Air Pollution Epidemiology
R.L.S.
02/05/09

1 Introduction and Background

To be written

2 Time Series Analysis

One of the major sampling designs for air pollution studies is *time series analysis*. In the air pollution context this phrase has a specific meaning: it refers to analyses where daily observations of some measure of health outcomes (e.g. deaths, hospital admissions, asthma attacks) are regressed against a variety of daily predictors. Common predictors include weather, day of week, and season, since health outcomes are known to vary with all of these factors. In addition, long-term trend terms are usually included to allow for such factors as people moving in an out of a city, changes in health care quality, etc. On top of all these, it is usual to add at least one indicator of air pollution, such as the daily measurement of ground-level ozone (O_3 or particulate matter. The latter is usually represented as either PM_{10} or $PM_{2.5}$, referring to airborne particles of diameter less than 10 or 2.5 microns respectively. For many years, PM_{10} was used as the main indicator of atmospheric particles, but since a revision of United Stated Environmental Protection Agency (USEPA) standards in 1997, this has largely been replaced by $PM_{2.5}$, which is widely regarded as the more potent indicator of particulate matter pollution, the theory being that small particles penetrate further into the lungs and therefore do more damage.

If the coefficient of the air pollution variable is positive and statistically significant, this is usually interpreted as evidence of a causal effect. Although the whole technique is called time series analysis, as usually practiced in epidemiology it does not involve traditional time series techniques such as autocorrelations and spectra. This is because, when applied to non-infectious diseases, there is no reason to expect autocorrelation, though it is still a good idea to check up on this as well as other assumptions of the model.

We now highlight a number of features that need to be taken into account in fitting a time series model.

2.1 Sampling distribution

Since the outcomes are counts (e.g. deaths), it is natural to use a Poisson distribution. For strict Poisson data, the variance and the mean are equal; when the variance exceed the mean, the data are called *overdispersed*. Both Poisson and overdispersed models for count data may be fitted using the `glm` function in R, as we shall see.

2.2 Meteorological covariates

The most common meteorological variables are temperature and dewpoint, since by combining these variables it is believed we can account for humidity. However, in both cases, the response function is expected to be nonlinear — deaths are greatest when it is either very hot or cold. Therefore, we would like to use a nonlinear function of both temperature and humidity. While simple parametric functions such as quadratic polynomials and piecewise linear functions are possible, these are not

particularly flexible. An alternative is to use expansions in terms of *basis functions*. Thus, a variable x (temperature, say) is included in the regression through a function of form

$$f(x) \quad = \quad \sum_{j=1}^{m} \beta_j B_j(x), \tag{1}$$

where $\beta_1, ..., \beta_m$ are unknown regression coefficients and $B_j(x),\ j = 1, ..., m$ are predetermined basis functions. The number of basis functions m is called the *degrees of freedom* of the model: the larger m is, the more accurately a general smooth nonlinear $f(x)$ may be represented, but this is at the costs of increasing variance in estimating the β_j's and possibly other problems such as multicollinearity. There is also the question of what $B_j(x)$ functions to take. Well known choices would be polynomials (e.g. $B_j(x) = x^j,\ j = 1, 2, 3, ...$) or Fourier expansions (where each $B_j(x)$ is either $\cos \omega_j x$ or $\sin \omega_j x$ for some frequency ω_j) but these have well-known disadvantages, e.g. polynomials tend to be numerically unstable as soon as m gets large (also, their extrapolation properties are terrible), while Fourier series expansions really only work well for periodic functions. An alternative is to use *cubic splines*, in which the $B_j(x)$ ar epiecewise cubic polynomials with coefficients chosen so that both the resulting function (1) and its first-order derivatives are continuous in x. In R, the `splines` package allows the user to create a matrix of spline basis functions through the operator `bs` (for B-splines) or `ns` (natural splines), which refer to two different methods for constructing piecewise cubic approximating functions. For example, if the variable x consists of a vector of n temperature values, whose individual values are $x_1, ..., x_n$, then the R operator

```
fx<-ns(x,df=m)
```

forms a $n \times m$ matrix (here called `fx`) whose (i, j) entry is $B_j(x_i)$. This matrix may then be inserted directly into a regression function.

We also need to consider lagged meteorology covariates (to be completed...)

2.3 Seasonality and long-term trend

2.4 Lagged air pollution variables: the constrained and unconstrained distributed lag models

2.5 Confounding and effect modification

3 Combining Data in Independent Linear Regressions: The "tlnise" Algorithm

Everson and Morris (2000) considered a "two-level normal" hierarchical model, as follows:

$$\mathbf{Y}_j \mid \boldsymbol{\theta}_j \quad \sim \quad N_p[\boldsymbol{\theta}_j, V_j] \quad (V_j \text{ known}, \quad j = 1, ..., J, \tag{2}$$

$$\boldsymbol{\theta}_j \quad \sim \quad N_p[W_j^T \boldsymbol{\gamma}, A] \tag{3}$$

where W_j is a $r \times p$ matrix given below ($r = pq$, some $q < J$), $\boldsymbol{\gamma}$ is an unknown r-dimensional vector, and A is a common unknown covariance matrix. It is assumed that W_j is given by

$$W_j = \begin{pmatrix} w_{j1} & & & & \\ \vdots & & & & \\ w_{jq} & & & & \\ & w_{j1} & & & \\ & \vdots & & & \\ & w_{jq} & & & \\ & & \ddots & & \\ & & & w_{j1} & \\ & & & \vdots & \\ & & & w_{jq} & \end{pmatrix} \tag{4}$$

where $\mathbf{w}_j = (w_{j1}, ..., w_{jq})^T$ is a vector of covariates for the j'th vector $\boldsymbol{\theta}_j$.

The usual context for application of this result is when $\mathbf{Y}_j$ is a vector of estimates of $\boldsymbol{\theta}_j$ from the j'th city or the j'th study (so $\mathbf{Y}_j$ could also be written $\hat{\boldsymbol{\theta}}_j$, but $\mathbf{Y}_j$ is more convenient notation when there is no danger of confusion with the original data from which $\hat{\boldsymbol{\theta}}_j$ was calculated). We assume the covariance matrices V_j are estimated from sufficiently large samples that they may in practice be treated as known. Also note the assumed multivariate normal distribution of $\mathbf{Y}_j$ (or $\hat{\boldsymbol{\theta}}_j$) given $\boldsymbol{\theta}_j$: this is exact in the case of a simple OLS regression, but is itself an approximation in more complex cases such as a generalized linear model regression with Poisson error structure.

The theory divides into two cases: first we consider the case where all of $V_1, ..., V_J$ are the same, then we extend it to the general case. The first case is not usually appropriate in practice, but it motivates the method used for the second case.

3.1 The equal covariance case

Suppose now $V_1 = ... = V_J = V_0$ in (2). Define

$$B_0 = V_0^{1/2}(V_0 + A)^{-1}V_0^{1/2} \tag{5}$$

where $V_0^{1/2}$ is the symmetric square root of V_0. Because A is non-negative definite, all the eigenvalues of B_0 lie in the interval $(0, 1]$, a condition we denote by writing $0 < B_0 \leq I$.

The marginal likelihood of B_0 is given by

$$L_0(B_0) \propto |B_0|^{(J-q)/2} \exp\left\{-\frac{1}{2}\mathrm{tr}(SB_0)\right\}, \quad 0 < B_0 \leq I. \tag{6}$$

Here

$$S = V_0^{-1/2}\left\{\sum(\mathbf{Y}_j - W_j^T\hat{\boldsymbol{\gamma}})(\mathbf{Y}_j - W_j^T\hat{\boldsymbol{\gamma}})^T\right\}V_0^{-1/2}, \tag{7}$$

$$\hat{\boldsymbol{\gamma}} = (\sum W_j W_j^T)^{-1}\sum W_j \mathbf{Y}_j^T. \tag{8}$$

The likelihood (6) has the form of a "constrained Wishart distribution", as follows.

First recall the standard definition of a Wishart distribution:

Definition 1. Let $X_1, ..., X_m$ be independent $MVN_p[0, \Sigma]$ and $M = \sum_{j=1}^m X_j X_j^T$. The M is said to have the Wishart distribution with m degrees of freedom and covariance matrix Σ, notation $M \sim W_p[\Sigma, m]$. The density is

$$f_M(M) = \frac{|M|^{(m-p-1)/2} \exp\left\{-\frac{1}{2}\text{tr}(\Sigma^{-1}M)\right\}}{2^{mp/2}\pi^{p(p-1)/4}|\Sigma|^{m/2}\prod_{j=1}^p \Gamma\left(\frac{m+1-j}{2}\right)}$$

with respect to Lebesgue measure on $\mathcal{R}^{p(p+1)/2}$, restricted to positive definite symmetric matrices M. The case $\Sigma = I_p$ (the $p \times p$ identity matrix) is known as *standard Wishart.*

We now turn to the definition of *constrained Wishart*, introduced by Everson and Morris.

Definition 2. A symmetric $p \times p$ random matrix X has a constrained Wishart distribution $\text{CWish}_p(\nu, \Sigma; Q)$, of dimension p, degrees of freedom $\nu > 0$, symmetric $p \times p$ scale matrix Σ and diagonal constraint matrix Q, if the density function $f(X)$ satisfies

$$f(X) \quad \propto \quad |X|^{(\nu-p-1)/2} \exp\left\{-\frac{1}{2}\text{tr}\left(\Sigma^{-1}X\right)\right\}, \quad 0 < X \leq Q. \tag{9}$$

Now consider a class of prior densities,

$$\pi(B_0) \propto |B_0|^{\nu_*-p-1)/2} \exp\left\{-\frac{1}{2}\text{tr}\left(S_* B_0\right)\right\}, \quad 0 < B_0 \leq I. \tag{10}$$

This is of the form of a constrained Wishart distribution, $B_0 \sim \text{CWish}_p(\nu_*, S_*^{-1}; I)$ provided $\nu_* > 0$, $S_* > 0$, but may also be extended to some cases of improper priors. The following theorem (Theorem 1 of Everson and Morris (2000)) established that the prior (10) is conjugate for the likelihood (6):

Theorem 1. Assume a two-level hierarchical model defined by equations (2) and (3), with all $V_j = V_0$, a known $p \times p$ symmetric positive definite matrix. Let $B_0 = V_0^{1/2}(V_0 + A)^{-1}V_0^{1/2}$ have the prior density (10) for $p \times p$ symmetric matrix $S_* \geq 0$ and some $\nu_* > -(J - q)$. Then, conditional on $\mathbf{Y}$,

$$B_0 \mid \mathbf{Y} \sim \text{CWish}_p\{J - q + \nu_*, (S + S_*)^{-1}; I\}, \tag{11}$$

with S as in (7).

As a result of this theorem, it is possible to construct a Monte Carlo sample from the posterior distribution of B_0 by drawing independent Wishart matrices and retaining only those for which $B_0 \leq I$. It is then possible to transform $A = V_0^{1/2}(B_0^{-1} - I)V_0^{1/2}$ to generate a posterior sample from the matrix A.

3.2 The unequal covariance case

Now consider the case where $V_1, ..., V_J$ are unequal. We still define a V_0 which is in principle arbitrary, but in practice usually taken to be the sample mean of $V_1, ..., V_J$. As before, let $B_0 = V_0^{1/2}(V_0+A)^{-1}V_0^{1/2}$. In this case the result is that a suitable constrained Wishart density "envelops" the posterior density of B_0 given $\mathbf{Y}$. For two densities f_0 and f_1, we say that f_0 envelops f_1 if the ratio f_1/f_0 is bounded above.

Theorem 2. For a given symmetric $p \times p$ matrix $V_0 > 0$, $f(B_0 \mid \mathbf{Y})$ is the posterior density function for $B_0 = V_0^{1/2}(V_0 + A)^{-1}V_0^{1/2}$, assuming some prior density $\pi(B_0)$ of the form (10) with

$\nu_* > -(J-q)$. Let f_0 be the density function for any CWish$(\nu, \Sigma; I)$ distribution. Then f_0 envelops $f(B_0 \mid \mathbf{Y})$ for any $\Sigma > 0$ and for any $\nu \leq J - q + \nu_*$.

