# Supplementary Appendix for

"Robust Distributed Lag Models via Data Adaptive Shrinkage"

**Yin-Hsiu Chen**[1], **Bhramar Mukherjee**[1]

**Sara D. Adar**[2], **Veronica J. Berrocal**[1], **Brent A. Coull**[3]

[1]Department of Biostatistics, University of Michigan

[2]Department of Epidemiology, University of Michigan

[3]Department of Biostatistics, Harvard University

## A.1 Connection Between $\boldsymbol{C}$ and $\boldsymbol{R}$ beyond Polynomial DLM

Denote

$\boldsymbol{C}$: a $(L+1)\times p$ transformation matrix
$\boldsymbol{R}$: a $(L+1-p)\times(L+1)$ constraint matrix
$\boldsymbol{C}_e$: a $(L+1)\times(L+1)$ matrix $[\boldsymbol{C}\ \mathbf{0}_{(L+1)\times(L+1-p)}]$ where $\mathbf{0}_{(L+1)\times(L+1-p)}$ is a $(L+1)\times(L+1-p)$ zero matrix
$\boldsymbol{R}_e$: a $(L+1)\times(L+1)$ matrix $\begin{bmatrix}\boldsymbol{R}\\ \mathbf{0}_{p\times(L+1)}\end{bmatrix}$ where $\mathbf{0}_{p\times(L+1)}$ is a $p\times(L+1)$ zero matrix

(1) $\boldsymbol{R}\to\boldsymbol{C}$
The $p$ basis functions corresponding to the $p$ columns of $\boldsymbol{C}$ span the solution space of $\boldsymbol{R}\boldsymbol{\beta}=0$ (or $\boldsymbol{R}_e\boldsymbol{\beta}=0$). $\boldsymbol{C}$ can be obtained by applying SVD on $\boldsymbol{R}_e$ (i.e. $\boldsymbol{R}_e=\boldsymbol{U}_R\boldsymbol{D}_R\boldsymbol{V}_R^T$). The last $p$ columns of $\boldsymbol{V}_R$ is one choice of $\boldsymbol{C}$.

(2) $\boldsymbol{C}\to\boldsymbol{R}$
$\boldsymbol{\beta}=\boldsymbol{C}\boldsymbol{\eta}$ and $\boldsymbol{R}\boldsymbol{\beta}=0$ so we have $\boldsymbol{R}\boldsymbol{C}\boldsymbol{\eta}=0$. Deriving $\boldsymbol{R}$ from $\boldsymbol{C}$ is equivalent of solving $\boldsymbol{C}^T\boldsymbol{R}^T=0$ (or $\boldsymbol{C}_e^T\boldsymbol{R}^T=0$). $\boldsymbol{R}$ can be obtained by applying SVD on $\boldsymbol{C}_e^T$ (i.e. $\boldsymbol{C}_e^T=\boldsymbol{U}_C\boldsymbol{D}_C\boldsymbol{V}_C^T$). The last $(L+1-p)$ rows of $\boldsymbol{V}_C^T$ is one choice of $\boldsymbol{R}$.

__Remark__: DLM solution is invariant to row operations on $\boldsymbol{R}$. For example, consider a piecewise linear distributed lag function with $L=6$ and only internal knot at 3. With basis functions 1, $\ell$, and $(\ell-3)_+$, $\boldsymbol{C}$ is given by

$$\boldsymbol{C}=\begin{bmatrix}1&0&0\\1&1&0\\1&2&0\\1&3&0\\1&4&1\\1&5&2\\1&6&3\end{bmatrix}.$$

Following the above procedure, $\boldsymbol{R}$ can be obtained as

$$\boldsymbol{R} = \begin{bmatrix} 0.000 & -0.346 & 0.693 & -0.030 & -0.255 & -0.439 & 0.377 \\ 0.000 & -0.218 & 0.436 & -0.290 & -0.238 & 0.691 & -0.381 \\ 0.000 & -0.090 & 0.179 & -0.549 & 0.779 & -0.180 & -0.139 \\ -0.560 & 0.727 & 0.226 & -0.275 & -0.157 & -0.039 & 0.079 \end{bmatrix}.$$

Through row operations, we can obtain

$$\boldsymbol{R} = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix}$$

as suggested. The solution of $\boldsymbol{R\beta} = \mathbf{0}$ is a piecewise linear function with internal knot at 3.

## A.2 Asymptotic Results for the Empirical Bayes estimator

We first derive the variance-covariance expression of $\hat{\boldsymbol{\psi}} = \hat{\boldsymbol{\beta}}_{UDLM} - \hat{\boldsymbol{\beta}}_{CDLM}$ and then obtain the asymptotic theory of $\hat{\boldsymbol{\beta}}_{EB1}$. Let $S_U^{(t)}(\boldsymbol{\beta})$ denote the first-order derivative of the unconstrained DLM likelihood for time $t$ (i.e. $(y_t - e^{-\boldsymbol{X}_t^T\boldsymbol{\beta}})\boldsymbol{X}_t$) $S_C^{(t)}(\boldsymbol{\theta})$ denote the first-order derivative of the constrained DLM likelihood for time $t$ (i.e. $(y_t - e^{-\boldsymbol{Z}_t^T\boldsymbol{\theta}})\boldsymbol{Z}_t$), and let $H_U(\boldsymbol{\beta})$ and $H_C(\boldsymbol{\theta})$ denote the Hessian matrices from the two models, respectively. Let $\boldsymbol{\beta}_0$ denote the true vector of lagged coefficients. By Taylor expansion,

$$\begin{aligned}
&\ell'(\hat{\boldsymbol{\beta}}_{UDLM}) = \ell'(\boldsymbol{\beta}_0) + \ell''(\boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}}_{UDLM} - \boldsymbol{\beta}_0) + o_p(|\hat{\boldsymbol{\beta}}_{UDLM} - \boldsymbol{\beta}_0|) \\
\Rightarrow &\hat{\boldsymbol{\beta}}_{UDLM} - \boldsymbol{\beta}_0 = [-\ell''(\boldsymbol{\beta}_0) + o_p(1)]^{-1}\ell'(\boldsymbol{\beta}_0) \\
\Rightarrow &\mathrm{Cov}(\hat{\boldsymbol{\beta}}_{UDLM}) = [-\ell''(\boldsymbol{\beta}_0)]^{-1}\mathrm{Cov}(\ell'(\boldsymbol{\beta}_0))[-\ell''(\boldsymbol{\beta}_0)]^{-1} + o_p(1) \\
\Rightarrow &[-\ell''(\boldsymbol{\beta}_0)]^{-1}\mathrm{Cov}(\ell'(\boldsymbol{\beta}_0))[-\ell''(\boldsymbol{\beta}_0)]^{-1} \xrightarrow{P} \mathrm{Cov}(\hat{\boldsymbol{\beta}}_{UDLM})
\end{aligned}$$

Since $\hat{\boldsymbol{\beta}}_{UDLM} \to \boldsymbol{\beta}_0$, $-H_U^{-1}(\hat{\boldsymbol{\beta}}_{UDLM})^{-1} \xrightarrow{P} [-\ell''(\boldsymbol{\beta}_0)]^{-1}$. Also, $\mathrm{Cov}(\ell'(\boldsymbol{\beta}_0))$ can be consistently estimated by empirical variance $\sum_{t=1}^T S_U^{(t)}(\hat{\boldsymbol{\beta}}_{UDLM})S_U^{(t)}(\hat{\boldsymbol{\beta}}_{UDLM})^T$. So,

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = H_U^{-1}(\hat{\boldsymbol{\beta}})[\sum_{t=1}^T S_U^{(t)}(\hat{\boldsymbol{\beta}})S_U^{(t)}(\hat{\boldsymbol{\beta}})^T]H_U^{-1}(\hat{\boldsymbol{\beta}})$$

Similarly,

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) = H_C^{-1}(\hat{\boldsymbol{\theta}})[\sum_{t=1}^T S_C^{(t)}(\hat{\boldsymbol{\theta}})S_C^{(t)}(\hat{\boldsymbol{\theta}})^T]H_C^{-1}(\hat{\boldsymbol{\theta}})$$

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}_{UDLM}, \hat{\boldsymbol{\theta}}) = H_U^{-1}(\hat{\boldsymbol{\beta}}_{UDLM})[\sum_{t=1}^T S_U^{(t)}(\hat{\boldsymbol{\beta}}_{UDLM})S_C^{(t)}(\hat{\boldsymbol{\theta}})^T]H_C^{-1}(\hat{\boldsymbol{\theta}})$$

