Environmental Exposures and Cardiovascular Morbidity in Scotland: A Study of the Effects of Air Pollution on Health

Abita Bhaskar

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

Department of Statistics

February 2009

Acknowledgements

First and foremost, I would like to thank my supervisors Prof. Marian Scott and Dr. Duncan Lee for their help and support throughout this research.

I would also like to thank the Information Services Division of the NHS for supplying me with the funding that allowed me to carry out this research and for providing me with the data used in this thesis. A special thank you goes to Kevin Greig and Carole Morris from ISD for helping me and answering any questions I had about the data.

Thank you to the girls in the office for helping me to relax when things weren't going well for me.

Finally a big thank you goes to my partner Dougie for letting me bounce ideas off him, proof reading my thesis, reminding me to eat and most of all, for helping me to stay sane!

Abstract

Air pollution has been an ongoing problem around the world for centuries. It was brought to the public's attention in the mid 1900s with the London Smog which resulted in approximately 3000 excess deaths. Since then, there have been numerous studies carried out to determine the extent to which air pollution is related to human health. There are two main aims to this thesis, the first of which is to investigate the effects of PM_{10} exposure on cardiovascular illness in Scotland, focussing on the three largest cities, Glasgow, Edinburgh and Aberdeen. As this study makes use of readily available data, the second aim is to determine whether or not such data can be used to accurately estimate the effects of air pollution.

Chapter 1 provides a detailed discussion of air pollution, focussing on the history of air pollution and the change in pollutants over time, and cardiovascular illnesses, giving a definition of cardiovascular disease, details of how they occur and giving incidence rates in Scotland. This chapter also gives an overview of the Information Services Division of the NHS (ISD), the Scottish Air Quality Website and the British Atmospheric Data Centre (BADC).

Chapter 2 is a review of the relevant literature covering the standard modelling approach used in air pollution and health studies and will also outline the data used in these studies and the covariates involved. This chapter focusses exclusively on the short-term effects of air pollution as this is the focus of this thesis.

Chapter 3 uses Poisson generalised linear models to explore the relationship between exposure to air pollution and cardiovascular admissions to hospital in Scotland, focusing specifically on Glasgow, Edinburgh and Aberdeen.

Chapter 4 comprises a set of subanalyses of these data focusing on the effects of air pollution on various subclasses of cardiovascular morbidity in Glasgow. All analyses will be implemented using a generalised linear model, within the statistical programming language R (*R 2.2.0 - A Language and Environment* (2005)).

Chapter 5 provides a summary of the results from the analyses. It also discusses the limitations associated with the use of routinely collected data and describes some of the dilemmas faced by researchers in this field.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Air pollution has been an ongoing problem around the world for centuries and was brought to the public's attention in 1952, when abnormally high concentrations resulted in a dense smog in London that lasted almost a week (5th - 9th December). During that time, concentrations of particulate matter were between 5 and 19 times higher than current standards recommend, while sulpher dioxide was between 12 and 23 times higher. A consequence of these abnormally high pollution concentrations was a sudden increase in the number of deaths and hospital admissions due to respiratory and cardiovascular diseases. In the weeks leading up to the smog the average mortality rate in London was about 1570 deaths per week, while the average rate of hospital admissions numbered approximately 750. However for the week ending 13th December there were approximately 3000 extra deaths, while hospital admissions increased by 48%.

In addition to the London smog, there were two other incidences that highlighted the need for investigation into the health effects of air pollution exposure. The first was in the Meuse Valley in Belgium in 1930, where there was a lethal smog that lasted from the 1st to the 5th December, comprising a mixture of pollutants including sulphur dioxide, sulphuric acid mists and fluoride gases. This resulted in an increased level of ill health during the final two days of the smog, as the mortality rate was over 10 times the norm. The second episode was in Donora,

Pennsylvania, in October 1948. On the 30th and 31st October atmospheric conditions in the town were so bad that 19 people died within a 24 hour period, which was 6 times the normal death rate. In addition, approximately 500 people became ill with symptoms of respiratory diseases. Prior to these episodes there were no major studies linking mortality and morbidity with air pollution, and the health impacts observed have resulted in a wealth of epidemiological studies and government legislation.

One of the most immediate consequences was the UK Clean Air Act, which was passed in 1956. It aimed to control sources of smoke pollution by introducing smokefree zones, in which only smokeless fuels could be burnt. It also gave local authorities the power to control emissions of dark smoke from industrial sites. It was very successful, and UK black smoke concentrations fell from over 200 $\mu g/m^3$ in the 1950s to 20 $\mu g/m^3$ by 1980. A similar trend has also been observed in other countries. However, a number of studies from America in the late 1980s reported that even this comparatively low concentration of pollution may have a substantial public health impact.

Since the 1950s the pollutant mix in the air has slowly evolved, with black smoke, being the main form of air pollution. Although it is still prevalent (more commonly measured as particulate matter), there are additional pollutants such as sulphur dioxide, carbon monoxide, ozone and nitrous oxides that are also common. These pollutants can be classified into two types, primary pollutants and secondary pollutants, the former of which is emitted directly from a source, such as sulphur dioxide or nitrous oxides, while a secondary pollutant is formed when primary pollutants react in the atmosphere. An example of the latter is ozone, which is formed when UV light reacts with nitrous oxides and hydrocarbons emitted from vehicles.

Although studies have shown that all the pollutants mentioned have been associated with adverse health, the effect of particles has recently become a major issue. Particulate matter (PM) is created from a number of sources, including both primary particles, such as particles from engine emissions, quarrying and construction, and secondary particles formed from emissions of ammonia, sulphur dioxide and oxides of nitrogen. However these particles vary widely in size, with PM larger than $10\mu m$ in diameter generally not passing into the lungs. Therefore for the purposes of health assessment it has become common to measure particles with a diameter of $10\mu m$ or less, which is denoted by PM_{10}.

The relationship between both long and short term exposure to PM_{10} and respiratory mortality has long been recognised, with the World Health Organisation stating that reducing PM_{10} concentrations in polluted cities could decrease deaths by up to 15% (World Health Organisation (2006)). Although the effects on respiratory morbidity, such as admissions to hospital due to asthma, have not been as well researched as those for mortality, there is compelling evidence to suggest that both short term and long term exposures to PM_{10} have detrimental effects. At present, a threshold level of PM_{10}, below which no adverse effects on health are observed, has not been established, although current legislation in the UK aims to reduce the annual mean concentration to $40\mu g/m^3$ by 31st December 2010, while the daily mean should not exceed $50\mu g/m^3$ more than 35 times a year. In Scotland the background concentrations of pollution are generally lower, and the objective is to reduce the annual mean concentration to $18\mu g/m^3$, while the daily mean should not exceed $50\mu g/m^3$ more than 7 times in a year.

In contrast with respiratory disease, the effects of air pollution exposure on cardiovascular diseases are not well understood. Cardiovascular diseases refer to all diseases of the heart and blood vessels including stroke. It is the number one cause of death globally and accounts for approximately 40% of all deaths in the

UK. Cardiovascular diseases generally occur in middle to old age, with comparatively few incidences in young people. The most common cardiovascular disease in the UK is Coronary Artery Disease (CAD), which is also the most frequent single cause of death (COMEAP (2006)). It is caused by a build-up of plaque in the arteries, which is a mixture of fat, cholesterol, calcium, and other substances found in the blood. This build-up narrows the arteries and reduces the flow of blood to your heart. This can lead to either a complete blockage of the artery (heart attack) or a rupture of the plaque, the latter of which leaves an ulcer on the artery on which a blood clot can form. Either of these events can cause the supply of blood to the heart to be seriously reduced, which in turn causes the heart to stop functioning, thus leading to the death of the individual.

In the UK, cerebrovascular disease (stroke) is the second most common cause of cardiovascular death, and the third most common cause of all deaths (COMEAP (2006)). There are two main causes of stroke, the first of which occurs when there is a blockage in the artery that carries blood to the brain. This may be caused by a blood clot in the main artery leading to the brain, a blood clot elsewhere in the body which is then carried in the bloodstream to the brain, or a blockage in the blood vessels deep within the brain. The second cause of stroke occurs when a blood vessel bursts and causes bleeding in the brain, which can happen either in the brain or on the surface of the brain.

Up to 50% of deaths from cardiovascular diseases occur outside hospital, with a significant proportion of these patients having no previous history of heart disease. This proportion is higher in the younger age groups and as age increases, the proportion of cardiovascular deaths in people who have previous history of such problems increases. However determining cause of death for these diseases can be problematic, especially in older people. The coding rules for death certificates assign one primary cause of death in each case, and while this is more likely to be correct in the younger age groups, older people may be suffering from more

than one illness, and hence many cases of cardiovascular illnesses go unrecorded.

In this thesis we present a study that investigates the effects of exposure to PM_{10} on cardiovascular morbidity in Scotland, focusing on the three largest cities - Glasgow, Edinburgh and Aberdeen. The study makes use of routinely available data that are collected for other purposes, and one aim of this work is to determine whether such data are adequate for this purpose

1.1 Data Sources

The data used in this study are routinely available from various government bodies and briefly summarised here.

1.1.1 Health Data

The health data available for this study were provided by the Information Services Division (ISD) of the NHS in Scotland. They collect data on a variety of health related topics including cardiovascular diseases, mental health and health services costs. This data is used by ISD to advise a number of organisations including local authorities, hospitals and general practitioners, on how to improve current standards. Many of the statistics are publically available on ISD's website at www.isdscotland.org, and include mortality and incidence rates, survival rates and waiting times.

An advisory group, comprising government officials and civil servants (known as The Steering Group), was set up to identify priority areas of the NHS, and as a result numerous programmes and projects have been implemented. This M.Sc thesis is part of the 'Linking Environment and Health Data' programme, and as a result ISD were able to provide the necessary health data. This came in the form of daily counts of health events for a given population and covers the seven year period from 2000 to 2006. Although the greater part of this thesis focuses

on cardiovascular morbidity, cardiovascular mortality data and asthma data are also available. The cardiovascular morbidity data consist of emergency admissions to hospital, and are split up by age, sex, main diagnosis (CHD or stroke) and whether or not it was a first incidence. The mortality data are split up by age, sex, main cause of death and location of death (in or out of hospital). The asthma data relate to daily counts of emergency admissions to hospital. These sub-classifications of the data are analysed separately in Chapter 4.

1.1.2 Pollution Data

The pollution data used in this thesis were obtained from the Scottish Air Quality website, www.scottishairquality.co.uk, which is funded by the Scottish Government and set up to provide users with all possible information relating to pollution in Scotland. The site describes various pollutants in Scotland, including fine particles (PM_{10}, $PM_{2.5}$ and PM_1), sulphur dioxide and ozone. These descriptions include how the pollutant is produced as well as its health impacts. The website also gives a range of other details related to pollution, such as the impact on the environment, the various methods of monitoring pollution, the set of locations pollution is currently monitored at and the air quality standard for each pollutant.

This study focuses on the effects of PM_{10}, as particles larger than this cannot easily pass into the lungs and so are not considered to be as detrimental to human health. The effects of $PM_{2.5}$ are also becoming a major public health issue, as it is thought that particles of this size can penetrate further into the lungs. However $PM_{2.5}$ cannot be used for this study as there is not enough data available in Scotland. The pollution data used in this study spans seven years, from 2000 to 2006, and comprise 13 sites in total. In Glasgow there were six sites with PM_{10} data, Glasgow centre, kerbside at Hope Street, Anderston, Byres Road, Battlefield Road and Waulkmill Glen Reservoir, the last of which is a rural site and so is not included in the analysis. Edinburgh has three sites that monitored PM_{10} concentrations, Edinburgh Centre, St. Leonards and Roseburn. Finally,

Aberdeen has data from four sites, Centre, Anderson Drive, Market Street and Union Street. In Edinburgh there are approximately six months worth of missing data between 2002 and 2003 and a further two months missing later in 2003. Aberdeen has approximately two months of missing data in 2002, while Glasgow has no missing data. For each city, average daily PM_{10} concentrations were calculated by taking the average over the sites, which is a standard approach in the air pollution and health literature (Lee et al. (2006)).

1.1.3 Temperature Data

Temperature data were available for this study, and were obtained from the British Atmospheric Data Centre (BADC). The BADC helps researchers to acquire and interpret atmospheric data, which include meteorological variables such as temperature, barometric pressure and wind speed. However other atmospheric conditions, such as pollution concentrations, are also available. While many of the data sets are freely available, others, including those from the met office, are restricted to authorised personnel. However as this study is for academic and not profit making purposes, we obtained the required meteorology data.

The data take the form of daily minimum and maximum temperatures and spans the period 2000 to 2006. These temperatures were available at three sites in Edinburgh, Blackford Hill, Botanic Gardens and Gogarbank, two sites in Glasgow, Pollock Country Park and Bishopton and only one site in Aberdeen, at the Mannofield Reservoir. In Edinburgh and Glasgow, an average maximum temperature was found by taking a mean of the maximum temperature at each site, while an average minimum temperature was found using an analogous method. These spatial averages contain no missing values during the time period, whereas in Aberdeen, there was no data available for the last two months of the study, as well as a few missing values in 2005.

1.2 Thesis Outline and Aims

The remainder of this thesis will be split into four chapters. The first of these is a review of the relevant literature covering the standard modelling approach used in air pollution and health studies and will also outline the covariates involved in these studies. Chapter 3 is a time series study of the effects of air pollution exposure on cardiovascular morbidity in Scotland, focusing on Glasgow, Edinburgh and Aberdeen. Due to the lack of pollution or temperature data for Dundee it is not possible to estimate the association between PM_{10} and cardiovascular health for this city. Chapter 4 comprises a set of subanalyses of these data focusing on the effects of air pollution on various subclasses of cardiovascular morbidity in Glasgow. All analyses will be implemented using a generalised linear model, within the statistical programming language R (*R 2.2.0 - A Language and Environment* (2005)). Where necessary, the additional 'Splines' package (*'Splines' 2.2.0* (2005)) will be used. Finally, chapter 5 will contain a concluding discussion.

The study described in this thesis has two main aims, the first of which is to investigate the effects of PM_{10} exposure on cardiovascular illness in Scotland, focusing on the three largest cities, Glasgow, Edinburgh and Aberdeen. As previously mentioned, this study makes use of readily available data, and the second aim is to determine whether or not such data can be used to accurately estimate the effects of pollution.

Chapter 2

Literature Review

There is a wealth of statistical, environmental and public health literature that describes the potential health effects of air pollution exposure, focusing on both the long and short term effects. This chapter critiques some of the more recent literature in this area, focusing exclusively on the short term effects because that is the main aspect of this thesis. While the relationship between air pollution and respiratory mortality and morbidity has been well researched, the effects on cardiovascular illnesses are relatively unknown and are therefore the focus of this thesis.

This chapter provides a critical review of short-term air pollution and health studies, focusing on various aspects of both the data and modelling. Section 2.1 discusses the standard modelling approach used in these studies, while section 2.2 provides an outline of the data available for these studies. Sections 2.3 and 2.4 focus on how the pollution and covariate data are incorporated into the model, while the final three sections discuss mortality displacement, multi-city studies and meta analyses, and the ecological fallacy (Sections 2.5 to 2.7).

2.1 Overall Modelling Approach

The modelling techniques for estimating the short-term association between pollution exposure and ill health have been developed mainly over the last couple of decades. Letting Y_t denote the number of events on day t, $\mathbf{z}_t$ denote a vector of p covariates and $\boldsymbol{\alpha}$ denote the associated regression parameters, early studies such as Schwartz & Marcus (1990) used Normal linear regression models, which have the general form

$$\begin{aligned} Y_t &\sim N(\mu_t, \sigma^2) \quad \text{for} \quad t = 1, ..., n \\ \mu_t &= \mathbf{z}_t^T \boldsymbol{\alpha} + x_{t-q}\gamma \end{aligned} \tag{2.1}$$

where x_{t-q} are the pollution concentrations and γ is the estimated effect of pollution. This was used because it was easy to implement. However this model is not appropriate because it assumes the count data come from a Normal distribution whereas a Poisson assumption is more plausible. Therefore a generalised linear model is more appropriate, and extends the linear model by replacing the Normal assumption with a probability distribution from an exponential family. In addition, it relates the response variable to the explanatory variables through a non-linear function called the link function, $g(\mu)$, thus giving the model the form

$$\begin{aligned} Y &\sim f(y_t|\mu_t) \\ g(\mu_t) &= \mathbf{z}_t^T \boldsymbol{\alpha} + x_{t-q}\gamma \end{aligned} \tag{2.2}$$

where f is a probability density function of an exponential distribution, and g is a link function. This link function differs according to the distribution the response follows, with the Binomial distribution having a Logit function while for Poisson data, the link function is Log.

In most short-term air pollution and health time series studies, the mortality

or morbidity data are only available in the form of daily counts $Y = (y_1, ..., y_n)$ for the population rather than individual level outcomes. This means that the standard modelling approach is based on Poisson generalised linear models, with the random variable Y_t denoting the number of deaths or admissions on day t, while x_{t-q} are the pollution concentrations. A general model is given by

$$\begin{aligned} Y_t &\sim \text{Poisson}(\mu_t) \quad \text{for} \quad t = 1, ..., n \\ \ln(\mu_t) &= \mathbf{z}_t^T \boldsymbol{\alpha} + x_{t-q}\gamma \end{aligned} \tag{2.3}$$

where $\mathbf{z}_t = (z_{t1}, ..., z_{tp})$ are a vector of p covariates for day t and $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_p)$ are the associated regression parameters. The pollution exposure is lagged by q days and the parameter of primary interest in this model is γ, which is the estimated effect of pollution. However results are often presented in the form of a relative risk for a $10\mu g/m^3$ increase in the pollutant of interest. The relative risk is the ratio of the expected number of hospital admissions given the current pollution concentrations divided by the expected numbers if the concentrations rose by $10\mu\text{g}/m^3$ and is calculated as shown below:

$$\begin{aligned} RR &= \frac{\mathbb{E}[\text{Number of deaths if } x \text{ increased by } 10\mu\text{g}/m^3]}{\mathbb{E}[\text{Number of deaths}]} \\ &= \frac{\exp(\mathbf{z}_t^T\hat{\boldsymbol{\alpha}} + (x+10)\hat{\gamma})}{\exp(\mathbf{z}_t^T\hat{\boldsymbol{\alpha}} + x\hat{\gamma})} \\ &= \frac{\exp(\mathbf{z}_t^T\hat{\boldsymbol{\alpha}} + x\hat{\gamma} + 10\hat{\gamma})}{\exp(\mathbf{z}_t^T\hat{\boldsymbol{\alpha}} + x\hat{\gamma})} \\ &= \frac{\exp(\mathbf{z}_t^T\hat{\boldsymbol{\alpha}} + x\hat{\gamma})\exp(10\hat{\gamma})}{\exp(\mathbf{z}_t^T\hat{\boldsymbol{\alpha}} + x\hat{\gamma})} \\ &= \exp(10\hat{\gamma}) \end{aligned} \tag{2.4}$$

A relative risk of one implies that pollution has no effect on health, while a relative risk less than one implies a negative effect (i.e. pollution is protective) and a relative risk greater than one implies a positive effect (i.e. pollution is harmful). A 95% confidence interval for the relative risk is typically also calculated, to

determine whether or not the relative risk is significantly different from one. Its general form is given by

$$\exp(10\hat{\gamma} \pm 1.96 \times \text{Standard Error})$$

where the standard error is calculated during the model fitting stage (for details see Dobson (1991)). If the interval contains one, the relative risk is nonsignificant, while a confidence interval that is completely greater than or less than one implies the relative risk is significant.

However Model (2.3) may not be adequate because it makes a number of possibly unrealistic assumptions, one of which being that the response variable, in this case admissions to hospital, follows a Poisson distribution. This enforces the mean number of admissions (μ_t) to equal the variance, which is unlikely to be the case. In addition, the number of admissions are assumed to be independent for each day, which may be unlikely for time series data of this type. If the mean is less than the variance, this leads to a phenomenon known as over-dispersion, which can be corrected by multiplying the standard error by the square root of the over-dispersion parameter,

$$\phi = \frac{1}{n-p} \sum_{t=1}^{n} \frac{(Y_t - \mu_t)^2}{\mu_t} \tag{2.5}$$

within a quasi-likelihood rather than Poisson linear model (Chardon et al. (2007)). In this case the model is only specified by its first two moments, namely $\mathrm{E}(Y_t) = \mu_t$ and $\mathrm{Var}(Y_t) = \phi\mu_t$, where $\mu_t = \exp(\boldsymbol{z}_t^T \boldsymbol{\alpha} + x_{t-q}\gamma)$. If ϕ equals one, this implies there is no overdispersion and the poisson assumption is adequate. If ϕ is greater than one, there is overdispersion in the model, while if ϕ is less than one, there is underdispersion. The extension from a Poisson to a quasi-likelihood to incorporate over-dispersion has been used by many authors such as Carder et al. (2008) and Samoli et al. (2007). A less common alternative to quasi-likelihood is to model Y_t as a Negative Binomial random variable (See for example Gwynn et al.

(2000)), where $Y_t \sim NB(\mu_t, \phi)$ and has mean μ_t and variance $\mu_t + \frac{\mu_t^2}{\phi}$.

In time series studies such as this, admissions to hospital or deaths on successive days are likely to be correlated, which is due to unmeasured risk factors being similar for days close together in time. Most studies attempt to remove this correlation by the addition of covariates and a smooth time trend to the model, although this approach is not always adequate. The presence of correlation can be checked by examining the autocorrelation function of the standardised residuals

$$r_t = \frac{Y_t - E(Y_t)}{\sqrt{\text{Var}(Y_t)}} \tag{2.6}$$

with high values indicating correlation is present. If correlation does occur, there are two main approaches to removing it, parameter and observation driven models. Parameter driven models remove correlation by adding a correlated latent process to Model (2.3), and was proposed by Zeger (1988) and West et al. (1985). The latter was adapted by Chiogna & Gaetan (2002) and more recently by Lee & Shaddick (2008) in air pollution and health studies, and their models had the general form

$$\begin{aligned} \ln(\mu_t) &= \mathbf{z}_t^T \boldsymbol{\alpha} + x_{t-q}\gamma + \beta_t \\ \text{where} \quad \beta_t &= 2\beta_{t-1} - \beta_{t-2} + \epsilon_t \\ \epsilon_t &\sim N(0, \lambda) \end{aligned} \tag{2.7}$$

The second approach is the observation driven method, which uses past outcomes as additional covariates, and was proposed by Zeger & Qaqish (1988b). This approach was used in an air pollution context by Xu et al. (1995), who used a model with the general form

$$\begin{aligned} Y_t &\sim \text{Poisson}(\mu_t) \quad for \quad t = 1, ..., n \\ \ln(\mu_t) &= \mathbf{z}_t^T \alpha + x_{t-q}\gamma + \sum_j^p g_j(D_t)\psi_j \end{aligned} \tag{2.8}$$

where ψ_j are unknown parameters, g_j are known functions and D_t are the set of past response outcomes. This is the approach used in this study when the residuals displayed residual correlation, and we used a model of the form

$$\ln(\mu_t) = \mathbf{z}_t^T \alpha + \beta \mathrm{Y}_{t-1} + \gamma X_{t-q} \tag{2.9}$$

A common problem in regression modelling arises when there are two or more variables in the model that are highly correlated, as this can lead to a phenomenon known as collinearity, which makes it difficult or impossible to estimate the regression coefficients reliably (Dobson (1991)). In air pollution studies this often happens when pollution concentrations on more than one day are added to the model. The choice of lag to use is an outstanding one in air pollution research, and is described in section 2.3.2. The solution used in this study is via a distributed lag model (Zanobetti et al. (2000)) which has the general form

$$\begin{aligned} \ln(\mu_t) &= \mathbf{z}_t^T \alpha + \sum_{q=0}^{n} z_{t-q} \gamma_q \\ \text{where} \quad \gamma_q &= \sum_{k=0}^{q} \eta_k q^k, \quad q = 0, ..., n \end{aligned} \tag{2.10}$$

This model includes multiple lags of exposure in the model but constrains the coefficients to follow a polynomial of degree q in lag number, thus reducing the effects of collinearity. When the polynomial constraint is taken to the power zero, the relative risks at each lag are all equal, meaning that effectively only a single parameter is being estimated, which removes any collinearity problems. At the other extreme, setting the degree of the polynomial equal to the number of lags results in no constraint, meaning that the estimates are the same as those in the multiple lag model and again suffer from collinearity problems. Therefore we are looking for a power that minimises collinearity but ensures the pattern of the relative risks as the lag increases is biologically plausible. Further details of such plausible relative risk versus lag shapes are described in section 2.5 in relation to the mortality displacement hypothesis.

After a model has been constructed it must be checked to determine whether it adequately describes the variation in the data. This is achieved by examining the standardised residuals r_t, where, for the poisson model, $\mathrm{r}_t = \frac{Y_t - \mu_t}{\sqrt{\mu_t}}$. If the model is a good fit,

- $\mathrm{E}(\mathrm{r}_t) = 0$
- $\mathrm{Var}(\mathrm{r}_t) = 1$
- There should be no obvious trend
- The residuals should be uncorrelated

These are checked by creating diagnostic plots and summaries of the residuals. To check for any residual trend, plot the residuals against time, while a plot of (μ_t, r_t) will show up evidence of non-constant variance. For the residuals to be considered acceptable, they should be evenly scattered about zero and have a constant variance. Residual correlation is checked using the autocorrelation function as previously mentioned, which is a plot of lag k against $\mathrm{corr}(\mathrm{r}_t, \mathrm{r}_{t-k})$ for $\mathrm{k} \geq 0$. From this plot, independence can be assumed if correlation at each lag is within the given 95% confidence intervals.

When comparing two different models, it can often be difficult to decide which model is the better fit of the data. A useful method for deciding between the two is by using Akaike's Information Criterion (AIC). This is defined as being

$$\mathrm{AIC} = -2\mathrm{logLik} + 2k \tag{2.11}$$

where logLik is the log likelihood of the model and k is the number of parameters fitted in the model. The lower the value of the AIC, the better the fit of the model.

2.2 Data Description

Data that are typically available for time series studies such as the one presented in this thesis are measures of mortality and morbidity, ambient pollution concentrations and a number of covariates, all of which are described in detail in the following sections.

2.2.1 Health Data

Data about mortality or morbidity events are typically only available as daily counts, aggregated over the population living within the region of interest, such as a city. Ideally, studies would be carried out using individual level data, but this violates personal confidentiality, so such data are not available. In some studies, the number of daily deaths or admissions to hospital for the population are small (less than five), so they are aggregated to the weekly level (Bell & Davis (2001)), which is also the case for some of the analyses in Chapters 3 and 4 in this thesis. All health events are classified using the International Classification of Diseases (ICD) with either the 9th revision (1977-2000) or the 10th revision (2000-Present) being used, depending on the time period the data relate to. The data analysed in this study are classified using the 10th revision (ICD-10) as they relate to the period 2000 to 2006. A number of mortality and morbidity classifications have been used in air pollution and health studies and a brief review is given below.

All Cause

The most general classification is all cause mortality or morbidity (ICD-9 <800) which has been used by numerous authors including Schwartz (2004). In that study the effects of particulate air pollution on daily deaths were investigated and, for an increase of $10\mu g/m^3$ in PM_{10} concentrations, a relative risk of 1.0036, with a 95% confidence interval of 1.0022 to 1.005 was reported. However such a general classification also includes a large number of deaths that are not pollution exposure related meaning that any estimated association may be biased by the

potentially large proportion of deaths not related to air pollution. Subsequently, the majority of studies focus on cause specific mortality and morbidity.

Respiratory Studies

The most commonly investigated health endpoint is respiratory disease, for which consistent positive associations have been found. For example Tellez-Rojo et al. (2000) studied the effects of PM_{10} on respiratory mortality (ICD-9 466, 480-487, 490-496), both in and outwith hospital, in Mexico city. Outwith hospital it was found that for a $10\mu g/m^3$ increase in PM_{10} concentrations the relative risk of dying was 1.024 at lag one with a 95% confidence interval of 1.004 to 1.045. At lags of two, three, four and five days the relative risks were 1.027 (95% CI: 1.007, 1.047), 1.029 (95% CI: 1.009, 1.049), 1.026 (95% CI: 1.006, 1.045) and 1.02 (95% CI: 1.006, 1.04) respectively, showing consistent evidence of a positive significant association. In addition a cumulative five day exposure showed a relative risk of 1.042 (95% CI: 1.017, 1.068). However within a medical unit, it was found that effects of PM_{10} on respiratory disease were only significant after a five day lag (RR: 1.024, 95% CI: 1.005, 1.042). A five day accumulated effect of PM_{10} exposure was also found, with a relative risk of 1.025 (95% CI: 1.001, 1.049).

In addition, this study found significant associations with subclassifications, such as deaths from Chronic Obstructive Pulmonary Disease (COPD) with a relative risk outwith hospital of 1.03 at a lag of one day, with a 95% confidence interval of 1.01 to 1.059. Again a five day cumulative exposure to PM_{10} resulted in a significant increase in mortality, with a relative risk of 1.061 (95% CI: 1.024, 1.099). Within hospital, a significant effect was only found at lag five, with a relative risk of 1.033 and a 95% confidence interval of 1.005 to 1.061.

Of increasing interest is the effect of various pollutants on admissions to hospital. For example Atkinson et al. (2001) published results from the APHEA

project, focusing on the effect of particulate air pollution on respiratory admissions. They separated admissions into four groups: (i) Asthma (ICD-9 493) for ages 0-14 years; (ii) asthma for people aged 15-64 years; (iii) COPD and asthma (ICD-9 490-496) for people aged 65 and over; (iv) all respiratory disease admissions (ICD-9 460-519) again for people aged 65+. Particulate air pollution was separated into $PM_{10}/PM_{2.5}$ and black smoke. For asthma admissions aged 0-14 years, it was found that for a $10\mu g/m^3$ increase in PM_{10} the relative risk was 1.012, with a 95% confidence interval of 1.002 to 1.023. For black smoke the relative risk was very similar at 1.013 (95% CI: 1.003, 1.024). For asthma admissions aged 15-64, the effects of PM_{10} were similar to that for children, with a relative risk of 1.011 (95% CI: 1.003, 1.018). In addition PM_{10} had a significant effect on COPD and asthma admissions for the elderly, with a relative risk of 1.01 (95%: 1.004, 1.015). Finally, admissions for all respiratory diseases for those aged 65 and over showed a relative risk of 1.009, with a 95% CI of 1.006 to 1.013. Black smoke showed no statistically significant effects on asthma admissions for those aged 15-64, admissions for COPD and asthma for the elderly or admissions from all respiratory diseases for the elderly. Numerous other studies have been conducted, (Chardon et al. (2007), Lee et al. (2006)) most of which have found similar small associations between respiratory health and pollution exposure.

Cardiovascular Studies

While the effects of air pollution on respiratory mortality and morbidity are well known, the effects on cardiovascular illnesses are less well documented. Recently, Ballester et al. (2006) studied the effects of various air pollutants on cardiovascular hospital admissions in Spain, focusing on all cardiovascular diseases (ICD-9 390-459) as well as heart diseases (ICD-9 410-414, 427, 428). The pollution data available included particulate matter (black smoke, Total Suspended Particles and PM_{10}), sulphur dioxide, nitrogen dioxide, carbon monoxide and ozone. They found that any significant effects of pollution occurred at lag zero to one, except for ozone, whose effects were significant two to three days later. For all

cardiovascular diseases, the effect of PM_{10} resulted in a relative risk of 1.0091 (95% CI: 1.0035, 1.015) whilst for heart diseases, the relative risk was 1.016 (95% CI:1.0082, 1.023). Carbon monoxide had the greatest effect on both all cardiovascular diseases (RR: 1.021, 95% CI: 1.0065, 1.0348) and heart disease (RR: 1.0415, 95% CI: 1.0131, 1.0708), with sulphur dioxide having a relative risk of 1.0133 for all cardiovascular diseases (95% CI: 1.0021, 1.0246) and 1.0172 for heart diseases (95% CI: 1.005, 1.0295). Nitrogen dioxide and ozone both had a small but significant effect with relative risks of 1.0038 and 1.0069 respectively for all cardiovascular diseases and 1.0086 and 1.0066 for heart diseases.

Touloumi et al. (2005) examined the effects of PM_{10} on cardiovascular mortality (ICD-9 390-459) in seven European cities, as part of the APHEA-2 project. In particular, they investigated whether influenza epidemics have a confounding effect in air pollution studies, and found that for a $10\mu g/m^3$ increase in PM_{10} concentrations, the relative risk of death from a cardiovascular disease was 1.0085 (95% CI: 1.0053, 1.0118) with no adjustment for influenza. They then used several different methods of controlling for influenza outbreaks, which resulted in relative risks of between 1.0086 (95% CI:1.0053, 1.0119) and 1.0106 (95% CI:1.0074, 1.0139) showing that influenza had little confounding effect in their study.

Age Specific Studies

There have also been numerous studies that analyse health data relating to specific age groups, particularly those that are thought to be susceptible to air pollution exposure. Hertz-Picciotto et al. (2007) studied the effects of Polycyclic Aromatic Hydrocarbons (PAHs) and $PM_{2.5}$ on respiratory illnesses in children, focusing particularly on bronchitis (ICD-10 J20-21). They found that the relative risks differed for various age groups, with children under the age of two having a relative risk of 1.29 (95% CI: 1.07, 1.54), for an increase of 100ng/m^3 ($0.1\mu g/m^3$) of PAHs. The relative risks for a 25μg/m^3 increase in $PM_{2.5}$ concentrations was similar, at 1.3, with a 95% confidence interval of 1.08 to 1.58. For children aged

two to four and a half years, the effect of PAHs was especially strong, with a relative risk of 1.56 (95% CI: 1.22, 2.00).

A second sub-population that is often studied is the elderly. Stankovic et al. (2007) studied the effects of black smoke and sulphur dioxide on cardiovascular mortality (ICD-10 I00-I99) amongst the elderly in Nis, Serbia. They found that a $10\mu g/m^3$ increase in sulphur dioxide concentrations resulted in a relative risk of 1.025 in people aged 65 and over. However they also found that this result was not statistically significant (95% CI: 0.99, 1.06). Similarly, an increase in black smoke concentrations showed a relative risk of 1.013, with a 95% confidence interval of 0.996 to 1.031. However this study only used pollution exposure on the same day as the death (lag 0), so there may have been an effect from previous days pollution or an accumulated effect of pollution over several days that this study did not investigate. Also this study used two temperature variables in the model, mean temperature at lag zero and mean temperature over lags zero to three (an average of the temperature values from the same day and three days previous) which are likely to be highly correlated, and therefore to be affected by collinearity, which may bias the results.

Gender Specific Studies

There have also been a number of studies carried out to investigate the effects of air pollution on males and females separately. Luginaah et al. (2005) studied the effects of various pollutants on respiratory admissions (ICD-9 460-519) in Windsor, Ontario. They found that only Coefficient of Haze had a significant effect on females of all ages after two days (RR: 1.067, 95% CI: 1.004, 1.135), with all other pollutant being unimportant. Conversely they found that there was no effect of pollution on male health. Chen et al. (2005) studied the effects of particulate pollution on cardiovascular mortality (ICD-9 410-414) and also found different results for males and females. For females, a $10\mu g/m^3$ increase in $PM_{2.5}$, resulted in a relative risk of 1.42 (95% CI: 1.06, 1.90) while for PM_{10} it was 1.22

(95% CI: 1.01, 1.47). In contrast, they also found no effect for males.

2.2.2 Air Pollution Data

Air pollution is a complex mixture of different types of pollutants, many of which are known to be harmful to humans. They are measured by automatic monitoring equipment and other types of monitors, situated at various points across the UK and Ireland. In this thesis we exclusively use data from the automatic monitors, which are located in a number of local environments, each of which has a different environmental objective. Table 2.1 gives a list of all the possible sites and their descriptions.

The most common monitoring environments are Urban Background and Roadside, with the former being distanced from pollution sources and so are generally thought to be a good indicator of background pollution concentrations. Roadside sites, however, are typically within 5m of the road, and are generally used to assess worst case population exposure and evaluate the impacts of vehicle emissions. Conversely, monitoring sites in a rural area are placed as far as possible from roads and populated or industrial areas, and are used to analyse the impact pollution has on the ecosystem. The monitors take hourly measurements throughout the day, allowing various statistics to be calculated, such as daily maximum and minimum measurements, daily mean, eight hour mean and annual mean. Many pollutants are measured in this way, including particulate matter (PM), sulphur dioxide, carbon monoxide, nitrous oxides and ozone.

Particulate matter is a mixture of primary and secondary pollutants, the former of which consist of particles from engine emissions, quarrying and construction, while the latter are formed from emissions of ammonia, sulphur dioxide and nitrous oxides. Recently, PM_{10} (which consists of particles that are less than $10\mu m$ in diameter) and $PM_{2.5}$ (particles less than $2.5\mu m$ in diameter) have become of particular interest as it is thought they may be especially detrimental to health,

because their small size means they are able to travel further into the lungs than other pollutants.

Site Type	Description
Urban	Urban
Kerbside	A site sampling within 1m of the kerb of a busy road
Roadside	Between 1m of the kerbside of a busy road and the back of the pavement, typically within 5m of the road, but could be up to 15m.
Suburban	A location type situated in a residential area on the outskirts of a town or city
Urban Background	An urban location distanced from sources, therefore broadly representative of city-wide background conditions
Urban Centre	An urban location representative of typical population exposure in towns or city centres
Urban Industrial	An area where industrial sources make an important contribution to the total pollution burden
Intermediate	20-30m from the kerb of a busy road
Airport	Monitoring within the boundary of an airport perimeter
Other	Any special source-orientated or location category covering monitoring undertaken in relation to specific emission sources such as power stations, car parks or tunnels
Rural	An open countryside location, in an area of low population density distanced as far as possible from roads, populated and industrial areas

Table 2.1. Table of monitoring sites and their descriptions (taken from the Scottish Air Quality website)

One concern in interpreting air pollution studies is that the reported association may be confounded by other pollutants which may be highly correlated. To this effect, a number of studies first look at the relationship between the various pollutants to consider whether or not collinearity may be a problem. For example, Cakmak et al. (2007) found a positive correlation between sulphur dioxide, carbon monoxide and PM_{10}, while ozone was negatively correlated with sulphur dioxide and carbon monoxide. Ballester et al. (2006) found similar results with positive

correlations between black smoke, PM_{10}, sulphur dioxide, carbon monoxide and nitrogen dioxide while ozone had a negative correlation with black smoke and carbon monoxide.

2.2.3 Meteorology Data

Meteorological data, in particular temperature, has long been known to have a confounding effect on epidemiological air pollution studies, especially when the health endpoint is mortality, with a higher deathtoll in winter (Carder et al. (2005)) and in summer during heatwaves (Huynen et al. (2001)) . Various different temperature metrics have been used to model this phenomenon, with most studies using daily mean values (Ballester et al. (2006)) rather than daily minimum (Prescott et al. (1998)) or maximum (Ye et al. (2001)). More than one temperature variable is generally not used in the same analysis as they are likely to be very highly correlated which would lead to collinearity. Interestingly, Carder et al. (2008) used both 'high' and 'low' temperatures in their model, where they defined 'high' temperatures as above 11°C and 'low' temperatures below 11°C. This allowed them to replace the typical linear relationship with a double linear extension, which they adopted because temperature is typically U-shaped, with two separate linear relations over different parts of the temperature range. The value of 11°C was chosen as the best fitting value, which was based on the log likelihoods of various models. Other meteorological covariates that are frequently used in air pollution studies are humidity (Bogdanovic et al. (2006)), barometric pressure (Stankovic et al. (2007)) and dew point temperature (Bell et al. (2004)).

2.2.4 Covariate Data

In addition to meteorological data, there are various other factors which may confound the results in any air pollution time series analysis. It is of vital importance to remove the influence of any possible confounders because if not removed, the

resulting association between air pollution and mortality or morbidity may be biased. Typical covariates that have been used are trend or seasonal variables, such as functions of time (Lee et al. (2006)), an indicator variable for 'day of the week' (Roberts (2004a)) and an indicator variable for season (Bogdanovic et al. (2006)). Other factors that have been controlled for include the existence of influenza outbreaks (Goodman et al. (2004)) and the presence of holidays (Zanobetti et al. (2003)). There have also been studies carried out which focused on whether or not there were socioeconomic aspects to air pollution mortality and morbidity. In particular Villeneuve et al. (2003) have found that there is a greater effect of air pollution amongst the lower social classes.

2.3 Pollution Modelling

As mentioned in section 2.2.2, air pollution is a complex mixture of components, many of which have been proved to be harmful to humans. Thus, in specifying the air pollution component of the model a number of choices need to be made, three of which are discussed below.

2.3.1 Pollutant

The first choice is the pollutant to use as the exposure, and there have been numerous studies carried out on a range of pollutants, a few of which are outlined below.

Ozone

Bell et al. (2004) studied the effects of a 10 parts per billion (ppb) volume increase in ozone concentrations on short-term respiratory and cardiovascular mortality in the United States from 1987 to 2000, using single lag models and both constrained and unconstrained distributed lag models. All models showed a significant effect of exposure, with the constrained distributed lag models showing a relative risk of approximately 1.0052 (95% CI: 1.0027, 1.0077).

Sulphur Dioxide

As part of the APHEA-II study (Air Pollution and Health: A European Approach), Sunyer et al. (2003a) studied the effects of daily sulphur dioxide concentrations on admissions to hospital with cardiovascular diseases. Cardiovascular diseases were defined to be Ischemic Heart Disease (IHD) and stroke. Analyses were carried out for IHD and stroke individually and together, and also for the subpopulations over and under 65 years of age. They found that a $10\mu g/m^3$ increase in sulphur dioxide concentrations resulted in a relative risk of 1.007 for all cardiovascular admissions (95% CI: 1.003 to 1.011). For IHD admissions in the under 65s, the relative risk was similar at 1.006 (95% CI: 1.002 to 1.011) while in the over 65s, the effect of sulphur dioxide was much more obvious, with a relative risk of 1.012 (95% CI: 1.008 to 1.016). However they reported a non-significant relationship between admissions to hospital with stroke and exposure to sulphur dioxide.

Nitrous Oxides

The APHEA-II project also considered the effect of nitrogen dioxide on mortality as investigated by Samoli et al. (2006). The mortality data included daily counts of all-cause mortality (ICD-9 <800), cardiovascular mortality (ICD-9 390-459) and respiratory mortality (ICD-9 460-519). Maximum hourly concentrations of nitrogen dioxide were used rather than daily means as the former were more widely available. Concentrations of nitrogen dioxide varied from city to city, ranging from 46-155$\mu g/m^3$, and for a $10\mu g/m^3$ increase in nitrogen dioxide concentrations the pooled relative risk for total mortality was 1.003 with a 95% confidence interval of 1.0022 to 1.0038. For cardiovascular mortality the associated relative risk was 1.004 (95% CI: 1.0029 to 1.0052) and for respiratory mortality the results were similar with a relative risk of 1.0038 (95% CI: 1.0017 to 1.0058).

Carbon Monoxide

Samoli et al. (2007) investigated the short-term effects of carbon monoxide on all-cause mortality (ICD-9 <800) and deaths from cardiovascular diseases (ICD-9 390-459). For a $1mg/m^3$ increase in carbon monoxide concentrations, a relative risk of 1.012 was associated with all-cause mortality (95% CI: 1.0063 to 1.0177), whilst for cardiovascular mortality the relative risk was 1.0125 (95% CI: 1.003 to 1.0221).

Particulate Matter

Particulate matter is made up of an accumulation of small particles, including engine emissions and dust. There are a number of different measurements of particulate matter that have shown a harmful effect on health, a few of which are mentioned below.

PM_{10} is defined to be particulate matter of diameter $10\mu m$ or less and Lee et al. (2006) investigated its effects on admissions to hospital with asthma amongst children under the age of 18 in Hong Kong. Their data spanned six years, from January 1997 to December 2002 during which time the mean PM_{10} concentration was $56.1\mu g/m^3$ and the median was $51.1\mu g/m^3$. The study examined the effect of PM_{10} concentrations on the same day of admission and the preceeding five days and for a $10\mu g/m^3$ increase, they found a significant effect at each lag, with the greatest increase in hospital admissions at lag four (RR: 1.0217, 95% CI: 1.0164 to 1.0271).

Ostro et al. (2006) studied the effects of $PM_{2.5}$ (particulate matter of diameter $2.5\mu m$ or less) on mortality in nine counties in the US, using a number of mortality classifications, including total deaths minus accidents and homicides, deaths from respiratory diseases (ICD-10 J00-J98) and deaths from cardiovascular diseases (ICD-10 I00-I99). Daily all-cause mortality was also calculated for

those over the age of 65. When the results were pooled, an average $PM_{2.5}$ concentration from lags zero and one led to significant effects for each health outcome. The results were especially strong for respiratory mortality, with a relative risk of 1.022 (95% CI: 1.006 to 1.039). Within cardiovascular mortality, although the pooled results were significant (RR of 1.006, 95% CI: 1.000 to 1.011), they were not significant for any of the individual counties. This suggests that the pooled effect may not be significant, and is actually a result of multiple testing.

Total Suspended Matter (TSM) is defined by the U.S. Environmental Protection Agency as "a method of monitoring airborne particulate matter by total weight". Goldberg et al. (2001b) investigate the effects of TSM on daily non-accidental mortality and find a significant effect with a relative risk of 1.0065 (95% CI: 1.00 to 1.013), at a lag of zero, but did not find any significant effects of exposure to TSM on any previous day.

Goldberg et al. (2003) define Coefficient of Haze (COH) as a measurement of organic and inorganic carbon. They studied the effects of several pollutants, including COH, on Congestive Heart Failure (CHF) in Montreal, Quebec, between 1984 and 1993. Separate analyses were carried for people who died of CHF (ISD-9 428) and people who were diagnosed with CHF before dying from another independent cause. Pollution concentrations on the day of death, the previous day and an average over lags zero to two days were investigated. For deaths from CHF, no significant effect of COH was found at any of the lags. However, among persons diagnosed with CHF before death, results were significant at all three exposure lag periods, and especially for the three-day mean, with a relative risk of 1.0432 (95% CI: 1.0095 to 1.078). In the above study. relative risks were calculated for an increase of 1.85 COH units.

Black smoke is a measure of the darkness of particles collected on a filter, with smaller particles being darker. Black smoke particles are generally considered

to be larger than PM_{10} but smaller than TSP (Dockery & Pope (1994)). Bogdanovic et al. (2006) studied the effects of black smoke on all non-accidental deaths (ICD-10 A00-R99), respiratory mortality (ICD-10, J00-J99) and cardiovascular mortality (ICD10, I00-I99). They lagged pollution for up to seven days and included all lags in one model. Their results were most significant at lag zero, where the relative risk for total mortality, given a $10\mu g/m^3$ increase in black smoke concentrations, was 1.0113 (95% CI: 1.0008 to 1.0220). The relative risk for cardiovascular mortality was slightly higher at 1.0125 (95% CI: 1.0053 to 1.0197) while the relative risk for respiratory mortality was not statistically significant. However due to the inclusion of eight, potentially highly correlated, pollution variables, collinearity may have been a problem.

Also of interest is the combined effect of pollutants on health, rather than individual pollutants. Yu et al. (2000) investigate the effects of particulate matter and carbon monoxide on asthma symptoms. They calculate a relative risk for a joint increase of $10\mu g/m^3$ in particulate matter concentrations and a 1ppm in carbon monoxide concentrations. Their results showed that although the individual relative risks are smaller, the joint effect of the pollutants is similar to the effect they found in single pollutant models. In contrast, Hong et al. (1999) introduced pollution indices, which are the sum of the individual pollutants, divided by their means, which reflects the variations amongst the pollutants. They reported larger relative risks from the multiple pollutant models than were found from the single pollutant models.

2.3.2 Lag

The second choice to make when specifying the pollution exposure in these studies is the value of the lag to use. There has been some debate over which is the most appropriate lag to use in these studies, with some studies investigating the effects at a lag of zero, for example Schwartz (2004), while others use single lags between one and five days (Kelsall et al. (1997)) However some more recent

studies take into account the accumulated effect of several days pollution, with Hertz-Picciotto et al. (2007) calculating the average exposure over lags zero to two. Other authors use multiple lag models, with the choice of lags varying from study to study. For example Cakmak et al. (2007) takes lags of up to five days, while Goodman et al. (2004) took lags of up to 40 days.

Another area of discussion is whether to include multiple lags in one model (Prescott et al. (1998)), or to run several models with different lags (Tellez-Rojo et al. (2000)), although some authors such as Roberts (2004a) adopt both methods. A multiple lag model has the problem that the lags are going to be very highly correlated, thus resulting in collinearity, which reduces the accuracy of the estimates and inflates the standard errors. Zanobetti et al. (2000) proposed a distributed lag model to overcome this problem, in which multiple lags are included but the effect estimates are constrained to reduce collinearity. However to date, no consensus has been reached as to the best approach for solving the 'lag problem'.

2.3.3 Dose-response Shape

The third subject of consideration is the shape of the relationship between air pollution and health. Most studies such as this one assume that the effect of pollution increases linearly as suggested in Figure 2.1. This may be a plausible theory as, to date, there has been no threshold concentration found below which pollution does not adversely affect cardiovascular or respiratory health.

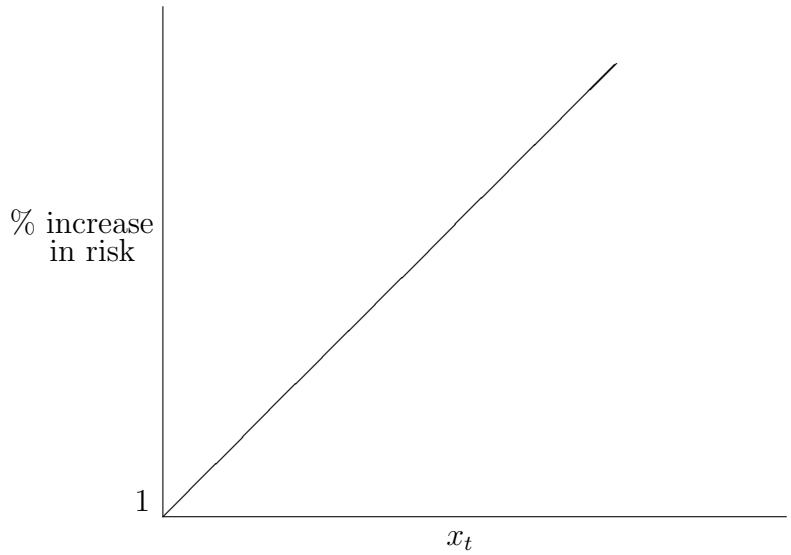

Figure 2.1. Linear effect of pollution

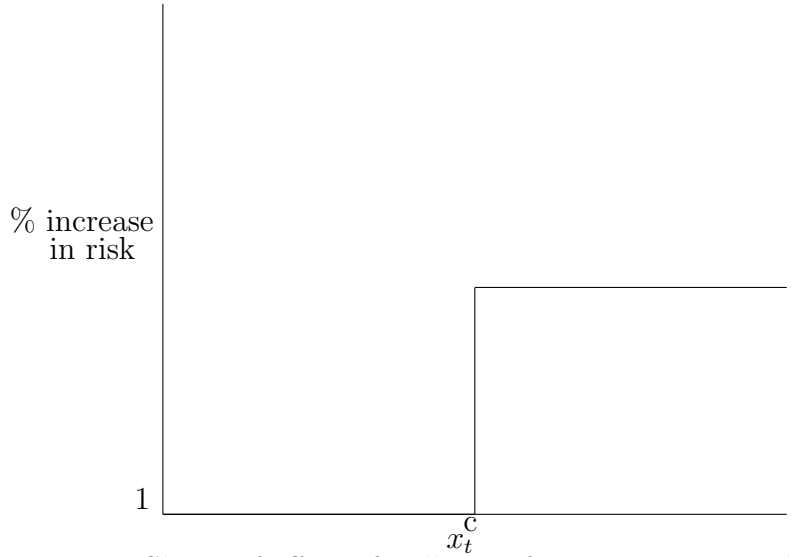

Figure 2.2. Shape of effect of pollution from a given exceedence value

Figure 2.2 shows the shape of the exposure response curve for concentrations that exceed a given value. This shape is somewhat implausible as it suggests that pollution has no adverse effect on health before this value, while after this value the effect jumps to a higher, unknown, risk after which it stays constant. This is unlikely as it implies the effect of pollution is discontinuous meaning that for concentrations below the set limit, there is no effect of pollution, while above this limit, there is an effect.

Recently Shaddick et al. (2008) proposed another possible shape of the dose-response curve. They suggested that a linear relationship may not be appropriate as there would eventually be an upper bound on the effect that air pollution has on health. Instead they recommended a function g that satisfies the following requirements: (i) boundedness; (ii) increasing monotonicity; (iii) smoothness; and (iv) no effect with no exposure

The European Directive on Air Quality recommends that the annual mean concentration for PM_{10} should be no higher than $40\mu g/m^3$. In addition, the daily mean concentration should exceed $50\mu g/m^3$ no more than 35 times a year, which should have been achieved by 1st January 2005. According to Defra (2007), this is also the concentrations that should have been achieved in the UK, except in Scotland, where the overall pollution concentrations are lower. There the objective was to reduce the annual mean PM_{10} concentrations to $18\mu g/m^3$ and the daily mean should exceed $50\mu g/m^3$ no more than seven times a year. However there has been extensive research done that suggests that even at this comparatively low concentration, pollution does have an adverse effect on health. For example Daniels et al. (2000) found that the relationship between PM_{10} and all cause mortality and cardiorespiratory morality in the US was linear even at the lowest concentrations, although for other causes, there was no noticeable effect

until $50\mu g/m^3$. In a more recently study, Daniels et al. (2004) found that at concentrations as low as $10\mu g/m^3$ a relationship between PM_{10} and cardiovascular-respiratory mortality could still be seen, thus implying there is no known concentration of PM_{10} which does not have an adverse effect on health.

2.4 Covariate Modelling

In addition to specifying the exposure, a number of confounding factors need to be modelled. There are a number of covariates that are regularly used in air pollution studies to remove potential confounding effects and so avoid biased estimates. A brief description is given below.

2.4.1 Meteorological Variables

Of the many covariates used in time series studies, meteorological variables are the most commonly used. These are usually measured by weather monitors situated at various points around a city, in particular at airports. The weather variables used most often in air pollution studies are described below.

Temperature

Temperature is recognised as having a confounding effect on air pollution and health studies, as it is thought to have a U-shaped relationship with ill health, with more people dying or becoming ill at extremely cold or hot temperatures. Lee et al. (2006) added mean daily temperature to their model after investigating its effect on mortality up to five days before admission to hospital. They eventually chose to add a smooth function of temperature on the day of admission, based on the minimisation of Akaike's Information Criteria (AIC). The mean temperature during the time period studied was 23.7°C.

Relative Humidity

Relative humidity measures how humid the air is in comparison with how humid it could be at that particular temperature, and is often added to air pollution studies as higher levels of humidity have been known to aggravate certain illnesses, including asthma. When Ostro et al. (2006) investigated the effects of air pollution in nine counties in California, they added a number of covariates, including a smooth function of average humidity, with three degrees of freedom, lagged by one day. Mean daily relative humidity varied from county to county, going from 55% in Fresno to 74% in San Diego.

Dew Point Temperature

Dew point temperature is often used as a replacement for relative humidity as this also measures the amount of humidity in the air. Bell et al. (2004) included dew point in their study of how ozone affects mortality in the US. They added the dew point on the day of death and also the average of the previous three days' dew points.

