Modelling Obesity in Scotland

Gary Napier

A Dissertation Submitted to the
University of Glasgow
for the degree of
Master of Science

School of Mathematics & Statistics

November 2010

Abstract

Overweight and obesity prevalence has been on the increase in Scotland since the 1990s when monitoring began with the Scottish Health Surveys. In this study models are developed that describe the prevalence of obesity in Scotland by analysing data from the Scottish Health Surveys. In particular, we study how the Body Mass Index (BMI) and waist-to-hip ratio (WHR) vary with gender, age and socio-economic status, and investigate whether there is a shift in the entire BMI/WHR distributions or simply a stretching-out of the upper tail (roughly corresponding to the overweight/obese categories).

Logistic regression is employed for modelling obesity prevalence, and generalised additive models are employed to examine the relationship between BMI/WHR and gender, age, socio-economic status and survey year. The odds of being overweight or obese increase with age, which is also the case for log(BMI) and WHR, with differing gender patterns. The rate of increase in BMI and WHR is at its greatest for individuals aged between 16 and 30, and gradually slows down before decreasing for males over the age of 55, but remains increasing for females of the same age. No significant difference in obesity prevalence is observed for males in social classes iii manual and iv & v, but males in social class iii non-manual are 1.27 times more likely to be obese in comparison to males in social classes i & ii. For females, the odds of being obese increase with each consecutive social class.

Quantile regression is used to study how the entire conditional distri-

butions of BMI and WHR vary with gender, age, socio-economic status and survey year. By specifying changes in the quantiles of the response (BMI/WHR), quantile regression highlights an uneven increase in the BMI and WHR over time; in each subsequent survey all quantiles shift to the right, but this increase is larger for the upper tail of the distribution. The effects of socio-economic status also vary across the quantiles of the BMI and WHR distributions, with males in each subsequent social class who lie at the lower end of the distribution having lower BMI values than males in social classes i & ii, but higher BMI values at the upper end of the distribution. Finally, subtle gender differences are observed in the relationship between BMI/WHR and age. In conclusion, quantile regression allows us to go beyond obesity prevalence and examine finer aspects of the BMI and WHR distributions.

Acknowledgements

First and foremost, I would like to thank my supervisors Dr Tereza Neocleous and Professor Stephen Senn for their help and support throughout the project. I acknowledge the Economic and Social Data Service (ESDS) and the UK Data Archive for providing the Scottish Health Survey data, and also, the School of Mathematics and Statistics for funding my research.

My thanks also go to family and friends for their support.

Declaration

I have prepared this thesis myself; no section of it has been submitted previously as part of any application for a degree. I carried out the work reported in it, except where otherwise stated.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation and Background

The increasing prevalence of obesity, a medical condition in which excess accumulation of body fat can have an adverse effect on one's health, has become a major concern to health organizations worldwide due to its associations with physical, psychological, public health and economic consequences. A recent study by Wang et al. (2008) that was carried out in the US yielded alarming results with the projected future prevalence of overweight or obese adults predicted to be a staggering 86%, and 51% respectively, by 2030. All US adults were predicted to be overweight or obese by as early as 2048. Such alarming figures have called into question the lifestyles of US adults with more importance needing to be placed on ways in which to get the population living more healthy and active lifestyles. According to the same study, the consequences of the ever-increasing prevalence of obesity would also lead to a huge impact on the US economy, with health care costs attributable to overweight and obese individuals estimated to double each decade, resulting in a predicted 960 billion US dollars by 2030, which would lead to a relatively significant proportion (16-18%) of total US health care costs being accountable to obesity alone. Direct health care costs related to obesity can include diagnostic, preventive and treatment measures,

while indirect costs can be attributed to morbidity and mortality. Increasing prevalence of obesity can also lead to increases in other health issues, such as heart disease, hypertension, diabetes, cancer, and many other health-related problems, which are discussed in Stein & Colditz (2004).

Other recent studies, which were carried out closer to home, by Zaninotto et al. (2009) and Mills (2008), analysed the trends in obesity among adults in England, and also made projections of the prevalence of obesity in future years. Zaninotto et al. (2009) estimate that since the turn of the century, obesity has been directly responsible for more than 9000 premature deaths, in England alone. The results of the study were similar, though not as pronounced, as those found in the study by Wang et al. (2008), with the prevalence of obesity significantly increasing from 14% to 24% and 17% to 24% from 1993 to 2004 among men and women, respectively. The projected prevalence of obesity in 2012 was estimated by Zaninotto et al. (2009) at 32% in men and 31% in women, if recent obesity trends continue. The study also took into consideration the age and social class of the study population with some interesting results. The prevalence of obesity was found to be significantly higher for both males and females within the manual (skilled/unskilled occupations requiring physical exertion) social class in comparison with those in the non-manual (professional/skilled occupations) social class, with the projected prevalence of obesity in 2012 for adults in manual social classes being 43%, and 35% for adults in non-manual social classes. A significant increase in the prevalence of obesity was found for both men and women in all age groups from 1993 to 2004, with higher obesity rates for individuals aged between 35-54, and 55-74. Due to the significant difference found in the prevalence of obesity between the manual and non-manual social classes, Zaninotto et al. (2009) believe that in order to reduce obesity a reduction in the social class inequalities would be required as it is the lower social classes that will be affected the most by the increasing prevalence of obesity. A

recent study on the increasing prevalence in China by Shankar (2010) uses quantile regression to show how the relationships between the BMI and several socio-economic and health-related variables can differ across the whole BMI distribution. For example, Shankar (2010) found that the upper tail of the BMI distribution showed a significant increase in BMI with energy intake, but the lower and middle portions of the distribution were relatively unaffected. In contrast, the relationship between BMI and smoking is entirely and relatively uniformly negative at the lower and middle parts of the distribution, while there is no change in BMI observed at the upper tail of the distribution. Such variation across the BMI distribution has led Shankar (2010) to believe that quantile regression can be extremely useful and beneficial in determining more efficient health policies in tackling obesity in China. We will study cross-sectional data from the Scottish Health Surveys of 1995, 1998, 2003 and 2008, which will allow us to compare the extent of obesity prevalence in Scotland with other nations and studies, such as those by Zaninotto et al. (2009), Mills (2008) and Shankar (2010).

1.2 Assessing Obesity

1.2.1 Body Mass Index

The majority of studies and statistical analyses of obesity tend to focus on using Body Mass Index (BMI), which was developed by Adolphe Quetelet (1796-1874) (Eknoyan (2008)), as an indicator of whether or not someone is classed as being obese. Anyone can obtain their BMI if they know their own weight in kilograms and height in metres as it is simply given as

$$\text{BMI} = \frac{\text{Weight (kg)}}{\text{Height}^2(\text{m}^2)}$$

BMI is used to estimate an ideal and healthy body weight of an indi-

vidual given their height. It is the most widely used indicator of weight categorisation due to the ease with which it can be attained, especially in large population studies. Table 1.1 shows the different BMI categories for adults. As the BMI is not gender-specific or age-specific for adults, in order to be deemed obese, a person's BMI has to be greater than or equal to 30 kg m^{-2}.

Table 1.1: The International Classification of adults underweight, overweight and obese according to the BMI (WHO expert consultation (2004))

Classification	BMI (kg/m^2)	
	Principal cut-off points	Additional cut-off points
Underweight	< 18.50	< 18.50
Severe thinness	< 16.00	< 16.00
Moderate thinness	$16.00 - 16.99$	$16.00 - 16.99$
Mild thinness	$17.00 - 18.49$	$17.00 - 18.49$
Normal range	$18.50 - 24.99$	$18.50 - 22.9$ $23.00 - 24.99$
Overweight	≥ 25.00	≥ 25.00
Pre-obese	$25.00 - 29.99$	$25.00 - 27.49$ $27.50 - 29.99$
Obese	≥ 30.00	≥ 30.00
Obese class I	$30.00 - 34.99$	$30.00 - 32.49$ $32.50 - 34.99$
Obese class II	$35.00 - 39.99$	$35.00 - 37.49$ $37.50 - 39.99$
Obese class III	≥ 40.00	≥ 40.00

The BMI has obvious advantages as a classifier of obesity, being a noninvasive, time effective, and easy way with which anyone can check if they are of healthy weight, but it also has its fair share of disadvantages.

Prentice & Jebb (2001) highlight many of the problems that can arise when using BMI as an indicator of obesity. One such issue arises with aging, due to an increase in body fat and diminishing muscle mass as we become older. However, during such changes many individuals are able to sustain a stable BMI, which suggests an age-dependent relationship between BMI and body fat.

Prentice & Jebb (2001) also give several examples of different ethnic groups, such as Asian-Indian and African-American, which show differing results in the relationship between body fat percentage and BMI between such ethnic groups. This would support the necessity for different BMI thresholds for different ethnic groups. For example, in order to have the same body fat proportion as Caucasians, the cut-off points for Asians would have to be lowered, with BMI ≥ 23.0 considered overweight for Asians, while the cut-off points would have to be raised for African-Americans.

BMI also runs into problems when dealing with individuals such as professional athletes and bodybuilders. This is due to the development of muscle mass through rigorous training methods, which the Body Mass Index does not take into account. This can lead to athletes with "overweight" and "obese" BMIs despite having low body fat percentages and being in peak physical condition. For example, Prentice & Jebb (2001) show how much leaner American footballers and shot-putters are than their BMIs would suggest.

Due to such flaws and criticisms in using the Body Mass Index as an

indicator of obesity, potential measures were proposed with which to replace BMI as the dominant indicator of choice when analysing obesity. One of these is the waist-to-hip ratio which is studied in greater detail in Chapter 6.

1.3 Scottish Health Survey

The data we have in order to model obesity in Scotland comes from the Scottish Health Surveys (Joint Health Surveys Unit (2009)) of 1995, 1998, 2003 and 2008, with the sample sizes of males and females surveyed with valid BMI measurements for each survey year shown in Table 1.2. The Scottish Health Survey is designed to contribute to the monitoring of health throughout Scotland by providing a clear picture of the health of the Scottish population living in private households. The main aim of the survey is to keep an eye on trends in Scotland's health. The Scottish Health Survey uses a two-stage interview process, with a personal interview from a trained interviewer followed up by a home visit from a nurse. This allows for in-depth and detailed information, such as self-assessed health and disability, health service use, smoking and drinking habits, as well as various other health related data, to be gathered from all individuals chosen to take part in the survey, which makes the Scottish Health Survey an extremely robust and rich source of information on all health related issues in Scotland. From 2008 onwards the survey became a two-stage approach involving a personal interview followed by a nurse visit, which unlike previous survey years, is only available to one sixth of the sample. Also, the survey that began in January 2008 will be continuously running from 2008-2011, with an annual report published for each year of the survey. Currently there is no commitment to continue the Scottish Health Survey after 2011.

Table 1.2: Sample sizes for each Scottish Health Survey.

	Survey Year			
	1995	1998	2003	2008
Males	3118	2986	2275	1774
Females	3740	3637	2816	2221

1.4 Overview of Thesis

In Chapter 2 we examine the data acquired from the Scottish Health Surveys in order to get an understanding of the extent of obesity prevalence in relation to age, socio-economic status and survey year for males and females, and also, inspect the BMI distribution. In Chapter 3 linear logistic regression models are fitted that describe the prevalence of overweight/obesity prevalence in Scotland through analysis of the data obtained from the Scottish Health Surveys. Generalized additive models, which allow for age to be treated as continuous in the aforementioned logistic regression models are implemented in Chapter 4 to examine both overweight/obesity prevalence, and also, average BMI values themselves, instead of concentrating on whether someone is above or below a predetermined cut-off point, e.g. 30 kg m^{-2} for a person to be classified as obese. Chapter 5 introduces quantile regression, which allows us to go beyond obesity prevalence and study the entire BMI distribution. Analysis of the waist-to-hip-ratio, a competing indicator of obesity, is carried out in Chapter 6 with potential differences between the BMI and waist-to-hip ratio explored, before reaching our conclusions in Chapter 7.

Chapter 2

Exploratory Analysis Using BMI

The aim of this chapter is to explore potential relationships between the Body Mass Index (BMI) and age and socio-economic status for both males and females, and also examine any BMI trends over the four years (1995, 1998, 2003 and 2008) of available data from the Scottish Health Survey. In addition we examine possible prevalence increases within each of the different BMI categories, namely the overweight and obese categories.

Figure 2.1 shows the percentage of males and females within each of the four main BMI categories from the Scottish Health Surveys of 1995, 1998, 2003 and 2008. The percentage of obese males increased from 17% to 27% from 1995 to 2008, which is a 59% increase in the prevalence of obesity amongst males. An increase of 56% in the prevalence of obesity amongst females was observed between 1995 and 2008, which is not as high an increase in obesity prevalence as was observed for males. However, a higher percentage of obese females is seen for each survey year, ranging from 18% to 29% from 1995 to 2008. In each survey year a higher percentage of overweight males were observed with the prevalence increasing by 2% (from 42% to 43%) from 1995 to 2008. A 9% increase in overweight prevalence was

observed in females from 1995 to 2008, with the percentage of overweight females increasing from 31% to 35%. At the same time the percentage of males and females within the normal weight category decreased by a factor of 29% and 27% respectively. Little change is observed in the percentage of males and females considered underweight, with the only difference by gender being an increase in the percentage of underweight males from 1995 to 2008, while a decrease is observed for females. The increase from 1995 to 2008 in the percentage of males and females who are overweight or obese could also be due to the age distributions from each survey year being quite different. With each survey fewer young people were surveyed, which may partially explain the observed increase in the proportion of overweight/obese members of the Scottish population, as people tend to gain weight with age.

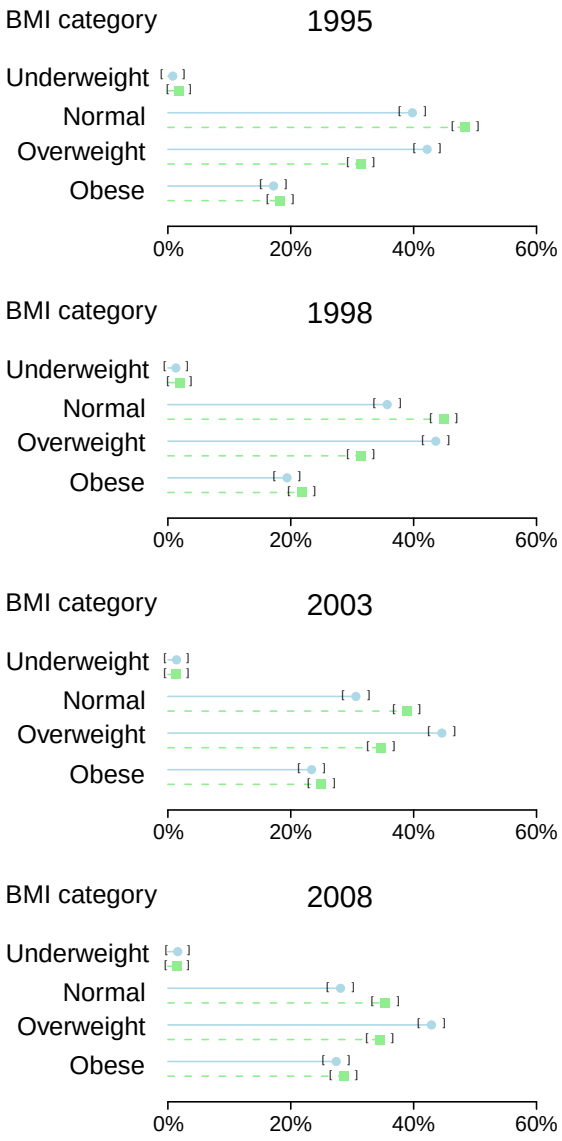

Figure 2.1: Percentage of population surveyed within each BMI category for each year of the Scottish Health Survey. Solid and dashed lines represent males and females, respectively. Parentheses show ± 1 standard error bounds.

The explanatory variable that has been chosen to represent any potential relationships between the prevalence of obesity and socio-economic status amongst males and females from the populations within the Scottish Health Surveys is the social class of the chief income earner, which has been grouped into four categories as follows:

- i & ii (professionals and managerial technical);
- iii non-manual (skilled non-manual);
- iii manual (skilled manual);
- iv & v (semi-skilled and unskilled manual)

The social class of the chief income earner was chosen as the socio-economic variable due to its availability in all years of the Scottish Health Survey, as other socio-economic status variables such as the National Statistics Socio-economic Classification (NS-SEC) were not introduced until 2001 (*The National Statistics Socio-economic Classification* (2010)).

Figure 2.2 shows the percentage of males and females who are considered to be obese within each of the different social class groupings for each survey year. The percentage of obese males in 1995 is very similar in all four social class groups with only 1% (17% vs 18%) difference between social classes i & ii and iv & v. In 1998 the difference is 2% (18% vs 20%) between social classes i & ii and iv & v, with the highest percentage of obese males observed within social classes iii non-manual and iii manual (22% and 20%, respectively). There is further disparity between social classes i & ii and iv & v in 2003, from 20% to 25%, with the highest percentage of obese males in 2003 lying within social class iii manual. A 1% difference (27% vs 28%) is observed in 2008 in the percentage of obese males between social classes i & ii and iii manual, with a smaller percentage observed within social classes iv & v (25% of males within said social class from the 2008 Scottish Health

Survey considered to be obese). The percentage of obese females varies from 14% to 22% in 1995 between social classes i & ii and iv & v. A similar difference in the percentage of obese females is observed from the highest to the lowest social classes in each subsequent survey year. Also, an increase in the percentage of obese females is observed from the Scottish Health Survey of 1995 through to the Scottish Health Survey of 2008 for each social class, with a slight decrease within social classes iv & v between 2003 and 2008.

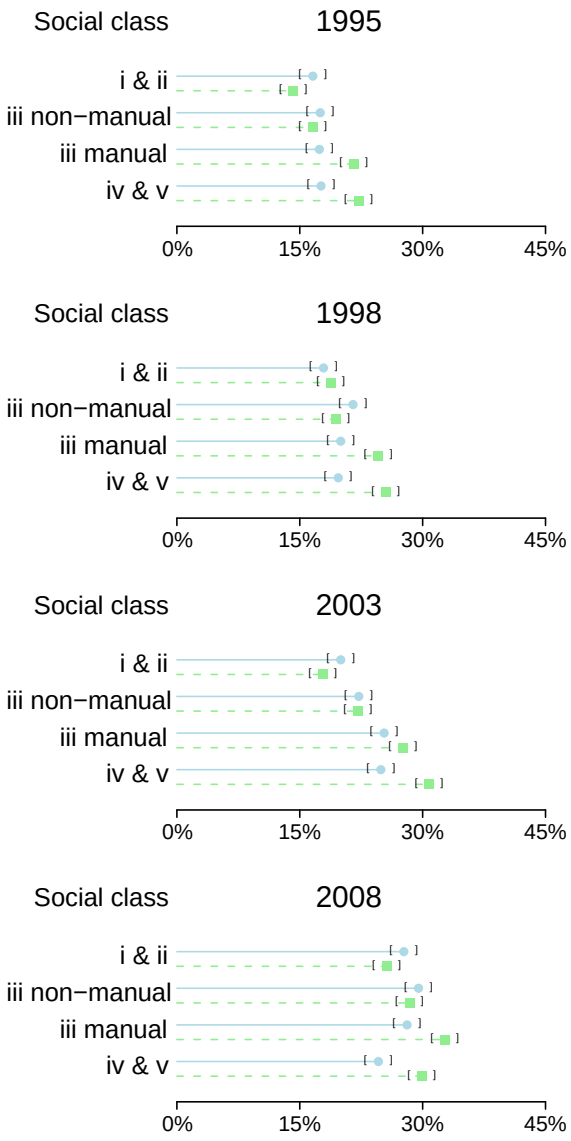

Figure 2.2: Percentage obese (population surveyed) within each social class for each year of the Scottish Health Survey. Solid and dashed lines represent males and females, respectively. Parentheses show ±1 standard error bounds.

