = 1, \ldots, J$, the linear model

$$\begin{aligned} X_i'(t_j) = \beta_0 + \beta_1(t_j)X_i(t_j) + \beta_2(t_j)W_i(t_j) + \tau_i \\ \tau_i \sim \mathcal{N}(0, \sigma_\tau^2) \end{aligned} \tag{6.5}$$

and test the hypothesis

$$\begin{aligned} H_0 &: \beta_2(t_j) = 0 \\ H_1 &: \beta_2(t_j) \neq 0. \end{aligned} \tag{6.6}$$

Testing this for each of the 500 data sets, using the normal linear model approach to computing p-values, allows us to estimate the power of this hypothesis test for each t_j. We can then compare the point wise power of our test, described in the preceding paragraph, with that obtained from the linear models test. Since we have generated data according to Gaussian distributions, we would anticipate that (6.5) holds (at least when error is

not included), and thus the standard linear model test of (6.6) would be a "gold standard". Of course, if (6.6) does not hold, then we should use the nonparametric approach outlined in Sections 4.4 and 4.5.

Figures 6.6 and 6.7 provide a comparison of the power of the tests of the hypotheses (6.4) and (6.6) as a function of the correlation between a and α_1 at a significance level of 5%. Predictably, for both tests, we observe much better performance for higher correlation between the generated observations.

For both the hypothesis test of the linear model and that of the nonparametric model we observe much higher power for younger ages when compared to older ages. While the joint distribution of $X'(t)$, $X(t)$ and $W(t)$ does depend on t in a complicated way, examining the contour plots for various ages, there does not appear to be an obvious reason for this higher power at younger ages.

Figure 6.8 shows the proportion of times the null hypothesis in (6.4) is rejected minus the proportion of times the null hypothesis in (6.6) is rejected, as a function of t. As seen in Figures 6.6 and 6.7, when the correlation between a and α_1 is low, neither test of the point-wise relationship has any sizable power and indeed, the tests are comparable, as can be seen in Figure 6.8. For high correlations, the linear model test has much higher power at young ages, but significantly lower power for $t \in (10, 60)$. This implies that for most ages, our method is an acceptable alternative to simply fitting linear models to the data. Moreover, the nonparametric method likely allows for the possibility of detecting nonlinear relationships in the data that a simple regression cannot.

In Table 6.3 we display, for each of the five correlation levels and a significance level of 0.05, the total number of times that each test correctly rejected, while the other test failed to reject, as well as when both or neither test rejected. As we can see, the majority of the time, neither of the tests reject. This indicates that to determine the significance of any relationship at a particular t_j, more work should be done to develop a more powerful alternative. That being said, we see that the test of (6.4) rejects a greater number of times than the test of (6.6). In particular, when the correlation is 0.8 there is a large difference in the total number of times the null hypotheses are rejected. It should be noted that, as mentioned above, the "gold standard" applies to the unsmoothed data with no error. Therefore, the actual linearity of the relationship may not perfectly hold after smoothing the noisy simulated data.

We conclude this section by describing some the potential issues with our study, as well as some ways it could be improved. Firstly, since each

simulated data set contains only 38 mice, we may observe greater power if the sample size were increased. Secondly, as mentioned in Section 4.5, our permutation method to approximate the null distribution of $\hat{S}$ only captures a subset of this null distribution. Using a different approach to approximating this underlying distribution may also lead to greater power. Further, due to computational restraints, we use only 200 permutations per data set when calculating the null distribution of $\hat{S}$. Increasing the number of permutations would increase the accuracy of the reported p-values. Finally, in Section 6.1 we specify a probability model for W. It is possible that this model is not the best representation of this process and therefore if a different model were used, it may make the test more powerful.

	Null Hypothesis Rejected				
Correlation	Linear Only	Nonparametric Only	Both	Neither	Total
0.2	1781	1923	387	35409	39500
0.4	2236	2638	517	34109	39500
0.6	3046	3400	1438	31616	39500
0.8	3344	5447	3799	26910	39500

Table 6.3: The total number (across all t_j) of times the null hypotheses are rejected for a significance level of 0.05 at each correlation level. The linear and nonparametric columns correspond to test (6.6) and (6.4) respectively.

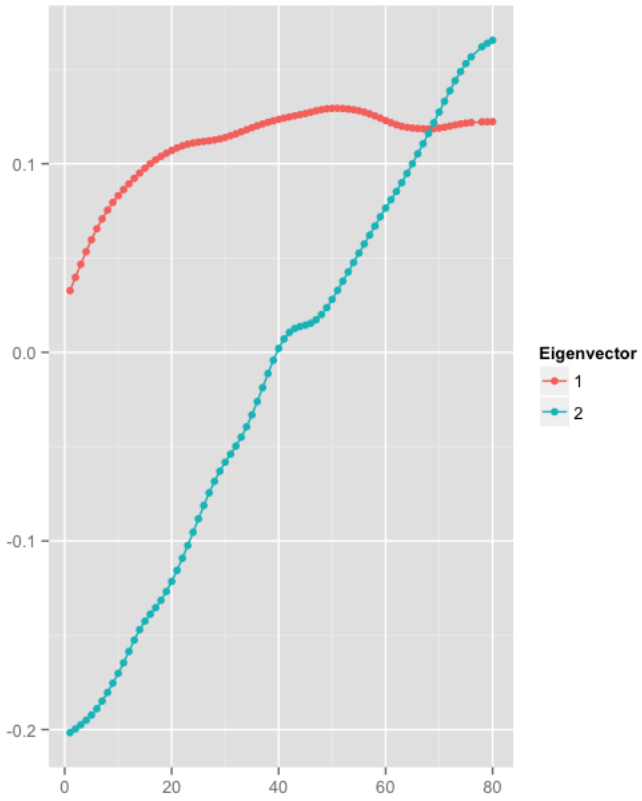


Figure 6.1: The first eigenvector (red), and second eigenvector (blue) from the principal component analysis of the data vectors $(x_{i1}, \ldots, x_{iJ})^t$ for $i = 1, \ldots, n$. These are used as φ_1 and φ_2 in the simulation study.