Therefore,

$$
\begin{aligned}
\mathrm{Cov}(\hat{\boldsymbol{\psi}}) &= \mathrm{Cov}(\hat{\boldsymbol{\beta}}_{UDLM} - \hat{\boldsymbol{\beta}}_{CDLM}) \\
&= \mathrm{Cov}(\hat{\boldsymbol{\beta}}_{UDLM} - \boldsymbol{C}\hat{\boldsymbol{\theta}}) \\
&= \mathrm{Cov}(\hat{\boldsymbol{\beta}}_{UDLM}) - 2\mathrm{Cov}(\hat{\boldsymbol{\beta}}_{UDLM}, \hat{\boldsymbol{\theta}})\boldsymbol{C}^T + \boldsymbol{C}\mathrm{Cov}(\hat{\boldsymbol{\theta}})\boldsymbol{C}^T \\
&= H_U^{-1}(\hat{\boldsymbol{\beta}}_{UDLM})[\sum_{t=1}^{T} S_U^{(t)}(\hat{\boldsymbol{\beta}}_{UDLM})S_U^{(t)}(\hat{\boldsymbol{\beta}}_{UDLM})^T]H_U^{-1}(\hat{\boldsymbol{\beta}}_{UDLM}) \\
&- 2H_U^{-1}(\hat{\boldsymbol{\beta}}_{UDLM})[\sum_{t=1}^{T} S_U^{(t)}(\hat{\boldsymbol{\beta}}_{UDLM})S_C^{(t)}(\hat{\boldsymbol{\theta}})^T]H_C^{-1}(\hat{\boldsymbol{\theta}})\boldsymbol{C}^T \\
&+ \boldsymbol{C}H_C^{-1}(\hat{\boldsymbol{\theta}})[\sum_{t=1}^{T} S_C^{(t)}(\hat{\boldsymbol{\theta}})S_C^{(t)}(\hat{\boldsymbol{\theta}})^T]H_C^{-1}(\hat{\boldsymbol{\theta}})\boldsymbol{C}^T
\end{aligned}
$$

Let $\boldsymbol{\beta}_{CDLM}$, $\boldsymbol{\beta}_{EB1}$, and $\boldsymbol{\psi}$ be the asymptotic limit of constrained DLM estimator, EB1 estimator, and bias, respectively. We have

$$\boldsymbol{\beta}_{EB1} = \boldsymbol{\beta}_0 + \boldsymbol{K}\boldsymbol{\psi}$$

where $\boldsymbol{K} = (\boldsymbol{V} \circ \boldsymbol{I}_{L+1})[(\boldsymbol{V} + \boldsymbol{\psi}\boldsymbol{\psi}^T) \circ \boldsymbol{I}_{L+1}]^{-1}$. When $\boldsymbol{\beta}_0 \neq \boldsymbol{\beta}_{CDLM}$ ($\boldsymbol{\psi} \neq \boldsymbol{0}$), we can use first-order Taylor expansion of $\hat{\boldsymbol{\beta}}_{EB1}$ at $(\boldsymbol{\beta}_0^T, \boldsymbol{\beta}_{CDLM}^T)^T$ and the fact that $\boldsymbol{V} = O_p(N^{-1})$ to obtain

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{EB1} - \boldsymbol{\beta}_{EB1}) = \boldsymbol{G} \times \sqrt{N}[(\hat{\boldsymbol{\beta}}_{UDLM}^T, \hat{\boldsymbol{\beta}}_{CDLM}^T)^T - (\boldsymbol{\beta}_0^T, \boldsymbol{\beta}_{CDLM}^T)^T] + o_p(1)$$

where

$$\boldsymbol{G} = [\mathrm{diag`}\{\frac{v_j(v_j - \psi_j^2)}{(v_j + \psi_j^2)^2}\}, \boldsymbol{I}_{L+1} - \mathrm{diag}\{\frac{v_j(v_j - \psi_j^2)}{(v_j + \psi_j^2)^2}\}]$$

where $v_j$s are the diagonal elements of $\boldsymbol{V}$ and $\psi_j$s are the elements of $\boldsymbol{\psi}$. Thus, $\hat{\boldsymbol{\beta}}_{EB1}$ is $\sqrt{N}$-consistent and asymptotically normal when $\boldsymbol{\beta}_0 \neq \boldsymbol{\beta}_{CDLM}$. Let $\hat{\boldsymbol{\Sigma}}$ denote the estimated variance-covariance matrix of $(\hat{\boldsymbol{\beta}}_{UDLM}^T, \hat{\boldsymbol{\beta}}_{CDLM}^T)^T$. With plug-in estimate of $\boldsymbol{G}$, the asymptotic variance of $\hat{\boldsymbol{\beta}}_{EB1}$ can be estimated as $\hat{\boldsymbol{G}}\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{G}}^T$.

## A.3 Equivalence of $(p-1)$-degree Polynomial DLM Estimator and GRR/HB Shrinkage Target Corresponding to $\boldsymbol{R}_{p-1}$

The general form of $(p-1)$-degree polynomial distributed lag function is $\beta(\ell) = \sum_{i=0}^{p-1} a_i \ell^i = \boldsymbol{C}\boldsymbol{\theta}$ for $\ell = 0, \cdots, L$ where $\boldsymbol{C}$ is a $(L+1) \times p$ transformation matrix with element $(\ell+1, j)$ equal to $\ell^{(j-1)}$ and $\boldsymbol{\theta} = (a_1, \cdots, a_p)^T$. Let $\boldsymbol{R}_{p-1}$ be the $(p-1)$-degree polynomial constraint matrix. We first show that $\boldsymbol{R}_{p-1}\boldsymbol{C} = \boldsymbol{0}$.

The corresponding constraints constructed in $\boldsymbol{R}_{p-1}$ are $\sum_{j=0}^{p}(-1)^j\binom{p}{j}\beta(\ell+j) = 0$ for $\ell = 0, ..., L-p$. Showing $\boldsymbol{R}_{p-1}\boldsymbol{C} = \boldsymbol{0}$ is the same as showing $\sum_{j=0}^{p}(-1)^j\binom{p}{j}[\sum_{i=1}^{p} a_i(\ell +$

$j)^{i-1}] = 0$ for $\ell = 0, ..., L-p$.

$$
\begin{aligned}
&\sum_{j=0}^{p}(-1)^j \binom{p}{j} [\sum_{i=1}^{p} a_i(\ell+j)^{i-1}] \\
=&\sum_{i=1}^{p}\sum_{j=0}^{p}(-1)^j \binom{p}{j} a_i(\ell+j)^{i-1} \\
=&\sum_{i=1}^{p} a_i[\sum_{j=0}^{p}(-1)^j \binom{p}{j} (\ell+j)^{i-1}]
\end{aligned}
$$

It is sufficient to show that $\sum_{j=0}^{p}(-1)^j \binom{p}{j}(\ell+j)^{i-1} = 0$ for $\ell = 0, ..., L-p$ and $1 \leq i \leq p$. It is well-known that each polynomial can be uniquely expressed as a linear combination of binomial coefficients. $\sum_{j=0}^{p}(-1)^j \binom{p}{j}(\ell+j)^{i-1} = 0$ corresponds to the binomial coefficient involved $(p-1)$-degree term of the characteristic polynomial $(\ell+j)^{i-1}$. We know that $i$ is at most $p$ so the coefficients of all the terms of degree larger than $p-1$ must be zero so we have $\sum_{j=0}^{p}(-1)^j \binom{p}{j}(\ell+j)^{i-1} = 0$ for $\ell = 0, ..., L-p$ and $1 \leq i \leq p$. Therefore, $\boldsymbol{R}_{p-1}\boldsymbol{C} = \boldsymbol{0}$.