Barometric Pressure

There have been some studies which suggested that extreme low or high pressure can be related to stroke incidences although more research is still needed in this field. Stankovic et al. (2007) added air pressure to their model when trying to relate air pollution to cardiovascular mortality in Nis, Serbia. The mean pressure between 2001 and 2005 was 993.8mBar. Air pressure on the same day of death was added to their analysis as a natural cubic spline with three degrees of freedom.

2.4.2 Trend and Seasonal Variation

Mortality and morbidity are often seen to be highly seasonal, with more deaths or admissions to hospital occurring in winter than in summer. Before trying to model any relationship between mortality or morbidity and air pollution, this

seasonality must be removed so as not to bias the parameter estimates. There are numerous ways of removing seasonality that has not been removed by the addition of meteorological variables. A commonly used method is adding sinusoidal curves to the model, such as the terms $\sin(2\pi \times \text{Time}/365.25)$ and $\cos(2\pi \times \text{Time}/365.25)$ which exhibit one peak a year. This method can be adapted by additionally including sinusoidal terms with different numbers of peaks a year, although due to the rigid nature of sinusoidal terms, they are not always suitable for removing seasonality as they do not allow for much variation. Therefore natural cubic splines are often used as an alternative. A natural cubic spline fits a number of polynomial functions joined together at knots. These knots are typically placed evenly throughout the variable of interest (e.g. time). The function is constrained to be continuous and has two additional knots at each end of the data. The number of knots determines how smooth the function will be.

To remove any overall trend in the data, a function of date is often added to the model with most studies adding this trend as a smoothed function of time. Various smoothing methods have been used, including natural cubic splines (Stankovic et al. (2007)) and loess functions (Ballester et al. (2006)). When using natural cubic splines, the number of degrees of freedom varies from study to study. In the study by Stankovic et al mentioned above, 30 degrees of freedom are used over a five year period, while Ostro et al. (2006) used 28 degrees of freedom (seven degrees of freedom per year of study).

2.4.3 Day of the Week Effect

A number of studies have recently established what is known as the 'Monday Effect', which refers to the increased numbers of deaths or admissions to hospital on a Monday. Evans et al. (2000) investigated cases of mortality from coronary heart disease in Scotland, while Witte et al. (2005) did a meta-analysis of 27 studies which reported a weekly variation in cardiovascular mortality and morbidity. Both studies found a marked increased incidence rate on Mondays. To

remove this effect, an indicator variable for Day of the Week has often been used (Samoli et al. (2007)).

2.5 Mortality Displacement

Although the detrimental effects of air pollution have become widely accepted, there is still some controversy over how strong these effects are. The theory of mortality displacement (also known as 'harvesting') suggests that it is mainly the frail proportion of the population who are at risk, and as such their deaths are brought forward by only a few days. If this is the case, the significance of air pollution in terms of public health is reduced. Zanobetti et al. (2002) claim that if the theory of mortality displacement is true, the correlation between air pollution and mortality would be positive during and immediately after exposure and would then be counterbalanced by a negative correlation at a later lag, as seen in Figure 2.3. In their study they use distributed lags to examine the effects of air pollution on 10 different European cities, and their results contradict the mortality displacement hypothesis. However recently Roberts & Switzer (2004) used simulations to investigate the properties of distributed lag models in the context of mortality displacement. Their results implied that distributed lag models are likely to give biased estimates and are thus likely to be misleading.

2.6 Meta Analyses and Multi-City Studies

There have been numerous air pollution studies carried out in cities around the world which have varied widely in terms of the pollutant analysed and the results obtained. Meta-analyses have often been used to combine and compare these results (Stieb et al. (2002)). However these meta-analyses are usually carried out on single-city analyses, and the quality of the individual studies varies widely, thus affecting the results.

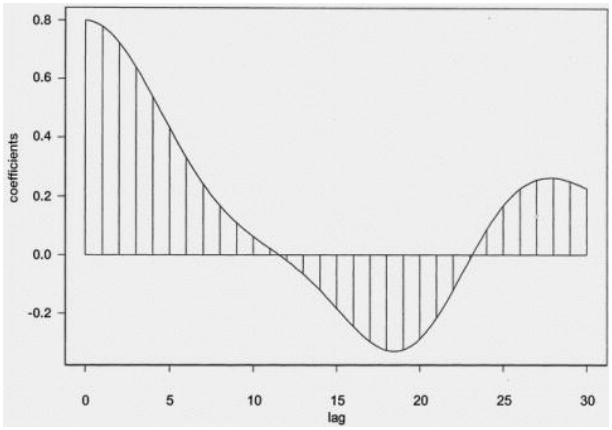

Figure 2.3. Suggested lag structure corresponding the mortality displacement effect (taken from Zanobetti et al. (2002))

An alternative approach is multi-city studies such as Air Pollution and Health: A European Approach (APHEA, Katsouyanni et al. (1995)) which studied the effects of various pollutants in 10 European cities. This was followed by APHEA-2 (Sunyer et al. (2003a), Sunyer et al. (2003b)) which increased the number of cities to 29 and the National Morbidity Mortality Air Pollution Study (NMMAPS, Dominici et al. (2000)) which examines 90 cities in America. These studies analyse the effects of air pollution in a number of cities using a standard protocol thus allowing the results to be compared directly. NMMAPS found that for a $10\mu g/m^3$ increase in PM_{10} concentrations there was on average, a relative risk of increased mortality of 1.0048 in the US (95% CI: 1.0005, 1.0092). When adjusted for ozone, this increased slightly to 1.0052 (95% CI: 1.0016, 1.0085). The APHEA 2 project found a slightly higher estimated effect of PM_{10} in Europe. For a $10\mu g/m^3$ increase, they found a relative risk of 1.009 (95% CI: 1.006, 1.013).

2.7 Ecological Fallacy

The Ecological Fallacy is defined by Delgado-Rodriguez & Llorca (2004) as being "a bias produced when analyses realised in an ecological (group level) analysis

are used to make inferences at the individual level". This can have major consequences in air pollution studies, where the effect of air pollution can be biased. One of the main problems with air pollution studies is the fact that pollution monitors are positioned at fixed points and so do not necessarily represent the true extent of pollution inhaled. In recent years, there have been a number of models proposed that try to overcome this problem, such as those by Lancaster & Green (2002) and Wakefield & Shaddick (2005).

Chapter 3

Multi-City Study

3.1 Introduction

In this chapter, we present a new epidemiological study of air pollution and health in urban Scotland, focusing on data from Aberdeen, Dundee, Edinburgh and Glasgow during the period 2000 to 2006 inclusive. These four cities are the largest in Scotland, with Glasgow being by far the largest with a population of approximately 580 000. Edinburgh is the second largest city with an approximate population of 475 000 people, while Aberdeen and Dundee are much smaller with populations of 200 000 and 145 000 respectively.

The health data for this epidemiological study are provided by the Information and Services Division (ISD) of the NHS, and comprise daily numbers of both acute and repeat admissions to hospital from cardiovascular diseases for Glasgow, Edinburgh, Aberdeen and Dundee from 2000 to 2006. Cardiovascular diseases are defined as being Coronary Heart Disease (CHD) or stroke, and are classified using the International Classification of Diseases 10th Revision (ICD-10) with CHD being ICD-10 codes I20 to I25 and stroke being codes I61, I63 and I64. In 2000 the standardised incidence rate per 100 000 of the Scottish population for CHD was 404.1, which fell to 307.5 in 2006. In comparison, the standardised incidence rate per 100 000 of the population for stroke was 218.8 in

2000, decreasing to 166 in 2006.

Data on several different pollutants were available to download from the Scottish Air Quality website (http://www.scottishairquality.co.uk), including carbon monoxide, nitrous oxides, particulate matter and ozone. However for this study, we focus exclusively on PM_{10}. Mean daily PM_{10} concentrations are available for Glasgow, Edinburgh and Aberdeen over the time period 2000 to 2006 while data for Dundee are only available from 2006 onwards. In Glasgow, PM_{10} concentrations were measured at six sites, Anderston, Battlefield Road, Byres Road, Centre, Kerbside and Waulkmill Glen Reservoir. However data for the whole period are only available from Glasgow Centre and Kerbside as the remainder did not start recording PM_{10} concentrations until 2005. In Edinburgh PM_{10} concentrations are only measured at Haymarket, Roseburn and St. Leonards, while in Aberdeen they were monitored at Aberdeen, Anderson Drive, Market Street and Union Street. However in common with Glasgow, data for Aberdeen over the whole period are only available from Aberdeen as the other sites did not have data available until 2005. To obtain a single measure of daily PM_{10} for each city, the values across the sites were averaged, an approach that is frequently adopted in these studies (Chardon et al. (2007), Lee et al. (2006)).

The meteorology data available for this study were provided by the British Atmospheric Data Centre (http://badc.nerc.ac.uk/home/index.html) and comprise daily maximum and minimum temperatures for one site in Aberdeen, two sites in Glasgow and three in Edinburgh for 2000 to 2006. There was, however, no data available for Dundee. For each meteorological variable, the measurements were averaged over all sites to produce a single daily measure for each city. The Aberdeen site was at Mannofield, about two miles outside the city centre and provided no data from October 2006 onwards, meaning that no temperature data are available for the last three months of the study in this city. The Glasgow sites were at Pollock Country Park and Bishopton and while there are some missing

data, there were no days where no data at all was collected. Finally the Edinburgh sites are at Blackford hill, the Botanic gardens and Gogarbank, and like Glasgow there are no days without a single measurement. We should be aware that the meteorological sites are far away from the pollution monitors, and thus may not give an accurate representation of the temperature at the pollution monitors.

As there was no pollution or temperature data available for Dundee, the analysis for this city cannot be carried. Therefore this chapter will focus estimating the association between PM_{10} and cardiovascular health in Glasgow, Edinburgh and Aberdeen.

3.2 Modelling Structure

This section will give a brief overview of the modelling strategy adopted in this thesis, which can be split into four parts, each of which are outlined below.

3.2.1 Exploratory Analysis

The first step is to produce numerical and graphical summaries of the data, which will aid the modelling process. These summaries will include plots of the data over time, comprising daily (or weekly) numbers of admissions to hospital, mean daily (or weekly) PM_{10} concentrations and maximum daily temperature (or mean weekly maximum temperature), which will aid in distinguishing any long term trends. These variables will also be plotted against each other to determine their inter-relationships.

3.2.2 Covariate Modelling

The second step is to model the prominent features in the health data, using any available covariates as well as artificial variables to remove any trend or seasonal structure. The response variable is admissions to hospital which come in the form

of daily counts, meaning that a Poisson generalised linear model is appropriate for these data. This model has the general form

$$\begin{aligned} Y_t &\sim \text{Poisson}(\mu_t) \\ \ln(\mu_t) &= \mathbf{z}_t^T \boldsymbol{\alpha} + \gamma \text{PM}_{10_{t-q}} \end{aligned} \tag{3.1}$$

where Y_t are the number of admissions to hospital on day t, $\mathbf{z}_t = (z_{t1}, ..., z_{tp})$ are a vector of p covariates for day t, $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_p)$ are the associated regression parameters and γ is the estimated effect of PM_{10}. In the model, pollution exposure is related to health at a lag of q days, which ranges between zero (exposure on the same day) and 14 (PM_{10} concentrations a fortnight before the health event) days in each city.

There are a number of external risk factors which can affect the health data in this study and may induce long-term trends and seasonal variation into these daily time series. Therefore the effects of these factors must be removed before adding PM_{10} to the model, otherwise its effect may be biased. Long-term time trends can be removed either by adding a parametric function of time to the model, such as a polynomial in t, or by using a flexible smooth function, such as natural cubic splines. To remove the seasonal variation, temperature is often added to the model as it has a well known seasonal shape (hotter in the summer than the winter), and has been shown to have a significant effect on health in numerous studies (Pauli & Rizzi (2006), Carder et al. (2005)). If temperature is not sufficient to model the seasonal variation, then sinusoidal fucntions can be added. However if this variation is not regular then flexible smooth functions such as natural cubic splines or lowess can be used. Initially fixed functions of time, such as linear and sinusoidal functions, will be used to model the data, and if this is not appropriate we revert to a flexible formulation via natural cubic splines.

3.2.3 Model Checking

Once the variation in the data has been removed the model should be assessed to determine whether it is an adequate description of the data. This is achieved by assessing numerical and graphical summaries of the standardised residuals, r_t, where $r_t = \frac{Y_t - \mu_t}{\sqrt{\mu_t}}$. If the model is a good fit to the data, $E(r_t) = 0$, $Var(r_t) = 1$, they should contain no obvious trend, structure or relationship with any covariates and should be uncorrelated. Residual correlation is checked using the autocorrelation function (ACF), where particularly high values (except at lag zero) suggest correlation is present. A further method of determining the fit of the model is to look at Akaike's Information Criterion (AIC) with lower values indicating a better fit to the data. This is often used for comparing two possible models and in this study it is used to determine the shape and smoothness of the trend model, for example for selecting the degrees of freedom for the natural cubic splines.

3.2.4 Pollution Modelling

Once a model has been constructed that appears to remove any trend, seasonality or correlation, pollution will then be added. First we fit a series of single lag models from zero to 14 days, to determine the latency with which pollution may effect cardiovascular health. For these results and the rest in this thesis, we present the associations between PM_{10} and health on the relative risk scale for a $10\mu g/m^3$ increase in PM_{10} concentrations. However it seems unlikely that air pollution exposure from only one day will affect admissions to hospital and it is more likely there is an accumulated effect over a number of days. Therefore we also fit a multiple lag model to the data containing PM_{10} concentrations from the 14 days prior to admission. However this model has the problem that the pollution concentrations on adjacent days are likely to be highly correlated, thus resulting in collinearity. One method of overcoming collinearity is to use a distributed lag model, which constrains the effects of pollution at each lag to follow

a polynomial. If the polynomial is of power zero, the relative risk at each lag will be the same, essentially meaning that only one parameter is being estimated, thus removing any problem of collinearity. At the other extreme, setting the power equal to the number of lags results in no constraint, which means the relative risks are the same as those from the multiple lag model and again suffer from the problem of collinearity. Therefore the power of the polynomial is chosen to minimise collinearity, whilst allowing the relative risks to exhibit a biologically plausible shape over the 14 lags.

The modelling structure described above will be used in the analyses in both this and the following chapter and we begin with daily admissions to hospital in Glasgow.

3.3 Glasgow

3.3.1 Exploratory Analysis

We start by estimating the association between daily PM_{10} concentrations and cardiovascular admissions to hospital, beginning with a series of descriptive plots of the data.

Health

Figure 3.1 shows the number of daily hospital admissions in Glasgow due to cardiovascular illness between 2000 and 2006, with a smoothed lowess line to show the underlying trend in the data. The figure shows an approximately linear decreasing trend in hospital admissions over the seven year period, with little discernible seasonal variation. This latter observation is surprising, as previous studies (Pauli & Rizzi (2006)) have found the admissions to hospital to be highly seasonal, with more in the winter than in the summer.

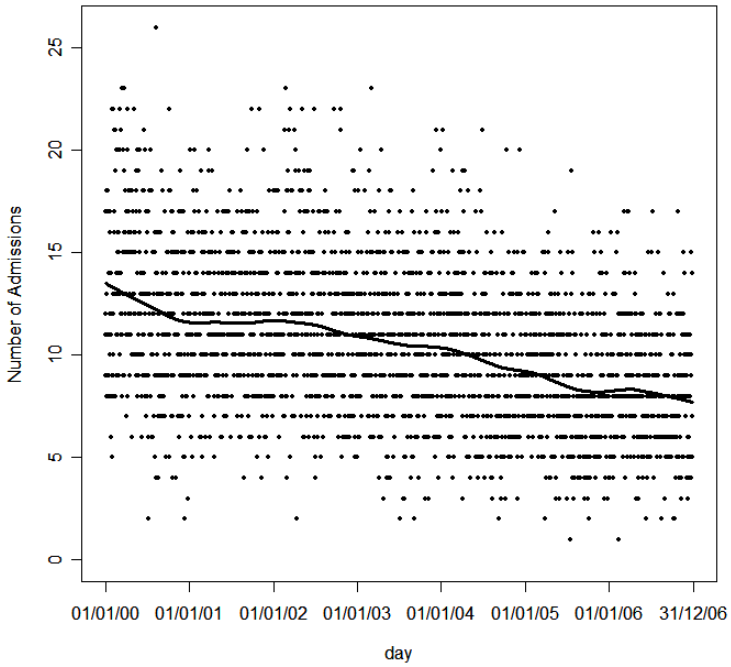

Figure 3.1. Daily hospital admissions in Glasgow from 2000 to 2006

Pollution

Figure 3.2 shows the daily PM_{10} concentrations for Glasgow from 2000 to 2006, while Figure 3.3 depicts the relationship between PM_{10} concentrations and admissions to hospital. Figure 3.2 shows that there were some unexpectedly high PM_{10} concentrations in the winters of 2000 to 2003, although the overall trend appears to stay fairly constant. From Figure 3.3 there appears to be a weakly positive relationship between PM_{10} concentrations and admissions to hospital in Glasgow, with admissions to hospital increasing as PM_{10} concentrations increase.

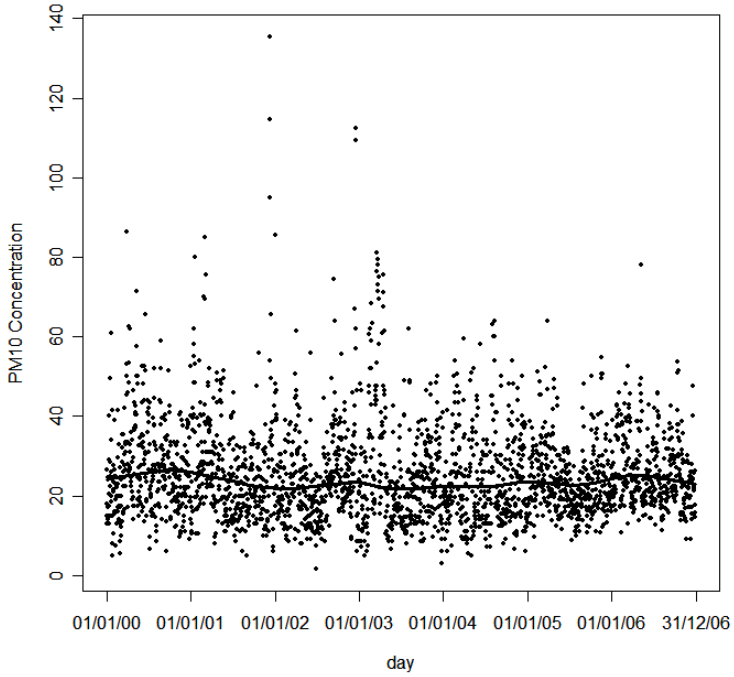

Figure 3.2. PM_{10} concentrations in Glasgow between 2000 and 2006

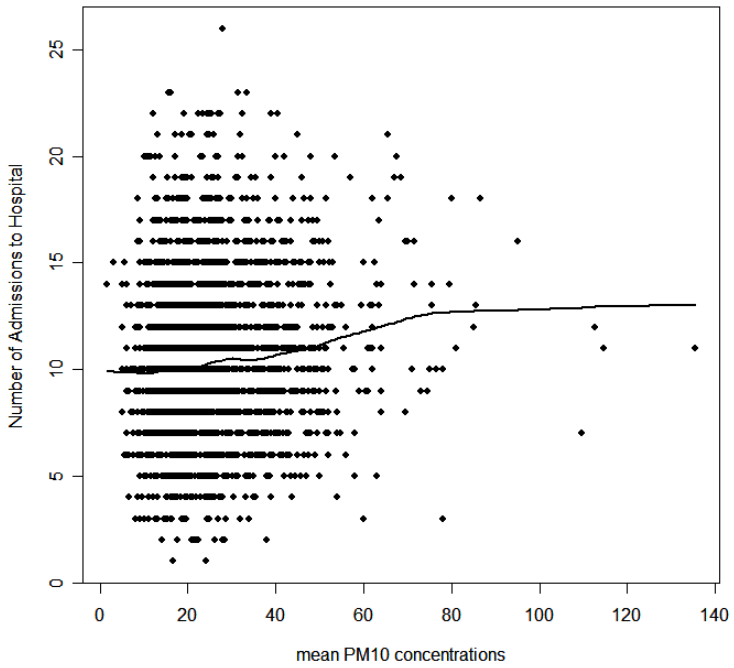

Figure 3.3. Relationship between hospital admissions and daily mean PM_{10} concentrations

Temperature

Figures 3.4, 3.5 and 3.6 are plots of the daily maximum temperatures over the seven year period, the first of which is against time, the second against hospital admissions and the third against PM_{10} concentrations. As expected, temperature is very seasonal, with peaks in summer and troughs in winter. However this seasonality was not observed in the admissions data (Figure 3.1) which is why there does not appear to be any significant relationship between admissions to hospital with cardiovascular diseases in Glasgow and temperature. Conversely there appears to be a slight quadratic curve between PM_{10} concentrations and temperature, with the former being higher for very low or high temperatures.

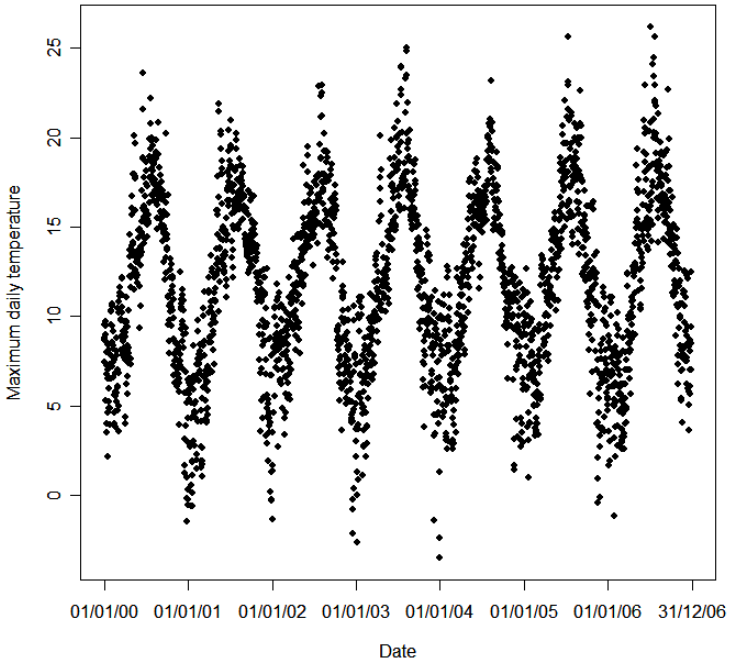

Figure 3.4. Maximum daily temperatures in Glasgow between 2000 and 2006

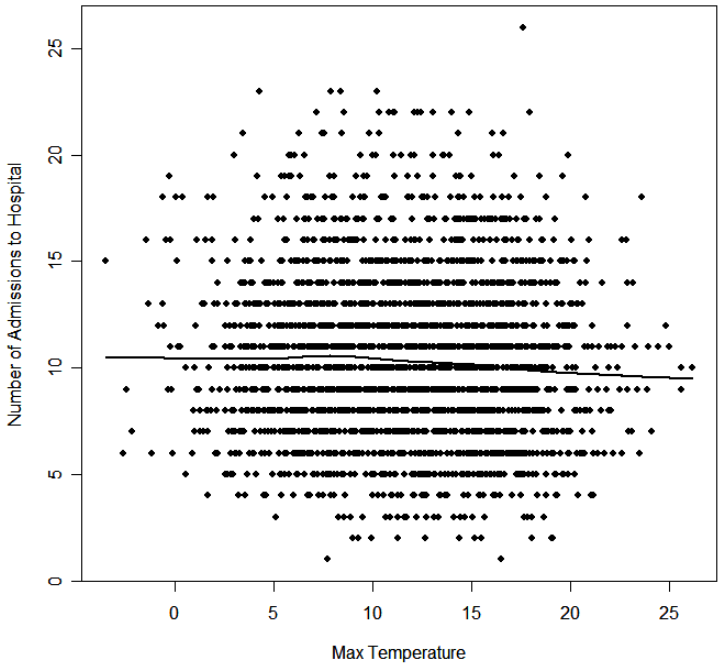

Figure 3.5. Relationship between hospital admissions and daily maximum temperature

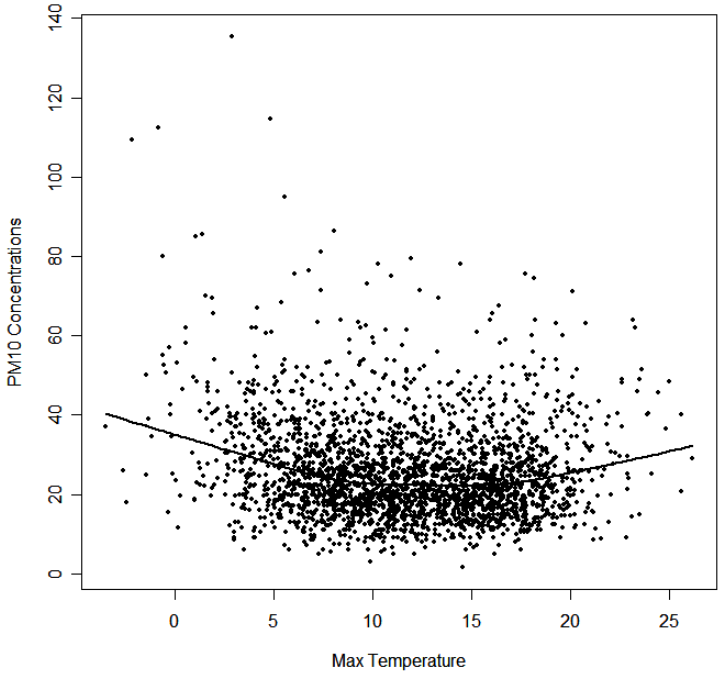

Figure 3.6. Relationship between PM_{10} concentrations and daily maximum temperature

3.3.2 Covariate Modelling

The next step in the analysis is to produce an adequate model for the daily admissions data, that has removed all trend, seasonal behaviour and other structure. As discussed in Section 3.2.2, the admissions to hospital are a time series of daily counts, meaning that a Poisson generalised linear model is appropriate (Model (3.1)). The main feature of the admissions data is a linearly decreasing trend (Figure 3.1), therefore as an initial model we include a linear time trend. Thus Model (3.2) is of the form

$$\ln(\mu_t) = \alpha_0 + \alpha_1 t \tag{3.2}$$

with a summary of the model fit provided in Table 3.1

Coefficient	Estimate	Standard Error	P-Value
Intercept	2.591	0.01151	$<2 \times 10^{-16}$
t	-0.0001954	8.318×10^{-06}	$<2 \times 10^{-16}$
AIC			13608

Table 3.1. Summary of Model (3.2)

The p-value for the linear time trend is very small (less than 0.05) meaning that time is highly significant in the model. The next step is to check the residuals r_t, to see how well this model fits the data. Recall that if the model is a good fit, $E(r_t) = 0$, $Var(r_t) = 1$, there should be no obvious pattern and the residuals should be uncorrelated. The residuals are plotted against time in Figure 3.7 and we can see an obvious parallel line pattern. However according to Nelder (1990) and Searle (1988), this is due to the fact that there is a limited number of admissions in any one day, and is to be expected. There also appears to be some signs of seasonality which suggests the initial model is not adequate.

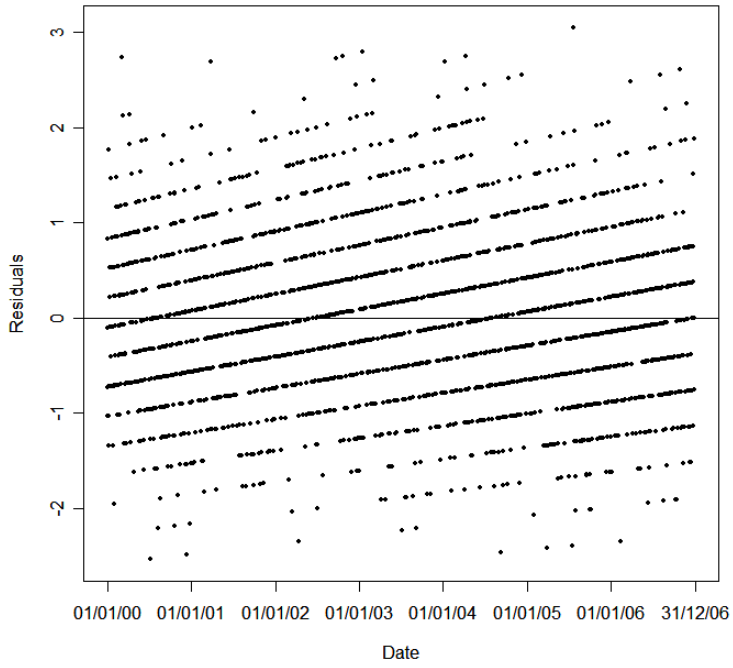

Figure 3.7. Residuals for Model (3.2)

Much of the literature available about short-term air pollution and health studies indicates that temperature has a seasonal effect on mortality and morbidity (Carder et al. (2005), Pattenden et al. (2003)), although no clear association is seen in Figure 3.5 Therefore to try and remove the small amount of seasonality within the residuals, temperature will be included in the next model. The temperature data available are maximum and minimum temperatures for two sites in Glasgow. An average maximum temperature is obtained by taking the mean of the maximum temperatures at each site and an average minimum temperature is similarly obtained. The maximum and minimum values are highly correlated as expected (0.86), meaning that only one is required. As the two temperatures variables are so highly correlated, it seems unlikely that they would give significantly different results. Therefore maximum temperature was chosen to be added

to the model giving the model

$$\ln(\mu_t) = \alpha_0 + \alpha_1 t + \alpha_2 \text{Max}_t \tag{3.3}$$

which is summarised in Table 3.2.

Coefficient	Estimate	Standard Error	P-Value
Intercept	2.621	0.01784	$<2 \times 10^{-16}$
t	-0.000193	8.343×10^{-06}	$<2 \times 10^{-16}$
Max	-0.002679	0.001237	0.0304
AIC			13601

Table 3.2. Summary of Model (3.3)

Table 3.2 shows that maximum temperature is significant in the model, although the p-value is much larger than that for the linear trend term. Also the AIC is lower than the AIC for the previous model, suggesting this model is a better fit of the data. However Figure 3.8 shows the residuals for this model, and suggests that the model is still not a good fit to the data as there is still seasonality in the residuals.

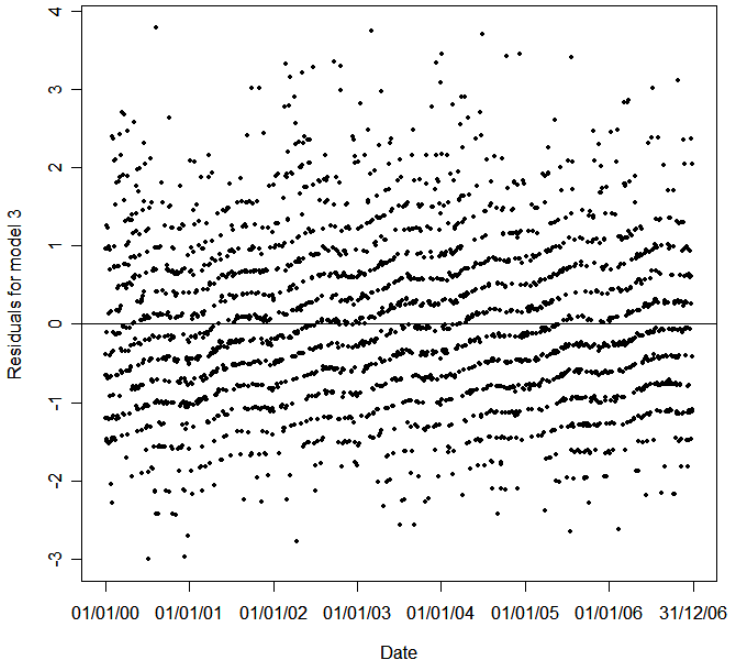

Figure 3.8. Residuals for model (3.3)

One approach is to remove it by adding sine and cosine terms (Keatinge & Donaldson (2001)), to the model. Such studies typically begin with the pair of variables [$\sin(2\pi \times \text{Time}/365.25)$, $\cos(2\pi \times \text{Time}/365.25)$] but as they exhibit one peak a year, they are highly correlated with temperature (-0.81), and are not required here. Instead we add the pair of terms [$\sin(4\pi \times \text{Time}/365.25)$, $\cos(4\pi \times \text{Time}/365.25)$] to the model to represent two peaks a year, as the residuals seem to show peaks that occur more than once a year. This gives Model (3.4):

$$\ln(\mu_t) = \alpha_0 + \alpha_1 t + \alpha_2 \text{Max}_t + \alpha_3 \sin(2\omega t) + \alpha_4 \cos(2\omega t) \tag{3.4}$$

where $\omega = \frac{2\pi}{365.25}$. Table 3.3 summarises the fit of Model (3.4).

Coefficient	Estimate	Standard Error	P-Value
Intercept	2.62	0.01788	$<2 \times 10^{-16}$
t	-0.0001944	8.359×10^{-06}	$<2 \times 10^{-16}$
Max	-0.002612	0.001243	0.0356
$\sin(2\omega t)$	-0.004908	8.658e-03	0.5708
$\cos(2\omega t)$	-0.01949	8.633e-03	0.0240
AIC			13600

Table 3.3. Summary of Model (3.4)

From Table 3.3 we see that the linear time trend, maximum temperature and the cosine term are statistically significant in the model. The sine term is not significant but because the cosine term is significant, it must be kept in the model. This is due to the fact that having one sine or cosine term in the model forces the troughs and peaks to stay in one position, whereas including both terms allows for more flexibility, thus meaning the model should be a better fit to the data. As with previous models, the residuals are displayed to check the fit of the model (Figure 3.9). The pattern in the residuals may be caused by short range seasonality or residual correlation, and a look at the autocorrelation function (Figure 3.10) should provide more insight.

From Figure 3.10, we can see that the residuals exhibit correlation at regular seven day intervals, suggesting it may be necessary to add a 'day of the week' term to the model. This then gives Model (3.5) below

$$\begin{aligned} \ln(\mu_t) &= \alpha_0 + \alpha_1 t + \alpha_2 \text{Max}_t + \alpha_3 \sin(2\omega t) + \alpha_4 \cos(2\omega t) \\ &+ \sum_{j=1}^{6} \alpha_{j+4} \text{DoW}_t^{(j)} \end{aligned} \tag{3.5}$$

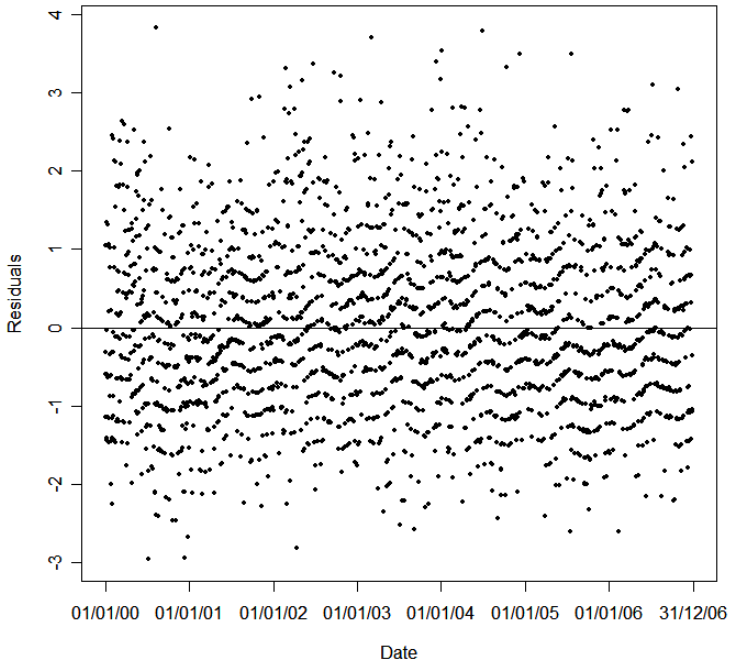

Figure 3.9. Residuals for Model (3.4)

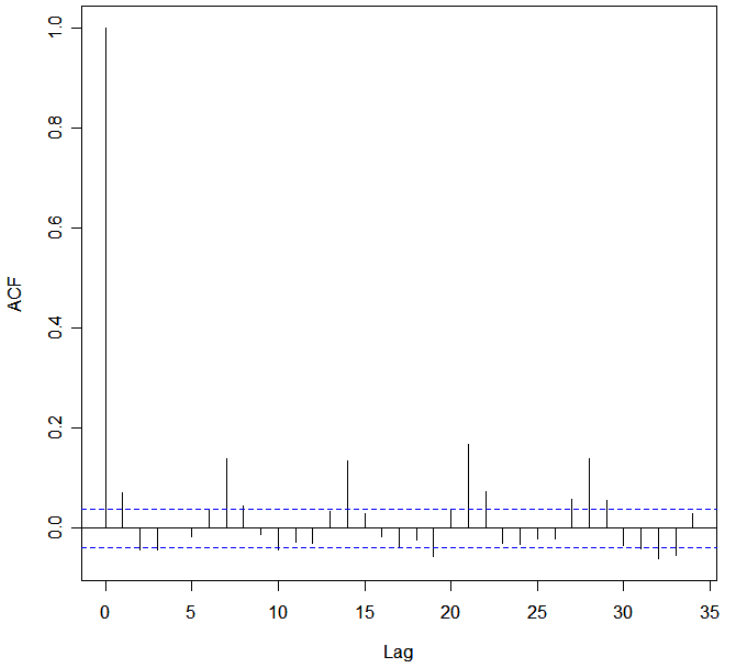

Figure 3.10. Autocorrelation function for residuals

where as before $\omega = \frac{2\pi}{365.25}$, and $\text{DoW}_t^{(j)}$ is an indicator variable for the jth day of the week, ranging from from Tuesday (j=1) to Sunday (j=6). The indicator variables equal 1 if day t is the day of the week in question, or 0 otherwise. A summary of Model (3.5) can be seen in Table 3.4 and again, the sine term is not significant in the model, but the cosine term is, therefore both must be left in. All other variables are significant in the model and Figures 3.11 and 3.12 suggest that the residuals resemble independent white noise, meaning that the model appears to be adequate. Also the AIC is much lower than the AIC in previous models, suggesting this model is the best fitting model to date.

Coefficients	Estimate	Standard Error	P-Value
Intercept	2.751	0.02258	$<2 \times 10^{-16}$
t	-0.000194	8.361×10^{-06}	$<2 \times 10^{-16}$
Max	-0.002938	0.001243	0.0181
$\sin(2\omega)$	-0.004792	0.008659	0.5800
$\cos(2\omega)$	-0.01939	0.008633	0.0247
tues	-0.08449	0.02190	0.0001
wed	-0.04255	0.02164	0.0493
thurs	-0.06566	0.02177	0.0026
fri	-0.08410	0.02188	0.0001
sat	-0.3479	0.02352	$<2 \times 10^{-16}$
sun	-0.3232	0.02337	$<2 \times 10^{-16}$
AIC			13194

Table 3.4. Summary of Model (3.5)

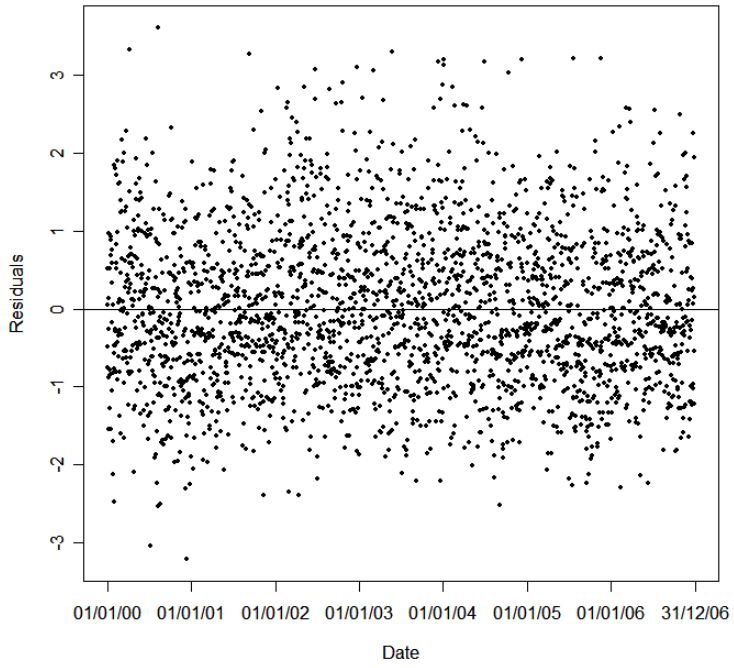

Figure 3.11. Residuals for Model (3.5)

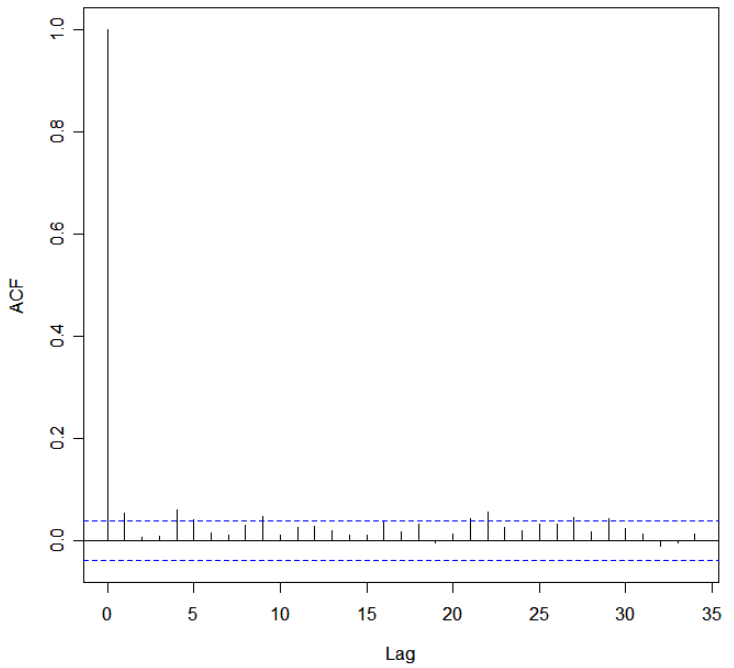

Figure 3.12. Autocorrelation Function for residuals

3.3.3 Pollution Modelling

The next step is now to add PM_{10} to the model, to estimate whether it is associated with admissions to hospital. We begin with a series of single lag models ranging from zero to 14 days, before examining a multiple lag approach. For a given lag q, the single lag model is seen as

$$\begin{aligned} \ln(\mu_t) &= \alpha_0 + \alpha_1 t + \alpha_2 \text{Max}_t + \alpha_3 \sin(2\omega t) + \alpha_4 \cos(2\omega t) \\ &+ \sum_{j=1}^{6} \alpha_{j+4} \text{DoW}_t^{(j)} + \gamma \text{PM}_{10_{t-q}} \end{aligned} \tag{3.6}$$

The associations between exposure to PM_{10} and health estimated in this thesis are presented as relative risks for a $10\mu g/m^3$ increase in ambient concentrations. The relative risk is the ratio of the expected number of hospital admissions given the current pollution concentrations divided by the expected numbers if the PM_{10} concentrations rose by $10\mu g/m^3$ and is defined in Chapter 2, equation (2.4).

During the course of this study, it was discovered that removing the indicator variables for day of the week made a significant impact on the estimated relative risks, which suggests there may be an interaction between pollution concentrations and day of the week, making interpretation of the results less straight-forward. Therefore Table 3.5 shows the relative risks and 95% confidence intervals for each single lag model with and without the 'day of the week' covariates included. The table shows that at each lag the relative risks for the model containing the 'day of the week terms' are all very close to one and non-significant at the 5% level, meaning that there does not appear to be any relationship between admissions to hospital from cardiovascular diseases and exposure to PM_{10}. Conversely, for the model without the 'day of the week' terms a number of the relative risks are significant. However the choice of lag is somewhat ad-hoc, and the effects of cumulative exposure over a few days may be of greater interest. In addition it is unlikely that the effect of air pollution on any one day is unrelated to pollution concentrations on adjacent days meaning that the estimated associations could be caused by

	Including DoW		Not Including DoW	
Lag	RR	95% CI	RR	95% CI
0	1.0008	(0.9991, 1.0011)	1.0197	(1.0100, 1.0294)
1	1.0004	(0.9991, 1.0010)	0.9949	(0.9853, 1.0047)
2	0.9940	(0.9984, 1.0004)	0.9859	(0.9763, 0.9957)
3	0.9913	(0.9981, 1.0001)	0.9907	(0.9810, 1.0005)
4	0.9945	(0.9985, 1.0004)	1.0042	(0.9946, 1.0139)
5	0.9962	(0.9986, 1.0006)	1.0075	(0.9977, 1.0171)
6	0.9969	(0.9987, 1.0007)	1.0208	(1.0112, 1.0305)
7	0.9977	(0.9988, 1.0007)	1.0178	(1.0082, 1.0276)
8	0.9925	(0.9983, 1.0002)	0.9972	(0.9875, 1.0069)
9	0.9935	(0.9984, 1.0003)	0.9910	(0.9813, 1.0008)
10	0.9960	(0.9986, 1.0006)	0.9949	(0.9853, 1.0047)
11	0.9954	(0.9985, 1.0005)	0.9967	(0.9870, 1.0064)
12	0.9948	(0.9985, 1.0005)	1.0009	(0.9912, 1.0106)
13	0.9977	(0.9878, 1.0077)	1.0190	(1.0094, 1.0287)
14	0.9962	(0.9864, 1.0061)	1.0234	(1.0138, 1.0332)

Table 3.5. Table showing relative risks for single lag models, both including (left) and excluding (right) the 'day of the week' covariates

confounding at a different lag. Therefore a multiple lag model containing the PM_{10} concentrations from the previous 14 days would be a sensible next step. Therefore Model (3.7) will be

$$\begin{aligned} \ln(\mu_t) &= \alpha_0 + \alpha_1 t + \alpha_2 \text{Max}_t + \alpha_3 \sin(2\omega t) + \alpha_4 \cos(2\omega t) \\ &+ \sum_{j=1}^{6} \alpha_{j+4} \text{DoW}_t^{(j)} + \sum_{q=0}^{14} \gamma_q \text{PM}_{10_{t-q}} \end{aligned} \tag{3.7}$$

the results from which (with and without the 'day of the week' variables) are presented in Table 3.6. We can see that the relative risks from the model containing the 'day of the week' term are, again, all non-significant and very close to 1, although the 95% confidence intervals are wider than in the single lag models. Most of the relative risks from the model that does not contain the 'day of the week' term are also non-significant, although there are a few that are significant, again with wider confidence intervals then in the single lag model. This phenomenon is known as collinearity, which occurs when there is a strong correlation between

	DoW		No DoW	
Lag	RR	95% CI	RR	95% CI
0	1.0017	(0.9889, 1.0015)	1.0250	(1.0118, 1.0384)
1	1.0067	(0.9914, 1.0222)	0.9868	(0.9718, 1.0021)
2	0.9946	(0.9794, 1.0101)	0.9886	(0.9735, 1.0040)
3	0.9919	(0.9767, 1.0073)	0.9931	(0.9780, 1.0084)
4	1.0001	(0.9848, 1.0157)	1.0054	(0.9901, 1.0209)
5	1.0004	(0.9850, 1.0159)	0.9844	(0.9694, 0.9997)
6	0.9996	(0.9843, 1.0151)	1.0113	(0.9960, 1.0268)
7	1.0061	(0.9908, 1.0215)	1.0266	(1.0112, 1.0423)
8	0.9914	(0.9762, 1.0068)	0.9890	(0.9739, 1.0043)
9	0.9975	(0.9822, 1.0131)	0.9795	(0.9645, 0.9947)
10	1.0026	(0.9872, 1.0181)	0.9929	(0.9777, 1.0084)
11	0.9985	(0.9832, 1.0140)	0.9881	(0.9729, 1.0035)
12	0.9936	(0.9783, 1.0091)	0.9920	(0.9768, 1.0073)
13	1.0043	(0.9889, 1.0199)	1.0207	(1.0054, 1.0363)
14	1.0018	(0.9890, 1.0148)	1.0114	(0.9986, 1.0245)

Table 3.6. Table showing relative risks for the multiple lag models, both including (left) and excluding (right) the 'day of the week' covariates

two or more variables, thus making it difficult or impossible to estimate their individual regression coefficients reliably. One solution is to adopt a distributed lag model, as described in Chapter 2, which has the form

$$\begin{aligned} \ln(\mu_t) = \alpha_0 \quad &+ \quad \alpha_1 t + \alpha_2 \mathrm{Max}_t + \alpha_3 \sin(2\omega t) + \alpha_4 \cos(2\omega t) \\ &+ \quad \sum_{j=1}^{6} \alpha_{j+4} \mathrm{DoW}_t^{(j)} + \sum_{q=0}^{14} \gamma_q \mathrm{PM}_{10_{t-q}} \\ \text{where} \quad \gamma_q \quad &= \quad \sum_{k=0}^{q} \eta_k q^k, \quad q = 0, ..., 14 \end{aligned} \tag{3.8}$$

where collinearity is reduced by constraining the 15 parameters $(\gamma_0, ..., \gamma_{14})$ to follow a polynomial of order q in lag order. The order of the polynomial is unknown, but we are looking for an order that minimises collinearity but whose relative risks give a biologically plausible shape of adjacent lags.

For the model containing the 'day of the week' terms, the AIC ranges from 12016 to 12023 for powers of zero to four, while the model without the 'day of the week' terms has an AIC of 12325 at power four. These are all much lower than the AICs from the single lag models (12963 to 13052) but higher than the AIC from the multiple lag model (12025). However the multiple lag model has the problem of collinearity, thus suggesting the distributed lag model is the best fit of the data.

Figures 3.13 and 3.14 show the shape of the constrained relative risks against lag for orders zero to four, where Figure 3.13 presents the results for the models including the 'day of the week' covariates while Figure 3.14 presents the results for the models without. In both cases the solid horizontal line is the relative risk at power zero which gives the overall estimate of the effect of PM_{10}.

In Figure 3.13 we can see that at each power the points for the distributed lag model are very close to the null value of one, thus giving the impression that the relative risks are not high (or low) enough to be significant. Conversely, when the 'day of the week' term has been left out of the model, PM_{10} does actually have a significant effect on admissions to hospital.

Figure 3.14 shows shape of the constrained relative risks against lag for orders zero to four for the model without the 'day of the week' term. At powers one, two and three the relative risks are very close to one, except at lag 14, although this is likely to be an effect of multiple testing rather than a significant result. However we can see that at power four, the relative risk at lag zero is particularly high. This reduces immediately and then starts to rise three days later.

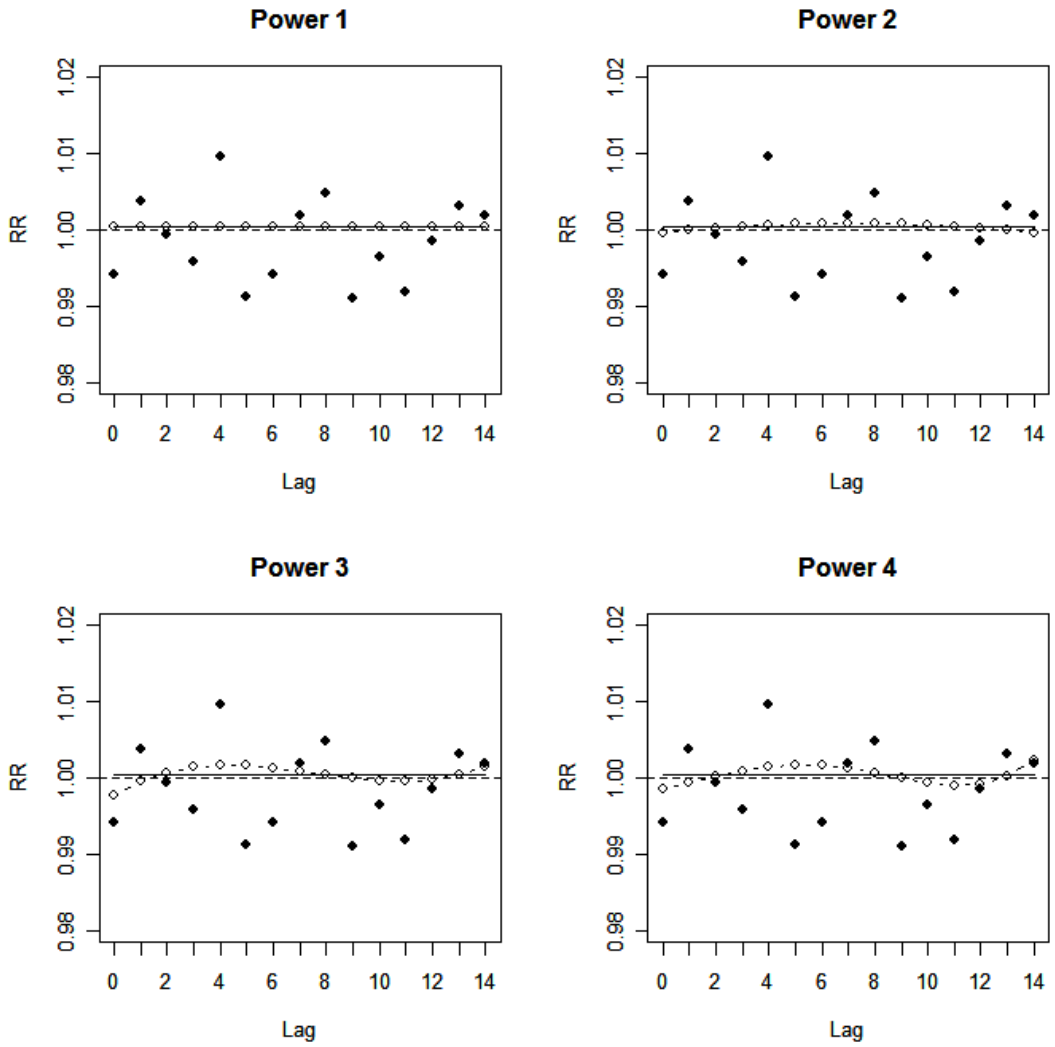

Figure 3.13. The solid line is the relative risk at power zero, the dashed line represents the null risk of 1, the filled diamonds are the relative risks from the unconstrained model and the unfilled diamonds are the relative risks from the distributed lag model

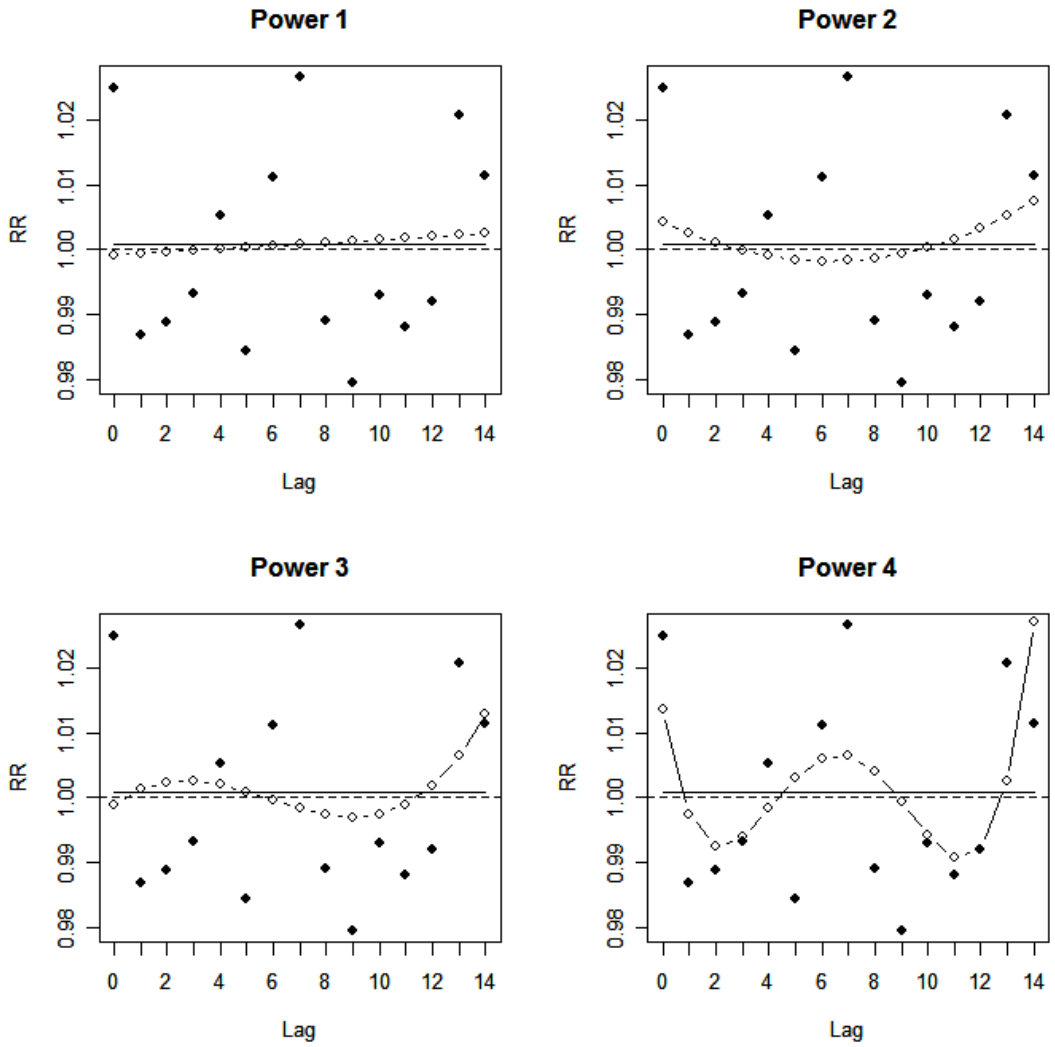

Figure 3.14. The solid line is the relative risk at power zero, the dashed line represents the null risk of 1, the filled diamonds are the relative risks from the unconstrained model and the unfilled diamonds are the relative risks from the distributed lag model

Figure 3.16 shows the relative risks from power four for a week. We can see that this lag structure is very similar to the one suggested by Zanobetti et al. (2002) when discussing the theory of mortality displacement (Figure 3.15), although the relative risk reduces much sooner then Zanobetti suggests. If this lag structure is correct, it could have important implications in Glasgow, as it suggests that admissions to hospital only occur on the same day as exposure to pollution and thus reducing the significance for public health. However Figure 3.14 shows a decrease in relative risk followed by another increase. This may simply be due to multiple testing, although a closer look at the relative risks in Tables 3.5 and 3.6 showed significant positive effects occurred approximately every seven days, which points to the lack of the 'day of the week term' as being a possible reason for the significant results, thus suggesting the theory of mortality displacement does not hold here.