Having looked at the percentage of obese individuals within each social class, we will now turn our attention to the relationship between BMI and age.

Figure 2.3 shows violin plots of BMI by age group for each year of the Scottish Health Survey separately for both males and females. Violin plots are similar to boxplots in that they display differences between populations - which in this case are the differences in BMI across each age group - but provide more information on the distribution, with the estimated density of the chosen variable plotted. Skewed distributions result in a "violin" shape. For males we see a steady increase in BMI as we go from the 16-24 to the 55-64 age group with a slight decrease in the average BMI within the 55-64 age group in comparison to the 45-54 age group. An increase in the average BMI of males is seen over the four years of the Scottish Health Survey from 26.3 in 1995 to 27.5 in 2008. For females there is an increase in BMI from the youngest to the oldest age group, and also an increase in mean BMI from 25.9 in 1995 to 27.6 in 2008.

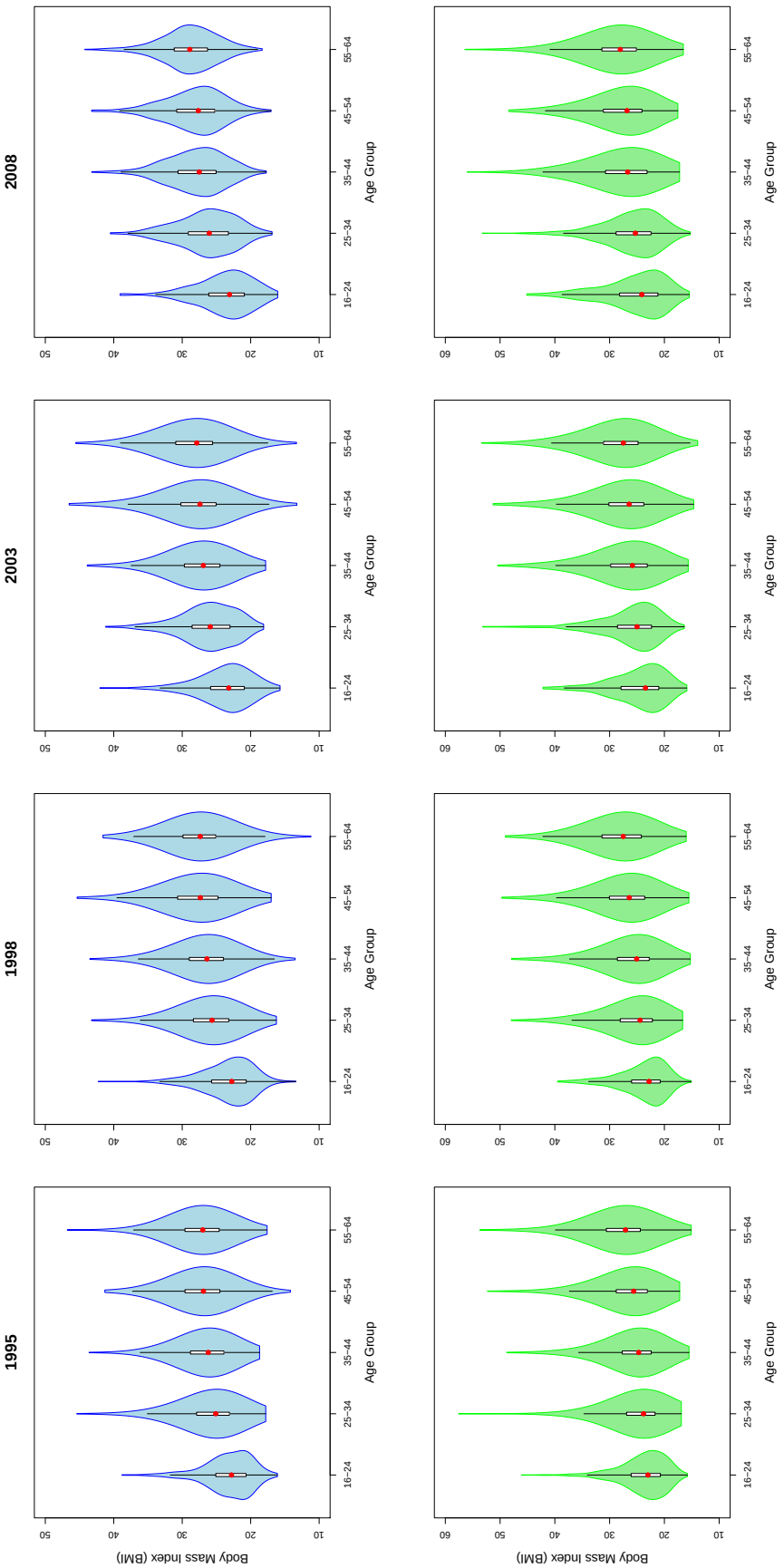

Figure 2.3: BMI by age group for each year of the Scottish Health Survey for males (top row) and females (bottom row).

2.1 BMI Distribution and the Rose Hypothesis

Over the past few decades the 'Population Strategy' proposed by Geoffrey Rose (Rose (1992)) has become key in the progression of epidemiological and preventional strategies in the field of medicine. The strategy requires that for public health problems a reduction in the prevalence of disease is handled by interventions to all individuals within the population, rather than targeting those who are at greater risk, that are found in the upper tail of the distribution. The purpose of such an approach is to shift downwards the entire population distribution of a risk factor, which is shown in plot A of Figure 2.4. However, there are those who are opposed to such an approach when considering the population distribution of the Body Mass Index (BMI).

A recent paper by Penman & Johnson (2006) highlights some of the problems associated with attempting to apply the population strategy to the distribution of BMI. One such problem is in the assumption that the population distribution of BMI is approximately symmetrical, much like the bell-shaped normal distribution. Studies by Penman & Johnson (2006) and Wang et al. (2008) among others have suggested that the population distribution of BMI is positively skewed, with proportionately much more shifting of the distribution at the upper end of the curve than the lower end over time as exemplified in plots B and C of Figure 2.4. Furthermore, if interventions were introduced to the entire population distribution of BMI in order to help reduce the prevalence of overweight and obese individuals in Scotland, that may lead to repercussions and health risks for individuals that lie within the lower tail of the distribution. For example, interventions such as encouraging individuals to consume fewer calories may put those who are already underweight, and possibly those of a perfectly healthy weight at risk. These issues would be far less likely to come to fruition if interventions were focused on

individuals that are at the upper end of the BMI distribution, *i.e.* overweight and obese individuals. In order to investigate whether the strategy proposed by Rose is appropriate for the data acquired from each of the four years of the Scottish Health Survey we examine the population distribution of BMI for males and females, respectively.

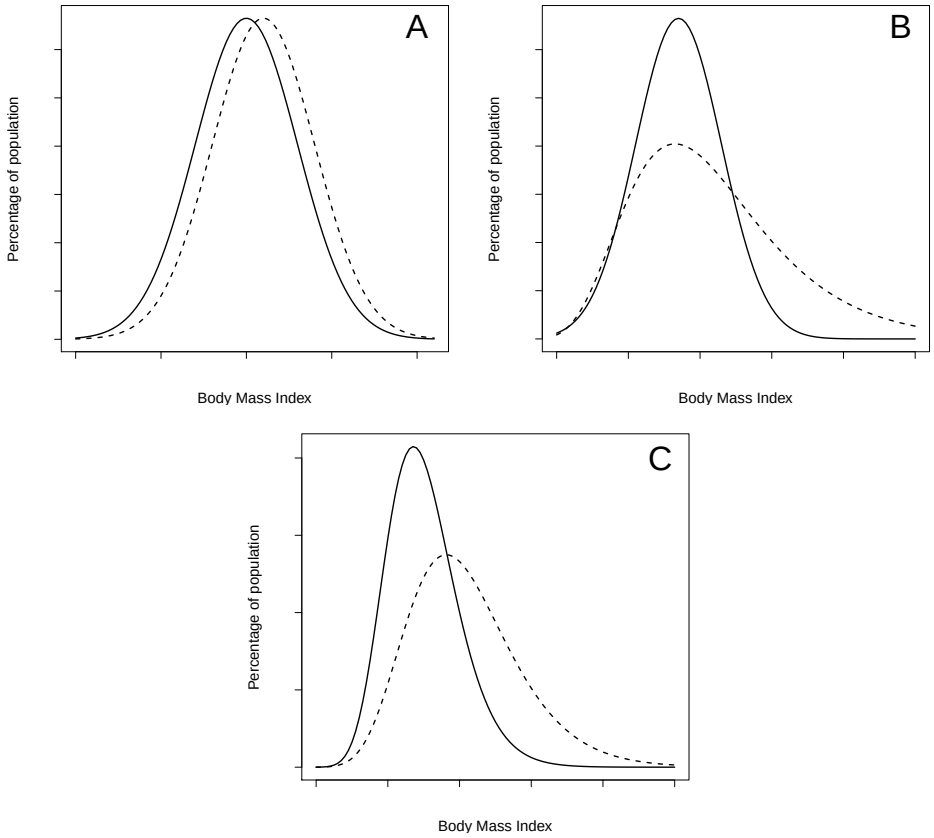

Figure 2.4: Population distribution of body mass index with possible changes over time. Plot A illustrates the 'Population Strategy' proposed by Rose with the dashed curve representing an upward shift in the entire distribution of BMI. Plots B and C show the positively skewed distribution of BMI (dashed curve) with the degree of skewing increasing over time.

Figure 2.5 displays density estimates of the survey population distributions of BMI for each of the Scottish Health Surveys for males and females, respectively. It is apparent that the BMI distributions follow the pattern described by Penman & Johnson (2006), in that the upper tail of the distribution is positively skewed, with skewness increasing from 1995 to 2008, which is particularly apparent in the population distributions of BMI for females. This also suggests that assuming that the BMI follows a normal distribution would be incorrect, and that either a log-normal distribution or non-parametric kernel density estimation would be more appropriate.

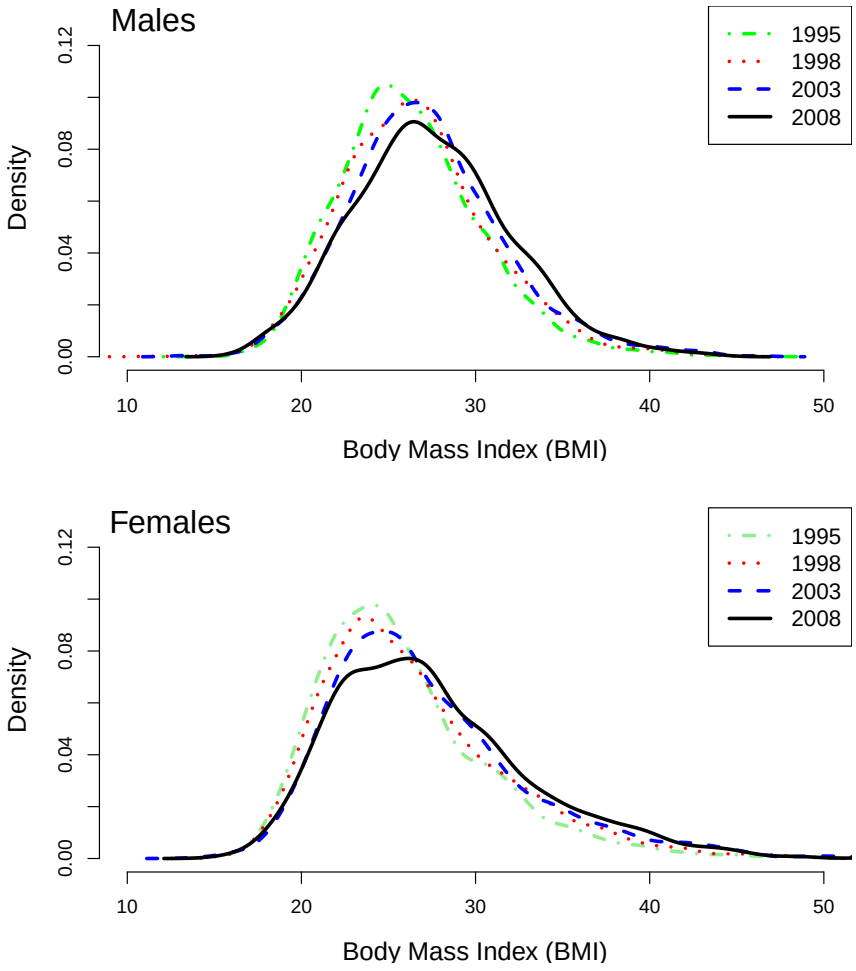

Figure 2.5: Density estimates of the survey population distributions of Body Mass Index (BMI). The BMI distributions in 1995, 1998, 2003 and 2008 are shown separately for males and females.

The BMI distributions shown in Figure 2.5 are representative of the survey population only, and therefore cannot necessarily be generalised to the population of Scotland. In order to display BMI distributions that are representative of the population of Scotland, a receiver operating characteristic (ROC) curve, incorporating survey weights, is used to illustrate the separation between BMI distributions in 1998, 2003 and 2008 in comparison to 1995 (Figure 2.6). The comparisons are not gender specific as the survey weights control for gender to correct for the under-representation of males. The x-axis contains the proportion of individuals within the Scottish population in 1995 with BMI values that are greater than specified threshold values, which in this case are BMI values that range between 18 and 60. Similarly, the y-axis contains the proportions of individuals within the Scottish populations in 1998, 2003 and 2008 with BMI values greater than the given thresholds. If the distributions of BMI in 1998, 2003 and 2008 were the same as in 1995 then we would expect the corresponding curves in Figure 2.6 to follow the dashed line, which is simply a straight line through zero to one, as that would be indicative of similar distributions. However, the fact that all three curves lie above the dashed line suggests that there has been an increase in the proportion of individuals with BMI values greater than the chosen threshold values in 1998, 2003 and 2008 when compared to the distribution of BMI in 1995. In other words, there has been a shift to the right in the population distributions of BMI, with an increasing shift in the distribution by survey year.

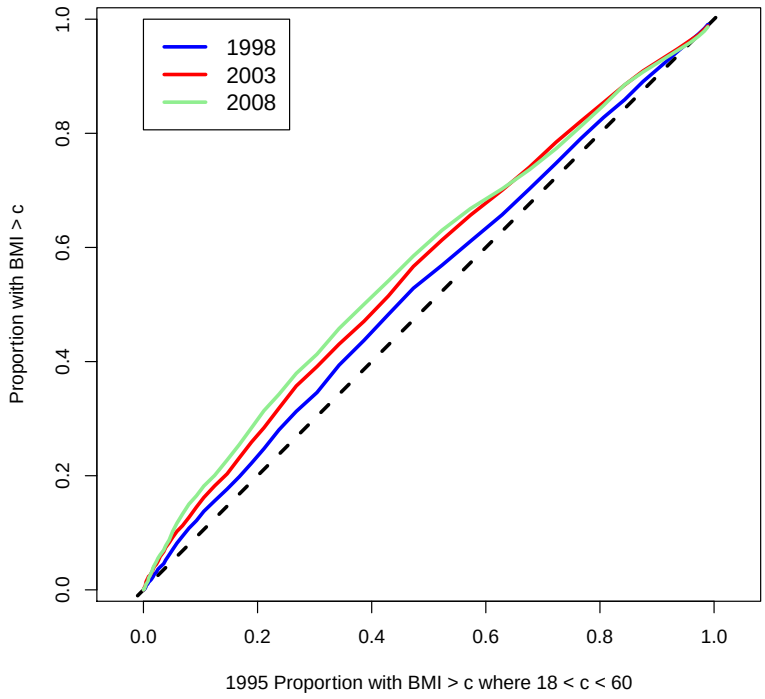

Figure 2.6: ROC curve illustrating the separation in the distributions of BMI. The distribution of BMI in 1995 provides a baseline for which comparisons can be made between 1995 and equivalent distributions in 1998, 2003 and 2008.

Chapter 3

Estimation of Overweight/Obesity Prevalence Using BMI

To study how obesity prevalence is related to various factors such as age and socio-economic status and how it changes over time, logistic regression models are fitted.

Binary logistic regression is a type of generalized linear model in which the outcome variable Y is measured on a binary scale, and depends on explanatory variables $(1, x_1, x_2, ..., x_p) = \mathbf{x}$. In this case, the binary response of the two categories is *obese* or *not obese*, and is defined as follows:

$$Y_i = \begin{cases} 1 & \text{if an individual is } \textit{obese} \text{ (BMI} \geq 30) \\ 0 & \text{if an individual is } \textit{not obese} \text{ (BMI} \leq 30) \end{cases}$$

with probabilities $Pr(Y_i = 1) = \pi(\mathbf{x_i})$ and $Pr(Y_i = 0) = 1 - \pi(\mathbf{x_i})$ from a Bernoulli distribution, where:

$$\log\left(\frac{\pi(\mathbf{x})}{1-\pi(\mathbf{x})}\right) = \boldsymbol{\beta}^T\mathbf{x} \tag{3.1}$$

for unknown parameters $\boldsymbol{\beta}^T = (\beta_0, \beta_1, ..., \beta_p)$ and $\mathbf{x}$, a design matrix with a column of 1s included. Alternatively, the logistic regression function can be expressed as

$$\pi(\mathbf{x}) = \frac{\mathbf{e}^{\beta^{\mathbf{T}}\mathbf{x}}}{\mathbf{1}+\mathbf{e}^{\beta^{\mathbf{T}}\mathbf{x}}} \tag{3.2}$$

where

$$\beta^{\mathbf{T}}\mathbf{x} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \tag{3.3}$$

The regression coefficients convey the magnitude of each contributing risk factor, with positive coefficients resulting in an increase in the probability of the chosen outcome for the given explanatory variable. More information on logistic regression methods can be found in Dobson & Barnett (2008).

3.1 Binary Logistic Regression for the Prevalence of Obesity

Using the statistical programming language **R** (R Development Core Team (2009)), logistic regression models of the form of equation (3.1) were fitted, separately for males and females. The response of each model is whether the individual is obese (BMI ≥ 30) or not. The covariates included were age, survey year and social class of chief income earner, with each treated as being categorical. The age categories are split up into 10 year age groups, which range from 16-24 to 55-64, with the corresponding baseline category being

individuals aged 16-24. These arbitrary groupings of age allow for a potential pattern or relationship between log(BMI) and age to be observed, before progressing onto using age as a continuous variable in our further analysis. For survey year, each year corresponds to that particular year of the SHS, and is therefore treated as categorical, with survey year 1995 acting as the baseline category. Similarly, social class of chief income earner is split into four categories (see Chapter 2 for details) with social classes i & ii considered as the baseline category. The interactions between the chosen covariates were found to be non-significant, and so only the main effects were included.

Figure 3.1 is a visual representation of the logistic regression models for the prevalence of obesity in males (top row) and females (bottom row), and displays the model-estimated odds of being obese for each combination of age group and social class for each of the Scottish Health Surveys in comparison to the baseline group, which is males/females aged 16-24 who are in social classes i & ii from the 1995 Scottish Health Survey.