The shrinkage target of GRR/HB estimator corresponding to $\boldsymbol{R}_{p-1}$ is the maximizer of likelihood function in (1) of main text subject to the constraint $\boldsymbol{R}_{p-1}\boldsymbol{\beta} = \boldsymbol{0}$. Let $\hat{\boldsymbol{\beta}}$ denote the GRR/HB estimator. Since $\hat{\boldsymbol{\beta}}$ conforms to the constraint, we have $\boldsymbol{R}_{p-1}\hat{\boldsymbol{\beta}} = \boldsymbol{0}$ and $\hat{\boldsymbol{\beta}}$ is an element in the kernel of $\boldsymbol{R}_{p-1}$. From above, we have $\boldsymbol{R}_{p-1}\boldsymbol{C} = \boldsymbol{0}$ and we know that the $p$ columns of $\boldsymbol{C}$ are linearly independent. Thus, the kernel of $\boldsymbol{R}_{p-1}$ is spanned by the $p$ columns of $\boldsymbol{C}$. Subsequently, every element in the kernel can be expressed as $\boldsymbol{C}\boldsymbol{\theta}$ so $\hat{\boldsymbol{\beta}}$ must be in the form of $\boldsymbol{C}\boldsymbol{\theta}$. Therefore, the maximizing the likelihood function in (1) in terms of $\boldsymbol{\beta}$ subject to the constraint $\boldsymbol{R}_{p-1}\boldsymbol{\beta} = \boldsymbol{0}$ is equivalent to maximizing the likelihood function in (2) in terms of $\boldsymbol{\theta}$ without any constraint. We then conclude that $(p-1)$-degree polynomial DLM estimator and GRR/HB shrinkage target corresponding to $\boldsymbol{R}_{p-1}$ are equivalent.

### A.4 Conditional Distributions of HB Estimator and Two-stage Shrinkage Estimator

The full conditional distributions of $\sigma_\pi^2$ and $\boldsymbol{\beta}$ for HB estimator are given by

$$
\begin{aligned}
f(\sigma_\pi^2|\boldsymbol{\beta}, \boldsymbol{Y}) &\propto IG(a_\pi + M/2, b_\pi + \boldsymbol{\beta}^T\boldsymbol{R}^T\boldsymbol{R}\boldsymbol{\beta}/2) \\
f(\boldsymbol{\beta}|\sigma_\pi^2, \boldsymbol{Y}) &\propto \prod_{t=1}^{T}[\exp(y_t\boldsymbol{X}_t^T\boldsymbol{\beta} - e^{\boldsymbol{X}_t^T\boldsymbol{\beta}})] \cdot \exp(-\frac{\boldsymbol{\beta}^T\boldsymbol{R}^T\boldsymbol{R}\boldsymbol{\beta}}{2\sigma_\pi^2}).
\end{aligned}
$$

For two-stage shrinkage approach, if we let $\boldsymbol{\omega} = (\omega_1, \omega_2)^T$ have a discrete uniform prior distribution, the full conditional distributions of $\boldsymbol{\beta}, \sigma^2$, and $\boldsymbol{\omega}$ are given by:

$$f(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}}_{EB}, \boldsymbol{\omega}, \sigma^2) \sim N([1/\sigma^2\boldsymbol{\Omega}(\boldsymbol{\omega})^{-1} + (\boldsymbol{G\Sigma G}^T)^{-1}]^{-1}(\boldsymbol{G\Sigma G}^T)^{-1}\hat{\boldsymbol{\beta}}_{EB}, [1/\sigma^2\boldsymbol{\Omega}(\boldsymbol{\omega})^{-1} + (\boldsymbol{G\Sigma G})^{-1}]^{-1})$$

$$p(\boldsymbol{\omega}|\hat{\boldsymbol{\beta}}_{EB}, \boldsymbol{\beta}, \sigma^2) = \frac{|\boldsymbol{\Omega}(\boldsymbol{\omega})|^{-1/2}\text{exp}[-\frac{1}{2\sigma^2}\boldsymbol{\beta}^T\boldsymbol{\Omega}(\boldsymbol{\omega})^{-1}\boldsymbol{\beta}]}{\sum_{\boldsymbol{\omega}^*}|\boldsymbol{\Omega}(\boldsymbol{\omega}^*)|^{-1/2}\text{exp}[-\frac{1}{2\sigma^2}\boldsymbol{\beta}^T\boldsymbol{\Omega}(\boldsymbol{\omega}^*)^{-1}\boldsymbol{\beta}]}$$

$$f(\sigma^2|\hat{\boldsymbol{\beta}}_{EB}, \boldsymbol{\beta}, \boldsymbol{\omega}) \sim IG(a_0 + (L+1)/2, b_0 + \boldsymbol{\beta}^T\boldsymbol{\Omega}(\boldsymbol{\omega})^{-1}\boldsymbol{\beta}/2).$$

## A.5 Analytical Results for the GRR Estimator

GRR estimator $\hat{\boldsymbol{\beta}}_{GRR}$ is given by

$$\hat{\boldsymbol{\beta}}_{GRR} = \arg\min_{\boldsymbol{\beta}}[-\sum_{t=1}^{T}[y_t\boldsymbol{\beta}^T\boldsymbol{X_t} - e^{\boldsymbol{\beta}^T\boldsymbol{X_t}} - \log(y_t!)] + \lambda\boldsymbol{\beta}^T\boldsymbol{R}^T\boldsymbol{R}\boldsymbol{\beta}]$$

and its asymptotic MSE $E[(\hat{\boldsymbol{\beta}}_{GRR} - \boldsymbol{\beta})^T(\hat{\boldsymbol{\beta}}_{GRR} - \boldsymbol{\beta})]$ can be decomposed into $f_1(\lambda) + f_2(\lambda)$ where

$$f_1(\lambda) = E[(\hat{\boldsymbol{\beta}}_{UDLM} - \boldsymbol{\beta})^T\boldsymbol{H}^T\boldsymbol{H}(\hat{\boldsymbol{\beta}}_{UDLM} - \boldsymbol{\beta})] = \text{trace}[(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-2}]$$

$$f_2(\lambda) = (\boldsymbol{H\beta} - \boldsymbol{\beta})^T(\boldsymbol{H\beta} - \boldsymbol{\beta}) = \lambda^2\boldsymbol{\beta}^T(\boldsymbol{R}^T\boldsymbol{R})(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-2}(\boldsymbol{R}^T\boldsymbol{R})\boldsymbol{\beta}.$$

We first show that $f_1(\lambda)$ is monotonic decreasing and $f_2(\lambda)$ is monotonic increasing and then we show that $f_1(\lambda) + f_2(\lambda)$ is convex.

$$\begin{aligned}
& df_1(\lambda) \\
=& d\text{trace}[(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X})(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-2}] \\
=& \text{trace}[2(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-1}(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X})d(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-1}] \\
=& \text{trace}\{-2(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-1}(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X})(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-1}[d(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})] \\
& \quad (\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X})(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-1}\} \\
=& \text{trace}[-2\boldsymbol{R}^T\boldsymbol{R}(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X})(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-3}]d\lambda
\end{aligned}$$

Since $\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X}$ and $\boldsymbol{R}^T\boldsymbol{R}$ are positive definite and $\lambda > 0$, $\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R}$ is positive definite (Weyl's inequality). It follows that $2\boldsymbol{R}^T\boldsymbol{R}(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X})(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-3}$ is positive definite and $\text{trace}[-2\boldsymbol{R}^T\boldsymbol{R}(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X})(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-3}] < 0$. Therefore, we have shown that $f_1(\lambda)$ is monotonic decreasing ($f_1^{'}(\lambda) < 0 \ \forall\lambda > 0$).

