Time-varying coefficient models for the analysis of air pollution and health outcome data

Duncan Lee and Gavin Shaddick

Abstract

In this paper, a time-varying coefficient model is developed for air pollution and mortality data, which allows for an interaction between the effect of air pollution and time. Such temporal effects may show a seasonal pattern which could be the result of an interaction with temperature, or show a long-term trend which could result from a change in the composition of individual pollutants over a number of years. Our model produces a smooth estimate of the changing effect of air pollution over time that does not follow an *a-priori* fixed parametric form set by the investigator. Instead, it follows an unspecified smooth function, the form of which is estimated from the data using penalised natural cubic splines. Poisson regression models, using both quasi-likelihood and Bayesian techniques, are developed, with estimation performed using an iteratively re-weighted least squares procedure and Markov chain monte carlo (MCMC) simulation respectively.

The models are applied to data from the National Morbidity, Mortality, and Air Pollution Study (NMMAPS). Daily total non-accidental mortality, particulate matter (PM_{10}) and weather data were obtained from four US cities (Cleveland, Detroit, Minneapolis, Pittsburgh) for a period of five years (1990-94). The results from the two approaches produced similar patterns for the effect of pollution over time, although the credible intervals produced using the Bayesian model were wider than their equivalent from the quasi-likelihood model as the latter ignored the inherent uncertainty associated with the choice of smoothing parameter and thus produced intervals which were too narrow.

1 Introduction

The potential health effects of exposure to ambient air pollution have been a major issue in public health for over fifty years. Numerous studies of air pollution and health data have shown a link between a component of air pollution and measures of public health. The majority of such studies have analysed the short-term (acute) effects of air pollution, using time series regression models such as generalised linear models (GLMs) (McCullagh and Nelder (1989)) and more recently, generalised additive models (GAMs) (Hastie and Tibshirani (1990)) to assess the association between daily measurements of pollution and counts of health outcome within a specified study region and period. Such models make a number of, possibly untenable, assumptions about the underlying mechanism that generates the series of health data. Relaxing these assumptions, in an attempt to make the regression models more biologically plausible, has been an active research topic for many years. Early studies (for example Schwartz (1991)) modelled mortality counts as arising from a Poisson distribution, meaning that the variance was equal to the mean. Common approaches to modelling also assumed that the responses were independent over time, that the mean was linearly dependent on the levels of air pollution, and that seasonal effects followed a fixed parametric form. Advances in statistical methods have allowed researchers to relax these assumptions and apply a wider class of regression models to these data.

The relationship between the mean and variance can be relaxed using a quasi-likelihood approach (McCullagh and Nelder (1989)), or adding a latent process to the model (Zeger (1988)). The assumption of independence is also no longer necessary, if serial correlation is present in the data it can be incorporated using parameter driven (Zeger (1988)) or observation driven (Zeger and Qaqish (1988)) methods. Earlier approaches to modelling underlying seasonality used fixed parametric functions (such as sine and cosine terms), which meant that the magnitude and location of yearly peaks in the health outcome had to be the same for each year of the study. The introduction of generalised additive models (GAM) with parametric (e.g. natural cubic splines) and non-parametric (e.g. LOESS smoothers) smoothers have allowed the effect of seasonality to vary between years. GAMs have also been used to explore the possibility that the relationship between air pollution and health may be non-linear (for example, Schwartz (1994), Dominici et al. (2002), Daniels et al. (2004)).

Comparatively little research has allowed the effects of air pollution to change over time. These effects may show a seasonal pattern, which could be the result of an interaction with temperature or with another pollutant exhibiting a seasonal pattern. Alternatively, the temporal effects may show a long-term trend, which could result from such interactions, a change in the composition of individual pollutants over a number of years, or from a change in the size and structure of the population at risk. The effect of particulate matter was allowed to vary seasonally by Moolgavkar et al. (1995) and by Peng et al. (2005), but in both cases the change was constrained to take a strict parametric form. A less rigid approach was taken by Chiogna and Gaetan (2002) who analysed data from Birmingham, Alabama using a dynamic generalised linear model (DGLM). They allowed the effect of air pollution to vary freely over time, but restricted the amount in which it could change by constraining it to follow a first order random walk. This approach was generalised by Lee and Shaddick (2005), who replaced the first order random walk with an autoregressive process. However, the estimated temporal effects of air pollution were contaminated by noise, which is a result of the weak autoregressive constraint used in the model.

In this paper, we propose an alternative approach to estimating the time-varying effects of air pollution. Our approach retains the flexibility of the model of Chiogna and Gaetan (2002), but smooth temporal effects of air pollution are estimated using a penalised natural cubic spline rather than the first order random walk constraint. Both likelihood based and hierarchical Bayesian models are presented and applied to data from four U.S. cities.

1.1 Motivating example

The models presented here are applied to data on daily mortality and particulate matter from four cities in the U.S. (Cleveland, Detroit, Minneapolis and Pittsburgh), and were obtained from the R package 'NMMAPSdata' (Peng and Welty (2004)). The cities and study period (1^{st} January 1990 until the 31^{st} December 1994) were chosen because they had daily PM_{10} and weather data with the fewest missing values of all the 108 NMMAPS cities over the fourteen years of available data (1987 - 2000). The response data, $\mathbf{y}$, comprise daily counts of total non-accidental mortality for all age groups, and the PM_{10} data are daily averages across a number of monitors in each city. The trend in the PM_{10} data from the NMMAPS package had been removed, but for this analysis the original series have been recovered using the method suggested in the documentation accompanying the data. The weather data include daily mean temperature and mean dewpoint temperature. Indicator variables for day of the week (denoted DOW) are included because previous analyses of this data set have shown them to be important explanatory variables (Dominici et al. (2000)). Data from all four cities will be analysed using the same set of confounders in line with the model suggested by Dominici et al. (2000), however in this analysis all three age groups are combined in a single analysis.

The time-varying effects of air pollution are estimated using quasi-likelihood and Bayesian hierarchical models and the results from the two approaches are compared, with particular interest in the differences between the estimated smooth time-varying effects of pollution and the associated confidence and credible intervals.

The remainder of this paper is organised as follows. Section 2 describes models used to assess the short-term effects of air pollution on human health, and proposes a penalised spline model for such analysis. Section 3 introduces the likelihood based estimation approach which uses an iteratively re-weighted least squares procedure, and also describes an alternative Bayesian formulation where inference is based on Markov Chain Monte Carlo (MCMC) simulation. In section 4 the methodology is applied to data from the four U.S. cities as previously described. Finally Section 5 gives a concluding discussion and suggests some future extensions.