Theorem 2 shows that it is possible to generate Monte Carlo samples from f_1 by *rejection sampling*. Suppose f_0 envelops f_1, with $f_1 \leq M f_0,\ M < \infty$. Generate a Monte Carlo X from f_0; "accept" X with probability $\frac{f_1(X)}{M f_0(X)}$; if reject, draw a new X and repeated the process until some X is accepted. This X then has density f_1. The process may be repeated as many times as desired, to generate an indepedentent, identically distributed sample of observations from f_1.

This has the advantage over the Hastings-Metropolis algorithm that the sample values are independent and have exactly the right distribution, whereas the Hastings-Metropolis algorithm produces samples that only asymptotically have those properties. On the other hand, we do need to know M. The proof in Everson and Morris (2000) actually includes a specific value of M, though it is not clear how sharp the bound is in practice. Two alternatives are *importance sampling*, which can be used to estimate the mean of any function of $X \sim f_1$ without directly generating Monte Carlo samples from f_1, or *sampling-importance resampling* (Rubin 1987), in which a large sample $X_1, ..., X_N$ is generated from f_0 and then, the values are *resampled* with probabilities proportional to $f_1(X_i)/f_0(X_i),\ 1 \leq i \leq N$. The disadvantage of the sampling-importance resampling is that it may be necessary to generate a sample size N much larger than the one that is actually required, but it is a way of performing rejection sampling without directly knowing M.

3.3 Technical Details

The joint density of $\{(\boldsymbol{\theta}_j, \mathbf{Y}_j),\ j = 1, ..., J\}$ is

$$\prod_{j=1}^{J} \left[|A|^{-1/2} \exp\left\{ -\frac{1}{2}(\boldsymbol{\theta}_j - W_j^T\boldsymbol{\gamma})^T A^{-1}(\boldsymbol{\theta}_j - W_j^T\boldsymbol{\gamma}) - \frac{1}{2}(\mathbf{Y}_j - \boldsymbol{\theta}_j)^T V_j^{-1}(\mathbf{Y}_j - \boldsymbol{\theta}_j) \right\} \right]. \qquad (12)$$

We integrate (12), first with respect to $\boldsymbol{\theta}_j$ for each j, then with respect to $\boldsymbol{\gamma}$.

The first integral requires completing the square in

$$\begin{aligned} & (\boldsymbol{\theta}_j - W_j^T\boldsymbol{\gamma})^T A^{-1}(\boldsymbol{\theta}_j - W_j^T\boldsymbol{\gamma}) + (\mathbf{Y}_j - \boldsymbol{\theta}_j)^T V_j^{-1}(\mathbf{Y}_j - \boldsymbol{\theta}_j) \\ = \ & \boldsymbol{\theta}_j^T(A^{-1} + V_j^{-1})\boldsymbol{\theta}_j - 2\boldsymbol{\theta}_j^T(A^{-1}W_j^T\boldsymbol{\gamma} + V_j^{-1}\mathbf{Y}_j) + \boldsymbol{\gamma}^T W_j A^{-1} W_j^T \boldsymbol{\gamma} + \mathbf{Y}_j^T V_j^{-1}\mathbf{Y}_j \end{aligned}$$

But the identity

$$\boldsymbol{\theta}^T B\boldsymbol{\theta} - 2\boldsymbol{\theta}^T\mathbf{c} + \mathbf{d} \ = \ (\boldsymbol{\theta} - B^{-1}\mathbf{c})^T B(\boldsymbol{\theta} - B^{-1}\mathbf{c}) + \mathbf{d} - \mathbf{c}^T B^{-1}\mathbf{c}$$

(where $\boldsymbol{\theta}$, $\mathbf{c}$ and $\mathbf{d}$ are p-dimensional vectors, B a $p \times B$ symmetric positive definite matrix) leads to

$$\int \exp\left\{ -\frac{1}{2}\left(\boldsymbol{\theta}^T B\boldsymbol{\theta} - 2\boldsymbol{\theta}^T\mathbf{c} + \mathbf{d} \right) \right\} d\boldsymbol{\theta} \ = \ (2\pi)^{p/2}|B|^{-1/2} \exp\left\{ -\frac{1}{2}\left(\mathbf{d} - \mathbf{c}^T B^{-1}\mathbf{c} \right) \right\}. \qquad (13)$$

Identifying B with $A^{-1} + V_j^{-1}$, $\mathbf{c}$ with $A^{-1}W_j^T\boldsymbol{\gamma} + V_j^{-1}\mathbf{Y}_j$ and $\mathbf{d}$ with $\boldsymbol{\gamma}^T W_j A^{-1} W_j^T\boldsymbol{\gamma} + \mathbf{Y}_j^T V_j^{-1}\mathbf{Y}_j$, we integrate (12) with respect to each of $\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_J$ to get

$$\prod_{j=1}^{J} \Bigg[|A|^{-1/2}(2\pi)^{p/2}|A^{-1} + V_j^{-1}|^{-1/2} \exp\Bigg\{ -\frac{1}{2}\Big(\boldsymbol{\gamma}^T W_j A^{-1} W_j^T\boldsymbol{\gamma} + \mathbf{Y}_j^T V_j^{-1}\mathbf{Y}_j \\ -(A^{-1}W_j^T\boldsymbol{\gamma} + V_j^{-1}\mathbf{Y}_j)^T(A^{-1} + V_j^{-1})^{-1}(A^{-1}W_j^T\boldsymbol{\gamma} + V_j^{-1}\mathbf{Y}_j) \Big) \Bigg\} \Bigg]$$

$$\propto \quad |A|^{-J/2}\Big(\prod_{j=1}^{J}|A^{-1}+V_j^{-1}|\Big)^{-1/2}\cdot\exp\Big[-\frac{1}{2}\sum_j\Big\{\boldsymbol{\gamma}^T W_j(A+V_j)^{-1}W_j^T\boldsymbol{\gamma} \\ -2\boldsymbol{\gamma}^T W_j(A+V_j)^{-1}\mathbf{Y}_j+\mathbf{Y}_j^T(A+V_j)^{-1}\mathbf{Y}_j\Big\}\Big]. \tag{14}$$

We now integrate (14) with respect to $\boldsymbol{\gamma}$, again using (13) with $\boldsymbol{\theta}$ replaced by $\boldsymbol{\gamma}$, where we identify B with the $r\times r$ matrix $\sum_j W_j(A+V_j)^{-1}W_j^T$, $\mathbf{c}$ with $\sum_j\{W_j(A+V_j)^{-1}\mathbf{Y}_j\}$ and $\mathbf{d}$ with $\sum_j\Big\{\mathbf{Y}_j^T(A+V_j)^{-1}\mathbf{Y}_j\Big\}$. The result is proportional to

$$|A|^{-J/2}\Big(\prod_{j=1}^{J}|A^{-1}+V_j^{-1}|\Big)^{-1/2}\Big|\sum_j W_j(A+V_j)^{-1}W_j^T\Big|^{-1/2}e^{-Q/2} \tag{15}$$

with

$$\begin{aligned} Q &= \sum_j\Big\{\mathbf{Y}_j^T(A+V_j)^{-1}\mathbf{Y}_j\Big\}-\Big[\sum_j\Big\{W_j(A+V_j)^{-1}\mathbf{Y}_j\Big\}\Big]^T\cdot \\ &\quad\cdot\Big\{\sum_j W_j(A+V_j)^{-1}W_j^T\Big\}^{-1}\Big[\sum_j\Big\{W_j(A+V_j)^{-1}\mathbf{Y}_j\Big\}\Big] \\ &= \sum_j(\mathbf{Y}_j-W_j^T\boldsymbol{\gamma}^*)^T(A+V_j)^{-1}(\mathbf{Y}_j-W_j^T\boldsymbol{\gamma}^*) \end{aligned} \tag{16}$$

where

$$\boldsymbol{\gamma}^* = \Big(\sum_j W_j(A+V_j)^{-1}W_j^T\Big)^{-1}\Big(\sum_j W_j(A+V_j)^{-1}\mathbf{Y}_j\Big). \tag{17}$$

The result of this calculation is therefore that the integrated or marginal likelihood is given by (15), with the general residual sum of squares Q defined by (16) and (17).

We note a some side consequences of this calculation:

1. The conditional distribution of $\boldsymbol{\gamma}$ given A is normal with mean $\boldsymbol{\gamma}^*$ and covariance matrix $\Big(\sum_j W_j(A+V_j)^{-1}W_j^T\Big)^{-1}$;

2. The conditional distribution of $\boldsymbol{\theta}_j$ given A and $\boldsymbol{\gamma}$ is normal with mean $\Big(A^{-1}+V_j^{-1}\Big)^{-1}\Big(A^{-1}W_j^T\boldsymbol{\gamma}+V_j^{-1}\mathbf{Y}_j\Big)$ and covariance matrix $\Big(A^{-1}+V_j^{-1}\Big)^{-1}$.

3. As a result of the two preceding points, the conditional distribution of $\boldsymbol{\theta}_j$ given A and is normal with mean $\Big(A^{-1}+V_j^{-1}\Big)^{-1}\Big(A^{-1}W_j^T\boldsymbol{\gamma}^*+V_j^{-1}\mathbf{Y}_j\Big)$ and covariance matrix $(A^{-1}+V_j)^{-1}A^{-1}W_j^T\Big(\sum_j W_j(A+V_j)^{-1}W_j^T\Big)^{-1}W_jA^{-1}(A^{-1}+V_j)^{-1}+\Big(A^{-1}+V_j^{-1}\Big)^{-1}$.

These facts are useful in calculating the posterior distributions of $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}_j$ respectively, which are also important for applications.

The final expression (15) is the *integrated likelihood* (also called the *restricted likelihood*) of A, and a simple REML estimator would choose A to maximize (15). Note that when $p=1$, A is a

scalar and this is a simple one-parameter optimization. Also up to this point, the calculation is completely general, making no restrictions on $V_1, ..., V_J$ or the form of W_j.

We now show how to simplify (15) when $V_1 = ... = V_J = V_0$ and W_j is of the form (4). Write $B_0 = V_0^{1/2}(V_0 + A)^{-1}V_0^{1/2}$. Then $A(A^{-1} + V_0^{-1}) = (V_0 + A)V_0^{-1} = V_0^{1/2}B_0^{-1}V_0^{-1/2}$ so $|A| \cdot |A^{-1} + V_0^{-1}| = |B_0|^{-1}$. Hence

$$|A|^{-J/2}\Big(\prod_{j=1}^{J} |A^{-1} + V_j^{-1}|\Big)^{-1/2} \quad = \quad |B_0|^{J/2}. \tag{18}$$

Next, we note some simple facts about *Kronecker products*. If A is a $p_1 \times p_2$ matrix and B is a $q_1 \times q_2$ matrix, and $A \otimes B$ is a $p_1q_1 \times p_2q_2$ matrix whose generic entry is $a_{ij}b_{k\ell},\ 1 \le i \le p_1,\ 1 \le j \le p_2,\ 1 \le k \le q_1,\ 1 \le \ell \le q_2$. Then,

1. If A, B, C, D are matrices such that the matrix products AC and BD are both well defined, then the matrix product of $A \otimes B$ and $C \otimes D$ is $AC \otimes BD$.
2. In particular, if $p_1 = p_2$ and $q_1 = q_2$, then the inverse of $A \otimes B$ is $A^{-1} \otimes B^{-1}$.
3. Also if $p_1 = p_2 = p$ and $q_1 = q_2 = q$, $|A \otimes B| = |A|^q|B|^p$.