Firstly, if we examine the relationship between the odds of obesity and age we see an increase in the odds of obesity with age, with the relationship differing between males and females. For males the odds of obesity are significantly greater in each subsequent age group in comparison to baseline (16-24) with a leveling out of the odds around the 55-64 age range. For females the odds of obesity are also significantly greater with each age group compared to baseline but increase at a slower rate within the lower age groups, with no apparent leveling off of the odds within the upper age ranges as observed for males.

The main effect fitted for survey year adequately describes the significantly increasing odds of obesity with each consecutive year in which the Scottish Health Survey was carried out, with survey year having a slightly

larger effect on the odds of obesity in females than it does in males.

For the social class of the chief income earner we see that, for males, the odds of obesity are greater in social classes iii non-manual, iii manual and iv & v in comparison to social classes i & ii. However, the odds are only significantly higher for males in the iii non-manual social class. There is no significant difference in the odds of obesity between social classes iii manual and iv & v. Also, the separation between the four social class groups becomes more apparent with age and the year in which the Scottish Health Survey was carried out. For females the pattern is similar, but the odds of obesity are significantly greater in each subsequent social class following social classes i & ii. The odds are at their highest in social classes iv & v, only slightly higher than those in the iii manual social class, with the separation between each of the social classes increasing with age and for each year of the Scottish Health Survey.

A goodness-of-fit test can be performed in order to check if the models for males and females are adequate fits to the data provided from the Scottish Health Surveys. One possible method of testing goodness-of-fit is the le Cessie-van Houwelingen test statistic for the unweighted sum of squared errors (le Cessie & van Houwelingen (1991)). In the case of the logistic regression model for the prevalence of obesity in males there is no evidence of a lack of fit based on this test with a p-value of 0.246 (p-value > 0.05). However, according to the le Cessie-van Houwelingen test, the main effects model for the prevalence of obesity in females is not a good fit to the data with a p-value of 0.003. The odds ratios obtained from the logistic regression model of obesity in females differ from empirical odds ratios acquired from the data. The odds ratios produced from the data are much greater than those obtained from the logistic regression model. For example, females aged 25-34 in social classes i & ii in 1995 are 12.3 times more likely to be

obese than females aged 16-24 from the same social classes and survey year (baseline group) according to the odds ratios obtained from the data, but are only 1.5 times more likely to be obese based on the model, which suggests that the model could well be underestimating the differences in the odds of obesity in females.

There is no evidence of lack of fit when all possible two-way interactions between age group, social class and survey year are included within the model fitted for females (suggesting that the inclusion of the interactions would be beneficial). However, the interactions between age group and survey year, and also, social class and survey year are non-significant, and therefore excluded from the model. Significant interactions between each age group and social classes iii non-manual and iii manual were observed, but non-significant interactions for social classes iv & v. Despite significant interactions between each age group and social classes iii non-manual and iii manual, the model still displays evidence of a lack of fit according to the le Cessie-van Houwelingen test, and also has a higher AIC (Akaike Information Criterion) than the main effects model, and therefore we will choose to keep the simpler main effects model.

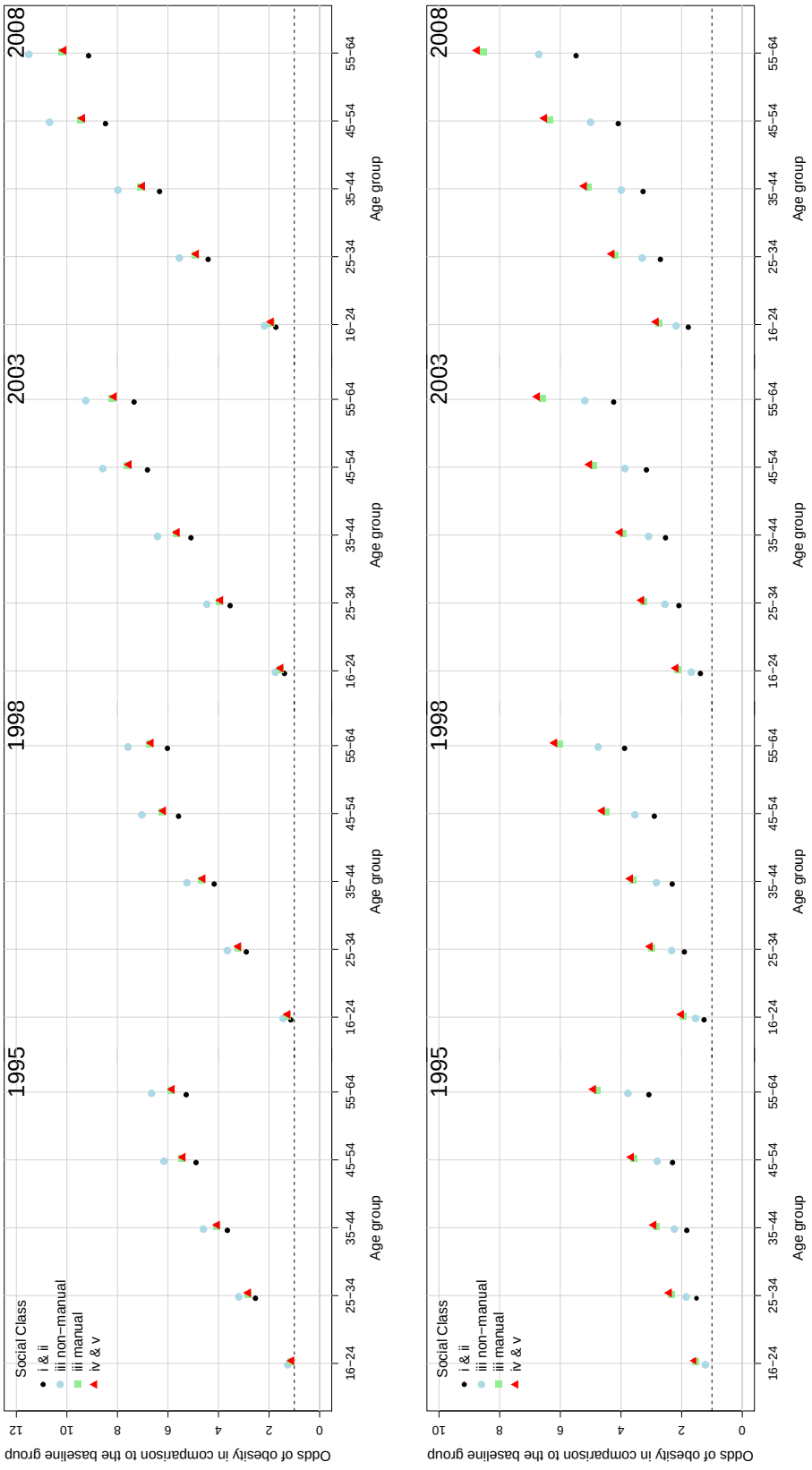

Figure 3.1: Model-estimated obesity prevalence in males (top row) and females (bottom row) by age group and socio-economic status for each year of the Scottish Health Survey. The dashed line indicates an odds ratio that is equal to 1. The corresponding estimated model coefficients can be found in Tables A.1 and A.2.

3.2 Binary Logistic Regression of Overweight and/or Obesity Prevalence

Models of the form of equation (3.1) were fitted for males and females separately, with the response this time being whether or not they had a BMI value greater than or equal to 25 (overweight or obese) with age (categorical), year and social class of chief income earner as covariates.

Figure 3.2 visually represents the logistic regression models for the prevalence of overweight and/or obesity in males (top row) and females (bottom row), with the odds of overweight and/or obesity in comparison to the baseline group (males/females aged 16-24 in social classes i & ii surveyed in 1995) displayed on the y-axis for each combination of age group and social class and for each year in which the Scottish Health Surveys were carried out.

For males, the odds of being overweight or obese significantly increase with each age group in a relatively linear fashion, whereas the relationship seen for females is similar to that from the previous model for the prevalence of obesity, with a slow rate of increase at the lower age ranges before increasing in the upper age ranges.

Again, as with the logistic regression models for the prevalence of obesity, the main effect was found to be adequate for survey year. The odds of being overweight or obese significantly increase with each year of the Scottish Health Survey, with survey year having a slightly stronger effect on the odds of females being overweight or obese than on the odds for males.

Studying the effects of social class we see that, for males, the odds of being overweight or obese are highest in social classes iii non-manual and iii manual but do not differ significantly from those in social classes i & ii.

However, the odds of being overweight or obese do differ significantly for males in social classes iv & v, with the odds of being lower for males in social classes iv & v in comparison to males in social classes i & ii. Also, there is an increase in the separation in the odds between each of the four social class groups with age and survey year. For females the highest odds of being overweight or obese are observed for those within the iii manual social class, social classes iv & v and then in social class iii non-manual, with each similar to one another, but significantly different from social classes i & ii.

The logistic regression models for the prevalence of overweight/obesity in males and females both fit the data well, according to the le Cessie-van Houwelingen goodness-of-fit test (p-values of 0.49 and 0.14, respectively).

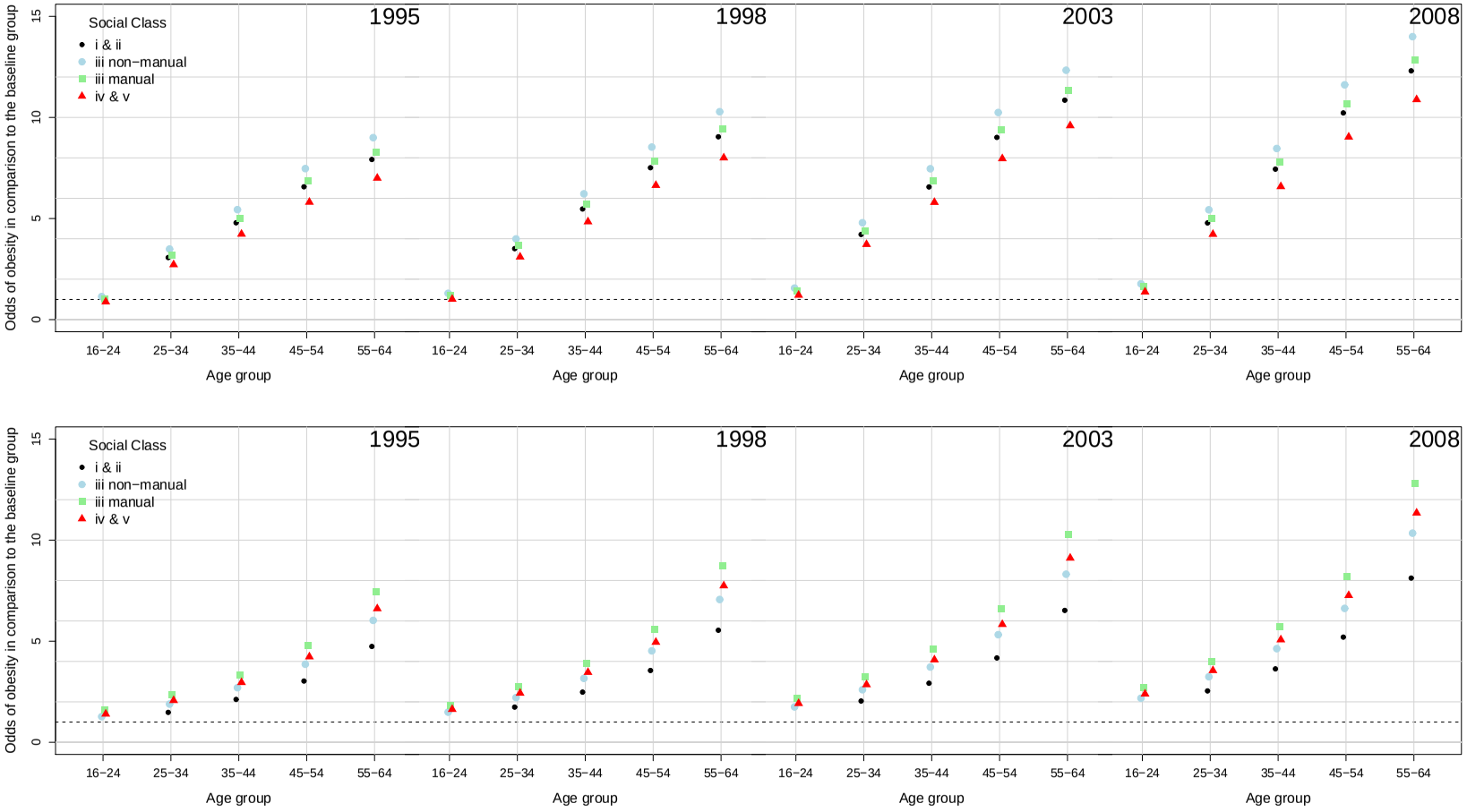

Figure 3.2: Model-estimated overweight/obesity prevalence in males (top row) and females (bottom row) by age group and socio-economic status for each year of the Scottish Health Survey. The dashed line indicates an odds ratio that is equal to 1. The corresponding estimated model coefficients can be found in Tables A.3 and A.4.

3.3 Summary of the Logistic Regression Results of Overweight/Obesity Prevalence

The logistic regression models show an increase in the odds of being overweight or obese as age increases, with differing patterns for males and females. In males, a leveling off of the odds of overweight/obesity prevalence was seen within the 55-64 age range, which is more apparent from the logistic regression model for the prevalence of obesity (Figure 3.1). This leveling off of the odds in later years is not present in females, with the odds of being overweight or obese increasing more slowly in early age, but more quickly in middle to later life.

Obesity prevalence is at its highest for males who are in social class iii non-manual, and less so in social classes iii manual and iv & v, with very little difference between these two social groups. Obesity prevalence in males is lowest within social classes i & ii, but this is not the case for the prevalence of overweight or obesity, were the lowest odds can be found within social classes iv & v. The odds of being overweight or obese for females follow a similar pattern to what was seen with males, however, the odds are at their highest in social classes iv & v, slightly higher than those in the iii manual group. For both males and females the separation in the odds between the social classes becomes more apparent with age, and also, the year in which the Scottish Health Survey was carried out, with overweight/obesity prevalence generally lower for those of higher socio-economic status.

As discussed in Sections 3.1 and 3.2, a main effect term was found to adequately capture the effect of survey year, with no significant interactions between survey year, social class and/or age group. The odds of obesity and overweight/obesity increase with each consecutive survey year, and have a slightly stronger effect on the odds of obesity in females than in males.

Chapter 4

Generalised Additive Models Using BMI

Having considered logistic regression with age as a categorical covariate, we now turn our attention to logistic regression using generalised additive models (GAMs). As age is not naturally categorical and its relationship to the log odds of obesity prevalence is unlikely to be linear, GAMs allow for the age covariate to be treated as continuous in the logistic regression models of Chapter 3, without restricting its functional form.

GAMs are generalised linear models in which the linear predictor comprises a sum of smooth functions of the covariates of interest. Generally, the structure of a generalised additive model is given by

$$g(\mu_i) \quad = \quad \boldsymbol{\beta^T x_i^*} + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}, x_{4i}) + \ldots$$

where

$\mu_i \equiv \mathbb{E}(Y_i)$, the response variable Y_i follows some exponential family distribution, $\boldsymbol{x_i^*}$ is a row of the design matrix for any strictly parametric model components, $\boldsymbol{\beta}$ is the corresponding parameter vector, and the f_j are smooth

functions of the covariates, $(x_1, x_2, \ldots, x_p)$.

Smoothing is performed using cubic smoothing splines. If the selected smoothing parameter is too high then the data will be over-smoothed, while if it is too low, the data will be under-smoothed, which means that, in both cases, the spline estimate $\hat{f}$ will not be close to the true function f. The ideal outcome is to choose the smoothing parameter so that $\hat{f}$ is as close as possible to f. The smoothing parameter is chosen using cross-validation, as described in Wood (2006), and implemented in the **mgcv** package (Wood (2006)) in **R** (R Development Core Team (2009)).

Here, the generalised additive models used for males and females are given by

$$g(\pi) = \boldsymbol{\beta^T x_i^*} + f(\text{age}_i) \tag{4.1}$$

where

$$g(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$$

is the log odds of being obese, and

$$\beta^{\mathbf{T}}\mathbf{x_i^*} = \beta_0 + \sum_{j=1}^{3} \beta_j \ (\text{survey year})_j + \sum_{j=4}^{6} \beta_j \ (\text{social class})_j.$$

4.1 Generalised Additive Models for the Prevalence of Obesity

Figure 4.1 displays the plots obtained from generalised additive models (4.1) for the prevalence of obesity in males and females, respectively. The first plot for males (top row of Figure 4.1) displays the relationship between

obesity prevalence and age, and highlights increasing obesity prevalence with increasing age. This increase in the prevalence of obesity appears to peak around 55-60 years of age. Obesity prevalence in social classes iii manual and iv & v does not differ significantly from social classes i & ii, which can be observed from the second plot of Figure 4.1 as the confidence bands contain zero. However, obesity prevalence is found to be significantly greater for males in social class iii non-manual. Each consecutive survey year resulted in an additive increase in the log odds of a male being obese, from 0.133 in 1998 to 0.544 in 2008, which can be observed from the estimated main effect of survey year in the final plot on the top row of Figure 4.1. The estimates for 1995 and social classes i & ii (baseline survey year and social class) are shown as zero, because the first level of a categorical variable is used as a reference category.

For females (bottom row of Figure 4.1) the relationship between obesity prevalence and age differs from that observed for males. The rate of increase in obesity prevalence is at its greatest amongst females aged 16-30, while a slower rate of increase is observed for females between the ages of 30 and 64. Unlike with males, a clear leveling off in obesity prevalence is not apparent for females over the age of 50. Also, at the lower and upper ends of the age range, the 95% confidence bands are at their widest, which is likely due to there being fewer observations at these points. The log odds of obesity significantly increase with each subsequent social class following social classes i & ii, with the greatest log odds of obesity observed for females within social classes iv & v. Again, for each consecutive year of the Scottish Health Survey the log odds of females being obese increase. Also, the impact of survey year is larger for females, especially in 1998, in comparison to males (estimates of 0.133 and 0.227 in males and females, respectively).

The social class effects observed are quite different to those found by

Zaninotto et al. (2009). Whereas Zaninotto et al. (2009) had socio-economic status split between manual and non-manual social classes with the prevalence of obesity higher in the manual social class, the social class relationship here is not as straightforward, especially for males. Here we see that males within the iii non-manual social class actually have the highest odds of being obese, whereas no significant increase in the odds was observed for males in social classes iii non-manual and iv & v. Perhaps such differences between the two studies would not occur if social classes i & ii and iii non-manual were combined to form a non-manual social class and likewise for the manual social classes, but such interesting differences between social classes would have been lost.