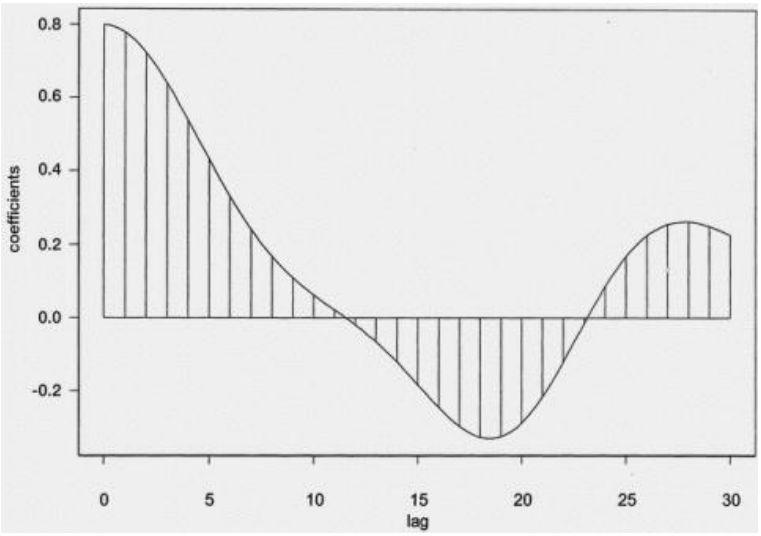

Figure 3.15. Suggested lag structure corresponding the mortality displacement effect (taken from Zanobetti et al. (2002))

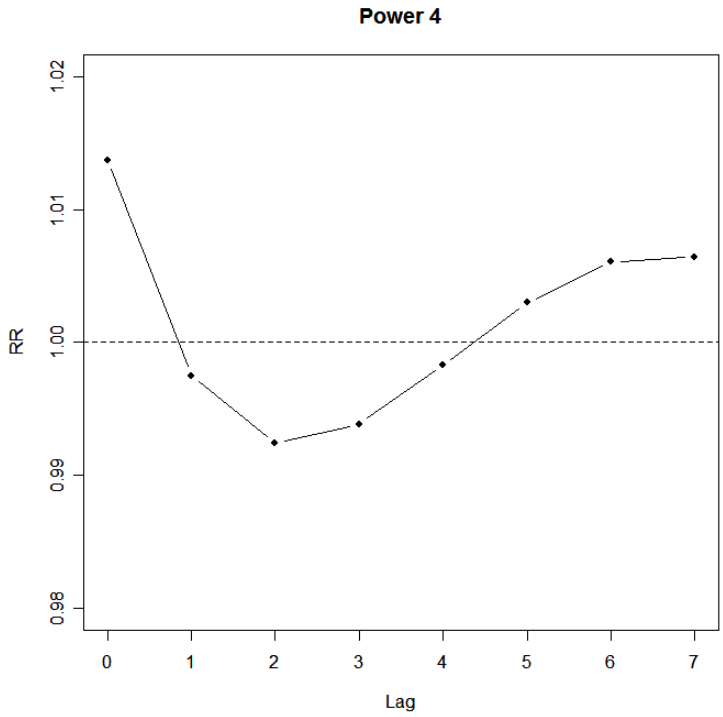

Figure 3.16. Lag structure found when modelling daily admissions to hospital in Glasgow, with no 'day of the week' term

3.3.4 Conclusions

There has been a steady decrease in cardiovascular admissions to hospital in Glasgow between 2000 and 2006, although there does not appear to be much of a seasonal element to the data. Also PM_{10} concentrations appear to have stayed fairly constant over the seven year period, although there were some unusually high concentrations. There does not appear to be any obvious relationship between PM_{10} concentrations and cardiovascular admissions to hospital although there does seem to be a quadratic relationship with temperature. During the analysis, it was discovered that when the 'day of the week' term is included in the model, PM_{10} does not appear to have any significant effect on cardiovascular admissions to hospital, although when it is omitted, PM_{10} becomes significant at some lags. This suggests there an interaction between PM_{10} and day of the week

that we do not yet understand.

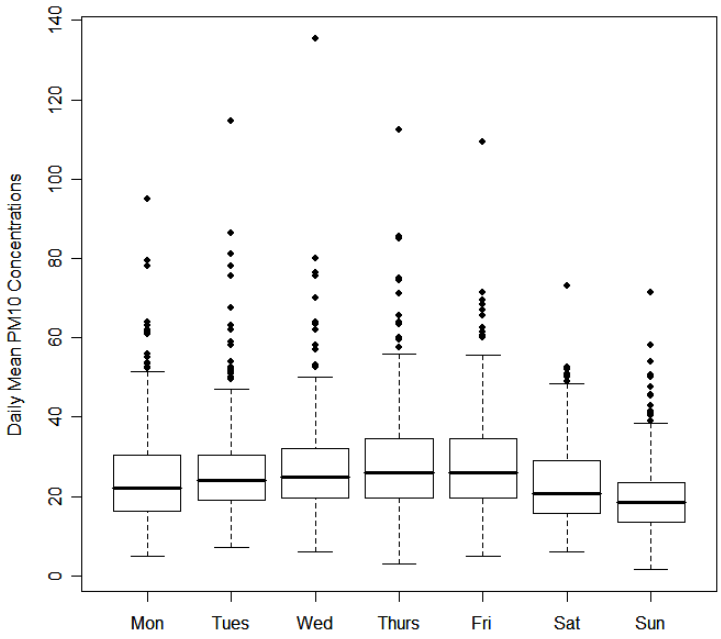

Figure 3.17. Boxplots of mean daily PM_{10} concentrations for each day of the week

Figure 3.17 shows boxplots of mean daily PM_{10} concentrations for each day of the week. We can see that there does seem to be some kind of relationship between PM_{10} concentrations and day of the week, with PM_{10} increasing from Monday to Thursday. On Friday the PM_{10} concentrations appear to be similar to the concentrations on Thursday, before decreasing over the weekend. This relationship makes interpreting whether or not PM_{10} affects admissions to hospital with cardiovascular diseases difficult. However, if PM_{10} does have an effect on cardiovascular admissions to hospital, it seems that only PM_{10} exposure on the day of admission and lags of one and two weeks affects health, which suggests there is no true effect of PM_{10}.

It could be argued that since the the numbers of admissions to hospital do not peak at the same time each year, the method of adding sine and cosine terms to the model is not appropriate for these data, as they are very rigid and do not allow for variation in the location of the peaks. Therefore, to verify these results, the analysis was redone using a different method of removing the trend and seasonality, using natural cubic splines. Many air pollution-mortality studies have used this approach including Dominici et al. (2003) who used seven degrees of freedom per year. However their data are highly seasonal, and as the data in this study have very little seasonality, it makes sense to use fewer degrees of freedom. Therefore 22 degrees of freedom over the seven year period were used, which was chosen by minimising the AIC. After removing the trend and seasonality, PM_{10} was added to the model using the previous single lag and multiple lag models. The results were very similar to those obtained in Tables 3.5 and 3.6, with no significant effect of PM_{10} in the models that contained the 'day of the week' term, and significant effects at lags zero, seven and 14 (with relative risks of approximately 1.021) in the models without the 'day of the week' term. As these results are very similar to those presented earlier, it suggests the estimates are not being confounded by poor seasonality control.

However there are a limited number of admissions to hospital in any one day (mean of 10.5) and so it is possible that any effect of PM_{10} is not being picked up. Therefore the analysis will be re-done using weekly admissions to hospital and mean weekly PM_{10} concentrations. This will also help to eliminate the debate concerning the 'day of the week' term.

3.3.5 Exploratory Analysis - Weekly Admissions

This section aims to estimate the association between mean weekly PM_{10} concentrations and weekly cardiovascular admissions to hospital in Glasgow. We begin with a series of descriptive plots of the data.

Health

Figure 3.18 shows the weekly admissions to hospital due to cardiovascular illness over the seven year period 2000 to 2006, with a smoothed lowess line running through the data to show the underlying trend. We can see that there is a very strong decreasing trend to the data, that appears to be broadly linear, and there are also signs of seasonality.

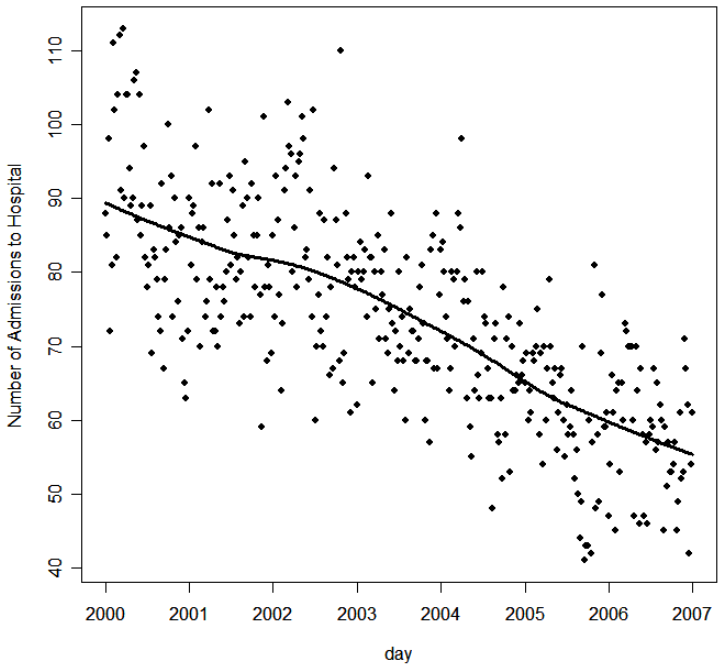

Figure 3.18. Weekly admissions to hospital in Glasgow from 2000 to 2006

Pollution

Figure 3.19 shows the mean weekly PM_{10} concentrations over the seven year period while Figure 3.20 illustrates the relationship between weekly admissions to hospital and mean weekly PM_{10} concentrations.

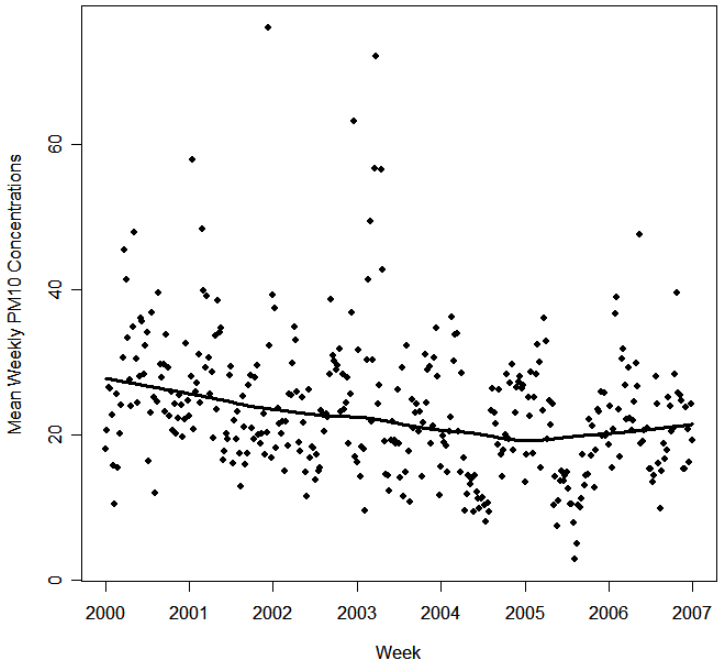

Figure 3.19. Mean weekly PM_{10} concentrations in Glasgow from 2000 to 2006

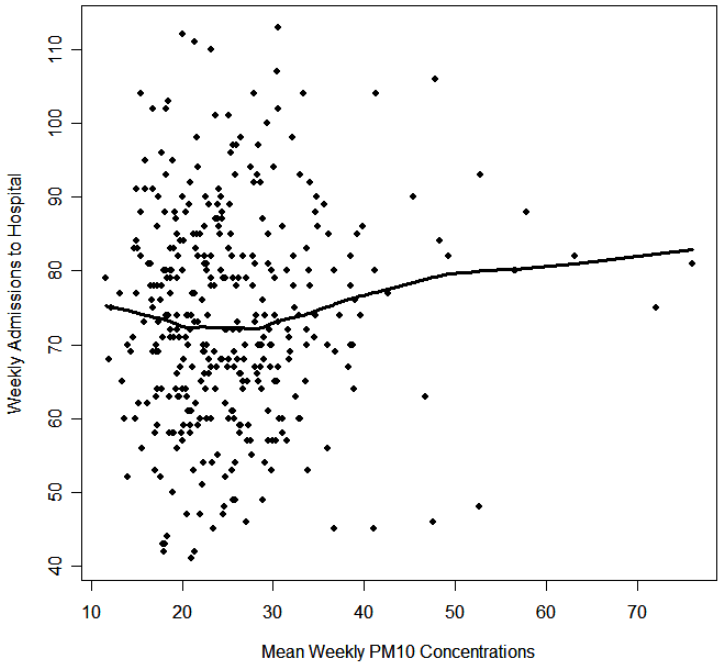

Figure 3.20. Relationship between weekly admissions to hospital and mean weekly PM_{10} concentrations

Figure 3.19 shows a steady decrease in PM_{10} concentrations from 2000 to 2005, after which it starts to rise slightly again. There also appears to be some seasonal variation, with peaks approximately every year. However from Figure 3.20, there does not appear to be any obvious relationship between weekly admissions to hospital and mean weekly PM_{10} concentrations.

Temperature

Figures 3.21, 3.22 and 3.23 below show the plots of mean weekly maximum temperature from 2000 to 2007, weekly admissions to hospital by temperature and mean weekly PM_{10} concentrations against temperature.

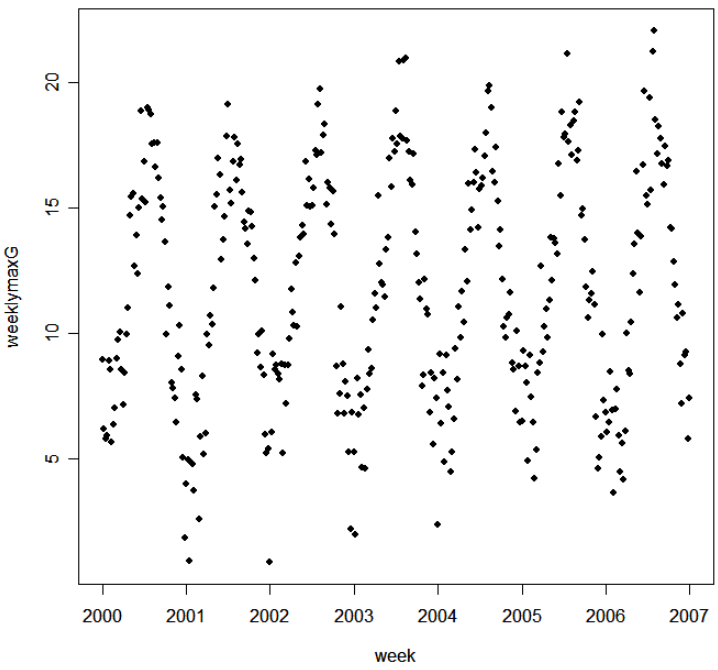

Figure 3.21. Mean weekly maximum temperature in Glasgow from 2000 to 2006

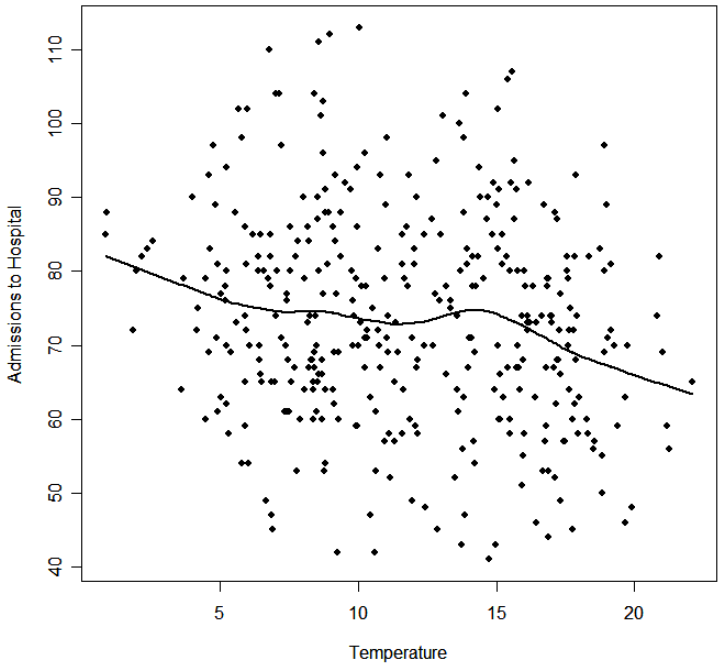

Figure 3.22. Relationship between weekly admissions to hospital and mean weekly maximum temperature

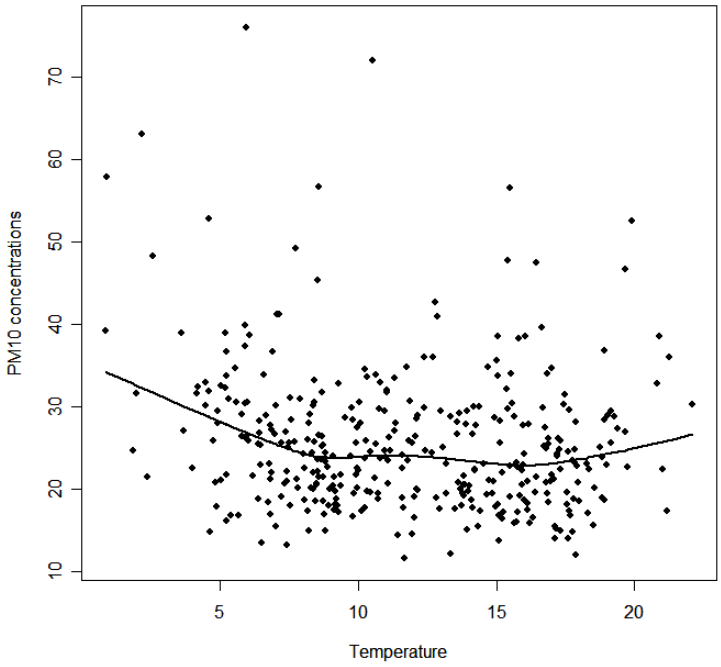

Figure 3.23. Relationship between mean weekly PM_{10} concentrations and mean weekly maximum temperature

Figure 3.21 shows peaks in temperature in summer and troughs in winter, as we would expect. However there does not appear to be any obvious relationship between admissions to hospital and maximum temperature. On the other hand, there appears to be a slight quadratic curve with PM_{10} concentrations decreasing as temperature increases and then decreasing again at higher temperatures.

3.3.6 Covariate Modelling

The next step in the analysis is to produce a model that adequately removes all trends, seasonality and correlation. As weekly admissions to hospital come in the form of counts the assumption of using a Poisson generalised linear model still applies. After following the same method used to analyse daily admissions to hospital, the final model we are left with is Model (3.9)

$$\ln(\mu_t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 \text{Max}_t + \alpha_4 \sin(\frac{\omega t}{2}) + \alpha_5 \cos(\frac{\omega t}{2}) \tag{3.9}$$

where $\omega = \frac{2\pi}{365.25}$ as before. A summary of Model (3.9) can be seen in Table 3.7.

Coefficient	Estimate	Standard Error	P-Value
t	-3.503 x 10^{-04}	2.360 x 10^{-04}	0.1377
t^2	-2.792 x 10^{-06}	6.448 x 10^{-07}	1.49 x 10^{-05}
Max	-0.005014	0.001331	0.0002
$\sin(\frac{\omega}{2})$	0.03249	0.009142	0.0004
$\cos(\frac{\omega}{2})$	0.02216	0.00858	0.0098
AIC			2671.5

Table 3.7. Summary of Model (3.9)

Adding the sine and cosine terms makes t insignificant, however as t^2 is still significant, t should be kept in the model. Figure 3.24 displays the residuals for this model and Figure 3.25 shows the autocorrelation function for the residuals. The residuals appear to be scattered evenly about zero with no obvious pattern and within the autocorrelation function, most of the lags are within the 95% confidence bands. Therefore we can assume that most of the correlation within

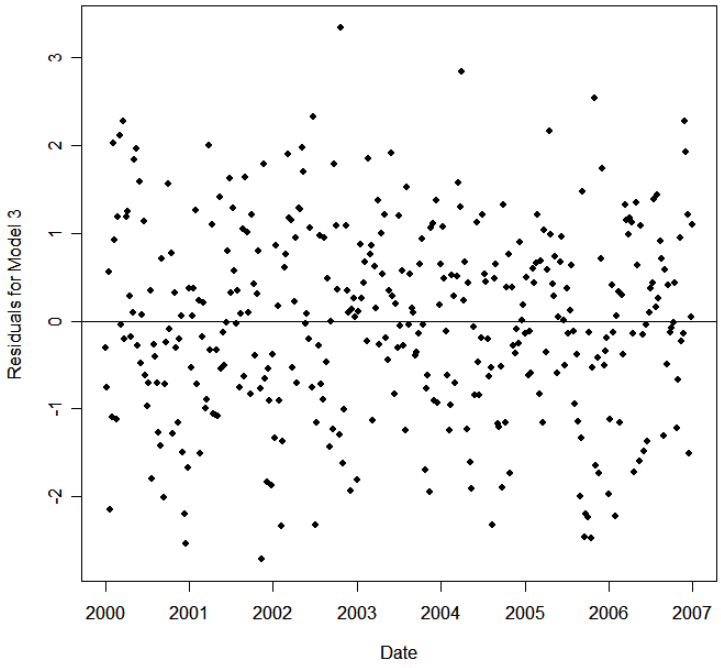

Figure 3.24. Residuals for Model (3.9)

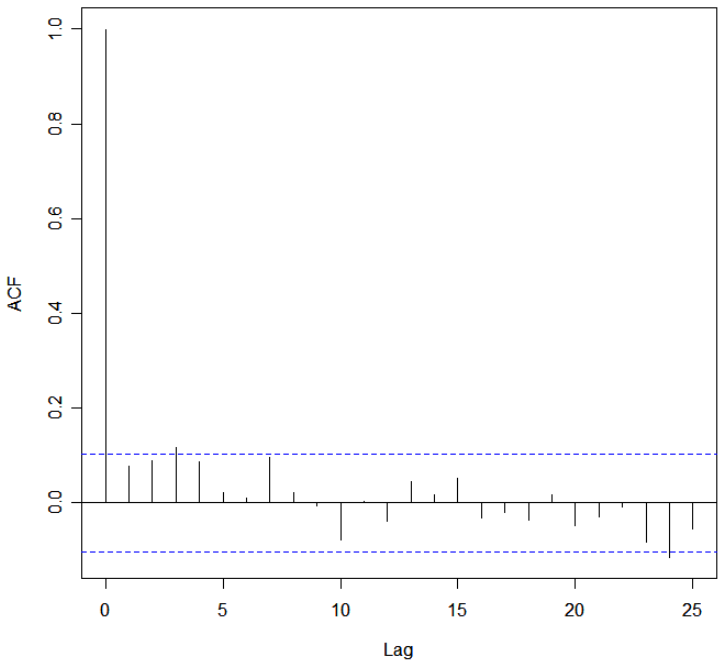

Figure 3.25. Autocorrelation function for the residuals

the residuals has now been removed. Also the AIC is lower than the AIC from the previous models, which ranged from 2687.1 to 2707.8, suggesting this is the best fitting model to date and so we can now continue with the analysis.

3.3.7 Pollution Modelling

The next step is to add mean weekly PM_{10} concentrations to the model to estimate its association with weekly admissions to hospital. This will be done using two different models. The first will be a series of single lag models ranging from zero to eight weeks before admission to hospital. For a given lag q, this model is seen as

$$\begin{aligned} \ln(\mu_t) &= \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 \text{Max}_t + \alpha_4 \sin(\frac{\omega t}{2}) + \alpha_5 \cos(\frac{\omega t}{2}) \\ &+ \gamma \text{PM}_{10_{t-q}} \end{aligned} \tag{3.10}$$

The relative risks and 95% confidence intervals associated with this model will be shown in Table 3.8. However there may be an accumulated effect of PM_{10} that is not being picked up by the single lag models. In addition, it is unlikely that the effects of air pollution are unrelated to pollution concentrations on adjacent weeks, meaning that the effects may be confounded by concentrations at a different lag. These problems can be overcome using a multiple lag model which includes PM_{10} at all the lags. This gives Model (3.11)

$$\begin{aligned} \ln(\mu_t) &= \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 \text{Max}_t + \alpha_4 \sin(\frac{\omega t}{2}) + \alpha_5 \cos(\frac{\omega t}{2}) \\ &+ \sum_{q=0}^{8} \gamma_q \text{PM}_{10_{t-q}} \end{aligned} \tag{3.11}$$

The relative risks and 95% confidence intervals associated with this model can be seen in Table 3.9. In Table 3.8, the relative risks from the single lag models are very close to one and non-significant at the 5% level meaning there does not appear to be evidence of a significant relationship between weekly admissions to hospital and mean weekly PM_{10} concentrations. Table 3.9 shows the relative

Lag	RR	95% CI
0	0.9990	(0.9855, 1.0129)
1	0.9877	(0.9741, 1.0016)
2	0.9957	(0.9820, 1.0095)
3	1.0014	(0.9877, 1.0153)
4	0.9990	(0.9853, 1.0128)
5	0.9975	(0.9838, 1.0113)
6	1.0054	(0.9918, 1.0193)
7	1.0019	(0.9882, 1.0159)
8	0.9967	(0.9830, 1.0106)

Table 3.8. Relative risks from the single lag models

Lag	RR	95% CI
0	1.0034	(0.9885, 1.0185)
1	0.9885	(0.9729, 1.0043)
2	0.9987	(0.9829, 1.0148)
3	1.0007	(0.9847, 1.0170)
4	1.0013	(0.9853, 1.0176)
5	0.9984	(0.9824, 1.0146)
6	1.0059	(0.9901, 1.0220)
7	0.9998	(0.9842, 1.0156)
8	0.9965	(0.9816, 1.0115)

Table 3.9. Relative risks from the multiple lag model

risks from the multiple lag model. Again these relative risks are very close to one and and are non-significant at the 5% level. However this model suffers from the problem of collinearity. To overcome this problem, a distributed lag model will be used to constrain the PM_{10} estimates to follow a polynomial. This gives Model (3.12)

$$\begin{aligned} \ln(\mu_t) = \beta_0 &+ \beta_1 t + \beta_2 t^2 + \beta_3 \text{Max}_t + \beta_4 \sin(\frac{\omega t}{2}) + \beta_5 \cos(\frac{\omega t}{2}) \\ &+ \sum_{q=0}^{8} \gamma_q \text{PM}_{10_{t-q}} \qquad (3.12) \\ \text{where} \quad \gamma_q &= \sum_{k=0}^{q} \eta_k q^k, \quad q = 0, ..., 8 \end{aligned}$$

The AICs for this model range from 2023.7 to 2030.0 which are much lower than the AIC from the single lag models, which ranged from 2612.3 to 2673.5 and the AIC from the multiple lag model which was 2608.4, which suggests the distributed lag model is a better fit of the data. Figure 3.26 shows the shape of the relative risks against lag for orders zero to four, where the solid horizontal line is the relative risk at power zero, which gives the overall estimated effect of PM_{10}.

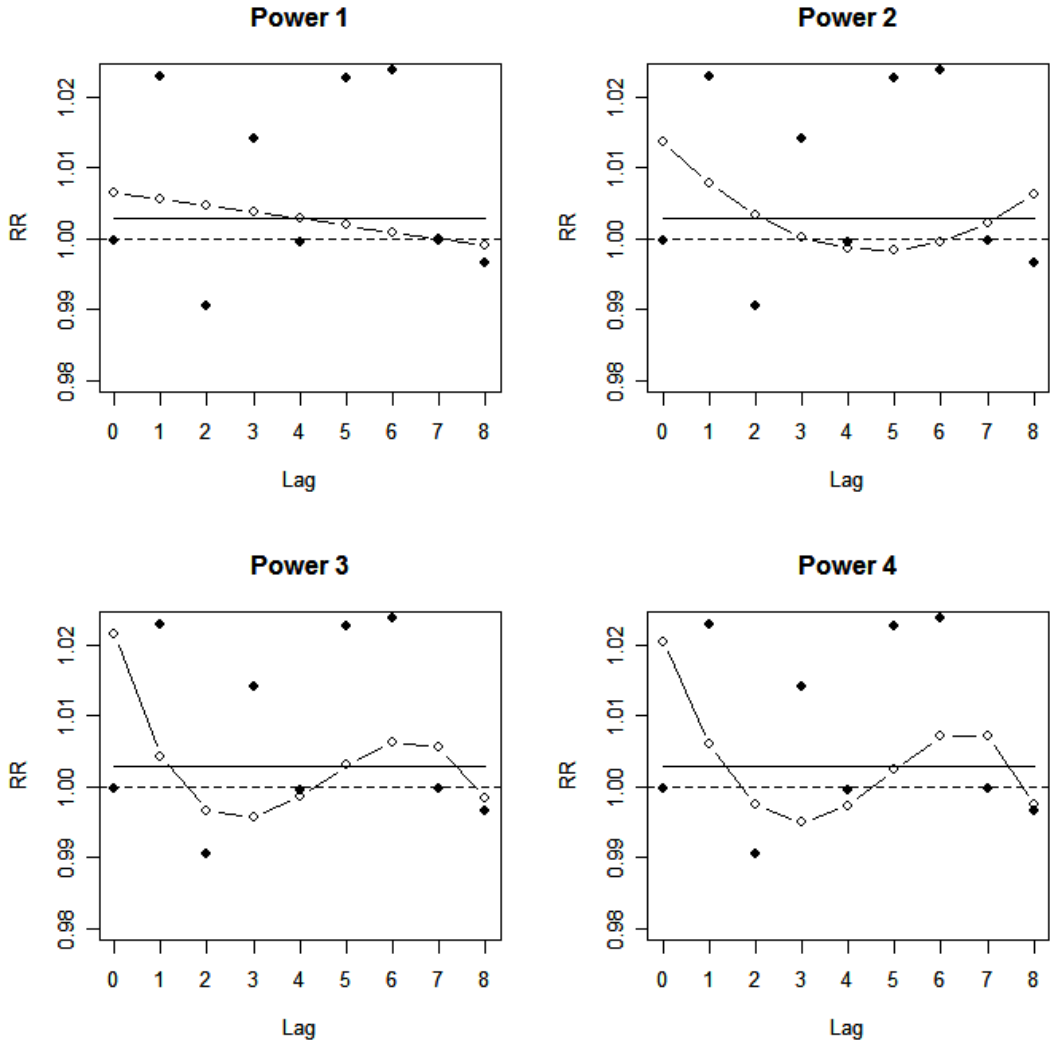

Figure 3.26. The solid line is the relative risk at power zero, the dashed line represents the null risk of 1, the filled diamonds are the relative risks from the unconstrained model and the unfilled diamonds are the relative risks from the distributed lag model

We can see that at powers three and four the relative risk at lag zero is quite high at approximately 1.02. However this reduces immediately and within two weeks PM_{10} apparently has a positive effect on admissions to hospital. If this were true, it would again suggest that the theory of mortality displacement is true. However Table 3.10 refutes this belief as none of the p-values for the relative risk coefficients are significant.

Coefficient	Estimate	Standard Error	P-Value
Intercept	1.717	0.3265	1.45 x 10^{-07}
t	0.04094	0.01206	0.0007
t^2	-1.051 x 10^{-04}	3.178 x 10^{-05}	0.0009
Max	0.001638	0.006311	0.7952
$\sin(\frac{\omega}{2})$	-3.3	1.041	0.0015
$\cos(\frac{\omega}{2})$	0.4796	0.2174	0.0273
η_0	0.002125	0.001531	**0.1652**
η_1	-0.002252	0.001931	**0.2436**
η_2	5.968 x 10^{-04}	5.887 x 10^{-04}	**0.3107**
η_3	-4.388 x 10^{-05}	4.844 x 10^{-05}	**0.3650**
AIC			2028.1

Table 3.10. Results from Model (3.12) using a power of three

3.3.8 Conclusions

In conclusion, weekly admissions to hospital have decreased sharply since 2000, with some signs of seasonality. Mean weekly PM_{10} concentrations also decreased between 2000 and 2005 before starting to rise again, but there does not appear to be any obvious relationship between the two variables. After removing as much of the trend and seasonality as possible, PM_{10} lagged between zero and eight weeks was added to the model, initially using a series of single lag models. However as there may have been an accumulated effect of PM_{10} that was not picked up by these model, a multiple lag model was next produced. Unfortunately this model has the problem of collinearity due to the PM_{10} concentrations being highly correlated. Therefore a distributed lag model was used, which constrains the PM_{10} estimates to follow a polynomial, thus reducing collinearity. The results found here imply there is no significant relationship between air pollution and weekly cardiovascular admissions to hospital in Glasgow. As this analysis was performed on the weekly level, it also eliminated any concerns regarding the 'day of the week' terms. However the pollution data used in this analysis was for mean weekly PM_{10} concentrations and therefore it may be worth redoing this analysis using maximum weekly PM_{10} values, rather than mean weekly values, as

any effects of air pollution are more likely to be seen when the concentrations are higher. After re-doing this analysis, there was still no significant effect of PM_{10} found. In the next section, the associations between PM_{10} concentrations and admissions to hospital in Edinburgh will be discussed.

3.4 Edinburgh

3.4.1 Exploratory Analysis

The exploratory analysis will show plots of daily admissions to hospital, daily mean PM_{10} concentrations and daily maximum temperatures.

Health

Figure 3.27 below shows the daily number of admissions to hospital for Edinburgh between 2000 and 2006, with a smoothed lowess line running through the plot to show the underlying trend of the data.

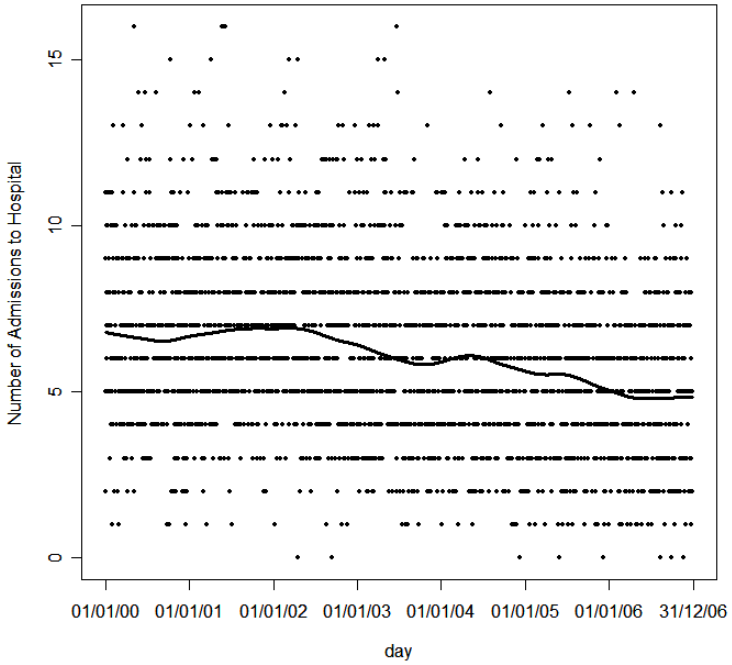

Figure 3.27. Daily hospital admissions in Edinburgh from 2000 to 2006

The figure shows an approximately linear decreasing trend in admissions to hospital over the seven years with little sign of any seasonal variation.

Pollution

Figure 3.28 shows the plots of mean weekly PM_{10} concentrations between 2000 and 2006, while Figure 3.29 shows the relationship between weekly admissions to hospital against mean weekly PM_{10} concentrations. In Figure 3.28, the first thing we notice is the large amount of missing data in 2003, which will be ignored in the meantime. The underlying trend appears to have decreased slightly over time and there is one especially high value in 2005 with a few other high values (concentrations of around 100 $\mu g/m^3$) in 2000 and 2001.

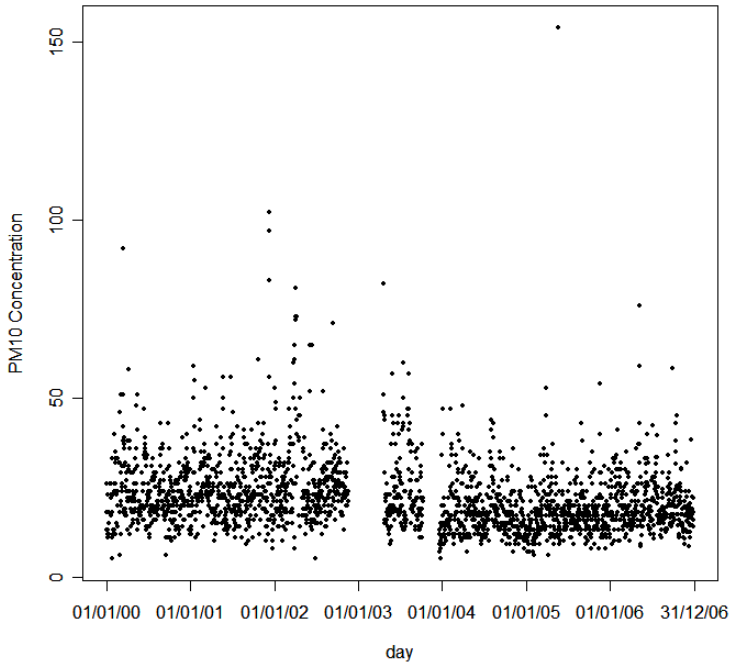

Figure 3.28. Daily mean PM_{10} concentrations in Edinburgh from 2000 to 2006

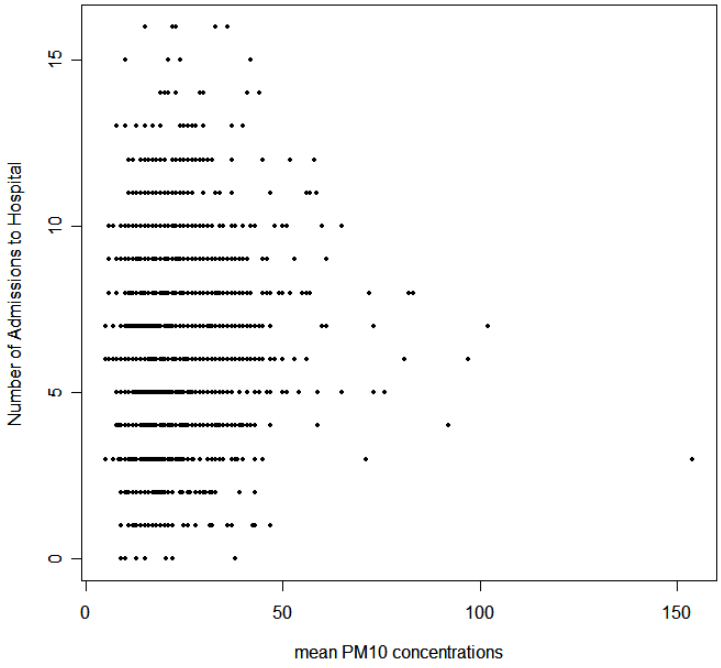

Figure 3.29. Relationship between daily admissions to hospital and mean daily PM_{10} concentrations

There also seem to be some signs of seasonality. However from Figure 3.29 there does not appear to be any relationship between PM_{10} concentrations and daily admissions to hospital in Edinburgh.

Temperature

Figures 3.30, 3.31 and 3.32 show the plots of daily maximum temperature, the first of which is over time, the second against admissions to hospital, and the third against daily mean PM_{10} concentrations.

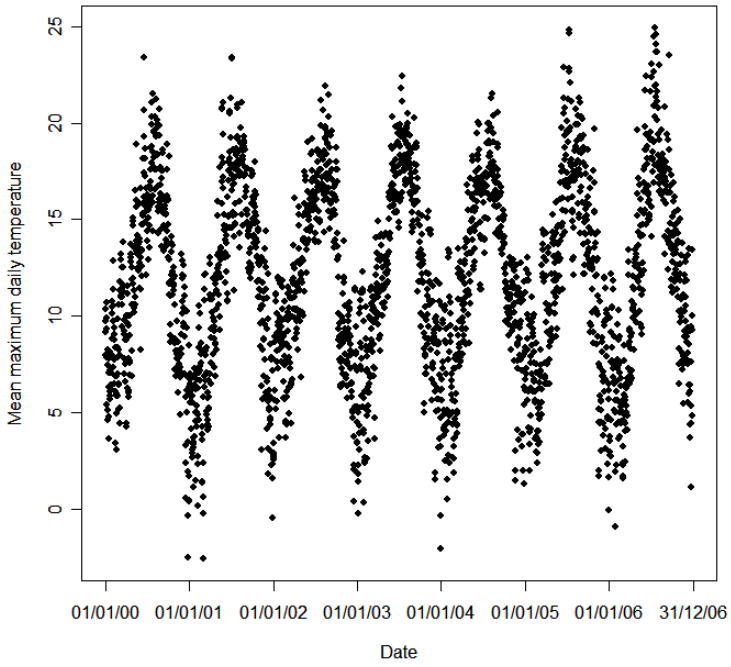

Figure 3.30. Daily maximum temperatures in Edinburgh between 2000 and 2006

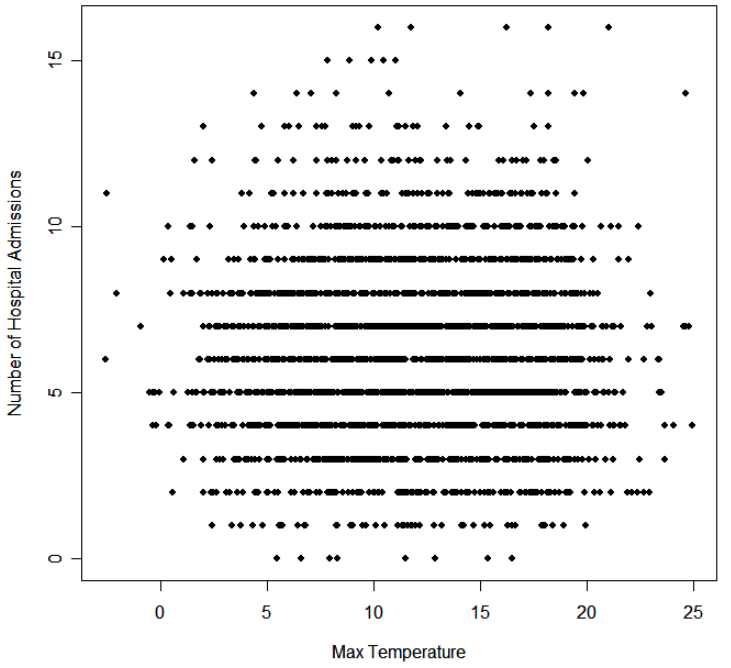

Figure 3.31. Relationship between daily admissions to hospital and daily maximum temperature

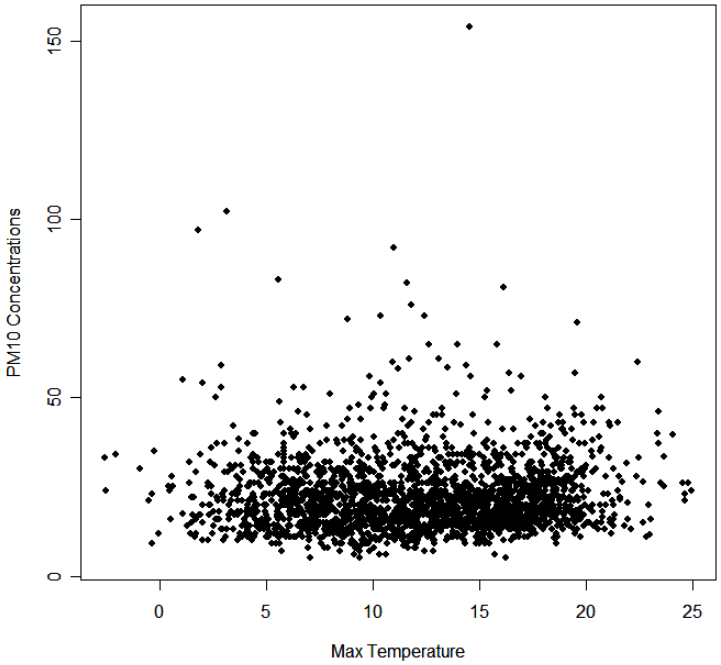

Figure 3.32. Relationship between daily mean PM_{10} concentrations and daily maximum temperature

Figure 3.30 shows the seasonality that we would expect, with higher temperatures in summer and lower temperatures in winter. From Figure 3.31, we can see that there does not appear to be any relationship between temperature and admissions to hospital in Edinburgh. In addition, from Figure 3.32 there does not appear to be any relationship between PM_{10} concentrations and temperature.

3.4.2 Covariate Modelling

The next step is to produce a model that adequately removes any trend, seasonal variation and correlation within the daily admissions. As admissions to hospital come in the form of daily counts, models will use the Poisson generalised linear model used in the previous section. The model produced was Model (3.13)

$$\ln(\mu_t) = \alpha_0 + \alpha_1 t + \sum_{j=1}^{6} \alpha_{j+1} \mathrm{DoW}_t^{(j)} \tag{3.13}$$

where $\text{DoW}_t^{(j)}$ is an indicator variable for day of the week. A summary of Model 3.13 is shown in Table 3.11.

Coefficient	Estimate	Standard Error	P Value
Intercept	2.110	0.02377	$<2 \times 10^{-16}$
t	-0.0001452	1.081×10^{-05}	$<2 \times 10^{-16}$
tues	-0.05249	0.02857	0.0662
wed	-0.06711	0.02868	0.0193
thurs	-0.06995	0.02870	0.0148
fri	-0.04163	0.02850	0.1440
sat	-0.2897	0.030452	$<2 \times 10^{-16}$
sun	-0.2727	0.03030	$<2 \times 10^{-16}$
AIC			11904

Table 3.11. Results from model 3.13

We can see that Saturday and Sunday are very significant in the model. Wednesday and Thursday are also significant, but Tuesday and Friday are not. However all the 'day of the week' terms must be kept in the model. Figure 3.33 shows the residuals for this model and Figure 3.34 displays the autocorrelation function for the residuals. From Figure 3.33 we can see that the data does not appear to have much seasonality left in the model and in Figure 3.34 most of the lags are within the 95% confidence bands, suggesting that much of the correlation within the residuals has now been removed and the model appears to be adequate. Also the AIC is lower than the AIC for the model containing only a linear function of time (12066), suggesting this is the best fitted model thus far.

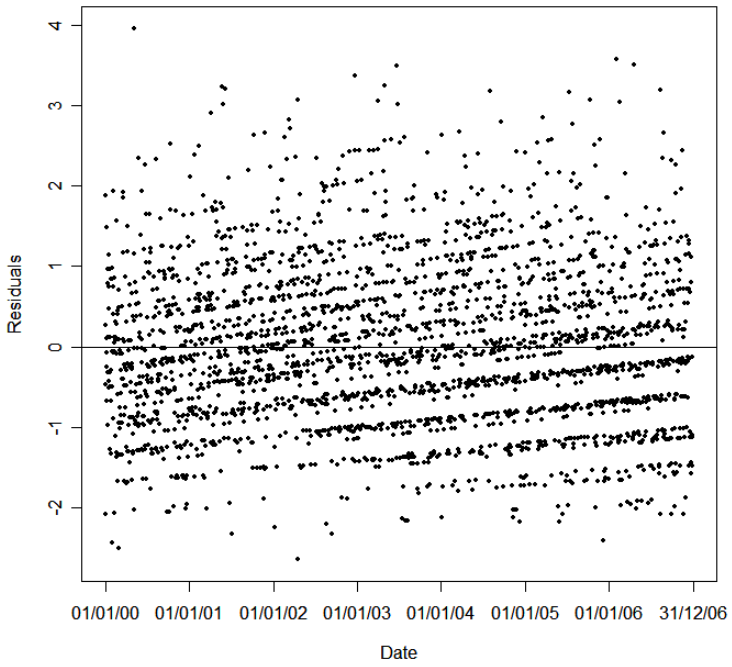

Figure 3.33. Plot of residuals for Model (3.13)

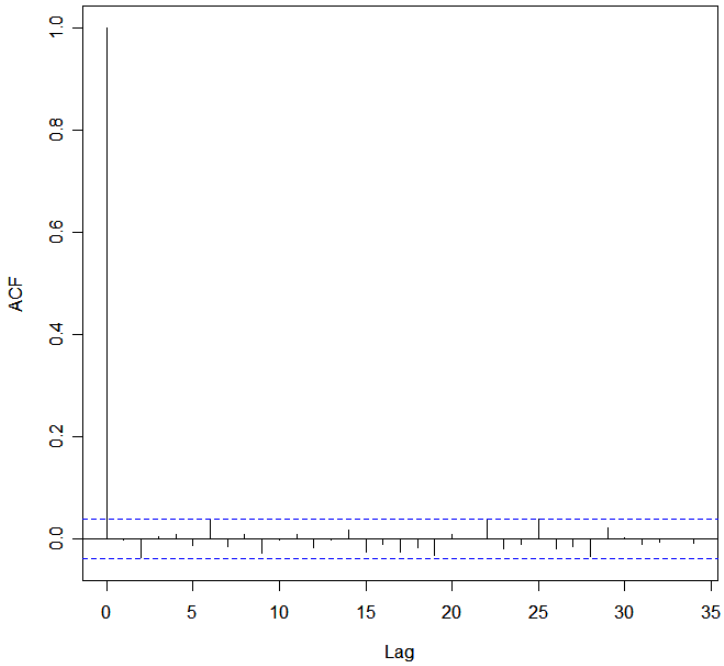

Figure 3.34. Autocorrelation function from the residuals

3.4.3 Pollution Modelling

The next step is to add PM_{10} to the model to estimate its association with admissions to hospital. We begin with a series of single lag models ranging from zero to fourteen days (Model (3.14)).

$$\ln(\mu_t) = \alpha_0 + \alpha_1 t + \sum_{j=1}^{6} \alpha_{j+1} \mathrm{DoW}_t^{(j)} + \gamma \mathrm{PM}_{10_{t-q}} \tag{3.14}$$

Table 3.12 shows the relative risks and associated 95% confidence interval at each lag. However, as previously mentioned, there may be an accumulated effect of PM_{10} that is not seen from a single lag model and it is unlikely that the the effect of air pollution on any one day is unrelated to pollution concentrations from adjacent days. Therefore a multiple lag model containing lags from the previous 14 days would be a sensible next step. This gives model 3.15.

$$\ln(\mu_t) = \alpha_0 + \alpha_1 t + \sum_{j=1}^{6} \alpha_{j+1} \mathrm{DoW}_t^{(j)} + \sum_{q=0}^{14} \gamma_q \mathrm{PM}_{10_{t-q}} \tag{3.15}$$

Table 3.13 shows the relative risks and the 95% confidence interval at each lag. In Table 3.12, the relative risks are all very close to one and the confidence intervals all contain one, meaning they are non-significant at the 5% level. Therefore, there does not appear to be a relationship between PM_{10} concentrations and cardiovascular admissions to hospital in Edinburgh. Table 3.13 shows the relative risks from the multiple lag model. We see that the relative risks are again quite close to one and all the confidence intervals contain one, again meaning they are not significant at the 5% level, again suggesting there is no significant effect of PM_{10} on cardiovascular admissions to hospital in Edinburgh. However these intervals are wider than the confidence intervals from the single lag model, which is most likely due to collinearity caused by the correlated pollution variables. To try to remove this problem, the same distributed lag model as used previously in the Glasgow analysis will be applied here, giving Model (3.16).

