 DOI: 10.1111/j.1541-0420.2007.01039.x

Bayesian Distributed Lag Models: Estimating Effects of Particulate Matter Air Pollution on Daily Mortality

L. J. Welty,[1,*] R. D. Peng,[2] S. L. Zeger,[2] and F. Dominici[2]

[1]Department of Preventive Medicine, Northwestern University, Feinberg School of Medicine, 680 North Lake Shore Drive, Suite 1102, Chicago, Illinois 60611, U.S.A.

[2]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, Maryland 21205, U.S.A.

[*]*email*: lwelty@northwestern.edu

Summary. A distributed lag model (DLagM) is a regression model that includes lagged exposure variables as covariates; its corresponding distributed lag (DL) function describes the relationship between the lag and the coefficient of the lagged exposure variable. DLagMs have recently been used in environmental epidemiology for quantifying the cumulative effects of weather and air pollution on mortality and morbidity. Standard methods for formulating DLagMs include unconstrained, polynomial, and penalized spline DLagMs. These methods may fail to take full advantage of prior information about the shape of the DL function for environmental exposures, or for any other exposure with effects that are believed to smoothly approach zero as lag increases, and are therefore at risk of producing suboptimal estimates. In this article, we propose a Bayesian DLagM (BDLagM) that incorporates prior knowledge about the shape of the DL function and also allows the degree of smoothness of the DL function to be estimated from the data. We apply our BDLagM to its motivating data from the National Morbidity, Mortality, and Air Pollution Study to estimate the short-term health effects of particulate matter air pollution on mortality from 1987 to 2000 for Chicago, Illinois. In a simulation study, we compare our Bayesian approach with alternative methods that use unconstrained, polynomial, and penalized spline DLagMs. We also illustrate the connection between BDLagMs and penalized spline DLagMs. Software for fitting BDLagM models and the data used in this article are available online.

Key words: Air pollution; Bayes; Distributed lag; Mortality; NMMAPS; Penalized splines; Smoothing; Time series.

1. Introduction

Distributed lag models (DLagMs; Almon, 1965) are regression models that include lagged exposure variables, or distributed lags (DLs), as covariates. They have recently been employed in environmental epidemiology for estimating short-term cumulative effects of environmental exposures on daily mortality or morbidity (e.g., Pope et al., 1991; Pope and Schwartz, 1996; Braga et al., 2001; Zanobetti et al., 2002; Kim, Kim, and Hong, 2003; Bell McDermott, Zeger, Samet, and Dominici, 2004; Goodman, Dockery, and Clancy, 2004; Welty and Zeger, 2005). DLagMs are specialized types of varying-coefficient models (Hastie and Tibshirani, 1993) and dynamic linear models (Ravines, Schmidt, and Migon, 2006).

For Poisson log-linear DLagMs that estimate the effects of lagged air pollution levels on daily mortality counts, the sum of the DL coefficients is interpreted as the percentage increase in daily mortality associated with a one unit increase in air pollution on each of the previous days. Because the time from exposure to event will almost certainly vary in a population, this sum is a more appropriate measure of the effect of short-term exposure than a single day's coefficient. Results from previous time series studies suggest that compared to DLagMs, models with single day pollution exposures might underestimate the risk of mortality associated with air pollution (Schwartz, 2000; Zanobetti et al., 2003; Goodman et al., 2004; Roberts, 2005).

Exposure variables, such as ambient air pollution levels, may be highly correlated over time, making DL coefficients difficult to estimate. A general solution is to constrain the coefficients as a function of lag. Common constraints include a polynomial (Almon, 1965) or a spline (Corradi, 1977). Estimating DLagMs as varying-coefficient models constrains the coefficients to follow a natural cubic spline (Hastie and Tibshirani, 1993). The DL function for air pollution and mortality has been estimated with polynomial constraints (e.g., Schwartz, 2000, Braga et al., 2001; Kim et al., 2003; Bell, Samet, and Dominici, 2004; Goodman et al., 2004), spline constraints (Zanobetti et al., 2000), and without constraints (Zanobetti et al., 2003).

Each type of constraint on the DL coefficients is an application of prior knowledge to model specification. In the context of air pollution and mortality, prior knowledge suggests that short-term risk of mortality varies smoothly as a function of lag and decreases to zero. Prior knowledge about the effects of air pollution on mortality at early lags is limited. There may be short delays in health effects after exposure,

as suggested by studies of single day pollution exposures that find the largest effect on mortality at lag day 1 (Zmirou et al., 1988; Katsouyanni et al., 2001; Dominici et al., 2003). In the scenario of mortality displacement (Schimmel and Murawsky, 1978), in which high air pollution levels may advance by several days the deaths of frail individuals, the DL function may be zero or positive at early lags, then decrease and become negative (Zanobetti et al., 2000, 2002). If there were both a delay in health effect and mortality displacement, hypotheses concerning the sign or smoothness of the DL function at early lags would be tenuous at best.

For more appropriate model specification and improved estimation, it may be advisable to formulate DLagMs so that (i) coefficients are constrained to approach zero smoothly with increasing lag and (ii) early coefficients are relatively unconstrained. Neither polynomial nor spline constraints, the most common methods for specifying DLagMs, include this prior information in estimation. In this article, we develop Bayesian DLagMs (BDLagMs) that incorporate our understanding of the relationship between short-term fluctuations of particulate matter (PM) air pollution and daily fluctuations in mortality counts. Our prior distribution specifies that as lag increases, the DL function will have increasing smoothness and approach zero. An advantage of our approach is that the degree of smoothness of the DL function is estimated from the data. We note that BDLagMs have been explored in economics (e.g., Leamer, 1972; Schiller, 1973; Ravines et al., 2006), and autoregressive priors have been used generally to smooth time-dependent coefficients in generalized linear models (e.g., Fahrmeir and Knorr-Held, 1997; Manda and Meyer, 2005). However, our prior is quite different from those using a constant degree of smoothness (Schiller, 1973), a particular parametric form (Leamer, 1972; Ravines et al., 2006), or an autoregressive structure (e.g., Fahrmeir and Knorr-Held, 1997; Manda and Meyer, 2005).

We apply our BDLagM to data from the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) to estimate the shape of the DL function between daily PM and daily deaths for Chicago, Illinois from 1987 to 2000. We examine the sensitivity of the estimated DL function to the specification of the BDLagM prior. We compare the air pollution effect estimated with the BDLagM to that estimated using unconstrained maximum likelihood (ML). We also compare air pollution effects estimated under the full formulation of the BDLagM, computed using a Gibbs sampler, to those estimated under an approximate formulation, computed using a closed form expression.

We also conduct a simulation study comparing BDLagMs to unconstrained, polynomial, and penalized spline DLagMs. For penalized spline DLagMs, we compare estimates obtained using generalized cross validation (GCV) and restricted maximum likelihood estimation (REML; Ruppert, Wand, and Carroll, 2003). We include DLagMs that are consistent with biological knowledge along with DLagMs for which our BDLagMs may be misspecified.

Because constraining DL coefficients is a way of smoothing, we consider how our Bayesian approach relates to penalized spline DLagMs. We demonstrate that BDLagMs are analogous to penalized spline DLagMs with a specific penalty matrix derived from the BDLagM prior.

Though our BDLagM formulation was motivated by a desire to model flexibly the DL function between lagged PM levels and daily mortality counts, it is relevant to situations in which the lagged effects of an exposure on an outcome are unknown for the first few lags but are believed to dissipate with lag. Using BDLagMs with repeated measures data would require extensions to our approach. For documentation and to encourage implementation, our BDLagM software is available online at `http://www.ihapss.jhsph.edu/software/BayesDLM/`.