Assume $\lambda_2 > \lambda_1 > 0$,

$$
\begin{aligned}
& f_2(\lambda_2) - f_2(\lambda_1) \\
=& \lambda_2^2 \boldsymbol{\beta}^T(\boldsymbol{R}^T\boldsymbol{R})(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda_2\boldsymbol{R}^T\boldsymbol{R})^{-2}(\boldsymbol{R}^T\boldsymbol{R})\boldsymbol{\beta} - \lambda_1^2\boldsymbol{\beta}^T(\boldsymbol{R}^T\boldsymbol{R})(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda_1\boldsymbol{R}^T\boldsymbol{R})^{-2}(\boldsymbol{R}^T\boldsymbol{R})\boldsymbol{\beta} \\
=& \boldsymbol{\beta}^T(\boldsymbol{R}^T\boldsymbol{R})[(\frac{1}{\lambda_2}\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \boldsymbol{R}^T\boldsymbol{R})^{-2} - (\frac{1}{\lambda_1}\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \boldsymbol{R}^T\boldsymbol{R})^{-2}](\boldsymbol{R}^T\boldsymbol{R})\boldsymbol{\beta} \\
=& \boldsymbol{\beta}^T(\boldsymbol{R}^T\boldsymbol{R})(\frac{1}{\lambda_2}\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \boldsymbol{R}^T\boldsymbol{R})^{-2}[(\frac{1}{\lambda_1}\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \boldsymbol{R}^T\boldsymbol{R})^2 - (\frac{1}{\lambda_2}\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \boldsymbol{R}^T\boldsymbol{R})^2] \\
& (\frac{1}{\lambda_1}\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \boldsymbol{R}^T\boldsymbol{R})^{-2}(\boldsymbol{R}^T\boldsymbol{R})\boldsymbol{\beta} \\
=& \boldsymbol{\beta}^T(\boldsymbol{R}^T\boldsymbol{R})(\frac{1}{\lambda_2}\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \boldsymbol{R}^T\boldsymbol{R})^{-2}[(\frac{1}{\lambda_1^2} - \frac{1}{\lambda_2^2})(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X})^2 + (\frac{1}{\lambda_1} - \frac{1}{\lambda_2})(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X})(\boldsymbol{R}^T\boldsymbol{R}))] \\
& (\frac{1}{\lambda_1}\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \boldsymbol{R}^T\boldsymbol{R})^{-2}(\boldsymbol{R}^T\boldsymbol{R})\boldsymbol{\beta} \\
=& \text{trace}\{\boldsymbol{\beta}\boldsymbol{\beta}^T(\boldsymbol{R}^T\boldsymbol{R})^2(\frac{1}{\lambda_2}\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \boldsymbol{R}^T\boldsymbol{R})^{-2}[(\frac{1}{\lambda_1^2} - \frac{1}{\lambda_2^2})(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X})^2 + (\frac{1}{\lambda_1} - \frac{1}{\lambda_2})(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X})(\boldsymbol{R}^T\boldsymbol{R}))] \\
& (\frac{1}{\lambda_1}\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \boldsymbol{R}^T\boldsymbol{R})^{-2}\} \\
=& \text{trace}(\boldsymbol{A}) \\
=& \sum_{\ell=1}^{L+1} \alpha_\ell
\end{aligned}
$$

where $\gamma_\ell$ is the $\ell^{th}$ eigenvalue of $\boldsymbol{B}$. Since $\boldsymbol{\beta}\boldsymbol{\beta}^T$, $\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X}$, and $\boldsymbol{R}^T\boldsymbol{R}$ are positive definite and $\lambda_2 > \lambda_1$, all of the terms that $\boldsymbol{A}$ is composed of are positive definite and so is $\boldsymbol{A}$. Hence, $f_2(\lambda_2) - f_2(\lambda_1) = \sum_{\ell=1}^{L+1} \alpha_\ell > 0$. Therefore, we have shown that $f_2(\lambda)$ is monotonic increasing.

$$
\begin{aligned}
& f_2^{'}(\lambda) \\
=& 2\lambda\boldsymbol{\beta}^T(\boldsymbol{R}^T\boldsymbol{R})(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-2}(\boldsymbol{R}^T\boldsymbol{R})\boldsymbol{\beta} - \lambda^2\text{trace}[2(\boldsymbol{R}^T\boldsymbol{R})^3\boldsymbol{\beta}\boldsymbol{\beta}^T(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-3}]
\end{aligned}
$$

$$
\begin{aligned}
& f_1^{''}(\lambda) + f_2^{''}(\lambda) \\
=& \frac{d\text{trace}[-2\boldsymbol{R}^T\boldsymbol{R}(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X})(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-3}]}{d\lambda} + 2\lambda\boldsymbol{\beta}^T(\boldsymbol{R}^T\boldsymbol{R})(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-2}(\boldsymbol{R}^T\boldsymbol{R})\boldsymbol{\beta} \\
& + 2\lambda\frac{d[\boldsymbol{\beta}^T(\boldsymbol{R}^T\boldsymbol{R})(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-2}(\boldsymbol{R}^T\boldsymbol{R})\boldsymbol{\beta}]}{d\lambda} - 2\lambda\text{trace}[2(\boldsymbol{R}^T\boldsymbol{R})^3\boldsymbol{\beta}\boldsymbol{\beta}^T(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-3}] \\
& - \lambda^2\frac{d[2(\boldsymbol{R}^T\boldsymbol{R})^3\boldsymbol{\beta}\boldsymbol{\beta}^T(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-3}]}{d\lambda} \\
=& \text{trace}[6(\boldsymbol{R}^T\boldsymbol{R})^2(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X})(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-4}] + 2\lambda\boldsymbol{\beta}^T(\boldsymbol{R}^T\boldsymbol{R})(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-2}(\boldsymbol{R}^T\boldsymbol{R})\boldsymbol{\beta} \\
& + 2\lambda\text{trace}[2(\boldsymbol{R}^T\boldsymbol{R})^3\boldsymbol{\beta}\boldsymbol{\beta}^T(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-3}] - 2\lambda\text{trace}[2(\boldsymbol{R}^T\boldsymbol{R})^3\boldsymbol{\beta}\boldsymbol{\beta}^T(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-3}] \\
& + \lambda^2\text{trace}[6(\boldsymbol{R}^T\boldsymbol{R})^4\boldsymbol{\beta}\boldsymbol{\beta}^T(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-4}] \\
=& \text{trace}[6(\boldsymbol{R}^T\boldsymbol{R})^2(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X})(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-4}] + 2\lambda\boldsymbol{\beta}^T(\boldsymbol{R}^T\boldsymbol{R})(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-2}(\boldsymbol{R}^T\boldsymbol{R})\boldsymbol{\beta} \\
& + \lambda^2\text{trace}[6(\boldsymbol{R}^T\boldsymbol{R})^4\boldsymbol{\beta}\boldsymbol{\beta}^T(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X} + \lambda\boldsymbol{R}^T\boldsymbol{R})^{-4}] > 0
\end{aligned}
$$

Therefore, we have shown that $f_1(\lambda) + f_2(\lambda)$ is convex.

## A.6 Simulation Settings and Evaluation Metrics

Table 1: Summary of the three simulation scenarios for comparing UDLM, CDLM, EB1, EB2, GRR, GADLM, BDLM, and HB in simulation study 1.

| Scenario | Working DL Function* | True Distributed Lag Coefficients |
|---|---|---|
| (1) Working DL Function Completely Matches True DL Function | Cubic | $\beta_j = (j^3 - 17j^2 + 70j)/400$ for $j = 0, ..., 10$ |
| (2) Working DL Function Moderately Departs from True DL Function | Quadratic | Slight Departure from $\beta_j = (-0.7j^2 + 2.3j + 50.8)/400$ for $j = 0, ..., 10$ |
| (3) Non-smooth True DL Function | Quadratic | Oscillating between 0.02 and 0.18 |
| (4) Mixture of 5 True DL Function | Cubic | (a) $\beta_j = 0$ for $j = 0, \cdots, 10$<br>(b) $\beta_j = 0.014(10 - j)$ for $j = 0, \cdots, 10$<br>(c) Same as (1)<br>(d) Same as (2)<br>(e) Same as (3) |

*The working distributed lag (DL) function in CDLM for CDLM, EB1, EB2, GRR, and HB.

Table 2: Metrics used for evaluating the estimation precision in simulation study 1.

| Metric | Lag Effects Vector ($\boldsymbol{\beta}$) | Total Effect ($\sum_{j=0}^{10} \beta_j$) |
|---|---|---|
| Squared bias | $(\hat{\bar{\boldsymbol{\beta}}} - \boldsymbol{\beta})^T(\hat{\bar{\boldsymbol{\beta}}} - \boldsymbol{\beta})$ | $[\sum_{j=0}^{10}(\hat{\bar{\beta}}_j - \beta_j)]^2$ |
| Variance | trace$[\frac{1}{1000}\sum_{i=1}^{1000}(\hat{\boldsymbol{\beta}}_i - \hat{\bar{\boldsymbol{\beta}}})(\hat{\boldsymbol{\beta}}_i - \hat{\bar{\boldsymbol{\beta}}})^T]$ | $\frac{1}{1000}\sum_{i=1}^{1000}(\sum_{j=0}^{10}\hat{\beta}_{ij} - \sum_{j=0}^{10}\hat{\bar{\beta}}_j)^2$ |
| Relative Efficiency[1] | $\frac{\sum_{i=1}^{1000}\|\|\hat{\boldsymbol{\beta}}_i^{UDLM} - \boldsymbol{\beta}\|\|_2^2}{\sum_{i=1}^{1000}\|\|\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}\|\|_2^2}$ | $\frac{\sum_{i=1}^{1000}(\sum_{j=0}^{10}\hat{\beta}_{ij}^{UDLM} - \sum_{j=0}^{10}\beta_j)^2}{\sum_{i=1}^{1000}(\sum_{j=0}^{10}\hat{\beta}_{ij} - \sum_{j=0}^{10}\beta_j)^2}$ |
| Distance[2] | $\frac{1}{1000}\sum_{i=1}^{1000}\|\|\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}\|\|_2$ | - |

[1]Relative efficiency with respect to UDLM in terms of mean squared errors (MSE)
[2]Mean distance to the true coefficient vector $\boldsymbol{\beta}$
*$\hat{\boldsymbol{\beta}}_i = (\hat{\beta}_{i0}, ..., \hat{\beta}_{i10})^T$: the estimated lag coefficients from the $i^{th}$ data set for a particular method
**$\hat{\bar{\boldsymbol{\beta}}} = \frac{1}{1000}\sum_{i=1}^{1000}\hat{\boldsymbol{\beta}}_i$ and $\hat{\bar{\beta}}_j = \frac{1}{1000}\sum_{i=1}^{1000}\hat{\beta}_{ij}$ for $j = 0, ..., 10$.