Lag	RR	95% CI
0	1.0026	(0.9859, 1.0197)
1	1.0077	(0.9910, 1.0247)
2	1.0075	(0.9907, 1.0246)
3	1.0053	(0.9885, 1.0224)
4	0.9918	(0.9749, 1.0090)
5	0.9891	(0.9722, 1.0063)
6	0.9944	(0.9774, 1.0116)
7	1.0021	(0.9853, 1.0192)
8	0.9980	(0.9811, 1.0151)
9	0.9901	(0.9732, 1.0074)
10	0.9844	(0.9674, 1.0016)
11	0.9954	(0.9785, 1.0126)
12	0.9924	(0.9756, 1.0096)
13	1.0053	(0.9884, 1.0224)
14	1.0105	(0.9938, 1.0276)

Table 3.12. Relative risks from the single lag models

Lag	RR	95% CI
0	0.9998	(0.9767, 1.0235)
1	0.9972	(0.9712, 1.0239)
2	1.0047	(0.9790, 1.0310)
3	1.0123	(0.9890, 1.0361)
4	0.9892	(0.9660, 1.0130)
5	0.9964	(0.9729, 1.0204)
6	0.9876	(0.9639, 1.0119)
7	1.0122	(0.9887, 1.0361)
8	0.9917	(0.9686, 1.0153)
9	0.9953	(0.9721, 1.0192)
10	0.9858	(0.9625, 1.0096)
11	0.9993	(0.9759, 1.0233)
12	0.9867	(0.9633, 1.0106)
13	1.0082	(0.9848, 1.0322)
14	1.0138	(0.9924, 1.0356)

Table 3.13. Relative risks from the multiple lag model

$$\ln(\mu_t) = \alpha_0 + \alpha_1 t + \sum_{j=1}^{6} \alpha_{j+1} \mathrm{DoW}_t^{(j)} + \sum_{q=0}^{14} \gamma_q \mathrm{PM}_{10_{t-q}} \qquad (3.16)$$

$$\text{where} \quad \gamma_q = \sum_{k=0}^{q} \eta_k q^k, \quad q = 0, ..., 14$$

The AICs from the above model are approximately 8446 for powers zero to four, which is much lower than the AICs from the single lag models (10443 to 10494) or the AIC from the multiple lag model (8487.7) thus suggesting the distributed lag model is the best fitting. Again we are looking for a power that minimises collinearity, but whose relative risks give a plausible representation of the lagged effects of PM_{10}. Figure 3.35 shows the shape of the constrained relative risks at each lag for powers zero to four.

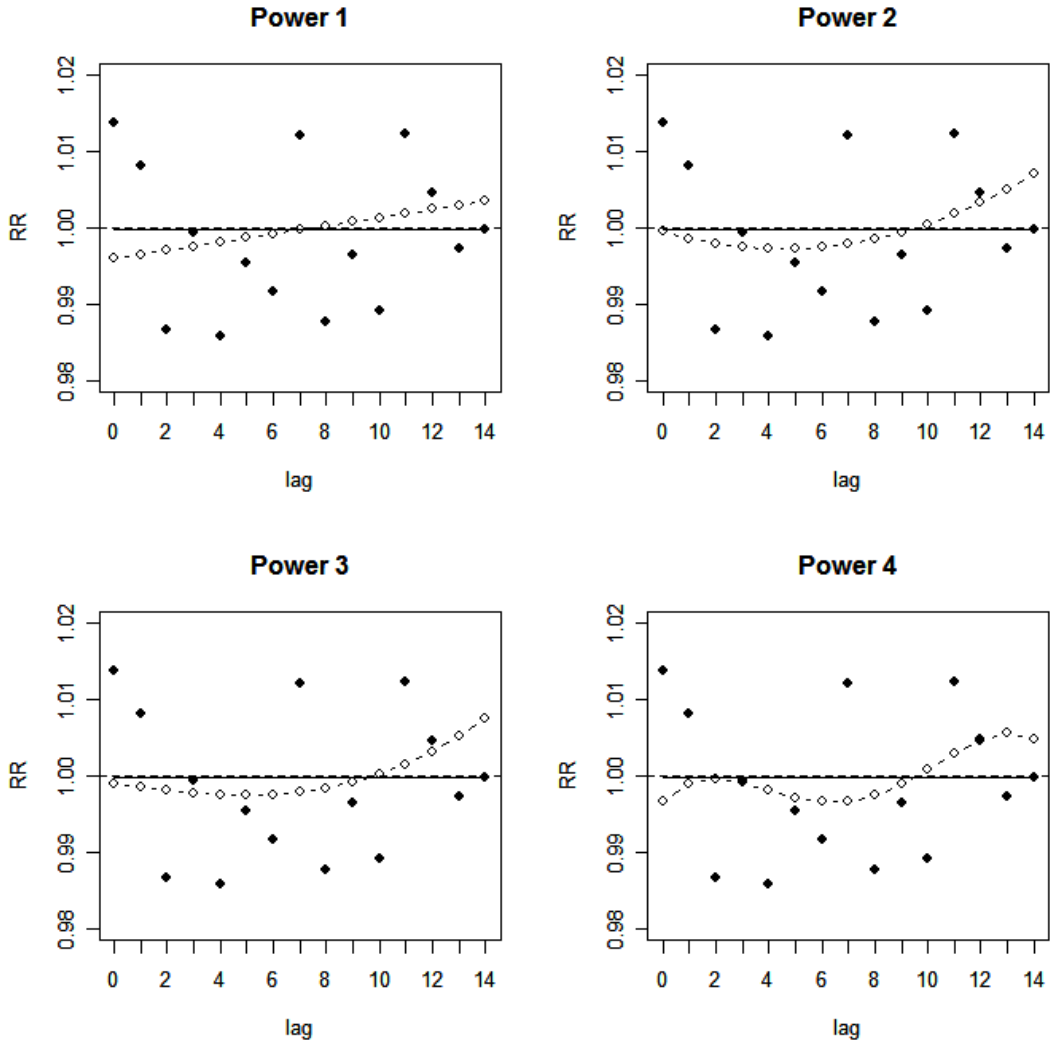

Figure 3.35. The solid line is the relative risk at power zero, the dashed line represents the null risk of 1, the filled diamonds are the relative risks from the unconstrained model and the unfilled diamonds are the relative risks from the distributed lag model

The solid horizontal line in the plots indicates the relative risk when the power is zero, which gives the overall estimated effect of PM_{10} on cardiovascular admissions to hospital. We can see that the the shape of the relative risks from the distributed lag models are very different from those found for the Glasgow data. In particular, at powers two and three the shapes are not biologically plausible. It seems unlikely that admissions to hospital decrease for the first few days after

exposure to PM_{10} and a week later increase again. We can see that the shape of the relative risks are similar at powers two and three. Also, the relative risks are always fairly close to one, thus implying there is no significant effect of exposure to PM_{10} on admissions to hospital with cardiovascular diseases.

3.4.4 Conclusions

Daily admissions to hospital in Edinburgh have been decreasing approximately linearly since 2000. PM_{10} concentrations also appear to have decreased although it is difficult to tell. However there does not appear to be any obvious relationship between admissions to hospital and PM_{10} concentrations. After removing as much of the trend and cyclicity as possible, PM_{10} lagged between zero and 14 days was added to the model, initially using a series of single lag models, then a multiple lag model. However as the multiple lag model suffers from collinearity, a distributed lag model was also used. From this, there does not appear to be any significant effect of exposure to PM_{10} on daily admissions to hospital in Edinburgh with cardiovascular illnesses. However, it is possible that due to the small number of admissions in any one day (mean of 6.2), a significant effect may not have been detected. Therefore the next step is to repeat this analysis with the admissions to hospital aggregated up to a weekly level and using mean weekly PM_{10} concentrations.

3.4.5 Exploratory Analysis - Weekly Admissions

The first step is to produce descriptive plots of weekly data in Edinburgh. These will include weekly admissions to hospital with cardiovascular illnesses, mean weekly PM_{10} concentrations and mean weekly maximum temperatures.

Health

Figure 3.36 shows the numbers of weekly admissions to hospital with a cardiovascular illness in Edinburgh between 2000 and 2006. We can see that between 2000

and 2002, the number of admissions increased slightly before decreasing sharply from 2002 onwards. There also appear to be some signs of seasonality, with peaks every two years in 2002, 2004 and 2006.

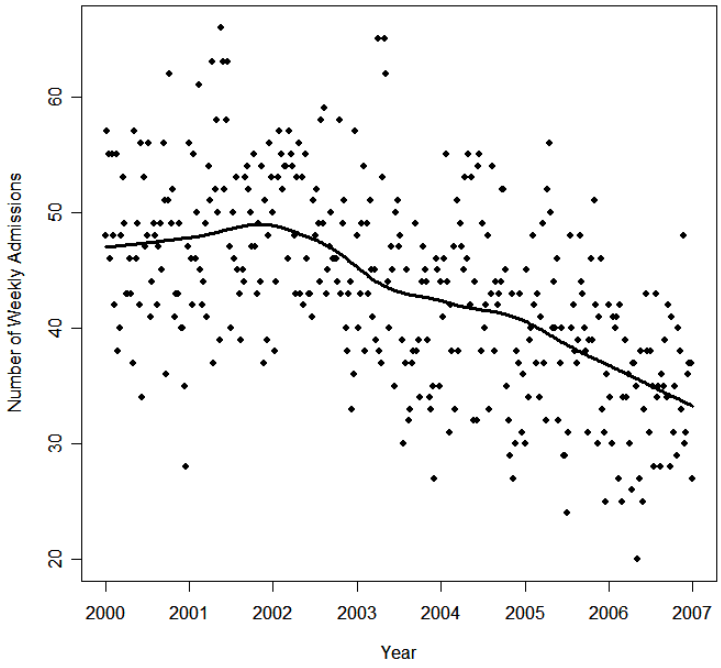

Figure 3.36. Weekly hospital admissions in Edinburgh from 2000 to 2006

Pollution

Since the health data for Edinburgh needed to be aggregated up to weekly admissions, it will also be necessary to aggregate up the mean daily PM_{10} concentrations to mean weekly concentrations. Figures 3.37 and 3.38 below show plots of mean weekly PM_{10} concentrations, the former over time and the latter against weekly hospital admissions.

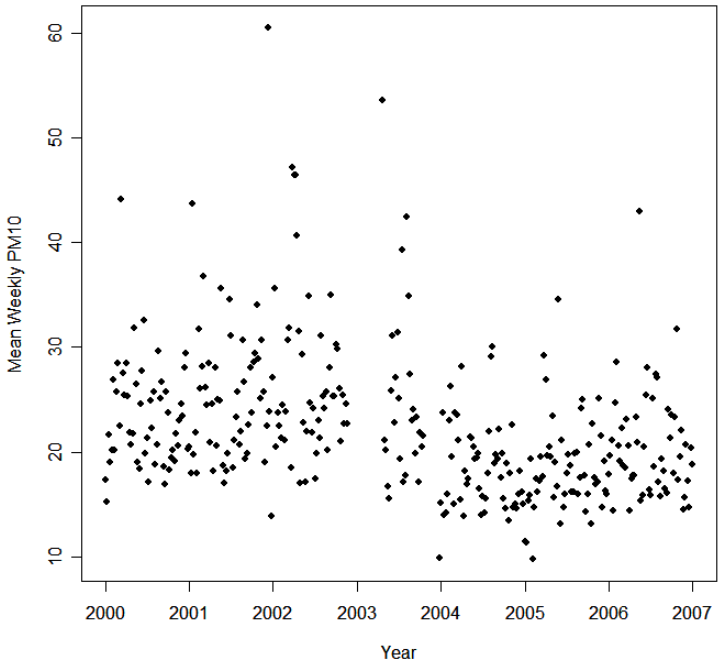

Figure 3.37. Mean weekly PM_{10} concentrations in Edinburgh

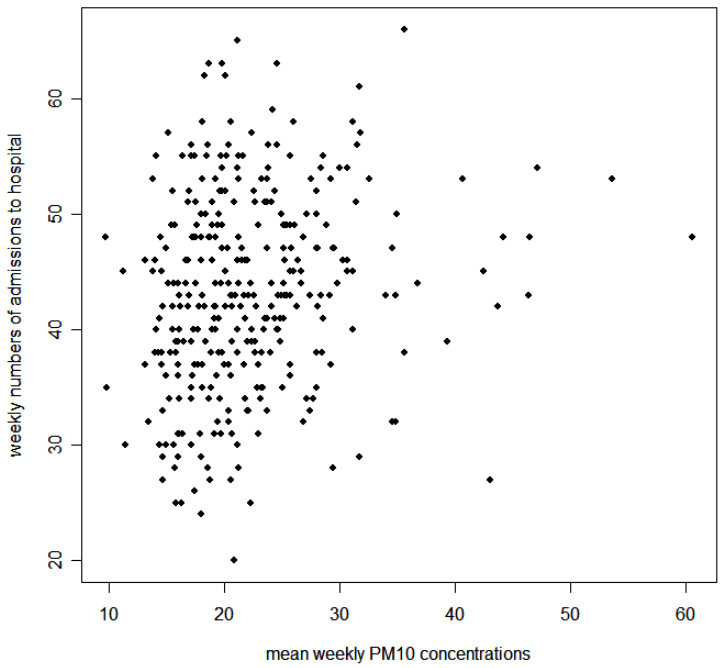

Figure 3.38. Relationship between weekly admissions to hospital and mean weekly PM_{10} concentrations

Figure 3.37 shows the change in mean weekly PM_{10} concentrations between 2000 and 2006. Pollution levels increased slightly from 2000 to 2002 when they started decreasing. However in 2005 the concentrations started rising again. There is quite a lot of missing data in 2003 and during the winter of 2003/04 so at that point it is difficult to interpret the graph. However the other years seem to show signs of seasonality. From Figure 3.38, there does not appear to be any relationship between mean weekly PM_{10} concentrations and weekly admissions to hospital.

Temperature

Figures 3.39, 3.40 and 3.41 show plots of the mean weekly maximum temperatures in Edinburgh, the first plot over time, the second against weekly admissions to hospital and the last against mean weekly PM_{10} concentrations.

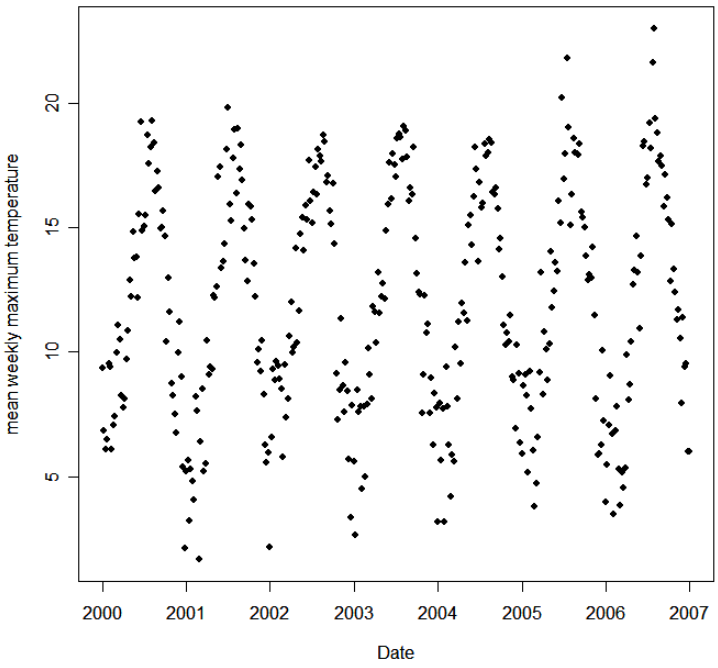

Figure 3.39. Mean weekly maximum temperature in Edinburgh from 2000 to 2006

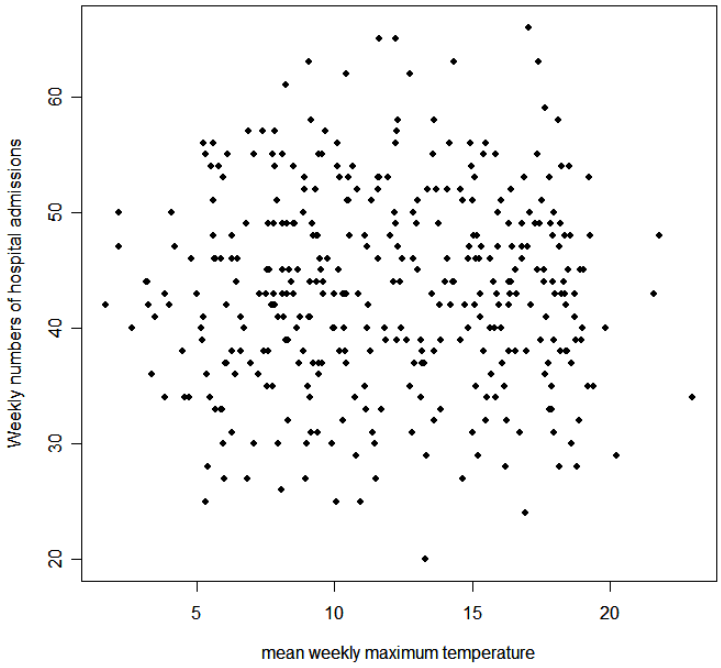

Figure 3.40. Relationship between weekly hospital admissions and mean weekly maximum temperature

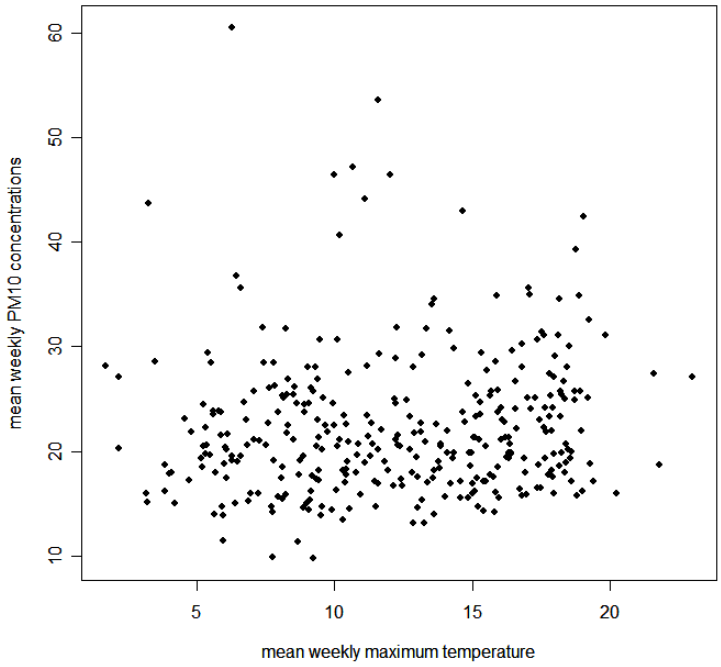

Figure 3.41. Relationship between mean weekly PM_{10} concentrations and mean weekly maximum temperature

From Figure 3.39, we can see that temperature is highly seasonal as expected, with higher values in summer and lower values in winter. From Figure 3.40 there does not appear to be any significant relationship between mean temperature and admissions to hospital with a cardiovascular disease in Edinburgh while Figure 3.41 shows a slight quadratic curve, with PM_{10} concentrations increasing at temperatures below approximately 10°C and above approximately 15°C.

3.4.6 Covariate Modelling

The next step is to adequately model the trend, seasonality and correlation in the data. Once again, a Poisson generalised linear model will be used. From the exploratory analysis, we saw that the smoothed line in the admissions to hospital seemed to increase slightly for the first two years, before showing an approximately linear decreasing trend. Therefore the usual parametric methods cannot be used here. Also, the cyclicity that can be seen is not regular enough to allow the use of sinusoidal curves. Instead natural cubic splines with 23 degrees of freedom will be used to model the trend and any seasonal variation. This value was chosen, as it minimised the AIC, thus giving Model (3.17)

$$\ln(\mu_t) = \alpha_0 + S_1(t; 23) \tag{3.17}$$

Figure 3.42 shows the residuals from Model (3.17) and Figure 3.43 shows the autocorrelation function for the residuals. From Figure 3.42, there does not appear to be any trend or seasonality left within the residuals and from Figure 3.43, most of the lags in the autocorrelation function are within the 95% confidence bands. Therefore it is feasible to suggest that the residuals now resemble white noise and so we can now proceed with the analysis.

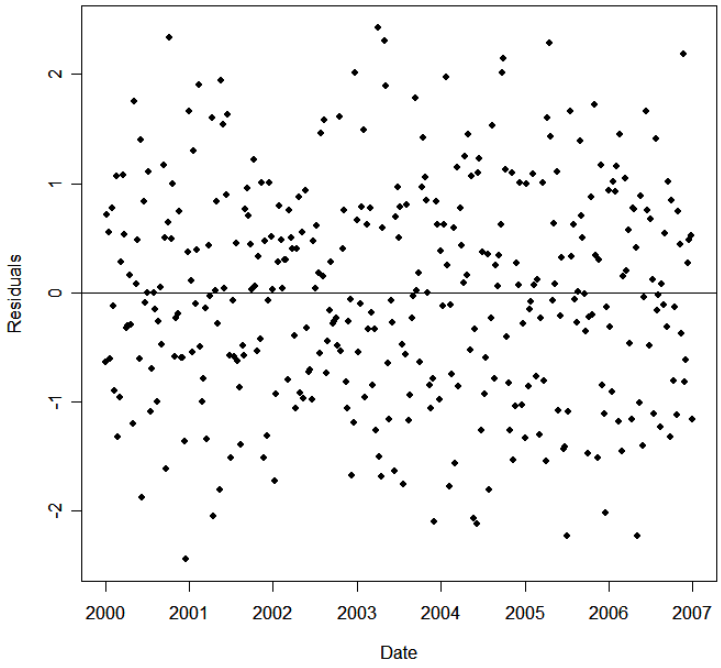

Figure 3.42. Residuals for Model (3.17)

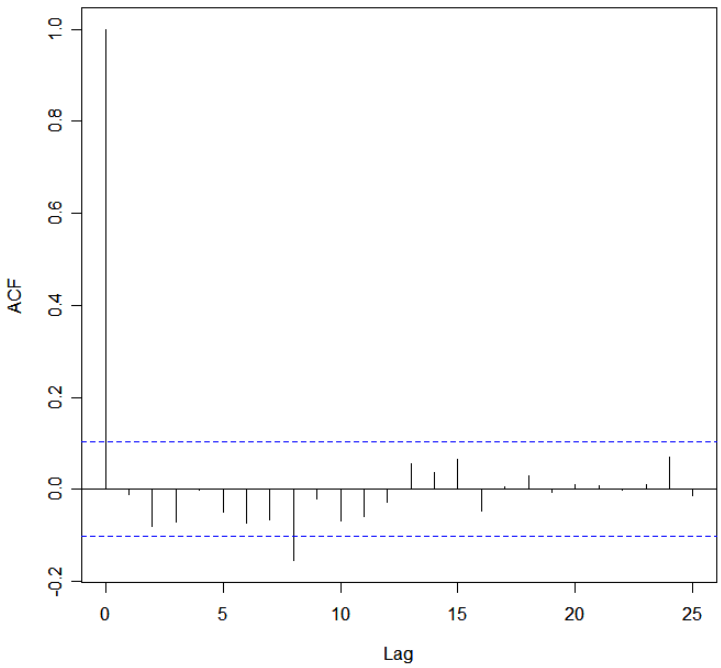

Figure 3.43. Autocorrelation function for the residuals

3.4.7 Pollution Modelling

The next step is now to add PM_{10} to the model to estimate its associations with admissions to hospital. First, a series of single lag models will be produced and the results will be presented in Table 3.14 as relative risks for a $10\mu g/m^3$ increase in PM_{10} concentrations. In addition, the 95% confidence intervals for each lag will be included in the table. For any given lag q, the single lag model is given as

$$\ln(\mu_t) = \alpha_0 + S_1(t; 23) + \gamma \mathrm{PM}_{10_{t-q}} \tag{3.18}$$

However as previously mentioned, a single lag model may not be adequate to model the associations between admissions to hospital and PM_{10} concentrations, therefore a multiple lag model containing the PM_{10} concentrations from the previous eight weeks will also be produced (Model (3.19))

$$\ln(\mu_t) = \alpha_0 + S_1(t; 23) + \sum_{q=0}^{8} \gamma_q \mathrm{PM}_{10_{t-q}} \tag{3.19}$$

The relative risks and 95% confidence intervals from this model are presented in Table 3.15

Lag	RR	95% CI
0	0.9888	(0.9621, 1.0162)
1	1.0057	(0.9787, 1.0334)
2	1.0243	(0.9970, 1.0525)
3	1.0059	(0.9787, 1.0339)
4	0.9993	(0.9721, 1.0273)
5	0.9978	(0.9706, 1.0257)
6	0.9982	(0.9710, 1.0260)
7	0.9979	(0.9706, 1.0260)
8	1.0199	(0.9923, 1.0483)

Table 3.14. Relative risks from single lag models

Lag	RR	95% CI
0	1.0094	(0.9787, 1.0410)
1	1.0074	(0.9769, 1.0388)
2	1.0300	(0.9992, 1.0617)
3	0.9987	(0.9686, 1.0297)
4	1.0109	(0.9810, 1.0418)
5	1.0108	(0.9808, 1.0417)
6	1.0036	(0.9738, 1.0344)
7	0.9789	(0.9496, 1.0090)
8	0.9947	(0.9663, 1.0239)

Table 3.15. Relative risks from multiple lag model

In Table 3.14, we can see that most of the relative risks are fairly close to one with confidence intervals that contain one, meaning the results are non-significant at the 5% level. Similarly, in Table 3.15, the relative risks are all close to one and

non-significant. However the second model suffers from the problem of collinearity, and therefore a sensible next step is to adopt the same distributed lag model that was used previously, thus giving Model (3.20)

$$\begin{aligned} \ln(\mu_t) &= \alpha_0 + S_1(t;23) + \sum_{q=0}^{8} \gamma_q \mathrm{PM}_{10_{t-q}} \\ \gamma_q &= \sum_{k=0}^{q} \eta_k q^k, \quad q = 0, ..., 8 \end{aligned} \tag{3.20}$$

The AICs for the single lag models range from 2182.0 to 2238.5 and the AIC for the multiple lag model is 2092.4. The AICs from the distributed lag model are lower, ranging from 2076.0 to 2080.5, suggesting this is the best fitting model to date. Figure 3.44 shows the shape of the relative risks at each lag for powers zero to four. Power zero is represented by the solid horizontal line in each plot and gives the overall estimated effect of PM_{10}. The shape of the relative risks at powers one and two are broadly linear while at powers three and four, the shape is more cubic in nature. At powers three and four, the relative risks at lags zero and eight are further from the null value of one than the other relative risks are (approximately 0.985 and 1.02 respectively), but we can see from Table 3.16 that they are actually non-significant, as none of the p-values for the relative risk coefficients are significant, thus suggesting there is no significant relationship between weekly admissions to hospital with a cardiovascular illness and mean weekly PM_{10} concentrations in Edinburgh.

Coefficient	Estimate	Standard Error	P-Value
Intercept	3.729	0.1425	$<2 \times 10^{-16}$
η_0	-0.001404	0.001409	**0.3192**
η_1	0.002141	0.001582	**0.1758**
η_2	-6.542×10^{-04}	4.754×10^{-04}	**0.1688**
η_3	5.524×10^{-05}	3.875×10^{-05}	**0.1540**
AIC			2078.5

Table 3.16. Summary of Model (3.20) with a power of three

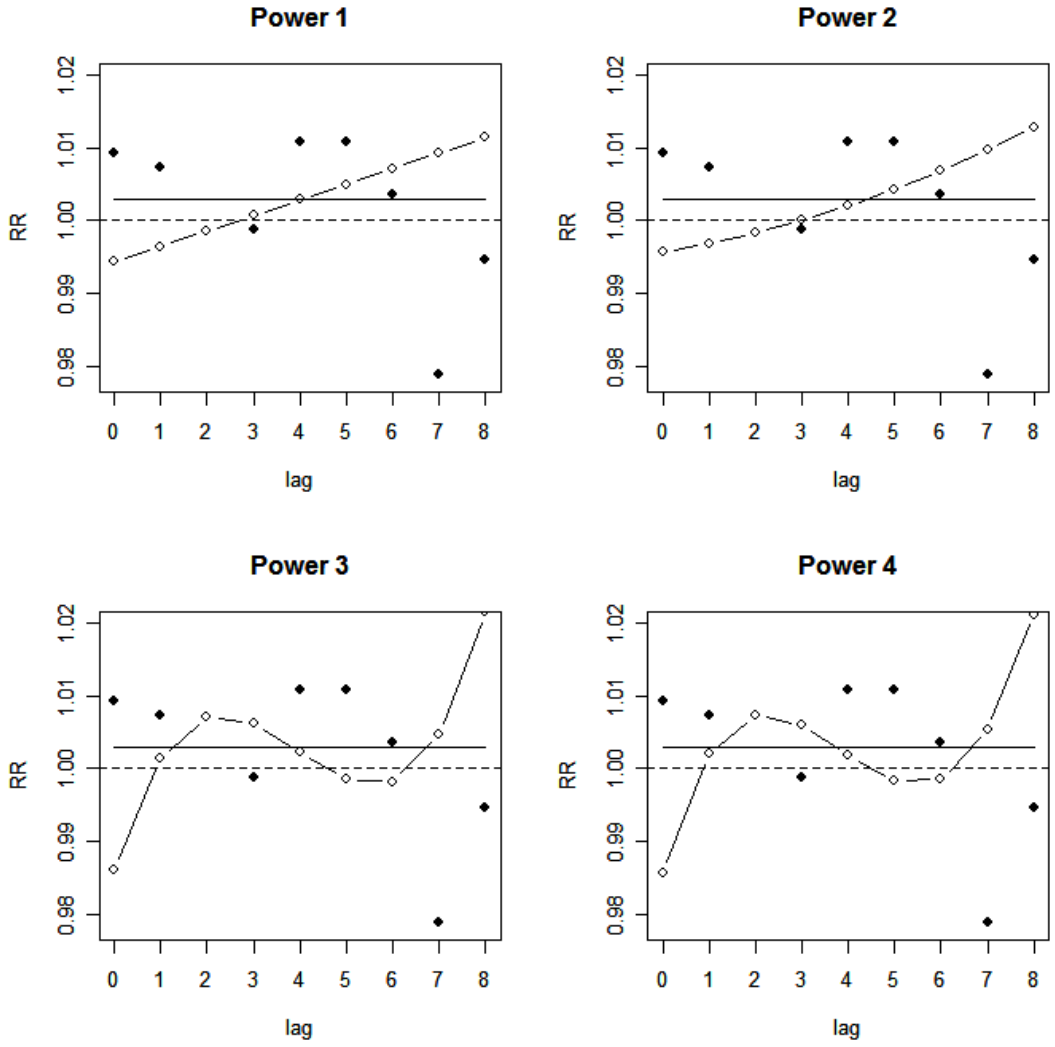

Figure 3.44. The solid line is the relative risk at power zero, the dashed line represents the null risk of 1, the filled diamonds are the relative risks from the unconstrained model and the unfilled diamonds are the relative risks from the distributed lag model

3.4.8 Conclusions

Weekly admissions to hospital increased slightly between 2000 and 2002 before decreasing. PM_{10} concentrations also appear to have decreased over the seven year period although initially there did not appear to be any relationship between

the two variables. The next step was to use natural cubic splines to remove the trend and variation. 23 degrees of freedom were chosen as this minimised the AIC. At this stage, PM_{10} lagged between zero and eight weeks was added to the model using a series of single lag models, a multiple lag model and a distributed lag model. From the results of these models, there does not appear to be any significant relationship between weekly admissions to hospital and exposure to PM_{10}. In the next section the relationship between PM_{10} concentrations and admissions to hospital in Aberdeen will be considered.

3.5 Aberdeen

3.5.1 Exploratory Analysis

The first step is to produce a series of descriptive plots of the data in Aberdeen, comprising admissions to hospital, PM_{10} concentrations and temperature.

Health

Figure 3.45 shows the daily admissions to hospital from cardiovascular illnesses in Aberdeen between 2000 and 2006. The median number of admissions in any one day over the seven years was three, with maximum and minimum numbers of 11 and zero respectively. However because the numbers are so low, it is difficult to gain any useful knowledge. Therefore the admissions will be aggregated up to weekly admissions. Figure 3.46 shows the weekly numbers of admissions to hospital in Aberdeen.

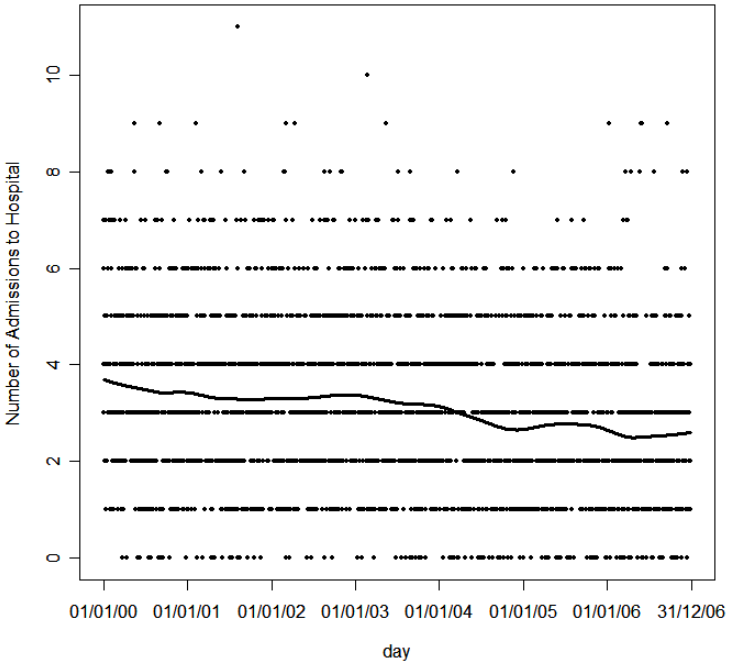

Figure 3.45. Daily admissions to hospital in Aberdeen between 2000 and 2006

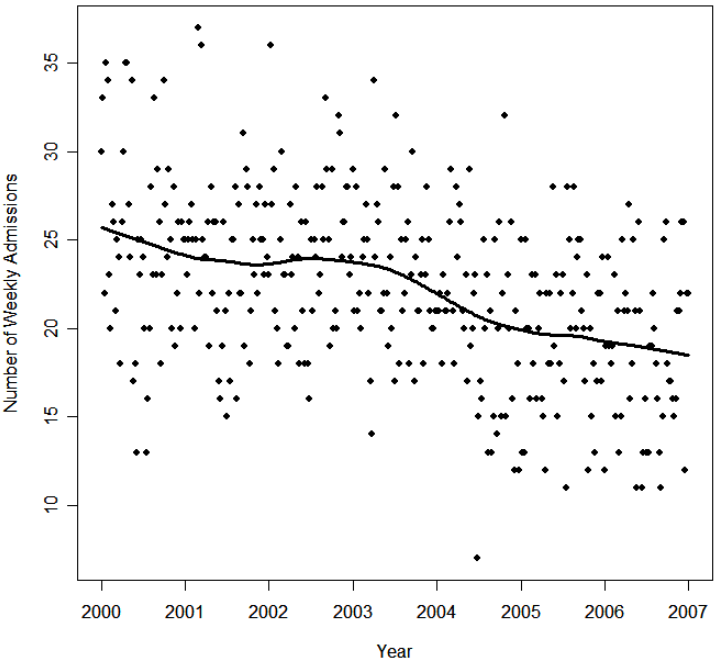

Figure 3.46. Weekly admissions to hospital in Aberdeen between 2000 and 2006

We now see an obvious downward trend to the data, apart from a period in 2002 where the admissions increase slightly. As with admissions to hospital in Glasgow and Edinburgh, there is very little seasonal variation within the data.

Pollution

Below are plots of mean weekly PM_{10} concentrations over the seven year period (Figure 3.47) and weekly admissions to hospital against mean weekly PM_{10} concentrations (Figure 3.48). From Figure 3.47 we can see that PM_{10} concentrations seem to decrease until 2002 when they start increasing again, whilst from Figure 3.48 we can see that there does not appear to be any relationship between weekly numbers of admissions to hospital and mean weekly PM_{10} concentrations in Aberdeen.

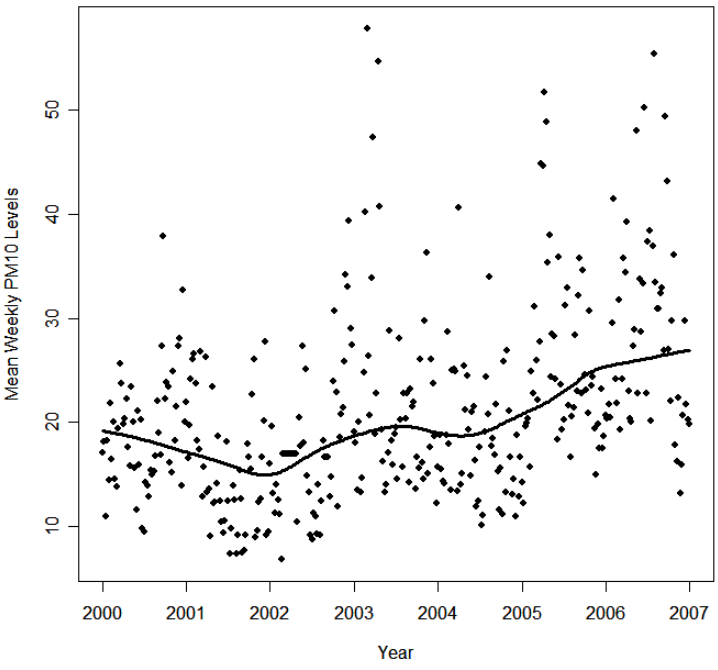

Figure 3.47. Mean weekly PM_{10} concentrations in Aberdeen from 2000 to 2006

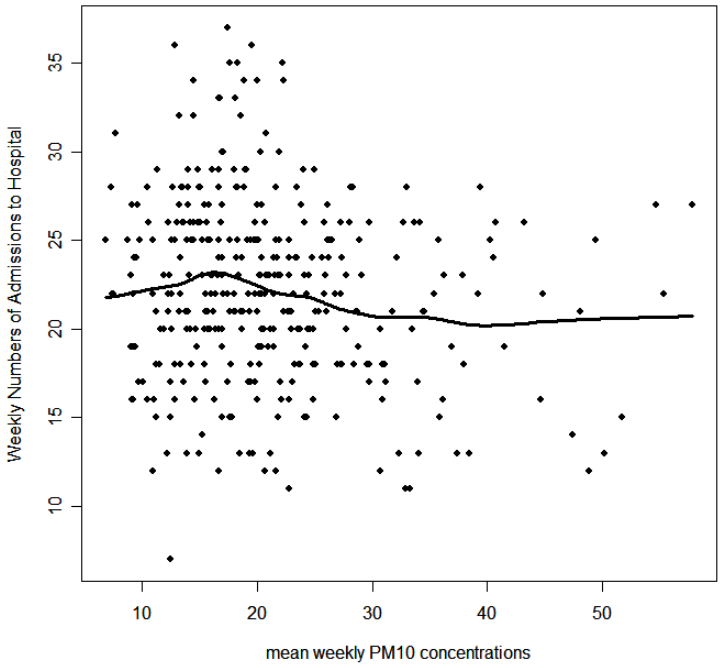

Figure 3.48. Relationship between weekly admissions to hospital and mean weekly PM_{10} concentrations

Temperature

Figures 3.49, 3.50 and 3.51 show plots of mean weekly maximum temperatures in Aberdeen, the first plot over time, the second against weekly admissions to hospital and the last against mean weekly PM_{10} concentrations. Figure 3.49 shows the seasonality we expect, with higher temperatures in summer and lower temperatures in winter. From Figure 3.50, we can see that there does not appear to be any significant relationship between weekly admissions to hospital and mean weekly maximum temperature. Finally, Figure 3.51 shows there does not appear to be any significant relationship between mean weekly PM_{10} concentrations and mean weekly maximum temperatures in Aberdeen.

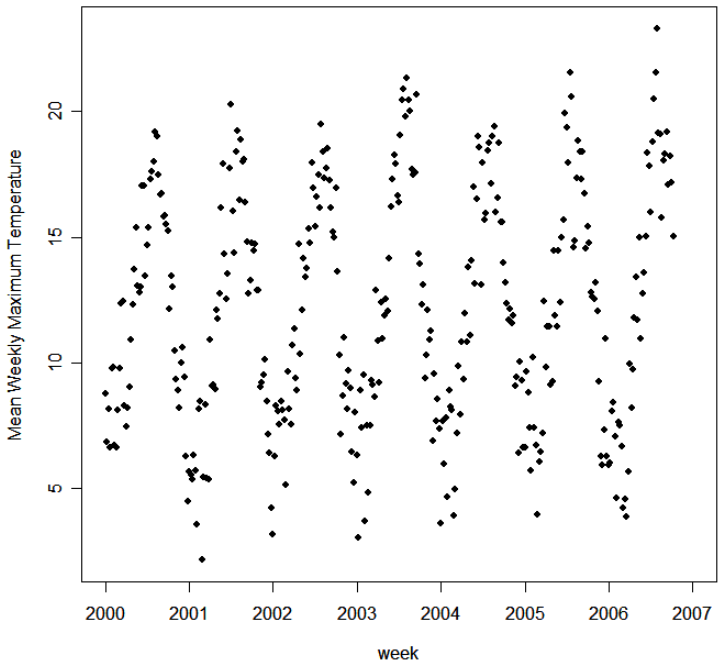

Figure 3.49. Mean weekly maximum temperatures in Aberdeen from 2000 to 2006

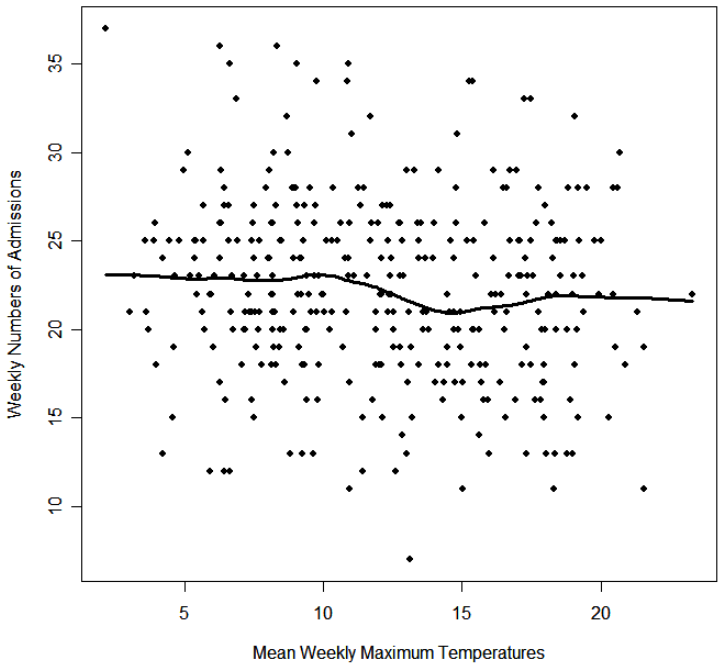

Figure 3.50. Relationship between weekly admissions to hospital and mean weekly maximum temperature

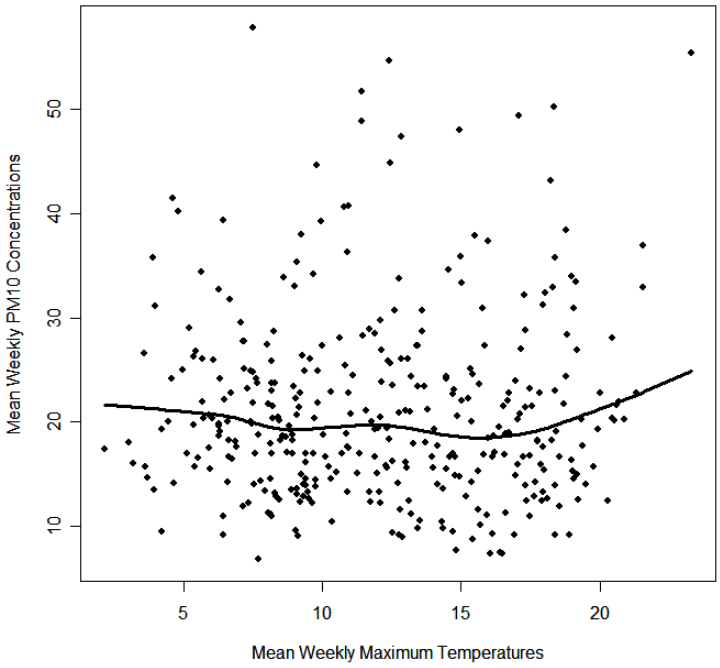

Figure 3.51. Relationship between mean weekly PM_{10} concentrations and mean weekly maximum temperature

3.5.2 Covariate Modelling

The next step is to produce a model that adequately removes any trend, seasonality and correlation from the data. As admissions to hospital in Aberdeen show a more subtle trend than either a linear or a quadratic trend the usual parametric methods will not work here. Also, there is very little seasonality, and what variation there is, is not regular. Therefore sine and cosine terms are not useful in this context. Instead, to remove the trend and seasonal variation, natural cubic splines will be used. 20 degrees of freedom were chosen as this was the model with the lowest AIC. Thus Model (3.21) will be

$$\ln(\mu_t) = \alpha_0 + S_1(t; 20) \tag{3.21}$$

Figure 3.52 shows the residuals from this model while Figure 3.53 shows the

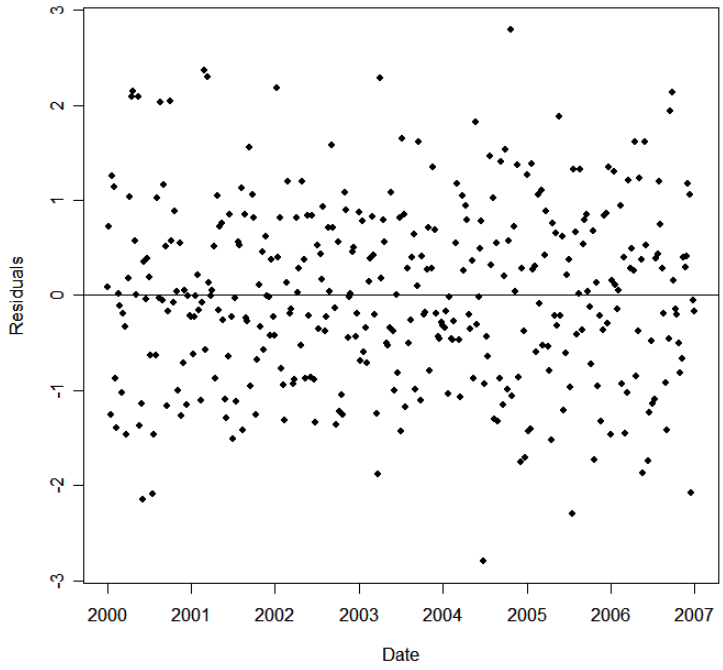

Figure 3.52. Plot of residuals from Model (3.21)

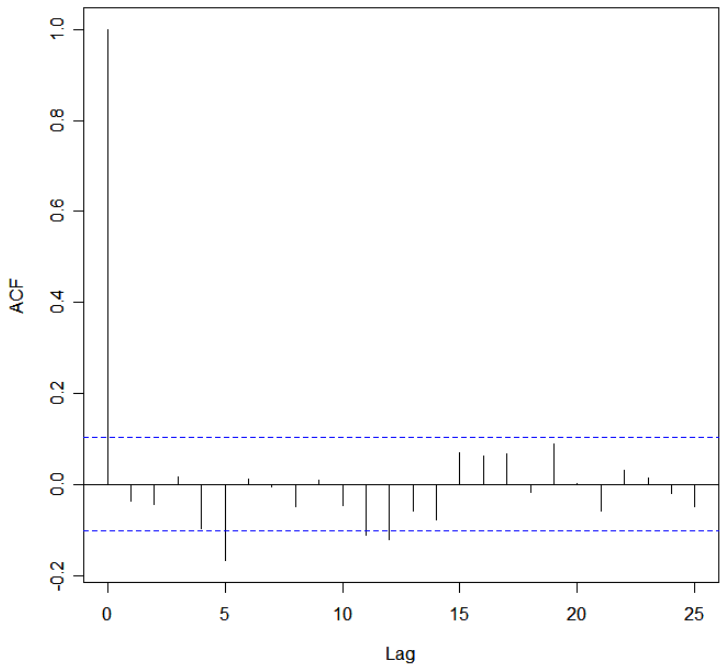

Figure 3.53. Autocorrelation function from the residuals

autocorrelation function for the residuals. From Figure 3.52 we can see that the residuals appear to be evenly scattered about zero with no trend or seasonality. From Figure 3.53 we can see that there are more significant negative correlations than there were for Glasgow and Edinburgh. However most of the lags are within the 95% confidence bands, thus suggesting most of the correlation within the residuals has been removed.

3.5.3 Pollution Modelling

The next step is to add PM_{10} to the model to estimate its association with admissions to hospital. First a series of single lag models, ranging from lag zero to lag eight, will be produced. At any given lag, the single lag models will be given by

$$\ln(\mu_t) = \alpha_0 + S_1(t; 20) + \gamma \mathrm{PM}_{10_{t-q}} \tag{3.22}$$

Table 3.17 shows the relative risks from this model. In addition, the 95% confidence intervals for the relative risks will be given for each lag. However, for reasons previously outlined, single lag models may not be adequate to model the effects of PM_{10}. Therefore a multiple lag model, with lags ranging from zero to eight, will also be produced, giving Model (3.23)

$$\ln(\mu_t) = \alpha_0 + S_1(t; 20) + \sum_{q=0}^{8} \gamma_q \mathrm{PM}_{10_{t-q}} \tag{3.23}$$

The relative risks for this model are presented in Table 3.18 along with the 95% confidence intervals at each lag.

Lag	RR	95% CI
0	1.0045	(0.9715, 1.0386)
1	1.0319	(0.9983, 1.0667)
2	1.0041	(0.9710, 1.0383)
3	0.9956	(0.9620, 1.0303)
4	0.9907	(0.9579, 1.0246)
5	1.0131	(0.9799, 1.0474)
6	1.0005	(0.9677, 1.0345)
7	0.9992	(0.9661, 1.0335)
8	0.9996	(0.9668, 1.0335)

Table 3.17. Relative risks from the single lag models

Lag	RR	95% CI
0	0.9971	(0.9612, 1.0344)
1	1.0403	(1.0023, 1.0798)
2	1.0018	(0.9646, 1.0404)
3	0.9949	(0.9577, 1.0335)
4	0.9834	(0.9474, 1.0208)
5	1.0130	(0.9762, 1.0512)
6	1.0038	(0.9670, 1.0421)
7	1.0077	(0.9707, 1.0460)
8	1.0029	(0.9675, 1.0396)

Table 3.18. Relative risks from the multiple lag model

In Table 3.17, the relative risks are all close to one and non-significant at the 5% level, meaning there does not appear to be any significant relationship between exposure to PM_{10} and cardiovascular admissions to hospital in Aberdeen. Also, most of the relative risks in Table 3.18 are all non-significant and close to one, apart from the relative risk at lag 1 which is 1.04. However the second model suffers from the problem of collinearity meaning the relative risks may not be reliable. Therefore a distributed lag model will be produced to try to reduce this problem. This gives Model (3.24)

$$\begin{aligned} \ln(\mu_t) &= \alpha_0 + S_1(t;20) + \sum_{q=0}^{8} \gamma_q \mathrm{PM}_{10_{t-q}} \qquad (3.24) \\ \text{where} \quad \gamma_q &= \sum_{k=0}^{q} \eta_k q^k, \quad q = 0, ..., 8 \end{aligned}$$

As for the Glasgow and Edinburgh data, the AICs from the distributed lag model (1895.4 to 1900.7) are lower than those for the single lag models (1970.2 to 2022.0) or the multiple lag model (1906.3) thus suggesting the distributed lag model is the best fitting. Figure 3.54 shows the shape of the relative risks for the distributed lag model at powers zero to four at each lag. The dashed horizontal line is the null value of one and the solid horizontal line gives the relative risks at power zero, which is an overall estimate of the effect of PM_{10}. The shape of the relative risks from powers two and three are very similar, while the shape of the relative

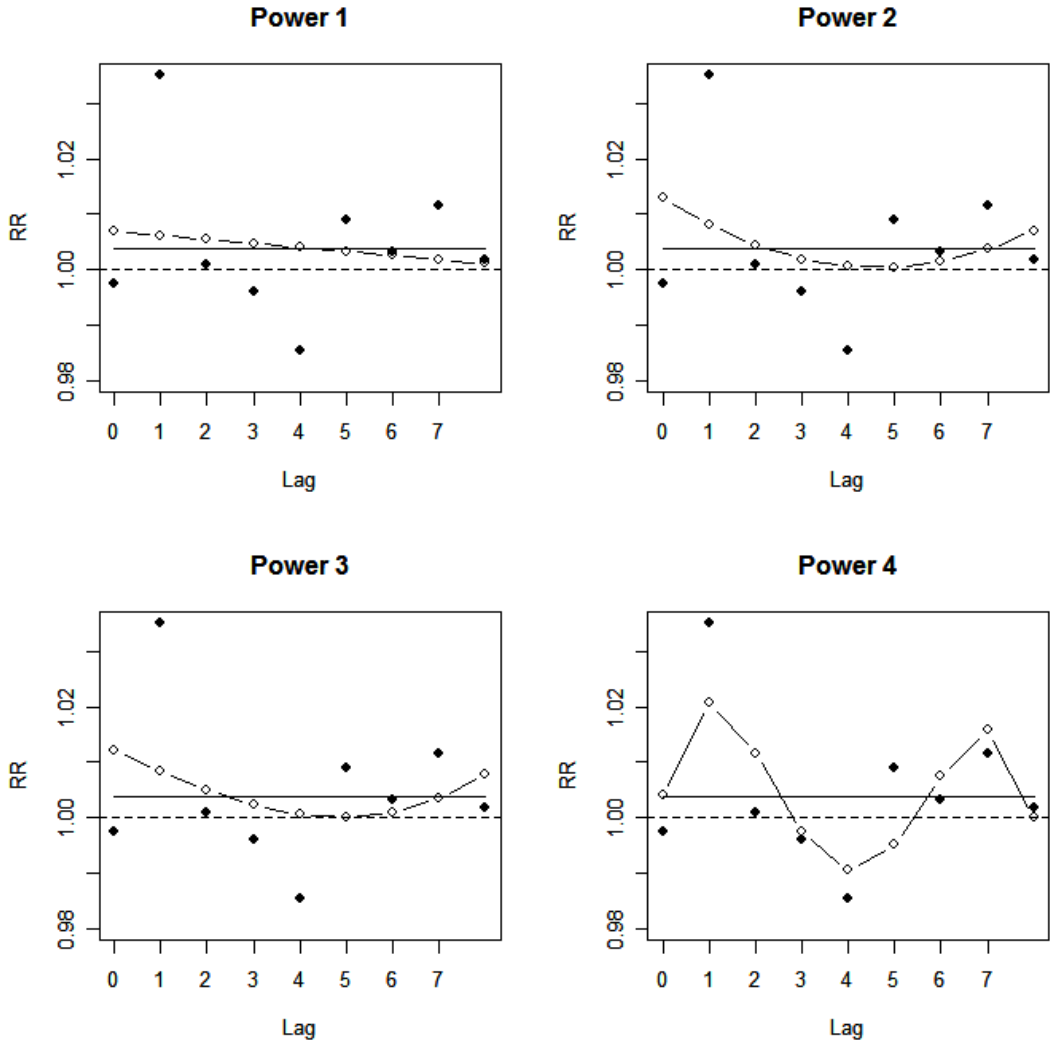

Figure 3.54. The solid line is the relative risk at power zero, the dashed line represents the null risk of 1, the filled diamonds are the relative risks from the unconstrained model and the unfilled diamonds are the relative risks from the distributed lag model

risks from powers one and four are different. In particular the shape of the relative risks at power four is biologically implausible as it is unlikely the relative risks would fluctuate as widely as they do here. All the relative risks at powers one, two and three are fairly close to one, thus implying there is no significant relationship between cardiovascular admissions to hospital and exposure to PM_{10}.

3.5.4 Conclusion

As the number of daily admissions in Aberdeen are so small, the data needed to be aggregated up to weekly levels. We could see an obvious decrease in weekly admissions over the years, apart from a period in 2002 when admissions increased slightly. Mean weekly PM_{10} concentrations initially decreased from 2000 to 2002, after which they increased considerably. At first, there does not appear to be any obvious relationship between admissions to hospital, PM_{10} concentrations or temperature. The next step was to try to remove as much of the trend and variation in the data as possible. This was done using natural cubic splines with 20 degrees of freedom, chosen to minimise the AIC. PM_{10} concentrations were then added for lags of zero up to eight weeks, using single lag models, a multiple lag model and a distributed lag model. The results from each model suggest there does not appear to be any significant relationship between cardiovascular admissions to hospital and exposure to PM_{10} in Aberdeen.