2. Bayesian DLagMs

Let y_t and x_t be the outcome and exposure time series. We consider a generalized linear DLagM $g(E[y_t \,|\, x_1, \ldots, x_t]) = \sum_{\ell=0}^{L} \theta_\ell x_{t-\ell}$ where L is the maximum lag and $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_L)'$ is the vector of the DL coefficients to be estimated. Initially we will consider the normal linear model $E[y_t \,|\, x_1, \ldots, x_t] = \sum \theta_\ell x_{t-\ell}$, with $\mathbf{Y}_t$ independent normal with constant variance.

The goal is to specify a prior on $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_L)'$ that is uninformative on the DL coefficients for small ℓ but that constrains the coefficients with larger ℓ to be smoother and approach zero. We assume $\boldsymbol{\theta} \sim N(0, \boldsymbol{\Omega})$, where $\boldsymbol{\Omega}$ is constructed so that for increasing lag the diagonal elements decrease to zero ($\mathrm{Var}(\theta_\ell) \to 0$) and the off–diagonal elements in its correlation matrix increase to one ($\mathrm{Cor}(\theta_{\ell-1}, \theta_\ell) \to 1$). Care must be taken to construct $\boldsymbol{\Omega}$ so that it remains positive definite. A natural approach is to define $\boldsymbol{\Omega} = \mathbf{ABA}$, where $\mathbf{AA^T}$ is the diagonal matrix of the individual variances of the θ_ℓs, and $\mathbf{B}$ is the correlation matrix for $\boldsymbol{\theta}$. Specifying an appropriate $\boldsymbol{\Omega}$ may then be achieved by setting $\mathbf{A}$ equal to the Cholesky decomposition of a diagonal matrix with the desired prior variances and setting $\mathbf{B}$ equal to the correlation matrix for increasingly correlated normal random variables.

To define $\mathbf{A}$, let the parameter σ^2 be the prior variance of θ_0, and set $\mathrm{Var}(\theta_1) = v_1\sigma^2, \ldots, \mathrm{Var}(\theta_L) = v_L\sigma^2$ where the v_ℓs are a decreasing sequence of weights such that $1 \geq v_1 \geq \cdots \geq v_L > 0$. We parameterize them by $v_\ell(\eta_1) = \exp(\eta_1 \ell)$, $\eta_1 \leq 0$, so that the hyperparameter η_1 governs how quickly the prior variances of the θ_ℓs approach zero. Choosing the exponential function is convenient but not required. Let $\mathbf{V}(\eta_1)$ be the diagonal matrix with entries 1, $v_1(\eta_1)^{1/2}, \ldots, v_L(\eta_1)^{1/2}$. We set $\mathbf{A} = \sigma\mathbf{V}(\eta_1)$.

To specify the correlation matrix $\mathbf{B}$, we similarly define $w_\ell(\eta_2) = \exp(\eta_2 \ell)$, $\eta_2 \leq 0$, to be a decreasing sequence of weights, and $\mathbf{M}(\eta_2)$ to be the $(L+1) \times (L+1)$ diagonal matrix with entries 1, $w_1(\eta_2), \ldots, w_L(\eta_2)$. We let $\mathbf{B} = \mathbf{W}(\eta_2)$, where $\mathbf{W}(\eta_2)$ is the correlation matrix derived from the covariance matrix $\mathbf{M}(\eta_2)\mathbf{M}(\eta_2)' + \{\mathbf{I}_{L+1} - \mathbf{M}(\eta_2)\}\mathbf{1}_{L+1} \times \mathbf{1}'_{L+1}\{\mathbf{I}_{L+1} - \mathbf{M}(\eta_2)\}'$, where by $\mathbf{1}_{L+1}$ we mean a $(L+1) \times 1$ vector of ones and by $\mathbf{I}_{L+1}$ we mean the $(L+1) \times (L+1)$ identity matrix. Then $\mathbf{W}(\eta_2)$ is the correlation matrix for the mixture of normal random variables $\mathbf{M}(\eta_2)\mathbf{X}_1 + \{\mathbf{I}_{L+1} - \mathbf{M}(\eta_2)\}\mathbf{1}_{L+1}X_2$ where $\mathbf{X}_1 \sim N(0, \mathbf{I}_{L+1})$ and $X_2 \sim N(0, 1)$. The first few elements of the independent $\mathbf{X}_1$ are weighted more heavily than the corresponding first few elements of the dependent $\mathbf{1}_{L+1}X_2$, and the latter elements of the dependent $\mathbf{1}_{L+1}X_2$ are weighted more heavily than the latter elements of the independent $\mathbf{X}_1$. The parameter η_2 controls how quickly the mixture moves from independent to dependent. The final

form for the prior on $\boldsymbol{\theta}$ is then $N(0,\, \sigma^2\boldsymbol{\Omega}(\boldsymbol{\eta}))$, where $\boldsymbol{\Omega}(\boldsymbol{\eta}) = \mathbf{V}(\eta_1)\mathbf{W}(\eta_2)\mathbf{V}(\eta_1)$ and $\boldsymbol{\eta} = (\eta_1,\, \eta_2)'$.

Let $\hat{\boldsymbol{\theta}}$ be the ML estimate of the unconstrained DL coefficients and let $\boldsymbol{\Sigma}$ be the sample covariance matrix. For a normal linear DLagM, $\hat{\boldsymbol{\theta}}$ is $N(\boldsymbol{\theta},\, \boldsymbol{\Sigma})$, so the posterior for $\boldsymbol{\theta}$ conditional on $\boldsymbol{\eta}$ and σ is

$$\boldsymbol{\theta} \,|\, \hat{\boldsymbol{\theta}}, \boldsymbol{\eta}, \sigma^2 \sim N\left(\left\{1/\sigma^2\boldsymbol{\Omega}(\boldsymbol{\eta})^{-1} + \boldsymbol{\Sigma}^{-1}\right\}^{-1}\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\theta}},\right.$$
$$\left.\left\{1/\sigma^2\boldsymbol{\Omega}(\boldsymbol{\eta})^{-1} + \boldsymbol{\Sigma}^{-1}\right\}^{-1}\right). \qquad (1)$$

For a general linear DLagM, the posterior distribution for $\boldsymbol{\theta}$ may not be available in closed form, but it may be computed through Gibbs sampling or other Markov chain Monte Carlo methods (e.g., Carlin and Louis, 2000). We discuss such an approach for our PM air pollution and mortality example, in which the $\mathbf{Y}_t$ are Poisson distributed daily mortality counts, $\log(E[y_t \,|\, x_1, \ldots, x_t]) = \sum_{\ell=0}^{L}\theta_\ell x_{t-\ell}$, and the likelihood for $\hat{\boldsymbol{\theta}}$ is Poisson.

The influence of the prior distribution in estimating $\boldsymbol{\theta}$ depends on the values of hyperparameters σ^2 and $\boldsymbol{\eta} = (\eta_1,\, \eta_2)'$. The hyperparameter σ^2, the prior variance of θ_0, can be viewed as a tuning parameter determining the starting point of the DL function. In practice there is little information in the data to jointly estimate σ^2 and $\boldsymbol{\eta}$. We therefore assume σ^2 is ten times the estimated statistical variance of θ_0 so that even for relatively large values of $\boldsymbol{\eta}$, the prior has little to no influence on the first few DL coefficients. We examine sensitivity of BDLagM estimates to choice of σ in Section 5.