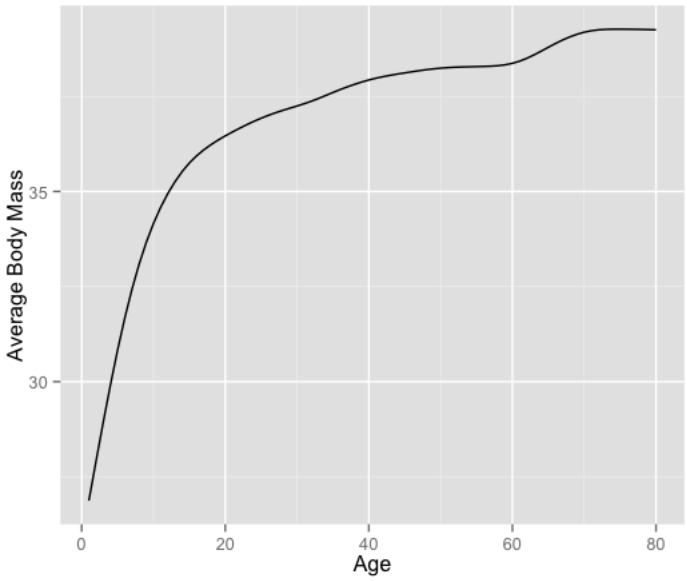


Figure 6.2: The mean body mass (grams) of the male active control mouse group, used as $\mu_X(t)$ in the simulation study.

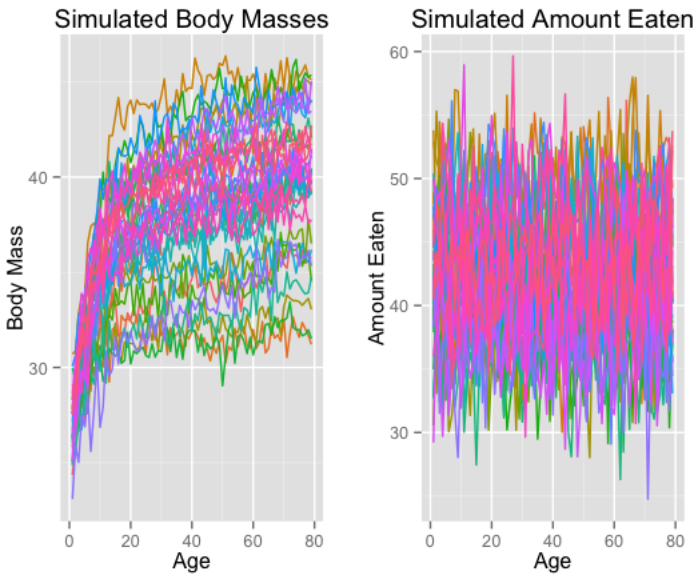


Figure 6.3: A simulated data set of body mass (grams) is given in the left pane, while a simulated data set of the amount eaten (grams) is given in the right. The correlation between the two processes has been set to 0.