The next section will give a comparison of the three cities, comprising both the data available for each city and the results from each city.

3.6 Comparison of the Three Cities

A direct comparison of Glasgow, Edinburgh and Aberdeen will be seen in this section. Only Glasgow and Edinburgh were analysed using daily data, while all three cities were studied using weekly data. Therefore the comparison will look at weekly admissions to hospital and mean weekly PM_{10} concentrations.

3.6.1 Comparison of Exploratory Plots

First the weekly admissions to hospital will be compared for the three cities, followed by a comparison of the mean weekly PM_{10} concentrations.

Health

Figures 3.55, 3.56 and 3.57 show boxplots of weekly admissions to hospital per 100 000 population for each year in the three cities, Glasgow, Edinburgh and Aberdeen.

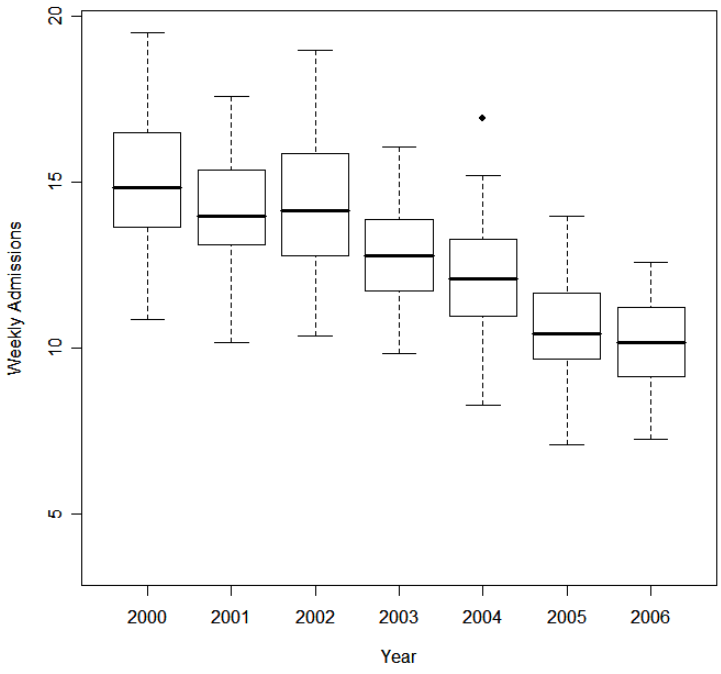

Figure 3.55. Yearly hospital admissions per 100 000 population in Glasgow

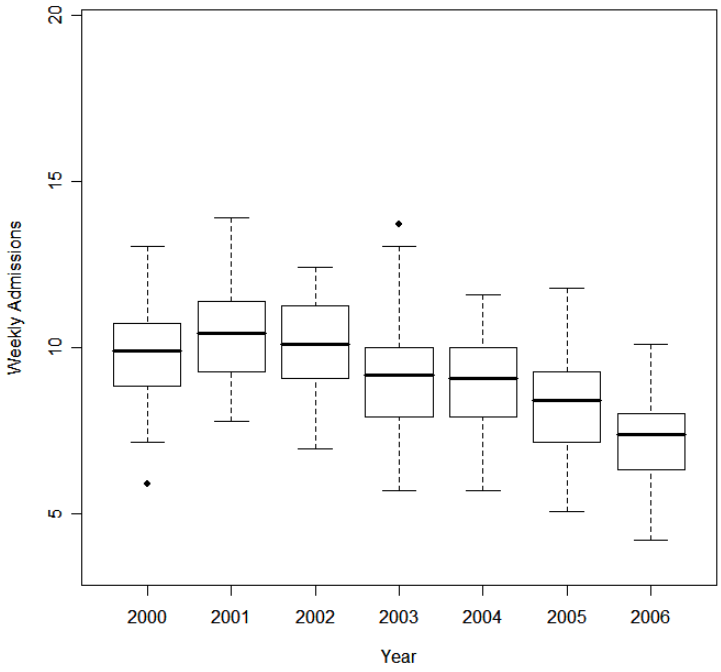

Figure 3.56. Yearly hospital admissions per 100 000 population in Edinburgh

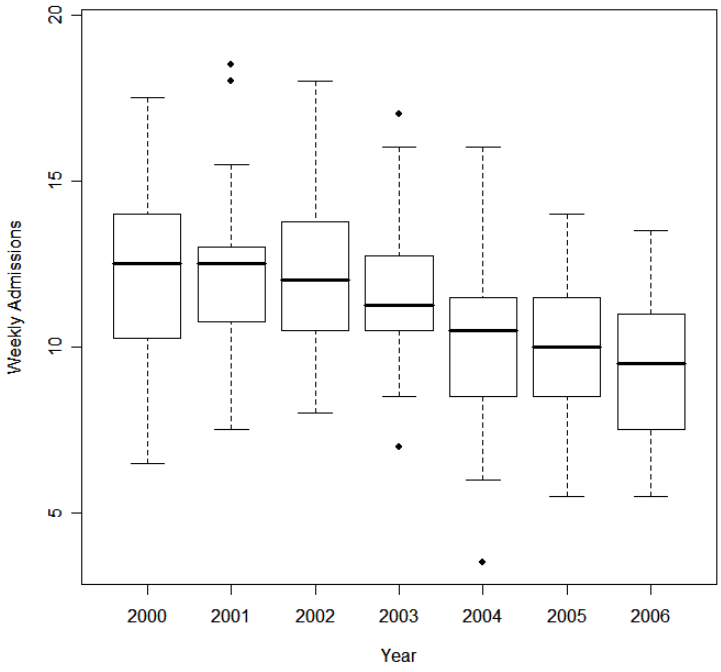

Figure 3.57. Yearly hospital admissions per 100 000 population in Aberdeen

We can see that in Glasgow the median number of admissions per 100 000 population in 2000 was approximately 15 with a maximum of 19 and minimum of 11. This reduces over the years to approximately 11 admissions in 2006 with a maximum and minimum of 13 and seven respectively. In Edinburgh, the admissions increased slightly from a median of 10 in 2000 (maximum and minimum of 13 and seven) to 11 in 2001 (maximum and minimum of 14 and eight) before decreasing again. By 2006 the median admissions per 100 000 had decreased to approximately eight with a maximum of 10 and minimum of four. Aberdeen shows a similar decrease in admissions to Glasgow with a median of 13 per 100 000 in 2000 decreasing to approximately nine in 2006. However Aberdeen also shows a much wider range of admissions. In 2000, the maximum number of admissions was approximately 17 and the minimum was seven, while in 2006 the maximum and minimum admissions were approximately 13 and four respectively.

Figure 3.58 shows a smoothed lowess line for the number of admissions per 100 000 population over time in the three cities. The solid line represents Glasgow, the dashed line is Edinburgh and the dotted line is for Aberdeen. We can see that although cardiovascular admissions to hospital have decreased in all three cities, Glasgow has the greatest number of admissions while Edinburgh has the least. However Glasgow's admissions have decreased more rapidly than either of the other two cities, and by the end of 2006, had a similar number of admissions to Aberdeen.

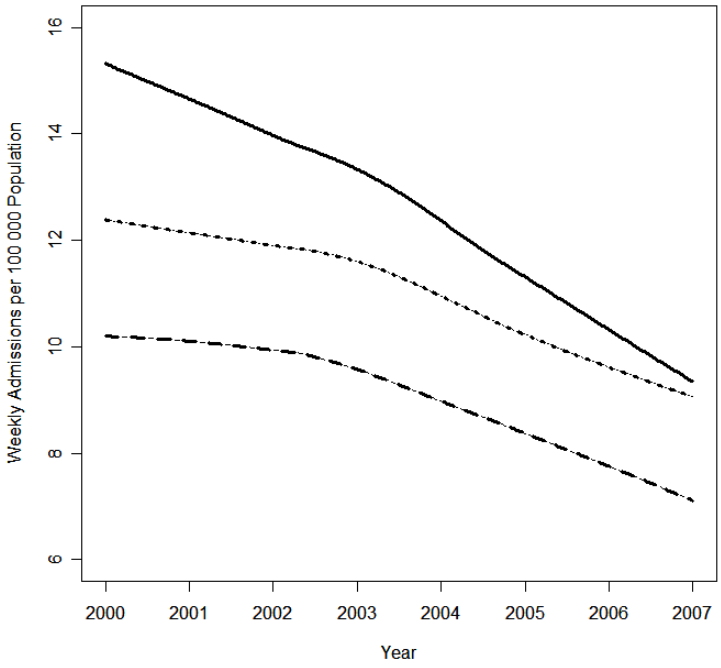

Figure 3.58. Admissions per 100 000 population over time in the three cities. The solid line represents Glasgow, the dashed line is Edinburgh and the dotted line is for Aberdeen.

Pollution

Figure 3.59 shows boxplots of the overall mean PM_{10} concentrations in Glasgow, Edinburgh and Aberdeen.

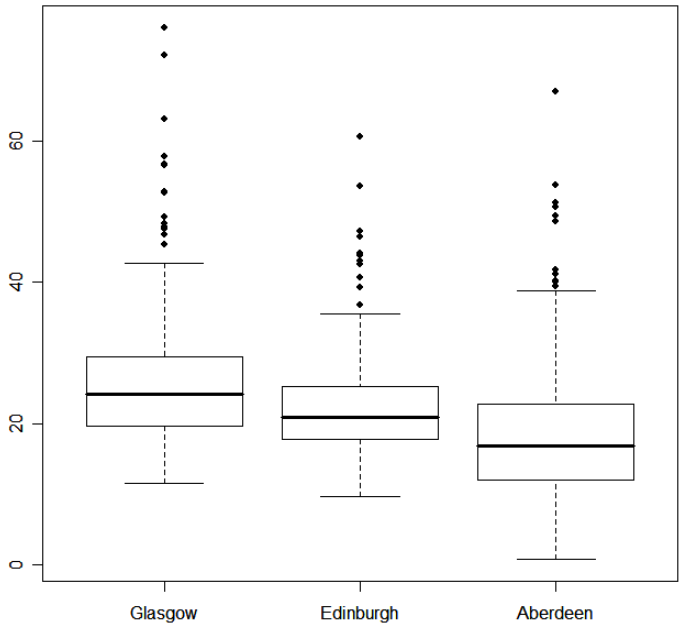

Figure 3.59. Boxplot of PM_{10} concentrations in the 3 Cities

We can see that the PM_{10} concentrations are very similar for the three cities although Glasgow is slightly higher than the other two cities. However these boxplots show the median pollution concentrations over the seven years and do not give any indication of the pollution levels in each individual year. Figures 3.60, 3.61 and 3.62 below show yearly boxplots of pollution concentrations for each of the three cities.

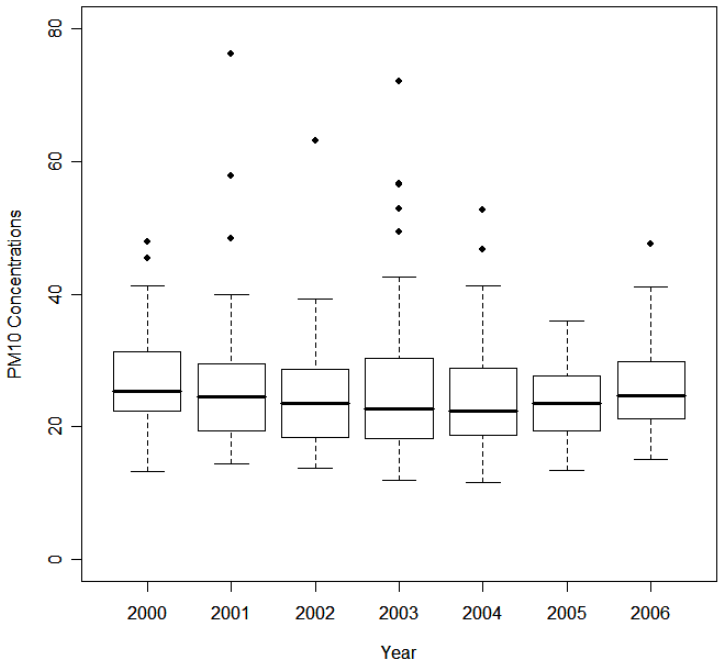

Figure 3.60. Yearly PM_{10} Concentrations in Glasgow

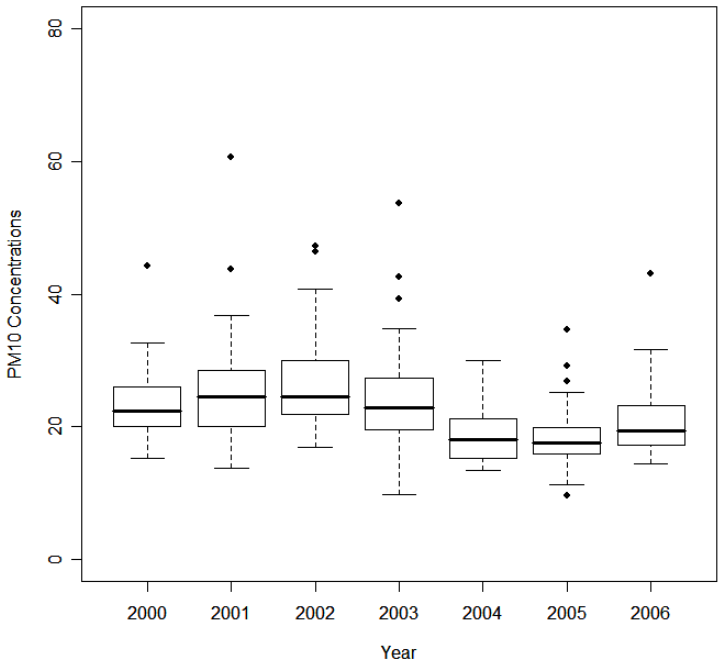

Figure 3.61. Yearly PM_{10} Concentrations in Edinburgh

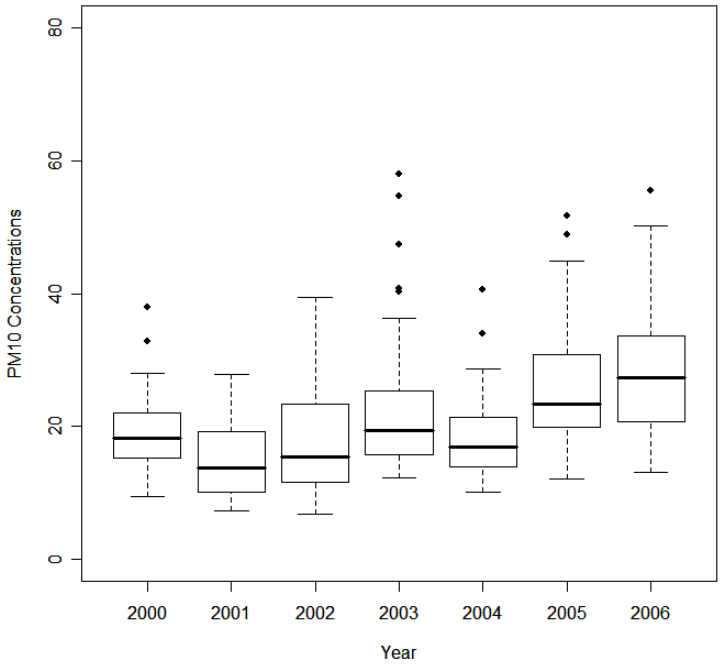

Figure 3.62. Yearly PM_{10} Concentrations in Aberdeen

We can see that PM_{10} concentrations do not appear to have changed much in Glasgow over the seven year period. They decreased slightly until 2003, after which they appear to have remained fairly steady. In Edinburgh the PM_{10} increased slightly between 2000 and 2002 after which they decreased again, staying fairly steady for a while before increasing slighly in 2006 while in Aberdeen the PM_{10} concentrations decreased between 2000 and 2001 when they remained steady before increasing again in 2003. It is unknown why pollution levels suddenly increased they way they did and is an area for further research.

3.6.2 Comparison of Results

This section compares the modelling process and the results for the three cities. Before adding PM_{10} to any of my models, it was necessary to remove as much of the seasonality and trend as possible so that any effect that could be seen would be the effect of pollution. In order to do a direct comparison, the results from the models using natural cubic splines in each city will be used. In Glasgow, 22 degrees of freedom were used to remove the trend and seasonality while in Edinburgh 23 degrees of freedom were used and the Aberdeen model used 20 degrees of freedom. Tables 3.19, 3.20 and 3.21 show the relative risks and 95% confidence intervals from the single lag models from each of the three cities.

Lag	RR	95% CI
0	0.9977	(0.9832, 1.0123)
1	1.0145	(1.0000, 1.0292)
2	1.0027	(0.9882, 1.0173)
3	1.0051	(0.9906, 1.0198)
4	1.0008	(0.9863, 1.0155)
5	1.0132	(0.9986, 1.0279)
6	1.0203	(1.0057, 1.0351)
7	0.9978	(0.9832, 1.0125)
8	0.9895	(0.9750, 1.0042)

Table 3.19. Relative risks from the single lag models for Glasgow

Lag	RR	95% CI
0	0.9888	(0.9621, 1.0162)
1	1.0057	(0.9787, 1.0334)
2	1.0243	(0.9970, 1.0525)
3	1.0059	(0.9787, 1.0339)
4	0.9993	(0.9721, 1.0273)
5	0.9978	(0.9706, 1.0257)
6	0.9982	(0.9710, 1.0260)
7	0.9979	(0.9706, 1.0260)
8	1.0199	(0.9923, 1.0483)

Table 3.20. Relative risks from the single lag models for Edinburgh

Lag	RR	95% CI
0	1.0045	(0.9715, 1.0386)
1	1.0319	(0.9983, 1.0667)
2	1.0041	(0.9710, 1.0383)
3	0.9956	(0.9620, 1.0303)
4	0.9907	(0.9579, 1.0246)
5	1.0131	(0.9799, 1.0474)
6	1.0005	(0.9677, 1.0345)
7	0.9992	(0.9661, 1.0335)
8	0.9996	(0.9668, 1.0335)

Table 3.21. Relative risks from the single lag models for Aberdeen

Glasgow is the only city to have any significant relative risks. Although Edinburgh and Aberdeen both have a particularly high relative risk at lags two and one respectively, they are non significant at the 5% level. The other relative risks are very close to one with narrow confidence intervals. The confidence intervals for the relative risks from the Aberdeen data are wider than those for the Glasgow or Edinburgh data, although this is likely due to the smaller numbers of admissions in this city. However it is unlikely that PM_{10} on only one day will have a significant effect on health so multiple lag models were produced, containing the mean weekly PM_{10} concentrations from the previous eight weeks. Tables 3.22, 3.23 and 3.24 show the relative risks from the multiple lag model for each of the three cities.

Lag	RR	95% CI
0	0.9993	(0.9835, 1.0153)
1	1.0203	(1.0041, 1.0368)
2	0.9952	(0.9787, 1.0119)
3	1.0115	(0.9949, 1.0284)
4	1.0000	(0.9837, 1.0167)
5	1.0106	(0.9941, 1.0273)
6	1.0206	(1.0039, 1.0375)
7	0.9985	(0.9823, 1.0148)
8	0.9957	(0.9798, 1.0119)

Table 3.22. Relative risks from the multiple lag model for Glasgow

Lag	RR	95% CI
0	1.0094	(0.9787, 1.0410)
1	1.0074	(0.9769, 1.0388)
2	1.0300	(0.9992, 1.0617)
3	0.9987	(0.9686, 1.0297)
4	1.0109	(0.9810, 1.0418)
5	1.0108	(0.9808, 1.0417)
6	1.0036	(0.9738, 1.0344)
7	0.9789	(0.9496, 1.0090)
8	0.9947	(0.9663, 1.0239)

Table 3.23. Relative risks from the multiple lag model for Edinburgh

Lag	RR	95% CI
0	0.9971	(0.9612, 1.0344)
1	1.0403	(1.0023, 1.0798)
2	1.0018	(0.9646, 1.0404)
3	0.9949	(0.9577, 1.0335)
4	0.9834	(0.9474, 1.0208)
5	1.0130	(0.9762, 1.0512)
6	1.0038	(0.9670, 1.0421)
7	1.0077	(0.9707, 1.0460)
8	1.0029	(0.9675, 1.0396)

Table 3.24. Relative risks from the multiple lag model for Aberdeen

Both Glasgow and Aberdeen have a significant relative risk at lag one, with Glasgow showing a 2% increase in cardiovascular admissions to hospital and Aberdeen showing a 4% increase. Also Glasgow shows a significant relative risk at lag six, which shows a 2% increase in admissions to hospital. Other than these, none of the results are significant, thus suggesting the significant relative risks are simply a result of multiple testing and are actually non-significant. It should be noted that the confidence intervals are wider than those from the single lag models. This is likely to be due to collinearity which was overcome using distributed lag models. However the results from the distributed lag models also showed no significant effect of exposure to PM_{10} on cardiovascular admissions to hospital.

In conclusion, after adding PM_{10} to the models, there was no significant effect on admissions to hospital in any of the three cities. This could simply be due to the fact that exposure to PM_{10} does not have any effect on admissions to hospital with cardiovascular diseases. However previous studies have found significant effects which suggests that either there is a threshold concentration below which PM_{10} does not have any significant effect on cardiovascular illnesses and PM_{10} concentrations in Glasgow, Edinburgh and Aberdeen and within this threshold, or the data available were not suitable for this study.

Chapter 4

Glasgow Sub-Category Analyses

In the previous chapter we found that there was no significant effect of exposure to PM_{10} on admissions to hospital with cardiovascular diseases in Glasgow, Edinburgh or Aberdeen. These analyses were carried out for CHD and stroke combined and it would be of interest to perform separate analyses for each of these two conditions. However this will involve subsetting the admissions data, and only Glasgow has a large enough number of admissions to give robust results. Therefore this chapter focuses exclusively on Glasgow. There were 26 857 admissions to hospital with a cardiovascular illness in Glasgow, between 2000 and 2006, of which 68.8% were from CHD and 31.2% were from stroke.

There have been a number of studies carried out assessing the effects of pollution on males and females separately, many of which have found a significant effect of air pollution in women, but not in men. Therefore in addition to a CHD/stroke split, two subanalyses will be carried out on the Glasgow data to investigate whether or not there is any evidence of a gender specific effect. The proportion of men and women being admitted into hospital with cardiovascular diseases between 2000 and 2006 was similar, with 52.6% of the total admissions being men and 47.4% women.

A third split of the data we consider concerns age. Elderly people are more

likely to be at risk of suffering from multiple illness, which makes it difficult to diagnose a primary cause of death or admission. Therefore it may also be insightful to analyse only data for the younger population. While many studies consider this age group to be under 65, there are not enough daily admissions to give robust results (40.7% of the total number of admissions). Therefore we will investigate the effects of PM_{10} on the under 75s, which consists of 67.5% of the available admissions.

Air pollution is commonly believed to aggravate symptoms in people who have already had a cardiovascular or respiratory illness. However we are interested in determining whether air pollution also has a negative effect on people who have never been admitted to hospital with a cardiovascular illness. Therefore we also focus on first time admissions only. Finally, although the majority of this thesis looks at the effects of air pollution on cardiovascular morbidity, cardiovascular mortality is also of interest. Therefore we investigate the effects of PM_{10} on mortality due to CHD and stroke separately for all ages.

This chapter will be split into five sections, the first of which will focus on analyses concerning admissions to hospital from both CHD and stroke, while Section 4.2 will investigate several analyses relating solely to CHD and Section 4.3 will consider analyses related to admissions to hospital from stroke. Section 4.4 will look at respiratory illnesses, in particular emergency admissions to hospital with asthma while the final section focuses on the effectiveness of the current air pollution limit values. Many of the subgroups to be analysed do not have large enough numbers of daily admissions to hospital. Therefore the analyses will be carried out using weekly data.

4.1 Breakdown of Cardiovascular Admissions

In this section, cardiovascular admissions to hospital will be broken down into groups, the first of which will be admissions to hospital amongst the male population while the second will investigate the relationship between PM_{10} and people who are being admitted to hospital with a cardiovascular illness for the first time. As the admissions to hospital come in the form of counts, the same Poisson generalised linear models will be used as in Chapter 3, for example equation (3.1).

4.1.1 Males

Exploratory Analysis

We start by producing a series of descriptive plots of the data for males. Figure 4.1 shows the weekly admissions to hospital among the male population between 2000 and 2006 with a smoothed lowess line to show the underlying trend in the data, while Figures 4.2 and 4.3 show the relationship between cardiovascular admissions to hospital among males and mean weekly PM_{10} concentrations and mean weekly maximum temperature respectively.

From Figure 4.1 we can see that there is an approximately linear decreasing trend in hospital admissions with little seasonal variation and Figures 4.2 and 4.3 show us that there does not appear to be any significant relationship between admissions to hospital among males and either mean weekly PM_{10} concentrations or mean weekly maximum temperature.

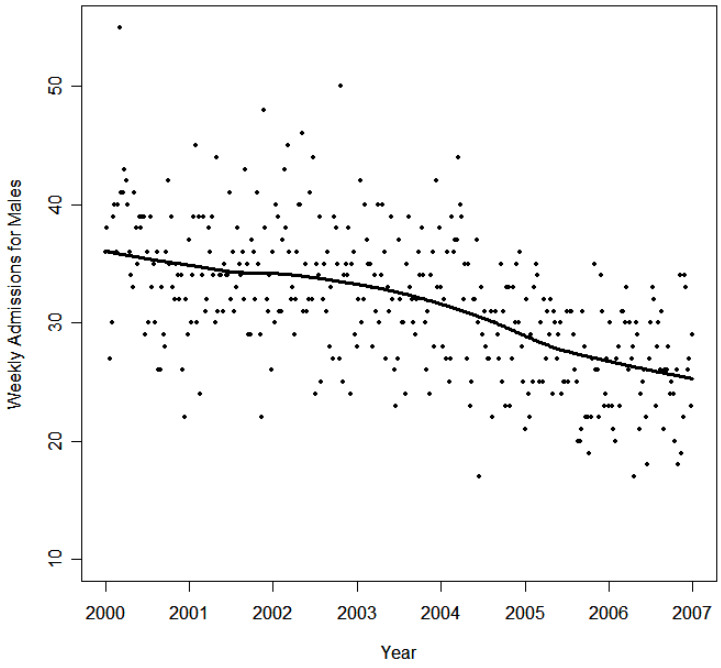

Figure 4.1. Weekly admissions to hospital among males

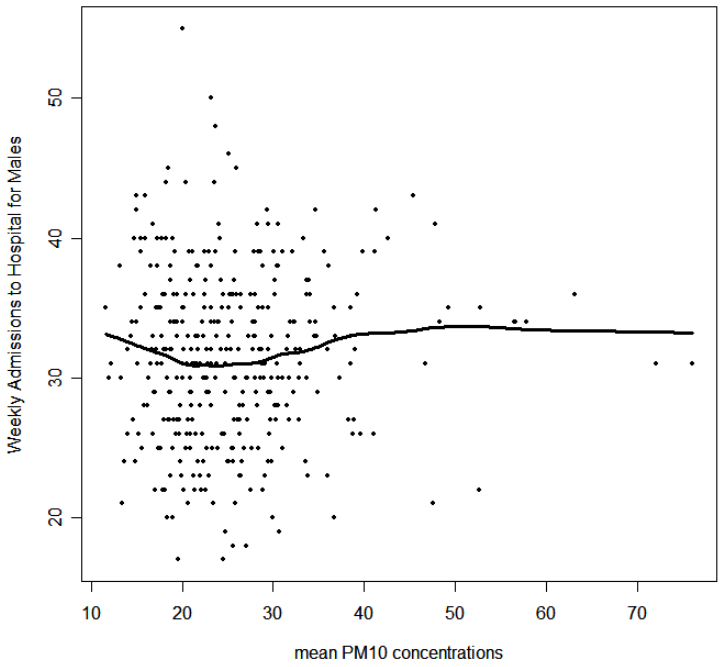

Figure 4.2. Relationship between weekly admissions to hospital among males and mean weekly PM_{10} concentrations

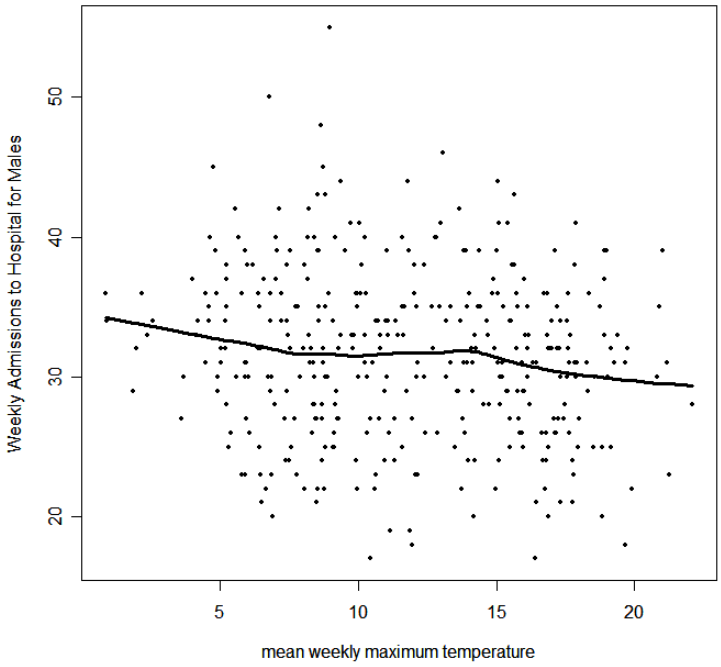

Figure 4.3. Relationship between weekly admissions to hospital among males and mean weekly maximum temperature

Covariate Modelling

The next stage is to remove the prominent trend in the data and initially a linear function of time was included in the model. However this did not remove all the underlying trend so instead natural cubic splines with 14 degrees of freedom were used. The 14 degrees of freedom were chosen as this minimised the AIC. Mean weekly maximum temperature was also added to the model as it was significantly related to weekly admissions. A summary of this model can be seen in Table 4.1.

Coefficient	Estimate	Standard Error	P-Value
Intercept	3.683356	0.063013	$<2 \times 10^{-16}$
max	-0.012542	0.004012	0.0018
AIC			2216.7

Table 4.1. Summary of the model for males

Figure 4.4 shows the residuals from this model, which appear to be evenly scattered about zero with no obvious pattern. In addition Figure 4.5 shows the autocorrelation function for the residuals in which most of the lags are within the 95% confidence bands, thus suggesting the model is adequate.

Pollution Modelling

The next step is to add mean weekly PM_{10} concentrations to the model. Previously, a series of single lag models were investigated followed by a multiple lag model and finally a distributed lag model. However for brevity in these subanalyses, only the distributed lag model is used, details of which are given in Chapter 3, equation (3.8). Figure 4.6 shows the shape of the relative risks for powers zero to four under the distributed lag model, where the solid horizontal line is the relative risk at zero, which forces the risks at all lags to be equal.

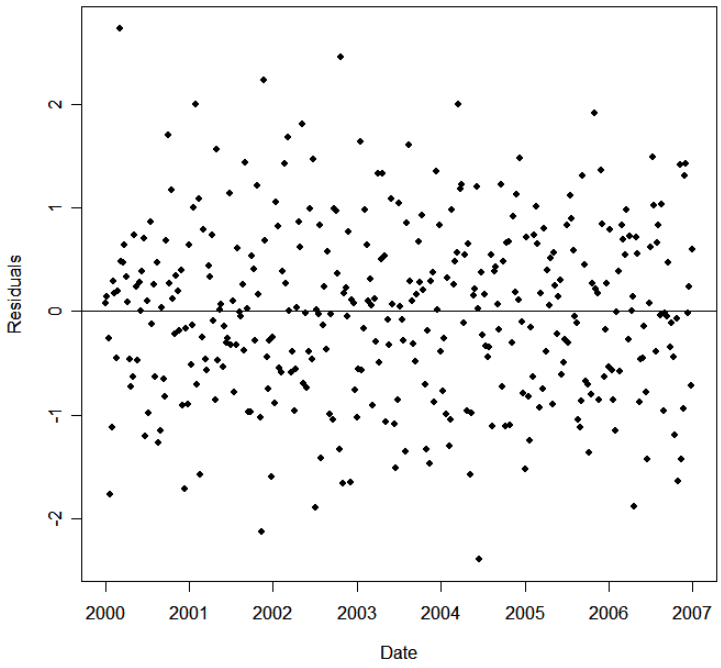

Figure 4.4. Residuals from the model of male admissions to hospital

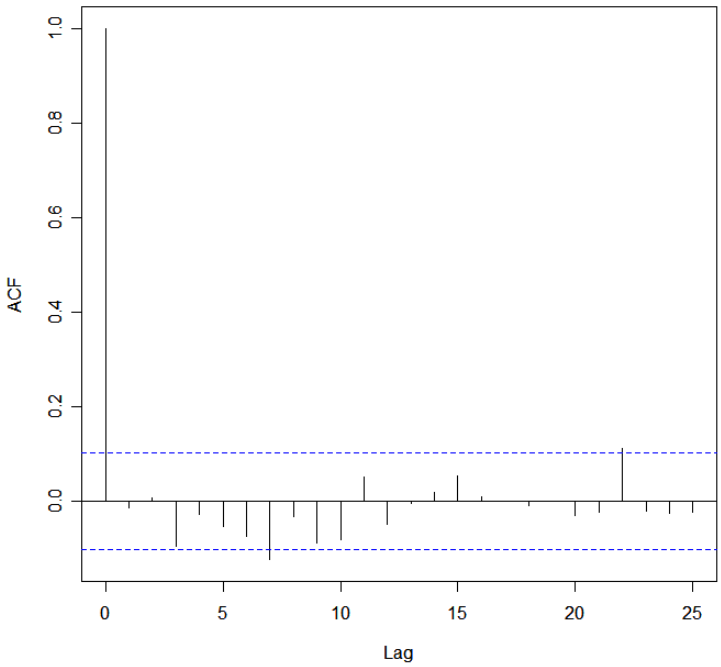

Figure 4.5. Autocorrelation function for the residuals

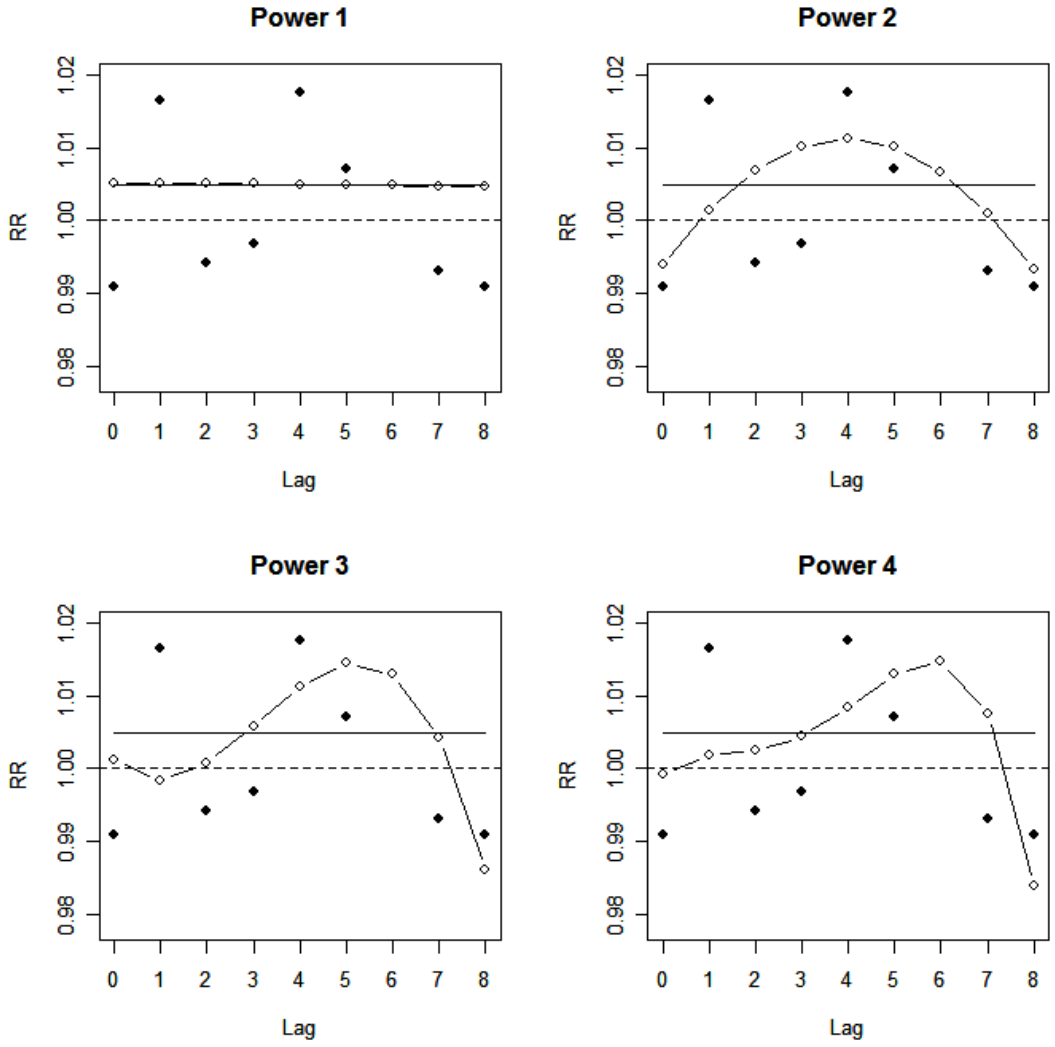

Figure 4.6. The solid line is the relative risk at power zero, the dashed line represents the null risk of 1, the filled diamonds are the relative risks from the unconstrained model and the unfilled diamonds are the relative risks from the constricted distributed lag model

We can see that most of the relative risks are close to one, although for powers three and four the relative risks are fairly high at about lags five and six, being approximately 1.015 (Ballester et al. (2006) reported a relative risk of 1.009). However none of the relative risk coefficients are significant, thus implying that the high relative risks are non significant.

Conclusions

From this analysis we conclude that weekly admissions to hospital among males have decreased between 2000 and 2006, but there does not appear to be any obvious relationship between admissions to hospital and mean weekly PM_{10} concentrations. At powers three and four, the distributed lag model showed some high relative risks at lags five and six although these results were non significant, suggesting that there does not appear to be any relationship between exposure to PM_{10} and hospital admissions due to cardiovascular diseases among males.

The relationship between exposure to PM_{10} and cardiovascular admissions to hospital among females was also investigated and again, no significant relationship was found. This analysis can be seen in Appendix A. The next section will try to determine whether or not there is a relationship between exposure to PM_{10} and people who are being admitted to hospital with a cardiovascular illness for the first time. This subgroup is restricted to people under 75 years of age, and comprises 43.5% of the overall numbers of admissions to hospital in Glasgow.

4.1.2 1st Admissions

Exploratory Analysis

As before, the first stage of the analysis is producing descriptive plots of the data. Figure 4.7 below shows the weekly numbers of patients who were admitted to hospital with a cardiovascular illness for the first time. As with the male admissions, the numbers have been decreasing approximately linearly from 2000 to 2006, with some discernable seasonality. Figures 4.8 and 4.9 show the relationships between first admissions to hospital and mean weekly PM_{10} concentrations and mean weekly maximum temperature respectively. We can see that there does not appear to be any obvious relationship between the variables.

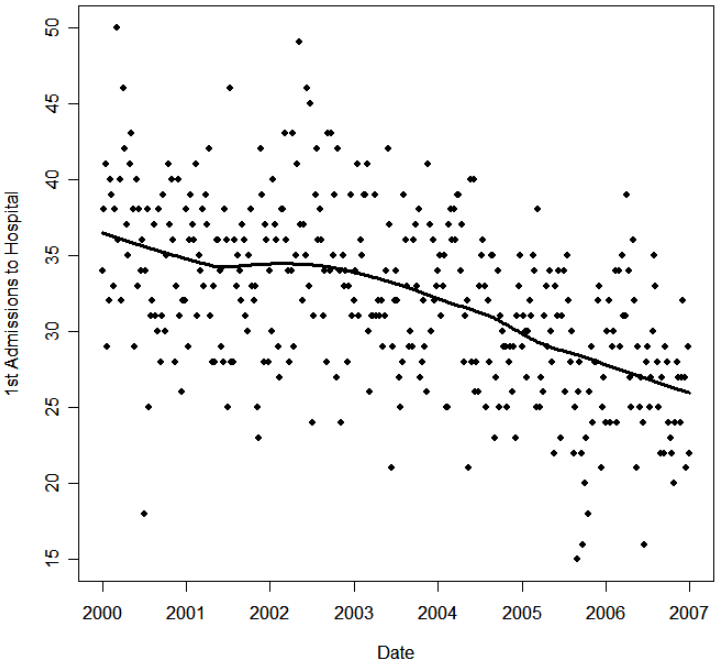

Figure 4.7. Weekly numbers of patients being admitted to hospital with a cardiovascular illness for the first time

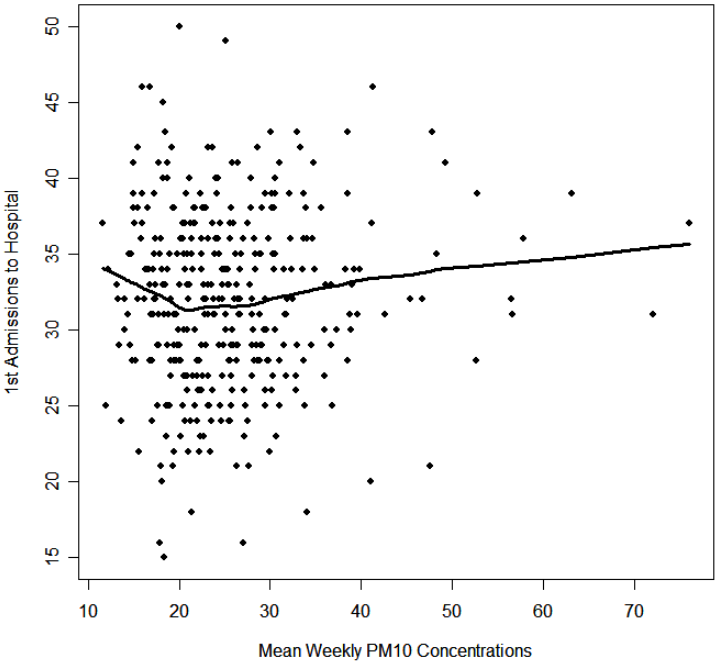

Figure 4.8. Relationship between first admissions to hospital and mean weekly PM_{10} concentrations

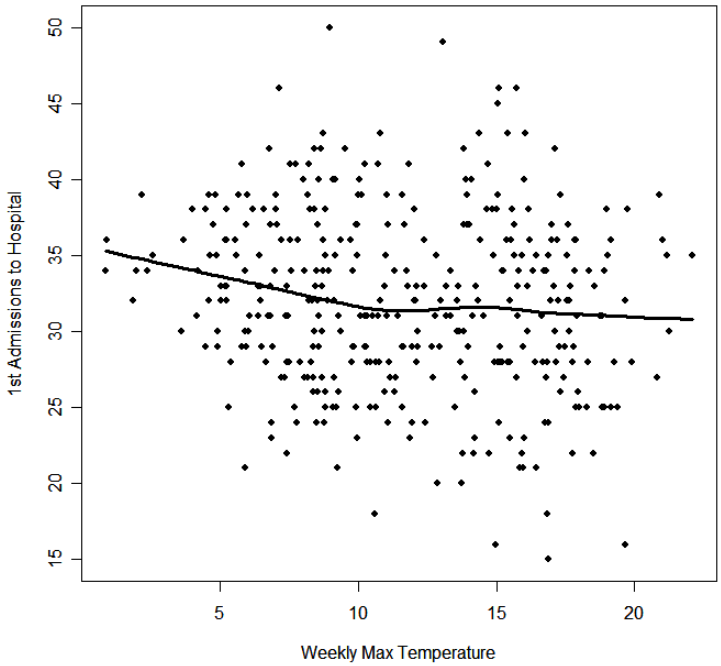

Figure 4.9. Relationship between first admissions to hospital and mean weekly maximum temperature

Covariate Modelling

The next step is to produce a model that removes the trend and seasonality in the data. As a linear time trend was not suitable in this case, natural cubic splines with 17 degrees of freedom were used to model the trend and seasonality. 17 degrees of freedom were chosen as this minimised the AIC. Figures 4.10 and 4.11 below show the plots of the residuals for this model and the autocorrelation function for the residuals. We can see that the residuals appear to be scattered evenly about zero with no obvious pattern, and within the autocorrelation function, most of the lags are within the 95% confidence bands, thus suggesting the residuals now resemble white noise. Therefore we can continue with the modelling process.

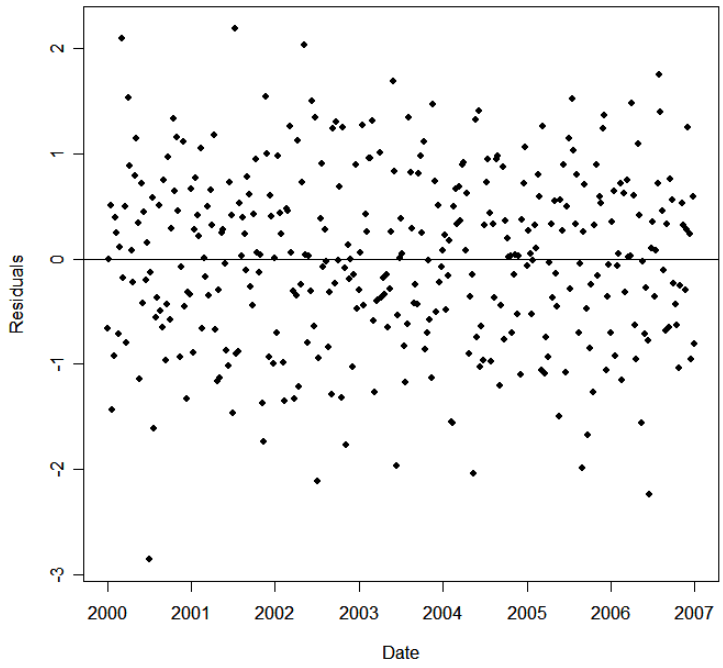

Figure 4.10. Residuals from the model of first admissions to hospital

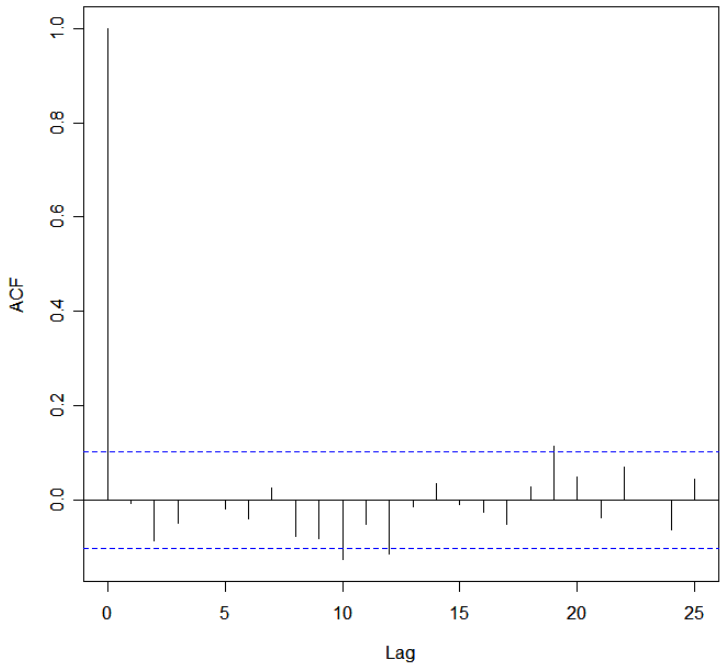

Figure 4.11. Autocorrelation function for the residuals

Pollution Modelling

The next step is to add PM_{10} concentrations to the model using a distributed lag model. Figure 4.12 shows the shape of the relative risks from the distributed lag model at powers zero to four.

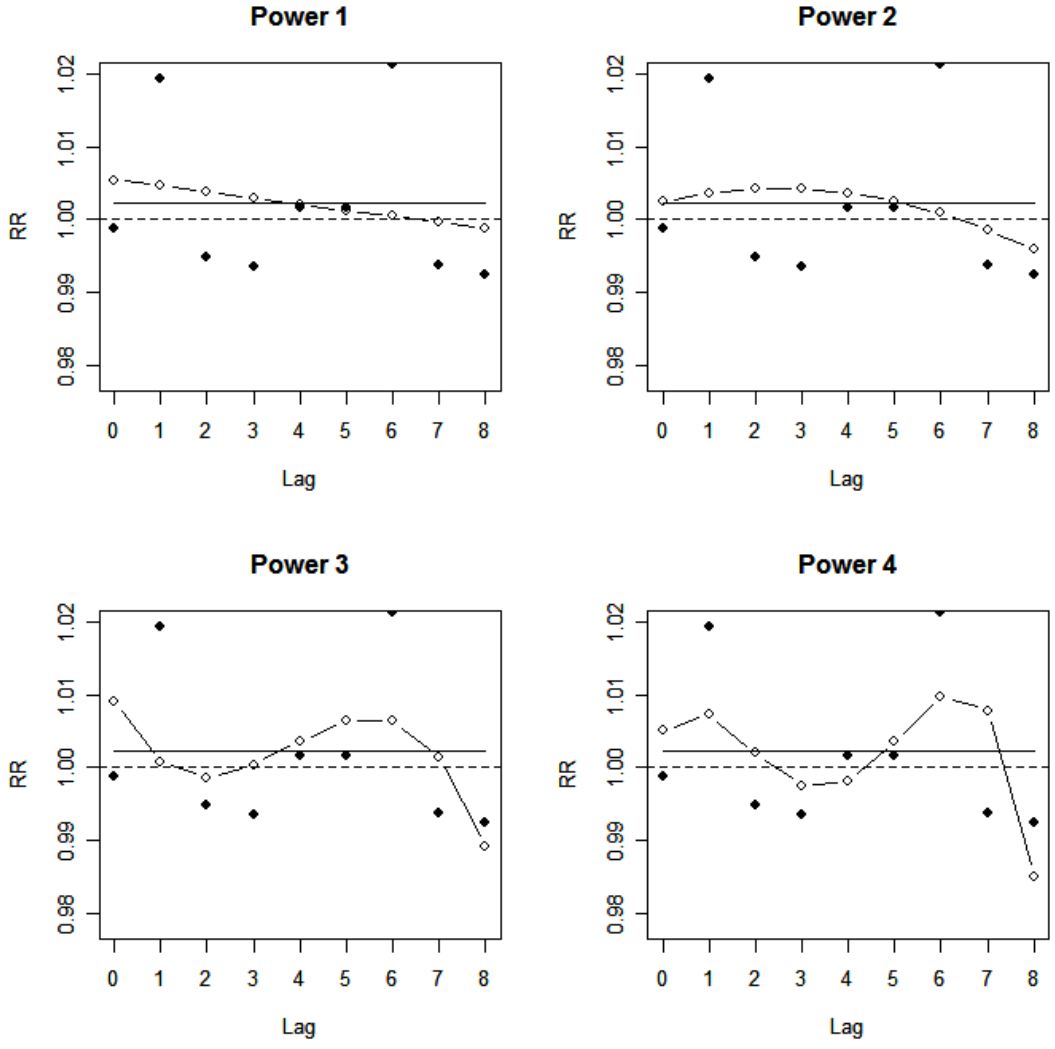

Figure 4.12. The solid line is the relative risk at power zero, the dashed line represents the null risk of 1, the filled diamonds are the relative risks from the unconstrained model and the unfilled diamonds are the relative risks from the distributed lag model

The solid horizontal line in each plot represents the relative risk at power zero, which is the overall estimated effect of exposure to PM_{10}. We can see that while most of the relative risks are fairly near one, at power three the relative risk at lag zero, and at power four the relative risk at lag 6 are comparatively high, being approximately 1.01. However the results are actually non-significant, meaning that there does not appear to be any significant relationship between exposure to PM_{10} and people who are being admitted to hospital with a cardiovascular illness for the first time.

Conclusions

In conclusion the weekly numbers of patients who have been admitted to hospital with a cardiovascular illness for the first time decreased between 2000 and 2006 and showed some signs of seasonality, but they did not appear to have any obvious relationship with mean weekly PM_{10} concentrations. After using natural cubic splines to model the trend, PM_{10} was added to a distributed lag model. Most of the relative risks from this model were close to one, although there were a couple of comparatively high relative risks. However these were non-significant, meaning that there does not appear to be any significant relationship between exposure to PM_{10} and first admissions to hospital with a cardiovascular illness.

4.2 CHD Data

This section contains analyses relating only to CHD. Section 4.2.1 shows the analysis for all admissions to hospital with CHD while Section 4.2.2 is related to mortality from CHD. The analysis of admissions to hospital with CHD among patients under 75 years of age is similar to the analysis of admissions to hospital with CHD among patients of all ages, while the analysis for people who have been admitted to hospital with CHD for the first time is similar to that of patients who have been admitted to hospital with a cardiovascular illness for the first time. Therefore these analyses will not be shown here, and can instead be seen

in Appendix A.

4.2.1 All Ages

Exploratory Analysis

We begin by producing a series of exploratory plots relating to admissions to hospital with CHD. Figure 4.13 shows the number of weekly admissions to hospital over time, while Figures 4.14 and 4.15 show the relationship between admissions to hospital and mean weekly PM_{10} concentrations and mean weekly maximum temperature respectively.