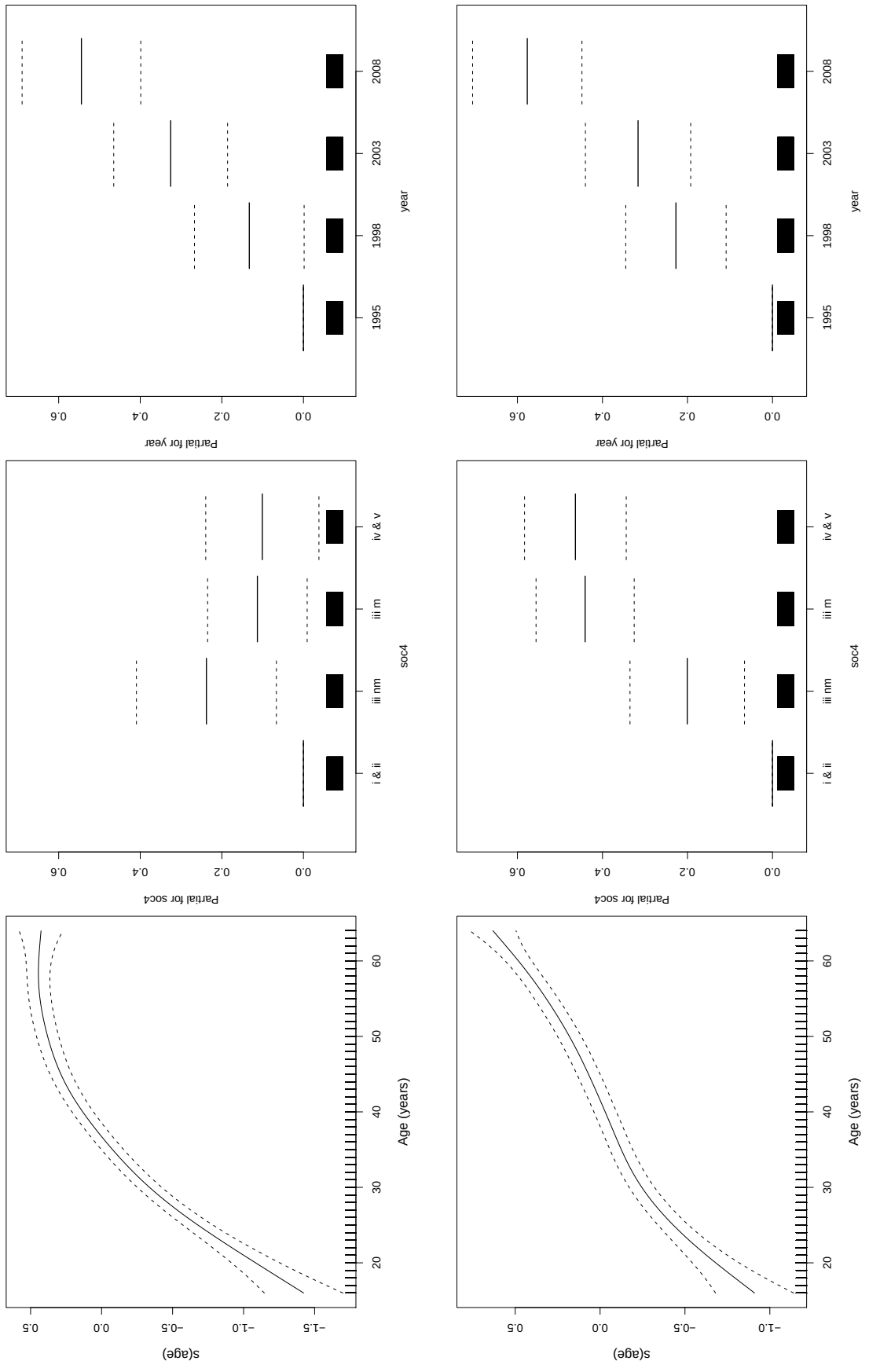

Figure 4.1: Generalised additive models for the prevalence of obesity in males (top row) and females (bottom row), respectively. The dashed lines represent 95% confidence intervals.

4.2 Generalised Additive Models for Overweight and/or Obesity Prevalence

Figure 4.2 shows the plots obtained from the generalised additive models described in equation (4.1) with the response being whether an individual has or does not have a BMI ≥ 25 (overweight or obese), for males and females, respectively. Again, the baseline level for survey year is 1995, and for social class is i & ii. The results obtained are very similar to those observed for the prevalence of obesity (Section 4.1) with a couple of exceptions. Firstly, for males, the prevalence of obesity was found be significantly higher for males in the iii non-manual social class and non-significant for the remaining social classes, but for the prevalence of overweight/obesity in males there is no significant difference observed between any of the social classes and social classes i & ii. For females a different relationship is also seen by socio-economic status. Females in social classes iv & v were found to have the highest odds of being obese (Figure 4.1), but for the prevalence of overweight or obese females we see that females in the iii manual social class have the highest odds of being overweight or obese.

Since the data obtained from the Scottish Health Surveys are cross-sectional, which is data gathered from subjects at a given point in time, we cannot be certain that BMI goes down for males when they approach 55-60 years of age. This is due to different subjects being sampled in each subsequent Scottish Health Survey after 1995, so for example, older males sampled in 2003 or 2008, say, could have been lighter (or "skinnier") than their counterparts in 1995, which could be a possible cause for the observed relationship between age and BMI. There could also be generational effects to take into consideration. For example, the increasing prevalence throughout the years of fast food chains is more than likely to have had an impact on the weight or diet of males born within the last 20-30 years or so, than it

has for the more mature males sampled.

Ultimately, if longitudinal data were available, in which the same individuals are observed over time, such as the study by Reas et al. (2007), then we would be able to better understand the relationship between age and BMI.

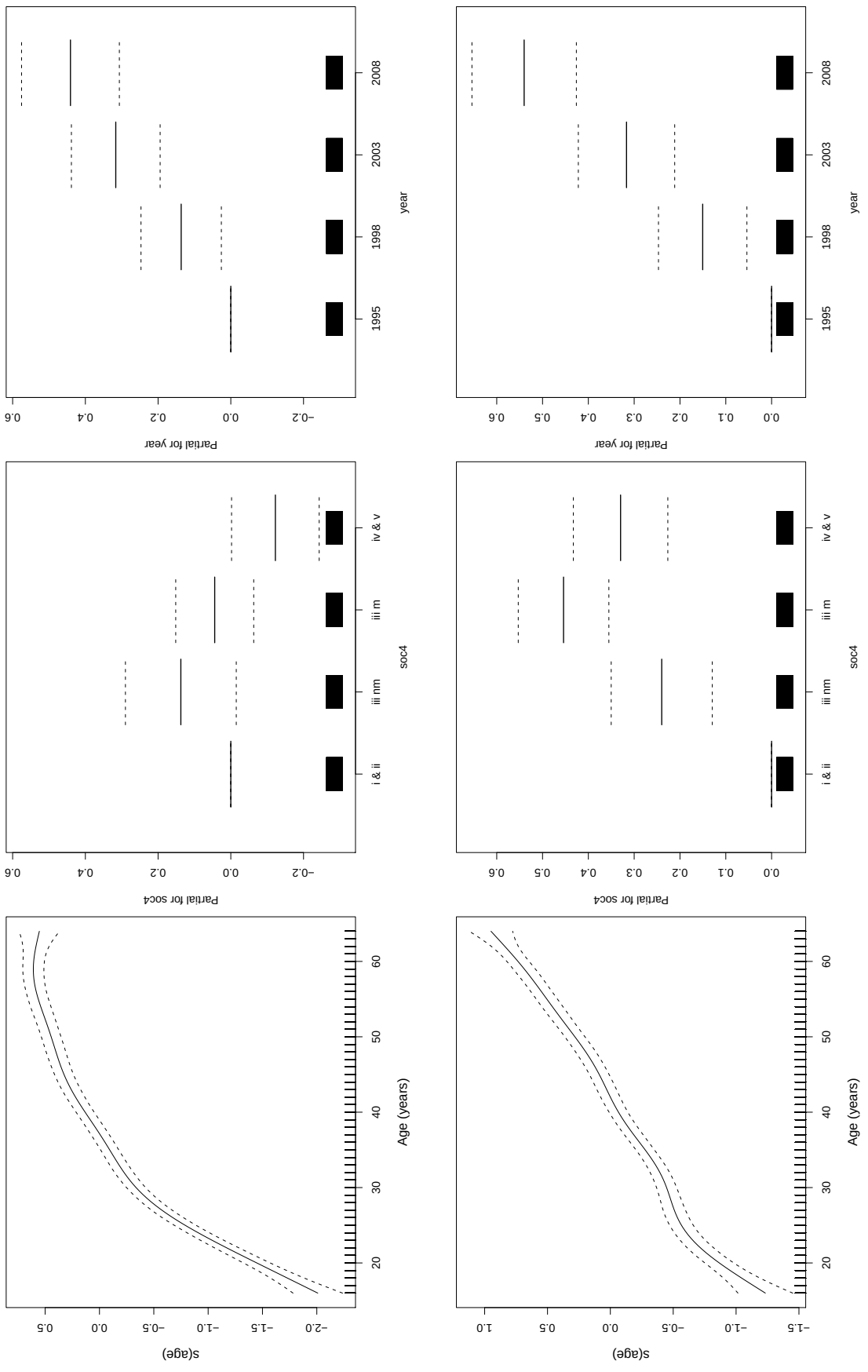

Figure 4.2: Generalised additive model plots for the prevalence of overweight/obesity in males (top row) and females (bottom row). The dashed lines represent 95% confidence intervals.

4.3 Comparison of Linear and Generalised Additive Models for Logistic Regression

For males, both linear and generalised additive models with the log odds of obesity as the response produced similar results. In particular, both showed an increase in the log odds of obesity with age, with a leveling off of the odds of obesity in older males. The effect of survey year was also similar, which was essentially an increase in the odds of obesity with each consecutive year in which the Scottish Health Survey was carried out. Similar results were also obtained by social class. For overweight or obese males both methods showed an increase in the log odds of being overweight or obese with age and survey year. However, there was a slight difference with age, with binary logistic regression suggesting an increase by age group, whereas for the GAM, a leveling off of the log odds was observed around 50-55 years of age. For obese females both modelling methods agree, with both showing an increase in the log odds of obesity with age, social class and survey year, with perhaps a slightly different pattern observed for age. Similar results were also seen for both modelling methods when looking at the prevalence of overweight or obese females.

The main difference between the two modelling methods is that the age covariate is likely to be more accurately represented within the GAMs which allow for it to be modelled as continuous and nonlinear. In contrast, the age covariate is categorized in the linear logistic regression model, but a a continuous and possibly nonlinear term is seldom realistic and age is not naturally a categorical variable. For example, it is very unlikely that the odds of a person being overweight or obese would suddenly increase overnight as they move from one age group to another, but they are assumed do so with the categorization of age.

4.4 Generalised Additive Model for the BMI

Just as it is more informative to treat age as a continuous variable, the same can be said for the BMI. Until now the BMI has been dichotomized, but it is possible, and more informative, to model it as a continuous response variable. Using mean BMI as the response instead of the probability of whether or not someone is obese will provide us with more information on BMI values themselves. Models of BMI for both males and females will be taking the natural logarithm of BMI as the response variable, since examination of the residual plots and distribution of BMI suggest that a log transformation would be more suitable. The log transformation has also been used in previous studies such as Chambers & Swanson (2010).

The fitted generalised additive model with log(BMI) as the response is given by

$$\text{mean } \log(\text{BMI}) = \beta_0 + \sum_{j=1}^{3} \beta_j \ (\text{survey year})_j + \sum_{j=4}^{6} \beta_j \ (\text{social class})_j + f(\text{age}_i)$$

and, as with previous models within Chapter 4, was fitted in **R** using the **mgcv** package.

Figure 4.3 shows the generalised additive models of log(BMI) for males and females respectively, and shows the relationship between log(BMI) and age for each social class separately for each year in which the Scottish Health Survey was carried out. For males (top row of Figure 4.3) the rate of increase in log(BMI) is found to be at its greatest in young males (between 16 and 30 years of age), and slows down in middle age before a decrease is observed for males aged 55 and over. Males in social classes iv & v have lower mean log(BMI) than males in the remaining social classes, however they do not differ significantly from males in social classes i & ii. Males in social class iii manual have higher mean log(BMI) than males in social classes i & ii, but again, the difference observed is non-significant. There is however a signifi-

cant difference in log(BMI) observed for males in social class iii non-manual, with males within this social class having, on average, significantly higher log(BMI) values than males within social classes i & ii.

For females (bottom row of Figure 4.3) the relationship between log(BMI) and age is more linear, with a slightly quicker increase in mean log(BMI) observed between the ages of 16 and 30, with no apparent maximum (or peak) reached. Females in social classes iii non-manual, iii manual and iv & v were found to have mean log(BMI) values significantly greater than those of females in social classes i & ii, females in social class iii manual found to have the highest mean log(BMI).

Also, for both males and females, each plot of Figure 4.3 shows an upward shift on the y-axis with survey year. For males, an additive increase of 0.01 units in log(BMI) is observed in 1998, which increases to 0.04 units in 2008. The increase with survey year is slightly greater for females, with an increase in log(BMI) of 0.02 units in 1998, and an increase of 0.06 units in 2008.

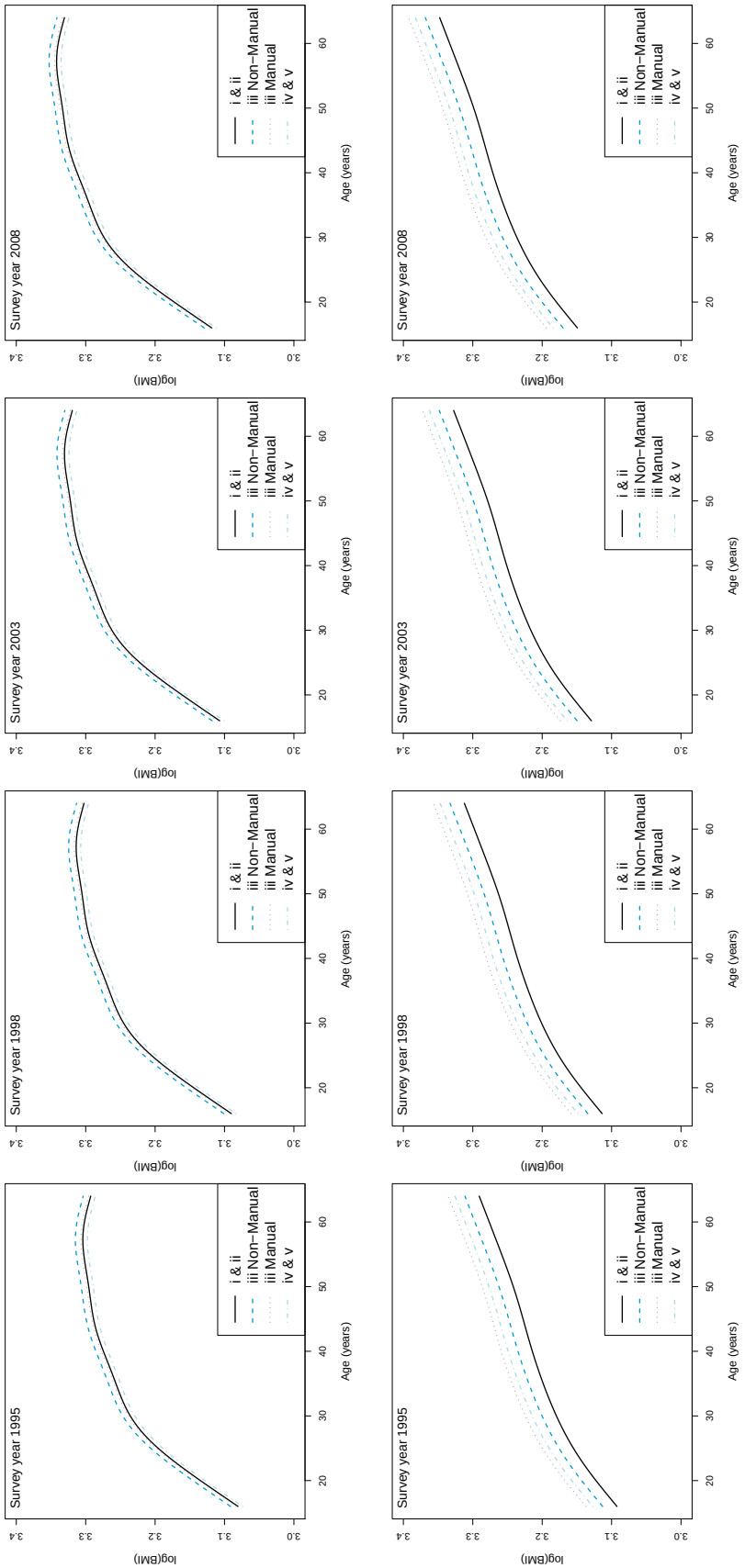

Figure 4.3: Generalised additive model plots of log(BMI) for males (top row) and females (bottom row) with the effect of social class and age shown for each year of the Scottish Health Survey.

Figure 4.4 displays the diagnostic plots for the generalised additive models of log(BMI) for males and females, respectively. Both normal q-q plots show that the assumptions of linearity and log-normality appear to hold for each of the models. Additional residual plots, not shown here, indicate that the assumption of constant variance is valid with no visible pattern or signs of heteroscedasticity.

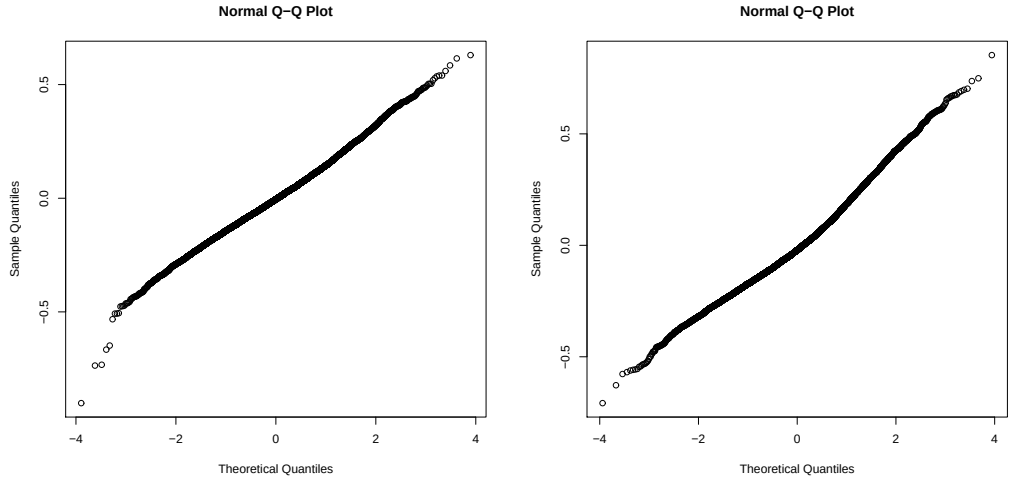

Figure 4.4: Diagnostic plots for the generalized additive models of log(BMI) for males (left) and females (right), respectively.

For both males and females, the generalised additive models of mean log(BMI) appear to qualitatively agree with the linear logistic regression models and generalised additive models with log odds of obesity and overweight or obesity as the response. Although the results yielded from modelling overweight/obesity prevalence and log(BMI) are very similar, from the interpretation point of view modelling log(BMI) allows for more information to be gained. Instead of focusing on predetermined BMI cut-off points for whether or not someone is overweight or obese, modelling with mean log(BMI) as the response variable allows us to study the effects that age, social class and survey year have on the log(BMI) values themselves.

4.5 Summary of the Generalised Additive Model Results

From the generalised additive models of obesity and overweight/obesity prevalence, both males and females are more likely to be overweight or obese with increasing age, with the odds reaching a maximum around 55-60 years of age in males. No significant difference in obesity prevalence was found for males in social classes iii manual and iv & v in comparison to social classes i & ii, but males in social class iii non-manual are 1.27 times more likely to be obese than males in social classes i & ii. The odds of being obese for females increased with each subsequent social class following social classes i & ii.