Rather than setting values for $\boldsymbol{\eta} = (\eta_1, \eta_2)'$ and directly determining the influence of the prior, we let $\boldsymbol{\eta} = (\eta_1, \eta_2)'$ have a discrete uniform prior on $\mathbf{N}_1 \times \mathbf{N}_2$, where $\mathbf{N}_1$ and $\mathbf{N}_2$ are finite sets of possible values for η_1 and η_2. Then the posterior distribution for $\boldsymbol{\theta}$ can be defined as the weighted sum $p(\boldsymbol{\theta} \,|\, \hat{\boldsymbol{\theta}}) = \sum_{\boldsymbol{\eta}} p(\boldsymbol{\theta} \,|\, \hat{\boldsymbol{\theta}}, \boldsymbol{\eta}) p(\boldsymbol{\eta} \,|\, \hat{\boldsymbol{\theta}})$, where p denotes a general probability density. Under the assumption that $\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$, the marginal posterior density of the hyperparameter $\boldsymbol{\eta}$ is available in closed form. For a given $\boldsymbol{\eta}^*$:

$$p(\boldsymbol{\eta}^* \,|\, \hat{\boldsymbol{\theta}}) = \frac{|\, \sigma^2\boldsymbol{\Omega}(\boldsymbol{\eta}^*)\boldsymbol{\Sigma}^{-1} + \mathbf{I}|^{-1/2} \exp\left[-\frac{1}{2}\hat{\boldsymbol{\theta}}'\left\{\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma}^{-1} + \frac{1}{\sigma^2}\boldsymbol{\Omega}(\boldsymbol{\eta}^*)^{-1}\right)^{-1}\boldsymbol{\Sigma}^{-1}\right\}\hat{\boldsymbol{\theta}}\right]}{\sum_{\boldsymbol{\eta}} |\, \sigma^2\boldsymbol{\Omega}(\boldsymbol{\eta})\boldsymbol{\Sigma}^{-1} + \mathbf{I}|^{-1/2} \exp\left[-\frac{1}{2}\hat{\boldsymbol{\theta}}'\left\{\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma}^{-1} + \frac{1}{\sigma^2}\boldsymbol{\Omega}(\boldsymbol{\eta})^{-1}\right)^{-1}\boldsymbol{\Sigma}^{-1}\right\}\hat{\boldsymbol{\theta}}\right]}. \qquad (2)$$

Sufficiently large ranges for $\mathbf{N}_1$ and $\mathbf{N}_2$ insure that the data drive the strength or weakness of the prior distribution and therefore the eventual smoothness of the estimated DL function.

3. Bayesian DLagMs and Penalized Splines

Following the well-established connection between nonparametric smoothing and Bayesian modeling (e.g., Silverman, 1985), we illustrate the relationship between normal linear BDLagMs and p-spline DLagMs. We show that estimating the normal linear DL function under model (1) is analogous to fitting a p-spline to DL coefficients with penalty derived from our prior. An advantage of this connection is that our method of putting a prior directly on the coefficients may be viewed as a transparent means for eliciting p-spline penalties, which are otherwise difficult to relate to biological or other prior knowledge.

Let $\boldsymbol{\theta} = \boldsymbol{U}\boldsymbol{\gamma}$, where $\boldsymbol{U}$ is a spline basis matrix and $\boldsymbol{\gamma}$ is a vector of spline coefficients. Let $\hat{\boldsymbol{\theta}}$ be the ML estimate of $\boldsymbol{\theta}$, and assume that $\hat{\boldsymbol{\theta}} = \boldsymbol{U}\boldsymbol{\gamma} + \boldsymbol{\nu}$, $\boldsymbol{\nu} \sim N(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the estimated covariance matrix for $\hat{\boldsymbol{\theta}}$. Under a p-spline approach, we estimate $\boldsymbol{\gamma}$ by minimizing the criterion $(\hat{\boldsymbol{\theta}} - \boldsymbol{U}\boldsymbol{\gamma})'\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{U}\boldsymbol{\gamma}) + \lambda\boldsymbol{\gamma}^T\boldsymbol{D}\boldsymbol{\gamma}$, where λ is a penalty parameter and $\boldsymbol{D}$ a positive semidefinite matrix (Eilers and Marx, 1996; Ruppert et al., 2003).

To show the connection between minimizing this criterion and estimating the BDLagM, (1), we reformulate the p-spline in its Bayesian form $\hat{\boldsymbol{\theta}} \,|\, \boldsymbol{\gamma} \sim N(\boldsymbol{U}\boldsymbol{\gamma}, \boldsymbol{\Sigma})$ and $\boldsymbol{\gamma} \sim N(0,\, \boldsymbol{\Gamma})$, where $\boldsymbol{\Gamma}$ is the prior covariance matrix of $\boldsymbol{\gamma}$. Because $\boldsymbol{\theta} = \boldsymbol{U}\boldsymbol{\gamma}$, the prior on $\boldsymbol{\gamma}$ translates to prior $\boldsymbol{\theta} \sim N(0,\, \boldsymbol{U}\boldsymbol{\Gamma}\boldsymbol{U}')$. In (1) we assume $\boldsymbol{\theta} \sim N(0,\, \sigma^2\Omega(\boldsymbol{\eta}))$, so we need $\boldsymbol{\Gamma}$ such that $\boldsymbol{U}\boldsymbol{\Gamma}\boldsymbol{U}' = \sigma^2\boldsymbol{\Omega}(\boldsymbol{\eta})$, or $\boldsymbol{\Gamma}(\boldsymbol{\eta}) = \boldsymbol{R}^{-1}\boldsymbol{Q}'\sigma^2\boldsymbol{\Omega}(\boldsymbol{\eta})\boldsymbol{Q}\boldsymbol{R}'^{-1}$ where $\boldsymbol{Q}\boldsymbol{R}$ is $\boldsymbol{U}$'s qr-decomposition.

Under this formulation the log posterior for $\boldsymbol{\gamma}$ is, up to a constant, $-\frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{U}\gamma)'\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{U}\gamma) - \frac{1}{2}\boldsymbol{\gamma}'\boldsymbol{U}'(\boldsymbol{U}\Gamma(\boldsymbol{\eta})W')^{-1}\boldsymbol{U}\boldsymbol{\gamma}$, and maximizing the log posterior for $\boldsymbol{\gamma}$ is equivalent to minimizing the above criterion with $\lambda = 1$ and $\boldsymbol{D} = \boldsymbol{U}'\,(\boldsymbol{U}\boldsymbol{\Gamma}(\boldsymbol{\eta})\,\boldsymbol{W}')^{-1}\,\boldsymbol{U}$ (Silverman, 1985; Green and Silverman, 1994). For a given value of the hyperparameter $\boldsymbol{\eta}$, the estimated DL coefficients are given by the posterior mean $\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{\Sigma}^{-1}\boldsymbol{U} + \boldsymbol{U}'(\boldsymbol{U}\boldsymbol{\Gamma}(\boldsymbol{\eta})\boldsymbol{U}')^{-1}\boldsymbol{U}^{-1})^{-1}\boldsymbol{U}'\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\theta}}$, and the equivalent degrees of freedom equal the trace of the smoother matrix $\boldsymbol{X}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{X} + \boldsymbol{X}^{\boldsymbol{T}}(\boldsymbol{X}\boldsymbol{\Gamma}(\boldsymbol{\eta})\boldsymbol{X}^{\boldsymbol{T}})^{-1}\boldsymbol{X}^{-1})\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\Sigma}^{-1}$ (Ruppert et al., 2003).

Though a prior on DL coefficients may be translated to a specific p-spline penalty, the spline approach requires that the DL function follow a specific form, $\boldsymbol{\theta} = \boldsymbol{U}\boldsymbol{\gamma}$. For our air pollution mortality example, we found that using a b-spline basis with $L + 1$ degrees of freedom produced estimates of $\boldsymbol{\theta}$ identical to those from the BDLagM. In the following simulation study, we compare BDLagMs to p-splines with penalties unrelated to the prior.