2 Modelling the relationship between daily measures of air pollution and health

For the purposes of this explanation, we assume that (acute) studies of air pollution and health comprise daily data describing the health outcomes, pollution and confounding factors from a specified study area from n consecutive days, $t = (1, \ldots, n)$. The data are a series of counts of health outcomes (such as respiratory mortality) denoted $\mathbf{y} = (y_1, \ldots, y_n)^T_{n\times 1}$, levels of air pollution $\mathbf{x} = (x_1, \ldots, x_n)^T_{n\times 1}$ and a matrix of covariates $Z = (\mathbf{z}_1^T, \ldots, \mathbf{z}_n^T)^T_{n\times q}$, such as weather variables, indicators for day of the week or an influenza epidemic, or basis functions for natural cubic splines that can be used to model long-term trends and seasonal variation. Covariates for day t are denoted by the vector $\mathbf{z}_t$ of length q. The parameters that model the effects of the confounding variables are collectively denoted $\delta_{q\times 1}$, and include any latent processes included in the model to allow for serial correlation or overdispersion.

Regression models for these data typically start with the assumption of an overdispersed Poisson model for the response, y_t, with the natural logarithm of the mean, μ_t, being linearly related to potential confounding variables $\mathbf{z}_t$ and air pollution x_t. The form specified for the air pollution and health relationship is a topic of much current research (see for example Dominici et al. (2002) and Daniels et al. (2004)), although a common general form can be expressed as

$$\begin{aligned} \log(\mu_t) &= \alpha + \beta_t x_t + \mathbf{z}_t^T \delta, \\ \beta_t &= S(t; \lambda), \end{aligned} \tag{1}$$

where $\beta = (\beta_1, \ldots, \beta_n)$, with β_t representing the effect of air pollution on day t, and S is an arbitrary smoothing function, parameterised by λ.

2.1 Specific forms of the air pollution and health relationship

A number of previously used forms for this function (i-v below) are now discussed, together with the time varying coefficient model proposed here (vi).

(i) $\beta_t = \beta$ for a constant effect of air pollution.

(ii) $\beta_t = 1$ and x_t is replaced with $f(x_t; \lambda)$ for a dose response relationship.

(iii) $\beta_t = \begin{cases} \beta_1 & \text{for spring days} \\ \beta_2 & \text{for summer days} \\ \beta_3 & \text{for autumn days} \\ \beta_4 & \text{for winter days} \end{cases}$ for a seasonally varying effect of air pollution.

(iv) $\beta_t = \gamma_0 + \gamma_1 \sin(2\pi t/365)/c_1 + \gamma_2 \cos(2\pi t/365)/c_2$ for a smooth seasonally varying effect of air pollution.

(v) $\beta_t = \beta_{t-1} + \omega_t \quad \omega_t \sim N(0, \sigma^2)$ for a time-varying effect of air pollution controlled by a first order random walk.

(vi) $\beta_t = S(t; \lambda)$ for a smooth time-varying effect of air pollution whose form is not fixed by the investigator *a-priori*.

Non time-varying effects (i)-(ii)
In the majority of studies, the covariate effect has been constant, $\beta_t = \beta$, see for example Mar et al. (2000) and Moolgavkar (2000), although a number of studies have allowed the effect of air pollution to vary as a non-linear function of the quantity of air pollution (see for example Schwartz (1994) and Daniels et al. (2004)).

Time-varying effects (iii)-(v)
Other extensions, as considered here, have allowed the relationship between air pollution and mortality to change over time. The relationship may exhibit seasonal behaviour or a long-term change over many years. Moolgavkar et al. (1995) estimated a different effect of air pollution for each season (model iii), constraining the parameter vector β to be a step function with only four distinct values. A step function is not an ideal model because it is implausible that the effects of air pollution are constant for three months and make a discontinuous jump at the end of each season. Peng et al. (2005) proposed a seasonal model (model iv), that forced β to be a sinusoidal function with a period of one year, which ensured a smooth time-varying effect, but the size and location of the seasonal effect are constant for each year of the study. Moreover, the underlying level of β was not allowed to exhibit a long-term trend. A more flexible model was suggested by Chiogna and Gaetan (2002) (model v), who did not force the effect of air pollution to take on a fixed parametric form, instead smoothness was enforced using a first order random walk for β_t. Lee and Shaddick (2005) proposed an extension that allowed β_t to follow a autoregressive process. However, although the estimate showed an underlying trend, it was contaminated with noise, a result of using a weak constraint that was not strong enough to force $\hat{\beta}$ to be smooth.

2.2 Time varying coefficient models *(vi)*

Time varying coefficient models (TVCM) are a special case of varying coefficient models (Hastie and Tibshirani (1993)) with the effect modifier being time. The model proposed here is that of equation (1), and the vector of air pollution effects are denoted by $\beta = (\beta_1, \ldots, \beta_n)_{n\times 1}$, where β_t is the effect on day t. These effects are modelled as an arbitrary smooth function $S(t; \lambda)$ using a penalised natural cubic spline, where the smoothness is controlled by the parameter λ.

A penalised spline is fully parametric and sits naturally within the generalised linear model framework. An advantage of this approach is that β is completely smooth, but its shape is determined by the data and not by an *a-priori* parametric form assumed by the investigator (see Peng et al. (2005)). The penalty allows the smoothness to be controlled by a single parameter, λ, rather than choosing the number and location of a set of knots. A larger number of knots are chosen than is necessary and the resulting estimate, $\hat{S}()$, is smoothed by enforcing a penalty on the coefficients of the spline, the form of which depends on the spline basis. A B-spline basis (Eilers and Marx (1996)) is used because it is numerically stable, and implementation of a penalty within a likelihood based analysis and a Bayesian framework is straightforward.

A cubic B-spline basis consists of p polynomial basis functions of degree 3. The jth basis function evaluated at time t is denoted $B_j(t) \quad j = 1, \ldots, p$. The matrix of spline basis functions is denoted $B_{n\times p}$, whose ijth element is $B_j(i)$. The penalised spline $S(t;\lambda)$ is a linear combination of the basis functions and parameters $\alpha = (\alpha_1, \ldots, \alpha_p)_{p\times 1}$, and at time t the effect of air pollution is modelled as

$$\beta_t = S(t;\lambda) = \sum_{j=1}^{p} \alpha_j B_j(t). \tag{2}$$

The smoothness of $S()$ is then ensured by penalising differences between the parameters of the spline. The form of the penalty term used in this paper is that suggested by Eilers and Marx (1996), who penalise kth order differences between the coefficients of the spline:

$$\lambda \sum_{j=k+1}^{p} (\Delta^k \alpha_j)^2, \tag{3}$$

where Δ denotes the difference operator. For example second order differences are given by

$$\sum_{j=3}^{p} (\Delta^2 \alpha_j)^2 \;=\; \sum_{j=3}^{p} (\alpha_j - 2\alpha_{j-1} + \alpha_{j-2})^2 \;=\; \alpha^T D \alpha, \tag{4}$$

where

$$D = \begin{pmatrix} 1 & -2 & 1 & & & & & & \\ -2 & 5 & -4 & 1 & & & & & \\ 1 & -4 & 6 & -4 & 1 & & & & \\ & 1 & -4 & 6 & -4 & 1 & & & \\ & & \vdots & \vdots & \vdots & \vdots & \vdots & & \\ & & & 1 & -4 & 6 & -4 & 1 & \\ & & & & 1 & -4 & 6 & -4 & 1 \\ & & & & & 1 & -4 & 5 & -2 \\ & & & & & & 1 & -2 & 1 \end{pmatrix}. \tag{5}$$