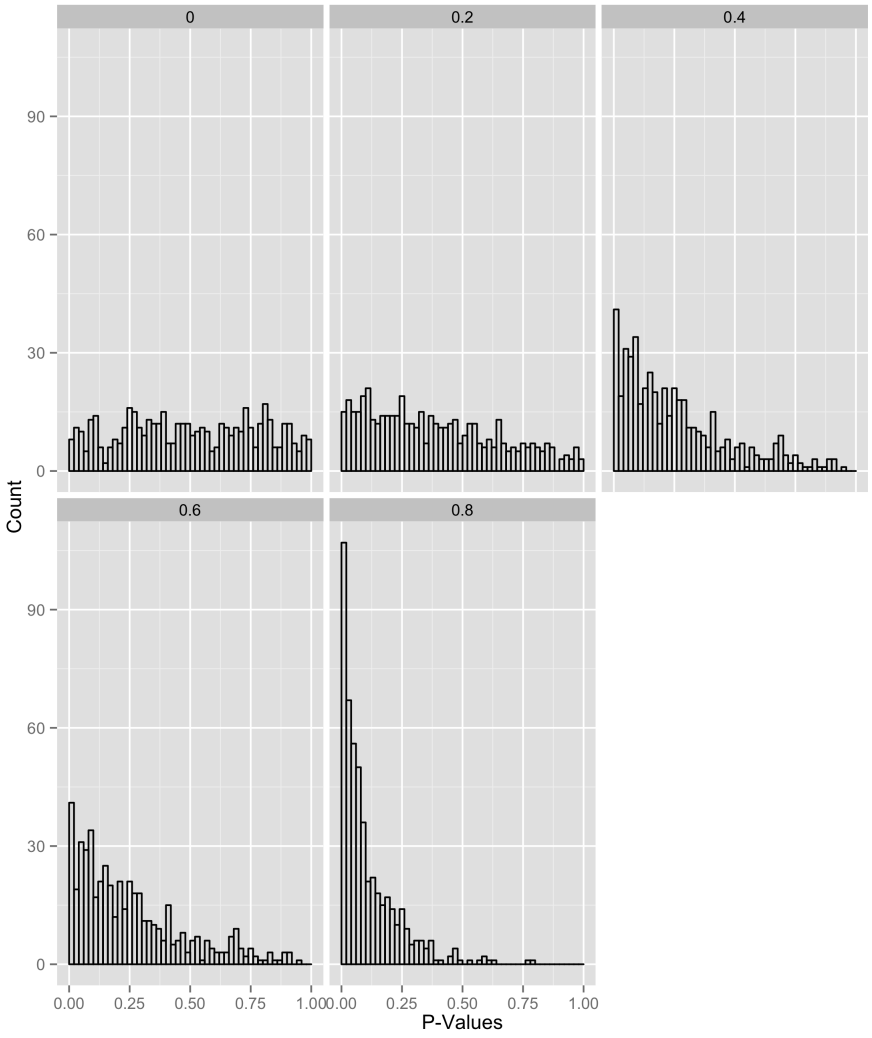


Figure 6.4: Histograms of the 500 p-values resulting from the nonparametric hypothesis test from Section 4.5. The correlation between α_1 and a is set to 0, 0.2, 0.4, 0.6 and 0.8.

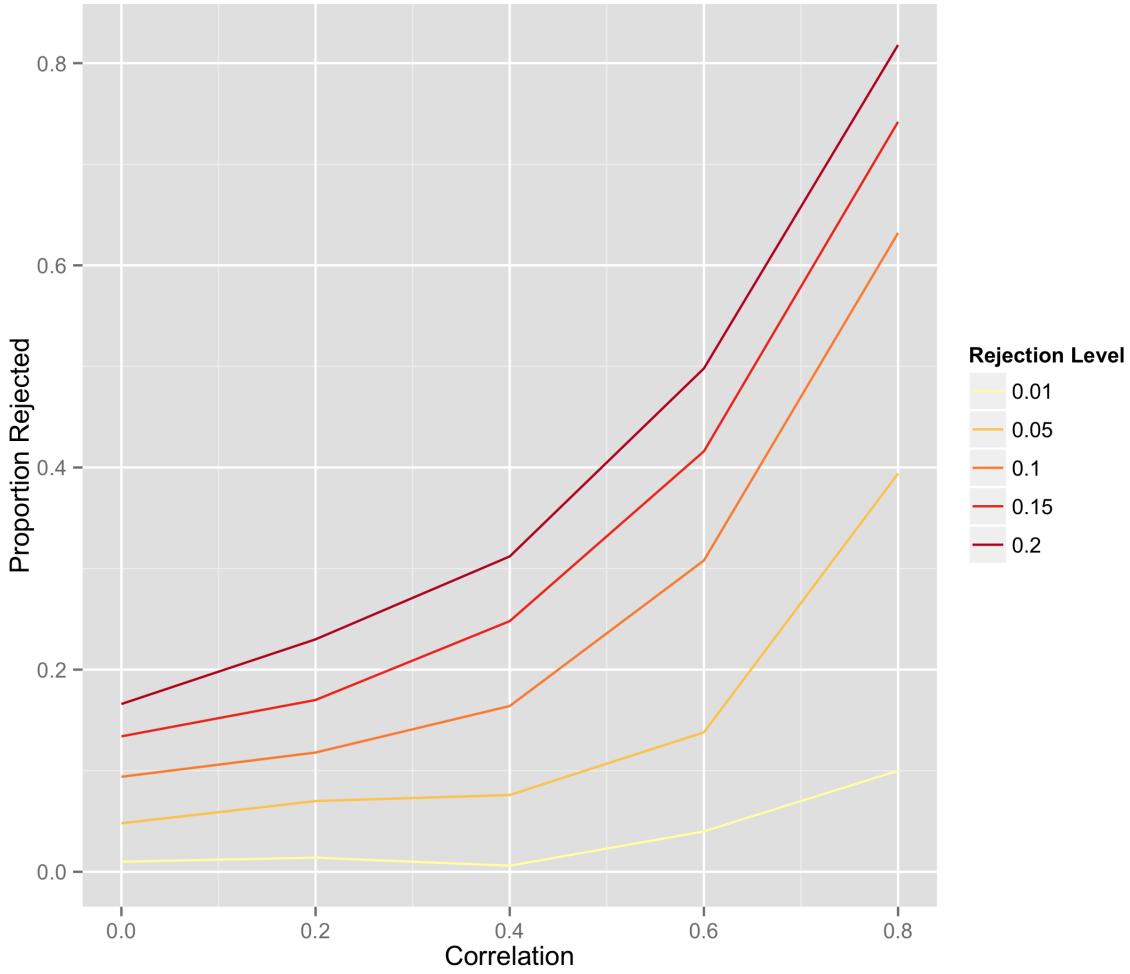


Figure 6.5: The power of the test statistic, $\hat{S}$, as a function of the correlation between α_1 and a. Each power curve corresponds to a fixed rejection level for the test.