The generalised additive models with mean log(BMI) as the response also provided similar results to those which were obtained from the GAMs for the prevalence of overweight/obesity. Mean log(BMI) increases with age, with the additive term fitted for social class being similar for males and females to that described for the GAMs of overweight/obesity prevalence above.

A main effect term was fitted for survey year in each of the generalised additive models, with no significant interactions found between any of the covariates. Much like the studies by Zaninotto et al. (2009) and Wang et al. (2008), the prevalence of overweight or obese individuals was found to increase over time. An increase in mean log(BMI) was also observed with each consecutive year in which the Scottish Health Survey was carried out. The effect survey year has on overweight/obesity prevalence and mean log(BMI) is slightly higher in females than it is in males.

Chapter 5

Quantile Regression

To obtain an even more complete picture of how the BMI varies with gender, age, socio-economic status and survey year, we will use quantile regression models. Quantile regression (Koenker & Bassett (1978)) is a method to formally study the entire population distribution of BMI, and was recently implemented in a study of obesity in China by Shankar (2010), and applied to BMI percentile curves by Li et al. (2010). Shankar (2010) highlights the benefits of using quantile regression when noting considerable cross-quantile variation between study variables and the BMI, which ordinary least squares (OLS) methods fail to detect. Quantile regression would allow us to assess whether the population strategy proposed by Rose (see Chapter 2) is an acceptable approach for assessing BMI, or whether other competing hypotheses, such as those posed by Penman & Johnson (2006), are more appropriate when describing the population distribution of BMI.

Quantile regression analysis is intended to estimate, and conduct inference about, either the median or other quantiles of a given response variable. Whereas previous regression methods have concentrated on either the log odds, or the conditional mean of the response variable, quantile regression models the relationship between given predictor variables and precise quantiles (or percentiles) of the chosen response variable, which in this case is

log(BMI).

By specifying changes in the quantiles of the response, quantile regression allows us to look at the whole BMI distribution in order to identify whether or not there is a shift in the entire distribution, and not only around the mean and also to study whether such a shift is uniform or of variable size, depending on the quantile of the distribution. This allows us to check if BMI is increasing for all and not only for those individuals who are considered overweight and/or obese.

We fit a linear model for the τ^{th} conditional quantile of log(BMI),

$$Q_{\log(\text{BMI})}(\tau|\mathbf{x}) \equiv \inf\{y : \Pr(\log(\text{BMI}) \leq y|\mathbf{x}) \geq \tau\}$$

$$\text{so that } \Pr(\log(\text{BMI}) \leq \beta(\tau)^T\mathbf{x}) = \tau$$

where

$$\begin{aligned}\beta(\tau)^T\mathbf{x} &= \beta_0(\tau) + \beta_1(\tau)\text{ year (1998)} + \beta_2(\tau)\text{ year (2003)} + \beta_3(\tau)\text{ year (2008)} \\ &+ \beta_4(\tau)\text{ social class (iii non-manual)} + \beta_5(\tau)\text{ social class (iii manual)} \\ &+ \beta_6(\tau)\text{ social class (iv \& v)} + \beta_7(\tau)\text{ age group (25-34)} \\ &+ \beta_8(\tau)\text{ age group (35-44)} + \beta_9(\tau)\text{ age group (45-54)} \\ &+ \beta_{10}(\tau)\text{ age group (55-64)}\end{aligned}$$

As quantile functions allow for monotonic transformations whilst preserving equivariant properties, the relationship between the quantiles of BMI on the logarithmic and original scale are the same. With log(BMI) as the response we observe a multiplicative effect for the covariates instead of an additive one but the equivariance property allows us to interpret these effects on the original BMI scale. Quantile regression models are fitted using the **quantreg** package (Koenker (2009)) in **R**.

5.1 Quantile Regression for Males

Figure 5.1 displays the quantile regression plots obtained for males, with each plot showing the distribution of log(BMI) ranging from the 5^{th} and 95^{th} percentiles for the coefficients of survey year, social class and age group. The grey bands highlight the 95% confidence intervals for values of τ between 0.05 and 0.95, and are obtained from pointwise confidence intervals using bootstrap. The solid red lines indicate the corresponding values in accordance with the least squares method, while the dashed red lines illustrate the 95% confidence bands for this least squares estimate. The first plot (labeled intercept) may be interpreted as the estimated cumulative distribution (or conditional quantile function) of log(BMI) for a male aged 16-24 who lies within social classes i & ii from the Scottish Health Survey of 1995. The remaining plots show, at any chosen quantile ranging from the 5^{th} to the 95^{th}, the differing log(BMI) values corresponding to males given each level of the chosen covariates.

Firstly, looking at survey year, according to the ordinary least squares estimate of the mean effect, males from the Scottish Health Survey of 1998 have, on average, BMI values that are 1.01 (exponential of 0.01) times that in 1995. However, from the quantile regression results, no change in log(BMI) is observed at the lower tail (bottom 25%) of the distribution, while at the upper end (top 20%) of the distribution BMI is 1.03 times greater for males in 1998 than in 1995. In 2003, with the exception of the bottom 10% of the distribution, there is a noticeable increase in log(BMI) for males from this survey year in comparison to 1995. The multiplicative increase in log(BMI) becomes more pronounced from the lower to the upper end of the distribution, were BMI is 1.05 times that observed in 1995. The disparity between log(BMI) values of males in 2008 and 1995 is even greater, and as with prior survey years, is highest at the upper end of the distribution. This increase

in BMI ranges from 1.02 to 1.06 times greater at the lower and upper tails of the distribution, respectively. Here, more so than any other survey year, the conventional least squares method poorly represents such changes in the multiplicative increase in log(BMI) observed across the entire distribution.

Now, if we examine the quantile regression plots for the various social classes, unlike with survey year, lower log(BMI) is found in the lower tails of the distributions in comparison to baseline social classes i & ii. BMI values observed at the bottom 10% and 20% of the distributions for social classes iii non-manual and iii manual, respectively, are significantly lower (95% confidence bands do not contain zero), and are 0.98 times those observed for social classes i & ii. For social classes iv & v BMI is significantly different at the bottom 40% of the distribution and is 0.96 times that observed in social classes i & ii. However, as we approach the median, BMI is 1.05 and 1.04 times greater for males in social classes iii non-manual and iii manual, respectively, with both estimates equivalent to the ordinary least squares estimate of the mean effect, but this observed increase in BMI does not differ significantly from that of males in social classes i & ii. Similarly, no significant change in median BMI is observed for social classes iv & v. The observed change in sign from the 5^{th} to the 50^{th} percentile highlights the benefits of quantile regression, as such changes cannot be captured using least squares regression. For the iii non-manual social class BMI is 1.02 times greater in the upper tail of the distribution, while for social class iii manual the remainder of the distribution remains stable and similar to the least squares estimate. For social classes iv & v, at the upper tail of the distribution, BMI is 1.015 times greater than that observed for social classes i & ii, which is not a significant increase in BMI (95% confidence bands contain zero).

Across the entire distribution of log(BMI) each age group differs significantly from males aged 16-24 (baseline age group). For males aged 25-34

the quantile regression results across the entire distribution of log(BMI) are very similar to the ordinary least squares estimate for the mean effect, with BMI estimated to be 1.12 times that of males aged 16-24. The increase in log(BMI) in the remaining age groups peaks around the 40^{th} percentile, but is at its lowest at the upper tail of the distribution. Also, an upward shift (on the y-axis) is found with each consecutive age group, which results in a greater increase in the multiplicative effect observed with age. For example, for males aged 35-44, at the upper tail of the distribution, BMI is 1.13 times that for males aged 16-24, but for males aged 55-64 the increase is even greater, with an increase in BMI that is 1.15 times that observed for males aged 16-24.

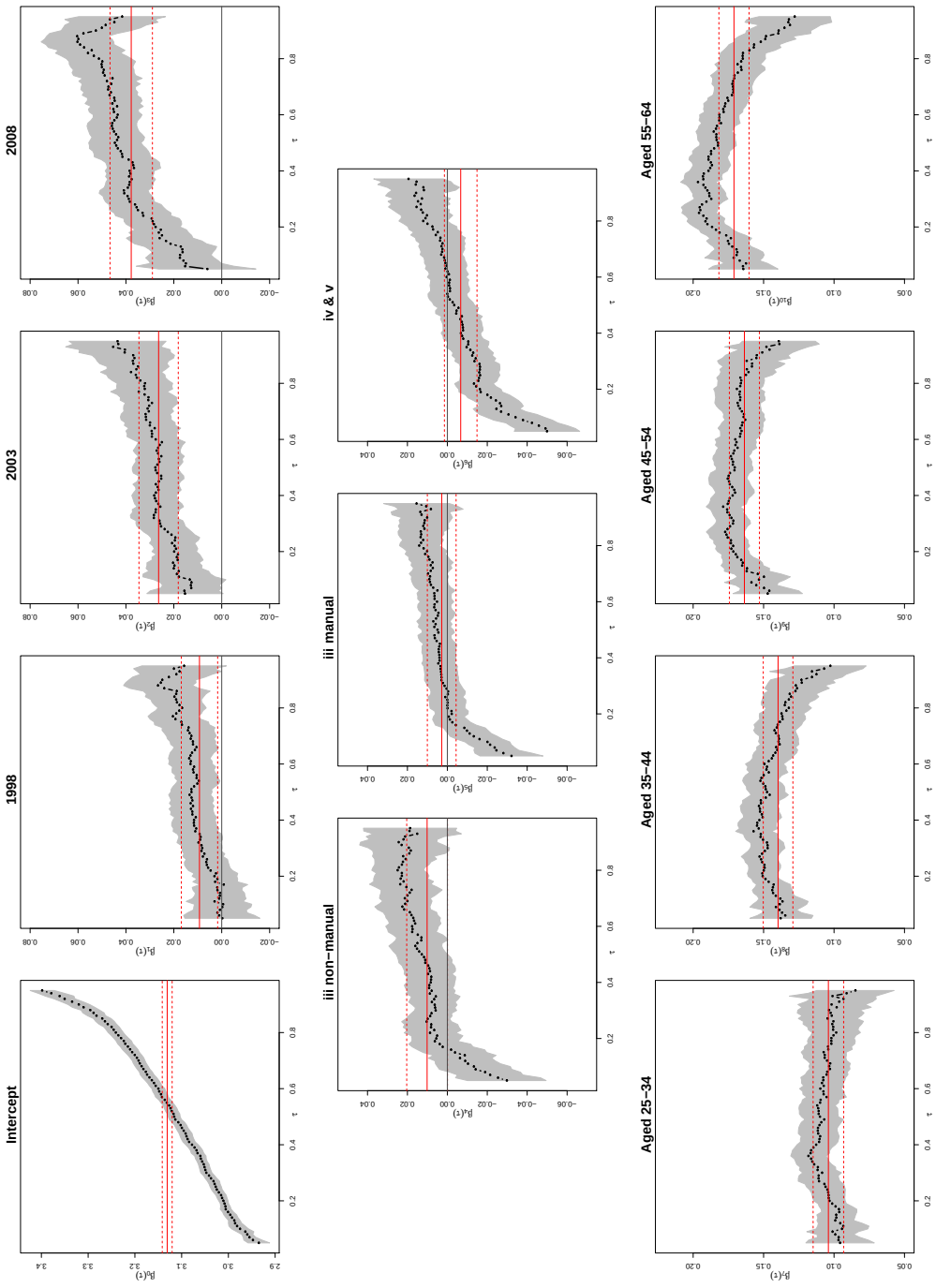

Figure 5.1: Survey year, social class and age group coefficients with 95% confidence bands for values of τ between 0.05 and 0.95 for males. The solid red lines indicate the value corresponding to the least squares coefficient. The dashed red lines are 95% confidence bands for this least squares estimate.

5.2 Quantile Regression for Females

Figure 5.2 shows the quantile regression results obtained for females, with each plot as described for males at the beginning of Section 5.1. The cross-quantile relationship between survey year and log(BMI) is very similar to that observed for males, but with a slightly higher increase (upward shift on the y-axis) in log(BMI) across the entire distribution. The biggest gender contrast found is in the relationship between social class and log(BMI). Instead of a change in sign across the distribution, as observed for males, there is a steady increase in log(BMI) with increasing τ for each subsequent social class following social classes i & ii, with the exception of the lower tail of the log(BMI) distribution for social classes iii non-manual and iv & v where no change in log(BMI) is observed. This results in females in social classes iv & v having the highest increase in BMI. The relationship between log(BMI) and age is very similar to that observed for males, with log(BMI) increasing with each subsequent age group, but remaining fairly uniform across the distribution, with the exception of females aged 55-64 where the highest observed increase is found between the 30^{th} and 40^{th} percentiles.

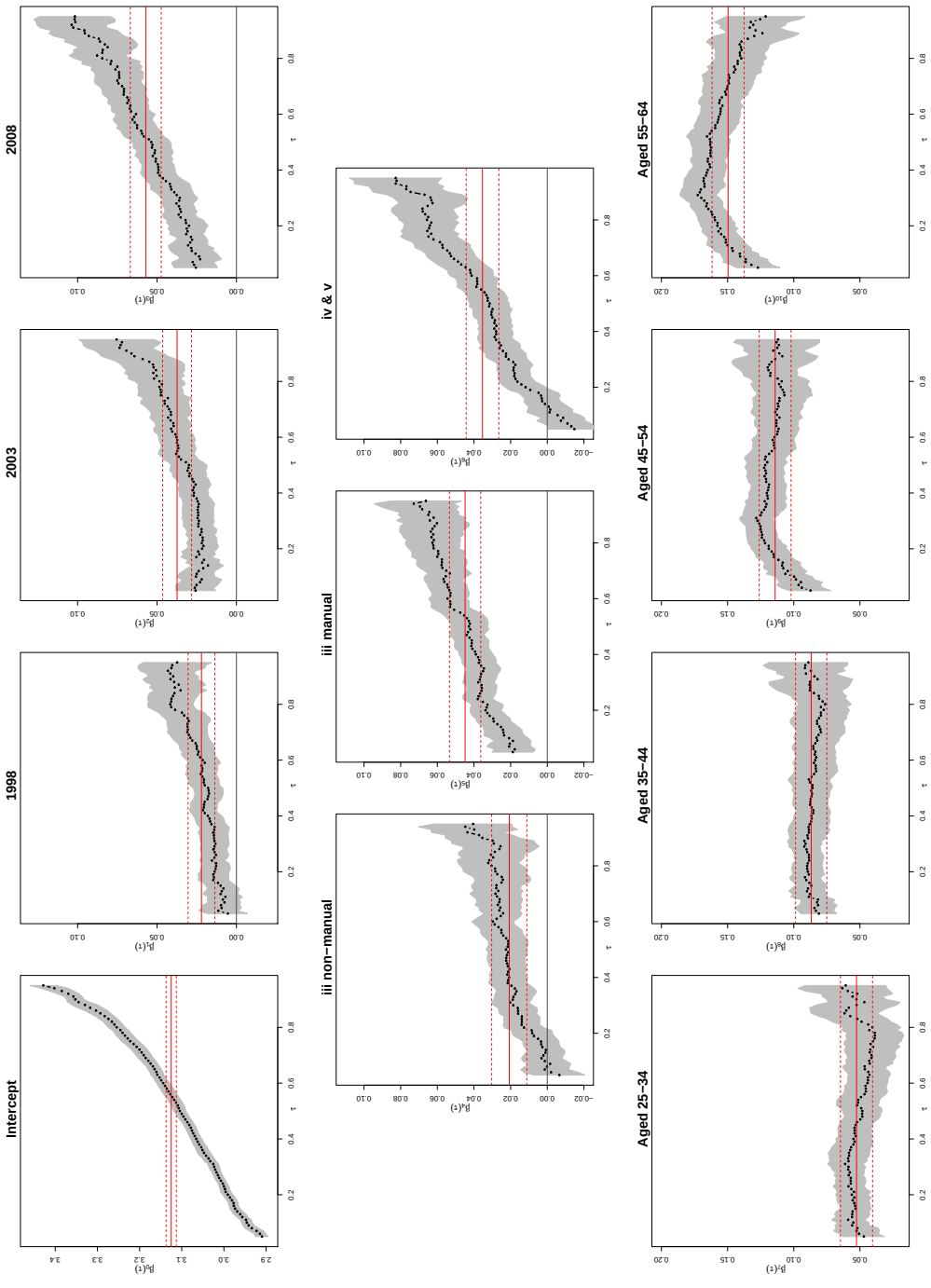

Figure 5.2: Survey year, social class and age group coefficients with 95% confidence bands for values of τ between 0.05 and 0.95 for females. The solid red lines indicate the value corresponding to the least squares coefficient. The dashed red lines are 95% confidence bands for this least squares estimate.

5.3 Quantile Regression with a Non-Parametric Term for Age

As with generalised additive models in Chapter 4, incorporating age as a continuous covariate into the quantile regression model will allow for a closer examination of the relationship between log(BMI) and age for each social class and survey year. One such strategy would be to use cubic basis splines (B-splines), which are a sequence of curve segments pieced together to form a single continuous curve, with knots acting as internal breakpoints between connected spline segments (de Boor et al. (1986)). The function **bs()** from **R** package **splines** (Bates & Venables (2009)) is used to construct the cubic B-spline basis functions, with the model given as:

$$\begin{aligned} Q_{\log(\text{BMI})}(\tau|\mathbf{X}) &= \beta_0(\tau) + \beta_1(\tau)\ \text{year (1998)} + \beta_2(\tau)\ \text{year (2003)} \\ &+ \beta_3(\tau)\ \text{year (2008)} + \beta_4(\tau)\ \text{social class (iii non-manual)} \\ &+ \beta_5(\tau)\ \text{social class (iii manual)} \\ &+ \beta_6(\tau)\ \text{social class (iv \& v)} + g_\tau\ \text{(age)} \end{aligned}$$

where $\tau \in [0, 1]$ and $g_\tau(\cdot)$ is a nonlinear function of age approximated by a linear combination of cubic B-spline basis functions with fixed knots at age 35 and 49 (the 33^{rd} and 66^{th} percentiles of the age distribution). Various numbers and positions of knots were examined before settling on the two chosen at the 33^{rd} and 66^{th} percentiles of the age distribution. These knots were chosen as they provide a relatively smooth term for age, with the degree of smoothness being similar to that in the GAMs in Section 4.4.

Figure 5.3 displays the quantile regression plots obtained for males with age incorporated as a continuous variable through the implementation of B-splines. The plots in Figure 5.3 corresponding to the survey year and social

class coefficients are very similar to those seen in Figure 5.1 and, as such, will not be discussed within this section (See Section 5.1 for details). An interesting feature can be see in the last plot of Figure 5.3, which displays the 10^{th}, 50^{th} and 90^{th} quantiles of the log(BMI) distribution as a function of age for survey year 2008, with similar relationships between log(BMI) and age observed for previous survey years.