We note that any matrix $A = A \otimes \mathbf{1}$ where $\mathbf{1}$ is the 1×1 matrix 1; also as a result of (4), $W_j = I_p \otimes \mathbf{w}_j$. The following statements (following Everson and Morris) are mainly repeated application of property 1 above (property 3 in the case of (21)):

$$\begin{aligned}
\sum_j W_j(V_0 + A)^{-1}\mathbf{Y}_j &= \sum_j (I_p \otimes \mathbf{w}_j)\{(V_0 + A)^{-1}\mathbf{Y}_j \otimes \mathbf{1}\} \\
&= \sum_j \{(V_0 + A)^{-1}\mathbf{Y}_j \otimes \mathbf{w}_j\}, \qquad (19) \\
\sum_j W_j(V_0 + A)^{-1}W_j^T &= \sum_j (I_p \otimes \mathbf{w}_j)\{(V_0 + A)^{-1} \otimes \mathbf{1}\}(I_p \otimes \mathbf{w}_j^T) \\
&= \sum_j \{(V_0 + A)^{-1} \otimes \mathbf{w}_j\mathbf{w}_j^T\}, \\
&= (V_0 + A)^{-1} \otimes \sum_j \mathbf{w}_j\mathbf{w}_j^T, \qquad (20) \\
|\sum_j W_j(V_0 + A)^{-1}W_j^T| &= |V_0 + A|^{-q}|\sum_j \mathbf{w}_j\mathbf{w}_j^T|^p. \qquad (21)
\end{aligned}$$

As a consequence of (21),

$$|\sum_j W_j(V_0 + A)^{-1}W_j^T|^{-1/2} \propto |B_0|^{-q/2}. \tag{22}$$

We also have

$$\begin{aligned}
\boldsymbol{\gamma}^* &= \left(\sum_j W_j(A + V_j)^{-1}W_j^T\right)^{-1}\left(\sum_j W_j(A + V_j)^{-1}\mathbf{Y}_j\right) \\
&= \left\{(V_0 + A) \otimes (\sum_j \mathbf{w}_j\mathbf{w}_j^T)^{-1}\right\}\sum_j \left\{(V_0 + A)^{-1}\mathbf{Y}_j \otimes \mathbf{w}_j\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_j \left\{ (V_0 + A) \otimes (\sum_j \mathbf{w}_j \mathbf{w}_j^T)^{-1} \right\} \left\{ (V_0 + A)^{-1} \mathbf{Y}_j \otimes \mathbf{w}_j \right\} \\
&= \sum_j \left\{ \mathbf{Y}_j \otimes (\sum_j \mathbf{w}_j \mathbf{w}_j^T)^{-1} \mathbf{w}_j \right\} \\
&= \sum_j \left\{ I_p \otimes (\sum_j \mathbf{w}_j \mathbf{w}_j^T)^{-1} \right\} (\mathbf{Y} \otimes \mathbf{w}_j) \\
&= \left\{ I_p \otimes (\sum_j \mathbf{w}_j \mathbf{w}_j^T)^{-1} \right\} \sum_j (\mathbf{Y} \otimes \mathbf{w}_j) \\
&= \left\{ \sum_j (I_p \otimes \mathbf{w}_j)(I_p \otimes \mathbf{w}_j^T) \right\}^{-1} \sum_j (I_p \otimes \mathbf{w}_j)(\mathbf{Y}_j \otimes \mathbf{1}) \\
&= \left(\sum_j W_j W_j^T \right)^{-1} \sum_j W_j \mathbf{Y}_j \\
&= \hat{\boldsymbol{\gamma}}. \qquad (23)
\end{aligned}
$$

Hence with Q defined by (16),

$$
\begin{aligned}
Q &= \sum_j (\mathbf{Y}_j - W_j^T \hat{\boldsymbol{\gamma}})^T (A + V_0)^{-1} (\mathbf{Y}_j - W_j^T \hat{\boldsymbol{\gamma}}) \\
&= \sum_j (\mathbf{Y}_j - W_j^T \hat{\boldsymbol{\gamma}})^T V_0^{-1/2} B_0 V_0^{-1/2} (\mathbf{Y}_j - W_j^T \hat{\boldsymbol{\gamma}}) \\
&= \mathrm{tr}\,(S B_0)) \,. \qquad (24)
\end{aligned}
$$

Combining (18), (22) and (24), (15) reduces to (6).

3.4 Choosing the Prior

The prior density (10) is proper if $\nu_* > 0$ and $S_* > 0$. However, it's possible to choose an improper prior so long as the posterior density is proper, which in view of (11), is true so long as $J - q + \nu_* > 0,\ S + S_* > 0$. So, for example, taking $\nu = -p - 1,\ S_* = 0$ is satisfactory so long as $J > q + p + 1$. This corresponds to a prior density for B_0 proportional to $|B_0|^{-p-1}$ which, on transforming back to A, is a uniform prior on $A > 0$.

The program `tlnise` is available as a downloadable package for R (written by Roger Peng of Johns Hopkins, based on earlier S-PLUS code of Philip Everson). This program assumes a prior for B_0 proportional to $|B_0|^{(\mathrm{prior}-p-1)/2}$ where "prior" may be specified as an argument. So prior=$-p-1$ (the default) corresponds to a uniform prior on $A > 0$. But prior=0 is the Jeffreys prior, and prior=$p+1$ would correspond to a uniform prior on B_0. So far as we know, there is no general theory to determine which of these is best.

3.5 Application to the NMMAPS ozone example

Assume a file `out1.txt` that contains the output of the individual model fits. In the form adopted here, this has four columns: Cols. 1 and 2 for the estimate and standard error of the ozone-mortality coefficient under the "all-year" model, and cols. 3 and 4 the same thing under the "summer only"

model. Note that some cities have missing data and these are denoted `NA`. The first few lines of the file are

```
-0.0001153294 0.001479885 -0.0001153294 0.001479885
0.0006965119 0.001674628 0.0009891532 0.001928025
NA NA NA NA
0.002951046 0.003074613 0.002951046 0.003074613
```

Units here are that this is the increase in $\log \mu_t$ (logarithm of the expected number of deaths on day t) associated with a 1 part per billion (ppb) rise in ozone. In practice, it is usual to express the result as percent rise in mortality associated with a 10 ppb rise in ozone. To convert to this scale, the above numbers must be multiplied by 1000.

A sample program is as follows:

```
out1<-matrix(scan('out1.txt'),ncol=4,byrow=T)
m1<-1000*out1[,1]
v1<-(1000*out1[,2])^2
# remove NAs
m1<-m1[!is.na(v1)]
v1<-v1[!is.na(v1)]
# apply tlnise - remember to load first
#
# there are a bunch of options but this is the basic form of the command
tln1<-tlnise(m1,v1)
#
# output "national" estimates
tln1$gamma
# output individual city posterior estimates and standard errors
tln1$theta
tln1$SDtheta
```

However, the original NMMAPS cities were also classified into seven regions, denoted Industrial Midwest, North East, North West, Southern California, South East (includes eastern Texas), South West and Upper Midwest. We can compute stratified estimates by region, using the following code:

```
# load regions data and convert to indicator variables
region<-scan('d:/r/c/props/API/work4/regions.txt')
regind<-matrix(0,ncol=7,nrow=108)
for(i in 1:108){
if(region[i]==1)regind[i,]<-c(1,0,0,0,0,0,0)
if(region[i]==2)regind[i,]<-c(0,1,0,0,0,0,0)
if(region[i]==3)regind[i,]<-c(0,0,1,0,0,0,0)
if(region[i]==5)regind[i,]<-c(0,0,0,1,0,0,0)
if(region[i]==6)regind[i,]<-c(0,0,0,0,1,0,0)
if(region[i]==7)regind[i,]<-c(0,0,0,0,0,1,0)
if(region[i]==8)regind[i,]<-c(0,0,0,0,0,0,1)
}
# load ozone-mortality coefficients
out1<-matrix(scan('d:/feb08/UNC/s890/nmmaps/out1.txt'),ncol=4,byrow=T)
```

```
m1<-1000*out1[,1]
v1<-(1000*out1[,2])^2
# remove NAs
regind<-regind[!is.na(v1),]
m1<-m1[!is.na(v1)]
v1<-v1[!is.na(v1)]
# apply tlnise - remember to load first
tln1<-tlnise(m1,v1,w=regind,intercept=F)
#
# output "national" estimates
tln1$gamma
# output individual city posterior estimates and standard errors
tln1$theta
tln1$SDtheta
```

As an example, Fig. 1 shows the posterior estimates of θ_j and 95% prediction intervals under the "national prior", in which we assume $\theta_j \sim N[\mu, \tau^2]$ for each city j under a common prior mean μ and variance τ^2 for all cities. Also shown on this figure are the posterior means (squares) from estimates computed under the "regional prior", in which $\theta_j \sim N[\sum_{r=1}^{7} \beta_r I_{r,j}, \tau^2]$, $\beta_1, ..., \beta_7$ having independent uniform priors on $(-\infty, \infty)$, $I_{r,j}$ denoting the regional indicator variable ($I_{r,j} = 1$ if city j is in region r, 0 otherwise).

Fig. 2 depicts the "population weighted average coefficient" for each region, with 95% prediction intervals.

Fig. 3 depicts various versions of a spatially interpolated reconstruction of the ozone-mortality coefficient. This uses spatial statistics and will be explained later in the course!

4 A Measurement Error Model for the Relationship Between Heart Rate and Exposure to Particulate Matter

4.1 Introduction

This section is based on a paper (Crooks *et al.*, 2008) that studied the relationship between ambient particulate matter (PM) and various measures of cardiovasular health. In particular, we consider whether exposure to PM has the effect of decreasing the RR interval (in effect, the reciprocal of heart rate). Decreases in RR (ms) or increases in HR (beat/sec) are associated with heightened activity of the sympathetic division of the autonomic nervous system, and this in turn is a risk factor for acute coronary heart disease (CHD) events. The data analysis was conducted as part of an ancillary study based on a large, geographically diverse population of U.S. women enrolled in the Women's Health Initiative (WHI) clinical trials, The Environmental Epidemiology of Arrhythmogenesis in WHI (EEAWHI).

The original dataset included 68,132 individuals studied between 1993 and 1998. At each study visit, an individuals RR interval (as well as a number of other measures of cardiovascular health, but we focus on RR here) were measured from an electrocardiograph. After applying various exclusion criteria, the main analysis was based on the first ECG recorded during the study period among 52,805 participants, grouped into 57 study centers.

Various individual attributed were included in the dataset, such as age, self-reported ethnicity (White/Non-Hispanic, Black/African-American, Hispanic/Latino, Asian/Pacific Islander or other);

Variable	Lag(s)	Estimate ($\times 10^4$)	SE ($\times 10^4$)	t statistic	2-sided p-value
PM_{10}	0	–1.58	0.55	–2.89	0.0039
PM_{10}	1	–1.51	0.57	–2.65	0.0080
PM_{10}	2	–0.26	0.60	–0.43	0.66
PM_{10}	3	0.43	0.59	0.72	0.47
PM_{10}	4	0.36	0.57	0.63	0.52
PM_{10}	5	0.26	0.55	0.47	0.64
PM_{10}	01	–2.06	0.64	–3.20	0.0014

Table 1: Simple linear regression models for 6 particulate matter measures.

education (whether or not a college graduate); indicators of diabetes, hypertension, etc.; whether or not the individual is a smoker; whether or not the individual uses beta-blockers; and chronic lung disease as measured by history of asthma, emphysema or lung cancer.

Meteorological data were collected; temperature at lag 1 day was included as a covariate in the regression analyses.

The principal novel feature of the dataset was that exposures to PM were not measured directly, but interpolated from monitor stations by kriging (to be described in more detail later in this course!). A previous paper (Liao *et al.*, 2006) established that the most efficient method of kriging was *log-normal kriging*, in which it is assumed that the *logarithm* of particulate matter has a normal distribution. This use of a lognormal distribution leads to some novel results for the effect of measurement error on the regression. For simplicity, in this discussion we consider only the concentrations of PM < 10 μm (PM_{10}) averaged over the day of and before each ECG (lag 0-1).

4.2 Simple Regression Analyses

As a preliminary analysis, we performed a simple regression. We used logarithm of RR as the response variable, and the following covariates: exam site as a factor variable with 57 levels; temperature; time of day, day of week, and season of ECG recording; plus each of the participants characteristics listed in Section 4.1.

In addition to the above, we included (one at a time) concentrations of PM_{10} at each of lags 0-5 days, as well as the average of lags 0 and 1. Results are in Table 1. From this we conclude that the strongest effects are at lags 0 and 1, with the strongest of all at the average of lags 0 and 1.

A second step was to include possible interactions between PM_{10} and participant characteristics. In particular, we divided the sample according to the presence of prior lung disease ("LD"), according to smoking status ("SM" = current smoker) and beta-blocker use ("BB") as well as various combinations of these variables. In all cases, the PM variable was taken to be the mean of PM_{10} at lags 0 and 1. Results are in Table 2.