## A.7 Quantification of the Uncertainty Ignored by Fixing Tuning Parameter GRR

Table 3: Average of the 1000 estimated variances as a percentage of the empirical variance of the 1000 estimates from 1000 repetitions for the 11 cumulative lag coefficient estimates based on GRR across the three scenarios in simulation study 1.

| Percentage | Scenario 1 | Scenario 2 | Scenario 3 |
|---|---|---|---|
| Lag 0 | 0.91 | 0.83 | 0.96 |
| Lag 1 | 0.97 | 0.88 | 0.93 |
| Lag 2 | 0.93 | 0.96 | 0.94 |
| Lag 3 | 0.92 | 0.92 | 1.02 |
| Lag 4 | 0.92 | 0.90 | 0.95 |
| Lag 5 | 0.93 | 0.99 | 0.89 |
| Lag 6 | 0.93 | 0.97 | 0.91 |
| Lag 7 | 0.95 | 0.95 | 0.93 |
| Lag 8 | 0.97 | 0.95 | 0.95 |
| Lag 9 | 0.95 | 0.95 | 0.91 |
| Lag 10 | 0.96 | 0.98 | 0.95 |

## A.8 NMMAPS Analysis

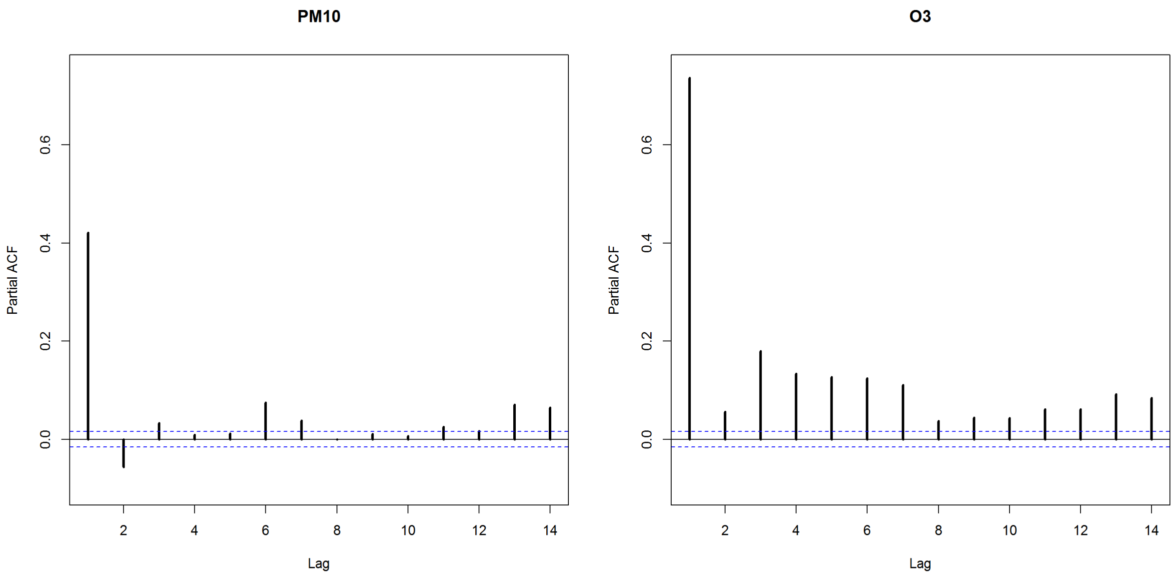


Figure 1: Partial autocorrelation function (PACF) plots for daily measurements of $PM_{10}$ (left) and $O_3$ (right) in Chicago, Illinois from 1987 to 2000 based on the National Morbidity, Mortality and Air Pollution Study (NMMAPS) data.

Table 4: Estimated mean and 95% confidence/credible interval of the cumulative lagged effect (% change in mortality count) up to 3, 7, and 14 days of $PM_{10}$ (upper) and $O_3$ (lower) on mortality with an interquartile range increase in exposure level ($PM_{10}$: $21.49 \mu g/m^3$, $O_3$: 14.65 ppb) in Chicago, Illinois from 1987 to 2000 based on the data from the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) under eight estimation methods.

| $PM_{10}$ | Up to Lag 3 (95% CI[1]) | Up to Lag 7 (95% CI[1]) | Up to Lag 14 (95% CI[1]) |
|---|---|---|---|
| UDLM | 0.75 (-0.01, 1.52) | 0.32 (-0.71, 1.36) | -0.75 (-2.24, 0.75) |
| CDLM | 0.51 (-0.22, 1.23) | 0.40 (-0.61, 1.40) | -0.87 (-2.36, 0.62) |
| EB1 | 0.81 (0.07, 1.56) | 0.41 (-0.54, 1.37) | -0.71 (-1.98, 0.56) |
| GRR | 0.67 (-0.06, 1.40) | 0.21 (-0.80, 1.22) | -0.74 (-2.23, 0.75) |
| BDLM | 0.57 (-0.24, 1.43) | -0.01 (-1.23, 1.16) | -1.05 (-2.76, 0.69) |
| HB | 0.63 (0.14, 1.12) | 0.26 (-0.41, 0.94) | -0.72 (-1.60, 0.15) |
| HB2-GRR | 0.48 (-0.21, 1.18) | 0.14 (-0.84, 1.12) | -0.43 (-1.91, 1.05) |
| HP-GRR | 0.97 (0.27, 1.67) | 0.48 (-0.45, 1.41) | -0.57 (-1.78, 0.64) |
| $O_3$ | Up to Lag 3 (95% CI[1]) | Up to Lag 7 (95% CI[1]) | Up to Lag 14 (95% CI[1]) |
| UDLM | 2.04 (0.98, 3.11) | 2.63 (1.31, 3.98) | 2.25 (0.53, 4.01) |
| CDLM | 2.03 (1.07, 3.00) | 2.52 (1.28, 3.77) | 2.10 (0.38, 3.85) |
| EB1 | 2.09 (0.82, 3.39) | 2.59 (0.97, 4.24) | 2.19 (0.11, 4.32) |
| GRR | 2.08 (1.10, 3.07) | 2.59 (1.33, 3.88) | 2.21 (0.48, 3.97) |
| BDLM | 1.91 (0.93, 2.90) | 2.32 (1.11, 3.56) | 2.26 (0.64, 3.91) |
| HB | 2.12 (1.18, 3.07) | 2.63 (1.41, 3.87) | 2.23 (0.61, 3.88) |
| HB2-GRR | 1.94 (1.01, 2.89) | 2.30 (1.10, 3.52) | 2.12 (0.46, 3.80) |
| HP-GRR | 1.83 (0.97, 2.70) | 2.18 (1.06, 3.31) | 2.12 (0.53, 3.73) |

[1]CI refers to confidence interval for UDLM, CDLM, EB1, GRR, and HP-GRR and refers to credible interval for BDLM, HB, and HB2-GRR.