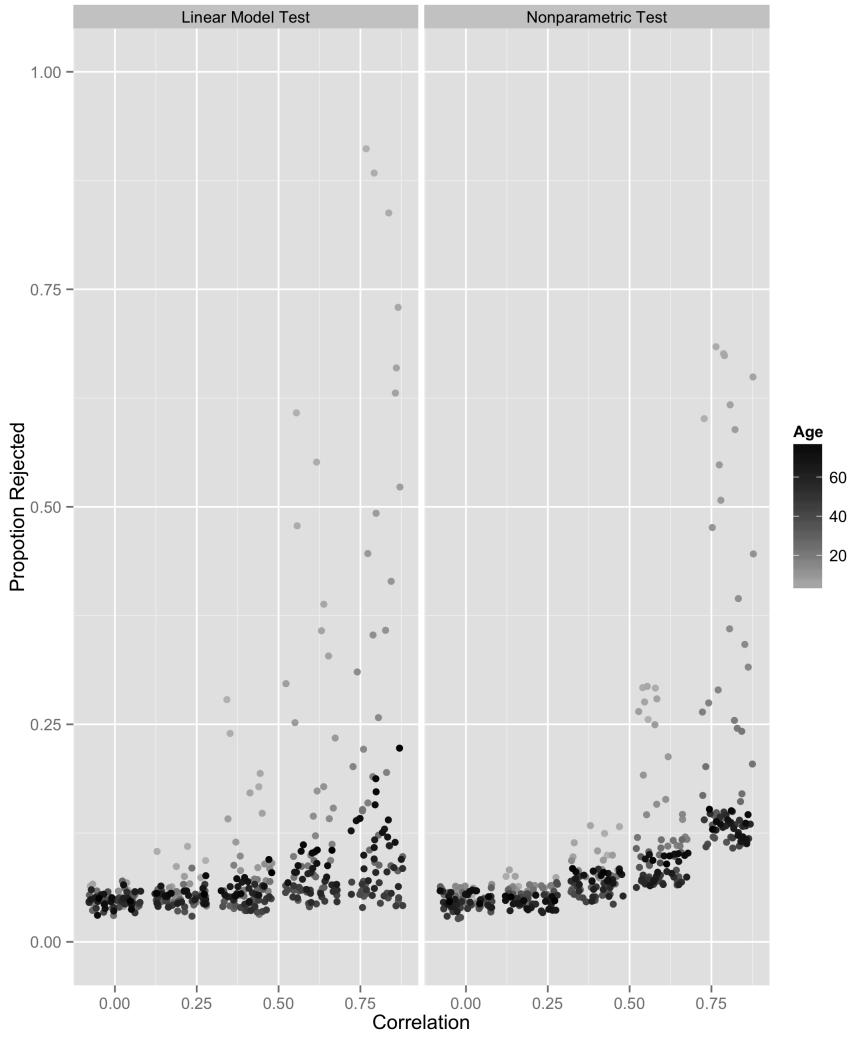


Figure 6.6: The point-wise power of the hypothesis test in (6.6) using a standard linear model (left) compared to the point-wise power of the test in (6.4) using our $\hat{r}(t)^2$ statistic, at a significance level of 5%. The x-axis indicates each of the five fixed correlations between a and α_1. The darker the color of a point, the greater the value of t_j.