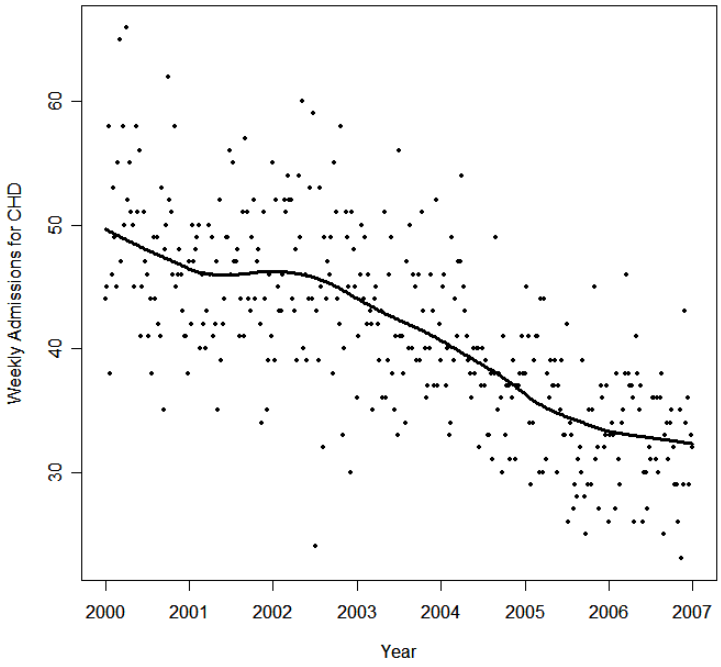

Figure 4.13. Weekly admissions to hospital with CHD

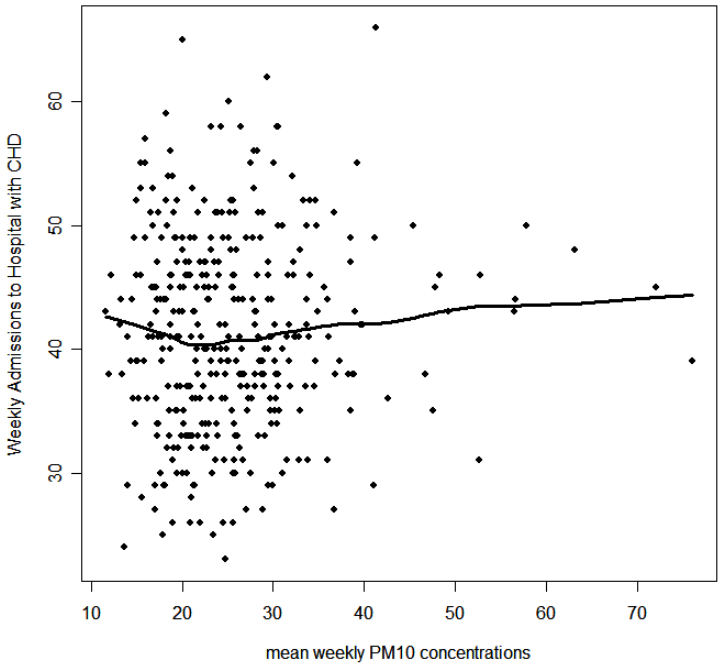

Figure 4.14. Relationship between weekly admissions to hospital with CHD and mean weekly PM_{10} concentrations

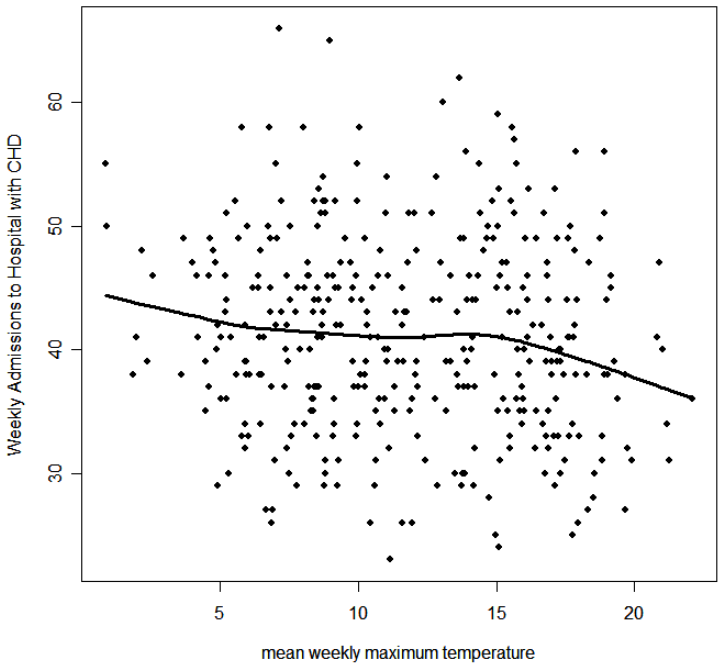

Figure 4.15. Relationship between weekly admissions to hospital with CHD and mean weekly maximum temperature

From Figure 4.13, we can see that weekly admissions to hospital with CHD decreased from 2000 to 2006, except for a period in 2001 where the number of admissions appeared to stay steady. Also there does not appear to be much seasonality in the plot. From Figures 4.14 and 4.15 we can see that there does not appear to be any obvious relationship between admissions to hospital with CHD and either mean weekly PM_{10} concentrations or mean weekly maximum temperature.

Covariate Modelling

The next step is to produce a model that adequately removes the trend and seasonality in the data. While there is a downward trend to the data, it is not a linear trend and so a linear function of time cannot be used. Instead, natural cubic splines with 16 degrees of freedom will be used to model this trend, with the number of degrees of freedom chosen to minimise the AIC. Figures 4.16 and 4.17 show the residuals for this model and the autocorrelation function from the residuals. PM_{10} to the model. From Figure 4.16 the residuals appear to be well scattered about zero, with constant variance. Also there does not seem to be any seasonality or trend. In Figure 4.17, most of the lags are within the 95% confidence band suggesting the correlation with the residuals has now been remove, and therefore we can now continue the analysis by adding

Pollution Modelling

The next step is to add PM_{10} to the model. This will be done using a distributed lag model to constrain the estimates and so reduce collinearity. Figure 4.18 shows the shape of the relative risks from the distributed lag model at powers zero to four, where the solid horizontal line in each plot is the relative risk at power zero, which gives the overall estimate of the effect of PM_{10}. At each power the relative risks from the distributed lag model are quite close to one apart from at lag eight for powers three and four, where the relative risks are approximately 0.985. However this is still fairly close to one and it seems unlikely that there

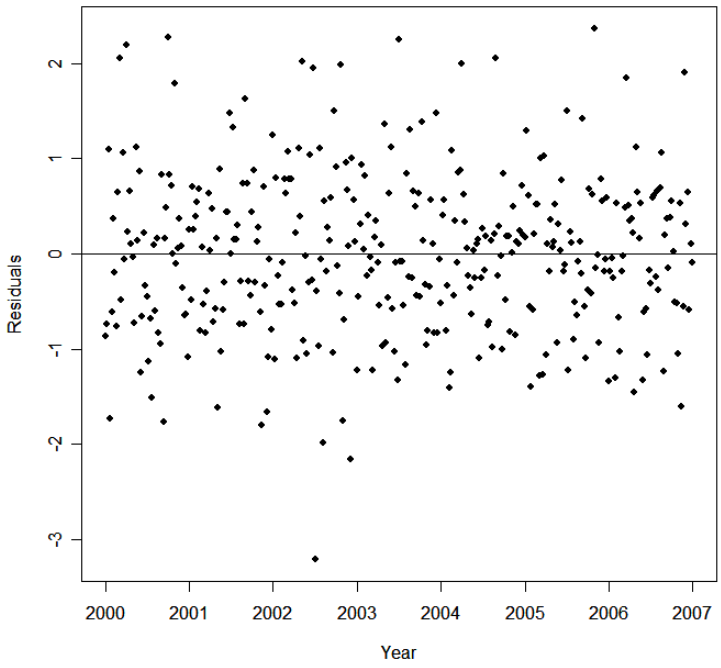

Figure 4.16. Residuals from the model of admissions to hospital with CHD

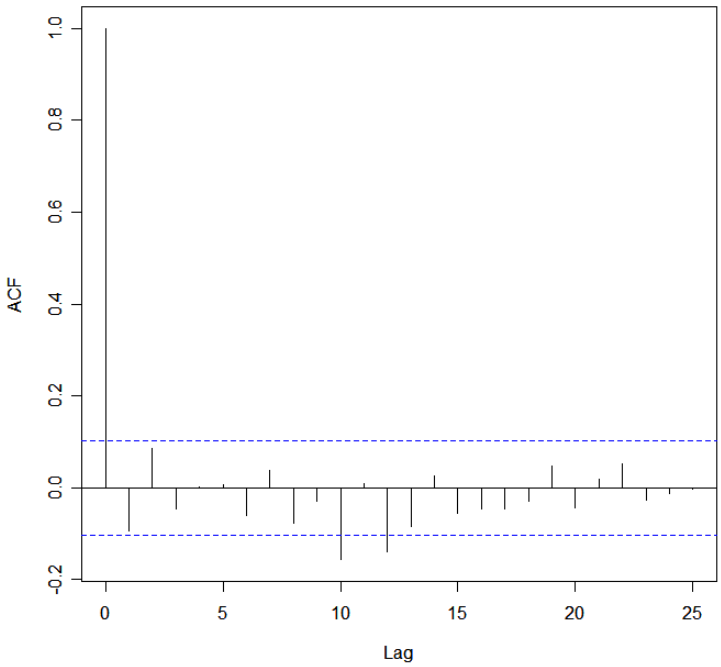

Figure 4.17. Autocorrelation function for the residuals

will be a strong effect of PM_{10} two months after exposure. Also, the relative risks are non-significant, and so we can conclude there is no significant relationship between exposure to PM_{10} and admissions to hospital with CHD.

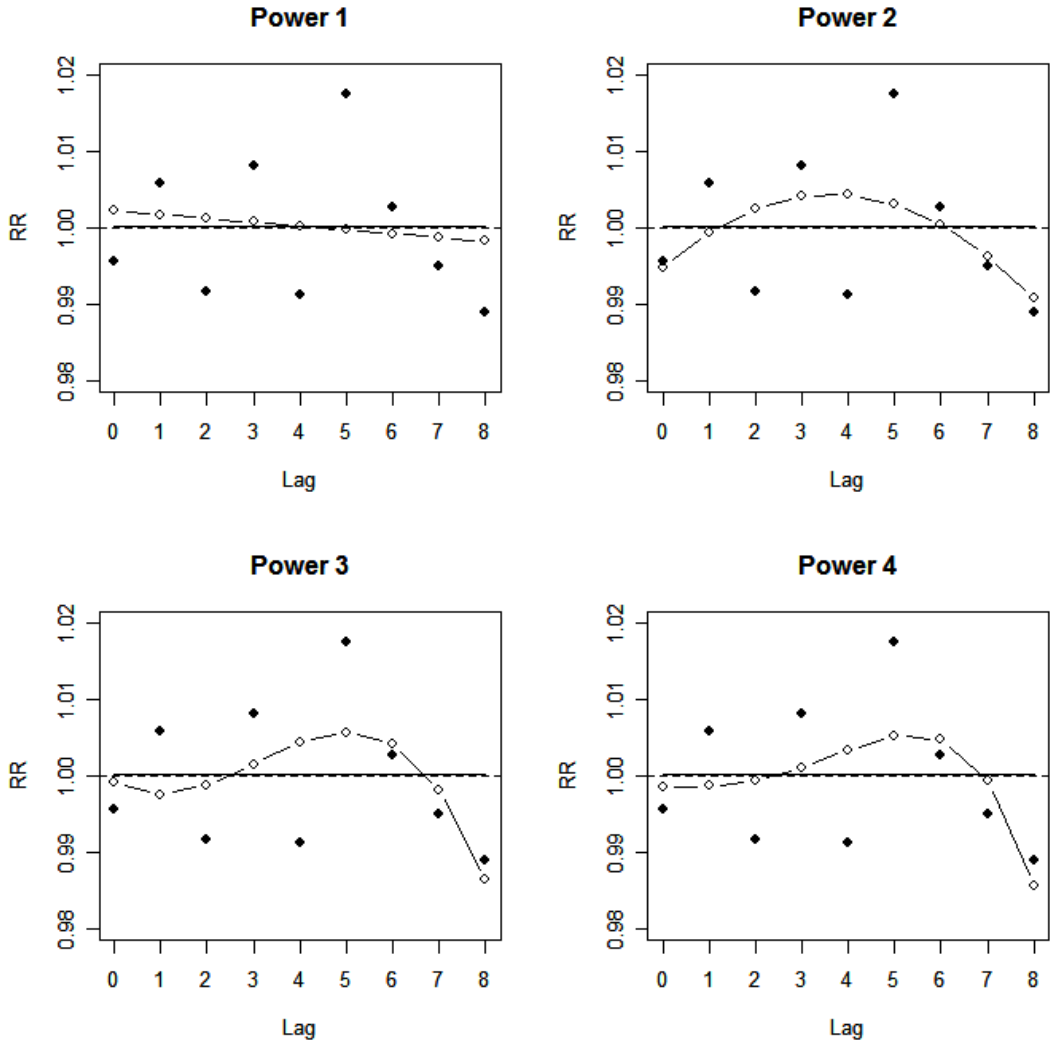

Figure 4.18. The solid line is the relative risk at power zero, the dashed line represents the null risk of 1, the filled diamonds are the relative risks from the unconstrained model and the unfilled diamonds are the relative risks from the distributed lag model

Conclusions

To conclude, weekly admissions to hospital with CHD decreased between 2000 and 2006 apart from a period in 2001 where admissions stayed steady. During the modelling process, natural cubic splines were used to remove the trend and seasonality. After adding PM_{10} concentrations to the distributed lag model, we saw that most of the relative risks were quite close to one, thus suggesting that there does not appear to be any statistically significant effect of PM_{10} on admissions to hospital with CHD in Glasgow. The next section will investigate the relationship between exposure to PM_{10} and mortality from CHD.

4.2.2 Mortality

Exploratory Analysis

Once again, the first step in this analysis is to produce a set of exploratory plots for mortality from CHD. Figure 4.19 shows the numbers of weekly deaths from CHD between 2000 and 2006 while Figures 4.20 and 4.21 show the relationships between weekly mortality from CHD and mean weekly PM_{10} concentrations and mean weekly maximum temperatures respectively. From Figure 4.19 we can see that mortality from CHD has decreased approximately linearly over the years and is seasonal with more deaths in winter than summer. Figure 4.21 shows a slight relationship between weekly mortality from CHD and mean weekly maximum temperature, where the number of deaths decreases as temperature increases, which could account for this seasonality. However Figure 4.20 shows no obvious relationship between weekly deaths from CHD and mean weekly PM_{10} concentrations.

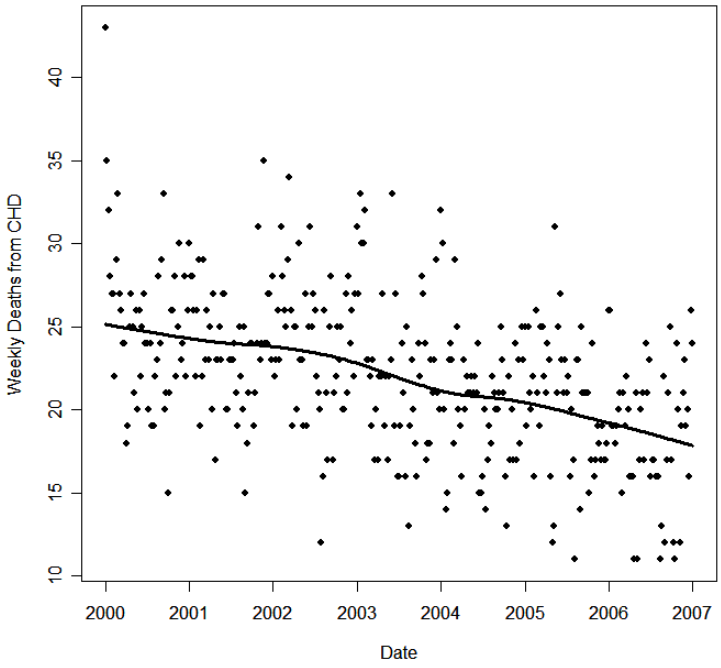

Figure 4.19. Weekly numbers of deaths from CHD

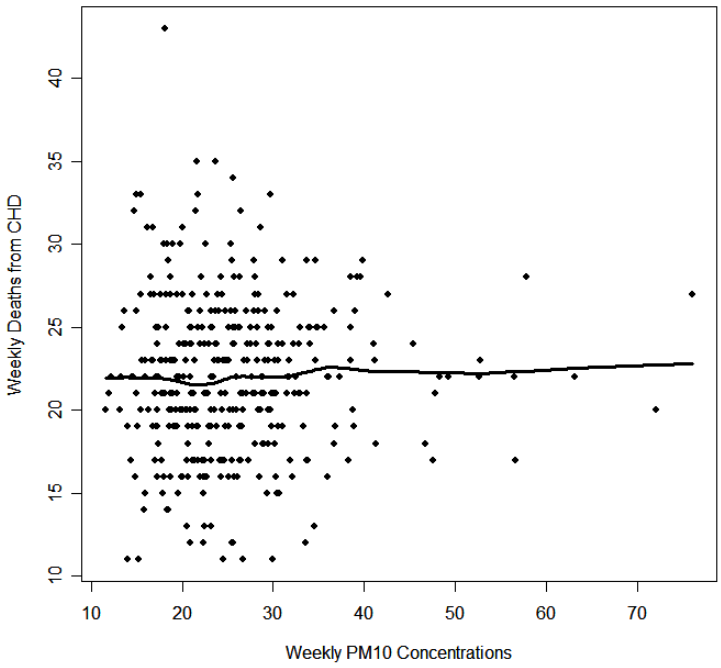

Figure 4.20. Relationship between weekly deaths from CHD and mean weekly PM_{10} concentrations

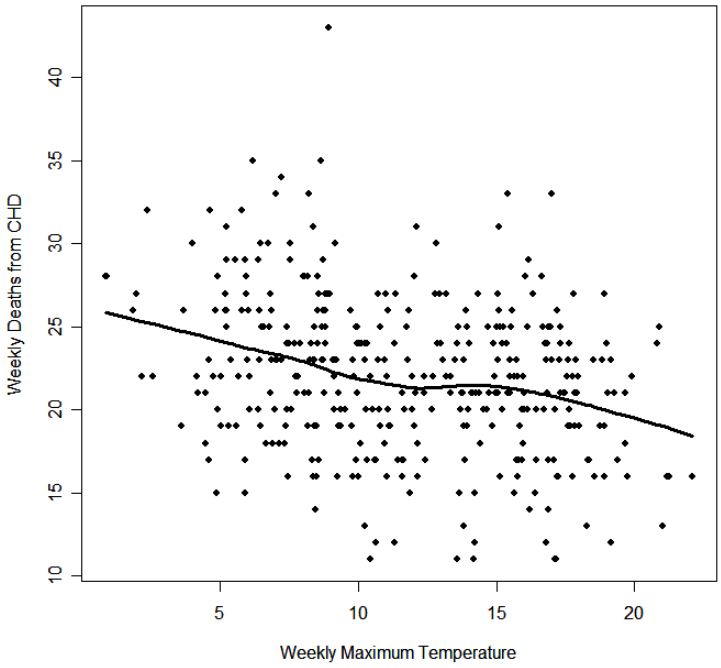

Figure 4.21. Relationship between weekly deaths from CHD and mean weekly maximum temperature

Covariate Modelling

Next we wish to produce a model that removes as much of the trend and seasonality in the data as possible. Originally fixed functions were added to the model, but this did not remove all of the seasonality and therefore natural cubic splines with 19 degrees of freedom were used instead to remove the trend and variation. Figure 4.22 shows the residuals from this model and Figure 4.23 shows the autocorrelation function for the residuals. We can see that the residuals are well scattered about zero, with constant variance and there does not appear to be any seasonality or trend, while in the autocorrelation function most of the lags are within the 95% confidence bands. This implies most of the correlation within the residuals has been removed and suggests the residuals now resemble white noise. Therefore we can continue with the modelling process.

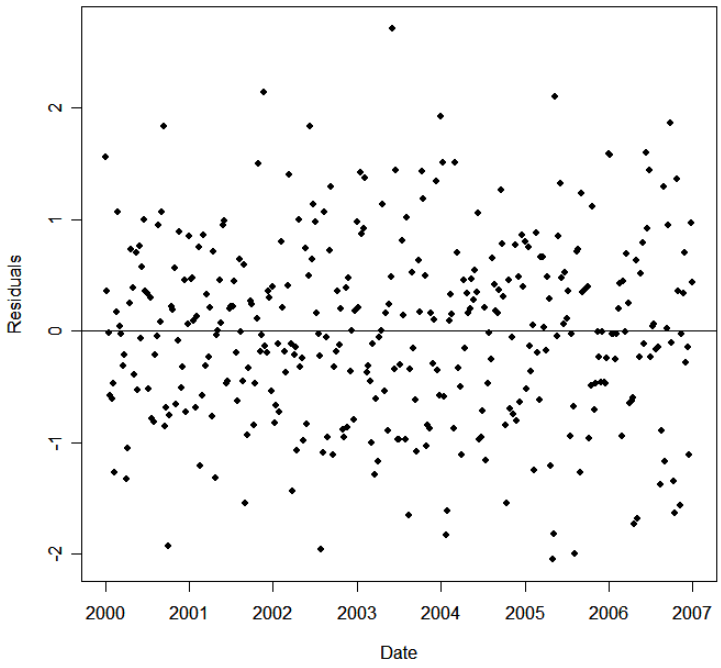

Figure 4.22. Residuals from the model of deaths from CHD

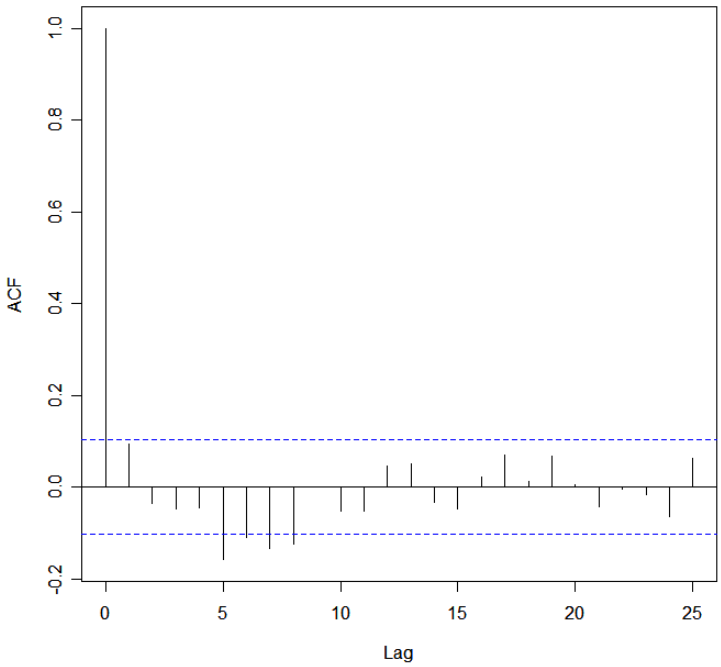

Figure 4.23. Autocorrelation function for the residuals

Pollution Modelling

The next step was to add PM_{10} to the model. As before, a distributed lag model is used. Figure 4.24 shows the shape of the relative risks from the distributed lag model at lags zero to eight, for powers zero to four. Power zero is represented by the solid horizontal line in each of the plots and gives the estimated effect of PM_{10}.

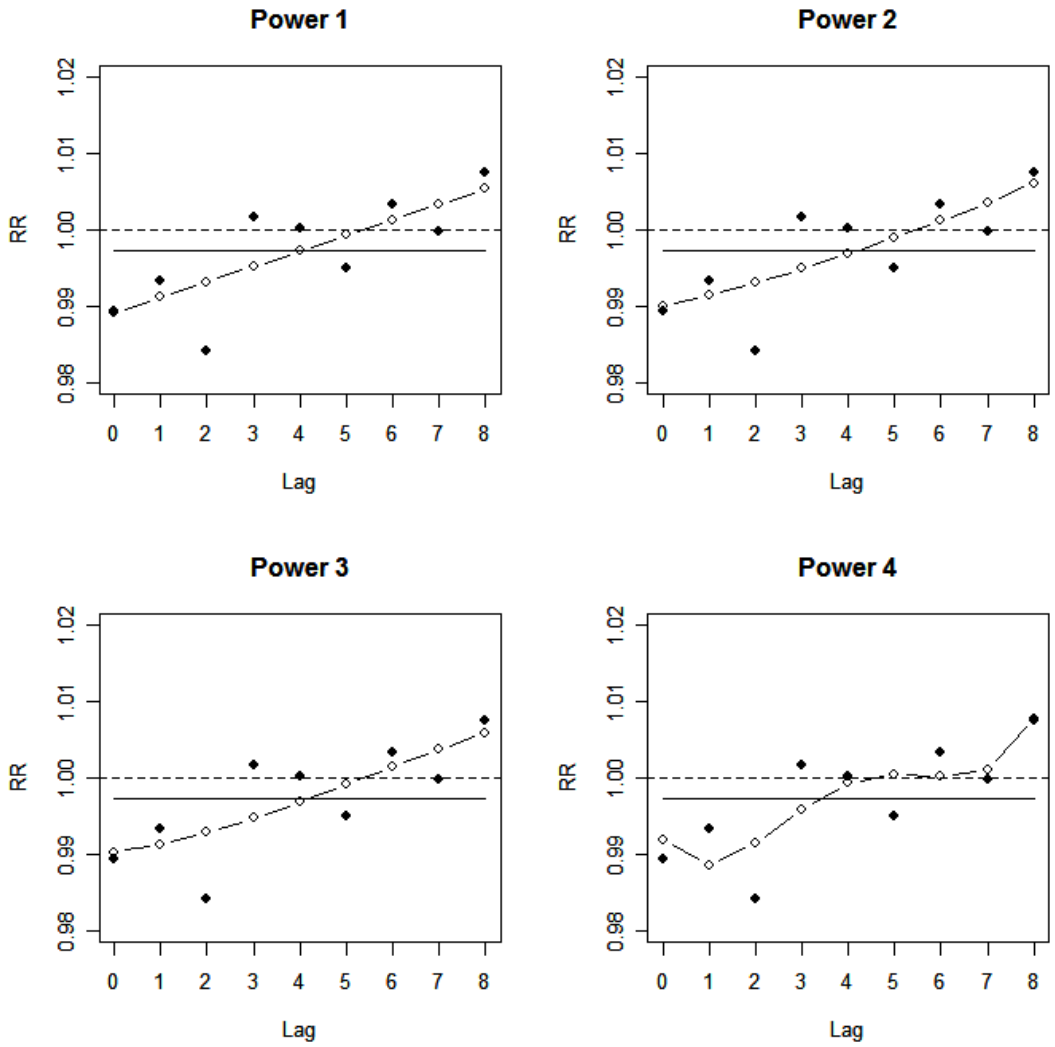

Figure 4.24. The solid line is the relative risk at power zero, the dashed line represents the null risk of 1, the filled diamonds are the relative risks from the unconstrained model and the unfilled diamonds are the relative risks from the distributed lag model

Around lags zero, one and two, the relative risks are fairly low, especially at power four, where they are approximately 0.99. However the relative risks are not significant, meaning that there does not appear to be any relationship between cardiovascular mortality and exposure to PM_{10}.

Conclusions

To conclude, the weekly numbers of deaths from CHD has decreased since 2000 and we saw that it was quite seasonal, with more deaths in winter than in summer. We also saw a slight relationship between weekly mortality from CHD and mean weekly maximum temperature, with the number of deaths decreasing with the higher temperatures. During the modelling process, natural cubic splines were used to remove the trend and seasonality, after which PM_{10} was added to a distributed lag model. The relative risks from the model shows some fairly low values at lags zero, one and two, particularly for powers three and four. However the overall effect of PM_{10} was fairly close to one, and the relative risks were non-significant, implying that there is no statistically significant effect of exposure to PM_{10} on mortality from CHD.

4.3 Stroke Data

This section covers analyses relating solely to admissions to hospital with stroke. Unfortunately, mortality from stroke and first admissions to hospital with stroke could not be analysed as the weekly numbers were too low. The median weekly number of deaths from stroke was 7 and the maximum was 19, while the median number of patients being admitted to hospital with stroke for the first time was 12 with a maximum of 23 per week. Therefore this section will only contain the analysis for admissions to hospital among patients of all ages. The analysis of admissions to hospital among patients under 75 years of age can be found in Appendix A.

4.3.1 All Ages

Exploratory Analysis

The first step is to produce a series of exploratory plots. Figure 4.25 shows the weekly admissions over time while Figures 4.26 and 4.27 show the relationships between weekly admissions to hospital and mean weekly PM_{10} concentrations and mean weekly maximum temperatures respectively. From Figure 4.25 we can see that admissions to hospital stayed steady until around 2004 after which there was a slight decrease. Also the admissions appear to be very seasonal with more admissions in the early part of the year. However from Figures 4.26 and 4.27 there does not appear to be any relationship between weekly admissions to hospital with stroke and either mean weekly PM_{10} concentrations or mean weekly maximum temperature.

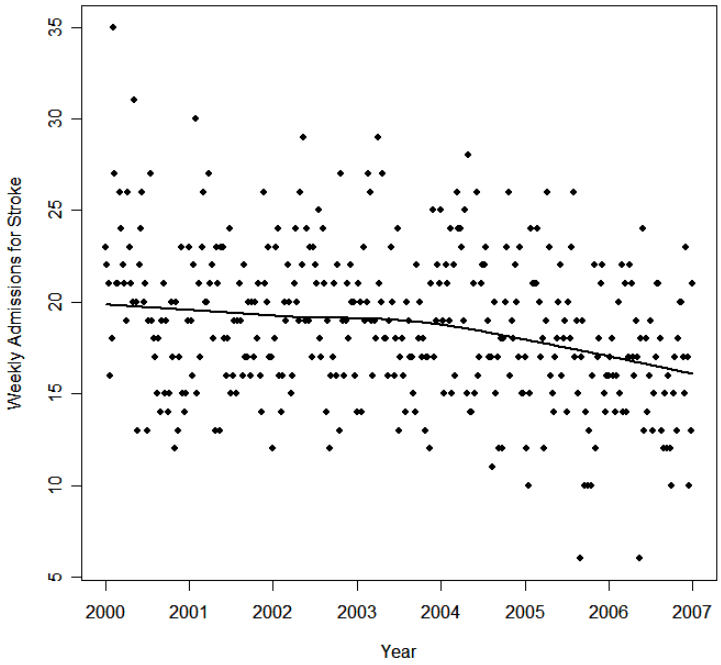

Figure 4.25. Weekly admissions to hospital with stroke

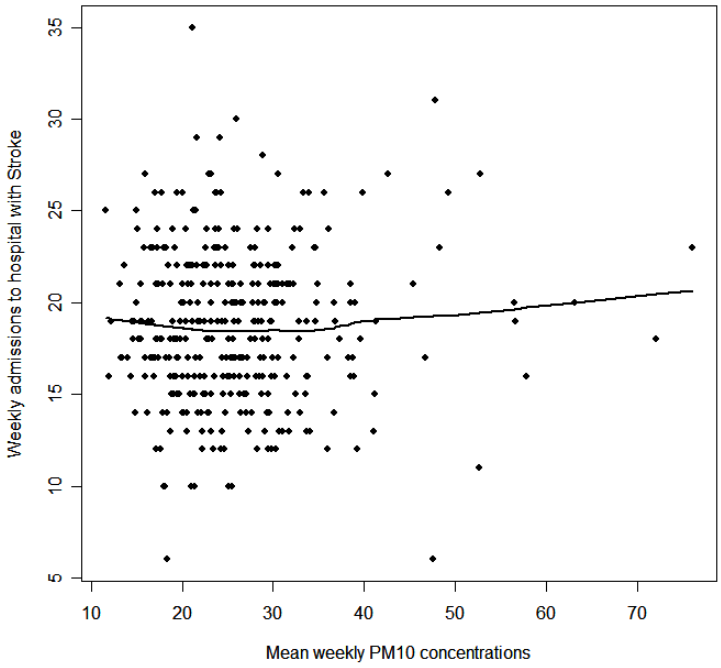

Figure 4.26. Relationship between weekly admissions to hospital with stroke and mean weekly PM_{10} concentrations

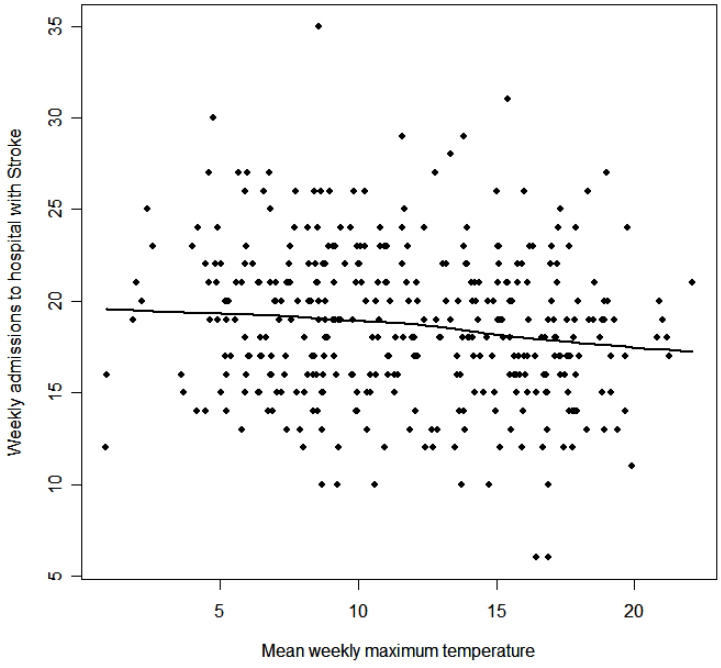

Figure 4.27. Relationship between weekly admissions to hospital with stroke and mean weekly maximum temperatures

Covariate Modelling

The next stage is to produce a model that adequately removes as much of the trend and seasonality in the data as possible. Admissions to hospital with stroke show a downward trend to the data, but it is not a linear trend and so a linear function of time cannot be used. Instead, natural cubic splines with 17 degrees of freedom will be used to model the trend and seasonality. The 17 degrees of freedom were chosen to minimises the AIC. Figure 4.28 shows the residuals from this model while Figure 4.29 shows the autocorrelation function for the residuals. The residuals seem to be scattered evenly about zero with no obvious pattern while most of the lags in the autocorrelation function are within the 95% confidence lags. This suggests the model is adequate and we can now continue with the modelling process.

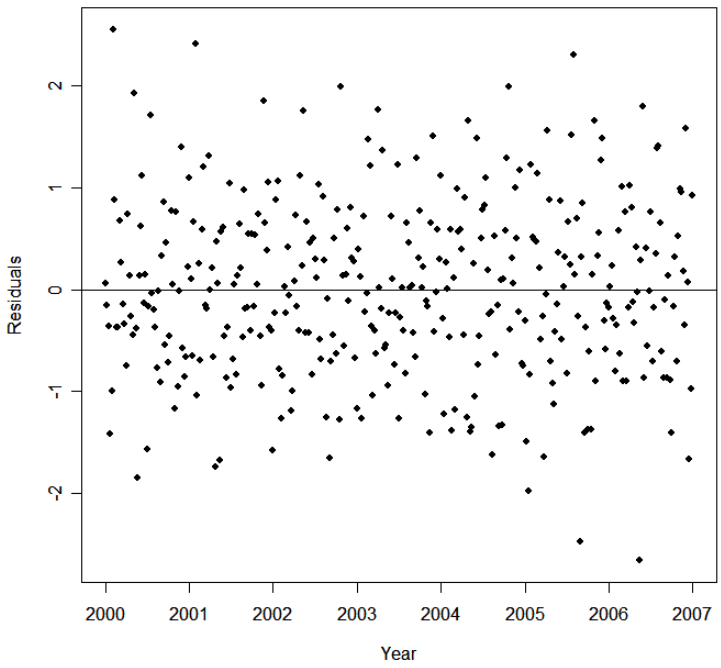

Figure 4.28. Residuals from the model of admissions to hospital with stroke

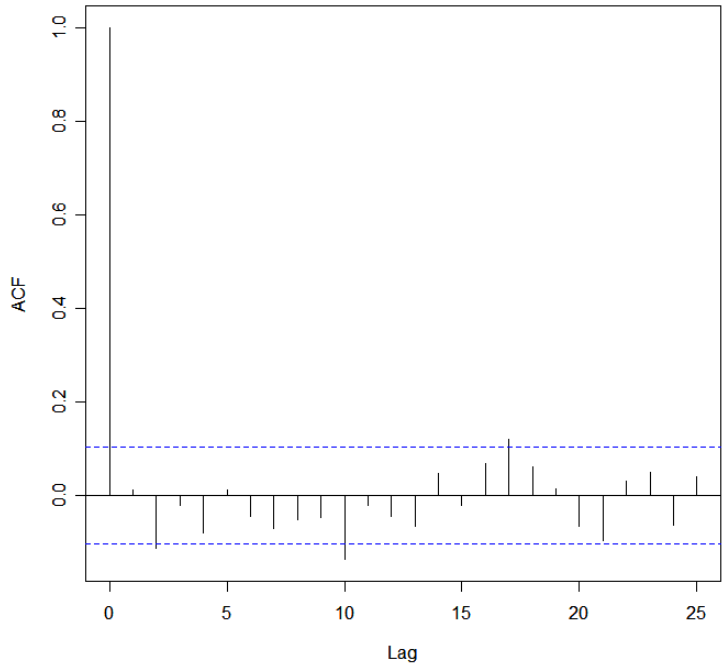

Figure 4.29. Autocorrelation function for the residuals

Pollution Modelling

The next step is to add PM_{10} to the model using a distributed lag model. Figure 4.30 shows the shape of the relative risks from the distributed lag model at lags zero to eight for powers zero to four. The relative risks for power zero are represented by the solid horizontal line in each plot and gives the estimated effect of exposure to PM_{10}. We can see that the relative risks from the distributed lag model are all higher than one except at lag eight, which is much lower at approximately 0.98. However the relative risk coefficients are non-significant meaning that there is no significant effect of exposure to PM_{10} on admissions to hospital with stroke.

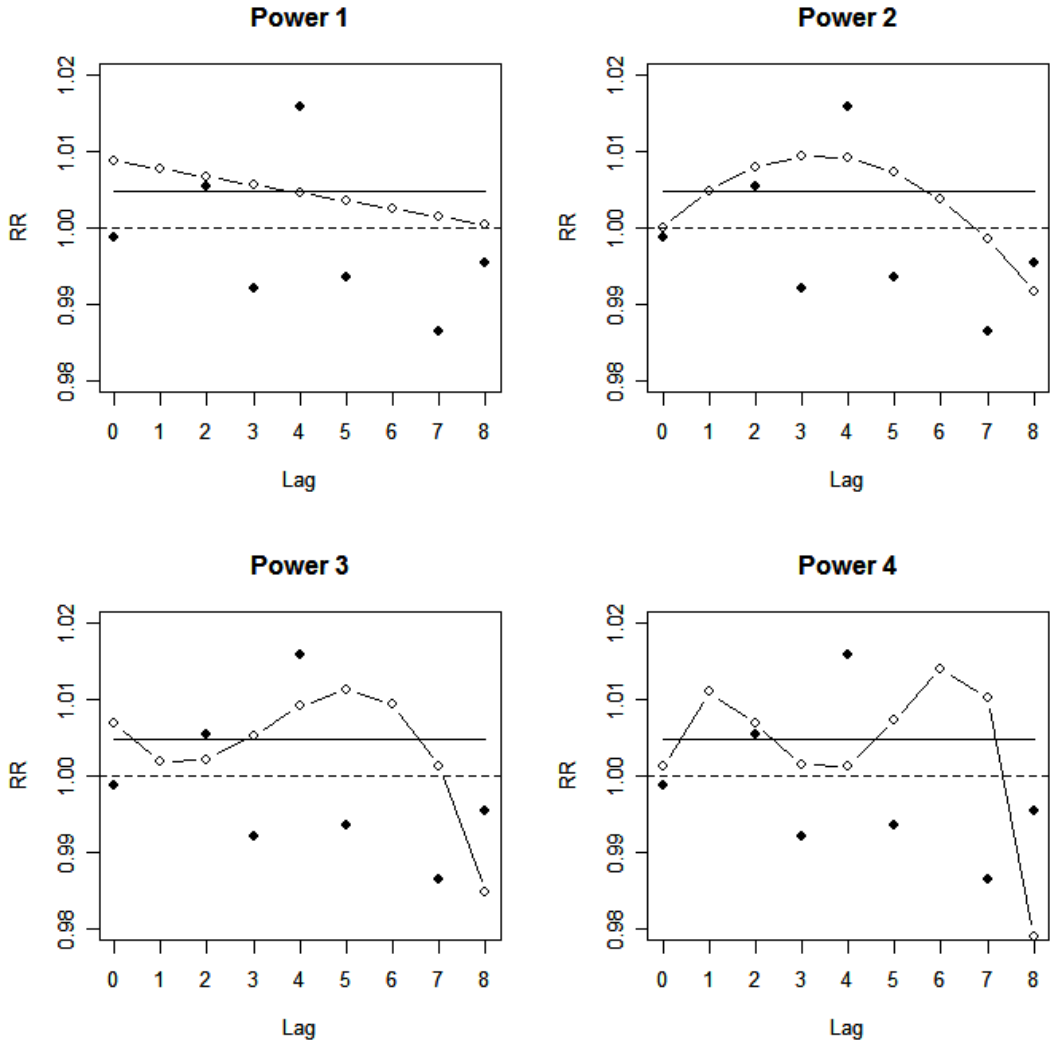

Figure 4.30. The solid line is the relative risk at power zero, the dashed line represents the null risk of 1, the filled diamonds are the relative risks from the unconstrained model and the unfilled diamonds are the relative risks from the distributed lag model

Conclusions

In conclusion, the weekly number of admissions to hospital with stroke have decreased over time and are very seasonal although there does not appear to be any obvious relationship between weekly admissions to hospital with stroke and mean weekly PM_{10} concentrations. After using natural cubic splines to remove

the trend and seasonality, PM_{10} was added to the model using a distributed lag model. The relative risks from this model were all greater than one, except at lag eight. However the relative risk coefficients were non-significant meaning that there was no statistically significant effect of PM_{10} on admissions to hospital with stroke seen in Glasgow.

4.4 Respiratory Admissions

4.4.1 Asthma

The analysis in this section investigates the relationship between weekly numbers of emergency admissions to hospital with asthma and mean weekly PM_{10} concentrations.

Exploratory Analysis

The first step in this analysis is to produce exploratory plots of emergency asthma admissions. Figure 4.31 below shows the number of weekly admissions between 2000 and 2006 while Figures 4.32 and 4.33 show the relationship between asthma admissions and mean weekly PM_{10} concentrations and mean weekly maximum temperature respectively. From Figure 4.31 the first thing we notice is that there appears to be a curved trend to the data, with weekly admissions decreasing until approximately 2003 and then increasing again. There also appears to be some signs of seasonality with the number of admissions to hospital increases in winter. However from Figures 4.32 and 4.33, there does not appear to be any obvious relationship between asthma admissions and either mean weekly PM_{10} concentrations or mean weekly maximum temperature.

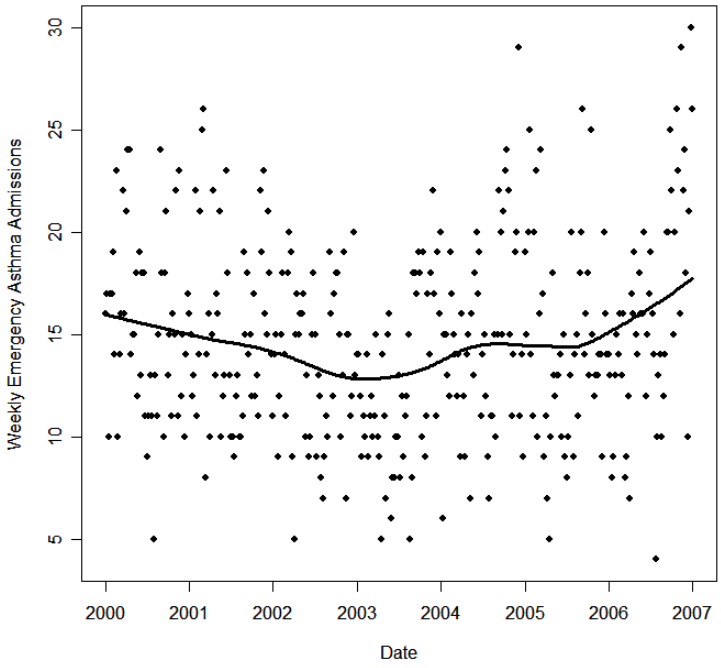

Figure 4.31. Weekly numbers of emergency admissions to hospital with asthma in Glasgow

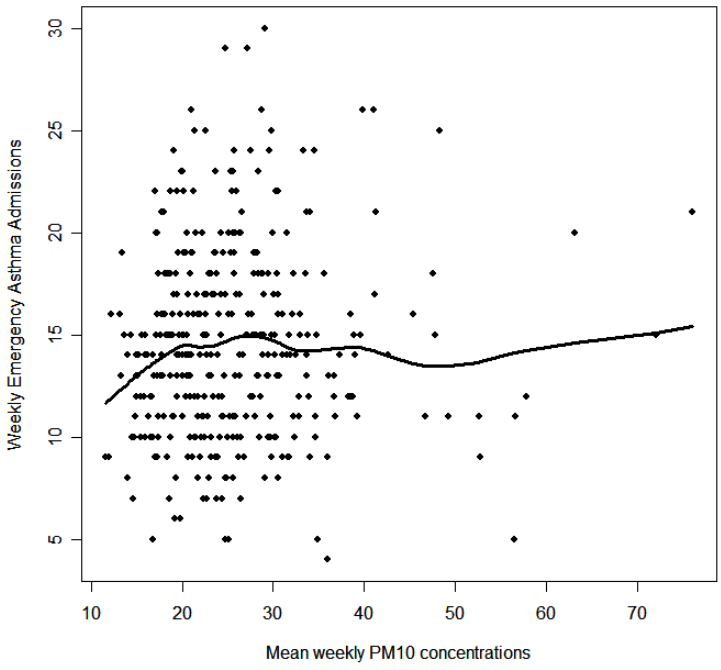

Figure 4.32. Relationship between weekly emergency asthma admissions and mean weekly PM_{10} concentrations

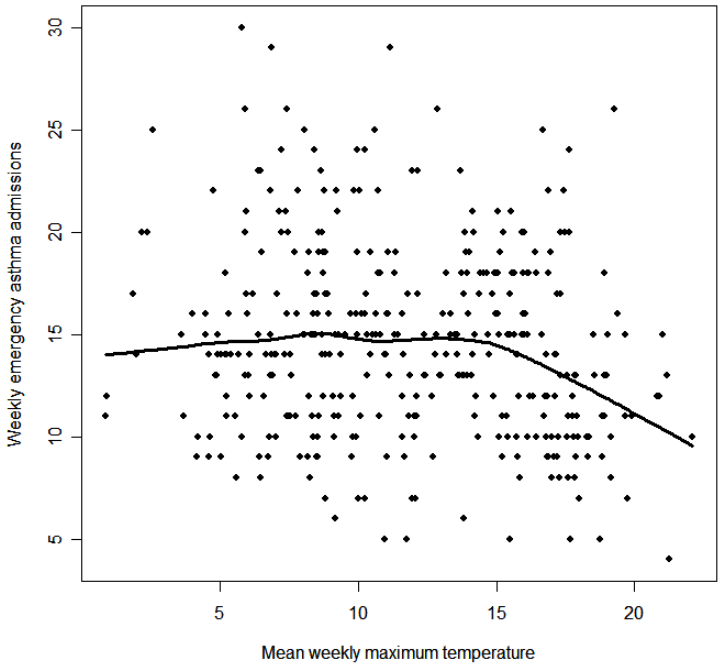

Figure 4.33. Relationship between weekly emergency asthma admissions and mean weekly maximum temperature

Covariate Modelling

The next step is to produce a model that removes the trend and seasonal variation. As the trend is neither linear nor quadratic, natural cubic splines with 16 degrees of freedom have been used. The number of degrees of freedom were chosen as this minimised the AIC. Maximum temperature was also added as it was significantly related to weekly admissions.

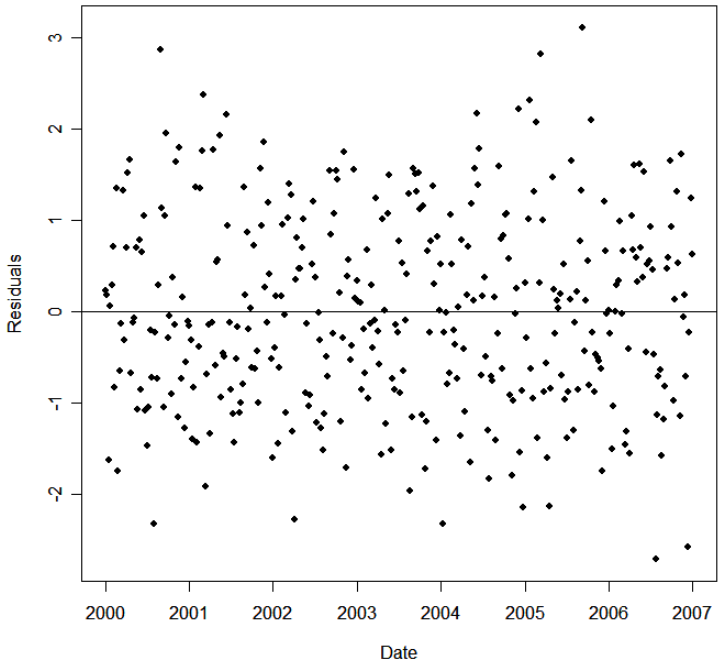

Figure 4.34. Plot of residuals from the first model of admissions to hospital with asthma in Glasgow

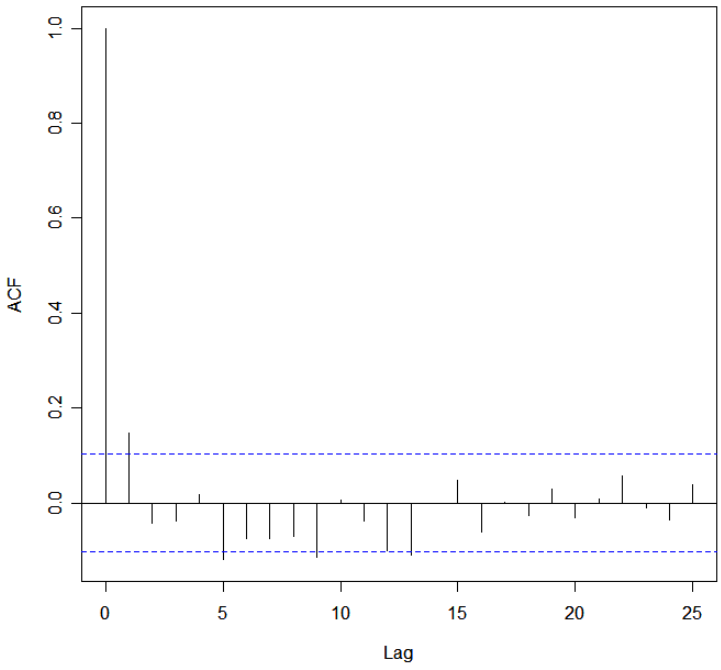

Figure 4.35. Autocorrelation function for the residuals

Figures 4.34 and 4.35 show the residuals from the model and the autocorrelation function for the residuals. The residuals appear to be evenly scattered about zero with no obvious pattern. However the autocorrelation function shows that there is still correlation in the residuals at lag one. There are several methods of removing this correlation, which have been outlined in Chapter 2. The method used here is the observation driven method, in which the previous week's admissions will be added as a covariate. The residuals for this model and the autocorrelation function for the residuals can be seen below (Figures 4.36 and 4.37). We can see that the residuals are still satisfactory and now the correlation at lag one has also been removed, meaning the residuals now resemble white noise and so we can continue with the analysis.

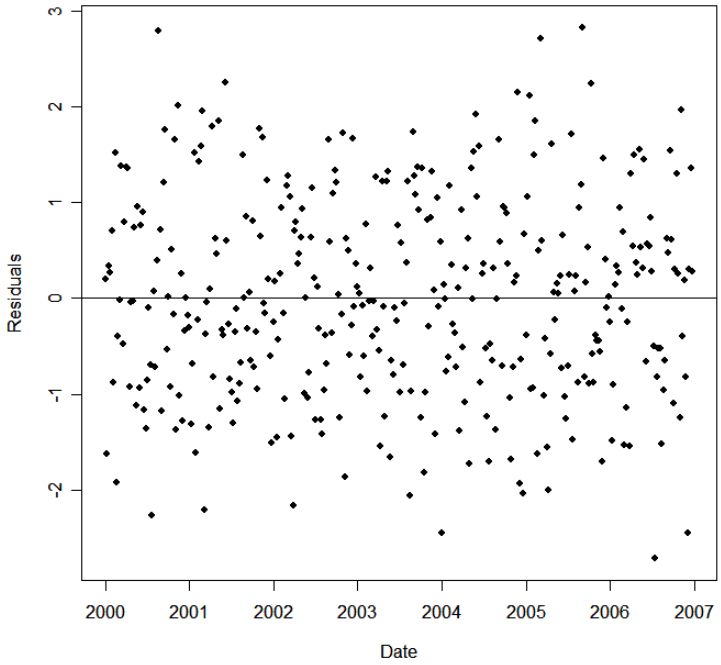

Figure 4.36. Plot of residuals from the second model of admissions to hospital with asthma in Glasgow

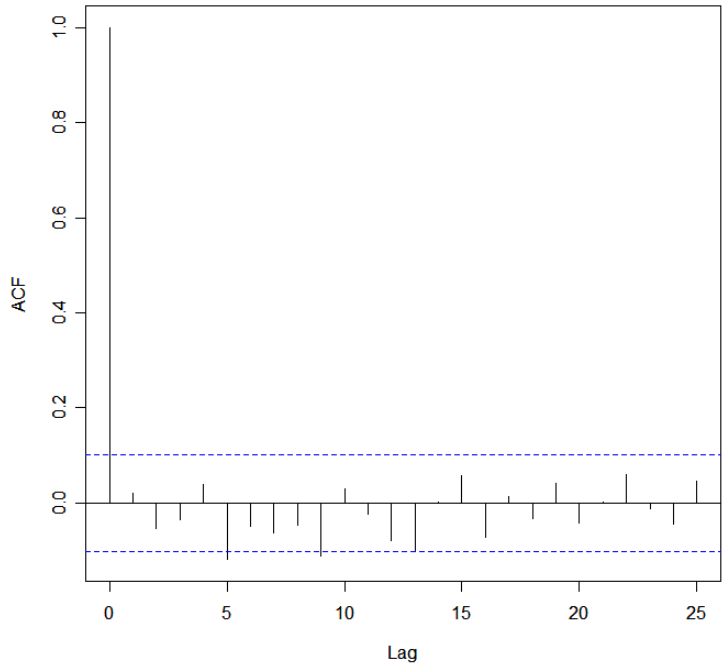

Figure 4.37. Autocorrelation function for the residuals

Pollution Modelling

The next step in the analysis is to add PM_{10} to the model using a distributed lag model. Figure 4.38 shows the shape of the relative risks from the distributed lag model at powers zero to four. The solid horizontal line in each plot represents the relative risk at a power of zero, which gives an overall estimate of the effect of PM_{10}. The dashed horizontal line represents the null value of one. The majority of relative risks are below one and are significant, suggesting that PM_{10} actually has a beneficial effect on asthma admissions. The relative risk for power zero is approximately 0.99, meaning that for a $10\mu g/m^3$ increase in PM_{10} concentrations, admissions to hospital with asthma decrease by approximately 1%. One possible reason for this decrease is that during particularly bad periods of air pollution, asthmatics are more aware of the unhealthy climate and thus are more likely to use inhalers or other medication. A consequence of this would be a reduction in

the number of emergency admissions to hospital.