Table 5: Estimated mean and 95% confidence/credible intervals (in parenthesis) for the lag effects (% change in mortality count) of an interquartile range increase of $PM_{10}$ ($21.49\mu g/m^3$) on mortality in Chicago, Illinois from 1987 to 2000 based on the data from the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) under eight estimation methods.

| | UDLM | CDLM | EB1 | GRR | BDLM | HB | HB2-GRR | HP-GRR |
|---|---|---|---|---|---|---|---|---|
| Lag 0 | -0.15 | -0.13 | -0.13 | -0.17 | -0.13 | -0.15 | -0.13 | -0.15 |
| | (-0.61, 0.31) | (-0.51, 0.26) | (-0.59, 0.34) | (-0.61, 0.27) | (-0.57, 0.31) | (-0.58, 0.28) | (-0.55, 0.29) | (-0.55, 0.25) |
| Lag 1 | 0.04 | 0.15 | 0.13 | 0.03 | 0.05 | -0.02 | 0.10 | 0.04 |
| | (-0.40, 0.48) | (-0.06, 0.35) | (-0.18, 0.43) | (-0.26, 0.33) | (-0.37, 0.46) | (-0.34, 0.31) | (-0.17, 0.37) | (-0.23, 0.31) |
| Lag 2 | 0.22 | 0.25 | 0.25 | 0.37 | 0.21 | 0.37 | 0.27 | 0.33 |
| | (-0.22, 0.65) | (0.06, 0.44) | (-0.23, 0.73) | (0.14, 0.60) | (-0.19, 0.60) | (0.11, 0.63) | (0.05, 0.48) | (0.12, 0.55) |
| Lag 3 | 0.65 | 0.23 | 0.56 | 0.44 | 0.54 | 0.48 | 0.25 | 0.37 |
| | (0.22, 1.08) | (0.05, 0.42) | (0.36, 0.77) | (0.22, 0.66) | (0.16, 0.92) | (0.21, 0.75) | (0.05, 0.45) | (0.18, 0.57) |
| Lag 4 | 0.22 | 0.15 | 0.15 | 0.19 | 0.13 | 0.21 | 0.07 | 0.12 |
| | (-0.20, 0.64) | (-0.01, 0.31) | (-0.17, 0.48) | (-0.01, 0.40) | (-0.22, 0.49) | (-0.02, 0.43) | (-0.10, 0.24) | (-0.06, 0.29) |
| Lag 5 | -0.27 | 0.03 | -0.17 | -0.13 | -0.23 | -0.17 | -0.11 | -0.15 |
| | (-0.69, 0.15) | (-0.11, 0.17) | (-0.32, -0.01) | (-0.32, 0.05) | (-0.55, 0.08) | (-0.39, 0.06) | (-0.26, 0.04) | (-0.31, 0.00) |
| Lag 6 | -0.41 | -0.09 | -0.31 | -0.29 | -0.27 | -0.31 | -0.17 | -0.23 |
| | (-0.83, 0.02) | (-0.24, 0.05) | (-0.46, -0.15) | (-0.47, -0.10) | (-0.56, 0.03) | (-0.52, -0.10) | (-0.32, -0.02) | (-0.37, -0.10) |
| Lag 7 | 0.03 | -0.19 | -0.08 | -0.23 | -0.11 | -0.21 | -0.14 | -0.19 |
| | (-0.40, 0.45) | (-0.34, -0.05) | (-0.24, 0.07) | (-0.42, -0.05) | (-0.36, 0.15) | (-0.42, 0.00) | (-0.28, 0.00) | (-0.32, -0.07) |
| Lag 8 | -0.25 | -0.26 | -0.26 | -0.12 | -0.15 | -0.10 | -0.10 | -0.16 |
| | (-0.70, 0.20) | (-0.40, -0.11) | (-0.71, 0.20) | (-0.31, 0.06) | (-0.38, 0.07) | (-0.31, 0.11) | (-0.22, 0.02) | (-0.27, -0.04) |
| Lag 9 | -0.03 | -0.28 | -0.14 | -0.08 | -0.13 | -0.09 | -0.09 | -0.14 |
| | (-0.48, 0.43) | (-0.41, -0.14) | (-0.28, 0.00) | (-0.27, 0.10) | (-0.33, 0.07) | (-0.32, 0.13) | (-0.20, 0.02) | (-0.25, -0.04) |
| Lag 10 | -0.27 | -0.25 | -0.25 | -0.14 | -0.14 | -0.15 | -0.09 | -0.14 |
| | (-0.73, 0.18) | (-0.40, -0.11) | (-0.75, 0.24) | (-0.34, 0.06) | (-0.32, 0.04) | (-0.38, 0.07) | (-0.20, 0.02) | (-0.23, -0.04) |
| Lag 11 | -0.12 | -0.20 | -0.19 | -0.24 | -0.13 | -0.24 | -0.08 | -0.13 |
| | (-0.58, 0.33) | (-0.37, -0.04) | (-0.55, 0.17) | (-0.44, -0.04) | (-0.29, 0.04) | (-0.48, 0.01) | (-0.18, 0.02) | (-0.21, -0.04) |
| Lag 12 | -0.33 | -0.13 | -0.22 | -0.31 | -0.12 | -0.31 | -0.07 | -0.12 |
| | (-0.79, 0.12) | (-0.30, 0.04) | (-0.39, -0.04) | (-0.52, -0.10) | (-0.28, 0.03) | (-0.55, -0.06) | (-0.17, 0.02) | (-0.19, -0.04) |
| Lag 13 | -0.25 | -0.08 | -0.14 | -0.23 | -0.10 | -0.24 | -0.07 | -0.10 |
| | (-0.70, 0.21) | (-0.25, 0.09) | (-0.33, 0.06) | (-0.50, 0.04) | (-0.24, 0.03) | (-0.55, 0.07) | (-0.16, 0.02) | (-0.17, -0.03) |
| Lag 14 | 0.18 | -0.07 | 0.07 | 0.18 | -0.08 | 0.18 | -0.06 | -0.09 |
| | (-0.24, 0.60) | (-0.38, 0.25) | (-0.27, 0.42) | (-0.21, 0.56) | (-0.21, 0.06) | (-0.22, 0.58) | (-0.14, 0.02) | (-0.15, -0.03) |

Table 6: Estimated mean and 95% confidence/credible intervals (in parenthesis) for the lag effects (% change in mortality count) of an interquartile range increase of $O_3$ (14.65 ppb) on mortality in Chicago, Illinois from 1987 to 2000 based on the data from the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) under eight estimation methods.