The influence of the penalty term is controlled by the smoothing parameter, λ, with larger values leading to the fitted curve being smoother. In a likelihood based approach, λ can be estimated using data driven methods such as generalised cross validation (GCV) or Akaike's information criterion (AIC). Penalised B-splines were adapted to a Bayesian setting by Lang and Brezger (2004), who replaced the difference penalty of order k in equation (3) with a kth order random walk prior for the parameters α_j, essentially replacing the kth order difference penalty by its stochastic analogue. For example, a second order difference penalty is replaced by a second order random walk prior,

$$\alpha_j|\alpha_{j-1}, \alpha_{j-2}, \tau^2 \sim \mathrm{N}(2\alpha_{j-1} - \alpha_{j-2}\ ,\ \tau^2) \qquad j = 3, \ldots, p. \tag{6}$$

The first two parameters (α_1, α_2) have non-informative priors, and the smoothing parameter is τ^2 which controls the magnitude of the differences between the parameters α. At the extremes, if $\tau^2 = 0$ all the parameters α_j would be equal which corresponds to maximal smoothing, and if $\tau^2 = \infty$ the parameters are not constrained which corresponds to no smoothing.

3 Statistical inference

3.1 Likelihood based analysis

The observations y_t come from an overdispersed Poisson distribution with the mean based on a measure of air pollution, x_t, and confounding variables $\mathbf{z}_t$. The effects of air pollution, $\beta = (\beta_1, \ldots, \beta_n)$, are modelled using a penalised spline (2) with a second order difference penalty (4). The full model is given by

$$\begin{array}{rcll} y_t & \sim & \mathrm{Poi}(\mu_t) & \quad t = 1, \ldots, n, \\ \mathrm{Var}(y_t) & = & \phi\mu_t, & \\ \log(\mu_t) & = & x_t\beta_t + \mathbf{z}_t^T\delta, & \\ \beta_t & = & B_t^T\alpha, & \end{array} \tag{7}$$

where $B_t^T = (B_1(t), \ldots, B_p(t))_{1\times p}$. The mean number of events on day t is denoted μ_t, with the vector of mean counts denoted by $\mu = (\mu_1, \ldots, \mu_n)_{n\times 1}$.

3.1.1 Parameter estimation

The vector of parameters (α, δ) can be estimated by iteratively re-weighted least squares with the addition of a penalty term, using the method of (Marx and Eilers (1998)). The model can be expressed as a generalised linear model by writing

$$\gamma = \begin{pmatrix} \delta \\ \alpha \end{pmatrix}_{q+p\times 1} \quad \text{and} \quad M = \begin{pmatrix} Z & XB \end{pmatrix}_{n\times q+p},$$

where the matrix $X = \mathrm{diag}(x_1, \ldots, x_n)_{n\times n}$ is multiplied by the matrix of spline basis functions $B_{n\times p}$, and the fitted value on day t is given by $\mu_t = \exp(M_t^T\gamma)$. The parameters γ are estimated using a penalised iteratively re-weighted least squares procedure with the second order difference penalty (4). The log-likelihood function for the Poisson data is given by

$$\mathcal{L}(\gamma; y) = \sum_{t=1}^{n} y_t \log(\mu_t) - \mu_t - \log(y_t!) - \lambda\gamma^T P\gamma,$$

where the penalty matrix P is given by

$$P = \begin{pmatrix} \mathbf{0}_{q\times q} & \mathbf{0}_{q\times p} \\ \mathbf{0}_{p\times q} & D_{p\times p} \end{pmatrix}.$$

Here $\mathbf{0}_{p\times q}$ denotes a matrix of zeros of appropriate dimension and D is given by equation (5). The maximum likelihood estimate $\hat{\gamma}$ is the solution to the iterative equation

$$\gamma^{k+1} = (M^T W(\gamma^k)M + \lambda P)^{-1} M^T W(\gamma^k)\tilde{\mathbf{y}}(\gamma^k),$$

where $W(\gamma) = \text{diag}(\mu_1(\gamma^k), \ldots, \mu_n(\gamma^k))$ and $\tilde{\mathbf{y}}(\gamma^k) = M\gamma^k + W(\gamma^k)^{-1}(\mathbf{y} - \mu(\gamma^k))$. The estimated asymptotic variance matrix for $\hat{\gamma}$ is given by the sandwich estimator

$$\widehat{Cov}(\hat{\gamma}) = \hat{\phi}(M^T W(\hat{\gamma})M + \lambda P)^{-1} M^T W(\hat{\gamma})M(M^T W(\hat{\gamma})M + \lambda P)^{-1},$$

with the overdispersion parameter given by

$$\hat{\phi} = \frac{1}{n - tr(H)} \sum_{t=1}^{n} \frac{(y_t - \mu_t)^2}{\mu_t}.$$

Here H denotes the 'Hat' matrix and is given by $H = M(M^T W(\hat{\gamma})M + \lambda P)^{-1} M^T W(\hat{\gamma})$, and the trace of the 'Hat' matrix is the approximate number of parameters used in the model. The approximate number of parameters for estimating the air pollution response function $\beta = B\alpha$ is given by $df_{poll} = tr(H) - q$ which is an alternative measure of the smoothness of the time-varying effect of air pollution. The parameter estimates $\hat{\gamma}$ and estimated covariance matrix are dependent on the smoothing parameter λ, which can be estimated by optimising data driven criteria, for example generalised cross-validation

(GCV) and Akaike's information criterion (AIC). For further details see Marx and Eilers (1998).

3.2 Bayesian analysis

The Bayesian hierarchical model assumes the sampling distribution of the health outcome, y_t, is Poisson, and the linear predictor depends on air pollution x_t and q confounding variables $\mathbf{z}_t$ with parameters δ. The coefficients of the spline, α, are penalised using second order random walk priors as described in section two, and the amount of smoothing is controlled by the variance parameter, τ^2. The proposed model is given by

$$\begin{aligned} y_t &\sim \text{Poi}(\mu_t) \qquad t = 1, \ldots, n, \\ \log(\mu_t) &= x_t \beta_t + \mathbf{z}_t \delta, \\ \beta_t &= B_t^T \alpha \\ \delta &\sim \text{N}(\mathbf{m}, \Sigma), \\ \alpha_j &\sim \text{N}(2\alpha_{j-1} - \alpha_{j-2}, \tau^2) \qquad \text{for } j = 3, \ldots, p, \\ \tau^2 &\sim \text{Inverse-Gamma}(e, f). \end{aligned} \tag{8}$$

Non-informative priors are given for (α_1, α_2), and the hyperparameters $(\mathbf{m}_{q\times 1}, \Sigma_{q\times q}, e, f)$ are fixed and chosen to make the priors non-informative.