4. Simulation Study

We conducted a simulation study to compare BDLagMs with four methods for estimating DL functions—unconstrained, polynomial, p-splines with penalty parameter chosen by GCV, and p-splines estimated with REML. We generated data under 25 different sets of true DL coefficients, including examples for which coefficients do not decrease to zero and smoothness does not increase with lag. We categorize the DL functions by four characteristics: (1) shape—decaying exponential (E), step function (St), or gamma distribution (G); (2) latency—0 or 2, the number of initial coefficients equal to zero; (3) oscillation—as described by $(-1)^\ell \bmod 2$, to mimic mortality displacement; and (4) maximum nonzero lag−7 or 14, the lag

by which the coefficients are less than 0.01. We also considered a null DL function with all zero coefficients. All DL functions included current day ($\ell = 0$). We set $L = 14$ as in the subsequent air pollution mortality example. Except for the null model, all the DL functions were normalized so the sum of squares of the DL coefficients is 1. We refer to the nonnull functions by [Shape]^{o} ([latency], [max lag]), where the superscript indicates oscillation.

Under each of the 25 scenarios, we generated 500 outcome series y_t from the model $y_t = \delta \sum_{\ell=0}^{14} \theta_\ell x_{t-\ell} + \epsilon_t$ where $\epsilon_t \sim$ i.i.d. N(0,1), and δ is a constant to balance signal and noise. For the exposure series x_t we used mean centered PM_{10} for 1996 from Chicago, Illinois because there were no missing observations and the autocorrelation is similar to what we experience when estimating the association between PM_{10} and mortality for Chicago for 1987–2000. For simplicity we take the ϵ_t to be independent $N(0, 1)$, noting that our simulations still apply to situations in which the ϵ_t are autocorrelated because application of an appropriate linear filter will result in a new DLagM with independent normal errors. We set $\delta = 0.25$ to generate moderate evidence for a total effect, $\sum \theta_\ell$, in nonnull models (we empirically determined that $\delta = 0.25$ generates y_t such that the t-statistic for the ML estimate for $\sum_\ell \theta_\ell$ is approximately two). Similarly we set $\delta = 0.475$ to generate strong evidence for total effect (we empirically determined that $\delta = 0.475$ generates y_t such that the t-statistic for the ML estimate for $\sum_\ell \theta_\ell$ is approximately four). For each simulated data set we compared the DL functions under five methods: (1) unconstrained ML; (2) the proposed Bayes' method (Bayes) using the normal posterior as in (1); (3) ML with a polynomial of degree four (Poly); (4) a penalized spline with penalty chosen by GCV (GCV); and (5) a penalized spline estimated with REML (REML). We also considered estimating the DL function using an AR-1 model. With the exception of the null model and $St^0(2, 14)$, the AR-1 model was not competitive, and was substantially worse when the DL function oscillates then goes to zero.

Figure 1 shows the estimated DL functions (white) averaged across the 500 simulations with the 95% confidence bands (gray) for 24 of the true DL functions (black) (results not pictured for null model). Results are reported for $\delta = 0.25$. Visual inspection of this figure indicates that the BDLagM performs consistently well and estimates the true DL function with narrower confidence bands than other methods.

To quantify the comparison, we summarize the mean squared errors of the estimated total effect ($\sum \theta_\ell$) and DL coefficients at lags 0, 7, and 14 under the five estimation methods and for the 25 scenarios. Table 1 summarizes the results for $\delta = 0.25$. Results for $\delta = 0.475$ are available in Web Table 1. Mean squared errors are expressed as percentages of the mean squared error of the corresponding unconstrained ML estimates. Values smaller than 100 favor the proposed estimation methods with respect to unconstrained ML.

When the DL function decreases to zero, BDLagM is 10 to 15% better at estimating the total effect than ML, whereas Poly, GCV, and REML perform comparably to ML. Results are similar for $\delta = 0.25$ and $\delta = 0.475$. The better performance of the Bayesian method with respect its competitors is mainly due to its greater flexibility in estimating the DL coefficients at the longer lags. Bayes is consistently 20–30% better than ML for lag 0; GCV and REML may be substantially better or substantially worse. However, Bayes consistently outperforms the others in estimating the lag 7 and the lag 14 coefficients for scenarios in which the coefficients go to zero by lag 7 or 14. When the BDLagM is misspecified and the DL coefficients do not decrease smoothly to zero, performance of the BDLagM is less predictable. Bayes may estimate the total effect only 5% worse than ML (and Poly and REML), or nearly 15% better (superior to Poly, GCV, REML).

Mortality counts are often modeled with Poisson log-linear regression, so we also examine how our results extend to the Poisson case. We simulated data from $Y_t \sim \text{Poisson}(\mu_t)$, $\log(\mu_t) = \log(100) + \Sigma_{\ell=0}^{\ell=14} x_{t-\ell}\theta_\ell/100$. The offset and division by 100 were determined empirically to approximate Chicago mortality levels in 1996. For each set of DL coefficients, we generated 1000 mortality series. We estimated the posterior distribution for $\boldsymbol{\theta}$ two ways—using (1) (approximating $\hat{\boldsymbol{\theta}}$ as normal) or a Gibbs sampler. Web Table 2 compares the mean squared errors of the total effects. The errors are comparable, suggesting that the simulation results for normal outcomes are not necessarily misleading for Poisson outcomes.

5. Application to Particulate Matter Air Pollution and Mortality

In this section, we apply BDLagMs to daily time series of PM with aerodynamic diameter less than 10 microns (PM_{10}) and nonaccidental deaths for Chicago, Illinois for the period 1987–2000. The data were collected from publicly available sources as part of the NMMAPS. NMMAPS contains daily time series of age classified mortality, temperature, dew point, and PM_{10} for 109 U.S. cities from 1987 to 2000. We analyzed the time series for Chicago because it is the largest U.S. city in NMMAPS with few missing PM_{10} values. Additional details regarding NMMAPS data assembly are available at `http://www.ihapss.jhsph.edu/` and are discussed in previous NMMAPS analyses (Samet, Zeger, Dominici, Curriero, Dockery, Schwartz, and Zanobetti, 2000; Samet, Zeger, Dominici, Schwartz, and Dockery, 2000; Dominici et al., 2003).

Poisson log-linear regression is frequently used to estimate the association between day-to-day variations in mortality counts and day-to-day variations in ambient air pollution levels. We accordingly assume that the mortality in Chicago on day t, $t = 1, \ldots, 5114$, is a Poisson random variable Y_t with expectation $E[Y_t] = \mu_t$. As above, we let $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_L)'$ be the unknown DL coefficients we wish to estimate. We let x_t denote the PM_{10} time series and for $t > L$ we let $\boldsymbol{x}_t$ denote the length $L + 1$ vector of lagged PM_{10} values $(x_t, \ldots, x_{t-L})'$.

Multisite time series studies of single day exposure PM_{10} and mortality have found strong evidence of an association between PM_{10} at lags $l = 0$, 1, and 2 and daily mortality (e.g., Zmirou et al., 1988; Burnett, Cakmak, and Brook, 1998; Katsouyanni et al., 2001; Dominici et al., 2003); single city studies with DLagMs have similarly found the largest effects in the first seven lags (e.g., Schwartz, 2000; Zanobetti et al., 2003; Goodman et al., 2004). Though lags beyond two weeks may have some influence on daily mortality (e.g., mortality displacement), it is unlikely that lags beyond 2 weeks have substantial influence on mortality compared to lags less than 2 weeks (Zanobetti et al., 2003). Models containing lags beyond 2 weeks are additionally difficult to estimate because long-term averages of PM_{10} have strong seasonal variation.