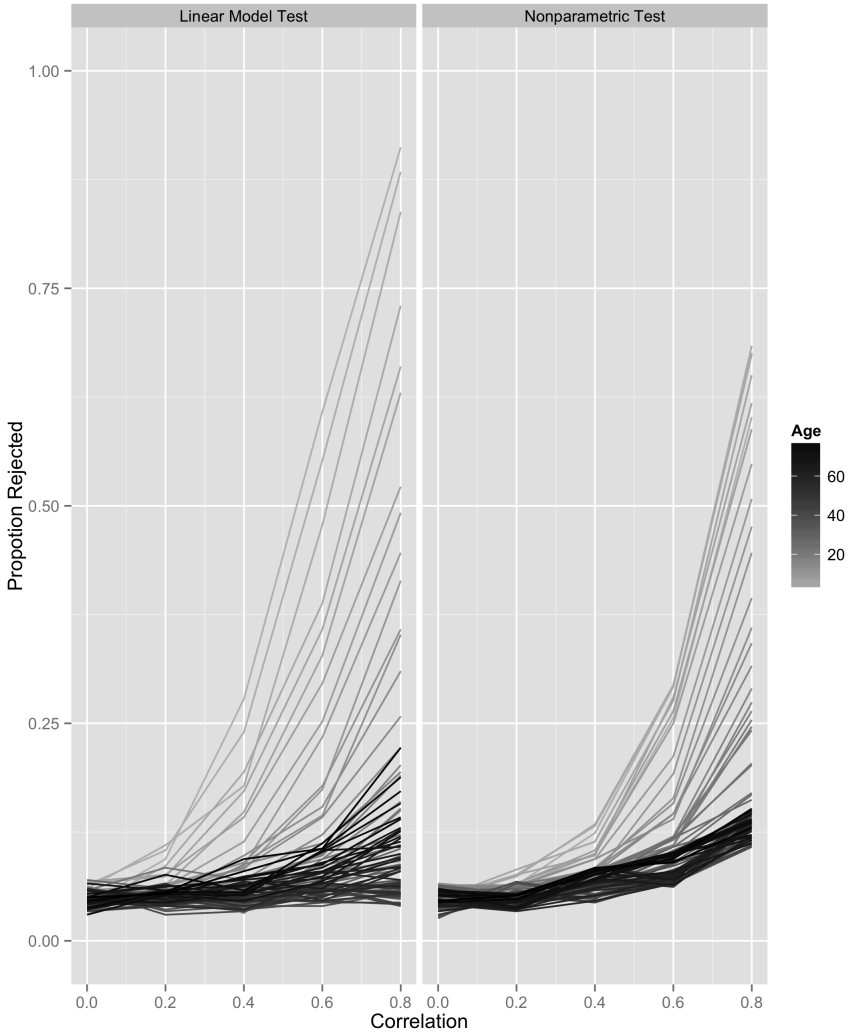


Figure 6.7: Point-wise power curves of the hypothesis test in (6.6) (left) using a standard linear model, and of the test in (6.4), using our $\hat{r}(t)^2$ statistic (right) as functions of the correlation between a and α_1. The significance level is 5% and the darker the color of a line, the greater the value of t_j.

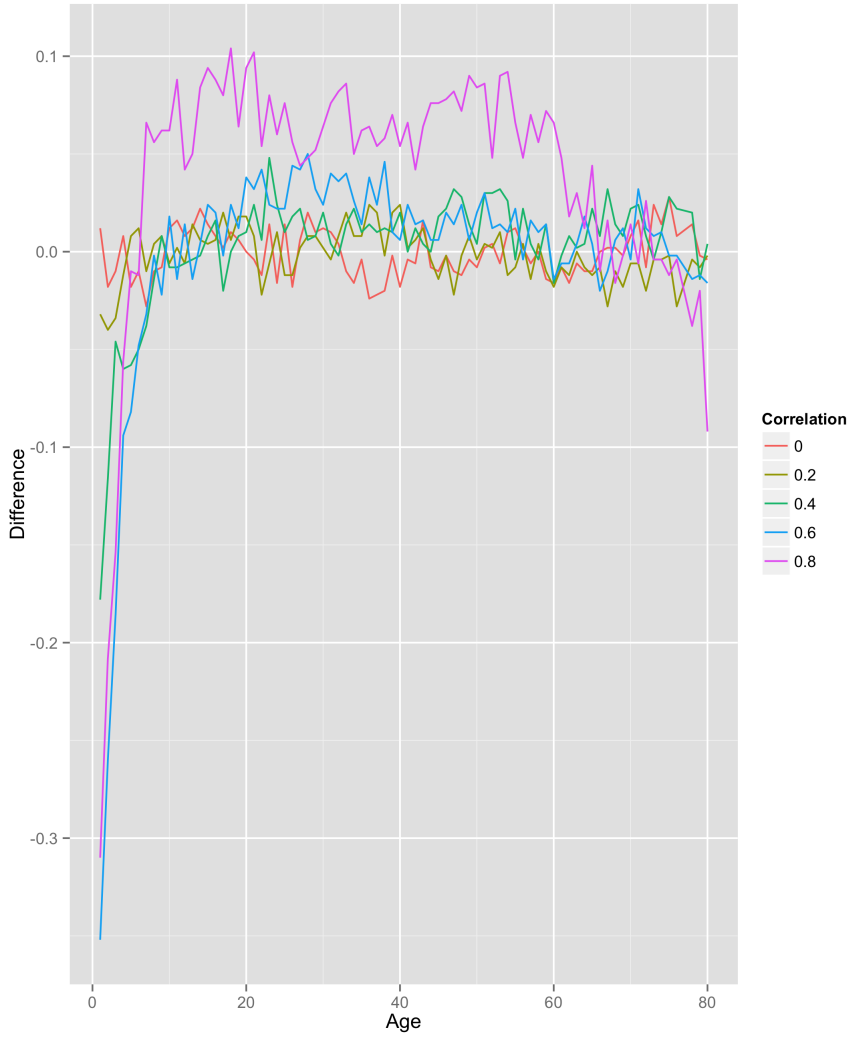


Figure 6.8: The proportion of times the null hypothesis in (6.4) is rejected minus the proportion of times the null hypothesis in (6.6) is rejected, as a function of t. The different colors represent the different correlation levels between a and α_1.

Chapter 7

Conclusion

In this thesis, we have proposed extensions to existing techniques for studying the instantaneous relationship between a process and its derivative. In many applications, this instantaneous relationship may be significantly influenced by an additional, related stochastic processes. To include an additional process in estimating the relationship between $X(t)$ and $X'(t)$, we use a two step smoothing procedure in the mold of Verzelen et al. (2012) as follows (Section 4.4). First, we smooth the data to obtain estimated trajectories of X and of its derivative, X'. Secondly, we fit a nonparametric regression model via bivariate kernel smoothing, where for each t, $X'(t)$ is regressed on $X(t)$ and $W(t)$.

Furthermore, we have developed a test statistic, $\hat{S}$, to determine whether the addition of $W(t)$ in the nonparametric regression provides a significantly better fit to using just $X(t)$ alone. Under the null hypothesis that $W(t)$ does not provide a significantly improved fit, we approximate the distribution of $\hat{S}$ through a permutation method (Section 4.5). We further use this permutation approach to attempt to determine at which specific time points or intervals the inclusion of $W(t)$ improves the fit significantly.

These techniques are applied to the data set described in Chapter 2, comprised of mouse growth data for eight distinct groups. These groups are characterized by gender (male/female), breeding design (selected/control) and access to an exercise wheel (sedentary/active). We carry out the two-step smoothing procedure with the estimated growth rate regressed on body mass only, and then on body mass and the previous weeks' amount eaten. (Chapter 5). These models are fit for each of the eight groups and we note a variety of observations and comparisons.