| | UDLM | CDLM | EB1 | GRR | BDLM | HB | HB2-GRR | HP-GRR |
|---|---|---|---|---|---|---|---|---|
| Lag 0 | 0.50 | 0.33 | 0.36 | 0.32 | 0.45 | 0.37 | 0.32 | 0.30 |
| | (-0.21, 1.20) | (-0.19, 0.85) | (-0.26, 0.98) | (-0.33, 0.98) | (-0.22, 1.11) | (-0.31, 1.00) | (-0.29, 0.94) | (-0.26, 0.85) |
| Lag 1 | 0.12 | 0.57 | 0.29 | 0.53 | 0.18 | 0.50 | 0.46 | 0.48 |
| | (-0.59, 0.83) | (0.29, 0.85) | (-0.08, 0.67) | (0.06, 1.00) | (-0.47, 0.83) | (0.00, 1.09) | (0.01, 0.91) | (0.08, 0.88) |
| Lag 2 | 1.20 | 0.61 | 1.04 | 0.68 | 1.00 | 0.69 | 0.63 | 0.59 |
| | (0.49, 1.91) | (0.35, 0.86) | (0.70, 1.38) | (0.34, 1.01) | (0.38, 1.62) | (0.26, 1.00) | (0.31, 0.95) | (0.30, 0.89) |
| Lag 3 | 0.22 | 0.51 | 0.39 | 0.54 | 0.27 | 0.54 | 0.52 | 0.45 |
| | (-0.48, 0.92) | (0.27, 0.75) | (0.09, 0.70) | (0.21, 0.86) | (-0.29, 0.84) | (0.17, 0.90) | (0.22, 0.82) | (0.18, 0.72) |
| Lag 4 | 0.36 | 0.35 | 0.35 | 0.31 | 0.25 | 0.30 | 0.27 | 0.24 |
| | (-0.34, 1.06) | (0.15, 0.55) | (-0.39, 1.10) | (0.01, 0.60) | (-0.24, 0.75) | (0.01, 0.70) | (0.01, 0.54) | (0.02, 0.46) |
| Lag 5 | 0.20 | 0.18 | 0.18 | 0.18 | 0.13 | 0.16 | 0.09 | 0.11 |
| | (-0.49, 0.90) | (0.00, 0.36) | (-0.62, 0.98) | (-0.11, 0.46) | (-0.30, 0.56) | (-0.14, 0.53) | (-0.16, 0.33) | (-0.08, 0.29) |
| Lag 6 | 0.02 | 0.03 | 0.03 | 0.10 | 0.03 | 0.10 | 0.01 | 0.02 |
| | (-0.68, 0.71) | (-0.16, 0.22) | (-0.79, 0.85) | (-0.19, 0.38) | (-0.34, 0.40) | (-0.28, 0.37) | (-0.22, 0.24) | (-0.14, 0.19) |
| Lag 7 | 0.01 | -0.08 | -0.07 | -0.07 | 0.00 | -0.05 | -0.02 | -0.03 |
| | (-0.68, 0.70) | (-0.27, 0.12) | (-0.74, 0.60) | (-0.35, 0.21) | (-0.33, 0.33) | (-0.45, 0.21) | (-0.23, 0.19) | (-0.18, 0.12) |
| Lag 8 | -0.08 | -0.12 | -0.12 | -0.27 | -0.04 | -0.27 | -0.06 | -0.04 |
| | (-0.77, 0.62) | (-0.31, 0.07) | (-0.82, 0.59) | (-0.55, 0.01) | (-0.35, 0.27) | (-0.55, 0.09) | (-0.26, 0.15) | (-0.17, 0.10) |
| Lag 9 | -0.57 | -0.11 | -0.40 | -0.22 | -0.07 | -0.26 | -0.07 | -0.02 |
| | (-1.26, 0.12) | (-0.29, 0.07) | (-0.60, -0.21) | (-0.51, 0.06) | (-0.40, 0.26) | (-0.46, 0.20) | (-0.26, 0.11) | (-0.13, 0.10) |
| Lag 10 | 0.18 | -0.05 | 0.02 | 0.12 | 0.03 | 0.12 | -0.02 | 0.00 |
| | (-0.51, 0.88) | (-0.24, 0.14) | (-0.26, 0.29) | (-0.17, 0.41) | (-0.20, 0.26) | (-0.23, 0.46) | (-0.19, 0.14) | (-0.11, 0.11) |
| Lag 11 | 0.57 | 0.02 | 0.41 | 0.32 | 0.05 | 0.36 | 0.02 | 0.00 |
| | (-0.13, 1.27) | (-0.21, 0.24) | (0.15, 0.67) | (0.01, 0.64) | (-0.20, 0.30) | (-0.15, 0.59) | (-0.14, 0.17) | (-0.09, 0.10) |
| Lag 12 | -0.19 | 0.05 | -0.03 | -0.03 | -0.01 | -0.02 | -0.01 | 0.00 |
| | (-0.88, 0.50) | (-0.18, 0.28) | (-0.36, 0.31) | (-0.34, 0.28) | (-0.20, 0.18) | (-0.38, 0.31) | (-0.15, 0.14) | (-0.09, 0.08) |
| Lag 13 | -0.55 | 0.00 | -0.39 | -0.54 | -0.03 | -0.57 | -0.03 | 0.00 |
| | (-1.23, 0.14) | (-0.22, 0.23) | (-0.64, -0.14) | (-0.97, -0.10) | (-0.22, 0.17) | (-0.89, 0.14) | (-0.18, 0.12) | (-0.08, 0.08) |
| Lag 14 | 0.28 | -0.20 | 0.13 | 0.24 | 0.01 | 0.26 | -0.01 | 0.00 |
| | (-0.34, 0.90) | (-0.64, 0.25) | (-0.36, 0.63) | (-0.33, 0.81) | (-0.14, 0.15) | (-0.43, 0.73) | (-0.14, 0.12) | (-0.07, 0.07) |

Table 7: Estimated mean and 95% confidence intervals (in parenthesis) for the cumulative lag effect (% change in mortality count) of an interquartile range increase of $PM_{10}$ ($21.49 \mu g/m^3$) across lags on mortality in Chicago, Illinois from 1987 to 2000 based on the data from the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) under eight estimation methods.

| | UDLM | CDLM | EB1 | GRR | BDLM | HB | HB2-GRR | HP-GRR |
|---|---|---|---|---|---|---|---|---|
| Lag 0 | -0.15 | -0.13 | -0.13 | -0.17 | -0.18 | -0.25 | -0.13 | -0.15 |
| | (-0.61, 0.31) | (-0.51, 0.26) | (-0.51, 0.25) | (-0.61, 0.27) | (-0.66, 0.24) | (-0.59, 0.08) | (-0.55, 0.29) | (-0.53, 0.23) |
| Lag 1 | -0.12 | 0.02 | 0.00 | -0.14 | -0.15 | -0.09 | -0.03 | 0.01 |
| | (-0.71, 0.48) | (-0.53, 0.58) | (-0.42, 0.42) | (-0.70, 0.42) | (-0.78, 0.41) | (-0.48, 0.31) | (-0.56, 0.51) | (-0.40, 0.43) |
| Lag 2 | 0.10 | 0.27 | 0.25 | 0.23 | 0.17 | 0.29 | 0.24 | 0.36 |
| | (-0.58, 0.78) | (-0.37, 0.92) | (-0.39, 0.88) | (-0.42, 0.89) | (-0.60, 0.91) | (-0.16, 0.73) | (-0.38, 0.86) | (-0.25, 0.96) |
| Lag 3 | 0.75 | 0.51 | 0.81 | 0.67 | 0.57 | 0.63 | 0.48 | 0.97 |
| | (-0.01, 1.52) | (-0.22, 1.23) | (0.07, 1.56) | (-0.06, 1.40) | (-0.24, 1.43) | (0.14, 1.12) | (-0.21, 1.18) | (0.27, 1.67) |
| Lag 4 | 0.97 | 0.65 | 0.97 | 0.86 | 0.71 | 0.78 | 0.55 | 1.12 |
| | (0.13, 1.81) | (-0.14, 1.45) | (0.10, 1.83) | (0.06, 1.67) | (-0.18, 1.69) | (0.23, 1.32) | (-0.22, 1.32) | (0.30, 1.94) |
| Lag 5 | 0.70 | 0.68 | 0.80 | 0.73 | 0.55 | 0.70 | 0.44 | 0.94 |
| | (-0.20, 1.61) | (-0.19, 1.55) | (-0.11, 1.71) | (-0.15, 1.61) | (-0.45, 1.59) | (0.11, 1.30) | (-0.39, 1.28) | (0.06, 1.82) |
| Lag 6 | 0.30 | 0.59 | 0.50 | 0.44 | 0.24 | 0.49 | 0.28 | 0.60 |
| | (-0.67, 1.27) | (-0.35, 1.52) | (-0.44, 1.44) | (-0.50, 1.39) | (-0.84, 1.35) | (-0.15, 1.13) | (-0.63, 1.18) | (-0.31, 1.51) |
| Lag 7 | 0.32 | 0.40 | 0.41 | 0.21 | -0.01 | 0.26 | 0.14 | 0.48 |
| | (-0.71, 1.36) | (-0.61, 1.40) | (-0.54, 1.37) | (-0.80, 1.22) | (-1.23, 1.16) | (-0.41, 0.94) | (-0.84, 1.12) | (-0.45, 1.41) |
| Lag 8 | 0.07 | 0.14 | 0.16 | 0.09 | -0.15 | 0.09 | 0.04 | 0.24 |
| | (-1.03, 1.18) | (-0.93, 1.21) | (-0.93, 1.25) | (-0.99, 1.16) | (-1.39, 1.13) | (-0.61, 0.79) | (-1.02, 1.09) | (-0.78, 1.26) |
| Lag 9 | 0.05 | -0.13 | 0.02 | 0.00 | -0.27 | -0.03 | -0.06 | 0.12 |
| | (-1.12, 1.22) | (-1.27, 1.01) | (-1.10, 1.13) | (-1.14, 1.15) | (-1.55, 1.13) | (-0.76, 0.70) | (-1.19, 1.07) | (-0.92, 1.16) |
| Lag 10 | -0.22 | -0.39 | -0.24 | -0.14 | -0.40 | -0.16 | -0.15 | -0.05 |
| | (-1.46, 1.01) | (-1.59, 0.82) | (-1.54, 1.07) | (-1.35, 1.08) | (-1.79, 1.05) | (-0.92, 0.60) | (-1.35, 1.06) | (-1.15, 1.05) |
| Lag 11 | -0.35 | -0.59 | -0.43 | -0.38 | -0.64 | -0.37 | -0.23 | -0.19 |
| | (-1.65, 0.95) | (-1.86, 0.68) | (-1.70, 0.84) | (-1.66, 0.91) | (-2.09, 0.91) | (-1.17, 0.42) | (-1.51, 1.05) | (-1.31, 0.92) |
| Lag 12 | -0.68 | -0.72 | -0.65 | -0.69 | -0.97 | -0.67 | -0.30 | -0.35 |
| | (-2.05, 0.68) | (-2.06, 0.62) | (-1.96, 0.66) | (-2.03, 0.66) | (-2.51, 0.65) | (-1.49, 0.16) | (-1.65, 1.05) | (-1.49, 0.79) |
| Lag 13 | -0.93 | -0.80 | -0.78 | -0.92 | -1.23 | -0.91 | -0.37 | -0.48 |
| | (-2.35, 0.49) | (-2.21, 0.61) | (-2.10, 0.53) | (-2.33, 0.50) | (-2.85, 0.45) | (-1.77, -0.04) | (-1.78, 1.05) | (-1.66, 0.69) |
| Lag 14 | -0.75 | -0.87 | -0.71 | -0.74 | -1.05 | -0.72 | -0.43 | -0.57 |
| | (-2.24, 0.75) | (-2.36, 0.62) | (-1.98, 0.56) | (-2.23, 0.75) | (-2.76, 0.69) | (-1.60, 0.15) | (-1.91, 1.05) | (-1.78, 0.64) |