3.2.1 Posterior simulation

Inference for the Bayesian hierarchical model (8) is performed using MCMC simulation. A hybrid block Metropolis-within Gibbs approach will be applied updating the parameters in blocks $[\delta = (\delta_1, \ldots, \delta_q), \alpha = (\alpha_1, \ldots, \alpha_p), \tau^2]$. The joint posterior distribution of all the parameters is given by

$$p(\alpha, \delta, \tau^2; y) \propto \prod_{t=1}^{n} \text{Poi}(y_t | \mu_t(\alpha, \delta)) \prod_{j=1}^{p} \text{N}(\alpha_j | 2\alpha_{j-1} - \alpha_{j-2}, \tau^2) \text{N}(\delta | m, \Sigma) \text{Inverse-Gamma}(e, f).$$

Simulation for this type of regression problem has been researched by Fahrmeir and Lang (2001), and the simulation algorithm described below is based on their work. The following three sub-sections describe the simulation procedure for α, δ and τ^2.

Simulation from $p(\alpha|\delta, \tau^2, y)$
The full conditional of $\alpha = (\alpha_1, \ldots, \alpha_p)$ is a product of n Poisson observations and a Gaussian second order random walk (RW(2)) smoothness prior. The prior is a product of $(p-2)$ RW(2) Gaussian distributions for α_j $j = 3, \ldots, p$, and flat priors for the first two parameters $p(\alpha_1)$ and $p(\alpha_2)$, which are both proportional to one. The full conditional for α is given by

$$\begin{aligned} p(\alpha|\delta, \tau^2, y) &\propto \prod_{t=1}^{n} p(y_t|\alpha, \delta) p(\alpha) \\ &\propto \prod_{t=1}^{n} \mathrm{Poi}(y_t|\mu_t(\alpha, \delta)) \prod_{j=3}^{p} \mathrm{N}(\alpha_j|2\alpha_{j-1} - \alpha_{j-2}\ ,\ \tau^2). \end{aligned}$$

The full conditional is non-standard and cannot efficiently be sampled from directly, so a block Metropolis-Hastings algorithm (Chib and Greenberg (1995)) is used. The proposal used by Fahrmeir and Lang (2001) is that suggested by Knorr-Held (1999), which bases the proposal density on the prior of α. The proposals are computationally cheap to generate and the Metropolis-Hastings acceptance probability is easy to calculate. The parameter vector, α, is sampled in blocks which is a sensible intermediate strategy between the low acceptance rates obtained from sampling high dimensional vectors, and the computational burden of sampling variables individually. The block size, g, is a tuning parameter to achieve the desired acceptance rates. Knorr-Held suggests that the algorithm is optimal when the acceptance rates are well below one. The proposal, $q()$, for each block, is based on the prior for α which can be re-written as a multivariate Gaussian distribution with a singular variance matrix:

$$\begin{aligned} p(\alpha|\tau^2) &= \prod_{j=3}^{p} \mathrm{N}(\alpha_j|2\alpha_{j-1} - \alpha_{j-2}, \tau^2) \\ &\propto \exp\left(-\frac{1}{2\tau^2}\alpha^T D\alpha\right). \end{aligned}$$

The $p \times p$ penalty matrix D is that given by equation (5), and the prior $p(\alpha|\tau^2)$ is not a proper multivariate Gaussian distribution because the precision matrix D is singular (the variance matrix does not exist). The proposal density $q()$ used to update a block of parameters $\alpha_{rs} = (\alpha_r, \ldots, \alpha_s)$, is the multivariate Gaussian conditional distribution for α_{rs} given all the other elements α, $\alpha_{-rs} = (\alpha_1, \ldots, \alpha_{r-1}, \alpha_{s+1}, \ldots, \alpha_p)$. For iteration j the proposal α^*_{rs} is drawn from the density

$$q[\alpha^{j-1}_{rs}, \alpha^*_{rs}] = p(\alpha_{rs}|\alpha_{-rs}, \tau^2) \;\sim\; \mathrm{N}(\mu_{rs}, \Sigma_{rs}).$$

The mean and variance can be obtained by using standard properties of the multivariate Gaussian distribution, and are given by

$$\begin{aligned} \mu_{rs} &= \begin{bmatrix} -D^{-1}_{rs}D_{1,r-1}\alpha_{1,r-1} & \text{if s=p} \\ -D^{-1}_{rs}D_{s+1,p}\alpha_{s+1,p} & \text{if r=1} \\ -D^{-1}_{rs}(D_{1,r-1}\alpha_{1,r-1} + D_{s+1,p}\alpha_{s+1,p}) & \text{otherwise} \end{bmatrix}, \\ \Sigma_{rs} &= \tau^2 D^{-1}_{rs}. \end{aligned}$$

Here D_{rs} is the square sub-matrix of D containing only rows and columns from r to s, and $D_{1,r-1}, D_{s+1,p}$ are the sub-matrices of D with rows r to s and columns 1 to $(r-1)$ and $(s+1)$ to p respectively. To avoid any mixing problems at the boundaries of each block, the first block will be of random length between 1 and g as suggested by Knorr-Held (1999). The acceptance probability of a move from α^{j-1}_{rs} to α^*_{rs} is given by

$$P_\alpha = \min\left\{1, \frac{\prod_{t=1}^{n} \mathrm{Poi}(y_t|\mu_t(\alpha^*, \delta^{j-1}))}{\prod_{t=1}^{n} \mathrm{Poi}(y_t|\mu_t(\alpha^{j-1}, \delta^{j-1}))}\right\},$$

where $\alpha^*_{-rs} = \alpha^{j-1}_{-rs}$ (for further details see Knorr-Held (1999)).

Simulation from $p(\delta|\alpha, \tau^2, y)$
The full conditional for δ is also non-standard because it is a product of n Poisson distributions for data y_t and a multivariate Gaussian prior for δ. The full conditional is given by

$$p(\delta|\alpha,\tau^2,y) \;\propto\; \prod_{t=1}^{n} \mathrm{Poi}(y_t|\mu_t(\alpha,\delta))\mathrm{N}(\delta|\mathbf{m},\Sigma).$$

The parameters of the prior $(\mathbf{m},\Sigma)$ are known and are chosen to be non-informative (i.e. the variance matrix Σ is diagonal with large variances). The full conditional is non-standard and simulation can be carried out using the block Metropolis-Hastings algorithm. The length of the blocks g can be altered as a tuning parameter to obtain good acceptance rates from the Metropolis-hastings simulation. A multivariate random walk proposal with diagonal variance matrix is used because it is computationally cheap to draw proposals from. The proposal distribution $q()$ for a block $\delta_{rs} = (\delta_r,\ldots,\delta_s)$ at iteration j is given by

$$q(\delta_{rs}^{j-1},\delta_{rs}^{*}) \sim \mathrm{N}(\delta_{rs}^{j-1}\ ,\ \nu\mathbf{I}),$$

where $\mathbf{I}$ denotes the identity matrix of appropriate size. The variance, ν, can be tuned to give good acceptance rates. The proposal distribution is symmetric with respect to $(\delta_{rs}^{j-1},\delta_{rs}^{*})$, so the acceptance probability is the ratio of full conditionals:

$$\begin{aligned} P_\delta &= \min\left\{1, \frac{p(\delta_{rs}^{*}|\delta_{-rs}^{*},\alpha^{j-1},(\tau^2)^{j-1})}{p(\delta_{rs}^{j}|\delta_{-rs}^{j},\alpha^{j-1},(\tau^2)^{j-1})}\right\} \\ &= \min\left\{1, \frac{\prod_{t=1}^{n}\mathrm{Poi}(y_t|\mu_t(\alpha^{j-1},\delta^{*}))\mathrm{N}(\delta^{*}|\mathbf{m},\Sigma)}{\prod_{t=1}^{n}\mathrm{Poi}(y_t|\mu_t(\alpha^{j-1},\delta^{j-1}))\mathrm{N}(\delta^{j-1}|\mathbf{m},\Sigma)}\right\}, \end{aligned}$$

where $\delta_{-rs}^{*} = \delta_{-rs}^{j-1}$

Simulation from $p(\tau^2|\alpha,\delta,y)$

Choosing a prior for a hierarchical variance parameter such as τ^2 is not straightforward and the subject has recently been reviewed by Gelman (2006). The conjugate prior distribution is Inverse-Gamma(e,f), but the choice of (e,f) may dramatically affect the posterior distribution of parameters with a large mass of probability near zero. Estimation of τ^2 is crucial because it controls the amount of smoothing, and common choices for a non-informative prior include Inverse-Gamma(e,e), where e is small and positive, or

Inverse-Gamma$(1, e)$. However, Gelman has shown that such priors can be highly informative, and decreasing e gives the prior and hence the posterior more probability mass close to zero. The result is that the posterior mean / mode of τ^2 becomes smaller and hence forces more smoothing. For the general case of an Inverse-Gamma(e, f) prior the full conditional of τ^2 is given by

$$\begin{aligned}
p(\tau^2|\alpha,\delta,y) &\propto \prod_{j=3}^{p} \mathrm{N}(\alpha_j|2\alpha_{j-1}-\alpha_{j-2}\ ,\ \tau^2)\text{Inverse-Gamma}(\tau^2|e,f) \\
&\propto (\tau^2)^{-(p-2)/2}\exp\left(-\frac{1}{2\tau^2}\alpha^T D\alpha\right)(\tau^2)^{-(e+1)}\exp(-f/\tau^2) \\
&\propto (\tau^2)^{-(e+\frac{p-2}{2}+1)}\exp\left[-\frac{f+\frac{1}{2}\alpha^T D\alpha}{\tau^2}\right] \\
&\propto \text{Inverse-Gamma}\left(e+\frac{p-2}{2}\ ,\ f+\frac{1}{2}\alpha^T D\alpha\right).
\end{aligned}$$

An alternative suggested by Gelman (2006) is to use a non-informative prior on the scale of standard deviation,

$$p(\tau)\ \propto\ 1 \qquad \Rightarrow \qquad p(\tau^2)\ \propto\ \tau^{-1}.$$

The resulting posterior distribution is still Inverse-Gamma and is given by

$$p(\tau^2|\alpha,\delta,y)\ \sim\ \text{Inverse-Gamma}\left(\frac{p-3}{2}-1\ ,\ \frac{1}{2}\alpha^T D\alpha\right).$$

4 Application

The likelihood and Bayesian models described in section three are applied to the data described in section 1.1. In the two models below (9 and 10), $S(var, df)$ denotes a natural cubic spline of the variable *var* with degrees of freedom *df*. The vector of confounding variables for both models are identical and are those chosen by Dominici et al. (2000) for the same data. The confounding variables include a set of indicator variables to model a *day of the week effect*, and splines to model seasonal variation and long-term trends in the mortality data. Splines are estimated for calendar time (t), and for temperature and dewpoint temperature (for the current days reading $S(temp_0)$, $S(dew_0)$ and an average of the previous three days readings $S(temp_{1-3})$, $S(dew_{1-3})$). The degrees of freedom for each spline were identical to those suggested by Dominici et al. (2000) who showed these adequately controlled for seasonal variation in these mortality data. The vector of parameters for the confounding variables $\delta_{60\times 1}$ includes an intercept δ_1, six indicator variables for a *day of the week effect* $\delta^*_{6\times 1}$, and the parameters of the five spline terms.

The pollution variable is the average PM_{10} measurement from the previous day (lag 1), $\mathrm{PM}_{10_{t-1}}$ as used by Dominici et al. (2000). The penalised cubic spline that estimates the time-varying effect of particulate matter is based on a set of 60 B-spline basis functions. Sensitivity analyses were carried out using up to 100 basis functions but the results showed no change. To force the estimated time-varying effects of air pollution to be smooth, the Bayesian model uses a second order random walk prior (6) for $\alpha_{60\times 1}$ with smoothing parameter τ^2, whereas the quasi-likelihood model uses a second order difference penalty (4) with smoothing parameter λ.

4.1 Quasi-likelihood model

The quasi-likelihood model is

$$
\begin{aligned}
y_t &\sim \mathrm{Poi}(\mu_t), \\
\mathrm{Var}(y_t) &= \phi\mu_t, \\
\log(\mu_t) &= \mathrm{PM}_{10_{t-1}}\beta_t + \delta_1 + S(t,35) + S(temp_0,6) + S(temp_{1-3},6), \\
&\quad + S(dew_0,3) + S(dew_{1-3},3) + \delta^* DOW. \\
\beta_t &= B_t^T\alpha
\end{aligned}
\tag{9}
$$

The autocorrelation (ACF) and partial autocorrelation (PACF) functions (not shown) show there is no serial correlation present in the residuals for any of the four data sets, so a specific component to model serial correlation is not required. The data also show little overdispersion (see Table 3).

4.2 Bayesian model

The Bayesian model is

$$
\begin{aligned}
y_t &\sim \mathrm{Poi}(\mu_t), \\
\log(\mu_t) &= \mathrm{PM}_{10_{t-1}}\beta_t + \delta_1 + S(t,35) + S(temp_0,6) + S(temp_{1-3},6), \\
&\quad +S(dew_0,3) + S(dew_{1-3},3) + \delta^* DOW, \\
\beta_t &= B_t^T \alpha \\
\delta &\sim N(\mathbf{m},\Sigma), \\
\alpha_j &\sim N(2\alpha_{j-1} - \alpha_{j-2}, \tau^2) \qquad \text{for } j = 3,\dots,60, \\
p(\tau) &\propto 1.
\end{aligned}
\tag{10}
$$

Sensitivity analyses were carried out using an Inverse-Gamma(e,e) prior for small values of e, and the results, in terms of τ^2, were sensitive to the value of e. Therefore the prior for τ^2 is non-informative on the standard deviation scale, as suggested for hierarchical variance parameters by Gelman (2006). The prior for δ is non-informative, with a diagonal variance matrix with large diagonal elements $\Sigma = \mathrm{diag}(1000,\dots,1000)_{60\times 1}$. The vector $\mathbf{m}$ is based on estimates from a generalised linear model of the same variables from earlier years (1987-1989).