Paramount to these comparisons is that, based on our testing methods, only the two sedentary female groups had a relationship between growth rates at age t and body mass at age t that was explained significantly better by including the amount eaten in week t. In future work, it would be interesting to carry out similar analyses while also taking into account the genetic dependence that results from the eight genetic lines of mice, as described in Chapter 2.

To improve our understanding of the statistical properties of our testing approach, a simulation study based on the mouse growth data was carried out in Chapter 6. The point-power of our method was compared to that of the standard test for coefficient significance in a linear model. Surprisingly, our method resulted in greater power, except when testing at smaller values of t.

As final remarks, it is important to discuss the limitations and possible continuations of this work. Firstly, our simulations and data analysis make clear that $\hat{S}$ works well for determining whether or not $W(t)$ is significant overall in explaining $X'(t)$, but that determining what values of t contribute to a significant value of $\hat{S}$ is challenging. Trying to refine the method to more clearly determine these values of t is a natural first step in any subsequent research. Further, although our testing method outlined in Section 4.5 is flexible, the permutation method only approximates a subset of the $\hat{S}$'s null distribution. A further understanding of this underlying distribution and better ability to approximate it may result in a more powerful test. Moreover, we have not developed any asymptotic properties for the components of our test statistic. Certainly, properties analogous to those in Verzelen et al. (2012) would be desirable and provide an exciting opportunity for future work. Another possible extension would be to explore the effects of additional processes on the relationship between $X(t)$ and $X'(t)$. Our current approach is restricted to the addition of just one process but could be extended to include more.

Bibliography

R. B. Ash and M. F. Gardner. *Topics in Stochastic Processes*. Academic Press New York, 1975.

A. W. Bowman and A. Azzalini. *R package* `sm`*: nonparametric smoothing methods (version 2.2-5.4)*. University of Glasgow, UK and Università di Padova, Italia, 2014. URL `URLhttp://www.stats.gla.ac.uk/~adrian/sm,http://azzalini.stat.unipd.it/Book_sm`.

J. Cao and J. O. Ramsay. Parameter cascades and profiling in functional data analysis. *Computational Statistics*, 22:335–351, 2007.

J. Cao, J. Huang, and H. Wu. Penalized nonlinear least squares estimation of time-varying parameters in ordinary differential equation. *Journal of Computational and Graphical Statistics*, 21:42–56, 2012.

S. P. Ellner, Y. Seifu, and R. H. Smith. Fitting population dynamics models to time series data by gradient matching. *Ecology*, 83:22562270, 2002.

T. Gasser, H. G. Müller, W. Kohler, L. Molinari, and A. Prader. Nonparametric regression analysis of growth curves. *The Annals of Statistics*, 12 (1):210–229, 1984.

P. Hall, J.S. Marron, and B. Park. Smoothed cross validation. *Probability Theory and Related Fields*, 92(1):1–20, 1992.

T. J. Hastie and R. J. Tibshirani. *Generalized Additive Models*. London: Chapman & Hall, 1990. ISBN 0412343908.

T. J. Hastie, R. J. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer New York Inc., New York, NY, USA, 2001.

N. Heckman. Comments on: Dynamic relations for sparsely sampled Gaussian processes. *Test*, 19(1):46–49, 2010.

N. Heckman and J. O. Ramsay. Penalized regression with model-based penalties. *The Canadian Journal of Statistics*, 28:241–258, 2000.

M. Kirkpatrick and N. Heckman. A quantitative genetric model for growth, shape, reaction norms, and other infinite dimensional characters. *Journal of Mathematical Biology*, 27(4):429–450, 1989.

P. Koteja, J. G. Swallow, P. A. Carter, and T. Garland. Energy cost of wheel running in house mice: Implications for coadaptation of locomotion and energy budgets. *Physiological and Biochemical Zoology*, 72(2):238–249, 1999.

P. Koteja, J.G. Swallow, P.A. Carter, and T. Garland. Maximum cold-induced food consumption in mice selected for high locomotor activity: Implications for the evolution of endotherm energy budgets. *Journal of Experimental Biology*, 204(6):1177–1190, 2001.

P. A. Koteja, P.and Carter, J. G. Swallow, and T. Garland. Food wasting by house mice: Variation among individuals, families, and genetic lines. *Physiology and Behavior*, 80(2):375–383, 2003.

H. Liang and H. Wu. Parameter estimation for differential equation models using a framework of measurement error in regression models. *Journal of the American Statistical Association*, 103(484):1570–1583, 2008.

B. Liu and H. G. Müller. Estimating derivatives for samples of sparsely observed functions, with application to online auction dynamics. *Journal of the American Statistical Association*, 104:704–717, 2009.

H. Miao, C. Dykes, L. M. Demeter, and H. Wu. Differential equation modeling of HIV viral fitness experiments: model identification, model selection, and multimodel inference. *Biometrics*, 65(1):292–300, 2009.

H. G Müller and W. J. Yang. Dynamic relations for sparsely sampled Gaussian processes. *Test*, 19:1–29, 2010.

H. G. Müller and F. Yao. Empirical dynamics for longitudinal data. *The Annals of Statistics*, 38:3458–3486, 2010.

E. A. Nadaraya. On estimating regression. *Theory of Probability and its Applications*, 9:141–142, 1964.

R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL `https://www.R-project.org/`.

