

# UC Irvine

## UC Irvine Previously Published Works

### Title

Distributed Lag Interaction Models with Two Pollutants.

### Permalink

<https://escholarship.org/uc/item/0bs9r5gn>

### Journal

Journal of the Royal Statistical Society. Series C, Applied statistics, 68(1)

### ISSN

0035-9254

### Authors

Chen, Yin-Hsiu  
Mukherjee, Bhramar  
Berrocal, Veronica J

### Publication Date

2019

### DOI

10.1111/rssc.12297

Peer reviewed



Published in final edited form as:

*J R Stat Soc Ser C Appl Stat.* 2019 January ; 68(1): 79–97. doi:10.1111/rssc.12297.

## Distributed Lag Interaction Models with Two Pollutants

Yin-Hsiu Chen, Bhramar Mukherjee, and Veronica J. Berrocal

Department of Biostatistics, University of Michigan

### Summary.

Distributed lag models (DLMs) have been widely used in environmental epidemiology to quantify the lagged effects of air pollution on a health outcome of interest such as mortality and morbidity. Most previous DLM approaches only consider one pollutant at a time. In this article, we propose distributed lag interaction model (DLIM) to characterize the joint lagged effect of two pollutants. One natural way to model the interaction surface is by assuming that the underlying basis functions are tensor products of the basis functions that generate the main-effect distributed lag functions. We extend Tukey's one-degree-of-freedom interaction structure to the two-dimensional DLM context. We also consider shrinkage versions of the two to allow departure from the specified Tukey's interaction structure and achieve bias-variance tradeoff. We derive the marginal lag effects of one pollutant when the other pollutant is fixed at certain quantiles. In a simulation study, we show that the shrinkage methods have better average performance in terms of mean squared error (MSE) across different scenarios. We illustrate the proposed methods by using the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) data to model the joint effects of  $PM_{10}$  and  $O_3$  on mortality count in Chicago, Illinois, from 1987 to 2000.

### Keywords

Shrinkage; Time series; Tukey's single df test for non-additivity; Two-dimensional distributed lag interaction models

## 1. Introduction

The association between air pollution and adverse health outcomes has been an important public health concern and a topic of extensive research in environmental epidemiology (Pope and Dockery, 2006). The short-term, or acute effects, of air pollution exposure on health outcomes, such as mortality and cardiovascular events, have been widely studied (Pope et al., 1995; Dominici et al., 2006). However, most studies so far have considered adverse health effects of exposure to a single pollutant (Dominici et al., 2010). When ambient concentration data are available for multiple pollutants, it is standard practice to analyze their effects one at a time by fitting multiple single pollutant models. However, the health burden from simultaneous exposure to multiple pollutants may differ from the sum of individual effects and the mode of action can be synergistic or antagonistic (Mauderly, 1993). A multi-pollutant approach that considers the joint effects of chemical mixtures of exposures is likely to yield more accurate assessment of health risk (Billonnet et al., 2012). A variety of approaches have been proposed to estimate the health effects of multiple pollutants (Sun et al., 2013), including least absolute shrinkage and selection operator

(LASSO) (Tibshirani, 1996), classification and regression tree (CART) (Hu et al., 2008), Bayesian kernel machine regression (BKMR) (Bobb et al., 2014). However, very few methods so far consider the problem of capturing the lagged effect of two pollutants and their potential interactions over a biologically meaningful time period. Single-day pollution measures might underestimate risk when there is a cumulative effect of air pollution over a time window preceding a health event (Roberts, 2005).