The results showed that PM significantly affects RR only among non-smokers without chronic lung disease. This suggested defining two risk groups, one consisting of non-smokers without chronic lung disease ("No SM, No LD") and the other of everyone else ("SM or LD"). To examine the effect of beta-blocker use, each of the prior groups was split into two further subgroups, labeled "No BB" or "BB". Within the "No SM, No LD" group where the effect of PM is significant, this further subdivision into BB and no BB showed about the same level of statistical significance as

Subgroup	Percent of total	Estimate ($\times 10^4$)	SE ($\times 10^4$)	t statistic	2-sided p-value
No LD	89.8	–2.41	0.67	–3.58	0.00034
LD	10.2	0.96	1.81	0.53	0.60
No BB	87.2	–1.72	0.68	–2.53	0.011
BB	12.8	–4.45	1.65	–2.70	0.0069
No SM	94.0	–2.29	0.66	–3.48	0.00051
SM	6.0	1.60	2.37	0.67	0.50
NoBB NoLD	77.9	–2.02	0.71	–2.85	0.0043
NoBB, LD	9.3	0.84	1.83	0.46	0.65
BB, NoLD	11.9	–5.00	1.66	–3.01	0.0026
BB, LD	1.0	2.05	3.05	0.67	0.50
NoBB, NoSM	81.8	–1.92	0.69	–2.76	0.0058
NoBB, SM	5.40	1.28	2.40	0.53	0.59
BB, NoSM	12.3	–4.91	1.66	–2.96	0.0031
BB, SM	0.6	3.63	3.71	0.98	0.33
NoSM, NoLD	84.6	–2.62	0.69	–3.83	0.00013
NoSM, LD	9.5	0.75	1.84	0.41	0.68
SM, NoLD	5.2	1.33	2.42	0.55	0.58
SM, LD	0.7	3.31	3.41	0.97	0.33
NoBB,NoSM,NoLD	73.1	–2.12	0.71	–2.98	0.0029
NoBB,SMorLD	14.0	0.48	1.37	0.35	0.73
BB,NoSM,NoLD	11.4	–5.35	1.67	–3.21	0.0013
BB,SMorLD	1.4	1.99	2.54	0.78	0.43

Table 2: Regression analysis in which the population is split into subgroups in various ways. Variable of interest is PM_{10}, mean of lags 0 and 1.

Subgroup	Pollutant	Posterior Mean ($\times 10^4$)	95% CI ($\times 10^4$)
NoBB,NoSM,NoLD	PM_{10}	–1.77	(–3.40,–0.14)
NoBB,SMorLD	PM_{10}	0.40	(–3.25,4.05)
BB,NoSM,NoLD	PM_{10}	–5.63	(–9.80,–1.46)
BB,SMorLD	PM_{10}	2.08	(–3.33,7.49)

Table 3: Posterior means and 95% credible intervals (CIs) for combined coefficient of PM_{10}, subdivided into four groups of participants, based on hierarchical analysis across sites but without taking into account PM measurement error.

measured by the t statistic, but a much larger magnitude of regression coefficient in the BB group, suggesting that PM has a greater effect among beta-blocker users than non-users.

For subsequent analysis, we concentrate on the split defined by the last box in Table 2, dividing the population first into two groups, one consisting of non-smokers without chronic lung disease, and the other consisting of everyone else. Then, each of those subgroups was further split according to beta-blockers use.

4.3 NMMAPS-Style Analysis

AS a first step towards generalizing the simple regression analysis, we considered an "NMMAPS-style analysis" in which the same regression model was fitted *separately* to data from each of the 57 exam sites, with individual counts ranging from 75 to 1972. For each site the coefficient and standard error corresponding to the particulate matter effect was retained. The results were combined across sites using the `tlnise` program. The analysis was restricted to the data split mentioned at the end of Section 4.2 (based on chronic lung disease, current smoking status and betablocker use), and to PM concentrations averaged across lags 0 and 1. Results are shown in Table 3.

The results reinforce the findings of Section 4.2. The PM-RR associations are statistically significant only among the non-smokers without chronic lung disease, but among that group, the effect is much stronger among beta-blocker users than among non-users.

4.4 Fully Bayesian Analysis

This section presents a fully Bayesian analysis, combining data across sites as in Section 4.3, but also including the effects of measurement error in PM_{10}. The analysis relies on a Bayesian hierarchical model, using Gibbs sampling to update the regression parameters and a Metropolis-Hastings update for the particulate matter variable.

Let y_{ij} denote the observed response of individual j at exam site i, and let $\{x_{ijk}\}$ denote the individual-level covariates. We assume site-level regression coefficients $\{\beta_{ik}\}$ and precisions $\{\kappa_i\}$ to have prior normal and gamma distributions respectively, where the prior mean of β_{ik} depends on site-level covariates $\{z_{it},\ t = 1, ..., T\}$. Parameters α_{kt} and ψ_k define the prior distributions of $\{\beta_{ik}\}$. This is the general formulation of the model, that allows for both site-level and individual-level covariates, but for the applications in this paper, we take $T = 1$, $z_{i1} = 1$ for all i, and write α_k instead of $\alpha_{kt},\ t = 1, ..., T$. We write P_{ij} for the true value of the pollution variable (PM_{10} or $PM_{2.5}$) for individual j in site i, and we assume that the log-normal kriging analysis specifies the

prior mean and standard deviation, V_{ij} and s_{ij}, of P_{ij}; the standard deviation was converted back to a logarithmic scale by a delta-function argument. In cases where the coefficient of P_{ij} is the same for everyone in the population, we just take $x_{ij1} = P_{ij}$ and then α_1 is the coefficient. For subgroup analyses (e.g. four subgroups, group 1 consisting of individuals who do not take beta-blockers and do not have chronic lung disease or smoking, etc.) we define $x_{ijk} = P_{ij}\delta_{ijk},\ k = 1, ..., 4$ where δ_{ijk} is the indicator for individual j at site i to be in subgroup k, and in that case α_k for $k = 1, ..., 4$ is the regression coefficient of P_{ij} in subgroup k. The full model is

$$\begin{aligned}
y_{ij} &\sim N\left[\sum_k x_{ijk}\beta_{ik}, \kappa_i^{-1}\right], \quad j = 1, ..., n_i,\ i = 1, ..., C, \\
\kappa_i &\sim \Gamma[\gamma_\kappa, \delta_\kappa],\ i = 1, ..., C, \\
\beta_{ik} &\sim N\left[\sum_{t=1}^{T} z_{it}\alpha_{kt}, \psi_k^{-1}\right], \quad k = 1, ..., K, \\
\psi_k &\sim \Gamma[\gamma_\psi, \delta_\psi],\ k = 1, ..., K, \\
\alpha_{kt} &\sim U[-\infty, \infty],\ k = 1, ..., K,\ t = 1, ..., T, \\
\gamma_\kappa, \delta_\kappa &\sim \Gamma[a_0, b_0], \quad (\text{e.g. } a_0 = b_0 = 0.001) \\
\log P_{ij} &\sim N\left[\log V_{ij}, u_{ij}\right] \quad (u_{ij} \text{ known}, = \frac{s_{ij}^2}{V_{ij}^2}),\ j = 1, ..., n_i,\ i = 1, ..., C.
\end{aligned}$$

The joint density of all observations is:

$$\begin{aligned}
&\prod_{i=1}^{C}\left[\kappa_i^{n_i/2}\exp\left\{-\frac{\kappa_i}{2}\sum_j\left(y_{ij} - \sum_k x_{ijk}\beta_{ik}\right)^2\right\}\cdot\left\{\frac{\delta_\kappa^{\gamma_\kappa}}{\Gamma(\gamma_\kappa)}\kappa_i^{\gamma_\kappa - 1}e^{-\delta_\kappa\kappa_i}\right\}\cdot\right. \\
&\cdot\prod_{k=1}^{K}\left\{\psi_k^{1/2}\exp\left(-\frac{\psi_k}{2}\left(\beta_{ik} - \sum_t z_{it}\alpha_{kt}\right)^2\right)\right\}\cdot \\
&\left.\cdot\prod_{j=1}^{n_i}\left\{\frac{1}{P_{ij}}\exp\left(-\frac{1}{2u_{ij}}(\log P_{ij} - \log V_{ij})^2\right)\right\}\right]\cdot \\
&\cdot\gamma_\kappa^{a_0 - 1}e^{-b_0\gamma_\kappa}\cdot\delta_\kappa^{a_0 - 1}e^{-b_0\delta_\kappa}.
\end{aligned}$$

Define $\boldsymbol{\alpha}_k$ to be the vector of $\alpha_{kt},\ t = 1, ..., T$, $\boldsymbol{\beta}_{i\cdot}$ to be the vector of $\beta_{ik},\ k = 1, ..., K$ for fixed i, $\boldsymbol{\beta}_{\cdot k}$ to be the vector of $\beta_{ik},\ i = 1, ..., C$ for fixed k, $\mathbf{z}_i$ to be the vector of $z_{it},\ t = 1, ..., T$, and $\mathbf{x}_{ij}$ to be the vector of $x_{ijk},\ k = 1, ..., K$ for fixed i and j. All these are column vectors. Also let Z be the matrix of z_{it} $(C \times T)$, A be the matrix of α_{kt} $(K \times T)$, and let Ψ be the $K \times K$ diagonal matrix with diagonal entries $(\psi_1, ..., \psi_K)$.

The conditional distributions are:

$$\begin{aligned}
\boldsymbol{\beta}_{i\cdot} \mid \text{rest} &\sim N\left[\left(\kappa_i\sum_j \mathbf{x}_{ij}\mathbf{x}_{ij}^T + \Psi\right)^{-1}\left(\kappa_i\sum_j \mathbf{x}_{ij}y_{ij} + \Psi A\mathbf{z}_i\right), \left(\kappa_i\sum_j \mathbf{x}_{ij}\mathbf{x}_{ij}^T + \Psi\right)^{-1}\right], \\
\boldsymbol{\alpha}_k \mid \text{rest} &\sim N\left[(Z^TZ)^{-1}Z^T\boldsymbol{\beta}_{\cdot k}, (\psi_k Z^TZ)^{-1}\right], \\
\kappa_i \mid \text{rest} &\sim \Gamma\left[\gamma_\kappa + \frac{n_i}{2}, \delta_\kappa + \frac{1}{2}\sum_j\left(y_{ij} - \sum_k x_{ijk}\beta_{ik}\right)^2\right],
\end{aligned}$$

$$\psi_k \mid \text{rest} \sim \Gamma\left[\gamma_\psi + \frac{C}{2}, \delta_\psi + \frac{1}{2}\sum_i \left(\beta_{ik} - \sum_t z_{it}\alpha_{kt}\right)^2\right],$$

$$\delta_\kappa \mid \text{rest} \sim \Gamma\left(a_0 + C\gamma_\kappa, b_0 + \sum_i \kappa_i\right).$$

This covers the "Gibbs sampling" part of the solution. The other elements must be updated by Metropolis sampling. In the case of P_{ij}, we propose the following algorithm. First note that in the subgroups analysis with four subgroups, each P_{ij} is one of $x_{ijk},\ k = 1, ..., 4$ depending on which subgroup contains individual j of site i; without loss of generality, let us assume $P_{ij} = x_{ij1}$ for a particular (i, j). Then

$$f(x_{ij1}) = \exp\left\{-\frac{\kappa_i}{2}\left(y_{ij} - \sum_k x_{ijk}\beta_{ik}\right)^2\right\} \cdot \exp\left\{-\frac{1}{2u_{ij}}(\log x_{ij1} - \log V_{ij})^2\right\}.$$

At step (i, j) define x'_{ij1} by replacing x_{ij1} with $x'_{ij1} = x_{ij1}e^{\Delta(U-\frac{1}{2})}$ where U is uniform on $[0, 1]$ and Δ is arbitrary; accept x'_{ij1} with probability $\min\left\{\frac{f(x'_{ij1})}{f(x_{ij1})}, 1\right\}$, otherwise keep x_{ij1} at its present value until the next iteration.

For γ_κ, define

$$g(\gamma_\kappa) = \frac{\left(\prod_i \kappa_i\right)^{\gamma_\kappa} \delta_\kappa^{C\gamma_\kappa}}{\Gamma(\gamma_\kappa)^C} \cdot \gamma_\kappa^{a_0} e^{-b_0\gamma_\kappa}.$$

Based on current γ_κ define new $\gamma'_\kappa = \gamma_\kappa e^{\Delta'(U-\frac{1}{2})}$ where U is uniform on $[0, 1]$ and Δ' is arbitrary; accept γ'_κ with probability $\min\left\{\frac{g(\gamma'_\kappa)}{g(\gamma_\kappa)}, 1\right\}$.