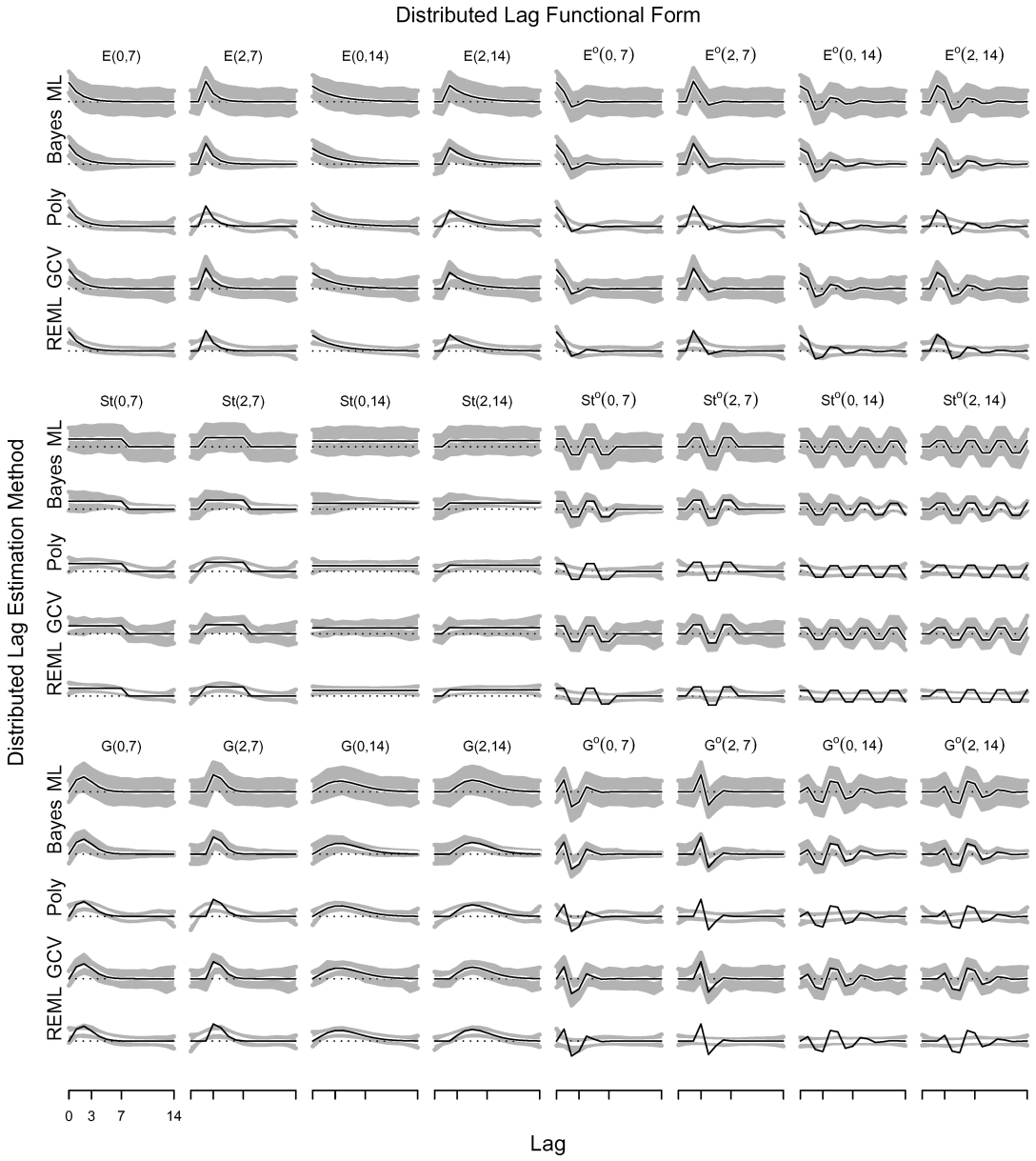

Figure 1. Mean estimated DL functions (white) and 95% posterior bands (gray) under five estimation methods—unconstrained ML, the proposed Bayesian method (Bayes), ML with a polynomial of degree four (Poly), a penalized spline with penalty chosen by GCV (GCV), and a penalized spline estimated with REML (REML). Outcome series were simulated under moderately strong evidence for the sum of the DL coefficients ($\delta = 0.25$).

We set $L = 14$ to capture the majority of short-term effects of PM_{10} on mortality without confounding estimation of DL coefficients with seasonal trends in mortality.

When estimating air pollution health effects from time series studies it is important to account for potential time-varying confounders such as weather, seasonality, and influenza epidemics (e.g., Schwartz, 1993; Samet et al., 1998; Braga, Zanobetti, and Schwartz, 2000; Samoli et al., 2001; Bell, Samet, and Dominici, 2004; Dominici, McDermott, and Hastie, 2004; Peng, Dominici, and Louis, 2005; Welty and Zeger, 2005). We let $\boldsymbol{z}_t$ denote the vector of time-varying covariates to include in the model, and we specify $\boldsymbol{z}_t$ as in previous NMMAPS analyses (Dominici et al., 2003). The exact specification is documented in the associated `R` code, available at `http://www.ihapss.jhsph.edu/software/BayesDLM/`. Our goal is to estimate the DL coefficients $\boldsymbol{\theta}$ as part of the generalized linear model

$$\log(\mu_t) = \boldsymbol{x}_t'\boldsymbol{\theta} + \boldsymbol{z}_t'\boldsymbol{\beta}. \tag{3}$$

The estimate for $1000 \times \theta_\ell$ corresponds to the percentage increase in daily mortality associated with a $10\mu g/m^3$ increase in PM_{10} at lag ℓ, and $1000 \times \sum_{\ell=0}^{14} \theta_\ell$ corresponds to the percentage increase in daily mortality associated with a $10\mu g/m^3$ increase in PM_{10} at lags $\ell = 0, \ldots, 14$.

Bayesian estimation of the generalized linear model in (3) with our proposed prior for the DL coefficients $\boldsymbol{\theta}$ requires two

Table 1

Mean squared errors of the estimates of the total effect and of the DL coefficients at lags 0, 7, *and* 14 *obtained under four estimation methods (Bayesian method* (B), *a polynomial with four degrees of freedom* (*P*), *a p-spline with penalty parameter chosen by GCV* (*G*), *and a p-spline estimated with REML* (*R*)) *and for the* 25 *true DL functions. These results are reported under the assumption of moderately strong evidence of a total effect* ($\delta = 0.25$). *Mean squared errors are expressed as percentages of the mean squared error of the corresponding ML estimates.*

		Total effect				Lag 0				Lag 7				Lag 14			
		B	P	G	R	B	P	G	R	B	P	G	R	B	P	G	R
E(0,7)		89	99	102	99	84	56	175	129	6	14	36	6	2	100	83	62
E(2,7)		91	99	100	99	78	47	59	77	9	11	31	16	2	135	94	102
E(0,14)		91	99	103	99	81	47	161	57	6	13	36	8	3	96	89	67
E(2,14)		95	99	99	99	78	70	56	62	8	11	22	15	6	108	95	98
E^o(0, 7)		89	99	100	99	81	58	119	167	6	22	42	10	1	129	92	78
E^o(2, 7)		89	99	100	99	77	43	70	76	7	16	47	12	2	141	96	89
E^o(0, 14)		89	99	100	99	80	48	74	162	15	50	37	26	2	134	96	70
E^o(2, 14)		88	99	100	99	74	44	81	49	11	50	58	18	3	124	102	83
St(0,7)		97	99	102	99	75	55	76	29	40	29	27	40	9	130	103	69
St(2,7)		99	99	98	99	74	88	40	49	50	38	23	38	10	126	86	75
St(0,14)		106	99	102	99	73	47	58	10	7	13	19	3	28	95	96	37
St(2,14)		105	99	96	99	72	59	29	24	7	13	25	6	30	95	76	61
St^o(0, 7)		87	100	100	99	82	67	68	113	98	206	41	187	4	96	99	50
St^o(2, 7)		87	100	100	100	73	61	72	24	46	179	51	220	5	97	102	37
St^o(0, 14)		86	99	100	99	81	52	65	84	72	183	70	135	180	355	99	248
St^o(2, 14)		86	99	99	99	73	43	65	15	33	133	31	142	188	316	93	339
G(0,7)		92	99	99	100	73	70	64	149	11	11	19	22	3	131	93	106
G(2,7)		92	100	99	100	75	187	55	94	16	28	27	33	4	96	86	84
G(0,14)		99	99	97	99	75	57	27	40	8	15	23	10	14	96	82	84
G(2,14)		99	99	100	99	75	89	25	71	18	18	27	11	20	143	93	71
G^o(0, 7)		88	100	100	99	71	73	86	63	7	27	60	9	2	134	106	42
G^o(2, 7)		87	99	99	99	74	42	85	13	10	15	69	5	3	103	108	38
G^o(0, 14)		87	99	100	99	76	50	80	41	63	180	59	109	3	100	96	40
G^o(2, 14)		86	100	99	99	71	48	74	20	47	205	48	259	5	115	92	35
Null		89	99	96	99	74	47	21	10	5	13	24	3	1	95	83	37