Distributed lag models (DLMs) are a class of models often used to simultaneously include lagged measures of concentration levels of an ambient air pollutant. Parametric DLM assumes that the lag effect coefficients are a function of the lags, such as lower-degree polynomials (Almon, 1965). Generalized additive DLM (Zanobetti et al., 2000) uses penalized regression splines (Marx and Eilers, 1998) to represent the distributed lag (DL) function in a more flexible manner. Bayesian DLM (BDLM) (Welty et al., 2009) was proposed to incorporate prior knowledge about the DL function through specification of the prior variance-covariance matrix of lag coefficients. Most of the discussion regarding DLM has been in the context of a single pollutant and only few distributed lag interaction models (DLIMs) with two pollutants have been attempted. Extensions to higher dimensions include bivariate constrained DLIM (Muggeo, 2007) (CDLIM) and high degree DLM (HDDLIM) (Heaton and Peng, 2014). The CDLIM paper jointly models the temperature and air particulate matter with aerodynamic diameter less than 10 microns ( $PM_{10}$ ) main effect in the same way as a parametric DLM with two separate sets of basis functions. Tensor products of the two are employed to characterize the joint DL surface for the temperature- $PM_{10}$  interaction. The HDDLIM paper extended the DLM framework to incorporate higher-order interactions between lagged predictors corresponding to a single exposure, using a Gaussian process prior as a dimension reduction tool.

Tukey's one degree-of-freedom test for non-additivity (Tukey, 1949) is a parsimonious approach to model the interaction term as a scaled product of its corresponding main effects (Chatterjee et al., 2006; Maity et al., 2009). In this paper, we extend Tukey's model to DLIMs where the interaction is parameterized as a scaled product of two DLM main effects. We will consider estimation and inference under such an extension in both frequentist and Bayesian framework. We also propose a Bayesian constrained DLIM (BCDLIM) approach to characterize the joint effect of two pollutants. Instead of shrinking all main effects and interaction effects toward zero, we set a pre-specified parametric CDLIM as the shrinkage target in this approach. BCDLIM is able to strike a desirable bias-variance tradeoff in a data-adaptive way.

The rest of the paper is organized as follows. In Section 2, we first review the existing methods, including (1) unconstrained DLIM (UDLIM) and (2) constrained DLIM (CDLIM). We then introduce the proposed new methods (1) Tukey's DLIM (TDLIM), (2) Bayesian Tukey's DLIM (BTDLIM), and (3) Bayesian constrained DLIM (BCDLIM). In Section 3, we conduct a simulation study to evaluate the operating characteristics of the five different methods. In Section 4, we illustrate the methods by analyzing data from the National, Morbidity, Mortality, and Air Pollution Study (NMMAPS) to estimate the lagged effects of particulate matter with diameter less than 10 microns ( $PM_{10}$ ) and ozone ( $O_3$ ) concentration

on mortality in Chicago, Illinois, from 1987 to 2000. We conclude with a discussion in Section 5.

There are several novel features of this article. First, we extend DLM to DLIM to handle two pollutants. We attempt to characterize the changes in a true DL function corresponding to one exposure when the other is fixed at different values. Extending the well-known Tukey's model for interaction to DLIM is another innovation. Finally, using data adaptive shrinkage to allow for an unconstrained interaction model to shrink towards a parametric DLIM structure is a new contribution to the literature. More broadly, the paper posits new ideas for thinking about interaction structures between a pair of time-series predictors with potential lagged effects on an outcome. This approach bears relevance beyond air pollution epidemiology.

## 2. Methods

Let  $x_{1t}$  denote the first exposure measured at time  $t$  (e.g.  $PM_{10}$ ),  $x_{2t}$  denote the second exposure measured at time  $t$  (e.g.  $O_3$ ),  $y_t$  denote the response measured at time  $t$  (e.g. daily mortality count), and  $z_t$  denote the vector of covariates at time  $t$ , such as temperature and humidity, in addition to a constant 1 corresponding to the intercept parameter. Let  $T$  be the length of the time series,  $L_1$  and  $L_2$  be the maximum number of lags considered for the first and second exposure, respectively. In addition, we denote with  $X_{1t} = (x_{1t}, \dots, x_{1,t-L_1})^\top$ ,  $X_{2t} = (x_{2t}, \dots, x_{2,t-L_2})^\top$ , the vector of lagged exposure and with  $X_{It} = X_{1t} \otimes X_{2t}$ , where  $\otimes$  is the Kronecker product, the  $(L_1 + 1)(L_2 + 1)$  elements that refer to the two-way interaction terms between the two exposures. The log-linear Poisson DLIM with all pairwise interactions between lagged measurements of the two exposures is described as