4.3 Results

The two sets of analyses were implemented using code written and developed in R by the authors. The Bayesian model (10) was implemented using two parallel chains, where starting points were random draws from overdispersed versions of the priors (e.g. Student-t distributions replacing Gaussian distributions). There are a moderately large number of parameters ($\alpha_{60\times 1}$, $\delta_{60\times 1}$, τ^2), so to aid convergence, all explanatory variables were standardised to have a mean of zero and a standard deviation of one before inclusion in the model (and were subsequently back-transformed when obtaining results from the posterior distribution). Convergence was monitored using the methods of Gelman et al. (2003). The first 40,000 iterations were discarded as 'burn-in', by which point convergence was assessed to have been reached, and a further 125,000 iterations were then completed for each chain, which were then thinned by 5 to reduce autocorrelation in the samples, resulting in 25,000 samples from each chain. The fitted values μ_t^s were calculated for each day $t = 1, \ldots, 1826$ and for each sample $s = 1, \ldots, 50,000$ as

$$\begin{aligned} \mu_t^s &= \exp[\text{PM}_{10_{t-1}} B_t^T \alpha^s + \delta_1^s + S^s(t, 35) + S^s(temp_0, 6) + S^s(temp_{1-3}, 6) \\ &\quad + S^s(dew_0, 3) + S^s(dew_{1-3}, 3) + \delta^{*^s} DOW], \end{aligned} \tag{11}$$

where $S^s()$ denotes the simulation of parameters that make up the spline from sample s. Plots of the respiratory mortality data and the median fitted values from the Bayesian model are shown for the four data sets in Figure 1. The fitted values from the quasi-likelihood models are similar to those from the Bayesian models (not shown). Both sets of models estimate the mortality data clearly and capture the underlying seasonal trend in mortality.

4.3.1 Time-varying coefficient for PM_{10}

The time-varying effects of the previous days PM_{10} for the four cities are shown in Figures two to five. Each figure gives the time-varying estimate from the quasi-likelihood model (part (a)) and from the Bayesian model (part (b)) with appropriate 95% pointwise confidence / credible intervals. Table one gives the maximum, minimum and median values for the estimated time-varying effects of PM_{10} for each of the four cities.

The estimated time-varying coefficients are not consistent over the four cities, although they all show a long-term trend in the effect of air pollution, with both increasing (e.g. Figure 3 - Detroit), and decreasing (e.g. Figure 5 - Pittsburgh) effects over the five year period. In particular, the time-varying effects show no seasonal pattern for any of the four cities which is in contrast to the work of Peng et al. (2005), who used a model which forced the effects to adopt a sinusoidal shape with a period of one year. The Bayesian and quasi-likelihood estimates are generally similar in shape, with the Bayesian estimate showing slightly more curvature over the five year period. The biggest difference between the two approaches is in Minneapolis, for which the Bayesian model estimates a peak in the effects of PM_{10} followed by a drop at the end, which is in contrast to the steady increase estimated by the quasi-likelihood approach. However, the two estimates for both Cleveland and Pittsburgh are almost identical.

Table 1 shows that the median of the estimated time-varying PM_{10} for all four cities is higher for the quasi-likelihood model than for the Bayesian model. The Bayesian estimates have a greater range of values for all cities, which corroborates the visual assessment of more curvature in the time-varying coefficients, compared with the quasi-likelihood models. The 95% pointwise confidence / credible intervals are generally wider for the Bayesian model compared with the quasi-likelihood model, but the differences are not large. This increased variability in the Bayesian model is likely to be due to the method of estimation for the smoothing parameter. The quasi-likelihood model estimates λ and assumes that its value is fixed when calculating the 95% confidence intervals. In contrast the Bayesian approach correctly feeds the variability in τ^2 through into the estimates of the time-varying effects of PM_{10}. The result is that the Bayesian 95% credible interval correctly allows for the extra uncertainty in the smoothing parameter whereas the quasi-likelihood method does not.

The posterior distributions of the Bayesian smoothing parameters τ^2, for each city, are summarised in table two. The majority of the posterior mass in each distribution is close to zero, which represents close to maximal smoothing, as the second order differences between the parameters α are close to zero. In the limit of maximal smoothing, the estimated time-varying coefficient behaves like a low order polynomial, which can be seen from Figures two to five. In the quasi-likelihood model, the smoothing parameter was chosen by generalised cross validation (GCV), as suggested by Eilers and Marx (1996). The estimated number of parameters, which is an alternative measure of the smoothness of the time-varying effect, are shown in Table 3, together with the estimated overdispersion parameters ϕ. The low effective number of parameters is consistent with the long-term trends exhibited by the time-varying effects of PM_{10} and the assumption that the time-varying effects follow a low order polynomial. The models show almost no overdispersion or serial correlation which is consistent with other studies of these data.

5 Discussion

The regression models proposed in this paper allow the effects of air pollution to vary smoothly over time without having to make any *a-priori* decisions about their temporal shape. These effects may show a seasonal pattern, which could be the result of an interaction with temperature or with another pollutant exhibiting a seasonal pattern. Alternatively, the temporal effects may show a long-term trend, which could result from a change in the composition of individual pollutants over a number of years, or from a change in the structure of the population at risk. The time-varying effects were modelled using a penalised natural cubic spline, with implementation in both quasi-likelihood and Bayesian settings. The models were illustrated using data from four cities in the U.S. obtained from the National Morbidity, Mortality and Air Pollution Study.

The results from the Bayesian and quasi-likelihood models were similar, with both producing clear estimates of the underlying trend in mortality counts over the five year period. The residuals from both models (not shown) showed no structure and little serial correlation, suggesting that any seasonal variation had been adequately modelled. Both models estimated similar median effects of PM_{10} over the five year period, with the quasi-likelihood estimates being slightly larger (Relative Risk of an increase of $10\mu g$ of PM_{10} for each city for the Bayesian and quasi-likelihood models respectively: Cleveland 1.004 and 1.006, Detroit 1.004 and 1.005, Minneapolis 1.004 and 1.005, and Pittsburgh 1.003 and 1.004). These median effect sizes are similar to those found by other investigators who analysed these data (see for example Samet et al. (2000)).

The Bayesian 95% credible intervals were generally wider than the corresponding quasi-likelihood confidence intervals, which probably results from the estimation of the respective smoothing parameters (λ, τ^2). The Bayesian model correctly allowed the uncertainty associated with the smoothing parameter to be incorporated into the model, leading to more realistic estimates of the variability in the effects of air pollution. In contrast the quasi-likelihood confidence intervals are likely to be too narrow because the smoothing parameter of the penalised spline is assumed to be estimated without error.

The estimated time-varying effects of PM_{10} were consistent between the two approaches, showing the same patterns of increasing or decreasing long-term trends across the four cities. These changes could be due to a confounding effect that varies over time, such as a change in the composition of PM_{10} and in particular the component that might be detrimental to human health. Noticeably, the time-varying coefficients did not show a seasonal effect for any of the four cities, which is in contrast to the results from Peng et al. (2005) and Moolgavkar et al. (1995), although in those cases, the structure of the time-varying coefficient enforced an interaction between PM_{10} and seasonality. Although the estimated effect of PM_{10} exhibited a long-term trend, the confidence / credible intervals were too wide to discard the possibility of a constant effect over time. This is probably the result of the small effect sizes that the models are trying to estimate.