extensions from the general approach outlined in Section 2. First, the likelihood for $(Y_t \,|\, \boldsymbol{x}_t,\, \boldsymbol{z}_t)$ is Poisson, so that $\hat{\boldsymbol{\theta}}$, the ML estimates of $\boldsymbol{\theta}$, will not be normal and the posterior distribution for $\boldsymbol{\theta} \,|\, \hat{\boldsymbol{\theta}}$ will not have a closed form expression. Second, usual Bayesian estimation requires specifying a joint prior for $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, an untenable approach given the size of the nonpollutant covariate matrix and its potential relationship with the pollutant covariate matrix.

We propose two approaches. The first is to fit (3) and treat the ML estimates $\hat{\boldsymbol{\theta}}$ as $N(\boldsymbol{\theta},\, \boldsymbol{\Sigma})$, where Σ is the sample covariance matrix. This approach ignores the uncertainty introduced by estimating $\boldsymbol{\beta}$ and relies on the asymptotic normality of the Poisson likelihood, but allows us to estimate $\boldsymbol{\theta}$ directly using its closed form posterior (1). The second approach is to fit the Poisson log-linear model using a Gibbs sampler; details and code are available at `http://www.ihapss.jhsph.edu/software/BayesDLM/`.

For both computational methods, we set the hyperprior on $\boldsymbol{\eta} = (\eta_1, \eta_2)$ to be a discrete uniform distribution over $\boldsymbol{N}_1 \times \boldsymbol{N}_2$, where $\boldsymbol{N}_1$ is a length 10 sequence ranging from -0.35 to -0.05 in equal intervals, and $\boldsymbol{N}_2$ is a length 10 sequence ranging from -0.37 to 0 in equal intervals. We selected the interval for $\boldsymbol{N}_1$ so that the ratio of the prior standard deviation of θ_0 to θ_L is bounded between 2 and 100. We selected the values for $\boldsymbol{N}_2$ so that the prior correlation of θ_{L-1} and θ_L is bounded approximately by 0 and 0.99. We also set $\sigma = 0.004$, slightly larger than the square root of ten times the estimated variance in the ML estimate of θ_0. The sensitivity of the estimated BDLagMs to choices of σ and $N_1 \times N_2$ is considered below. We ran the Gibbs sampler for $K = 5000$ iterations, discarding the first 1000 as burn-in. Diagnostic checks suggested that the algorithm converged.

Figure 2 shows the posterior mean and the 95% posterior region of the DL function for the association between PM_{10}

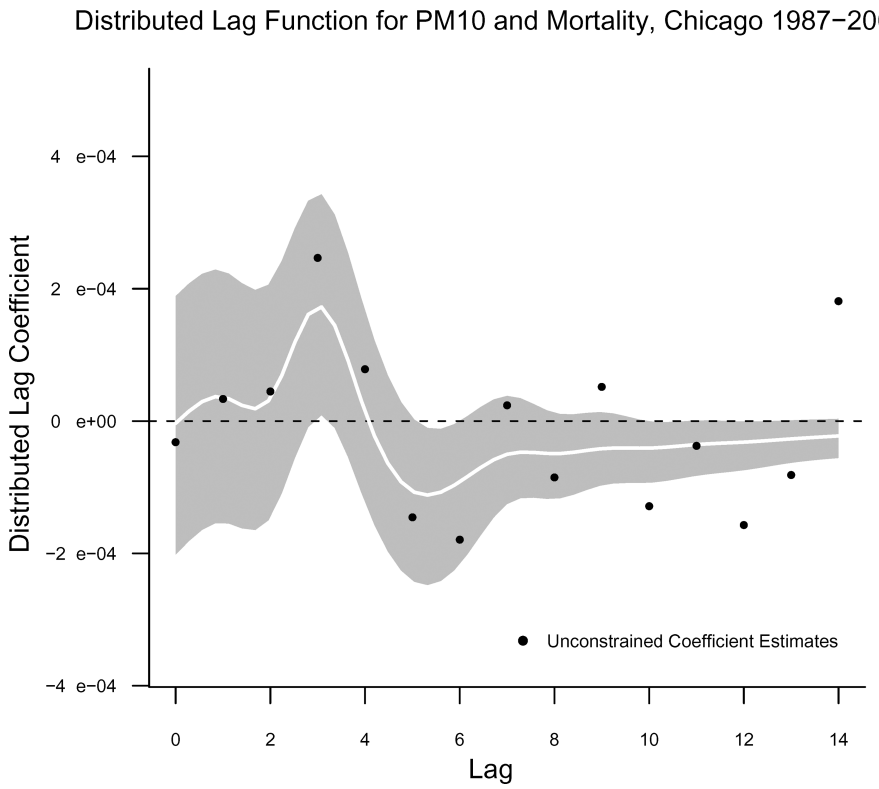

Figure 2. Posterior mean (white) of the DL function for the effect of PM_{10} on mortality for Chicago, Illinois from 1987 to 2000, using the last 4000 of 5000 iterations of the Gibbs sampler. The gray shaded region denotes the 95% posterior region. Black dots indicate ML estimates for the unconstrained DL coefficients.

and mortality in Chicago from 1987 to 2000. The black dots indicate the unconstrained ML estimates of the DL coefficients. The strongest association between PM and mortality occurs at lag 3: a $10\mu g/m^3$ increase in PM_{10} at lag 3 is associated with a 0.17% increase in mortality (95% posterior interval [PI] 0.01%, 0.34%), all other lagged PM_{10} levels remaining constant. The drop in relative risk from lag 3 to lag 5 suggests the possibility of mortality displacement. We estimate a total effect of −0.24% (95% PI −0.73%, 0.23%). The estimated total effect using unconstrained ML, −0.19%, is similar, but has a wider 95% confidence interval (−0.86%, 0.48%). The joint posterior distribution of $\boldsymbol{\eta} = (\eta_1, \eta_2)$ (see Web Figure 1) favored models for which $\text{Var}(\theta_\ell) \to 0$ quickly and $\text{Cor}(\theta_\ell, \theta_{\ell+1}) \to 1$ moderately or quickly.

Figure 3 compares posterior distributions of DL coefficients from the Gibbs sampler (black) and the normal approximation (gray). The estimates from the two methods differ for more moderate lags but are similar for early and later lags and for the overall sum of DL coefficients. This pattern of agreement and discrepancy is not surprising, given that we expect the normal approximation and the true posterior distribution to be most similar where the prior is weakest and the data drive estimation (early lags) and where the prior is strongest and drives estimation (later lags). The normal approximation was computationally faster than the Gibbs sampler (on an AMD Opteron 848 system with a 2.2 GHz processor, 8.6 seconds versus 15.5 hours for 5000 iterations).