$$y_t | z_t, X_{1t}, X_{2t}, X_{It} \sim \text{Poisson}(\mu_t) \quad (1)$$

$$\begin{aligned} \log(\mu_t) &= z_t^\top \alpha + X_{1t}^\top \beta_1 + X_{2t}^\top \beta_2 + X_{It}^\top \gamma \\ &= z_t^\top \alpha + \sum_{i=0}^{L_1} x_{1,t-i} \beta_{1i} + \sum_{j=0}^{L_2} x_{2,t-j} \beta_{2j} + \sum_{i=0}^{L_1} \sum_{j=0}^{L_2} \gamma_{ij} x_{1,t-i} x_{2,t-j} \end{aligned} \quad (2)$$

where  $\alpha$  represents the effect of covariates,  $\beta_1 = (\beta_{10}, \dots, \beta_{1L_1})^\top$  is the  $(L_1 + 1)$ -vector of lagged main effects of the first exposure,  $\beta_2 = (\beta_{20}, \dots, \beta_{2L_2})^\top$  is the  $(L_2 + 1)$ -vector of lagged main effects of the second exposure, and  $\gamma = \text{vec}(\Gamma) = (\gamma_{00}, \gamma_{01}, \dots, \gamma_{L_1 L_2})^\top$  where  $\Gamma$  is the  $(L_1 + 1) \times (L_2 + 1)$  matrix of interaction effects. Our primary goal is to estimate the main effects

$\beta_1$  and  $\beta_2$  and the interaction effects  $\gamma$ . For simplicity, we leave out  $z_t^\top \alpha$  in subsequent presentations.

**Remark:** (1) and (2) model the conditional mean response at a time point  $t$  given the current and past measurements of the two exposures. Non-null interaction effect in (2) implies that the lagged effects of the first exposure depend on the level of the second exposure, and vice versa. It is noted that the interaction effects in (2) are not symmetric, namely  $\gamma_{ij} \neq \gamma_{ji}$  for  $i \neq j$ . A natural quantity of interest is the marginal effect of one exposure at a certain lag given the other exposure fixed at a certain level such as median or a specified quantile.

Algebraically, if we fix the second exposure at  $x_2^*$  across all lags, the marginal lag effects of the first exposure at lag  $i$  can be written as  $\beta_{1i}^*(x_2^*) = \beta_{1i} + x_2^* \sum_{j=0}^{L_2} \gamma_{ij}$  for  $i = 0, \dots, L_1$ . The vector representation is

$$\beta_1^m(x_2^*) = \beta_1 + x_2^* \cdot \Gamma \mathbf{1} \quad (3)$$

where  $\mathbf{1}$  is a vector of 1s. Similarly, if we fix the first exposure at  $x_1^*$ , the marginal lag effects of the second exposure at lag  $j$  can be written as  $\beta_{2j}^*(x_1^*) = \beta_{2j} + x_1^* \sum_{i=0}^{L_1} \gamma_{ij}$  for  $j = 0, \dots, L_2$  with vector representation  $\beta_2^m(x_1^*) = \beta_2 + x_1^* \cdot \Gamma^\top \mathbf{1}$ . Throughout the rest of this paper, we will summarize the estimates of  $\beta_1$ ,  $\beta_2$ , and  $\gamma = \text{vec}(\Gamma)$  based on the above expressions.

## 2.1. Existing Methods

**2.1.1. Unconstrained Distributed Lag Interaction Model (UDLIM)**—UDLIM does not impose any constraints on coefficients  $\psi = (\beta_1^\top, \beta_2^\top, \gamma^\top)^\top$  in (2). The UDLIM coefficients can be simply estimated via maximum likelihood estimation (MLE).

$$\hat{\psi}_{UDLIM} = \arg \max_{\psi} \sum_{t=1}^T [y_t X_t^\top \psi - e^{-X_t^\top \psi} - \log(y_t!)],$$

where  $X_t = (X_{1t}^\top, X_{2t}^\top, X_{It}^\top)^\top$ . Standard frequentist inference based on large sample theory of MLEs can be drawn subsequently. However, due to the collinearity between serially measured exposure levels and the large number of parameters (i.e.  $L_1 + L_2 + 2$  main effect terms and  $(L_1 + 1)(L_2 + 1)$  interaction terms), the lagged effect estimates may be less efficient with inflated variance and the estimated DL functions could be highly variable.