We could treat γ_ψ, δ_ψ in a similar manner to $\gamma_\kappa, \delta_\kappa$ (i.e. defining a prior distribution using hyperparameters) but we prefer not to for the following reason: the κ_i are exchangeable (they represent equivalent measurements taken in different exam sites, but we believe the sites are similar) so it makes sense to estimate a distribution across sites that could, for example, be used to predict responses at a new site should one ever be added to the dataset. The same argument does not apply to the ψ_k parameters, which represent qualitatively different covariates. It is doubtful that this distinction has much impact on the results, but we did find in preliminary analysis that some care was needed in handling γ_ψ and δ_ψ. For the following analysis we took $\gamma_\psi = 0.01, \delta_\psi = 10^{-6}$. The value of γ_ψ ensures that the prior distribution is highly diffuse, while our choice of δ_ψ then ensures that the prior mean of the ψ_k is about 10^4. That would fit in with the fact that the standard errors of the α parameters, estimated through a conventional regression approach, are of the order of 0.01 $(= 1/\sqrt{10^4})$.

The values of Δ and Δ' are arbitrary but we took $\Delta = 5$ for the measurement error analysis with $M = 1$ and $\Delta = 10$ for the measurement error analysis with $M = 2$. These choices were guided by the criteria of Gelman *et al.* (1996) for optimal acceptance rates in Hasting-Metropolis sampling. We took $\Delta' = 1$.

The Bayesian hierarchical analysis was run using the average for lags 0 and 1 of either PM_{10} or $PM_{2.5}$, with a multiplier $M = 0$, 1 or 2 applied to the kriging error. Since the concentrations were kriged on a daily basis we calculated the error for this average using the kriging errors from lags 0 and 1 and the lag-1 autocorrellation of PM_{10} (0.57) and $PM_{2.5}$ (0.67) calculated from our

		$M = 0$		$M = 1$		$M = 2$	
Subgroup	PM	Mean ($\times 10^4$)	95% CI ($\times 10^4$)	Mean ($\times 10^4$)	95% CI ($\times 10^4$)	Mean ($\times 10^4$)	95% CI ($\times 10^4$)
NoBB,NoSM,NoLD	10	–2.15	(–3.93,–0.37)	–1.41	(–3.03,0.14)	–0.33	(–1.29,0.61)
NoBB,SMorLD	10	0.12	(–2.88,3.02)	–0.44	(–2.98,2.20)	0.15	(–1.35,1.56)
BB,NoSM,NoLD	10	–5.48	(–9.13,–1.96)	–5.72	(–9.45,–1.92)	–2.34	(–4.17,–0.62)
BB,SMorLD	10	1.83	(–3.75,7.56)	–0.54	(–6.30,5.73)	0.62	(–2.76,3.95)

Table 4: Posterior means and 95% credible intervals, for combined coefficient of mean PM_{10} at lag 0-1, subdivided into four groups of participants, based on fully Bayesian model for cases $M = 0$, $M = 1$, $M = 2$.

Variable	M	NoBB,NoSM,NoLD ($k = 1$)	NoBB,SMorLD ($k = 2$)	BB,NoSM,NoLD ($k = 3$)	BB,SMorLD ($k = 4$)
PM_{10}	0	0.991	0.465	0.999	0.266
PM_{10}	1	0.962	0.637	0.999	0.580
PM_{10}	2	0.750	0.409	0.997	0.354

Table 5: Posterior probabilities of $\alpha_k < 0$, $k = 1, 2, 3, 4$, for PM_{10} and three noise multiplies ($M = 0$, 1 and 2). Based on 50000 iterations, the first 20000 iterations discarded as burn-in.

data. The introduction of M is intended to allow examination of different levels of measurement error. $M = 0$ ignores the measurement error and gives results fairly comparable to Section 4.3 (see Table 4). $M = 1$ is the case of primary interest, while $M = 2$ is included to allow examination of the effect of under-estimating the true PM measurement error. As before, the analysis included the participant characteristics listed in Section 4.1.

The participants were divided into the same four subgroups, as discussed in Section 4.2, based on chronic lung disease, current smoking status and beta-blocker use. As defined earlier, α_k is the regression coefficient of PM_{10} or $PM_{2.5}$ on log RR in subgroup k, $k = 1, 2, 3, 4$, where, for example, α_1 refers to the "no BB, no SM, no LD" subgroup. For each analysis, a total of 50,000 MCMC iterations was run, divided into 1000 loops of size 50. The first 20,000 iterations were discarded as burn-in, the rationale for this choice being discussed further in Section 4.8.

4.5 Results

Figures **??** and **??** show the posterior density estimates of α_k, $k = 1, ..., 4$. The comparison between the posterior density curves for $M = 0$ and $M = 1$ show that in all cases the density is shifted as a result of taking account of measurement error, though not necessarily towards 0. It appears that the shift in the posterior pdf as a result of taking account of measurement error is greater with PM_{10} than with $PM_{2.5}$. In all cases, however, doubling the measurement error standard deviation results in a much more marked shift, towards a posterior density that is highly peaked near 0. This feature of the results was initially unexpected, but there is a natural explanation for it.

Table 4 shows the posterior means and 95% credible intervals corresponding to the posterior density curves in Figure **??**. The $M = 0$ results here should be compared with those of Table 3 and show the effect of using a fully Bayesian approach approach to the hierarchical analysis as compared with the NMMAPS approach. In Table 3 the credible intervals were calculated assuming a normal

distribution based on the posterior mean and posterior standard deviation that are produced by the `tlnise` program; in Table 4 they are based directly on the MCMC output with boundaries at the 2.5 and 97.5 percentiles of the posterior distribution. The results clearly show some differences between the two approaches, though they are not so great as to affect the epidemiological interpretation of the results. Table 5 shows the posterior probabilities of $\alpha_k < 0,\ k = 1, ..., 4$, for three values of the error multiplier M, and two pollutants. These probabilities have a similar interpretation to that of a p-value in classical statistics: a posterior probability close to 1 indicates a high level of confidence that the true value of α_k is negative. The results demonstrate that for the "No SM, No LD" groups, the posterior probability that $\alpha_k < 0$ is near 1 in every case for which $M = 0$ or $M = 1$; in fact, with one exception (NoBB,NoSM,NoLD; PM_{10}) the same is also true when $M = 2$. We therefore see that for this dataset, the inference that there is an inverse PM-RR association is quite robust against measurement error.

4.6 Subgroup Differences

Though we originally split the study population into four subgroups based on scientific plausibility it is worth revisiting whether the differences are borne out in the analysis. In a frequentist context, a test of the null hypothesis $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$ would be interpreted as a test of interaction between subgroup number and PM. In the present Bayesian context, if at least one of the posterior probabilities that $\alpha_{k_1} - \alpha_{k_2} < 0$ (for difference $k_1, k_2 \in \{1, 2, 3, 4\}$) is very close to 0 or 1, we conclude that interactions exist.

For example, our results suggest that α_1 and α_3 are both negative but $\alpha_3 < \alpha_1$, implying that PM has a greater relative effect on beta-blocker users than non-users among non-smokers who do not have chronic lung disease. In fact, the posterior probability that $\alpha_3 < \alpha_1$ is 0.96, 0.98, and 0.98 for $M = 0, 1, 2$ respectively. In all cases there appears to be strong evidence that $\alpha_3 < \alpha_1$, but interestingly, the evidence is stronger when measurement error is explicitly modeled ($M = 1$ or 2) that when it is ignored ($M = 0$).

On the other hand, among participants who do not use beta-blockers, the posterior probabilities that the PM effect is stronger for non-smokers who do not have chronic lung disease than for others are 0.92, 0.75, and 0.72 for PM_{10}. In this case there appears to be strong evidence for a difference between groups only if kriging error is ignored.

4.7 Rao-Blackwellization

Recall one of the conditional distributions used in the Gibbs sampler:

$$\boldsymbol{\alpha}_k \mid \text{rest} \quad \sim \quad N\left[(Z^T Z)^{-1} Z^T \boldsymbol{\beta}_{\cdot k}, (\psi_k Z^T Z)^{-1}\right]. \tag{25}$$

As written here, the conditional distribution is for a vector $\boldsymbol{\alpha}_k$ but in practice our main interest has been in scalar parameters $\alpha_1, ..., \alpha_4$, which are a subset of (25).

Suppose we want to find the posterior mean of α_1. Two possible approaches are:

1. Generate a Monte Carlo sample of size B, say $\alpha_1^{(b)},\ 1 \le b \le B$, then calculate $\frac{1}{B}\sum_{b=1}^{B} \alpha_1^{(b)}$, as our estimator of the posterior mean of α_1.

2. Generate a random sample of $\boldsymbol{\beta}_{\cdot k}$, say $\boldsymbol{\beta}_{\cdot k}^{(b)},\ 1 \le b \le B$; calculate $\frac{1}{B}\sum_{b=1}^{B} (Z^T Z)^{-1} Z^T \boldsymbol{\beta}_{\cdot k}^{(b)}$.

Because of (25), these two procedures lead to the same result in large samples. However, method 2 is more efficient, essentially because whenever X and Y are jointly distributed random variables with finite variances, $\mathrm{Var}(\mathrm{E}(X|Y)) < \mathrm{Var}(X)$.

This trick is known as *Rao-Blackwellization*, essentially because the Rao-Blackwell theorem of classical statistical inference relies on the same inequality.

Rao-Blackwellization is also useful for displaying posterior densities. Suppose we want to show the posterior density of α_1. One approach is to generate a Monte Carlo sample $\alpha_1^{(b)},\ 1 \le b \le B$ and construct a density estimate, e.g. using a kernel density estimate. However that has various disadvantages, including the need to select a kernel bandwidth. Rao-Blackwellization provides an alternative approach: for each $b \in \{1, 2, ..., B\}$, construct the normal density based on (25) (based on $\boldsymbol{\beta}_{\cdot k}^{(b)}$ and $\psi_k^{(b)}$). The final plotted posterior density is the average over these B conditional normal density curves.

4.8 Convergence Diagnostics for MCMC

The package CODA (Plummer *et al.* 2006) is a package of programs written in R, designed for post-processing of MCMC output. For this discussion, we focus on two of the procedures included as part of the package: the Gelman-Rubin procedure of monitoring convergence, and the Heidelberger-Welch method for estimating confidence intervals from an autocorrelated time series.

The Gelman-Rubin procedure was originally proposed by Gelman and Rubin (1992), and extended to multivariate distributions by Brooks and Gelman (1998). As described by Gelman and Rubin (1992), it consists of seven steps.

1. Simulate $m \ge 2$ sequences, each of length $2n$, with starting points drawn from a distribution that is believed to be overdispersed relative to the true posterior distribution. Discard the first n iterations of each sequence, retaining the last n.

2. For each scalar parameter of interest, calculate B/n as the variance between the m sequence means $\bar{x}_{i\cdot}$, each based on n values of x, so that $B/n = \sum_{i=1}^{m}(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})^2/(m-1)$; and also W, the average of the m within-sequence variances s_i^2, each of which has $n-1$ degrees of freedom — thus $W = \sum_{i=1}^{m} s_i^2/m$.

3. Estimate the target mean μ as the mean over mn values of x, so $\hat{\mu} = \bar{x}_{\cdot\cdot}$.

4. Estimate the target variance $\sigma^2 = \int (x-\mu)^2 P(x)dx$ by a weighted average of W and B, namely

$$\hat{\sigma}^2 \quad = \quad \frac{n-1}{n}W + \frac{1}{n}B.$$

   This will be an overestimate of σ^2 if the starting distribution is appropriately overdispersed, but is unbiased for σ^2 if the starting distribution is the true stationary distribution. Also, for finite n, W should underestimate σ^2, because the individual Markov chains have not had time to range over the full stationary distribution; however as $n \to \infty$, both $\hat{\sigma}^2$ and W should converge to σ^2.