We examined the sensitivity of the BDLagM estimates to the specification of the prior on $\boldsymbol{\eta}$ and the selection of the value for σ (Web Figure 2). The value for σ^2 was initially set to 10 times the estimated variance of θ_0. Larger values of σ result in BDLagMs that more closely followed the unconstrained ML estimates at longer lags. Smaller values of σ resulted in BDLagMs with latter DL coefficients shrunk to zero. For $\sigma =$ 0.04, 0.004, 0.0004, the initial DL coefficient estimates were indistinguishable. The original discrete uniform prior on η_1 was set so that the ratio of the prior standard deviation of θ_0 to θ_L ranged approximately from 2 to 100. We considered two new priors for η_1 so that the ratio ranged from approximately 2 to 50 (more restrictive) or from 2 to 200 (less restrictive). We did not consider alternate priors on η_2 because the prior was already constructed to be as broad as possible without creating numerical instability. The BDLagMs estimated across different prior distributions for $\boldsymbol{\eta}$ and $\sigma = 0.004$ were remarkably similar. We concluded that the estimated BDLagM is not driven strongly by the range of values for $\boldsymbol{\eta}$.

6. Discussion

We introduce a Bayesian approach to estimate DL functions in time series models of air pollution and mortality. This formulation uses prior knowledge about the shape of the DL function, and allows the degree of smoothness of the DL function to be estimated from the data. We illustrate in a simulation study that when prior assumptions are valid, BDLagMs estimate DL coefficients with smaller mean squared errors than three common methods—polynomial, spline, and unconstrained DLagMs.

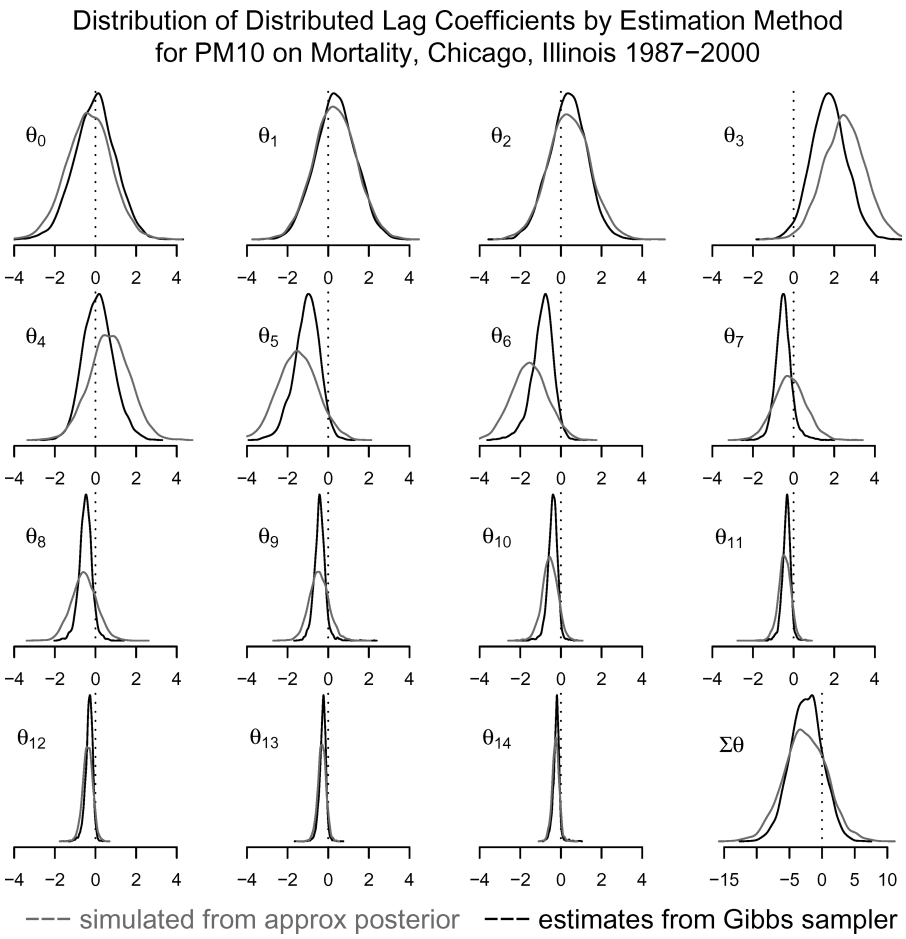

Figure 3. Comparison of estimation methods for DL coefficients of the effect of PM_{10} on mortality for Chicago, Illinois from 1987 to 2000 by estimation method. Distributions of DL coefficient estimates, by lag, and sum of DL coefficients (all in units of 10^{-4}) are shown for (i) the DL coefficient vector simulated from the normal approximate posterior distribution (gray) and (ii) the estimates of DL coefficients from last 4000 iterations of the Gibbs sampler (black).

We also show that our approach relates to using penalized splines to estimate DL functions. Specifically, fitting a penalized spline DLagM with a specific penalty matrix is analogous to using a BDLagM with a normal prior on the DL coefficients. An advantage of using the Bayesian approach is the simplicity of formulating a prior distribution on DL coefficients rather than specifying a penalty matrix.

Using the proposed BDLagM we estimated the association between lagged exposures of PM_{10} and mortality for Chicago, Illinois from 1987 to 2000. We found that the largest effect of PM_{10} on mortality occurs at lag 3 and that the total effect is equal to -0.21% (95% PI -0.86%, 0.41%). The shape of the DL function is consistent with mortality displacement.

For the Chicago data we found that the BDLagM estimated using the normal approximation to the likelihood (with a posterior distribution for $\boldsymbol{\theta}$ available in closed form) and the Poisson likelihood (with a Gibbs sampler) yielded similar estimates for the total effect and for early and later DL coefficients. The relatively large number of daily deaths in Chicago (on average, 116) as well as the length of the time series may account for the agreement between the two methods. For applications with outcome distributions that are not approximately normal, we anticipate less agreement between the two estimation methods and that the normal approximate posterior would be a less efficient proposal distribution.

The BDLagM formulated for a single city time series study may be naturally extended to a multicity framework. Multicity studies of mortality and air pollution use hierarchical models to pool individual city relative risks across multiple cities or counties, and have provided strong evidence for the association between air pollution and mortality (Zmirou et al., 1988; Burnett et al., 1998; Schwartz, 2000, Katsouyanni et al., 2001; Samoli et al., 2001; Zanobetti et al., 2002, 2003; Dominici et al., 2003). The hierarchical models used to date have estimated risk for single lag PM exposures or the total effect, which may not fully describe the relationship between short-term health risk and air pollution exposure. Estimating our BDLagM for multiple cities in a hierarchical model of an overall DL function between air pollution and mortality would provide additional understanding of the relationship between air pollution and health (Peng, Dominici, and Welty, 2007).

A challenge to estimating our BDLagMs for multiple cities is missing data. For many U.S. cities, PM air pollution is measured 1 in every 6 days. Before estimating the outlined

BDLagMs for multicity studies, it will be necessary to develop a version that estimates the DL coefficients in the presence of missing data. Accounting for missingness in the exposure series would expand the applicability of the proposed BDLagMs.

Given the equivalence between estimating DL functions using a penalized spline and putting a prior directly on the DL coefficients, our Bayesian method may be viewed as a means for eliciting a penalty matrix. P-spline penalties can be interpreted as the size of jumps of a smooth's third or higher derivatives, which may be difficult to relate to biological or other prior knowledge. Our method may be viewed as a transparent or intuitive means for eliciting penalties that are consistent with prior knowledge of the objective function. Our approach is not limited to functions that increase in smoothness as they approach zero; it could also be applied, for instance, to monotonic functions. However, given the necessity of choosing a value for $\sigma^2 = \text{Var}(\theta_0)$, it could be imprudent to use our approach to estimate DL functions about which there is no prior knowledge about the range of θ_0.

7. Supplementary Materials

Web Tables and Figures referenced in Sections 4 and 5 are available under the Paper Information link on the *Biometrics* website at `http://www.biometrics.tibs.org`.