**2.1.2. Constrained Distributed Lag Interaction Model (CDLIM)**—Parametric DLIM imposes a smooth structure on lagged effect coefficients by assuming each lag coefficient to be a linear combination of known basis functions measured at its lag index. CDLIM extends this configuration to two-dimensional scenarios. Assume  $B_{11}(\cdot), \dots, B_{1p_1}(\cdot)$

are the  $p_1$  basis functions applied to  $\beta_1$  and  $B_{21}(\cdot), \dots, B_{2p_2}(\cdot)$  are the  $p_2$  basis functions applied to  $\beta_2$ . The main effects coefficients are assumed to be of the form  $\beta_{1i} = \sum_{m=1}^{p_1} B_{1m}(i)\theta_{1m}$  for  $i = 0, \dots, L_1$  and  $\beta_{2j} = \sum_{n=1}^{p_2} B_{2n}(j)\theta_{2n}$  for  $j = 0, \dots, L_2$  where  $\{\beta_{1i}\}$  and  $\{\beta_{2j}\}$  are elements of  $\beta_1$  and  $\beta_2$ , respectively, and  $\{\theta_{1m}\}$  and  $\{\theta_{2n}\}$  are free parameters to be estimated. In order to smooth the interaction surface, Muggeo (2007) utilizes tensor products of marginal basis functions. The element corresponding to the interaction between  $x_{1,t-i}$  and  $x_{2,t-j}$  can be expressed as

$$\gamma_{ij} = \sum_{m=1}^{p_1} \sum_{n=1}^{p_2} B_{1m}(i)B_{2n}(j)\theta_{1mn}.$$

Define  $C_1$  as a  $(L_1 + 1) \times p_1$  transformation matrix (Gasparri et al., 2010) where the element  $(i+1, m)$  is  $B_{1m}(i)$  and similarly, define  $C_2$  as a  $(L_2+1) \times p_2$  transformation matrix where the element  $(j+1, n)$  is  $B_{2n}(j)$ . Denote  $\theta_1 = (\theta_{11}, \dots, \theta_{1p_1})$ ,  $\theta_2 = (\theta_{21}, \dots, \theta_{2p_2})$ , and  $\theta_I = (\theta_{I11}, \theta_{I12}, \dots, \theta_{I p_1 p_2})$  the CDLIM coefficients can be written in terms of the free parameters to be estimated as

$$\beta_1 = C_1\theta_1, \beta_2 = C_2\theta_2, \gamma = (C_1 \otimes C_2)\theta_I. \quad (4)$$

The free parameters  $\theta_1$ ,  $\theta_2$ , and  $\theta_I$  can be obtained by maximizing the log likelihood function

$$\sum_{t=1}^T [y_t \{W_{1t}^T \theta_1 + W_{2t}^T \theta_2 + W_{It}^T \theta_I\} - e^{-W_{1t}^T \theta_1 + W_{2t}^T \theta_2 + W_{It}^T \theta_I} - \log(y_t!)]$$

where  $W_{1t} = C_1^T X_{1t}$ ,  $W_{2t} = C_2^T X_{2t}$ , and  $W_{It} = (C_1 \otimes C_2)^T X_{It}$ . Let  $\Theta = (\theta_1^T, \theta_2^T, \theta_I^T)^T$ , a vector of length  $p_1 + p_2 + p_1 p_2$ , and  $C = \text{diag}[C_1, C_2, C_1 \otimes C_2]$ . The CDLIM estimator can be written as  $\hat{\psi}_{CDLIM} = C\hat{\Theta}$  and  $\text{Cov}(\hat{\psi}_{CDLIM}) = CC\text{Cov}(\hat{\Theta})C^T$ .