Table 8: Estimated mean and 95% confidence intervals (in parenthesis) for the cumulative lag effect (% change in mortality count) of an interquartile range increase of $O_3$ (14.65 ppb) across lags on mortality in Chicago, Illinois from 1987 to 2000 based on the data from the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) under eight estimation methods.

| | UDLM | CDLM | EB1 | GRR | BDLM | HB | HB2-GRR | HP-GRR |
|---|---|---|---|---|---|---|---|---|
| Lag 0 | 0.50 | 0.33 | 0.36 | 0.32 | 0.45 | 0.37 | 0.32 | 0.30 |
| | (-0.21, 1.20) | (-0.19, 0.85) | (-0.26, 0.98) | (-0.33, 0.98) | (-0.22, 1.11) | (-0.28, 1.03) | (-0.29, 0.94) | (-0.26, 0.85) |
| Lag 1 | 0.61 | 0.90 | 0.65 | 0.85 | 0.63 | 0.87 | 0.78 | 0.78 |
| | (-0.19, 1.42) | (0.16, 1.64) | (-0.27, 1.58) | (0.10, 1.61) | (-0.15, 1.40) | (0.11, 1.64) | (0.06, 1.51) | (0.10, 1.45) |
| Lag 2 | 1.82 | 1.51 | 1.70 | 1.53 | 1.63 | 1.57 | 1.42 | 1.38 |
| | (0.87, 2.77) | (0.65, 2.38) | (0.59, 2.82) | (0.66, 2.41) | (0.74, 2.53) | (0.73, 2.42) | (0.58, 2.27) | (0.60, 2.16) |
| Lag 3 | 2.04 | 2.03 | 2.09 | 2.08 | 1.91 | 2.12 | 1.94 | 1.83 |
| | (0.98, 3.11) | (1.07, 3.00) | (0.82, 3.39) | (1.10, 3.07) | (0.93, 2.90) | (1.18, 3.07) | (1.01, 2.89) | (0.97, 2.70) |
| Lag 4 | 2.40 | 2.39 | 2.45 | 2.39 | 2.17 | 2.42 | 2.22 | 2.08 |
| | (1.27, 3.55) | (1.34, 3.44) | (0.86, 4.07) | (1.31, 3.48) | (1.12, 3.23) | (1.39, 3.47) | (1.19, 3.26) | (1.14, 3.02) |
| Lag 5 | 2.61 | 2.57 | 2.63 | 2.57 | 2.29 | 2.59 | 2.31 | 2.18 |
| | (1.41, 3.83) | (1.45, 3.69) | (1.13, 4.16) | (1.42, 3.73) | (1.18, 3.42) | (1.47, 3.71) | (1.22, 3.41) | (1.19, 3.19) |
| Lag 6 | 2.63 | 2.59 | 2.66 | 2.67 | 2.32 | 2.69 | 2.32 | 2.21 |
| | (1.36, 3.91) | (1.42, 3.78) | (1.01, 4.34) | (1.46, 3.89) | (1.16, 3.50) | (1.51, 3.88) | (1.17, 3.48) | (1.15, 3.27) |
| Lag 7 | 2.63 | 2.52 | 2.59 | 2.59 | 2.32 | 2.63 | 2.30 | 2.18 |
| | (1.31, 3.98) | (1.28, 3.77) | (0.97, 4.24) | (1.33, 3.88) | (1.11, 3.56) | (1.41, 3.87) | (1.10, 3.52) | (1.06, 3.31) |
| Lag 8 | 2.55 | 2.39 | 2.47 | 2.32 | 2.28 | 2.35 | 2.25 | 2.14 |
| | (1.18, 3.95) | (1.08, 3.72) | (0.84, 4.11) | (1.00, 3.66) | (1.02, 3.57) | (1.07, 3.65) | (0.99, 3.52) | (0.95, 3.34) |
| Lag 9 | 1.97 | 2.28 | 2.05 | 2.09 | 2.21 | 2.09 | 2.17 | 2.12 |
| | (0.54, 3.41) | (0.91, 3.68) | (0.38, 3.75) | (0.71, 3.49) | (0.87, 3.57) | (0.74, 3.45) | (0.85, 3.51) | (0.86, 3.40) |
| Lag 10 | 2.15 | 2.23 | 2.07 | 2.22 | 2.24 | 2.21 | 2.15 | 2.13 |
| | (0.66, 3.67) | (0.80, 3.69) | (0.32, 3.85) | (0.77, 3.68) | (0.85, 3.66) | (0.80, 3.64) | (0.77, 3.55) | (0.79, 3.48) |
| Lag 11 | 2.73 | 2.25 | 2.48 | 2.55 | 2.29 | 2.57 | 2.17 | 2.13 |
| | (1.17, 4.32) | (0.75, 3.77) | (0.63, 4.37) | (1.02, 4.09) | (0.84, 3.76) | (1.10, 4.06) | (0.71, 3.64) | (0.73, 3.55) |
| Lag 12 | 2.53 | 2.30 | 2.46 | 2.52 | 2.28 | 2.55 | 2.16 | 2.13 |
| | (0.91, 4.18) | (0.73, 3.89) | (0.45, 4.50) | (0.93, 4.13) | (0.77, 3.81) | (1.04, 4.09) | (0.64, 3.70) | (0.66, 3.62) |
| Lag 13 | 1.97 | 2.31 | 2.06 | 1.97 | 2.26 | 1.96 | 2.13 | 2.12 |
| | (0.30, 3.67) | (0.67, 3.97) | (-0.02, 4.18) | (0.31, 3.65) | (0.69, 3.85) | (0.39, 3.57) | (0.54, 3.74) | (0.59, 3.68) |
| Lag 14 | 2.25 | 2.10 | 2.19 | 2.21 | 2.26 | 2.23 | 2.12 | 2.12 |
| | (0.53, 4.01) | (0.38, 3.85) | (0.11, 4.32) | (0.48, 3.97) | (0.64, 3.91) | (0.61, 3.88) | (0.46, 3.80) | (0.53, 3.73) |

## A.9 Computation Times

Table 9: Computation times of applying eight estimation methods to National Morbidity, Mortality, and Air Pollution Study (NMMAPS) data on an Intel i7-2600 CPU with a single 3.4GHz core.

| Methods | Time |
|---|---|
| UDLM | 1.7 seconds |
| CDLM | 1.6 seconds |
| EB1 | 5.8 seconds |
| GRR[1] | 63.7 seconds |
| BDLM[2] | 5.4 seconds |
| HB[3] | 1.1 hours |
| HB2-GRR[1,4] | 13.1 hours |
| HP-GRR[1,4] | 14.1 hours |

[1] Tuning parameter is chosen from a grid of 100 equally-spaced values
[2] Asymptotic normality of the Poisson likelihood is applied
[3] Gibbs sampler is based on 10000 iterations
[4] Standard error estimates are based on 1000 bootstrap samples