The rate of increase in log(BMI) with age is at its greatest in the early years of adulthood (say, between the ages of 16 and 30), with the rate of increase gradually diminishing before a decrease is observed at approximately 55-60 years of age. This increase is most prominent at the upper quantiles (90^{th} quantile), where the separation between social classes is also at its greatest. Males at the lower quantiles of the log(BMI) distribution (10^{th} quantile) in social classes i & ii have higher log(BMI) values than those in each subsequent social class. However, at the median, were the separation between social classes is minimal, males in social classes i & ii have log(BMI) values that are approximately the same as those for males in social classes iii manual and iv & v, but slightly less than those in social class iii non-manual. At the upper quantiles (90^{th} quantile) of the distribution males in social classes i & ii have lower log(BMI) values than males in each subsequent social class. Note, however, that as the data are not longitudinal, we cannot distinguish between generational effects and aging.

Using age as a categorical variable with quantile regression allows for the examination of cross-quantile variation within each age group, which although interesting and informative, does not provide us with as detailed a relationship between log(BMI) and age as a continuous variable, which was also observed in the transition from linear to generalised additive models for logistic regression (Chapter 4). Also, the GAMs with mean log(BMI) as the response show a similar relationship to that observed at the median, but

quantile regression allows us to study the full distribution of BMI in greater detail.

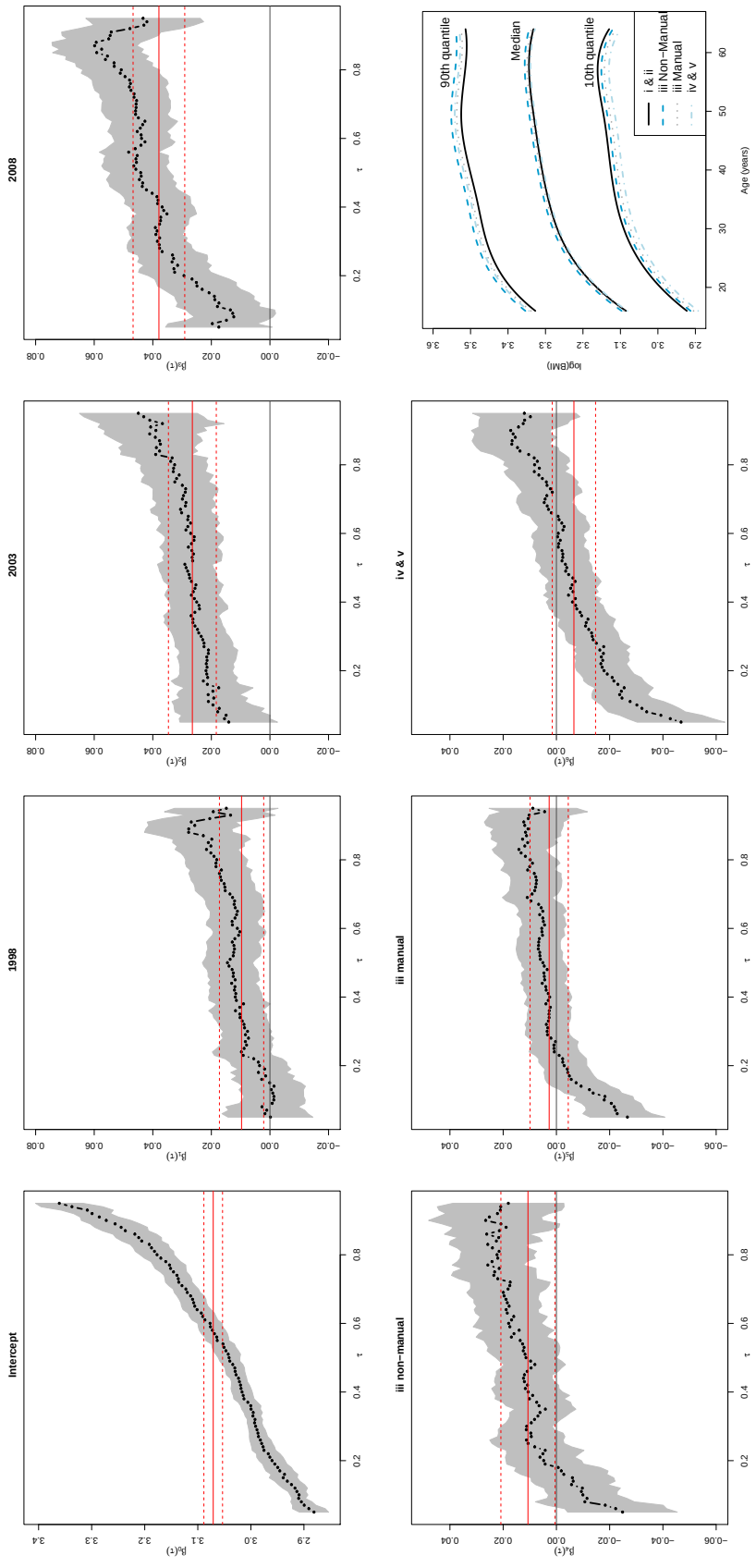

Figure 5.3: Survey year and social class coefficients with 95% confidence bands for values of τ between 0.05 and 0.95 for males. The solid red lines indicate the value of the corresponding least squares coefficient, while the dashed red lines are 95% confidence bands for this least squares estimate. The final plot displays the 10^{th}, 50^{th} and 90^{th} fitted log(BMI) quantiles as a function of age for survey year 2008.

The quantile regression plots for females in Figure 5.4 corresponding to the survey year and social class coefficients are very similar to those obtained from the quantile regression model with age as a categorical variable (see Figure 5.2) and therefore will not be discussed, with our focus turning to log(BMI) as a function of age.

Firstly, we can see that females have higher log(BMI) values than males at the lower, middle and upper quantiles of the distribution, with no discernable decrease in log(BMI) observed for females aged 55-60. The rate of increase in log(BMI) in younger females is not as great as that found in males at the 10^{th} quantile and median, with a relatively steady increase observed for females aged 16-64, but at the upper quantiles of the distribution the rate of increase in log(BMI) is found to be greater for females aged between 16 and 30. At the 10^{th} quantile there is no discernable difference in log(BMI) between females in social classes i & ii, iii non-manual and iv & v, with females in social class iii manual having the highest log(BMI) values. However, at the median the separation between social classes is more pronounced, with females in social classes i & ii having lower log(BMI) values than females in social class iii non-manual followed by social classes iv & v and iii manual, respectively. The separation in social classes is even more pronounced at the upper quantiles of the distribution than at the median, with females in social classes i & ii having the lowest log(BMI) values, while females in social classes iv & v have the highest log(BMI) values.

In comparison to the GAM of mean log(BMI) (Figure 4.3), the relationship between log(BMI) and age, and also the separation between social classes is very similar to that observed at the median, with perhaps a slightly quicker increase in log(BMI) for females between the ages of 16 and 30.

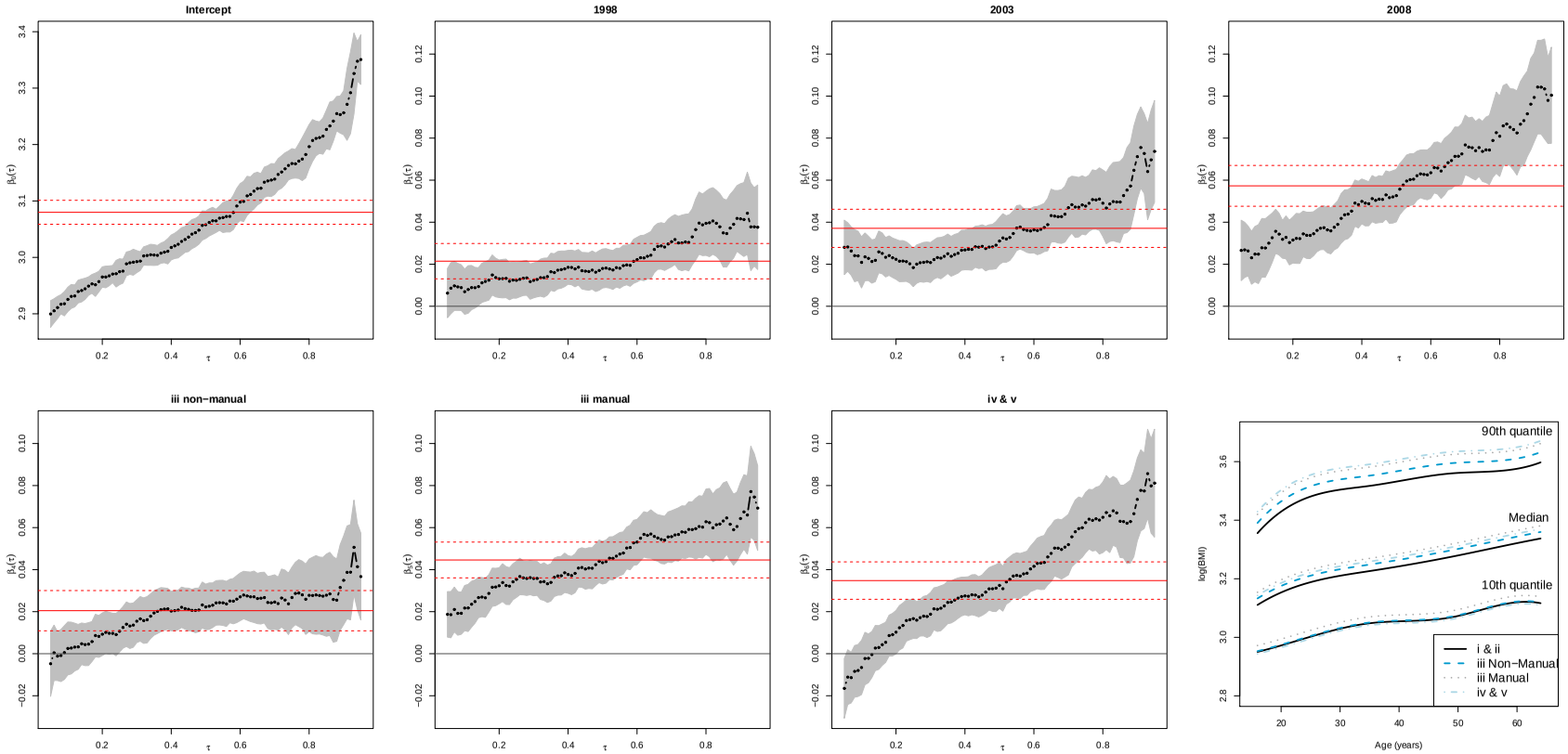

Figure 5.4: Survey year and social class coefficients with 95% confidence bands for values of τ between 0.05 and 0.95 for females. The solid red lines indicate the value of the corresponding least squares coefficient, while the dashed red lines are 95% confidence bands for this least squares estimate. The final plot displays the 10^{th}, 50^{th} and 90^{th} fitted log(BMI) quantiles as a function of age for survey year 2008.

5.4 Summary of Quantile Regression Results

Controlling for age and social class, median BMI for males in 2008 is 1.04 times that in 1995. However, the increase at the 10^{th} percentile (underweight males) is lower than that observed at the median, with BMI values in 2008 that are 1.02 times those in 1995. At the same time, at the 90^{th} percentile (obese males) of the distribution have BMI values that are 1.06 times greater than those of males surveyed in 1995. For females, median BMI in 2008 is 1.06 times that of females surveyed in 1995. Again, as with males, the increase in BMI observed is at its lowest at the lower tail (10^{th} percentile) of the distribution, with females in 2008 having BMIs which are 1.03 times those in 1995. At the upper tail (90^{th} percentile), BMI values of females in 2008 are 1.11 times those of females surveyed in 1995.

Controlling for age and survey year, median BMI values for males in social classes iii non-manual, iii manual and iv & v did not differ significantly from those of males in social classes i & ii, with imperceptible difference in BMI observed. However, a prominent lowering of the BMI is observed at the lower tail (10^{th} percentile) in each social class in comparison to social classes i & ii. For example, 'underweight' males in social classes iii non-manual and iii manual have BMIs which are 0.98 times those of males in social classes i & ii. Underweight males in social classes iv & v have BMI values that are 0.96 times those in social classes i & ii. Interestingly, despite the lower BMI values at the lower end of the distributions for each social class, and little to no change around the median, higher BMI values were observed at the upper end of the distributions. The highest BMI values at the upper end of the distribution are found for males in social class iii non-manual, with obese males having BMI values that are 1.025 times those of males in social classes i & ii. For females, BMI values in all social classes are higher in comparison to social classes i & ii, and increasing from the lower to the upper end of the

distributions with each subsequent social class. For example, median BMI for females in social classes iv & v is 1.03 times that of females in social classes i & ii, but is 1.07 times greater at the 90^{th} percentile of the distribution.

For both males and females, given survey year and social class, an increase in BMI was observed across the entire distribution for all age groups. For males aged 25-34, the rise in BMI observed was similar to that of the least squares estimate throughout the whole distribution, while a dip in the observed rise in BMI was found at the upper end of the distributions in each subsequent age group, and was most prominent in males aged 35-44 and 55-64. Also, the rise in BMI observed became greater with age. For example, median BMI for males aged 25-34 is 1.13 times that of males aged 16-24, while the median BMI for males aged 55-64 is 1.2 times greater than that of males aged 16-24. For females aged 25-34, 35-44 and 45-54, the increase in BMI observed is relatively similar across the entire distribution. For females aged 55-64 there was a rise in the increase in BMI around the 40^{th} percentile, and a slight drop at the upper tail of the distribution. Again, as with males, the increase in BMI observed became more prominent with age. The median BMI for females aged 25-34 is 1.05 times greater than that for females aged 16-24, but is 1.19 times greater for females aged 55-64.

The cross-quantile variation observed also highlights the advantages of using quantile regression, with ordinary least squares (OLS) methods unable to capture such changes, which was also noted in the study by Shankar (2010), who believes that this is very helpful when considering the best means of introducing intervention to reduce obesity in China.

Chapter 6

Waist-to-Hip Ratio

The waist-to-hip ratio (WHR) is the ratio of the smallest measurable circumference of the waist divided by the widest circumference of the hips. Ideally, to maintain a low risk level of encountering health problems associated with obesity the WHR should be no greater than 0.95 for males and 0.80 for females, with moderate and high health risk cut-off points shown in Table 6.1.

Table 6.1: Health risk based on waist-to-hip ratio measurements

Male	Female	Health Risk
≤ 0.95	≤ 0.80	Low Risk
$0.96 - 1.00$	$0.81 - 0.85$	Moderate Risk
> 1.00	> 0.85	High Risk

Body fat location is also an important health risk factor. Fat stored in the abdomen is the most likely to lead to health problems associated with obesity, with individuals who store fat in the abdomen known to have apple-shaped bodies. People who store fat mostly around their hips and buttocks, described as pear-shaped, are less likely to encounter obesity-related health problems, and in turn have lower waist-to-hip ratio measurements than those

who are apple-shaped (Gesta et al. (2007)).

A study carried out by Yusuf et al. (2005) discovered that using WHR instead of BMI as an indicator of obesity resulted in a three-fold increase in population attributable risk for myocardial infarction. Also, research by Srikanthan et al. (2009) indicated that the waist-to-hip ratio is better than the BMI for measuring obesity in the elderly. This is due to physical changes that are a part of the aging process having an effect on the height and weight measurements used to define the BMI.

6.1 Exploratory Analysis of Waist-to-Hip Ratio Data

First of all, the number of males and females surveyed with valid waist-to-hip ratio measurements was lower than those surveyed with valid BMIs, with the resulting sample sizes for waist-to-hip ratio data shown in Table 6.2, which also displays the sample sizes of those surveyed with valid BMI measurements (as seen in Table 1.2). Due to the reduced sample sizes the results obtained from the models using the BMI cannot be directly compared with those for the waist-to-hip ratio.

Table 6.2: Males and females surveyed with valid BMI and WHR measurements for each Scottish Health Survey. The sample sizes shown in **bold** are those for males and females with valid BMI measurements.

	Survey Year			
	1995	1998	2003	2008
Males	2761 (**3118**)	2503 (**2986**)	1569 (**2275**)	319 (**1774**)
Females	3247 (**3740**)	3014 (**3637**)	1920 (**2816**)	406 (**2221**)

Figure 6.1 shows the percentage of males and females who are at low, moderate and high risk of encountering health-related issues (as described in Table 6.1) associated with obesity for each year in which the Scottish Health Survey was carried out. As there are no predefined weight categories in accordance with waist-to-hip ratio measurements we cannot determine the percentage of males and females surveyed who would be deemed obese. However, it is clear that the percentage of males and females surveyed with moderate and high health risks associated with obesity has increased from 1995 to 2008. It is also worth noting that for each survey year there are a larger percentage of females with moderate and high health risks, with considerable differences observed in 2003 and 2008 between males and females with high health risk. This may be due to the lower waist-to-hip ratio cut-off points (shown in Table 6.1) for females (with males having slightly more leeway with their waist-to-hip ratio measurements). Also, as noted for the BMI, the increase in WHR, and in turn, the increase in the percentage of individuals with high health risk could partially be accounted for by the differing age distributions for each survey year. With fewer young people surveyed with each subsequent SHS, an increase in WHR may be expected, with people gaining weight with age.

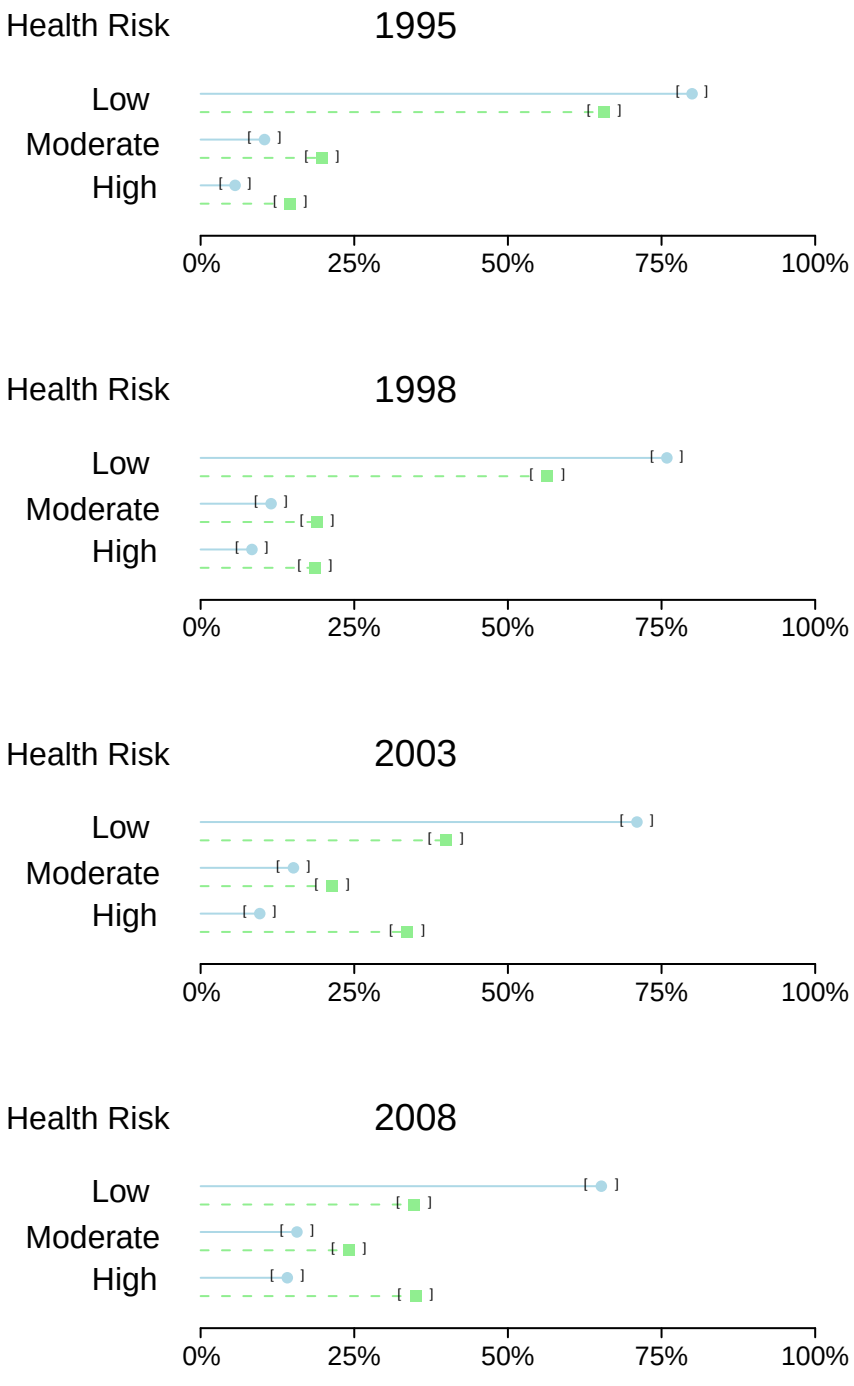

Figure 6.1: Percentage of population surveyed by health risk according to waist-to-hip ratio for each year of the Scottish Health Survey. Solid and dashed lines represent males and females, respectively. Parentheses show ±1 standard error bounds.