## 2.2. Proposed Methods

**2.2.1. Tukey’s Distributed Lag Interaction Model (TDLIM)**—The underlying foundation of Tukey’s model for interaction is a latent variable framework (Chatterjee et al., 2006). Suppose we define a surrogate variable for each exposure that aggregates the temporal lagged effect of the exposure through weighted sum at time  $t$ . Namely,

$$s_{1t} = \sum_{i=0}^{L_1} w_{1i} x_{1,t-i}, s_{2t} = \sum_{j=0}^{L_2} w_{2j} x_{2,t-j}. \quad (5)$$

If we assume that the association between  $y_t$ ,  $X_{1t}$  and  $X_{2t}$  is through the interaction model

$$\log(E[y_t]) = \mu_0 + \mu_1 s_{1t} + \mu_2 s_{2t} + \mu_I s_{1t} s_{2t}. \quad (6)$$

Substituting (5) in (6), we can obtain

$$\begin{aligned} \log(E[y_t]) &= \mu_0 + \sum_{i=0}^{L_1} \mu_1 w_{1i} x_{1,t-i} + \sum_{j=0}^{L_2} \mu_2 w_{2j} x_{2,t-j} + \sum_{i=0}^{L_1} \sum_{j=0}^{L_2} \mu_I w_{1i} w_{2j} x_{1,t-i} x_{2,t-j} \\ &= \mu_0 + \sum_{i=0}^{L_1} \beta_{1i} x_{1,t-i} + \sum_{j=0}^{L_2} \beta_{2j} x_{2,t-j} + \sum_{i=0}^{L_1} \sum_{j=0}^{L_2} \gamma_{ij} x_{1,t-i} x_{2,t-j} \end{aligned}$$

where  $\beta_{1i} = \mu_1 w_{1i}$ ,  $\beta_{2j} = \mu_2 w_{2j}$ , and  $\gamma_{ij} = \mu_I w_{1i} w_{2j}$ . Note that we can express the interaction coefficient as  $\gamma_{ij} = \beta_{1i} \beta_{2j} \left( \frac{\mu_I}{\mu_1 \mu_2} \right)$ , a scaled product of the corresponding main-effect

coefficients. This motivates the use of Tukey's style interaction in our context. The surrogate variables  $s_{1t}$  and  $s_{2t}$  represent summary exposures over all the lags of the two exposures, respectively. Coefficients  $\mu_0$ ,  $\mu_1$ ,  $\mu_2$ , and  $\mu_I$  characterize the overall combined effects of the two exposures in association with outcome measurement at lag 0. The lag measurements of the two exposures interact through the two surrogate variables in the simple pairwise interaction model described in (6). Estimating the lagged effects in this model is the same as estimating the relative weights to combine the exposure lagged measurements into a summary surrogate variable. To extend the classical Tukey interaction structure to DLIMs, we now assume that the main effects are specified in the same way as in CDLIM with constrained parameterization such that  $\beta_1 = C_1 \theta_1$  and  $\beta_2 = C_2 \theta_2$  as in (4). In matrix form, the interaction coefficients can be expressed under Tukey's model as

$$\gamma = \eta \cdot (\beta_1 \otimes \beta_2) = (C_1 \otimes C_2) [\eta (\theta_1 \otimes \theta_2)].$$

Note that the interaction structure corresponding to TDLIM is a special case of CDLIM with  $\theta_I = \eta (\theta_1 \otimes \theta_2)$ . The number of parameters used for modeling the interaction effect reduces from  $p_1 p_2$  to 1. The model without interaction is nested within the Tukey's structure with the scalar parameter set to zero, assuming non-null main effects. The free parameters  $\theta_1$ ,  $\theta_2$ , and  $\eta$  can be estimated by maximizing the log likelihood function

$$\sum_{t=1}^T \left\{ y_t [W_{1t}^T \theta_1 + W_{2t}^T \theta_2 + \eta \cdot W_{It}^T (\theta_1 \otimes \theta_2)] - e^{W_{1t}^T \theta_1 + W_{2t}^T \theta_2 + \eta \cdot W_{It}^T (\theta_1 \otimes \theta_2)} - \log(y_t!) \right\}. \quad (7)$$