5. Calculate the target distribution with mean $\hat{\mu}$, scale $\sqrt{\hat{V}} = \sqrt{\hat{\sigma}^2 + B/(mn)}$ and degrees of freedom $\mathrm{df} = 2\hat{V}^2/\hat{\mathrm{var}}(\hat{V})$ where

$$\hat{\mathrm{var}}(\hat{V}) \quad = \quad \left(\frac{n-1}{n}\right)^2 \cdot \frac{1}{m}\hat{\mathrm{var}}(s_i^2) + \left(\frac{m+1}{mn}\right)^2 \cdot \frac{2}{m-1}B^2$$

$$+2\frac{(m+1)(n-1)}{mn^2}\cdot\frac{n}{m}[\hat{\text{cov}}(s_i^2,\bar{x}_{i\cdot}^2)-2\bar{x}_{\cdot\cdot}\hat{\text{cov}}(s_i^2,\bar{x}_{i\cdot})]$$

where the estimated variances and covariances are obtained from the m sample values of $\bar{x}_{i\cdot}$ and s_i^2.

6. Calculate the "potential scale reduction" as $\sqrt{\hat{R}}=\sqrt{(\hat{V}/W)df/(df-2)}$. If $\hat{R}$ is much above 1, then we have reason to believe the inference will be improved by continuing the simulation.

7. Once $\hat{R}$ is near 1, it is desirable to summarize the target distribution by a set of simulations rather than normal-theory approximations.

To apply these methods to our present application, the preceding analysis was repeated four times, with 50,000 MCMC iterations in each trial. To save overall computing time, this was not done for all the different analyses, but only for the one that seems of greatest interest: taking PM_{10} as the pollution covariate of interest, and $M = 1$.

To achieve overdispersed initial distributions, taking into account that the greatest uncertainty seems to be in the posterior distributions of the PM_{10} values themselves, we multiplied the initial PM_{10} estimates by four multiplicative factors — 0.2, 1, 5 and 25 — as starting values for the MCMC procedure, and also varying the seed of the random number generator. In all other respects, the four MCMC simulations, of 50,000 iterations each, were identical.

The parameters of primary interest are α_k for $k = 1, 2, 3, 4$, which were generated through the sequence of conditional distributions

$$\alpha_k \mid \text{rest} \quad\sim\quad N\left[(Z^TZ)^{-1}Z^T\boldsymbol{\beta}_{\cdot k},(\psi_kZ^TZ)^{-1}\right], \tag{26}$$

where $\boldsymbol{\beta}_{\cdot k}$ is the set of $\beta_{i,k},\ i = 1, ..., C$ coefficients associated with the k'th covariate and ψ_k is the corresponding scale parameter.

The calculations were subdivided as follows:

4.8.1 Posterior mean of α_k

According to Equation (26), the conditional posterior mean of α_k, given all the values of β_{ik}, is $(Z^TZ)^{-1}Z^T\boldsymbol{\beta}_{\cdot k}$. Unconditionally, the posterior mean may be calculated by averaging this quantity over all MCMC iterations, after discarding initial burn-in iterations. Therefore, for each of $k = 1, 2, 3, 4$, $(Z^TZ)^{-1}Z^T\boldsymbol{\beta}_{\cdot k}$ was generated for each iteration of the MCMC and for each of the four replications, and subsequent analyses was based on the resulting time series.

The diagnostic of Gelman and Rubin (1992) is based on a "potential scale reduction factor" R, and the CODA package calculates both a median value and an upper confidence limit for R — in the present discussion, the confidence limit has been calculated as the 97.5% quantile of the distribution. A value of R that is much above 1 is taken to indicate non-convergence of the algorithm.

In Figure 5, we show a plot of the Gelman-Rubin diagnostic (function `gelman.plot` in CODA) as a function of the number of iterations for estimating the posterior means of each of the parameters $\alpha_k,\ k = 1, 2, 3, 4$. In each case, we note that there is a sharp reduction in R over the first 3,000 iterations and that it settles down to a value less than 1.2 by about the 15,000'th iteration. However, the convergence is not uniformly fast over all four parameters — in particular, for $k = 2, 3, 4$, it

	$k=1$	$k=2$	$k=3$	$k=4$
Run 1, Estimate	–1.45	–0.50	–5.45	–0.30
Run 2, Estimate	–1.41	–0.44	–5.72	–0.54
Run 3, Estimate	–1.36	–0.57	–5.88	–0.92
Run 4, Estimate	–1.41	–0.51	–5.70	–0.55
Run 1, 95% CI	(–1.51,–1.39)	(–0.77,–0.22)	(–5.78,–5.12)	(–0.76,0.17)
Run 2, 95% CI	(–1.48,–1.35)	(–0.65,–0.24)	(–6.11,–5.32)	(–1.11,0.03)
Run 3, 95% CI	(–1.43,–1.28)	(–0.80,–0.34)	(–6.18,–5.57)	(–1.38,–0.47)
Run 4, 95% CI	(–1.48,–1.34)	(–0.75,–0.26)	(–6.05,–5.35)	(–1.10,–0.01)

Table 6: Estimates and 95% confidence intervals for the posterior means of each of the α_k parameters, $k = 1, 2, 3, 4$, based on four parallel MCMC runs of length 50,000, discarding the first 20,000 iterations as burn-in, and using the Heidelberger-Welch procedure to take account of autocorrelation. All results have been multiplied by 10^4 for ease of numerical display.

seems to be slower than for $k = 1$. This may reflect the inherent uncertainty of these parameters (of the four subgroups of the population, group 1 is by far the largest, therefore the group about which one would expect the most precise inferences). Guided by these results, in subsequent analyses we have discarded the first 20,000 iterations as burn-in, and calculate output statistics based on iterations 20,001 through 50,000.

When reporting results from a simulation, it is conventional to calculate both the mean of the quantity of interest (in this case, the posterior mean of α_k) and a standard error. Since simulation outputs are autocorrelated, however, it is necessary to correct for the autocorrelation. The method of Heidelberger and Welch (1981) is based on a nonparametric estimate of the spectral density at zero frequency, and is implemented in CODA through the function `spectrum0`, after discarding the first 20,000 iterations as previously described. This procedure leads to the point estimates and 95% confidence intervals for the posterior means given in Table 6.

From the table, we can draw some conclusions about how accurately the posterior means have been estimated. Clearly α_1 is the best estimated, all four confidence intervals within the range –1.51 to –1.28. For α_3, the range is wider (–6.18 to –5.12) also much further from 0 — in both cases, it is clear that the posterior mean is less than 0 (consistent with an adverse effect of PM on RR). For α_2 and α_4, the confidence intervals are much wider. Indeed for α_4, we cannot even state the sign of the posterior mean.

4.8.2 Posterior probability that $\alpha_k < 0$

According to Equation (26), conditionally on ψ_k and all the values of β_{ik}, the posterior probability that $\alpha_k < 0$ is

$$\Phi\left(-\frac{(Z^TZ)^{-1}Z^T\boldsymbol{\beta}_{\cdot k}}{\sqrt{(Z^TZ)^{-1}/\psi_k}}\right) \tag{27}$$

with $\Phi(\cdot)$ the standard normal distribution function (note that Z^TZ is scalar in this instance, so Equation (27) makes sense). Therefore, we can calculate the unconditional posterior probability that $\alpha_k < 0$ by averaging Equation (27) over all MCMC runs, discarding burn-in.

For the analysis, the quantity Equation (27) was calculated for each iteration, and the resulting values analyzed using the same tests as in Section 4.8.1. The Gelman-Rubin test easily confirmed

	$k=1$	$k=2$	$k=3$	$k=4$
Run 1, Estimate	0.963	0.640	0.9958	0.542
Run 2, Estimate	0.962	0.637	0.9993	0.580
Run 3, Estimate	0.953	0.659	0.9996	0.629
Run 4, Estimate	0.957	0.631	0.9988	0.576
Run 1, 95% CI	(0.955,0.972)	(0.573,0.707)	(0.987,1.004)	(0.481,0.603)
Run 2, 95% CI	(0.955,0.969)	(0.571,0.703)	(0.999,1)	(0.510,0.649)
Run 3, 95% CI	(0.943,0.962)	(0.592,0.726)	(0.999,1)	(0.567,0.691)
Run 4, 95% CI	(0.949,0.965)	(0.562,0.701)	(0.997,1.001)	(0.507,0.644)

Table 7: Estimates and 95% confidence intervals for the posterior probability that $\alpha_k < 0$ for $k = 1, 2, 3, 4$.

	$k=1$	$k=2$	$k=3$	$k=4$
1	(–3.02,0.14)	(–3.36,2.48)	(–8.85,–1.85)	(–6.02,5.47)
2	(–3.03,0.14)	(–2.98,2.20)	(–9.45,–1.92)	(–6.30,5.73)
3	(–2.94,0.23)	(–3.23,2.10)	(–9.44,–2.23)	(–6.61,4.94)
4	(–3.05,0.20)	(–3.37,2.27)	(–9.38,–2.20)	(–6.52,5.58)

Table 8: 95% credible intervals for $\alpha_k,\ k = 1, 2, 3, 4$, computed from each of four MCMC runs.

that a burn-in of 20,000 iterations is adequate for convergence of the MCMC. The Heidelberger-Welch test was applied to find confidence intervals for the desired probabilities, with results in Table 7.

As a result of this calculation, we can see that the posterior probabilities that $\alpha_k < 0$ are close to 1 in the cases $k = 1$ and 3. (The slightly anomalous confidence intervals for $k = 3$ are probably explained by the non-normality of the posterior distribution of $\Pr\{\alpha_k < 0\}$.) Conversely, for $k = 2$ and $k = 4$, none of the boundaries of the confidence intervals are very close to 0 or 1. Despite the obvious ambiguities that remain, it seems safe to say that α_1 and α_3 are statistically significant (< 0), while α_2 and α_4 are not.

4.8.3 Posterior densities

Figure 6 shows the posterior density of $\alpha_k,\ k = 1, 2, 3, 4$ computed from each of the four runs. The four runs are in excellent agreement for $k = 1$; less so for $k = 2, 3, 4$, though the agreement is still good. Based on these posterior densities, from each run a 95% equal-tailed credible interval is computed for each α_k; see Table 8. These results reinforce that there is strong evidence that both $\alpha_1 < 0$ and $\alpha_3 < 0$, but for both α_2 and α_4, all four intervals cover 0 and therefore do not indicate a significant result. On the other hand, even for α_1 and α_3, the 95% prediction intervals are wide relative to the magnitudes of the posterior means, implying that even though we have high confidence that these parameters are < 0, it would still not be possible to make precise risk calculations based on their numerical values.

4.8.4 Summary of Convergence Diagnostics

The Gelman-Rubin diagnostics indicate that reasonable convergence has been achieved by iteration 20,000 at latest. The other statistics show generally good but not perfect agreement across the

four runs of the MCMC. We have very strong reason to believe that both α_1 and α_3 are negative (indicating a detrimental effect of PM_{10} in the two subgroups consisting of non-smokers without chronic lung disease) but the prediction intervals for the parameters themselves are still wide relative to the respective posterior means.

4.9 Summary and Conclusions

The present study differs from previous work by assuming a log-normal model, which is usually a better fit than the linear model in air pollution contexts. However, under this model, measurement error cannot be accounted for through simple adjustments to the regression coefficients. We have demonstrated a method that accounts for this measurement error using a computationally intensive but conceptually straightforward Bayesian technique.

Our results contradict some common conceptions about this kind of analysis. First, including measurement error did not necessarily cause the means of the posterior distributions to shift toward zero. Indeed, at $M = 1$ there was no discernable pattern to the direction of "bias" compared to $M = 0$. However, when the kriging error was doubled all the posteriors shifted toward zero. Second, our posteriors did not widen systematically after accounting for measurement error, and in fact *narrowed* in the $M = 2$ scenario. The original paper (Crooks *et al.*, 2008) presented evidence from a simulation study that these results are due to the log-normal measurement error distribution and not some other aspect of our model.

Our primary aim was to determine whether the epidemiological results were robust to ignoring kriging uncertainty. We have shown that, because of the way the posterior distributions change under the effect of log-normal emeasurement error, the credibility of the result is robust, though the effect magnitudes generally are not. We stress, though, that this final conclusion should not be taken to be generally true of log-normal measurement error models though it may be true in individual studies.

5 References

Brooks, S.P., and Gelman, A. (1998), General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics***7**, 434-455.

Everson, P.J. and Morris, C.N. (2000), Inference for multivariate normal hierarchical models. *Journal of the Royal Statistical Society B* **62**, 399–412.

Gelman, A. and Rubin, D.B. (1992), Inference from iterative simulation using multiple sequences. *Statistical Science* **7**, 457-511.