At present, the models are applied separately to data from the individual cities. Future development could extend the methodology to deal with data from multiple cities simultaneously. The model could then combine data over several cities to produce an overall effect (which can exhibit temporal variability), which would be based on larger amounts of data. This could be done by adding a random effects term (which would fit naturally within the Bayesian hierarchical model), and if there was evidence of correlation between the effects from cities in close proximity, then the covariance structure could be formulated to incorporate a spatial structure.

References

Chib, S. and E. Greenberg (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician 49*, 327–335.

Chiogna, M. and C. Gaetan (2002). Dynamic generalized linear models with applications to environmental epidemiology. *Applied Statistics 51*, 453–468.

Daniels, M., F. Dominici, S. Zeger, and J. Samet (2004). The National Morbidity, Mortality, and Air Pollution Study Part III: Concentration-Response Curves and Thresholds for the 20 Largest US Cities. *HEI Project 96-97*, 1–21.

Dominici, F., M. Daniels, S. Zeger, and J. Samet (2002). Air Pollution and Mortality: Estimating Regional and National Dose-Response Relationships. *Journal of the American Statistical Association 97*, 100–111.

Dominici, F., J. Samet, and S. Zeger (2000). Combining evidence on air pollution and daily mortality from the 20 largest US cities: a hierarchical modelling strategy. *Journal of the Royal Satistical Society Series A 163*, 263–302.

Eilers, P. and B. Marx (1996). Flexible Smoothing with B-splines and Penalties. *Statistical Science 11*, 89–1218.

Fahrmeir, L. and S. Lang (2001). Bayesian inference for generalized additive mixed models based on Markov ranodm field priors. *Applied Statistics 50*, 201–220.

Gelman, A. (2006). Prior distirbutions for variance parameters in hierarchical models. *Bayesian Analysis to appear*.

Gelman, A., J. Carlin, H. Stern, and D. Rubin (2003). *Bayesian Data Analysis* (2nd ed.). Chapman and Hall.

Hastie, T. and R. Tibshirani (1990). *Generalized Additive Models* (1st ed.). Chapman and Hall.

Hastie, T. and R. Tibshirani (1993). Varying-coefficient Models. *Journal of the Royal Satistical Society Series B 55*, 757–796.

Knorr-Held, L. (1999). Conditional Prior Proposals in Dynamic Models. *Scandinavian Journal of Statistics 26*, 129–144.

Lang, S. and A. Brezger (2004). Bayesian P-Splines. *Journal of Computational and Graphical Statistics 13*, 183–212.

Lee, D. and G. Shaddick (2005). Modelling the effects of air pollution on health using bayesian dynamic generalised linear models. *University of Bath technical report series 2*.

Mar, F., G. Norris, J. Koenig, and T. Larson (2000). Associations between Air Pollution and Mortality in Phoenix, 1995-1997. *Environmental Health Perspectives 108*, 347–353.

Marx, B. and P. Eilers (1998). Direct generalized additive modelling with penalized likelihood. *Computational Statistics and Data Analysis 28*, 193–209.

McCullagh, P. and J. Nelder (1989). *Generalized Linear Models* (2nd ed.). Chapman and Hall.

Moolgavkar, S. (2000). Air Pollution and Daily Mortality in Three U.S. Counties. *Environmental Health Perspectives 108*, 777–784.

Moolgavkar, S., H. Luebeck, T. Hall, and E. Anderson (1995). Air Pollution and Daily Mortality in Philadelphia. *Epidemiology 6*, 476–484.

Peng, R., F. Dominici, R. Pastor-Barriuso, S. Zeger, and J. Samet (2005). Seasonal Analyses of Air Pollution and Mortality in 100 U.S. Cities. *American Journal of Epidemiology 161*, 585–594.

Peng, R. D. and L. J. Welty (2004). The NMMAPSdata package. *R News 4*(2), 10–14.

Samet, J., S. Zeger, F. Dominici, F. Curriero, I. Coursac, D. Dockery, J. Schwartz, and A. Zanobetti (2000). The National Morbidity, Mortality, and Air Pollution Study Part II: Morbidity and Mortality, from Air Pollution in the United States. *HEI Project 96-97*, 5–47.

Schwartz, J. (1991). Particulate Air Pollution and Daily Mortality in Detroit. *Environmental Research 56*, 204–213.

Schwartz, J. (1994). Nonparametric smoothing in the analysis of air pollution and respiratory illness. *The Canadian Journal of Statistics 22*, 471–487.

Zeger, S. (1988). A regression model for time series of counts. *Biometrika 75*, 621–629.

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

Figure 1: **Counts of respiratory mortality (*) and estimates from the Bayesian model (-)**

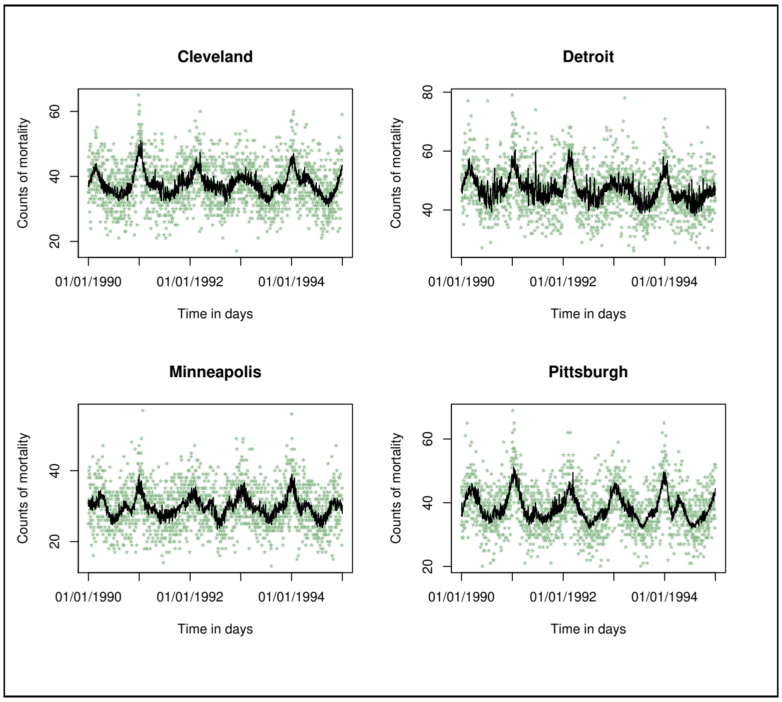

Table 1: **Quantiles of the time-varying coefficients for the four cities**

City	**Quasi-likelihood**			**Bayesian**		
	min	median	max	min	median	max
Cleveland	-0.000171	0.000597	0.000737	-0.000148	0.000398	0.000880
Detroit	-3.58^{-7}	0.000498	0.000863	-9.77^{-7}	0.000420	0.00107
Minneapolis	-3.98^{-7}	0.000510	0.000890	-0.0011	0.000410	0.00146
Pittsburgh	-0.000532	0.000352	0.00119	-0.000725	0.000348	0.00106