TDLIM is a nonlinear regression model where the objective function (7) involves products of the parameters. Linear approximation using first-order Taylor series expansion can be applied for parameter estimation and statistical inference. However, empirically, we found that the approximation accuracy using first-order approximation is poor and the asymptotic variance is far from the empirical variance. We therefore consider an iterative approach for

estimation (details provided in Supplementary Appendix A.1). The value of the objective function decreases at each step and the solution is guaranteed to converge. We recognize that the likelihood function (7) is non-convex in terms of the parameters so the convergence to a global maximum is not guaranteed by the iterative procedure. However, in our numerical studies, when the main effects are bounded away from zero, the choice of various initial values did not affect the final parameter estimates. When at least one of the main effects are close to the null value, the parameter  $\eta$  is not identifiable and estimation instability occurs in these cases. For statistical inference, we consider a standard vanilla bootstrap by resampling observations with replacement to obtain standard errors and confidence intervals.

**2.2.2. Bayesian Tukey's Distributed Lag Interaction Model (BTDLIM)**—In the proposed BTDLIM, the main effects are parametrically specified in the same way as in (4) and the interaction effects are modelled in the spirit of TDLIM. The distinction from the presentation in the previous section is that BTDLIM allows departure from Tukey's interaction structure in a data-adaptive way. BTDLIM assumes that the scalar parameter can vary across different interaction terms through the following prior specification

$$\gamma = \eta \odot (\beta_1 \otimes \beta_2), \eta \sim N(\mathbf{0}, \sigma^2 \Sigma(\omega))$$

where  $\eta = (\eta_{00}, \eta_{01}, \dots, \eta_{L_1 L_2})^\top$  is the vector of scalars,  $\odot$  is the operator denoting element-wise multiplication,  $\sigma^2$  is the common variance, and  $\Sigma$  is the correlation matrix parameterized by a single parameter  $\omega > 0$ . The correlation between  $\eta_{ij}$  and  $\eta_{i^*j^*}$  is given by  $\omega \sqrt{(i-i^*)^2 + (j-j^*)^2}$  assuming exponential structure. The prior on  $\eta$  relaxes the strict specification of Tukey's interaction structure. The amount of departure from Tukey's model is controlled by the parameter  $\omega$ . At one extreme, when  $\omega = 0$ , no structure is imposed on the interaction effects. The interaction coefficients are simply a reparametrization of the UDLIM coefficients in (2). At the other extreme when  $\omega = 1$ , the model degenerates to TDLIM and enforces the interaction coefficients to follow the Tukey's structure exactly. When  $\omega$  approaches 1, the correlation between neighboring coefficients is larger, resulting in a smoother interaction surface.

To complete the model specifications, we assign  $\theta_1 \sim N(\mathbf{0}, 100^2 \mathbf{I})$  and  $\theta_2 \sim N(\mathbf{0}, 100^2 \mathbf{I})$  as vague priors for the main effects coefficients. We assume a non-informative prior (Gelman et al., 2006) on the variance parameter  $\sigma^2 \sim IG(a = 0.001, b = 0.001)$  where  $a$  and  $b$  are the shape and scale parameters of the Inverse-Gamma (IG) distribution. To alleviate computational burden and keep the prior uninformative, we let  $\omega$  have a discrete uniform prior on  $\{0.1, 0.2, \dots, 1\}$ . The marginal posterior density of  $\beta_1$ ,  $\beta_2$ , and  $\gamma$  is not available in closed form. We use Metropolis-Hastings algorithm within a Gibbs sampler to approximate the posterior distribution and obtain the BTDLIM estimator as the posterior mean with the corresponding highest posterior density (HPD) interval as the corresponding credible interval. The full conditional distributions are presented in Supplementary Appendix A.2.



### 2.2.3. Bayesian Constrained Distributed Lag Interaction Model (BCDLIM)—

CDLIM is a fully parametric model. The dimension reduction from  $(L_1+1)+(L_2+1)+(L_1+1)(L_2+1)$  parameters to  $p_1 + p_2 + p_1p_2$  parameters results in efficiency gain in estimation. However, the benefit can be counterbalanced by potential bias when the underlying structure for the DL functions/surface is mis-specified. We propose a Bayesian constrained DLIM (BCDLIM) to shrink UDLIM estimates in a smooth manner toward a pre-specified CDLIM.