J. O. Ramsay and C. J. Dalzell. Some tools for functional data analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(3): 539–572, 1991.

J. O. Ramsay and B. W. Silverman. *Applied Functional Data Analysis*. SpringerLink ebook, New York City, NY, 2002.

J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. SpringerLink ebook, New York City, NY, 2005.

J. O. Ramsay, N. Heckman, and B. W. Silverman. Spline smoothing with model-based penalties. *Behavior Research Methods, Instruments, & Computers*, 29(1):99–106, 1997.

J. O. Ramsay, G. Hooker, D. Campbell, and J. Cao. Parameter estimation for differential equations: a generalized smoothing approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(5): 741–796, 2007.

S. K. Reddy and M. Dass. Modeling online art auction dynamics using functional data analysis. *Statistical Science*, 21(2):179–193, 2006.

J. A. Rice and B. W. Silverman. Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 233–243, 1991.

Y. Sakamoto, M. Ishiguro, and G. Kitagawa. *Akaike Information Criterion Statistics*. Tokyo : KTK Scientific Publishers, 1986.

J. G. Swallow, P. A. Carter, and T. Garland. Artificial selection for increased wheel-running behavior in house mice. *Behavior Genetics*, 28(3):227–237, 1998.

Patrick A. Carter Theodore J. Morgan, Theodore Garland. Ontogenies in mice selected for high voluntary wheel-running activity. i. mean ontogenies. *Evolution*, 57(3):646–657, 2003.

N. Verzelen, W. Tao, and H. G. Müller. Inferring stochastic dynamics from functional data. *Biometrika*, 99:533–550, 2012.

M. P. Wand. *KernSmooth: Functions for Kernel Smoothing*, 2015. URL `http://CRAN.R-project.org/package=KernSmooth`. R package version 2.23-15.

M.P. Wand and M. C. Jones. *Kernel Smoothing*. Monographs on statistics and applied probability. Chapman & Hall/CRC, Boca Raton (Fla.), London, New York, 1995.

G. S. Watson. Smooth regression analysis. *Sankhyā Ser.*, 26:359–372, 1964.

H. Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009. ISBN 978-0-387-98140-6. URL `http://had.co.nz/ggplot2/book`.

F. Yao, H. G. Müller, and J. L. Wang. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100 (470):577–590, 2005.

Appendix A

Calculation of Conditional Expectation for Simulations

Let $X(\cdot), W(\cdot)$ be Gaussian processes with mean and covariance functions given by $\mu_X(\cdot), C_X(\cdot,\cdot)$ and $\mu_W(\cdot), C_W(\cdot,\cdot)$, respectively.

For a fixed t, $X(t), W(t)$ and $X'(t)$ are jointly normal, $(X(t), W(t), X'(t))^T \sim \mathcal{N}(\boldsymbol{\mu}(t), \boldsymbol{\Sigma}(t))$, where $\boldsymbol{\mu}(t) = (\mu_X(t), \mu_W(t), \mu_{X'}(t))^T$ and $\boldsymbol{\Sigma}(t)$ is the covariance matrix of $X(t), W(t)$ and $X'(t)$. We will first state a useful result for computing the conditional expectation of $X'(t)$ given $X(t), W(t)$.

Proposition 1. *Let $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_n)^T$ and $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_m)^T$ be two multivariate normal distributions of size n and m, respectively, such that $\boldsymbol{Z}$ and $\boldsymbol{Y}$ are jointly normal with covariance matrix partitioned as*

$$\boldsymbol{\Sigma} = Cov\left[\begin{pmatrix}\boldsymbol{Z}\\ \boldsymbol{Y}\end{pmatrix}; \begin{pmatrix}\boldsymbol{Z}\\ \boldsymbol{Y}\end{pmatrix}\right] = \begin{pmatrix}\boldsymbol{\Sigma_{ZZ}} & \boldsymbol{\Sigma_{ZY}}\\ \boldsymbol{\Sigma_{ZY}}^T & \boldsymbol{\Sigma_{YY}}\end{pmatrix}.$$

Then the conditional expectation of $\boldsymbol{Y}$ on $\boldsymbol{Z}$ is given by

$$E[\boldsymbol{Y}|\boldsymbol{Z}] = \mu_{\boldsymbol{Y}} + \boldsymbol{\Sigma_{ZY}}^T \boldsymbol{\Sigma_{ZZ}}^{-1}(\boldsymbol{Z} - \mu_{\boldsymbol{Z}}),$$

where $\mu_{\boldsymbol{Z}}$ and $\mu_{\boldsymbol{Y}}$ are the means of Z and Y, respectively.