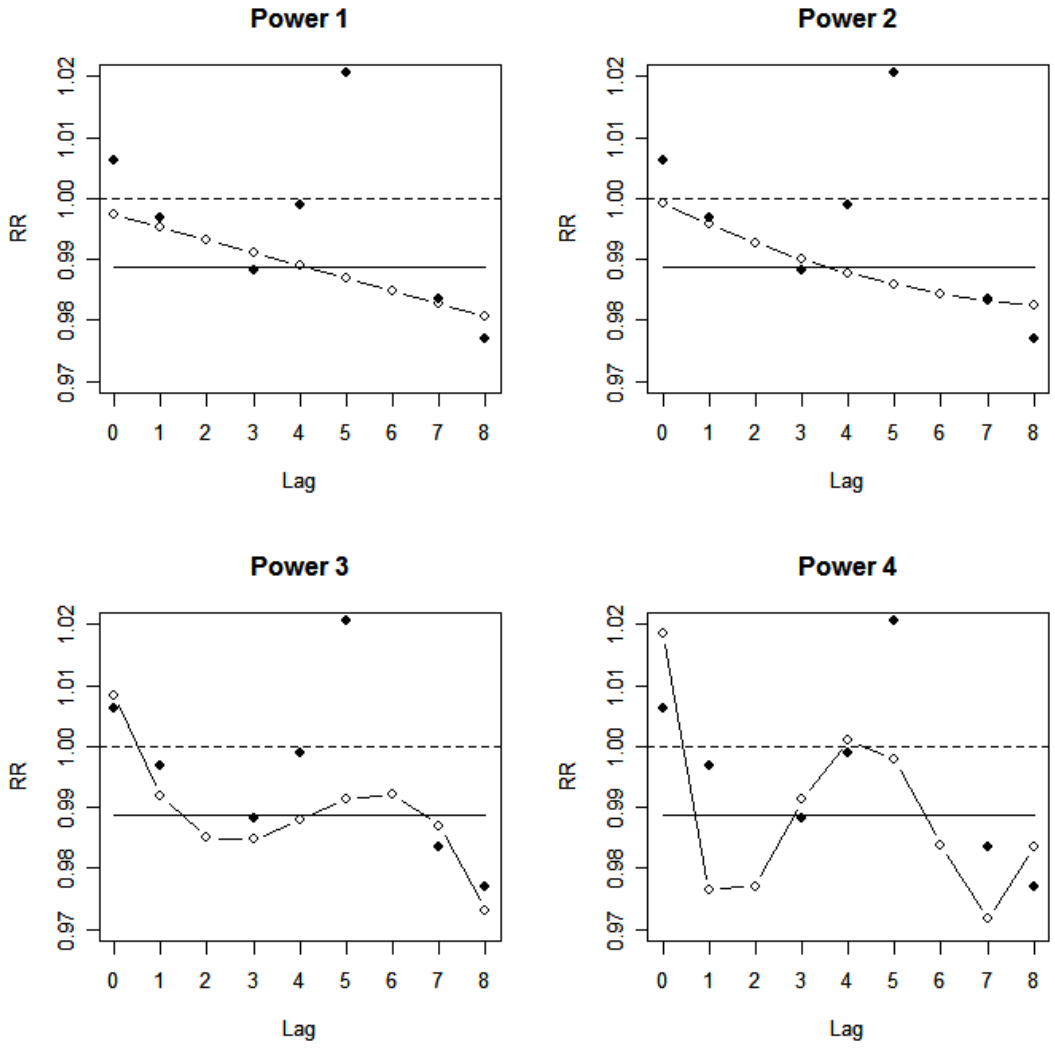

Figure 4.38. The solid line is the relative risk at power zero, the dashed line represents the null risk of 1, the filled diamonds are the relative risks from the unconstrained model and the unfilled diamonds are the relative risks from the distributed lag model

Conclusion

Admissions to hospital with asthma showed a curved trend with a decrease in admissions between 2000 and 2003 and an increase after this point. To remove this trend, natural cubic splines with 16 degrees of freedom were added to the

model. Also, mean weekly maximum temperature and was added as it was significantly related to the data. However this was not enough to remove the temporal correlation within the residuals so the observations from the previous week were also added to the model. This did remove the temporal correlation and so PM_{10} concentrations could be added, using a distributed lag model. The relative risks from this model were mainly below one, with the relative risks for power zero being approximately 0.99 meaning that for a $10\mu g/m^3$ increase in PM_{10} concentrations, admissions to hospital decrease by approximately 1%. Also, the relative risks were significant in this model, suggesting that PM_{10} is, in fact, beneficial to health. One possible reason for these results is that asthmatics may be more sensitive to the variations in pollution concentrations and during periods of particularly high pollution concentrations, they are more likely to use medication, thus reducing the numbers of admissions to hospital. There was a possibility that adding the previous week's observations to the model could amount to over adjustment, thus masking any possible effects of air pollution. However the analysis was repeated without the observations and very similar relative risks were found, thereby suggesting there was no problem with over adjustment.

4.5 Exceedences

One major area of concern is whether or not there is a threshold concentration, above which PM_{10} has a significant effect on health. To this effect, legislation has been put in place in Scotland that aims to reduce annual mean PM_{10} concentrations to $18\mu g/m^3$ and the daily mean should exceed $50\mu g/m^3$ no more than seven times a year. This subanalysis aims to determine whether values that exceed $50\mu g/m^3$ are related to cardiovascular admissions to hospital. For this analysis, daily data will be used rather than weekly data.

Exploratory Analysis

The plot of daily admissions to hospital between 2000 and 2006 in Glasgow can be seen in the previous chapter. Figure 4.39 shows the daily PM_{10} concentrations in Glasgow, with the horizontal line denoting the limit of $50\mu g/m^3$. From this we can see that there were some especially high PM_{10} concentrations in 2002 and 2003, but the number of times the PM_{10} concentrations exceed $50\mu g/m^3$ decreases over time, with very few exceedences from 2004 onwards.

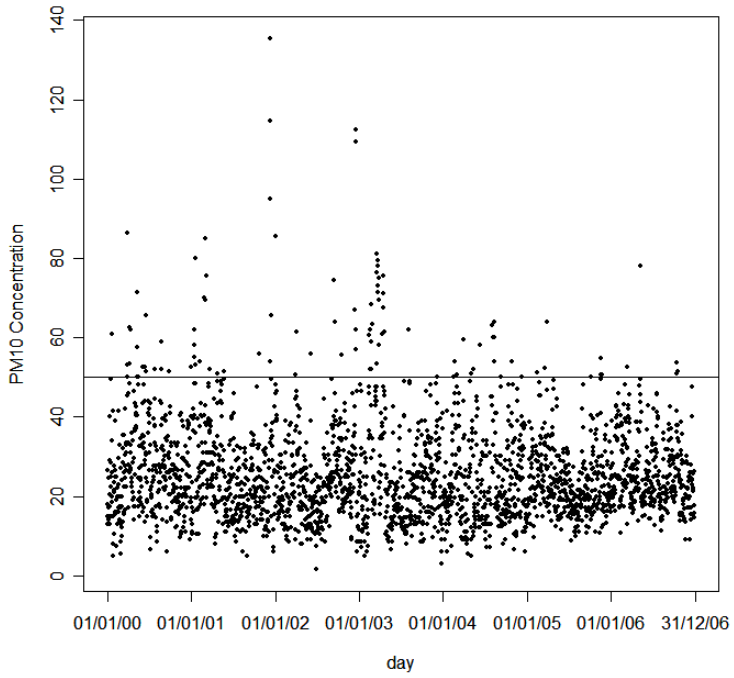

Figure 4.39. Daily mean PM_{10} concentrations in Glasgow where the horizontal line denotes the limit of $50\mu g/m^3$

Covariate Modelling

The first step in the analysis is to produce a model that effectively removes as much of the trend, seasonality and correlation as possible. As this is the same data that was used in the analysis of daily admissions to hospital in Glasgow

in Chapter 3, the summary of this model and the residuals and autocorrelation function can be seen in that chapter (Model (3.5), Table 3.4 and Figures 3.11 and 3.12).

Pollution Modelling

As the model now adequately removes the trend and seasonality, the next step is to add PM_{10} to the model using a distributed lag model with pollution being added as an indicator variable, $I(PM_{10_{t-q}})$, which equals 1 if PM_{10} concentrations are greater than or equal to 50 on day t and 0 otherwise. The relative risks for this model are presented as the increased risk of being admitted to hospital if PM_{10} concentrations cross the threshold value of $50\mu g/m^3$ and can be calculated as $\exp(\hat{\gamma})$. Figure 4.40 shows the shape of the constrained relative risks against lag for orders zero to four. The dashed horizontal line represents the baseline value of one while the solid horizontal line is the relative risk at power zero which gives the overall estimate of the effect of PM_{10} concentrations that are greater than $50\mu g/m^3$. We can see that while most of the relative risks are fairly near one, the relative risks at lags two and three for a power of 4 are particularly high, at approximately 1.03, which could suggest a significant effect of excessively high values of PM_{10}. However, these results are non-significant thus suggesting that PM_{10} concentrations greater than $50\mu g/m^3$ do not have a significant effect on cardiovascular admissions to hospital.

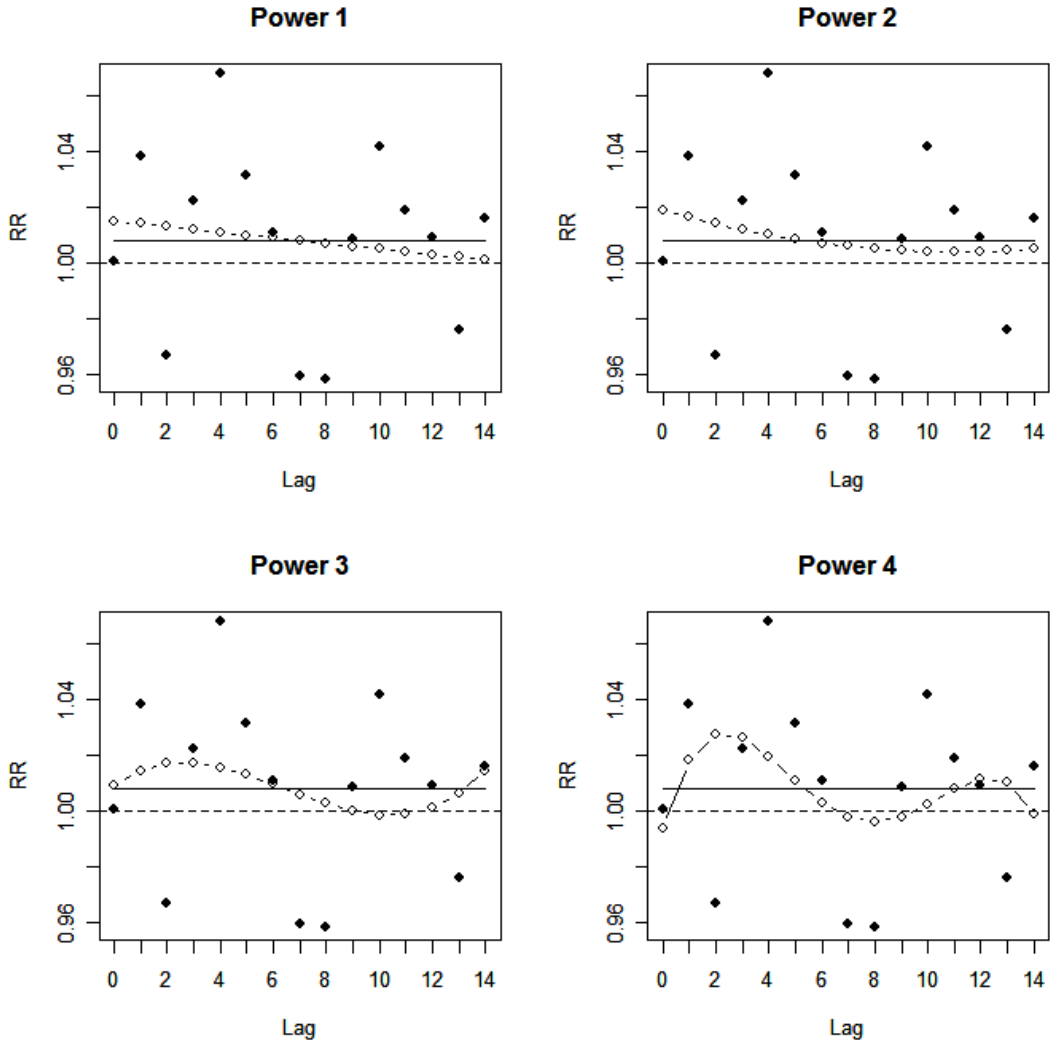

Figure 4.40. The solid line is the relative risk at power zero, the dashed line represents the null risk of 1, the filled diamonds are the relative risks from the unconstrained model and the unfilled diamonds are the relative risks from the distributed lag model

Conclusion

Daily numbers of admissions to hospital with a cardiovascular illness have decreased between 2000 and 2006, as have the number of times PM_{10} exceeded $50\mu g/m^3$. During the modelling process, a linear function of time was added to remove the time trend while maximum temperature and sinusoidal terms were

added to remove the seasonality. Also indicator variables for 'day of the week' were added to remove the temporal correlation. PM_{10} was then added using a distributed lag model. Most of the relative risks from this model were close to one, although at lags two and three for a power of four, the relative risks were higher. However the relative risks were all non-significant, meaning that there does not appear to be any statistically significant effect of PM_{10} concentrations over $50\mu g/m^3$ on admissions to hospital with cardiovascular illnesses in Glasgow.

4.6 Conclusions and Discussion

After completing the analysis of the effects of air pollution on cardiovascular morbidity in Glasgow, Edinburgh and Aberdeen, a number of further subanalyses were carried out. These could only be performed on the Glasgow data as it involved splitting up the data, and only Glasgow had enough admissions for this to be feasible. Even so, there were still two analyses that could not be carried due to the small numbers. First admissions to hospital with stroke had a median of seven admissions per week, with a maximum of 14, and mortality from stroke also had a median of seven deaths a week, with a maximum of 19.

The first analysis carried out in this chapter was for admissions to hospital due to cardiovascular illnesses among males. This was followed by the analysis relating to first admissions to hospital with a cardiovascular illness. Neither of these analyses showed any significant effect of exposure to PM_{10}. An analysis relating to admissions to hospital with a cardiovascular illness among females was also carried out and again no significant effect of exposure to PM_{10} was found. This analysis can be seen in Appendix A. The next section contained the analyses relating to admissions to hospital and mortality from CHD. The first of these was admissions to hospital with CHD among patients of all ages, followed by the analysis relating to mortality from CHD. Again, neither of these analyses showed any significant effect of exposure to PM_{10}. The analyses of admissions to hospital

with CHD among the under 75s and first admissions to hospital with CHD also showed no effect of exposure to PM_{10} and can be found in Appendix A. Next, the analysis of admissions to hospital with stroke among patients of all ages was shown and again showed no significant effect of PM_{10}. The analysis of admissions to hospital with stroke among the under 75s also showed no effect of exposure to PM_{10} and can be seen in Appendix A.

The next analysis was carried out on respiratory data, specifically emergency admissions to hospital with asthma. After analysing these data, it was found that PM_{10} actually has a negative effect, i.e. the higher the PM_{10} levels, the fewer emergency admissions to hospital there would be. One suggestion as to why this was found, is that asthmatics would notice the increased pollution and use inhalers and other medication more often, thus reducing the number of emergency admissions.

Recently a paper was published by Carder et al. (2008) which was also interested in the relationship between air pollution and health in Scotland. Their data spanned 21 years from 1981 to 2001. The health data related to all cause mortality, respiratory mortality, cardiovascular mortality and non-cardiorespiratory mortality while the pollution data related to mean daily black smoke concentrations. The results found in this study are consistent with those found by Carder et al. (2008). They also found no effect of PM_{10} on cardiovascular mortality. However while they found a relationship between PM_{10} and respiratory mortality, they found a harmful effect on health. This discrepancy may be due to the fact that their data related to mortality, while the data in this study related to morbidity.

The final analysis carried out in this section was relating to PM_{10} concentrations that exceeded $50\mu g/m^3$. This analysis used an indicator variable for PM_{10}, where it was equal to 1 if the PM_{10} concentration on day t was $50\mu g/m^3$ or more

and 0 otherwise. The results showed no significant effect of exposure to particularly high concentrations of PM_{10} on admissions to hospital with a cardiovascular illness in Glasgow.

Chapter 5

Conclusions and Discussion

This thesis has two main aims the first of which is to investigate the effects of PM_{10} exposure on cardiovascular illnesses in Scotland, focusing specifically on the three largest cities, Glasgow, Edinburgh and Aberdeen in Chapter 3. This multi-city study made use of data that are routinely available from various government bodies including the Information Services Division of the NHS (ISD) and the Scottish Air Quality website. The health data used in this project were provided by ISD and took the form of daily counts of health events, over the seven year period from 2000 to 2006 for each of Glasgow, Edinburgh and Aberdeen. Cardiovascular illnesses were considered to be either CHD (ICD-10 I20-I25) or stroke (ICD10 I61, I63, I64). The greater part of this thesis focused on admissions to hospital due to cardiovascular morbidity, although cardiovascular mortality and asthma data were also available. The pollution data were available from the Scottish Air Quality website (http://www.scottishairquality.co.uk/) and comprise mean daily PM_{10} concentrations from 2000 to 2006 for each city. PM_{10} was chosen as the pollutant of interest as the effect of particles has become a major issue recently due to their small size, which enables them to penetrate further into the lungs. In addition, temperature data were provided by the British Atmospheric Data Centre (http://badc.nerc.ac.uk/) and comprise daily maximum and minimum temperatures for the seven year period for each of the three cities. The second aim of the thesis is to determine whether routinely collected data are suitable for

estimating the effects of pollution exposure on cardiovascular health.

The first two chapters in this thesis describe the background to this study as well as air pollution and health studies in general. In particular Chapter 1 gives details of the data while Chapter 2 critiques some of the more recent literature relating to air pollution and health studies. Chapter 3 presents the multi-city study for Glasgow, Edinburgh and Aberdeen, focusing on all cardiovascular morbidity. In contrast Chapter 4 analyses sub-sets of these hospital admissions, including splitting up the data by age, sex, main diagnosis (CHD or stroke) and whether or not it was a first incidence. In addition, Chapter 4 also provides analyses for subsets of the mortality data. The last analysis relating to the cardiovascular data is concerned with the effects of particularly high concentrations of PM_{10}. Finally this chapter also investigates the effects of pollution on admissions to hospital with asthma.

5.1 Overall Results

As previously mentioned, Chapter 3 describes the multi-city study into the effects of PM_{10} on cardiovascular health, which was implemented using Poisson generalised linear models. The first part of the analysis was for Glasgow. After constructing a model that contained a linear function of time, maximum daily temperature, indicator variables for day of the week and sinusoidal curves, we found that the relative risks from the single lag models ranged from 0.991 to 1.0008 while the relative risks from the multiple lag models ranged from 0.991 to 1.006. In all cases, the confidence intervals contained one, suggesting there is no significant relationship between admissions to hospital and PM_{10} exposure.

During the course of the investigation, it was discovered that if the 'day of the week' term was left out of the model, there were significant relative risks at lags zero, seven and 13, with the relative risks being approximately 1.02. Also the

relative risks for lags zero to seven exhibit a similar shape to that proposed by Zanobetti et al. (2002) with regards the theory of mortality displacement (discussed further in Section 5.2). If this theory is valid, it means that the public health impacts of air pollution are significantly lower. However a closer look at the relative risks showed that the results were only significant approximately every seven days, which points to the lack of the 'day of the week' term as being the reason for the significant results. To overcome the 'day of the week' problem, the admissions were aggregated up to weekly levels and the analysis was redone, resulting in relative risks between 0.987 and 1.005, where the confidence intervals again contained one, meaning no significant relationship between admissions to hospital and PM_{10} exposure being found, which suggests there is no real effect of exposure to PM_{10}.

The next analysis to be carried out was for daily admissions to hospital in Edinburgh. The relative risks found ranged from 0.984 to 1.013, although as the confidence intervals all contained one, it suggests there is no significant relationship between exposure to PM_{10} and admissions to hospital with cardiovascular illnesses. However the numbers of admissions were lower than those for Glasgow (mean daily admissions of 6.2 in Edinburgh as opposed to 10.5 in Glasgow), so one possibility was that any effect of PM_{10} may not have been seen due to the small numbers. Therefore, the data were aggregated up to weekly levels and the analysis redone. The single lag models produced relative risks of between 0.988 and 1.024, while the relative risks from the multiple lag models were similar at 0.978 to 1.03. Again none of the confidence intervals were significant resulting in no significant relationship between admissions to hospital and PM_{10} being found.

The final analysis carried out in this chapter focused on Aberdeen. After examining the data, it was immediately obvious the data would need to be aggregated to weekly levels as the number of admissions in any one day was very low (mean of 3.2). This analysis was carried out using weekly admissions to hospital and mean

weekly PM_{10} concentrations. The single lag models showed no significant effect of PM_{10}, with relative risks ranging from 0.99 to 1.03, with the confidence intervals all containing one. The relative risks from the multiple lag models ranged from 0.983 to 1.04 with a significant risk at lag one. However this was the only significant relative risk and the multiple lag models were known to suffer from collinearity, therefore it was assumed that this result was the effect of multiple testing, and in fact there is no significant relationship between cardiovascular admissions to hospital in Aberdeen and exposure to PM_{10}.

Chapter 4 was split up into five sections, the first of which contained analyses relating to admissions to hospital with both illnesses. The second section was concerned with CHD and the third with stroke. Section 4.4 contained the analysis of emergency admissions to hospital with asthma and the final section was concerned with PM_{10} concentrations that exceeded $50\mu g/m^3$. This value was chosen because the National Air Quality Strategy recommends reducing the daily mean concentrations so that they exceed $50\mu g/m^3$ no more than seven times a year. For brevity in these analyses, only the distributed lag models were presented in this thesis, as it was thought the single lag models would not show the effect of cumulative exposure to air pollution and because it seems unlikely that the effect of air pollution on any one day is unrelated to pollution concentrations on adjacent days, while the multiple lag model had the problem of collinearity.

In section 4.1, the first of the subanalyses to be carried out related to admissions to hospital among males, followed by admissions to hospital among females. The next analysis to be carried out was for patients who were being admitted to hospital with a cardiovascular illness for the first time. Again, no effect of PM_{10} exposure was seen. In all these analyses, the relative risk at power zero was approximately 1.005 and non-significant, suggesting there is no significant effect of PM_{10}.

In section 4.2, the next analyses were for admissions to hospital with CHD among patients of all ages, where the relative risks at power zero were extremely close to one at 1.0003. This was followed by the analysis for admissions to hospital among the under 75s which gave relative risks at power zero of 1.001. Neither of these analyses showed any significant relationship between admissions to hospital and exposure to PM_{10}. These were followed by the analysis of patients being admitted to hospital with CHD for the first time, where the relative risks at power zero were also very close to one at 1.002 and also showed no significant effect of PM_{10}. The final analysis in this section was related to mortality from CHD. This gave relative risks at power zero of 0.997 which were shown to be non-significant, again meaning there was no significant effect of PM_{10}. Section 4.3 contained the analyses of admissions to hospital with stroke among all patients and among patients aged 75 and under. Both of these analyses showed relative risks at power zero of approximately 1.005 and were non-significant, meaning there does not appear to be any significant relationship between admissions to hospital and exposure to PM_{10}.

Section 4.4 contained the analysis of emergency admissions to hospital with asthma. After analysing these data, a significant effect of PM_{10} was found. However this effect was a negative one, with a relative risk of approximately 0.99, suggesting that an increase of $10\mu g/m^3$ in PM_{10} concentrations results in a 1% decrease in admissions to hospital. One possible reason for this is that asthmatics may be more likely to use medication during periods of high pollution concentrations, thus reducing the numbers of hospital admissions.

The final analysis carried out was to determine whether PM_{10} concentrations higher than $50\mu g/m^3$ have a significant effect on cardiovascular admissions to hospital. PM_{10} was transformed to an indicator variable, with concentrations of $50\mu g/m^3$ or higher coded as 1, while concentrations under $50\mu g/m^3$ were coded as 0. This analysis gave comparatively high relative risks at power zero of 1.008, but

they were non-significant meaning there was no significant relationship between PM_{10} concentrations and cardiovascular admissions to hospital. This model is somewhat unrealistic as it suggests that if there is an effect of PM_{10} at higher concentrations, then this effect is the same at all concentrations above $50\mu g/m^3$. Also, it implies that the relationship between PM_{10} and cardiovascular admissions to hospital is not continuous, with no effect at concentrations below $50\mu g/m^3$ while at concentrations above this threshold, there is a sudden effect.

There are a number of possible reasons for the lack of significant effects of PM_{10} the first of which is that there may be no relationship between exposure to PM_{10} and cardiovascular health in Glasgow, Edinburgh or Aberdeen. A second reason is that there may be a threshold level below which PM_{10} does not affect the cardiovascular system, and in Scotland PM_{10} concentrations are below this level. However a third possibility is that the data available were not suitable for this analysis. One major concern in this regard is that the pollution data available are not representative of the pollution intake of the population as they are at fixed points around the city while it is unlikely that anyone will remain in the vicinity of a monitoring site throughout the day. Also this does not take into account people who work in one city but live in another and so will be allocated the average pollution level in the city they live in, even if they spend more time in the city they work in. Another area of concern is that these data do not account for the time spent indoors, where the pollution levels may be different from external pollution concentrations.

Many studies have shown a significant effect of PM_{10} on cardiovascular morbidity and mortality. For a $10\mu g/m^3$ increase in PM_{10} concentrations Ballester et al. (2006) found that admissions to hospital with CVD increased by 0.9% (RR: 1.009, 95% CI: 1.004, 1.015) while Touloumi et al. (2005) found that cardiovascular mortality increased by 0.48% (RR: 1.0048, 95% CI: 1.0027, 1.007). However there have also been studies that have found no significant effect of PM_{10} on

cardiovascular health. For example Carder et al. (2008) studied the effects of black smoke on health in Scotland between 1981 and 2001 and found no significant effect on cardiovascular mortality, which reinforces the results found in this study. However the significant results found in many other studies suggest that the lack of significant results in this study are due to either a threshold value or unsuitable data. Recently Daniels et al. (2004) found that for concentrations as low as $10\mu g/m^3$ there was still a significant effect of PM_{10} on cardiovascular-respiratory mortality which suggests that the theory of a threshold concentration below which PM_{10} does not affect cardiovascular health is unlikely to be true.

This then points to the data being unsuitable for this type of study. There have been a number of recent studies carried out that address the problem of estimating pollution exposure, using two main methods. The first is by trying to simulate the actual pollution exposure of individuals during their day and estimating their cumulative exposure, such as the study by Zidek et al. (2005). The second method is to have certain individuals carry a pollution monitor with them throughout the day, which measures their exact pollution exposure. However this is very expensive and time consuming and so such studies are not usually carried out, although Dominici et al. (2000b) used personal PM_{10} exposures from five validation data sets to implement a multi-stage Poisson regression model.

Another issue to consider is the problem of the ecological fallacy which occurs because our analyses are carried out at an ecological (group) level when the quantity of interest, namely the effects of pollution on health, are at the individual level. Therefore the estimated effect of air pollution at an ecological level will not be equal to the estimated effect of air pollution at the individual level. Recently, a number of models have been proposed that try to overcome this problem, such as those by Lancaster & Green (2002) and Wakefield & Shaddick (2005).

When carrying out an investigation into the effects of air pollution on health, one

important decision to make is which pollutant to use. This study was interested in the effects of PM_{10} on cardiovascular health, but there are other pollutants in the atmosphere, many of which have been known to have a detrimental effect on human health. Samoli et al. (2007) investigated the effects of a 1 mg/m^3 increase in carbon monoxide concentrations on mortality and found significant associations with cardiovascular mortality (RR: 1.0125, 95% CI: 1.003, 1.0221) while Elliott et al. (2007) studied the relationship between sulphur dioxide and mortality in Great Britain and found that a 10ppb increase resulted in a relative risk of mortality from a cardiovascular illness of 1.008 (95% CI: 1.006, 1.009). Also of interest is the effect of nitrogen oxides on health. For a $10\mu g/m^3$ increase in nitrogen dioxide, Samoli et al. (2006) found a significant increase in cardiovascular mortality (RR: 1.004, 95% CI: 1.0029, 1.0052). Finally Ballester et al. (2006) investigated the effects of several pollutants on cardiovascular admissions to hospital, including ozone. They found a $10\mu g/m^3$ increase resulted in a significant increase in the number of admissions to hospital with cardiovascular illnesses (RR: 1.0069, 95% CI: 1.0034, 1.0103). Given the significant results in these studies, it may be that there is a significant relationship between cardiovascular admissions to hospital in Glasgow, Edinburgh or Aberdeen and some pollutant other than PM_{10}.

5.2 Lag Problem

One area of discussion in time-series studies such as this one, is the issue of lags. There has been some debate over which is the 'best' lag to use. Some authors used single lag models, with lags ranging from zero (Schwartz (2004)) to five (Kelsall et al. (1997)) while authors such as Hertz-Picciotto et al. (2007) take into account the accumulated effect of pollution and use the average pollution concentrations over a number of days. Other authors add pollution concentrations from several days to one model; for example Cakmak et al. (2007) takes lags of up to five days, while Goodman et al. (2004) took lags of up to 40 days.

Another issue is whether or not to add multiple lags in one model (Prescott et al. (1998)) or to run numerous models each with pollution concentrations from a different lag (Tellez-Rojo et al. (2000)). A number of authors such as Roberts (2004a) use both methods. However a multiple lag model has the problem that the pollution concentrations from the lags are likely to be highly correlated, which would result in collinearity, thus reducing the accuracy of the estimates and inflating the confidence intervals. To overcome this problem, Zanobetti et al. (2000) proposed a distributed lag model in which the pollution estimates are constrained to reduce collinearity. To date, no consensus has been reached as to the best approach for solving the problems surrounding lags, therefore this study made use of several approaches. We first used single lag models, which was followed by multiple lag models. However as the multiple lag models had the problem of collinearity, distributed lag models were also used.

As mentioned in Section 5.1, the shape of the relative risks from the analysis of the daily admissions to hospital in Glasgow appeared to follow the shape suggested by Zanobetti et al. (2002) with regard to the theory of mortality displacement. Mortality displacement suggests that it is mainly the frail proportion of the population who are adversely affected by air pollution, and as such their deaths are brought forward by only a few days. If this is the case, the public health impact of air pollution is significantly reduced. However recently there have been studies carried out, such as the study by Roberts & Switzer (2004), which investigated the properties of distributed lag models in the context of mortality displacement. They found that the estimates from distributed lag models are likely to be biased.

5.3 Shape of relationship

Many studies such as this one assume that the effects of air pollution on health increase linearly. For cardiovascular and respiratory health this may be plausible, as to date no threshold concentration below which air pollution does not affect health has been found. However recently Shaddick et al. (2008) suggested that a linear relationship may not be appropriate as there would eventually be an upper bound on the effect that air pollution has on health. Instead they recommended that the pollution-health relationship should satisfy the following requirements: (i) boundedness; (ii) increasing monotonicity; (iii) smoothness; and (iv) no effect with no exposure.

In contrast, studies that are interested in the effects of pollution concentrations that exceed a given concentration follow a different shape. These studies assume that there is no effect of pollution before this concentration yet when pollution reaches this concentration, there is a sudden effect that then stays constant, an example of which is presented in Figure 2.2.

The majority of the analyses in this thesis assume the shape of the effect of pollution to be linear although the final analysis follows the exceedence shape.

5.4 Future Work

There is still a great deal of further research to be carried out in this field, in particular relating to pollution exposure. One issue that is becoming more appealing to researchers is carrying out spatio-temporal studies, such as the study by Shin et al. (2008), which investigate the effect of air pollution over 17 years in 24 cities in Canada. Until recently, studies of this kind have been infeasible as the computational power needed was extremely high. However with the advances in technology, these have now become possible. Another area where further research is needed is finding a pollution concentration that adequately models an

individual's exposure. To date, many studies use an average pollution concentration over the whole area of interest. However this is unlikely to be representative of a person's actual pollution exposure and so new methods of modelling the pollution exposure are needed.

Appendix A

Additional Analyses

A.1 Females

A.1.1 Exploratory Analysis

The first step is to produce a series of exploratory plots. Figures A.1, A.2 and A.3 are plots relating to cardiovascular admissions to hospital among females, the first of which is over time, the second is against mean weekly PM_{10} concentrations and the last against mean weekly maximum temperatures. Admissions to hospital showed a marked decrease between 2000 and 2006 which is similar to the trend observed in males admissions where Figures A.2 and A.3 show no obvious relationship between weekly cardiovascular admissions to hospital and either mean weekly PM_{10} concentrations or mean weekly maximum temperatures.

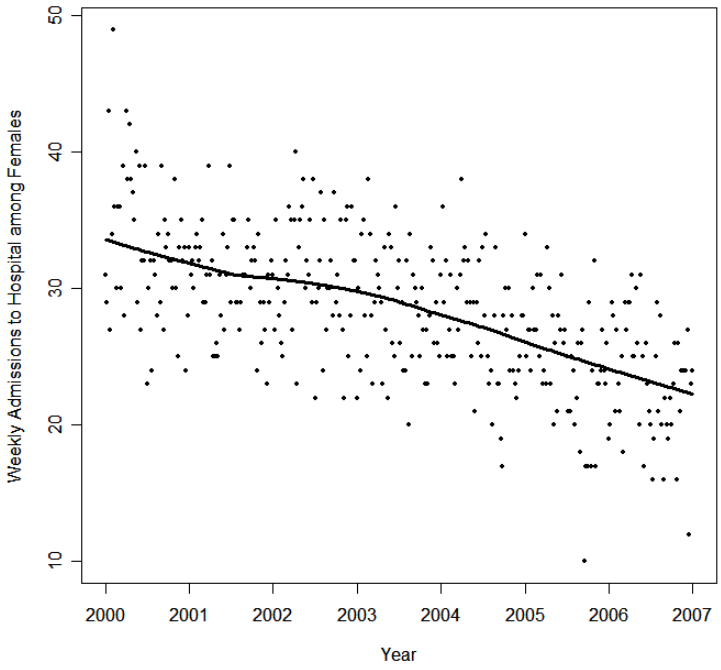

Figure A.1. Admissions to hospital among females

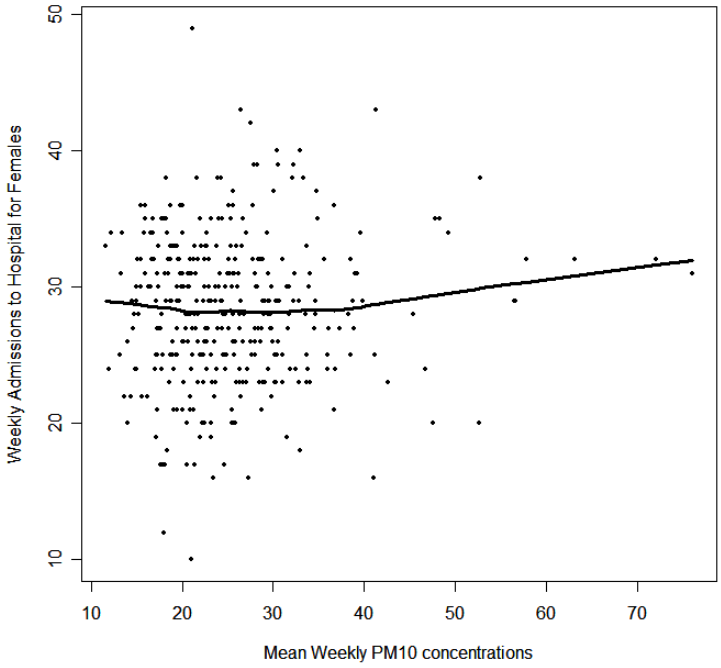

Figure A.2. Relationship between weekly admissions to hospital among females and mean weekly PM_{10} concentrations

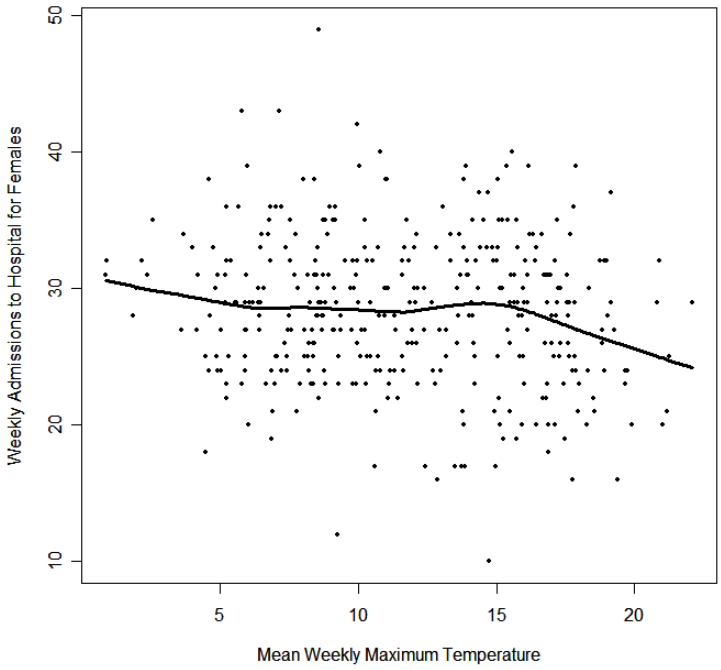

Figure A.3. Relationship between weekly admissions to hospital among females and mean weekly maximum temperatures

A.1.2 Covariate Modelling

We attempt to remove the trend using fixed functions of time such as linear and sinusoidal functions. A summary of the model can be seen in Table A.1.

Coefficient	Estimate	Standard Error	P-Value
Intercept	3.537	0.01889	$<2 \times 10^{-16}$
t	-0.001088	9.422×10^{-05}	$<2 \times 10^{-16}$
$\sin(\omega t)$	0.04498	0.01401	0.00132
$\cos(\omega t)$	-0.001515	0.01389	0.91312

Table A.1. Summary of model

Figures A.4 and A.5 respectively show the residuals for this model and the auto-correlation function for the residuals.

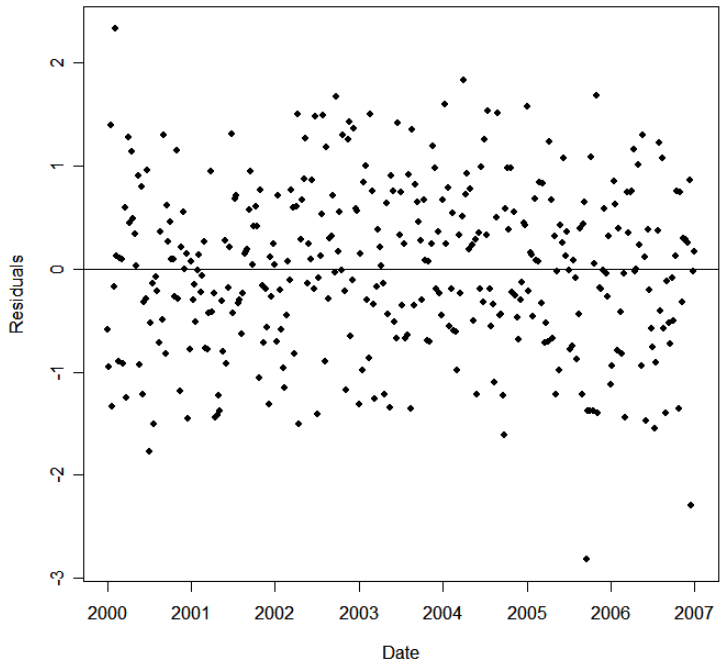

Figure A.4. Residuals from the model of admissions to hospital among females

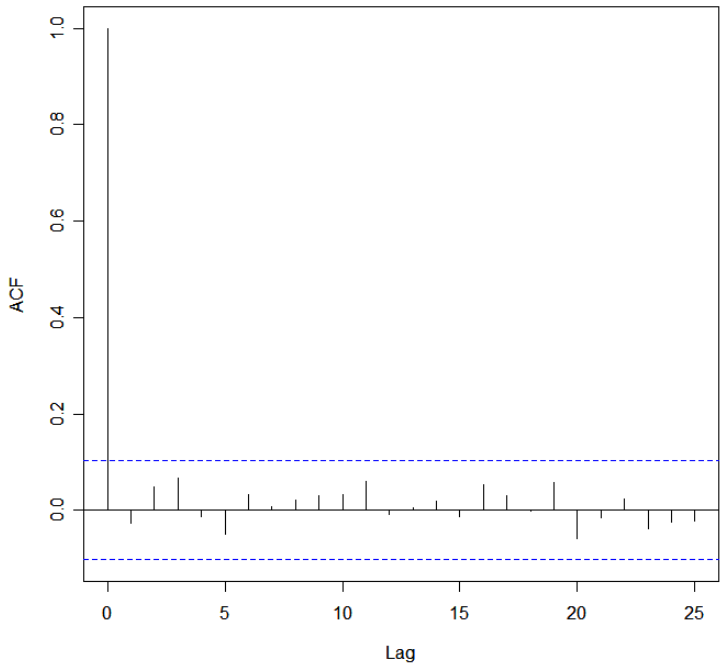

Figure A.5. Autocorrelation function for the residuals

From Figure A.4 we can see that the residuals appear to be evenly scattered about zero with no obvious pattern or trend, while Figure A.5 shows us that each of the lags are within the 95% confidence lags, thus suggesting the residuals now resemble white noise, meaning that the model is adequate and we can now continue the modelling.

A.1.3 Pollution Modelling

The next step is to add mean weekly PM_{10} concentrations to the model using a distributed lag model. Figure A.6 shows the shape of the relative risks from the distributed lag model at powers zero to four. The solid horizontal line in each plot represents the relative risk at zero, which gives the overall estimated effect of exposure to PM_{10}. The relative risks are all very close to one, except at lag eight. However the relative risks are non-significant, suggesting there does not appear to be any significant effect of PM_{10} on admissions to hospital among females.

A.1.4 Conclusions

In conclusion, weekly admissions to hospital have decreased among women from 2000 to 2006 with some signs of seasonality. After using fixed functions of time to remove the trend and seasonal variation, PM_{10} concentrations were added to a distributed lag model. As with the male data, the relative risks from pollution exposure are all very close to one, except at lag eight. However again the relative risks are non-significant, therefore suggesting that this result is an effect of multiple testing. In conclusion, there does not appear to be any significant relationship between admissions to hospital among females and exposure to PM_{10}.

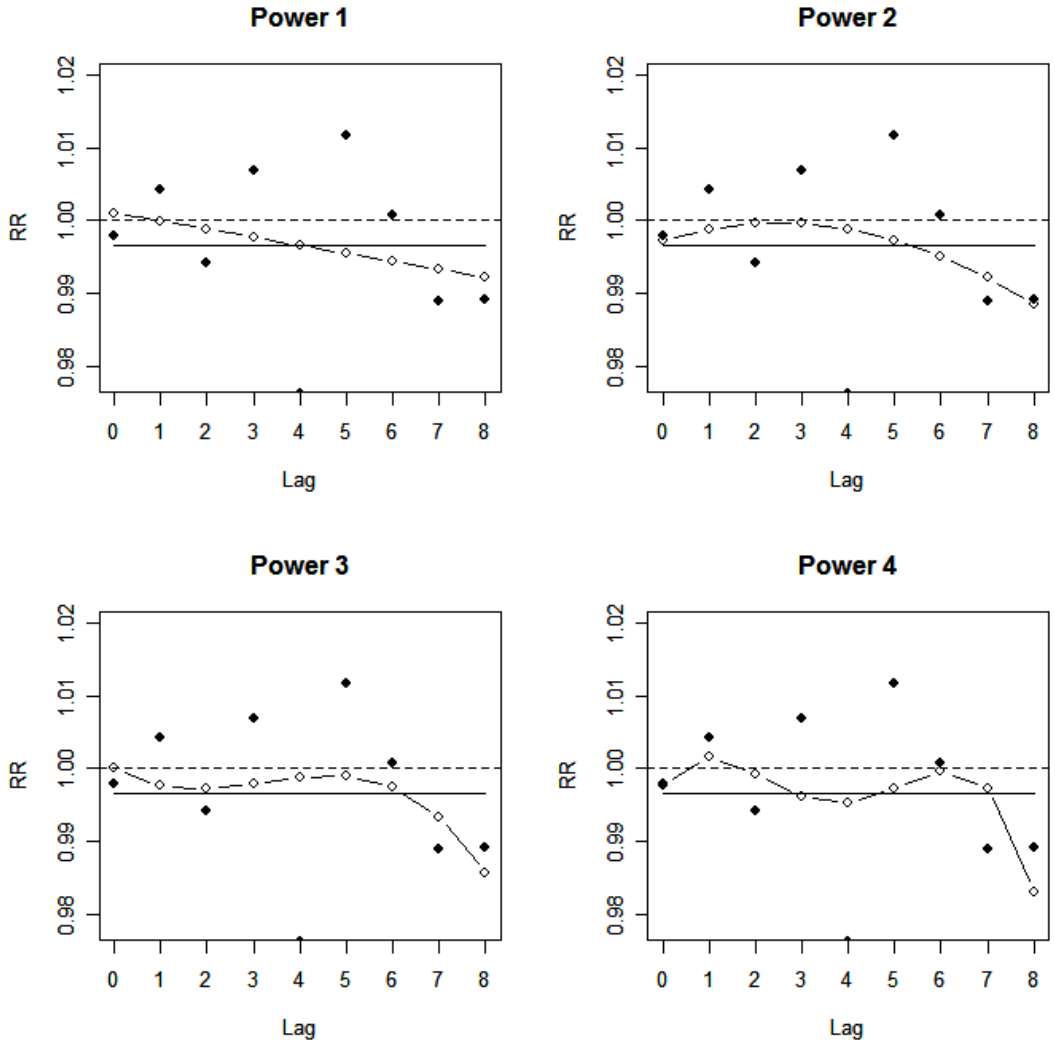

Figure A.6. The solid line is the relative risk at power zero, the dashed line represents 1, the filled diamonds are the relative risks from the unconstrained model and the unfilled diamonds are the relative risks from the distributed lag model

A.2 1st Admissions with CHD

A.2.1 Exploratory Analysis

The first stage in the analysis is producing a series of exploratory plots. Figures A.7, A.8 and A.9 show plots relating to patients who have been admitted to hospital with CHD for the first time. The first is over time, the second against mean weekly PM_{10} concentrations, and the last against mean weekly maximum temperature.

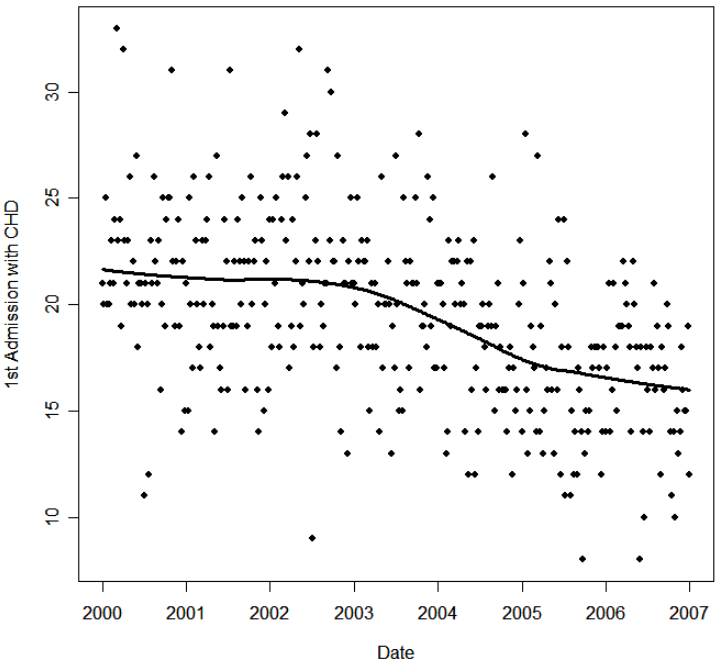

Figure A.7. First admissions to hospital with CHD

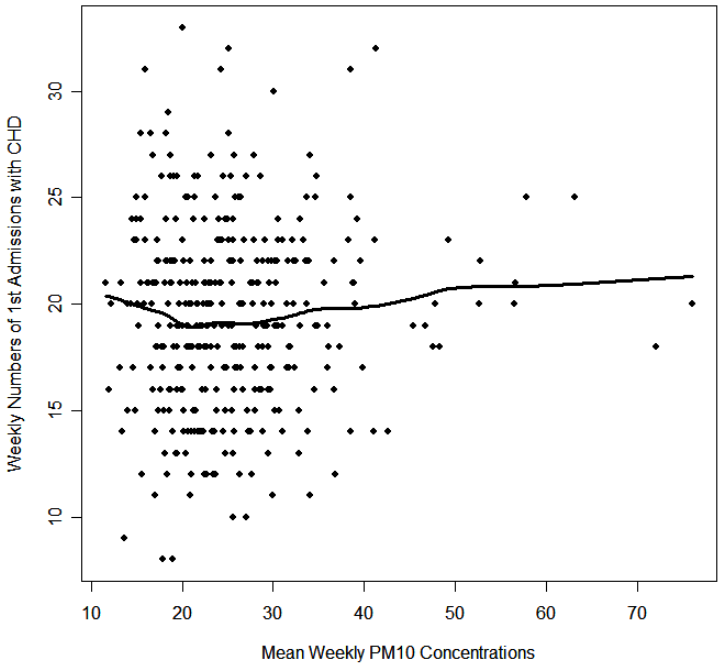

Figure A.8. Relationship between first admissions to hospital with CHD and mean weekly PM_{10} concentrations

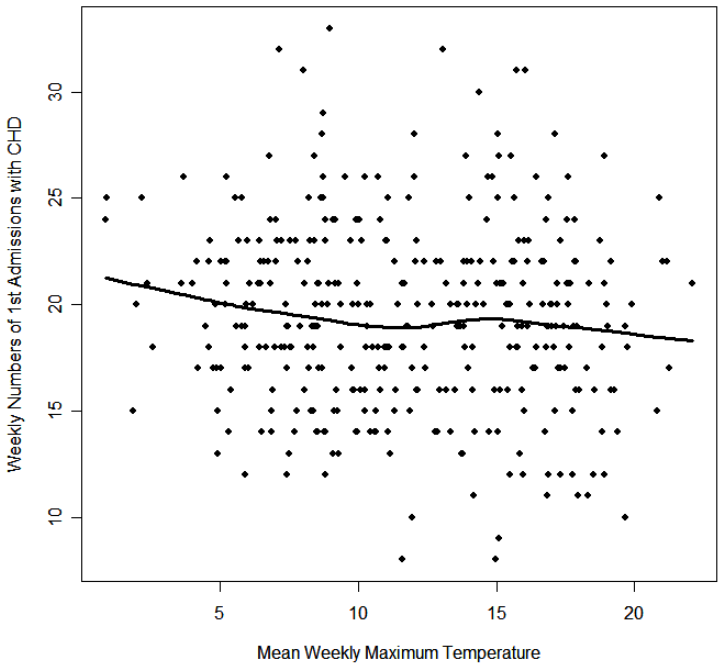

Figure A.9. Relationship between first admissions to hospital with CHD and mean weekly maximum temperatures

From Figure A.7 we can see that the number of people being admitted to hospital with CHD for the first time stayed approximately level between 2000 and 2003 after which the numbers rapidly decreased until 2005. From 2005 to 2006, although the number of admissions were still decreasing, it was less rapidly than before. Figures A.8 and A.9 show no obvious relationship between first admissions to hospital with CHD and either mean weekly PM_{10} concentrations or mean weekly maximum temperature.

A.2.2 Covariate Modelling

The next stage of the analysis is to produce a model that removes most of the trend and variation with in the data. To do this, natural cubic splines with 11 degrees of freedom were used, with 11 degrees of freedom minimising the AIC. Figure A.10 shows the residuals for the model and Figure A.11 shows the autocorrelation function for the residuals.

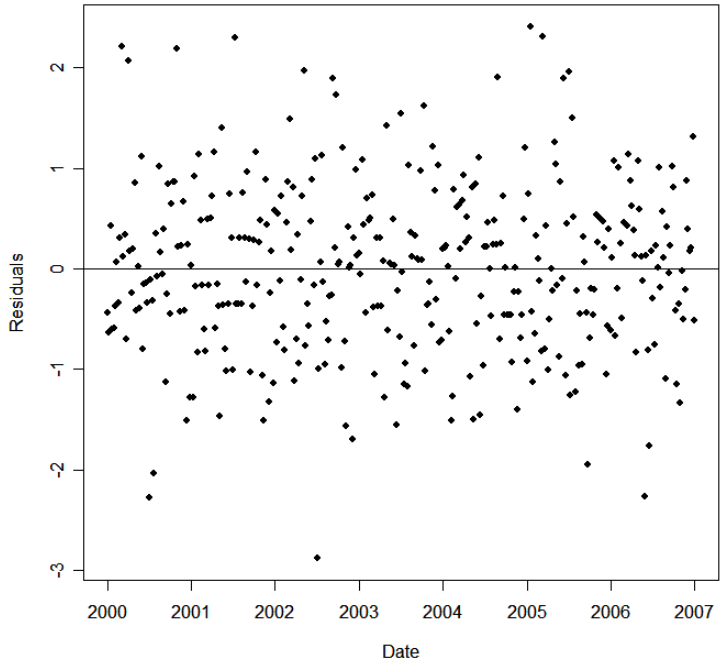

Figure A.10. Residuals from the model of first admissions to hospital with CHD

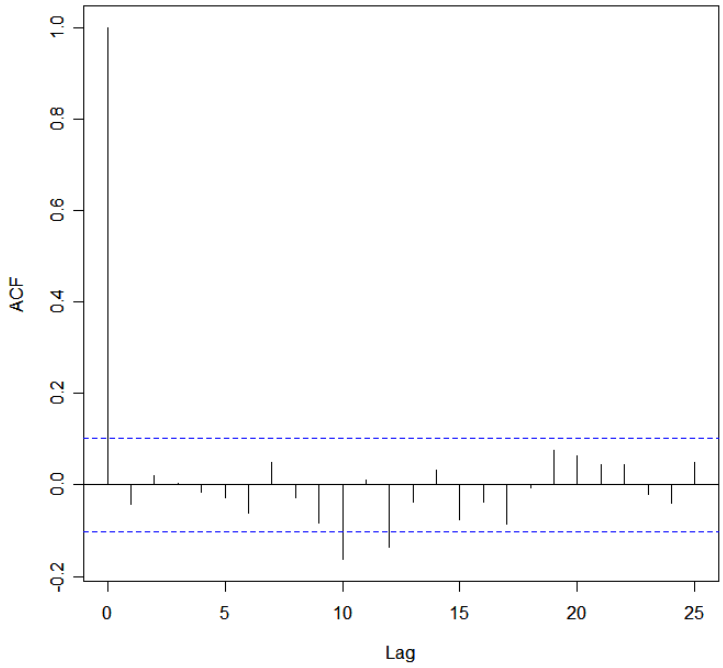

Figure A.11. Autocorrelation function for the residuals

From Figure A.10 we can see that the residuals appear to be evenly scattered about zero and have no obvious trend or pattern while from Figure A.11 we see that most of the lags are within the 95% confidence bands, thus suggesting that the residuals now resemble white noise.

A.2.3 Pollution Modelling

The next step is to add PM_{10} to the model using a distributed lag model. Figure A.12 shows the shape of the relative risks from the distributed lag model, for powers zero to four. Power zero can be seen as the solid horizontal line in each plot and gives the overall estimated effect of exposure to PM_{10}. The relative risks are all fairly close to one, although the relative risk at lag eight is lower than the others. However the relative risks are non-significant and thus suggest there is no significant relationship between first admissions to hospital with CHD and

exposure to PM_{10}.