Heidelberger, P. and Welch, P.D. (1981). A spectral method for confidence interval generation and run length control in simulations. *Communications of the ACM* **24**, 233-245.

Liao, D., Peuquet, D. J., Duan, Y., Whitsel, E. A., Dou, J., Smith, R. L., Lin, H. M., Chen, J. C. and Heiss, G. (2006), GIS approaches for the estimation of residential-level ambient PM concentrations. *Environ Health Perspect***114(9)**, 1374-1380.

Plummer, M., Best, N., Cowles, K. and Vines, K.(2006), CODA: Convergence Diagnosis and Output Analysis for MCMC. *R NEWS***6**, 7-11

Ripley, B.D. (1987), *Stochastic Simulation.* Wiley, Chichester.

Rubin, D.B. (1987), Comment on "The calculation of posterior distributions by data augmentation" by Tanner and Wong. *Journal of the American Statistical Association* **82**, 543–546.

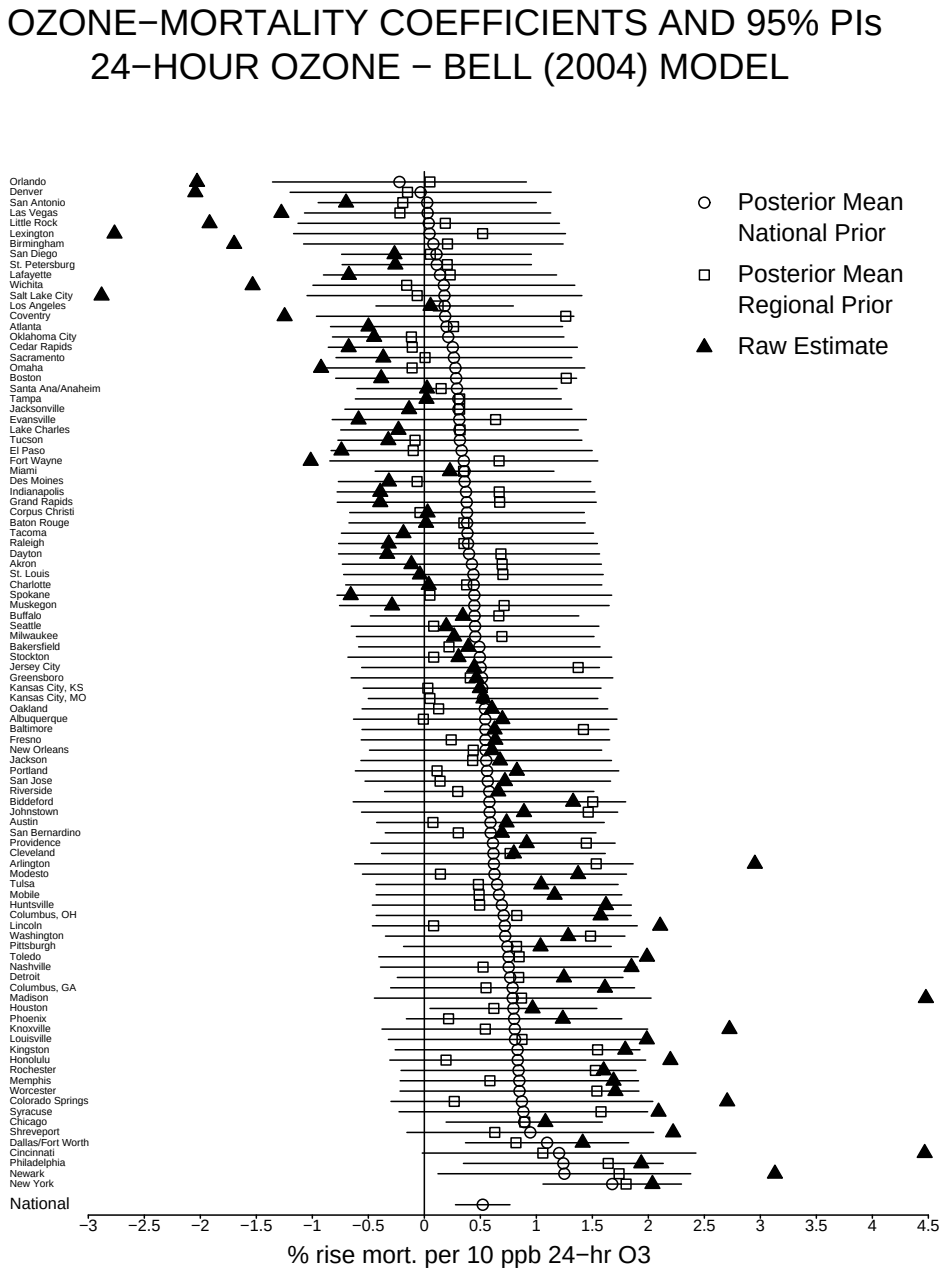

Fig. 1: 95% posterior intervals for the ozone-mortality coefficients, all-year data, by the hierarchical Bayesian method as in Fig. 2 of Bell et al. (2004). The Bayesian posterior estimates under the "national prior" (circles) are shown alongside those for the "regional prior" (squares) and the raw maximum likelihood estimates (triangles).

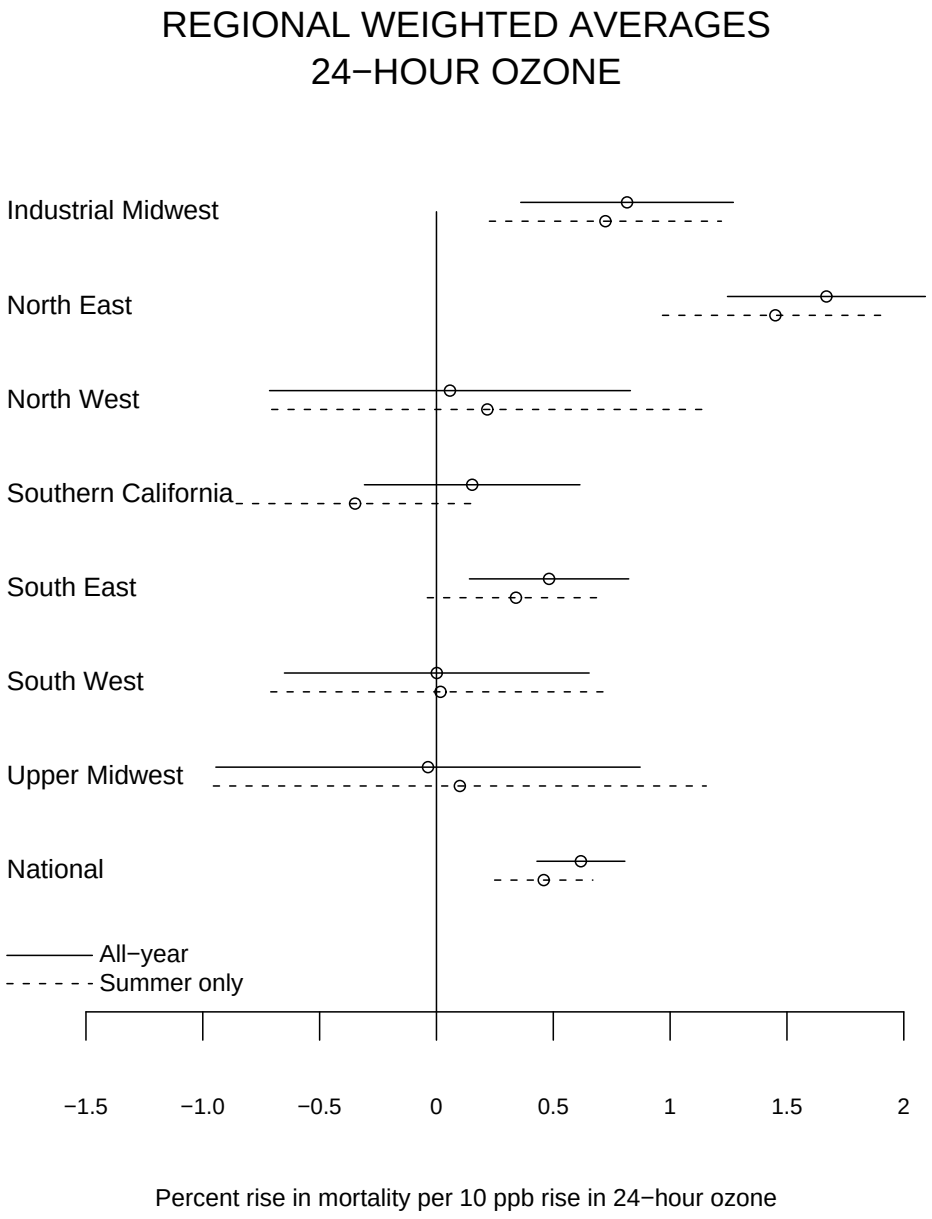

Fig 2: Regional estimates of population-weighted average regression coefficients based on 24-hour ozone, with 95% PIs.

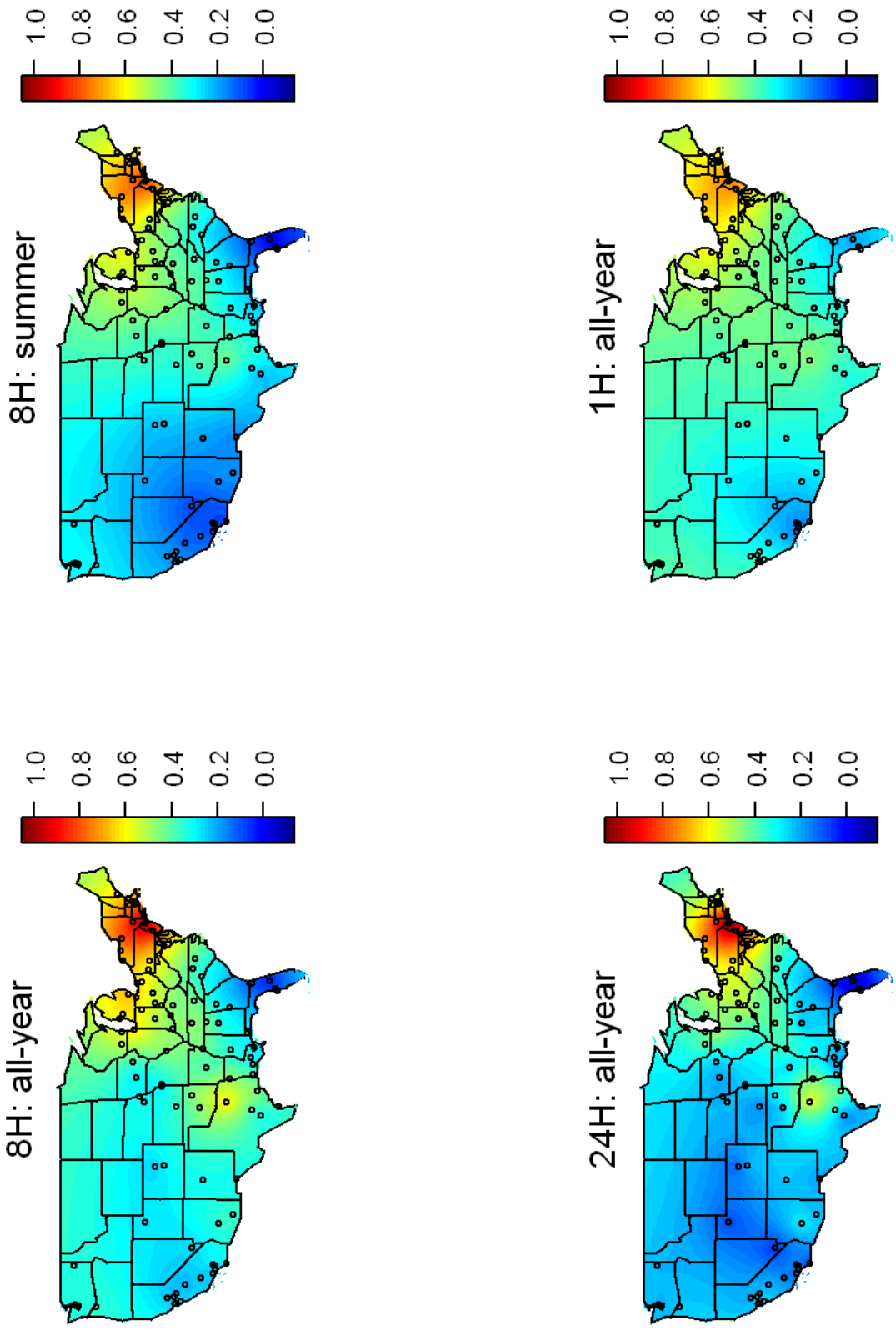

Fig 3: Map of spatially dependent ozone-mortality coefficient for 8-hour ozone (all-year data), 8-hour ozone (summer data), 24-hour ozone (all-year data) and 1-hour ozone (all-year data).