Figure 2: **Time-varying coefficient for Cleveland: (a) Quasi-likelihood model, (b) Bayesian model. The shading are 95% confidence / credible intervals.**

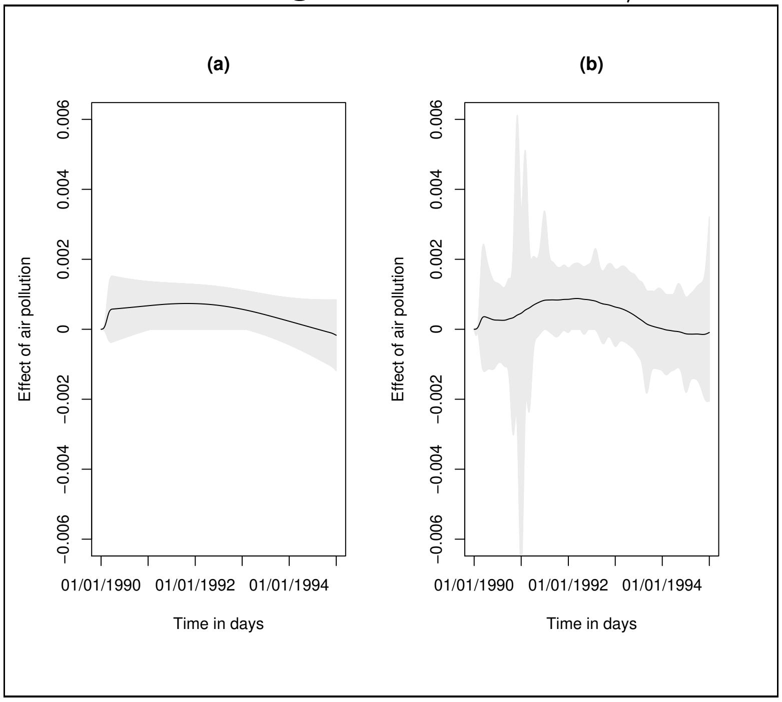

Table 2: **Posterior estimates of the smoothing parameter τ^2**

City	**Posterior quantiles**				
	2.5%	25%	median	75%	97.5%
Cleveland	2.06^{-08}	3.26^{-07}	1.13^{-06}	3.92^{-06}	0.040
Detroit	1.17^{-08}	1.56^{-07}	5.32^{-07}	1.63^{-06}	1.35^{-05}
Minneapolis	2.28^{-08}	4.88^{-07}	1.69^{-06}	5.06^{-06}	2.86^{-05}
Pittsburgh	7.98^{-09}	1.78^{-07}	6.08^{-07}	1.83^{-06}	1.12^{-05}

Figure 3: **Time-varying coefficient for Detroit: (a) Quasi-likelihood model, (b) Bayesian model. The shading are 95% confidence / credible intervals.**

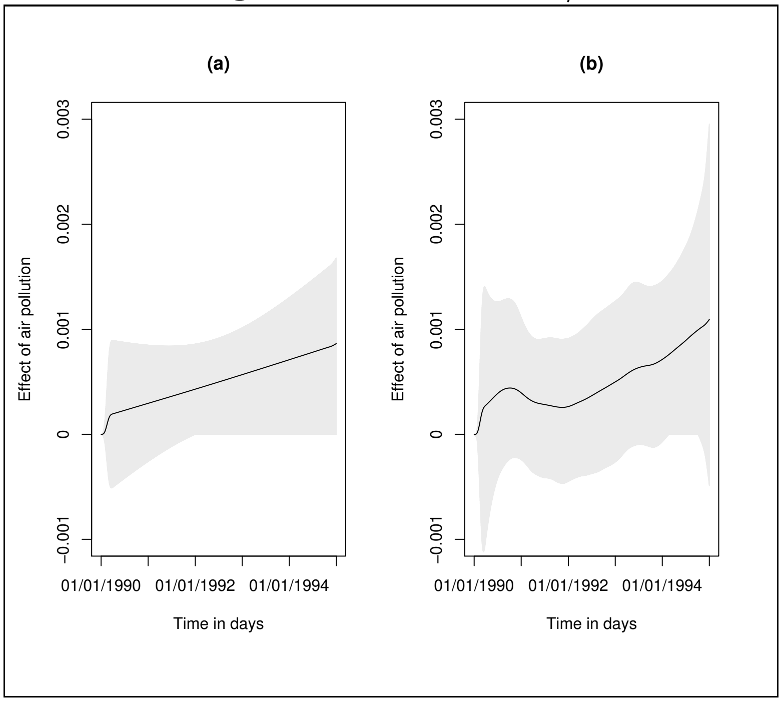

Table 3: **Estimated number of parameters used to estimate the time-vary coefficient**

City	**Parameters**	ϕ
Cleveland	2.6	1.03
Detroit	2.0	1.06
Minneapolis	2.0	0.99
Pittsburgh	2.0	0.98

Figure 4: **Time-varying coefficient for Minneapolis: (a) Quasi-likelihood model, (b) Bayesian model. The shading are 95% confidence / credible intervals.**

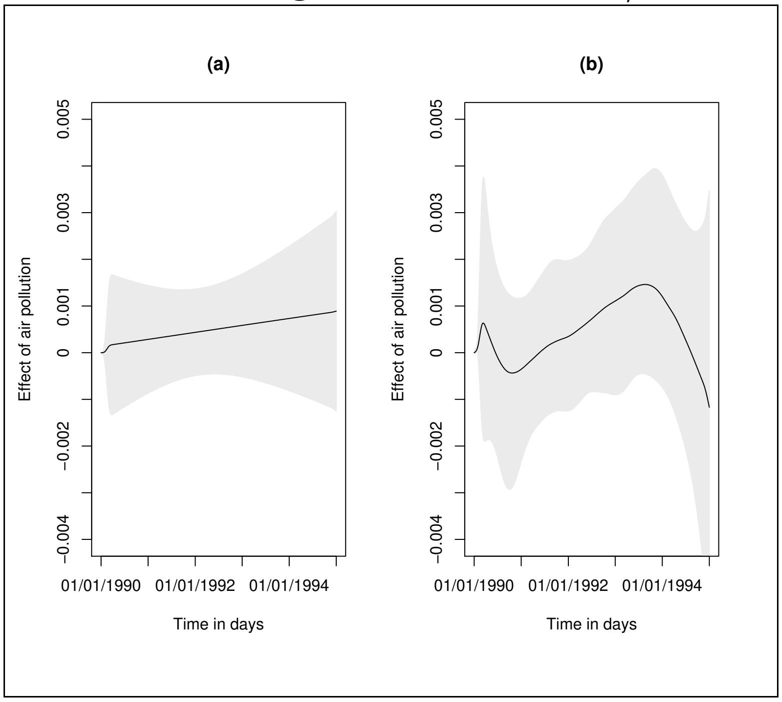

Figure 5: **Time-varying coefficient for Pittsburgh: (a) Quasi-likelihood model, (b) Bayesian model. The shading are** 95% **confidence / credible intervals.**