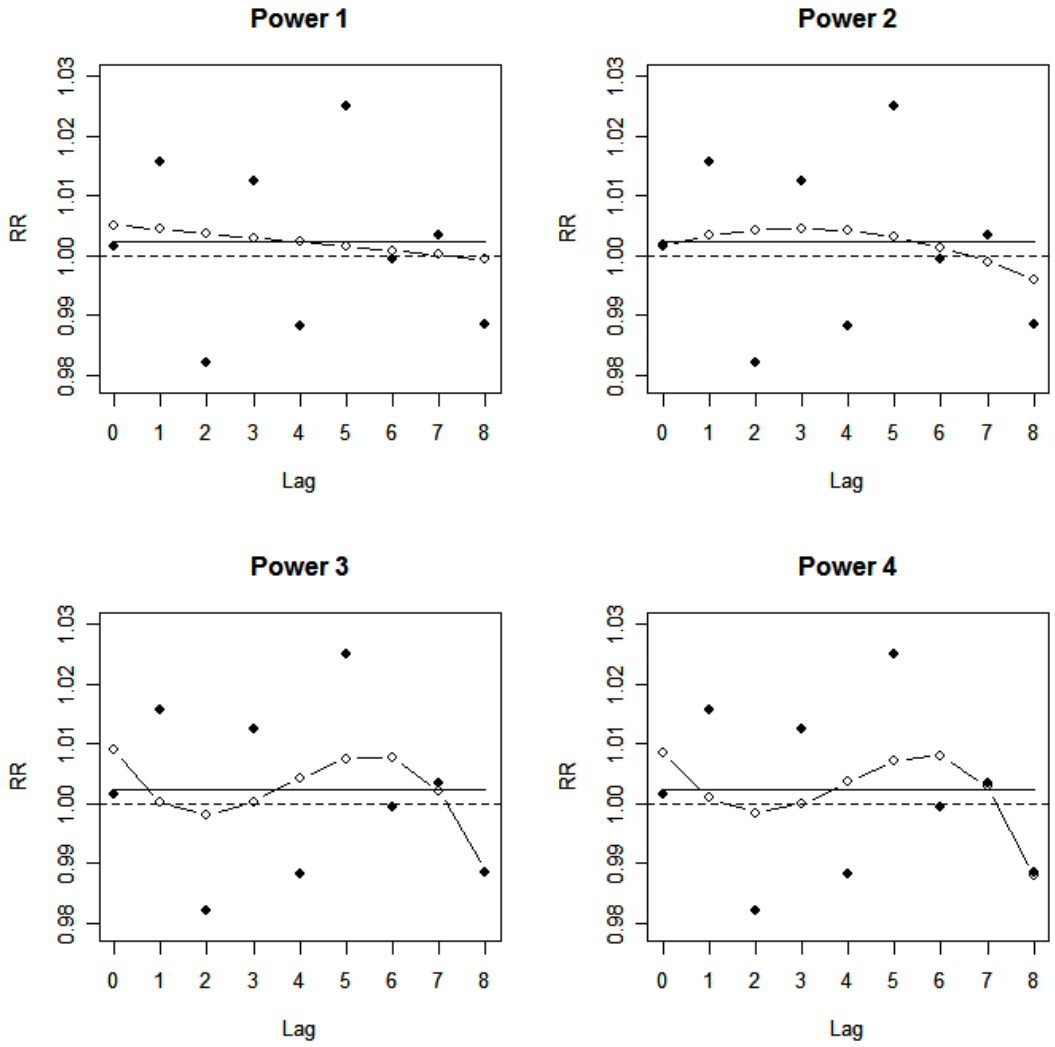

Figure A.12. The solid line is the relative risk at power zero, the dashed line represents 1, the filled diamonds are the relative risks from the unconstrained model and the unfilled diamonds are the relative risks from the distributed lag model.

A.2.4 Conclusions

In conclusion, admissions to hospital stayed steady between 2000 and 2003, before decreasing sharply until 2005. After 2005, although the number of admissions

were still decreasing, it was not as rapidly as before. Natural cubic splines were used to remove the trend and seasonality before adding PM_{10} to the model, using a distributed lag model. The relative risks from the model were all fairly close to one but non-significant, suggesting there there is no statistically significant relationship between first admissions to hospital with CHD and exposure to PM_{10}.

A.3 CHD Under 75s

A.3.1 Exploratory Analysis

The first stage of this analysis is to produce exploratory plots of admissions to hospital with CHD among the under 75s. Figure A.13 is the plot of weekly admissions to hospital over time while Figures A.14 and A.15 show the relationships between admissions to hospital with CHD among the under 75s and mean weekly PM_{10} concentrations and mean weekly maximum temperature respectively.

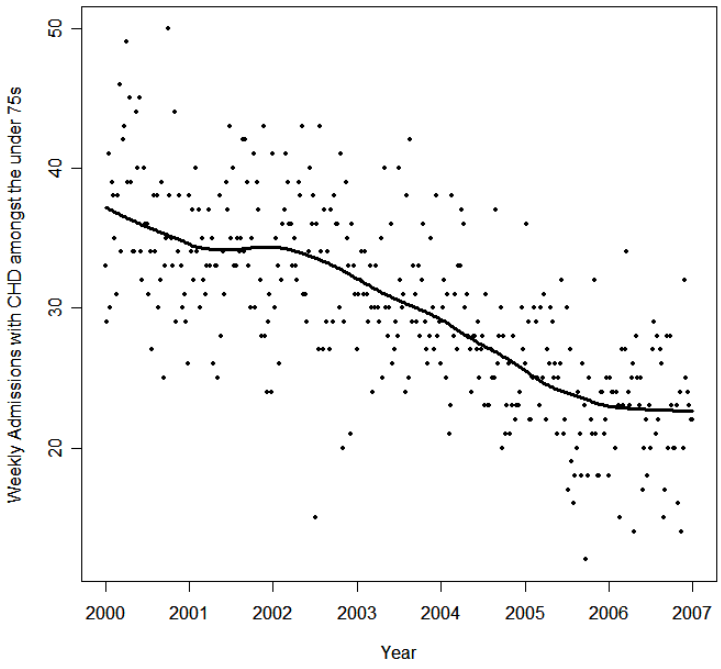

Figure A.13. Weekly admissions to hospital with CHD among the under 75s

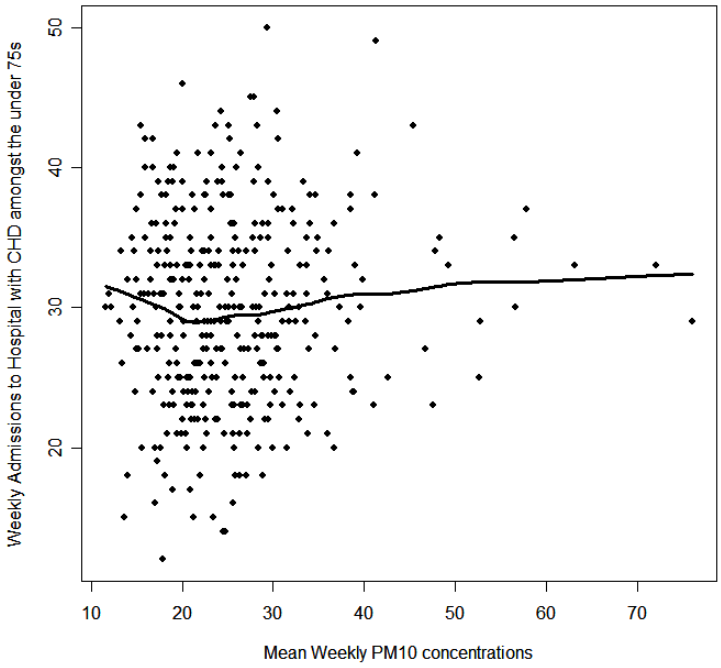

Figure A.14. Relationship between weekly admissions to hospital with CHD among the under 75s and mean weekly PM_{10} concentrations

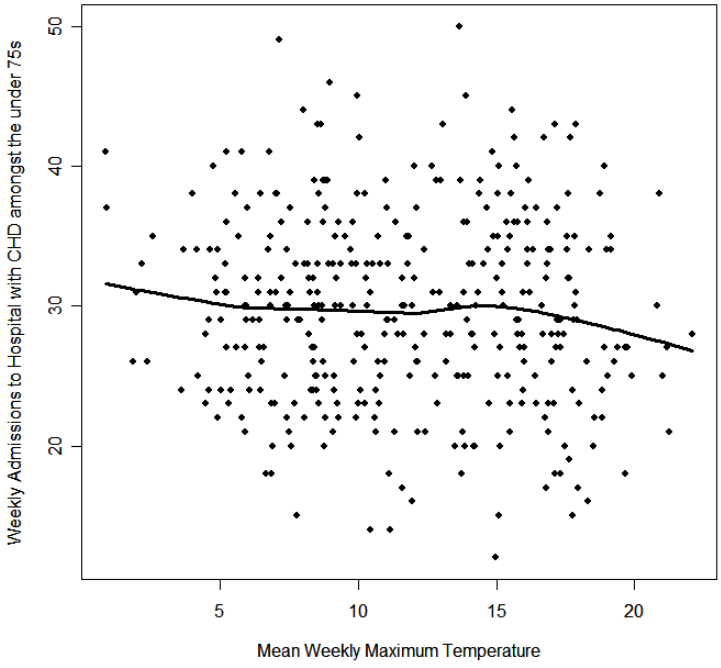

Figure A.15. Relationship between weekly admissions to hospital with CHD among the under 75s and mean weekly maximum temperature

The plot of admissions to hospital with CHD among the under 75s displays a similar trend to the plot of admissions to hospital among all patients. Admissions decreased between 2000 and 2006, with a period in 2001 where the numbers stayed fairly constant. There was little evidence of seasonal variation and there did not appear to be any obvious relationship between admissions to hospital with CHD among the under 75s and either mean weekly PM_{10} concentrations or mean weekly maximum temperature.

A.3.2 Covariate Modelling

The next step is to produce a model that adequately removes the trend and variation within the data. As the time trend was not linear, a linear function of time could not be used. Instead, natural cubic splines with 11 degrees of freedom were used to remove the trend and seasonality. Figures A.16 and A.17 show the residuals for this model and the autocorrelation function for the residuals. The residuals appear to be evenly scattered around zero with a constant variance and there does not appear to be any obvious pattern. In the autocorrelation function, most of the lags are within the 95% confidence bands, suggesting that most of the correlation within the residuals has been removed and therefore we can continue with the analysis.

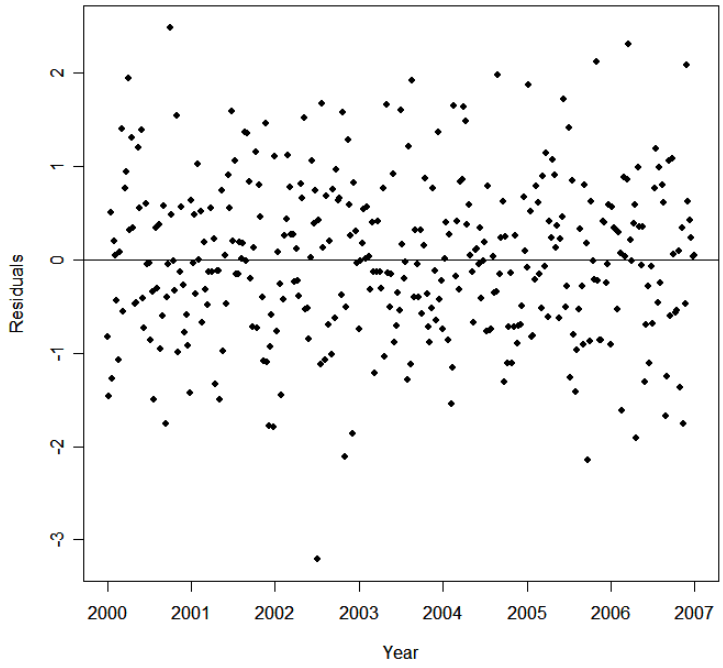

Figure A.16. Residuals from the model of admissions to hospital among the under 75s

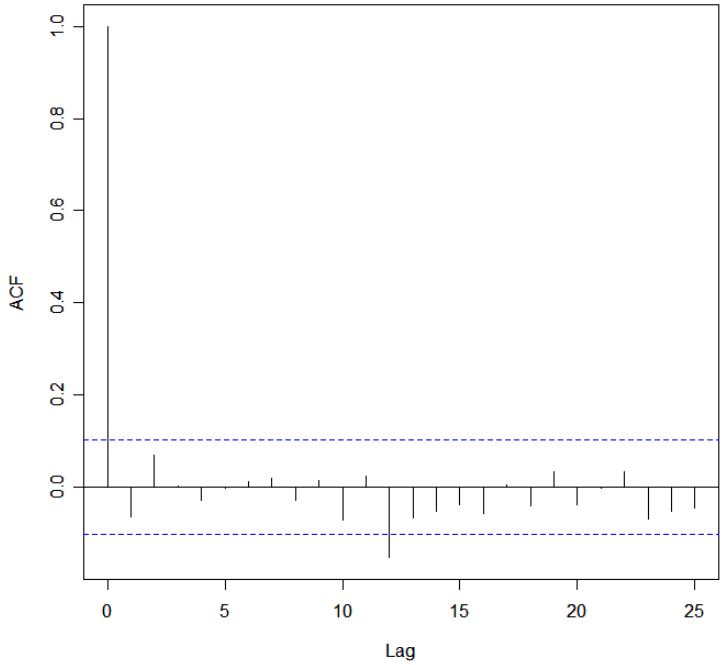

Figure A.17. Autocorrelation function for the residuals

A.3.3 Pollution Modelling

The next step in the analysis is to add PM_{10} to the model using a distributed lag model. Figure A.18 shows the shape of the relative risks from the distributed lag model at powers zero to four. Power zero is represented as the solid horizontal line in each plot and gives the overall estimated effect of PM_{10}.

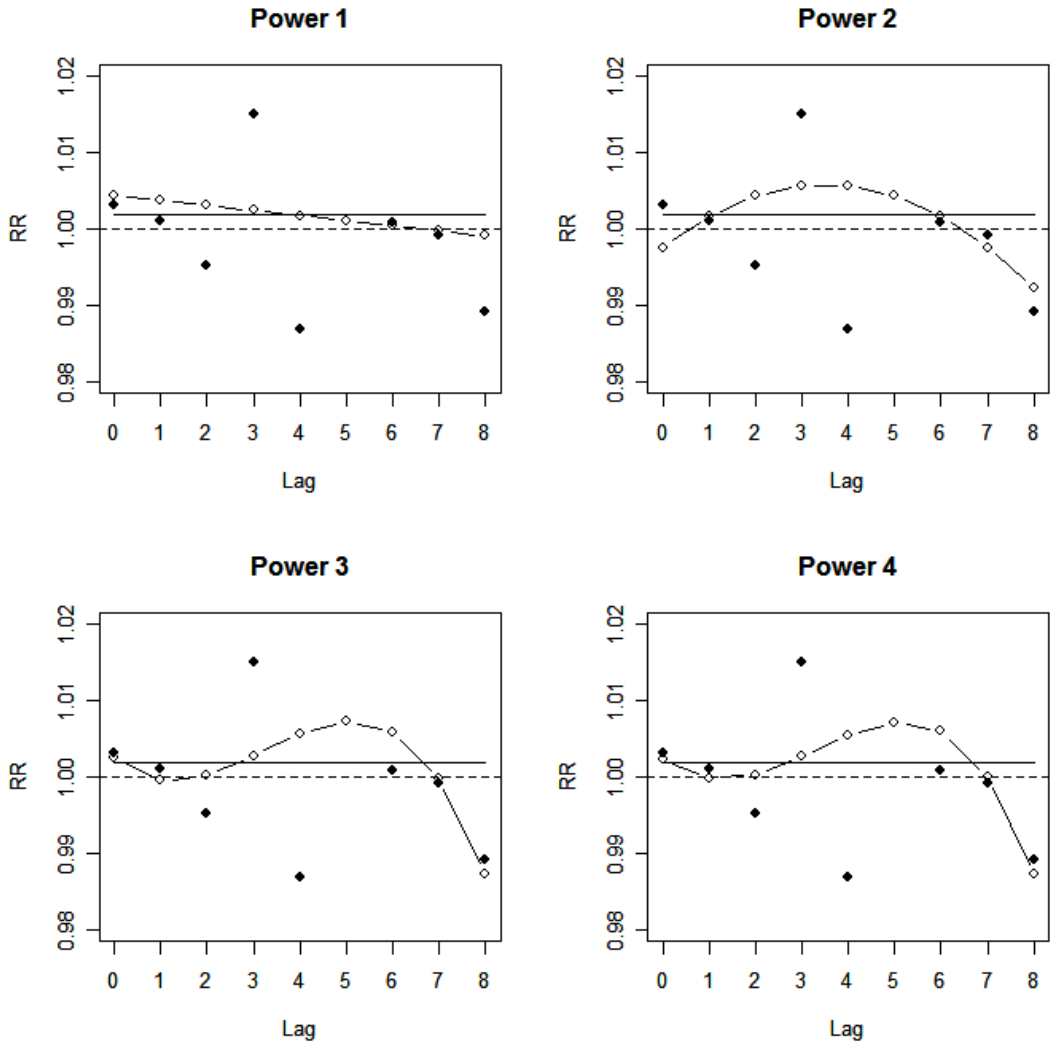

Figure A.18. The solid line is the relative risk at power zero, the dashed line represents 1, the filled diamonds are the relative risks from the unconstrained model and the unfilled diamonds are the relative risks from the distributed lag model

We see a very similar pattern to that of admissions to hospital with CHD for all ages. At each power, the points for the distributed lag model are close to one apart from at lag eight for powers three and four, where the relative risks are approximately 0.985, which is still quite close to one. Also, none of the relative risk coefficients are significant, which suggests there is no significant relationship between admissions to hosptial among patients under 75 years of age and exposure to PM_{10} concentrations.

A.3.4 Conclusions

In conclusion, the analysis for admissions to hospital with CHD among the under 75s was very similar to that for all admissions to hospital with CHD. The admissions decreased from 2000 to 2006 with a period in 2001 where they stayed steady. Natural cubic splines were used to model the data and then PM_{10} was added using a distributed lag model. The relative risks from this model were all quite close to one, except at lag eight for powers three and four. However the relative risks were non-significant, meaning there does not appear to be any statistically significant relationship between admissions to hospital with CHD among the under 75s and exposure to PM_{10}.

A.4 Stroke Under 75s

A.4.1 Exploratory Analysis

First a set of exploratory plots relating to admissions to hospital among the under 75s will be produced. Figures A.19 shows the weekly admissions to hospital among stroke patients under the age of 75 over time, A.20 displays the relationship between weekly admissions to hospital and mean weekly PM_{10} concentrations and A.21 shows the relationship between weekly admissions to hospital and mean weekly maximum temperatures.

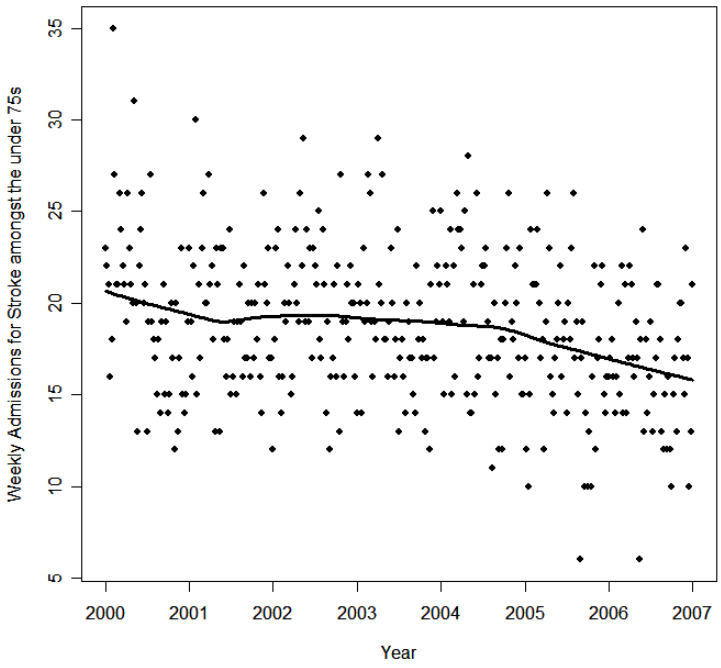

Figure A.19. Weekly admissions to hospital with stroke among the under 75s

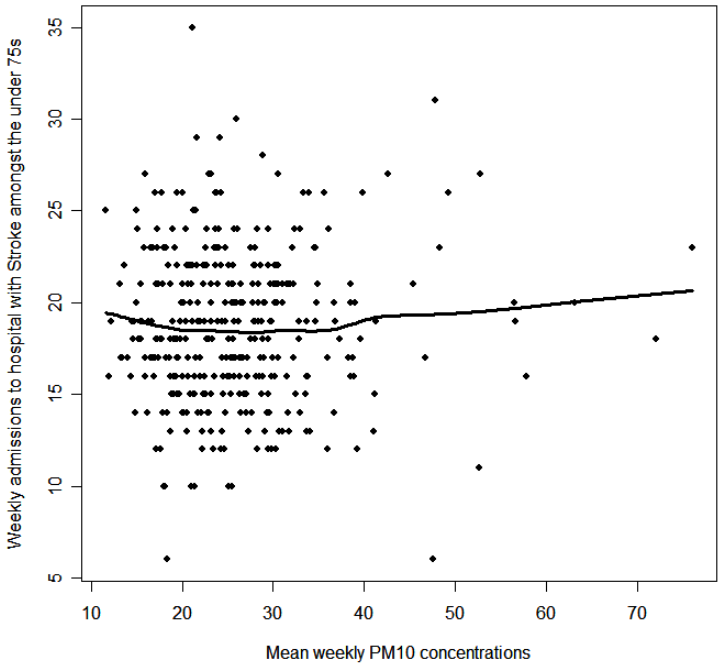

Figure A.20. Relationship between weekly admissions to hospital among the under 75s and mean weekly PM_{10} concentrations

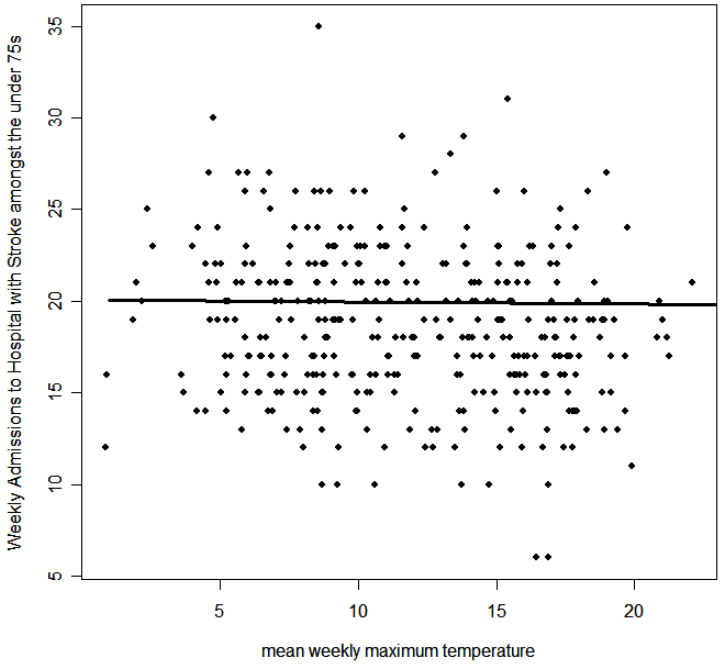

Figure A.21. Relationship between weekly admissions to hospital among the under 75s and mean weekly maximum temperature

From Figure A.19 we can see that stroke admissions among the under 75s decreased slightly between 2000 and 2002 before levellilng out. From 2002 until approximately 2005 the numbers stayed steady before decreasing again in 2005. One point of interest is that the admissions are highly seasonal with more in winter than in summer. However this does not appear to be due to temperature as we can see from Figure A.21 that there does not appear to be any obvious relationship between weekly admissions to hospital with stroke among the under 75s and mean weekly maximum temperature. There also does not appear to be a relationship between stroke admissions and mean weekly PM_{10} concentrations.

A.4.2 Covariate Modelling

The next stage is to produce a model that removes as much of the variation in the data as possible. Natural cubic splines with 17 degrees of freedom were used, with the number of degrees of freedom chosen to minimised the AIC. Figure A.22 is a plot of the residuals for this model. We can see that the residuals appear to be evenly scattered about zero with no apparent trend or pattern. Figure A.23 shows the autocorrelation function for the residuals, in which we can see that most of the lags are within the 95% confidence intervals, implying that most of the correlation in the residuals has been removed and the model is now adequate.

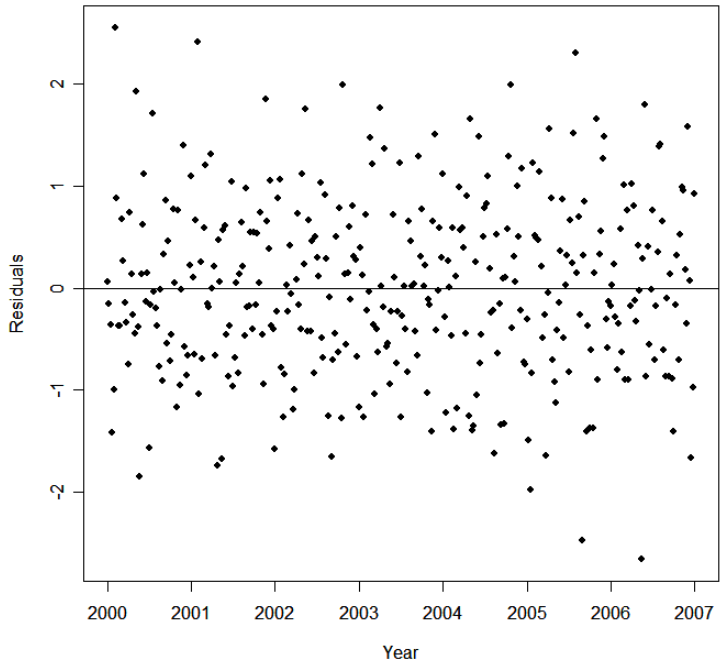

Figure A.22. Residuals from the model of admissions to hospital with Stroke among the under 75s

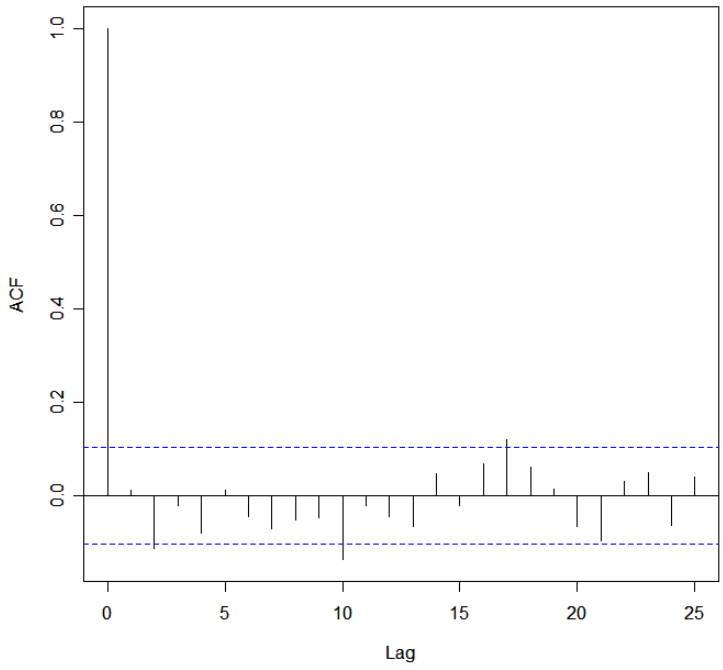

Figure A.23. Autocorrelation function for the residuals

A.4.3 Pollution Modelling

The next step is to add PM_{10} to the model using a distributed lag model. Figure A.24 shows the shape of the relative risks from the distributed lag model at powers zero to four, where power zero is seen as the solid horizontal line in each plot and gives the overall estimated effect of PM_{10}.

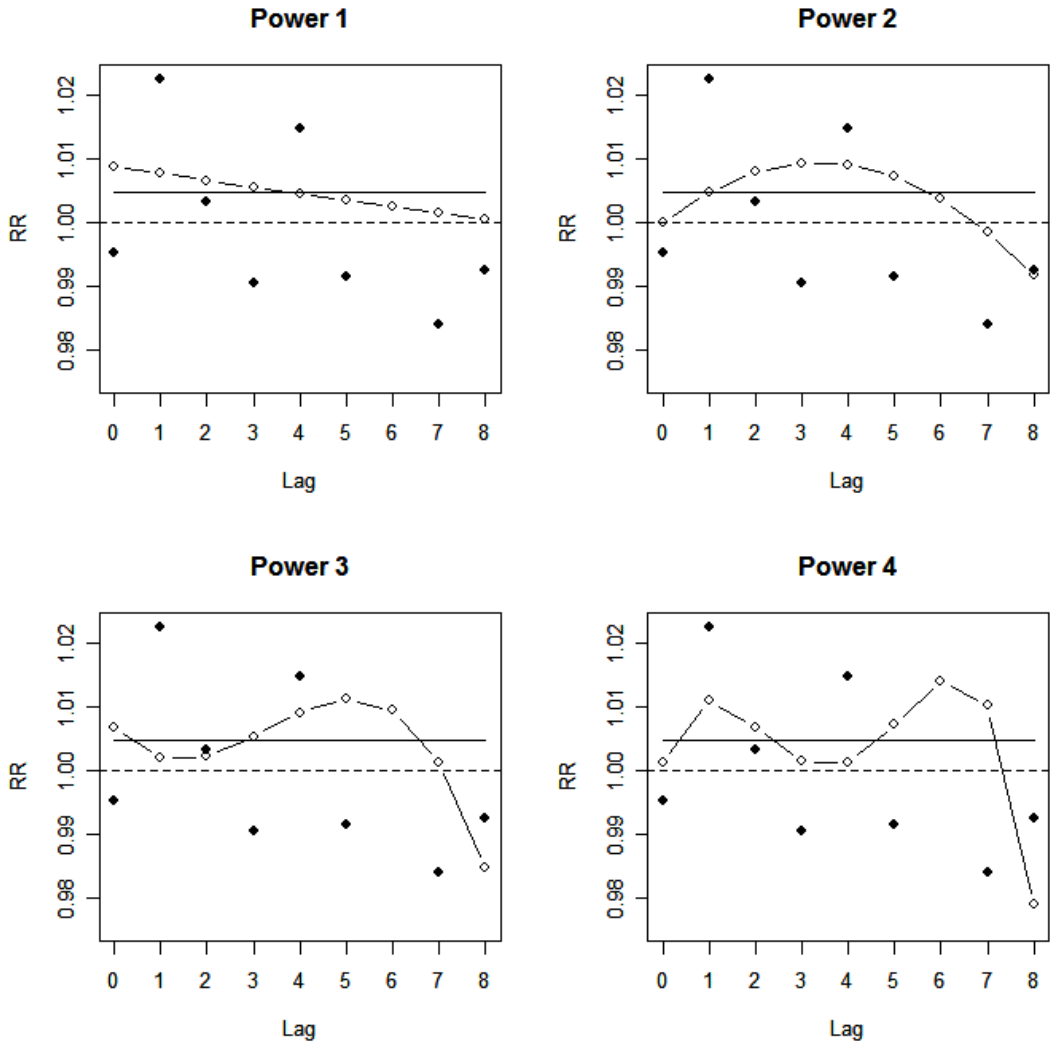

Figure A.24. The solid line is the relative risk at power zero, the dashed line represents 1, the filled diamonds are the relative risks from the unconstrained model and the unfilled diamonds are the relative risks from the distributed lag model

The relative risks are, in most cases, fairly close to one although at powers three and four, the relative risk at lag eight is is especially low. However the relative risk coefficients are non-significant, meaning there does not appear to be any significant relationship between admissions to hospital among patients aged under 75 years of age and exposure to PM_{10}.

A.4.4 Conclusions

Admissions to hospital with stroke among the under 75s has decreased slightly over the years. The main point of interest is the admissions are highly seasonal with more admissions in winter than in summer, however this did not appear to be due to temperature as there did not seem to be any obvious relationship between the two variables. Natural cubic splines were used to remove the trend and seasonality before adding PM_{10} concentrations to a distributed lag model. Although some of the relative risks from this model were quite high, none of them are statistically significant thus implying there does not appear to be any significant relationship between exposure to PM_{10} and admissions to hospital with Stroke among the under 75s.

Bibliography

Anderson, H. R., Spix, C., Medina, S., Schouten, J. P., Castellsague, J., Rossi, G., Zmirou, D., Touloumi, G., Wojtyniak, B., Ponka, A., Bacharova, L., Schwartz, J. & Katsouyanni, K. (1997), 'Air pollution and daily admissions for chronic obstructive pulmonary disease in 6 European cities: results from the APHEA project', *European Respiratory Journal* **10**, 1064–1071.

Annesi-Maesano, I., Agabiti, N., Pistelli, R., Couilliot, M.-F. & Forastiere, F. (2003), 'Subpopulations at increased risk of adverse health outcomes from air pollution', *European Respiratory Journal* **21**(Supplement 40), 57s–63s.

Artinano, B., Querol, X., Salvador, P., Rodriguez, S., Alonso, D. G. & Alastuey, A. (2001), 'Assessment of airborne particulate levels in Spain in relation to the new EU-directive', *Atmospheric Environment* **35**(Supplement No. 1), S43–S53.

Atkinson, R. W., Anderson, H. R., Sunyer, J., Ayres, J., Baccini, M., Vonk, J. M., Boumghar, A., Forastiere, F., Forsberg, B., Touloumi, G., Schwartz, J. & Katsouyanni, K. (2001), 'Acute Effects of Particulate Air Pollution on Respiratory Admissions', *American Journal of Respiratory and Critical Care Medicine* **164**, 1860–1866.

Ballester, F., Rodriguez, P., C., I., Saez, M., Daponte, A., Galan, I., Taracide, M., Arribas, F., Bellido, J., Cirarda, F. B., Canada, A., Guillen, J. J., Guillen-Grima, F., Lopez, E., Perez-Hoyos, S., Lertxundi, A. & Toro, S. (2006), 'Air Pollution and Cardiovascular Admissions Association in Spain: Results

Within the EMECAS Project', *Journal of Epidemiology and Community Health* **60**, 328–336.

Bell, M. L. & Davis, D. D. (2001), 'Reassessment of the Lethal London Fog of 1952: Novel Indicators of Acute and Chronic Consequences of Acute Exposure to Air Pollution', *Environmental Health Perspectives* **109**(Supplement 3), 389–394.

Bell, M. L., McDermott, A., Zeger, S. L., Samet, J. M. & Dominici, F. (2004), 'Ozone and Short-Term Mortality in 95 US Urban Communities, 1987-2000', *Journal of the American Medical Association* **292**(19), 2372–2378.

Bogdanovic, D. C., Nikic, D. S., Milosevic, Z. G. & Stankovic, A. M. (2006), 'Black Smoke Air Pollution and Daily Non-Accidental Mortality in Nis, Serbia', *Central European Journal of Medicine* **1**(3), 292–297.

Braga, A., M.-L. P. G. S. P. P. L. (2006), 'Gender as Effect Modifier of the Association Between Air Pollution and Cardio-Respiratory Diseases', *Epidemiology* **17**(6), S249.

Brockwell, P. J. & Davis, R. A. (1998), *Introduction to Time Series and Forecasting*, 1st edn, Springer.

Brumback, B. A., Ryan, L. M., Schwartz, J. D., Neas, L. M. & Stark, P. C. (2000), 'Transitional Regression Models, With Application to Environmental Time Series', *Journal of American Statistical Association* **95**(449), 16–27.

Brunekreef, B. & Holgate, S. T. (2002), 'Air Pollution and Health', *The Lancet* **360**, 1233–1242.

Cakmak, S., Dales, R. E. & Vidal, C. B. (2007), 'Air Pollution and Mortality in Chile: Susceptibility among the Elderly', *Environmental Health Perspectives* **115**(4), 524–527.

Carder, M., McNamee, R., Beverland, I., Elton, R., Cohen, G. R., Boyd, J. & Agius, R. M. (2005), 'The Lagged Effect of Cold Temperature and Wind Chill

on Cardiorespiratory Mortality in Scotland', *Occupational and Environmental Medicine* **62**, 702–710.

Carder, M., McNamee, R., Beverland, I., Elton, R., Van Tongeren, M., Cohen, G. R., Boyd, J., W., M. & Agius, R. M. (2008), 'Interacting Effects of Particulate Pollution and Cold Temperature on Cardiorespiratory Mortality in Scotland', *Occupational and Environmental Medicine* **65**, 197–204.

Chardon, B., Lefranc, A., Granados, D. & Gremy, I. (2007), 'Air Pollution and Doctors' House Calls for Respiratory Diseases in the Greater Paris Area (2000-3)', *Occupational and Environmental Medicine* **64**(5), 320–324.

Chen, L. H., Knutset, S. F., Shavlik, D., Beeson, W. L., Petersen, F., Ghamsary, M. & Abbey, D. (2005), 'They Association between Fatal Coronary Heart Disease and Ambient Particulate Air Pollution: Are Females at Greater Risk?', *Environmental Health Perspectives* **113**(12), 1723–1729.

Chiogna, M. & Gaetan, C. (2002), 'Dynamic Generalised Linear Models with Applications to Environemental Epidemiology', *Applied Statistics* **51**(4), 453–468.

COMEAP (2006), Cardiovascular Disease and Air Pollution. A Report by the Committee on the Medical Effects of Air Pollution, Technical report, Department of Health.

Curriero, F. C., Heiner, K. S., Samet, J. M., Zeger, S. L., Strug, L. & Patz, J. A. (2002), 'Temperature and Mortality in 11 Cities of the Eastern United States', *American Journal of Epidemiology* **155**(1), 80–87.

Daniels, M. J., Dominici, F., Samet, J. M. & Zeger, S. L. (2000), 'Estimating Particulate Matter-Mortality Dose-Response Curves and Threshold Levels: An Analysis of Daily Time-Series for the 20 Largest US Cities', *American Journal of Epidemiology* **152**(5), 397–406.

Daniels, M. J., Dominici, F., Zeger, S. L. & Samet, J. M. (2004), The National Morbidity, Mortality, and Air Pollution Study Part III: PM_{10} Concentration-Response Curves and Thresholds for the 20 Largest US Cities, Research Report 94 (Part III), Health Effects Institute.

Defra (2007), The Air Quality Strategy for England, Scotland, Wales and Northern Ireland, Technical report, Department for Environment, Food and Rural Affairs.

Delgado-Rodriguez, M. & Llorca, J. (2004), 'Bias', *Journal of Epidemiological and Community Health* **58**, 635–641.

Dobson, A. J. (1991), *An Introduction to Generalized Linear Models*, Chapman and Hall.

Dockery, D. W. & Pope, C. A. (1994), 'Acute Respiratory Effects of Particulate Air Pollution', *Annual Reviews of Public Health* **15**, 107–132.

Dominici, F., Mc Dermott, A., Zeger, S. L. & Samet, J. (2003), 'Airborne Particulate Matter and Mortality: Timescale Effects in Four US Cities', *American Journal of Epidemiology* **157**, 1055–1065.

Dominici, F., Samet, J. M. & Zeger, S. L. (2000), 'Combining Evidence on Air Pollution and Daily Mortality from the the Twenty Largest US Cities: A Hierarchical Modeling Strategy', *Journal of the Royal Statistical Society Series A* **163**(3), 263–302.

Dominici, F., Zeger, S. L. & Samet, J. M. (2000b), 'A Measurement Error Model for Time-Series Studies of Air Pollution and Mortality', *Biostatistics* **1**(2), 157–175.

Elliott, P., Shaddick, G., Wakefield, J. C., De Hoogh, C. & Briggs, D. J. (2007), 'Long-Term Associations of Outdoor Air Pollution with Mortality in Great Britain', *Thorax* **62**, 1088–1094.

Evans, C., Chalmers, J., Capewell, S., Redpath, A., Finlayson, A., Boyd, J., Pell, J., McMurray, J., KacIntyre, K. & Graham, L. (2000), '"I Don't Like Mondays" - Day of the Week of Coronary Heart Disease Deaths in Scotland: Study of Routinely Collected Data', *British Medical Journal* **320**, 218–219.

Forastiere, F., staffoggia, M., Tasco, C., Picciotto, S., Agabiti, N., Cesaroni, G. & Perucci, C. A. (2007), 'Socioeconimic Status, Particulate Air Pollution, and Daily Mortality: Differential Exposure or Differential Susceptibility', *American Journal of Industrial Medicine* **50**, 208–216.

Goldberg, M. S., Burnett, R. T., Bailar, J. C., Brook, J., Bonvalot, Y., Tamblyn, R., Singh, R. & Valois, M.-F. (2001b), 'The Association between Daily Mortality and Ambient Air Particle Pollution in Montreal, Quebec', *Environmental Research* **186**(1), 12–25.

Goldberg, M. S., Burnett, R. T., Bailar, J. C., Tamblyn, R., Ernst, P., Flegel, K., Brook, J., Bonvalot, Y., Singh, R., Valois, M.-F. & Vincent, R. (2001a), 'Identification of Persons with Cardiorespiratory Conditions who are at Risk of Dying from the Acute Effects of Ambient Air Particles', *Environmental Health Perspectives* **109**, 487–494.

Goldberg, M. S., Burnett, R. T., Valois, M.-F., Felgel, K., Bailar, J. C., Brook, J., Vincent, R. & Radon, K. (2003), 'Associations between ambient air pollution and daily mortality among persons with congestive heart failure', *Environmental Research* **91**, 8–20.

Goodman, P. G., Dockery, D. W. & Clancy, L. (2004), 'Cause-Specific Mortality and the Extended Effects of Particulate Pollution and Temperature Exposure', *Environmental Health Perspectives* **112**(2), 179–185.

Gwynn, R. C., Burnett, R. T. & Thurston, G. D. (2000), 'A Time-Series Analysis of Acidic Particulate Matter and Daily Mortality and Morbidity in the Buffalo, New York, Region', *Environmental Health Perspectives* **108**(2), 125–133.

Hertz-Picciotto, I., Baker, R. J., Yap, P.-S., Dostal, M., Joad, J. P., Lipsett, M., Greenfield, T., Herr, C. E. W., Benes, I., Shumway, R. H., Pinkerton, K. E. & Sram, R. (2007), 'Early Childhood Lower Respiratory Illness and Air Pollution', *Environmental Health Perspectives* **115**(10), 1510–1518.

Hoek, G., Schwartz, J., Groot, B. & Eilers, P. (1997), 'Effects of Ambient Particulate Matter and Ozone on Daily Mortality in Rotterdam, the Netherlands', *Archives of Environmental Health* **52**(6), 455–463.

Hong, Y., Leem, J., Ha, E. & Christiani, D. C. (1999), 'PM_{10} Exposure, Gaseous Pollutants, and Daily Mortality in Inchon, South Korea', *Environmental Health Perspectives* **107**(11), 873–878.

Huynen, M. M. T. E., Martens, P., Schram, D., Weijenberg, M. P. & Kunst, A. E. (2001), 'The Impact of Heat Waves and Cold Spells on Mortality Rates in the Dutch Population', *Environmental Health Perspectives* **109**, 463–470.

Katsouyanni, K., Zmirou, D., Spix, C., Sunyer, J., Schouten, J. P., Ponka, A., Anderson, H. R., Le Moullec, Y., Wojtyniak, B., Vigotti, M. A. & Bacharova, L. (1995), 'Short-Term Effects of Air Pollution on Health: a European Approach Using Epidemiological Time-Series Data', *European Respiratory Journal* **8**, 1030–1038.

Keatinge, W. E. & Donaldson, G. C. (2001), 'Mortality Related to Cold and Air Pollution in London After Allowance for Effects of Associated Weather Patterns', *Environmental Research* **86**, 209–216.

Kelsall, J. E., Samet, J. M., Zeger, S. L. & Xu, J. (1997), 'Air Pollution and Mortality in Philadelphia, 1974-1988', *American Journal of Epidemiology* **146**(9), 750–762.

Lancaster, G. & Green, M. (2002), 'Deprivation, Ill-Health and the Ecological Fallacy', *Journal of the Royal Statistical Society, Series A* **165**(2), 263–278.

Lee, D. & Shaddick, G. (2008), 'Modelling the Effects of Air Pollution on Health using Bayesian Dynamic Generalised Linear Models', *Environmetrics* .

Lee, S. L., Wong, W. H. S. & Lau, Y. L. (2006), 'Association Between Air Pollution and Asthma Admission among Children in Hong Kong', *Clinical and Experimental Allergy* **36**, 1138–1146.

Lippmann, M. & Ito, K. (1995), 'Separating the Effects of Temperature and Season on Daily Mortality from those of Air Pollution in London: 1965-1972', *Inhalation Toxicology* **7**, 85–97.

Luginaah, I. N., Fung, K. Y., Gorey, K. M., Webster, G. & Wills, C. (2005), 'Association of Ambient Air Pollution with Respiratory Hospitalization in a Government-Designated "Area of Concern": The Case of Windsor, Ontario', *Environmental Health Perspectives* **113**(3), 290–296.

Nafstad, P., Haheim, L. L., Wisloff, T., Gram, F., Oftedal, B., Holme, I., Hjermann, I. & Leren, P. (2004), 'Urban Air Pollution and Mortality in a Cohort of Norwegian Men', *Environmental Health Perspectives* **112**(5), 610–615.

Nelder, J. A. (1990), 'Nearly Parallel Lines in Residual Plots', *The American Statistician* **44**(3), 221–222.

Ostro, B., Broadwin, R., Green, S., Feng, W.-Y. & Lipsett, M. (2006), 'Fine Particulate Air Pollution and Mortality in Nine Californian Counties: Results from CALFINE', *Environmental Health Perspectives* **114**(1), 29–33.

Parliamentary Office of Science and Technology (1996), Fine particles and health, Technical Report 82, UK Parliament.

Pattenden, S., Nikiforov, N. & Armstrong, B. G. (2003), 'Mortality and Temperature in Sofia and London', *Journal of Epidemiology and Community Health* **57**, 628–633.

Pauli, F. & Rizzi, L. (2006), 'Statistical Analysis of Temperature Impact on Daily Hospital Admissions: Analysis of Data from Udine, Italy', *Environmetrics* **17**, 47–64.

Pearce, N. (2000), 'The Ecological Fallacy Strikes Back', *Journal of Epidemiological and Community Health* **54**, 326–327.

Pope, C. A. (2000), 'Epidemiology of Fine Particulate Air Pollution and Human Health: Biologic Mechanisms and Who's at Risk?', *Environmental Health Perspectives* **108**(Suppl 4), 713–723.

Pope, C. A., Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K. & Thurston, G. D. (2002), 'Lung Cancer, Cardiopulmonary Mortality and Long-Term Exposure to Fine Particulate Air Pollution', *The Journal of the American Medical Association* **287**, 1132–1141.

Pope, C. A., Burnett, R. T., Thurston, G. D., Thun, M. J., Calle, E. E., Krewski, D. & Dodleski, J. J. (2004), 'Cardiovascular Mortality and Long-Term Exposure to Particulate Air Pollution', *Circulation* **109**, 71–77.

Prescott, G. J., Cohen, G. R., Elton, R. A., Fowkes, F. G. R. & Agius, R. M. (1998), 'Urban Air Pollution and Cardiopulmonary Ill Health: A 14.5 Year Time Series Study', *Occupational and Environmental Medicine* **55**, 697–704.

R 2.2.0 - A Language and Environment (2005). The R Development Core Team.

Roberts, S. (2004a), 'Interactions Between Particulate Air Pollution and Temperature in Air Pollution Mortality Time Series Studies', *Environmental Research* **96**, 328–337.

Roberts, S. & Switzer, P. (2004), 'Mortality Displacement and Distributed Lag Models', *Inhalation Toxicology* **16**(14), 897–888.

Ruppert, D., Wand, M. P. & Carroll, R. J. (2003), *Semiparametric Regression*, Cambridge University Press.

Samet, J., Zeger, S., Kelsall, J., Xu, J. & Kalkstein, L. (1998), 'Does Weather Confound or Modify the Association of Particulate Air Pollution with Mortality?', *Environmental Research, Section A* **77**, 9–19.

Samoli, E., Aga, E., Touloumi, G., Nisiotis, K., Forsberg, B., Lefranc, A., Pekkanen, J., Wojtyniak, B., Schindler, C., Niciu, E., Brunstein, R., Fikfak, M. D., Schwartz, J. & Katsouyanni, K. (2006), 'Short-Term Effects of Nitrogen Dioxide on Mortality: An Analysis within the APHEA Project', *European Respiratory Journal* **27**(6), 1129–1137.

Samoli, E., Touloumi, G., Schwartz, J., Anderson, H. R., Schindler, C., Forsberg, B., Vigotti, M. A., Vonk, J., Kosnik, M., Skorkovsky, J. & Katsouyanni, K. (2007), 'Short-Term Effects of Carbon Monoxide on Mortality: An Analysis within the APHEA Project', *Environmental Health Perspectives* **115**(11), 1578–1583.

Schwartz, J. (1994), 'Nonparametric Smoothing in the Analysis of Air Pollution and Respiratory Illness', *The Canadian Journal of Statistics* **22**(4), 471–487.

Schwartz, J. (2000a), 'The Distributed Lag between Air Pollution and Daily Deaths', *Epidemiology* **11**, 320–326.

Schwartz, J. (2000b), 'Assessing Confounding, Effect Modification and Thresholds in the Association between Ambient Particles and Daily Deaths', *Environmental Health Perspectives* **108**(6), 563–568.

Schwartz, J. (2004), 'The Effects of Particulate Air Pollution on Daily Deaths: A Multi-City Case Crossover Analysis', *Occupational and Environmental Medicine* **61**, 956–961.

Schwartz, J. & Marcus, A. (1990), 'Mortality and Air Pollution in London: A Time Series Analysis', *American Journal of Epidemiology* **131**(1), 185–194.

Schwartz, J., Spix, C., Touloumi, G., Bacharova, L., Barumamdzadeh, T., le Tertre, A., Piekarksi, T., Ponce de Leon, A., Ponka, A., Rossi, G., Saez,

M. & Schouten, J. P. (1996), 'Methodological issues in studies of air pollution and daily counts of deaths or hospital admissins', *Journal of Epidemiology and Community Health* **50**(Suppl 1), S3–S11.

Searle, S. R. (1988), 'Parallel Lines in Residual Plots', *The American Statistician* **42**(3), 211.

Shaddick, G., Lee, D., Zidek, J. V. & Salway, R. (2008), 'Estimating Exposure Response Functions using Ambient Pollution Concentrations', *Annals of Applied Statistics, to appear* .

Shin, H. H., Stieb, D. M., Jessiman, B., Goldberg, M. S., Brion, O., Brook, J., Ramsay, T. & Burnett, R. T. (2008), 'A Temporal, Multicity Model to Estimate the Effects of Short-Term Exposure to Ambient Air Pollution on Health', *Environmental Health Perspectives* **116**(9), 1147–1153.

Shumway, R. H. & Stoffer, D. S. (2006), *Time Series Analysis and Its Applications*, 2nd edn, Springer.

'Splines' 2.2.0 (2005). The R Development Core Team.

Stankovic, A., Dragana, N., Nikolic, M. & Bogdanovic, D. (2007), 'Short-Term Effects of Air Pollution on Cardiovascular Mortality in Elderly in Nis, Serbia', *Central European Journal of Public Health* **15**(3), 95–98.

Stieb, D. M., Judek, S. & Burnett, R. T. (2002), 'Meta-Analysis of Time-Series Studies of Air Pollution and Mortality: Effects of Gases and Particles and the Influence of Cause of Death, Age, and Season', *Journal of Air and Waste Management Association* **52**, 470–484.

Sunyer, J., Atkinson, R., Ballester, F., Le Tertre, A., Ayres, J. G., Forastiere, F., Forsberg, B., Vonk, J. M., Bisanti, L., Anderson, R. H. & Katsouyanni, K. (2003a), 'Respiratory Effects of Sulphur Dioxide: A Hierarchical Multicity Analysis in the APHEA 2 Study', *Occupational and Environmental Medicine* **60**, e2.

Sunyer, J., Ballester, F., Le Tertre, A., Atkinson, R., Ayres, J. G., Forastiere, F., Forsberg, B., Vonk, J. M., Bisanti, L., Tenias, J. M., Medina, S., Schwartz, J. & Katsouyanni, K. (2003b), 'The Association of Daily Sulfur Dioxide Air Pollution Levels with Hospital Admissions for Cardiovascular Diseases in Europe (The Aphea-II Study)', *European Heart Journal* **24**, 752–760.

Tellez-Rojo, M. M., Romieu, I., Ruiz-Valasco, S., Lezana, M.-A. & Hernandez-Avila, M.-M. (2000), 'Daily Respiratory Mortality and PM_{10} Pollution in Mexico City: Importance of Considering Place of Death', *European Respiratory Journal* **16**, 391–396.

Touloumi, G., Samoli, E., Quenel, P., Paldy, A., Anderson, R. H., Zmirou, D., Galan, I., Forsberg, B., Schindler, C., Schwartz, J. & Katsouyanni, K. (2005), 'Short-Term Effects of Air Pollution on Total and Cardiovascular Mortality: The Confounding Effect of Influenza Epidemics', *Epidemiology* **16**(1), 49–57.

Villeneuve, P. J., Burnett, R. T., Shi, Y., Krewski, D., Goldberg, M. S., Hertzman, C., Chen, Y. & Brook, J. (2003), 'A Time-Series Study of Air Pollution, Socioeconomic Status and Mortality in Vancouver, Canada', *Journal of Exposure Analysis and Environmental Epidemiology* **13**, 427–435.

Wakefield, J. & Shaddick, G. (2005), 'Health-Exposure Modelling and the Ecological Fallacy', *Biostatistics* **1**(1), 1–19.

West, M., Harrison, J. & Migon, H. S. (1985), 'Dynamic Generalized Linear Models and Bayesian Forecasting', *Journal of the American Statistican* **80**(389), 73–83.

Witte, D. R., grobbee, D. E., Bots, M. L. & Hoes, A. W. (2005), 'A Meta-Analysis of Excess Cardiac Mortality on Monday', *European Journal of Epidemiology* **20**, 401–406.

World Health Organisation (2006), 'WHO Challanges World to Improve Air Quality', News Release (website).

Xu, X., Li, B. & Huang, H. (1995), 'Air Pollution and Unscheduled Hospital Outpatient and Emergency Room Visits', *Environmental Health Perspectives* **103**, 286–289.

Ye, F., Piver, W. T., Ando, M. & Portier, C. J. (2001), 'Effects of Temperature and Air Pollutants on Cardiovascular and Respiratory Diseases for Males and Females Older than 65 Years of Age in Tokyo, July and August 1980-1995', *Environmental Health Perspectives* **109**, 355–359.

Yu, O., Sheppard, L., Lumley, T., Koenig, J. Q. & Shapiro, G. G. (2000), 'Effects of Ambient Air Pollution on Symptoms of Asthma in Seattle-Area Children Enrolled in the CAMP Study', *Environmental Health Perspectives* **108**(12), 1209–1214.

Zanobetti, A., Schwartz, J., Samoli, E., Gryparis, A., Touloumi, G., Atkinson, R., Le Tertre, A., Bobros, J., Celko, M., Goren, A., Forsberg, B., Michelozzi, P., Rabczenko, D., Ruiz, E. A. & Katsouyanni, K. (2002), 'The temporal pattern of mortality responses to air pollution: A multicity assessment of mortality displacement', *Epidemiology* **13**(1), 87–93.

Zanobetti, A., Schwartz, J., Samoli, E., Gryparis, A., Touloumi, G., Peacock, J., Anderson, R. H., Le Tertre, A., Bobros, J., Celko, M., Goren, A., Forsberg, B., Michelozzi, P., Rabczenko, D., Hoyos, S. P., Wichmann, H. E. & Katsouyanni, K. (2003), 'The Temporal Pattern of Respiratory and Heart Disease Mortality in Response to Air Pollution', *Environmental Health Perspectives* **111**, 1188–1193S.

Zanobetti, A., Wand, M. P., Schwartz, J. & Ryan, L. M. (2000), 'Generalised Additive Distributed Lag Models: Quantifying Mortality Displacement', *Biostatistics* **1**(3), 279–292.

Zeger, S. L. (1988), 'A Regression Model for Time Series of Counts', *Biometrika* **75**(4), 621–629.

Zeger, S. L. & Qaqish, B. (1988b), 'Markov Regression Models for Time Series: A Quasi-Likelihood Approach', *Biometrics* **44**, 1019–1031.

Zidek, J. V., Shaddick, G., White, R., Meloche, J. & Chatfield, C. (2005), 'Using a Probabilistic Model (pCNEM) to Esimate Personal Exposure to Air Pollution', *Environmetrics* **16**, 481–493.