Acknowledgements

Funding for the authors was provided by NIEHS RO1 grant (ES012054-01), and by NIEHS Center in Urban Environmental Health (P30 ES 03819).

References

Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica* **33,** 178–196.

Bell, M. L., Samet, J. M., and Dominici, F. (2004). Time-series studies of particulate matter. *Annual Review of Public Health* **25,** 247–280.

Bell, M. L., McDermott, A., Zeger, S. L., Samet, J. M., and Dominici, F. (2004). Ozone and short-term mortality in 95 US urban communities, 1987–2000. *Journal of the American Medical Association* **19,** 2372–2378.

Braga, F., Zanobetti, A., and Schwartz, J. (2000). Do respiratory epidemics confound the association between air pollution and daily deaths? *European Respiratory Journal* **16,** 723–728.

Braga, F., Luis, A., Zanobetti, A., and Schwartz, J. (2001). The time course of weather-related deaths. *Epidemiology* **12,** 662–667.

Burnett, R., Cakmak, S., and Brook, J. (1998). The effect of the urban ambient air pollution mix on daily mortality rates in 11 Canadian cities. *Canadian Journal of Public Health* **89,** 152–156.

Carlin, B. P. and Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. Boca Raton, Florida: Chapman and Hall.

Corradi, C. (1977). Smooth distributed lag estimators and smoothing spline functions in Hilbert spaces. *Journal of Econometrics* **5,** 211–220.

Dominici, F., McDermott, A., Daniels, M. J., Zeger, S. L., and Samet, J. M. (2003). *Revised Analysis of the National Morbidity Mortality Air Pollution Study: Part II*. Cambridge, Massachusetts: The Health Effects Institute.

Dominici, F., McDermott, A., and Hastie, T. (2004). Improved semi-parametric time series models of air pollution and mortality. *Journal of the American Statistical Association* **468,** 938–948.

Eilers, P. and Marx, B. (1996). Flexible smoothing with b-splines and penalties. *Statistical Science* **1,** 89–121.

Fahrmeir, L. and Knorr-Held, L. (1997). Dynamic discrete-time duration models: Estimation via Markov chain Monte Carlo. *Sociological Methodology* **27,** 417–452.

Goodman, P. G., Dockery, D. W., and Clancy, L. (2004). Cause-specific mortality and the extended effects of particulate pollution and temperature exposure. *Environmental Health Perspectives* **112,** 179–185.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*. Boca Raton, Florida: Chapman and Hall.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society, Series B* **4,** 757–796.

Katsouyanni, K., Toulomi, G., Samoli, E., Gryparis, A., Le Tertre, A., Monopolis, Y., Rossi, G., Zmirou, D., Ballester, F., Boumghar, A., Anderson, H. R., Wojtyniak, B., Paldy, A., Braunstein, R., Pekkanen, J., Schindler, C., and Schwartz, J. (2001). Confounding and effect modification in the short-term effects of ambient particles on total mortality: Results from 29 European cities within the APHEA2 project. *Epidemiology* **12,** 521–531.

Kim, H., Kim, Y., and Hong, Y. (2003). The lag-effect pattern in the relationship of particulate air pollution to daily mortality in Seoul, Korea. *International Journal of Biometeorology* **48,** 25–30.

Leamer, E. E. (1972). A class of informative priors and distributed lag analysis. *Econometrica* **40,** 1059–1081.

Manda, S. and Meyer, R. (2005). Age at first marriage in Malawi: A Bayesian multilevel analysis using a discrete time-to-event model. *Journal of the Royal Statistical Society, Series A* **168,** 439–455.

Peng, R. D., Dominici, F., and Louis, T. A. (2005). Model choice in time series studies of air pollution and mortality. *Journal of the Royal Statistical Society, Series A* **169,** 179–203.

Peng, R. D., Dominici, F., and Welty, L. J. (2007). *A Bayesian hierarchical model for constrained distributed lag functions: Estimating the time course of hospitilization associated with air pollution exposure*. Technical Report 128, Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland.

Pope, C. A. and Schwartz, J. (1996). Time series for the analysis of pulmonary health data. *American Journal of Respiratory and Critical Care Medicine* **154,** S229–S233.

Pope, C. A. R., Dockery, D. W., Spengler, J. D., and Raizenne, M. E. (1991). Respiratory health and pm10 pollution. A daily time series analysis. *American Review of Respiratory Diseases* **144,** 668–674.

Ravines, R. R., Schmidt, A. M., and Migon, H. S. (2006). Revisiting distributed lag models through a Bayesian perspective. *Applied Stochastic Models in Business and Industry* **22,** 193–210.

Roberts, S. (2005). An investigation of distributed lag models in the context of air pollution and mortality time series analysis. *Journal of the Air and Waste Management Association* **55,** 273–282.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge, U.K.: Cambridge University Press.

Samet, J., Zeger, S., Kelsall, J., Xu, J., and Kalkstein, L. (1998). Does weather confound or modify the association of particulate air pollution with mortality. *Environmental Research* **77,** 9–19.

Samet, J. M., Zeger, S. L., Dominici, F., Curriero, F., Dockery, D. W., Schwartz, J., and Zanobetti, A. (2000). *The National Morbidity Mortality Air Pollution Study: Part II*. Cambridge, Massachusetts: The Health Effects Institute.

Samet, J. M., Zeger, S. L., Dominici, F., Schwartz, J., and Dockery, D. W. (2000). *The National Morbidity Mortality Air Pollution Study: Part I*. Cambridge, Massachusetts: The Health Effects Institute.

Samoli, E., Schwartz, J., Wojtyniak, B., et al. (2001). Investigating regional differences in short-term effects of air pollution on daily mortality in the APHEA project: A sensitivity analysis for controlling long-term trends and seasonality. *Environmental Health Perspectives* **109,** 349–353.

Schiller, R. J. (1973). A distributed lag estimator derived from smoothness priors. *Econometrica* **41,** 775–788.

Schimmel, B. and Murawsky, T. (1978). The relation of air pollution to mortality. *Journal of Occupational Medicine* **18,** 316–333.

Schwartz, J. (1993). Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions. *American Journal of Epidemiology* **137,** 1136–1147.

Schwartz, J. (2000). The distributed lag between air pollution and daily deaths. *Epidemiology* **11,** 320–326.

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society, Series B* **47,** 1–52.

Welty, L. J. and Zeger, S. L. (2005). Are the acute effects of particulate matter on mortality in the National Morbidity, Mortality, and Air Pollution study the result of inadequate control for weather and season? A sensitivity analysis using flexible distributed lag models. *American Journal of Epidemiology* **162,** 80–88.

Zanobetti, A., Wand, M. P., Schwartz, J., et al. (2000). Generalized additive distributed lag models: Quantifying mortality displacement. *Biostatistics* **1,** 279–292.

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

Zanobetti, A., Schwartz, J., Samoli, E., Gryparis, A., Touloumi, G., Peacock, J., Anderson, R. H., LeTertre, A., Bobros, J., Celko, M., Goren, A., Forsberg, B., Michelozzi, P., Rabczenko, D., Hoyos, S. P., Wichmann, H. E., and Katsouyanni K. (2003). The temporal pattern of respiratory and heart disease mortality in response to air pollution. *Environmental Health Perspectives* **111,** 1188–1193.

Zmirou, D., Schwartz, J., Saez, M., Zanobetti, A., Wojtyniak, B., Touloumi, G., Spix, C., Ponce de León, A., Le Moullec, Y., Bacharova, L., Schouten, J., Pönkä, A., and Katsouyanni, K. (1988). Time-series analysis of air pollution and cause-specific mortality. *Epidemiology* **9,** 495–503.

Received December 2005. *Revised January* 2008.
Accepted January 2008.