Figure 6.2 displays density estimates of the survey population distributions of waist-to-hip ratio for each survey year, separately for males and females. As was seen with the BMI distributions (Figure 2.5), the waist-to-hip ratio distribution has become increasingly positively skewed with each survey year, with the observed shift more pronounced for females, especially in survey years 2003 and 2008. Also, for females, a large separation is found between the waist-to-hip ratio distributions in 1995/1998 and 2003/2008.

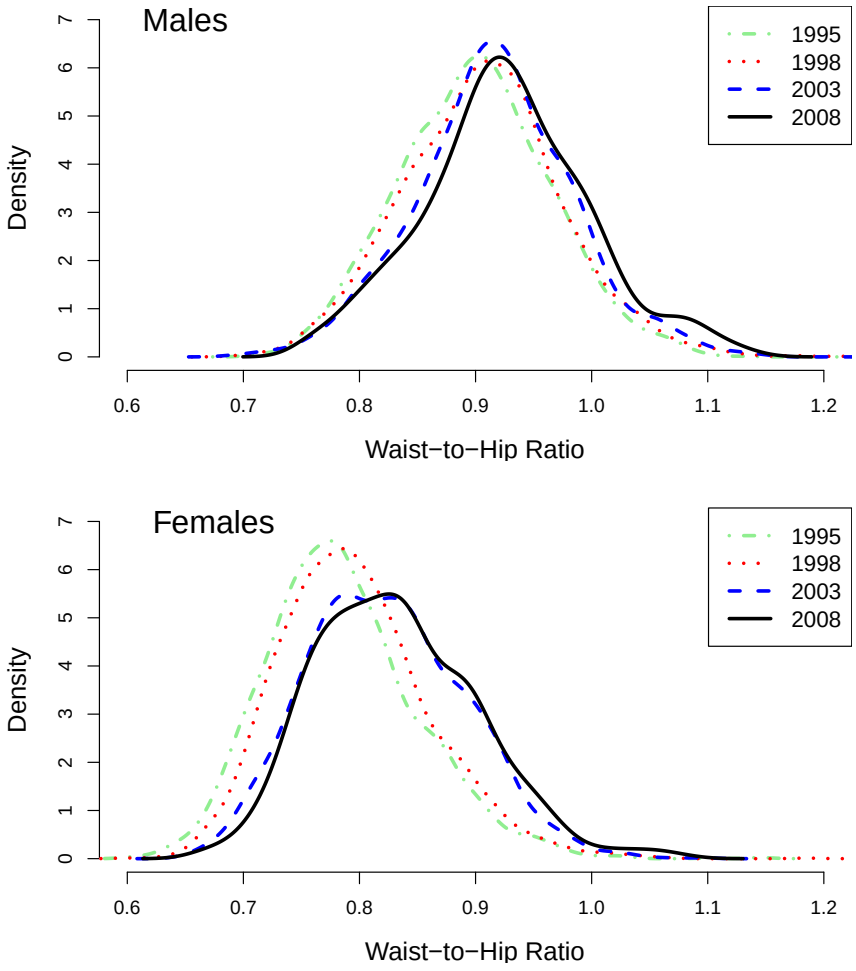

Figure 6.2: Density estimates of the survey population distributions of waist-to-hip ratio. Waist-to-hip ratio distributions in 1995, 1998, 2003 and 2008 are shown separately for males and females.

Figure 6.3 is a receiver operating characteristic curve for the distributions of waist-to-hip ratio measurements in survey years 1998, 2003 and 2008 and shows the separation between these distributions and that observed in 1995 (as described in Section 2.1 for the BMI). The displayed pattern is very similar to that observed for the distributions of BMI (as seen in Figure 2.6), but with increased separation found between survey years 2003 and 2008 in comparison to 1995, and also in comparison to the waist-to-hip ratio distribution in 1998.

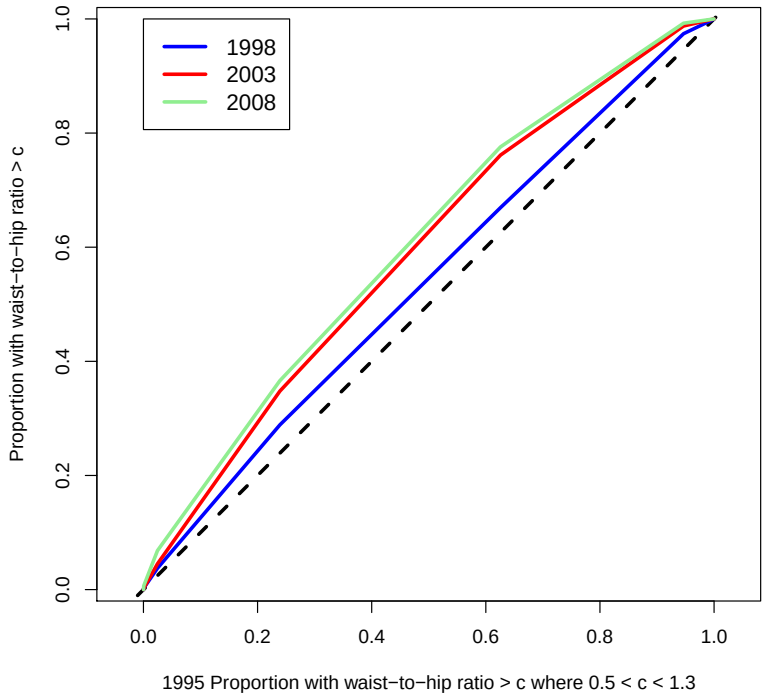

Figure 6.3: ROC curve illustrating the separation in the distributions of waist-to-hip ratio. The distribution of waist-to-hip ratio in 1995 provides a baseline for which comparisons can be made between 1995 and equivalent distributions in 1998, 2003 and 2008.

6.2 Generalised Additive Models for the Waist-to-Hip Ratio

Using generalised additive models provides us with information on the relationship between mean waist-to-hip ratio measurements and age, social class and survey year. The generalised additive models of WHR fitted separately for males and females using the **mgcv** package in **R** are given by

$$\text{mean WHR} = \beta_0 + \sum_{j=1}^{3} \beta_j \ (\text{survey year})_j + \sum_{j=4}^{6} \beta_j \ (\text{social class})_j + f(\text{age}_i)$$

and are of the same form as the generalised additive models fitted for log(BMI) (Section 4.4). A model with log(WHR) as the response was also considered, but due to negligible differences in the results obtained between WHR and log(WHR), the transformation was not employed, for ease of interpretation.

Figure 6.4 displays the results obtained for the generalised additive models of waist-to-hip ratio measurements for males and females respectively, and shows the relationship between the waist-to-hip ratio and age for each social class, separately for each survey year. For males (top row of Figure 6.4) the observed rate of increase in waist-to-hip ratio measurements is at its quickest between the ages of 16 and 30 before gradually slowing down, much like what was found with log(BMI) (Figure 4.3). However, unlike with log(BMI), there is no apparent decrease in waist-to-hip ratio observed for males over the age of 55. The social class pattern observed for waist-to-hip ratio measurements differs from that found using log(BMI) and is what may have initially been expected. Males in each subsequent social class following social classes i & ii have statistically significantly higher waist-to-hip ratio measurements, with males in social classes iv & v having the highest waist-to-hip ratio measurements, with no discernable difference found between males in social classes

iii non-manual and iii manual.

For females (bottom row of Figure 6.4) the relationship between waist-to-hip ratio and age is more or less linear (more so than was observed for log(BMI) (Figure 4.3)), with waist-to-hip ratio constantly increasing with age. The relationship between waist-to-hip ratio and social class is similar to that observed for log(BMI), with females in each subsequent social class following social classes i & ii having subsequently higher waist-to-hip ratio measurements.

With each survey year an increase in waist-to-hip ratio measurements is observed for both males and females, which is similar to that found for log(BMI), with the observed increase slightly greater for females.

With the results obtained for the BMI and WHR proving to be relatively similar to one another, with the exception of a few noted differences, it is not easy to see what could be learned from using the WHR, within the context of this study. For one, due to more data being available for the BMI and the ease with which it can be attained make it quite difficult to recommend the WHR in this context, which again, is further heightened by the results obtained for the WHR being similar to those found for the BMI. However, what we do see is that, regardless of the chosen classifier, an increase in the prevalence of obesity is observed.

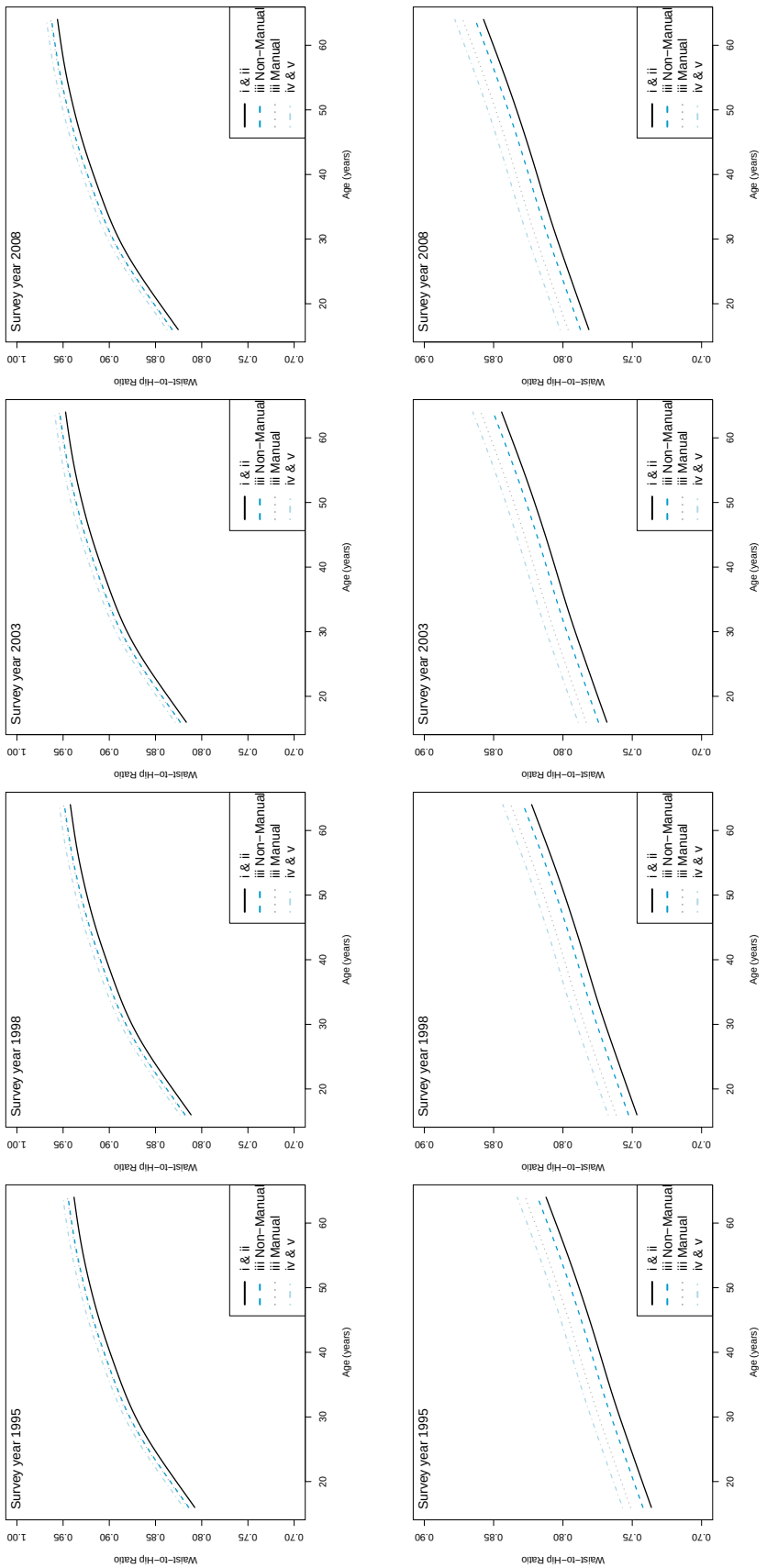

Figure 6.4: Generalised additive model plots of waist-to-hip ratio for males (top row) and females (bottom row) with the effect of age and social class shown for each year of the Scottish Health Survey.

6.3 Quantile Regression with Waist-to-Hip Ratio

We will now use quantile regression (see Chapter 5) to formally assess the entire waist-to-hip ratio distribution, with cubic B-splines used to incorporate age as a continuous variable as described in Section 5.3. The quantile regression models fitted separately for males and females are given by

$$\begin{aligned} Q_{\text{WHR}}(\tau|\mathbf{X}) &= \beta_0(\tau) + \beta_1(\tau)\ \text{year (1998)} + \beta_2(\tau)\ \text{year (2003)} \\ &+ \beta_3(\tau)\ \text{year (2008)} + \beta_4(\tau)\ \text{social class (iii non-manual)} \\ &+ \beta_5(\tau)\ \text{social class (iii manual)} \\ &+ \beta_6(\tau)\ \text{social class (iv \& v)} + g_\tau\ \text{(age)} \end{aligned}$$

where $\tau \in [0, 1]$ and $g_\tau(\cdot)$ is a nonlinear function of age approximated by a linear combination of cubic B-spline basis functions with fixed knots at age 35 and 49 (the 33^{rd} and 66^{th} percentiles of the age distribution). As waist-to-hip ratio is the response and not the logarithm, the effect of survey year and social class can be interpreted as additive.

Figure 6.5 displays the results obtained from the quantile regression model for males, and shows percentiles ranging from the 5^{th} and 95^{th} of the waist-to-hip ratio distribution for the coefficients of survey year and social class. The final plot shows the relationship between the waist-to-hip ratio and age at the 10^{th}, 50^{th} and 90^{th} percentiles of the distribution.

A uniform pattern is observed for males surveyed in 1998, with waist-to-hip ratio measurements not differing significantly from males surveyed in 1995 across the entire distribution, with the exception of a slight increase

found between the 25^{th} and 45^{th} percentiles. A similarly uniform pattern is observed for males surveyed in 2003, but this time the waist-to-hip ratio measurements are significantly different to those of males in 1995, with an observed waist-to-hip ratio increase of 0.01, and a slightly higher increase at the upper end of the distribution (increase of 0.015). The pattern observed in 2008 is quite different, with waist-to-hip ratio increasing with increasing values of τ until up to a 0.03 increase at the upper end of the distribution. Also, the increase in waist-to-hip ratio becomes greater with each subsequent survey year.

The waist-to-hip ratio measurements for males in social class iii non-manual do not differ from those for males in social classes i & ii across the lower half of the distribution (values of $\tau < 0.5$), but do so at the upper end of the distribution, where a difference in waist-to-hip ratio of 0.01 is found. With the exception of the lower end of the distribution, males in social class iii manual have waist-to-hip ratios that are significantly greater than those of males in social classes i & ii with the observed increase remaining fairly stable across the distribution. Again, with the exception of the lower end of the distribution, males in social classes iv & v have waist-to-hip ratios that are significantly higher than those of males in social classes i & ii. However, the observed difference is much greater at the upper end of the distribution. For example, at the median ($\tau = 0.5$) the difference in waist-to-hip ratio is close to 0.01, but at the upper end of the distribution it is closer to 0.03.

Studying the relationship between waist-to-hip ratio and age (final plot of Figure 6.5) we see that, the rate of increase in waist-to-hip ratio is slightly quicker in young males, before slowing down while continuing to increase (at the median). At both the 10^{th} and 90^{th} quantiles, a slight decrease in waist-to-hip ratio is observed for males over the age of 60. Also, the separation in social classes is most apparent at the upper end of the distribution.

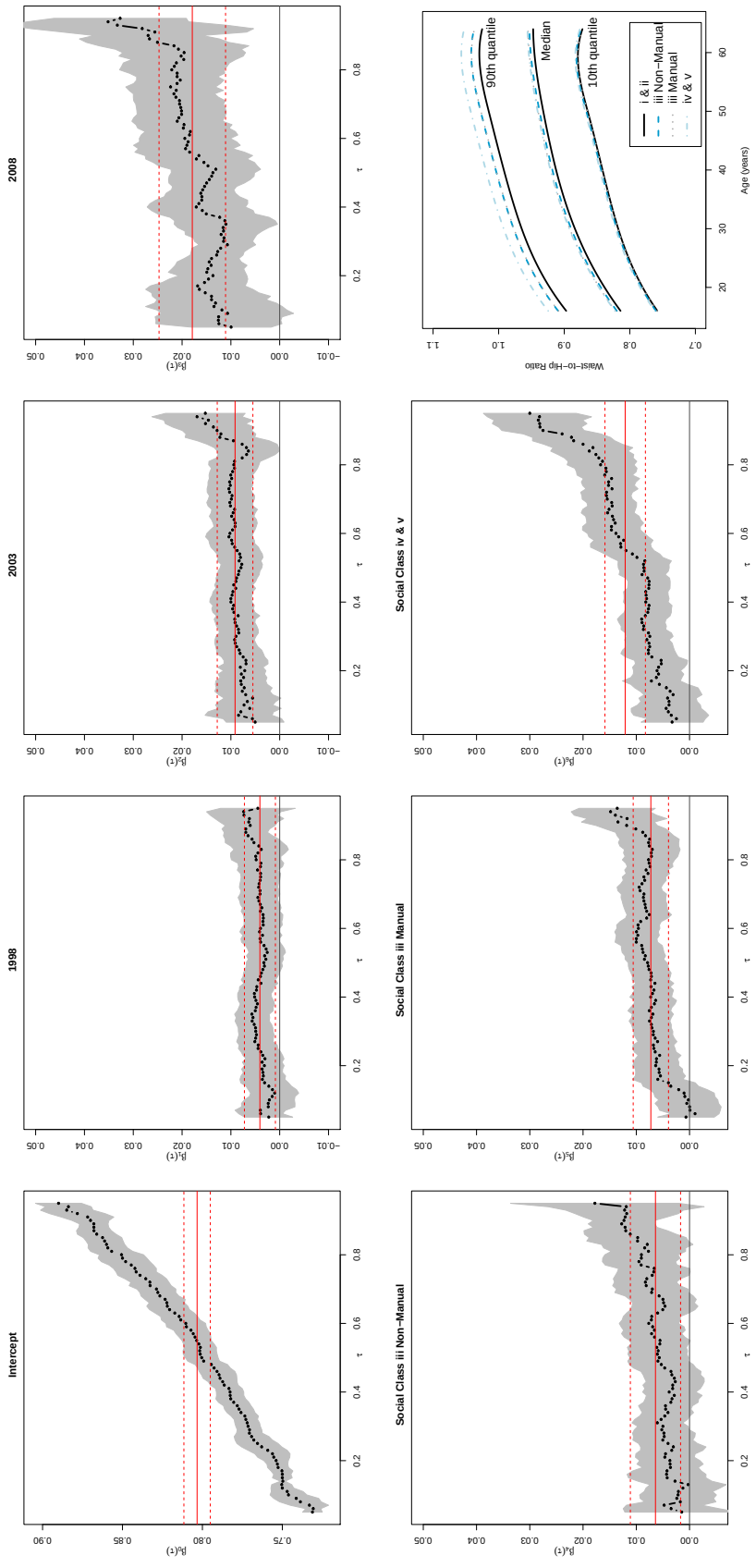

Figure 6.5: Survey year and social class coefficients with 95% confidence bands for values of τ between 0.05 and 0.95 for males. The solid red lines indicate the value of the corresponding least squares coefficient, while the dashed red lines are 95% confidence bands for this least squares estimate. The final plot displays the 10^{th}, 50^{th} and 90^{th} fitted waist-to-hip ratio quantiles as a function of age for survey year 2008.