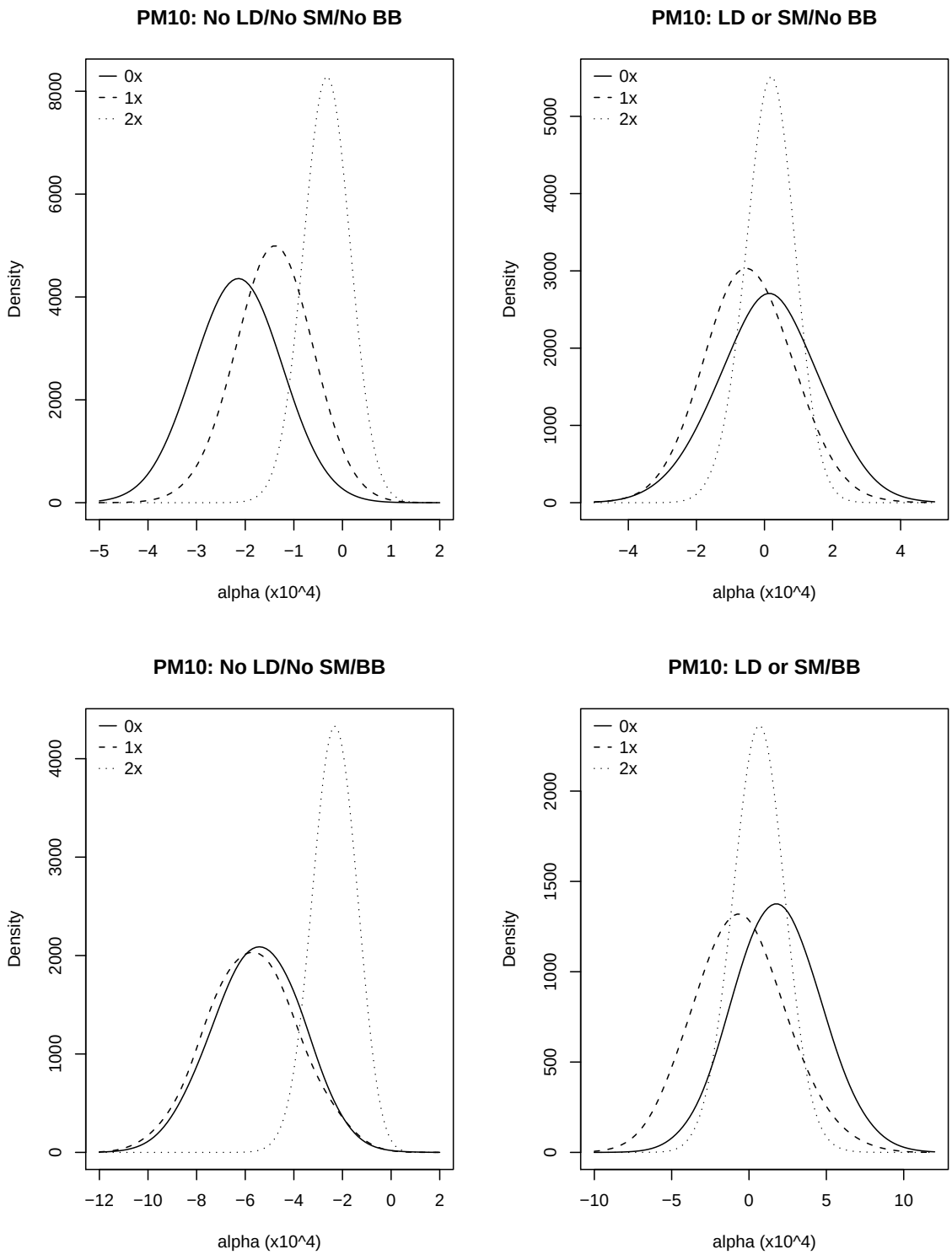

Fig 4: Posterior densities for PM_{10} regression coefficient by subgroup.

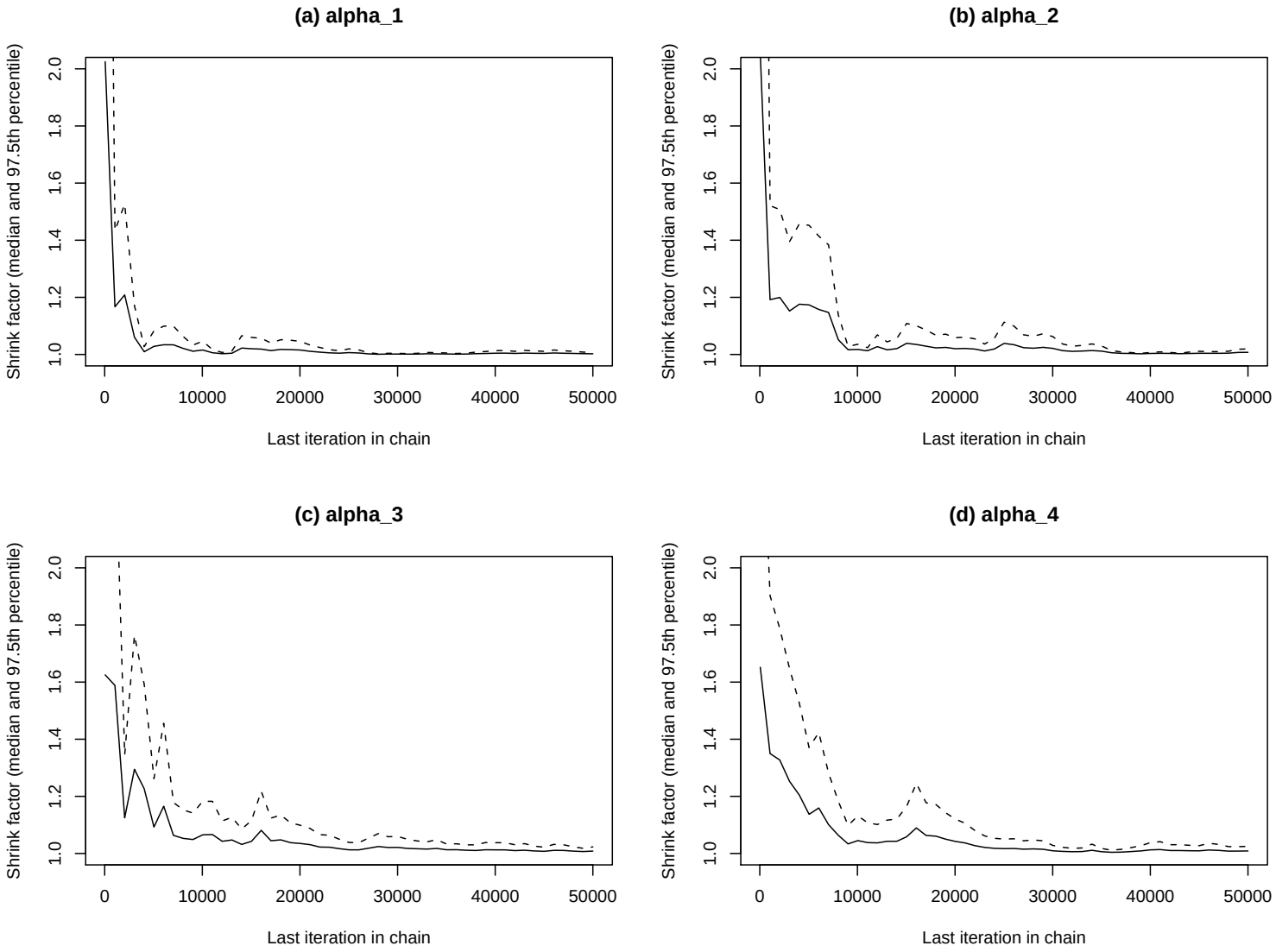

Fig 5: Gelman-Rubin diagnostics for convergence of MCMC. The solid curve represents the median value of Gelman-Rubin's R statistic; the dashed curve is the 97.5% quantile. All curves are for $M = 1$ and PM_{10} as the pollutant of interest.

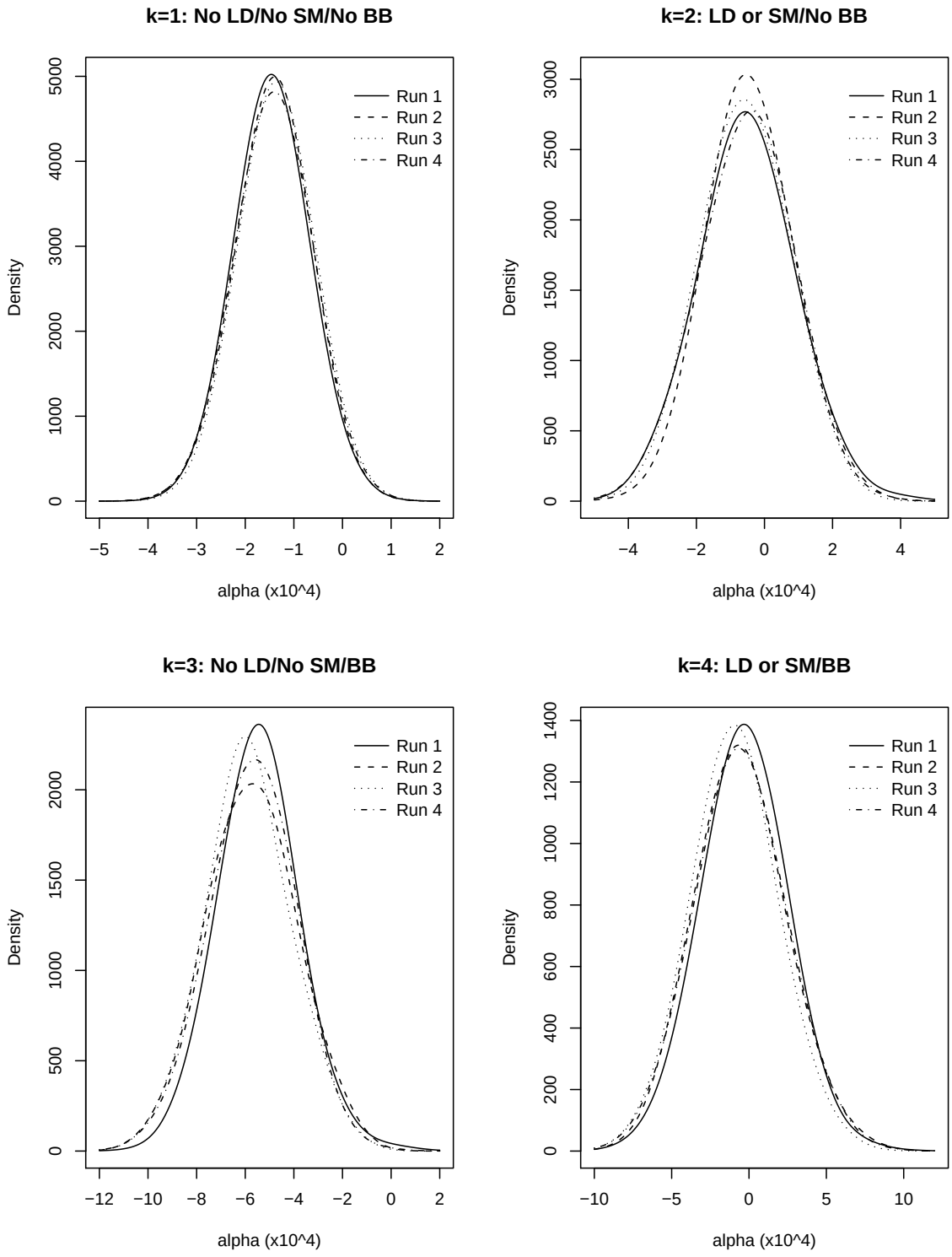

Fig 6: Posterior densities for PM_{10} regression coefficient by subgroup: 4 runs of MCMC.