Proof. Appendix A in Statistics for High-Dimensional Data. Methods, Theory and Applications, P. Buhlmann and S. van de Geer, Springer 2011." □

For a fixed t, we can now apply this result with $\boldsymbol{Z} = (X(t), W(t))$ and $\boldsymbol{Y} = X'(t)$, to find the conditional expectation of $X'(t)$ on $X(t)$ and $W(t)$. After some simple matrix multiplication we obtain:

$$E[X'(t)|X(t), W(t)] = \mu_{X'}(t) +$$
$$\frac{1}{\mathcal{D}(t)}\{\mathrm{Var}(W(t))\mathrm{Cov}(X(t), X'(t)) - \mathrm{Cov}(W(t), X'(t))\mathrm{Cov}(X(t), W(t))\}\{X(t) - \mu_X(t)\} +$$
$$\frac{1}{\mathcal{D}(t)}\{\mathrm{Var}(X(t))\mathrm{Cov}(W(t), X'(t)) - \mathrm{Cov}(X(t), X'(t))\mathrm{Cov}(X(t), W(t))\}\{W(t) - \mu_W(t)\},$$

where

$$\mathcal{D}(t) = \det \begin{bmatrix} \mathrm{Var}(X(t)) & \mathrm{Cov}(X(t), W(t)) \\ \mathrm{Cov}(X(t), W(t)) & \mathrm{Var}(W(t)) \end{bmatrix}.$$

We can calculate

$$\begin{aligned} C(W(s), X'(t)) &= \frac{\partial}{\partial t} C(W(s), X(t)), \\ C(X(s), X'(t)) &= \frac{\partial}{\partial t} C(X(s), X(t)) \text{ and} \\ C_{X'}(s,t) &= \frac{\partial^2 C_X}{\partial s \partial t}. \end{aligned}$$

If we want $E[X'(t)|X(t), W(t)]$ to have no dependence on $W(t)$ then we require $\mathrm{Var}(X(t))\mathrm{Cov}(W(t), X'(t)) - \mathrm{Cov}(X(t), X'(t))\mathrm{Cov}(X(t), W(t)) = 0$.

Appendix B

Calculation of Covariance Structure for Simulations

We wish to simulate two correlated stochastic processes, X and W. In the following, we assume that we can write X as

$$X(t) = \sum_{j=1}^{J} \alpha_j \varphi_j(t), \tag{B.1}$$

where the α_j's are uncorrelated, mean 0 random variables with the variance of α_j equal to λ_j, and W as

$$W(t) = a + \sigma\epsilon(t), \tag{B.2}$$

where a is a mean zero random variable with variance σ_a^2, uncorrelated with $\epsilon(t)$ for all t, and $\mathrm{E}(\epsilon(t)) = 0$, $\mathrm{Cov}(\epsilon(s), \epsilon(t)) = \mathcal{I}(s = t)$, the indicator of s equal to t. With these assumed decompositions and some additional stated assumptions, we will define the proper covariance structure for the processes observed at a fixed set of time points.

For a fixed set of time points, $t_1, \ldots, t_k$, we will generate observations of X and W at each t_i. Let $\mathbf{t} = (t_1, \ldots, t_k)^t$. Then we can write

$$X(\mathbf{t}) = (X(t_1), \ldots, X(t_k))^t = \mathbf{\Phi}\boldsymbol{\alpha} \tag{B.3}$$

where $\mathbf{\Phi} = (\varphi_1(\mathbf{t}), \ldots, \varphi_J(\mathbf{t}))$ is a $k \times J$ matrix and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_J)^t$. Similarly, we can write

$$W(\mathbf{t}) = (W(t_1), \ldots, W(t_k))^t = a\mathbf{1}_k + \sigma\boldsymbol{\epsilon}, \tag{B.4}$$

where $\boldsymbol{\epsilon} = (\epsilon(t_1), \ldots, \epsilon(t_k))^t$. Therefore, to simulate $X(\mathbf{t})$ and $W(\mathbf{t})$ we will generate values of $\boldsymbol{\alpha}, a,$ and $\boldsymbol{\epsilon}$. Having a proper covariance structure for $X(\boldsymbol{t})$ and $W(\boldsymbol{t})$ is equivalent to the $J + 1 + k$ dimensional covariance matrix of $\boldsymbol{\alpha}, a,$ and $\boldsymbol{\epsilon}$ being positive definite. We further assume that

$$\mathrm{Cov}(\boldsymbol{\alpha}, \boldsymbol{\epsilon}) = 0$$

and

$$\mathrm{Cov}(a, \boldsymbol{\epsilon}) = 0.$$

It follows from these assumptions that to have the covariance matrix of $\boldsymbol{\alpha}, a,$ and $\boldsymbol{\epsilon}$ positive definite, we need only have the covariance matrix of $\boldsymbol{\alpha}$ and a positive definite; that is, for an arbitrary $\boldsymbol{v} \in \mathbb{R}^J$, and $z \in \mathbb{R}$ such that $(\boldsymbol{v}, z) \neq \mathbf{0}$, we must have

$$(\boldsymbol{v}^t, z)\Sigma(\boldsymbol{v}^t, z)^t > 0, \tag{B.5}$$

where Σ is the covariance matrix of $(\boldsymbol{\alpha}^t, a)$. A simple calculation yields the following equivalent condition for positive definiteness:

$$\sum_{j=1}^{J} \lambda_j v_j^2 + 2z \sum_{j=1}^{J} v_j c_j + \sigma_a^2 z^2 > 0, \tag{B.6}$$

where $c_j = \mathrm{Cov}(\alpha_j, a)$. A sufficient condition for (B.6) is to have

$$\frac{c_j^2}{\lambda_j \sigma_a^2} < \frac{1}{J} \text{ or equivalently } \mathrm{Corr}^2(\alpha_j, a) < \frac{1}{J}. \tag{B.7}$$

Given values of λ_j and σ_a^2, one can thus set the c_j's according to the desired strength of covariance between X and W. If a subset of size I of the c_j's is set to zero, then condition (B.7) can be replaced by

$$\mathrm{Corr}^2(\alpha_j, a) < \frac{1}{J - I}, \tag{B.8}$$

for each j such that $c_j \neq 0$.