Figure 6.6 shows the quantile regression plots obtained for females with waist-to-hip ratio as the response. In comparison to males, the females see a significant rise in waist-to-hip ratio across the whole distribution between survey years 1998 and 2003, and also a less uniform pattern in survey years 2003 and 2008, with the rise in WHR increasing with τ. Much greater separation is observed between social classes in comparison to males with the increase in WHR increasing with τ and with each subsequent social class (upward shift on the y-axis). Also, as was observed for the BMI, the relationship between WHR and age is more linear for females, with WHR increasing steadily with age.

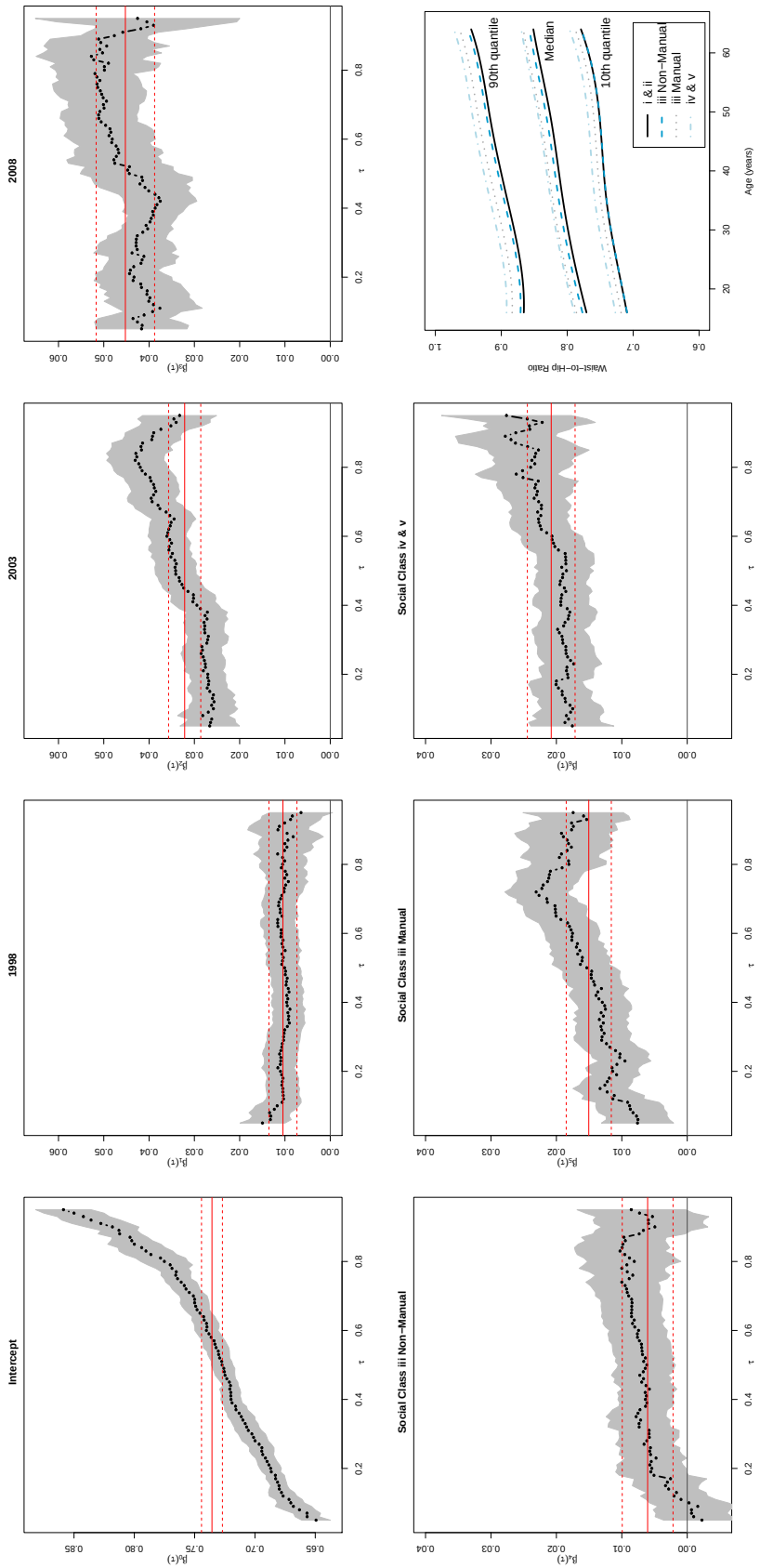

Figure 6.6: Survey year and social class coefficients with 95% confidence bands for values of τ between 0.05 and 0.95 for females. The solid red lines indicate the value of the corresponding least squares coefficient, while the dashed red lines are 95% confidence bands for this least squares estimate. The final plot displays the 10^{th}, 50^{th} and 90^{th} fitted waist-to-hip ratio quantiles as a function of age for survey year 2008.

6.4 Summary of Waist-to-Hip Ratio Results

Waist-to-hip ratio was found to increase with each survey year, with the increase fairly consistent across the entire distribution for males, while a greater increase was observed at the upper tail for females. A significant increase in WHR between survey years 1998 and 2003 was also found for females. Males in social classes iii manual and iv & v were found to have higher WHR measurements than males in social classes i & ii, with the separation between social classes greater at the upper end of the distribution. Greater disparity was observed between social classes for females, with WHR increasing with each subsequent social class, and also with increasing τ. WHR was also found to increase with age, with subtle gender differences. For males, the observed rate of increase was at its quickest in young males (between the ages of 16 and 30) before gradually slowing down, while a positive linear relationship was observed for females.

While the WHR results obtained are somewhat similar to those found for the BMI, there were a few noticeable differences. In contrast to the relatively uniform pattern observed for each survey year for the WHR, considerably more disparity was observed for log(BMI), with greater increase observed at the upper end of the distribution in comparison to the lower end for each survey year. Contrasting results were also observed with male socio-economic status. Unlike with the WHR, a change in sign across the log(BMI) distribution was observed with males lying at the lower end of the distribution in social classes iii non-manual, iii manual and iv & v having lower log(BMI) values than males in social classes i & ii, whilst males at the upper end of the distribution were found to have higher log(BMI) values. However, as the results for each classifier cannot be directly compared due to differing sample sizes (as shown in Table 6.2), it is not easy to determine whether such differences between the BMI and WHR actually exist.

Chapter 7

Discussions and Conclusions

7.1 Summary

Data from the Scottish Health Surveys carried out in 1995, 1998, 2003 and 2008 illustrated the ever-increasing prevalence of overweight/obese individuals in Scotland. Linear and generalized additive models for logistic regression highlighted associations between obesity prevalence and age, socio-economic status and survey year, with subtle gender differences. The odds of being overweight or obese were found to increase with age, plateauing for males over the age of 55, but continuing to increase for females in a relatively linear fashion. Variations in obesity prevalence in association with socio-economic status were observed by gender, with males in social class iii non-manual accruing the highest odds of being obese, whilst females in social classes iv & v were the most likely to be obese. Increasing mean BMI with each Scottish Health Survey was also found to follow similar patterns to those already mentioned for age, socio-economic status and gender. Furthermore, the observed increase in BMI and WHR with age, socio-economic status and survey year was seen across the entire distributions, but was larger at the upper end of the distributions (corresponding to overweight/obese individuals).

7.2 Study Limitations

The data acquired from the Scottish Health Surveys is cross-sectional, meaning that the same individuals were not followed up at each time point (or survey year), which can provide potential complications when interpreting the data. For example, the relationship between obesity prevalence, and also the BMI with age might not necessarily be as apparent as they seem when taking into account the nature of the data collection, with potential generational effects needing to be taken into consideration. However, if the same individuals were followed up over time, as is the case with longitudinal data, then we would be able to see how BMI changes as a person ages, but the data as it is can still provide us with information on the state of obesity in Scotland from 1995 to 2008. Also, if the data from the Scottish Heath Survey was gathered over more consistent and regular time intervals then, not only would there be more data available, but a more robust interpretation of the prevalence of obesity over time could be made. Other potential explanatory variables that were not examined here include smoking status, where an inverse association between smoking and BMI was observed by Shankar (2010), and also psychological well-being and self-esteem, as examined by Sweeting et al. (2008).

7.3 Future Work

Further research into associations between the Body Mass Index (BMI) and waist-to-hip ratio (WHR) could be performed using conditional bivariate quantile contours (Wei (2008)). In using quantile contours, the BMI and WHR could be directly compared for specific quantiles across the distributions, allowing for much closer examination of potential relationships and/or disparities in covariates between the BMI and WHR. Knot selection when using splines for age as continuous variable could also be examined more closely as currently there is no logical number or placements of knots, with

specific quantiles or evenly spaced knots having been considered. Also, unlike the logistic regression and generalised additive models there is not a formal goodness-of-fit test for quantile regression models, but they still provide us with detailed information on the cross-quantile variation in BMI/WHR with age, survey year and social class.

7.4 Conclusion

With obesity prevalence becoming an ever-increasing concern to health organisations worldwide, new and improved methods of analysing indicators of obesity would be beneficial in attempting to reduce the problem. Using quantile regression to model obesity, instead of standard linear and logistic methods, can help in aiding understanding of obesity, as it allows for thorough examination of the entire distribution of the Body Mass Index (BMI), or any other chosen obesity indicator for that matter. What this means is that it allows for potential fluctuations and cross-qantile variation in the BMI distribution to be captured, which simply cannot be ascertained through the use of conventional least squares regression. For example, this allows for us to not only examine whether the BMI of overweight/obese individuals is increasing, but to check if the BMI is generally increasing for the entire population, and then quantify any potential cross-quantile variation. This then provides us with information on individuals who are at greatest risk of either becoming obese, or from suffering more serious health related problems associated with obesity, to be the main focus of intervention campaigns introduced by health organisations in attempts to tackle obesity prevalence, and has also been noted by Shankar (2010).

Appendix A

Estimated Coefficients for the Logistic Regression Models

Table A.1: Logistic regression model for the prevalence of obesity in males.

	Estimate	Std. Error	P-value	Odds
(Intercept)	-2.919	0.128	0.000	0.05
25-34	0.933	0.130	0.000	2.54
35-44	1.296	0.125	0.000	3.65
45-54	1.587	0.125	0.000	4.89
55-64	1.663	0.124	0.00	5.27
iii nm	0.231	0.086	0.007	1.26
iii m	0.110	0.061	0.071	1.12
iv & v	0.102	0.069	0.143	1.12
1998	0.133	0.067	0.049	1.14
2003	0.331	0.070	0.000	1.39
2008	0.550	0.073	0.000	1.73

Table A.2: Logistic regression model for the prevalence of obesity in females.

	Estimate	Std. Error	P-value	Odds
(Intercept)	-2.441	0.097	0.000	0.087
25-34	0.416	0.096	0.000	1.52
35-44	0.607	0.093	0.000	1.83
45-54	0.832	0.092	0.000	2.30
55-64	1.125	0.091	0.000	3.08
iii nm	0.201	0.067	0.003	1.22
iii m	0.440	0.058	0.000	1.55
iv & v	0.466	0.060	0.000	1.59
1998	0.232	0.059	0.000	1.26
2003	0.320	0.062	0.000	1.38
2008	0.577	0.064	0.000	1.78

Table A.3: Logistic regression model for the prevalence of overweight and/or obesity in males.

	Estimate	Std. Error	P-value	Odds
(Intercept)	-1.037	0.079	0.000	0.35
25-34	1.123	0.078	0.000	3.07
35-44	1.566	0.077	0.000	4.79
45-54	1.884	0.080	0.000	6.58
55-64	2.069	0.082	0.000	7.92
iii nm	0.128	0.076	0.091	1.14
iii m	0.043	0.054	0.419	1.04
iv & v	-0.1223	0.060	0.041	0.88
1998	0.133	0.055	0.016	1.14
2003	0.315	0.061	0.000	1.37
2008	0.4411	0.067	0.000	1.55

Table A.4: Logistic regression model for the prevalence of overweight and/or obesity in females.

	Estimate	Std. Error	P-value	Odds
(Intercept)	-1.064	0.071	0.000	0.34
25-34	0.395	0.070	0.000	1.48
35-44	0.751	0.068	0.000	2.12
45-54	1.109	0.070	0.000	3.03
55-64	1.554	0.072	0.000	4.73
iii nm	0.243	0.055	0.000	1.27
iii m	0.455	0.049	0.000	1.58
iv & v	0.334	0.051	0.000	1.40
1998	0.157	0.048	0.001	1.17
2003	0.320	0.053	0.000	1.38
2008	0.539	0.057	0.000	1.71

Bibliography

Bates, D. M. & Venables, B. W. N. (2009), *Splines*. R package version 2.9.1. **URL:** *http://CRAN.R-project.org*

Chambers, J. A. & Swanson, V. (2010), 'A Health Tool for Multiple Risk Factors for Obesity: Age and Sex Differences in the Predicition of Body Mass Index', *British Journal of Nutrition* **104**, 298–307.

de Boor, C., Lorentz, G. G., Chui, C. K. & Schumaker, L. L. (1986), 'Splines as Linear Combinations of B-splines. A Survey'.

Dobson, A. J. & Barnett, A. G. (2008), *An Introduction to Generalized Linear Models*, Chapman and Hall.

Eknoyan, G. (2008), 'Adolphe Quetelet (1796-1874) - The Average Man and Indices of Obesity', *Nephrology Dialysis Transplantation* **23**(1), 47–51.

Gesta, S., Tseng, Y. & Khan, C. R. (2007), 'Development Origin of Fat: Tracking Obesity to Its Source', *Cell* **131**, 242–256.

Joint Health Surveys Unit (2009), 'University College London and Medical Research Council. Social and Public Health Sciences Unit; Scottish Health Survey, 1995-2008 Colchester, Essex: UK Data Archive'.

Koenker, R. (2009), *quantreg: Quantile Regression*. R package version 4.43. **URL:** *http://CRAN.R-project.org/package=quantreg*

Koenker, R. & Bassett, G. (1978), 'Regression Quantiles', *Econometrica* **46**(1), 33–50.

le Cessie, S. & van Houwelingen, J. C. (1991), 'A Goodness-of-Fit Test for Binary Regression Models, Based On Smoothing Methods', *Biometrics* **47**, 1267–1282.

Li, Y., Graubard, B. I. & Korn, E. L. (2010), 'Application of Nonparametric Quantile Regression to Body Mass Index Percentile Curves from Survey Data', *Statistics in Medicine* **29**, 558–572.

Mills, T. (2008), 'Forecasting Obesity Trends in England', *J. R. Statist. Soc. A* **171**, 107–117.

Penman, A. D. & Johnson, W. D. (2006), 'The Changing Shape of the Body Mass Index Distribution Curve in the Population: Implications for Public Health Policy to Reduce the Prevalence of Adult Obesity', *Preventing Chronic Disease* **3**(3).

Prentice, A. M. & Jebb, S. A. (2001), 'Beyond Body Mass Index', *Obesity Reviews* **2**, 141–147.

R Development Core Team (2009), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: *http://www.R-project.org*

Reas, D. L., Nygård, J. F., Svensson, E., Sørensen, T. & Sandanger, I. (2007), 'Changes In Body Mass Index by Age, Gender, and Socio-Economic Status Among a Cohort of Norwegian Men and Women (1990-2001)', *BMC Public Health* **7**, 269.

Rose, G. (1992), *The Strategy of Preventive Medicine*, New York (NY): Oxford University Press.

Shankar, B. (2010), 'Obesity in China: The Differential Impacts of Covariates Along the BMI Distribution', *Obesity* **18**, 1660–1666.

Srikanthan, P., Seeman, T. & Karlamangla, A. S. (2009), 'Waist-Hip-Ratio as a Predictor of All-Cause Mortality in High-Functioning Older Adults', *Annals of Epidemiology* **19**, 724–731.

Stein, C. J. & Colditz, G. A. (2004), 'The Epidemic of Obesity', *The Journal of Clinical Endocrinology & Metabolism* **89**, 2522–2525.

Sweeting, H., West, P. & Young, R. (2008), 'Obesity Among Scottish 15 Year Olds 1987-2006: Prevalence and Associations With Socio-Economic Status, Well-Being and Worries About Weight', *BMC Public Health* **8**, 404.

The National Statistics Socio-economic Classification (2010). Accessed 14 July 2010.
URL: *http://www.ons.gov.uk/about-statistics/classifications/current/ns-sec/index.html*

Wang, Y., Beydoun, M. A., Liang, L., Caballero, B. & Kumanyika, S. K. (2008), 'Will All Americans Become Overweight or Obese? Estimating the Progression and Cost of the US Obesity Epidemic', *Obesity* **16**(10), 2323–2330.

Wei, Y. (2008), 'An Approach to Multivariate Covariate-Dependent Quantile Contours with Application to Bivariate Conditional Growth Charts', *Journal of the American Statistical Association* **103**, 397–409.

WHO expert consultation (2004), 'Appropriate Body-Mass Index for Asian Populations and its Implications for Policy and Intervention Strategies', *The Lancet* **363**, 157–163.

Wood, S. N. (2006), *Generalized Additive Models: An Introduction With R*, Chapman and Hall.

Yusuf, S., Hawken, S. & Ôunpuu, S. (2005), 'Is Waist-To-Hip Ratio A Better Marker of Cardiovascular Risk Than Body Mass Index?', *Lancet* **366(9497)**, 1640–9.

Zaninotto, P., Head, J., Stamatakis, E., Wardle, H. & Mindell, J. (2009), 'Trends in Obesity Among Adults in England From 1993 to 2004 by Age and Social Class and Projections of Prevalence to 2012', *J Epidemiol Community Health* **63**, 140–146.