Let  $B_{11}^+(\cdot), \dots, B_{1, L_1+1}^+(\cdot)$  be  $L_1 + 1$  basis functions for the first exposure. For example, B-spline basis functions of degree 3 (cubic) with intercept and  $L_1 - 3$  equispaced internal knots positioned between 0 and  $L_1$ . Note that the basis functions describe the non-linearity in the DL function, but the exposure effect at each lag is still assumed to be linear. Let  $T_1$  be the corresponding  $(L_1+1) \times (L_1+1)$  transformation matrix. Let  $T_2$  denote the square transformation matrix with dimension  $(L_2 + 1) \times (L_2 + 1)$ , constructed in a similar manner for the second exposure, and let the transformation matrix for the interaction parameter be  $T_I = (T_1 \otimes T_2)$  with dimension  $(L_1 + 1)(L_2 + 1) \times (L_1 + 1)(L_2 + 1)$ . If we apply the transformation operators  $T_1, T_2$ , and  $T_I$  to CDLIM, the resulting estimator would be identical to UDLIM estimator since a full-rank transformation on the coefficients does not change the model fit. However, if we imposed shrinkage on the coefficients using a  $L_2$  penalty, the CDLIM estimator and UDLIM estimators would be different since the shrinkage is employed in different parameter spaces. The UDLIM estimator can be viewed as choosing  $B_{1m}^+(i) = I(m = i + 1)$  for  $m = 1, \dots, L_1 + 1$  and  $B_{2n}^+(j) = I(n = j + 1)$  for  $n = 1, \dots, L_2 + 1$ , where  $I(\cdot)$  is an indicator function, corresponding to  $T_1 = I$  and  $T_2 = I$ . Although the two sets of estimates share the same shrinkage target (i.e. the zero line), the solution paths are different. If the basis functions selected for  $T_1$  and  $T_2$  are smooth, CDLIM with shrinkage leads to smooth estimates.

Instead of shrinking the model coefficients toward 0, we consider shrinking them to a non-null target, determined by the transformation matrices  $C_1, C_2$ , and  $C_I = (C_1 \otimes C_2)$  for CDLIM defined in (4). Without loss of generality, we only describe how to construct the non-null shrinkage target for the first exposure. We first separate  $T_1$  into two parts –  $C_1$  and  $C_1^c$  where  $C_1^T C_1^c = 0$ . We make use of this orthogonal decomposition to obtain  $C_1^c$  whose columns span the complementary column space of  $C_1$ .  $C_1$  and  $C_1^c$  define the decomposition of the transformations corresponding to shrinkage toward a pre-specified target and shrinkage toward 0, respectively. The orthogonal projection of  $T_1$  onto the complementary column space of  $C_1$  is given by  $P_1 = [I - C_1(C_1^T C_1)^{-1} C_1^T] T_1$ . Using singular value decomposition (SVD), we can write  $P_1 = U_1 D_1 V_1^T$  where  $U_1$  contains the columns of left-singular vectors,  $D_1$  is a diagonal matrix with eigenvalues of  $P_1$ , and  $V_1$  contains the columns of right-singular vectors. Since the rank of  $P_1$  is  $L_1 + 1 - p_1$ , we can write  $U_1 = [U_{11} \ U_{12}]$  where  $U_{11}$  is a  $(L_1 + 1) \times (L_1 + 1 - p_1)$  matrix with columns of singular vectors corresponding to nonzero eigenvalues in  $D_1$ , while  $U_{12}$  is a  $(L_1 + 1) \times p_1$  matrix with columns of singular vectors corresponding to the eigenvalues of 0. We consider  $C_1^c = U_{11}$ . It is easy to show that  $C_1^T C_1^c = \mathbf{0}$  and the  $p_1$  columns of  $C_1$  and the  $L_1 + 1 - p_1$  columns of  $C_1^c$

span the entire  $\mathbb{R}^{L_1+1}$ . In other words, shrinkage through the columns of  $C_1^c$  defines CDLIM estimate as the shrinkage target. The complementary matrices  $C_2^c$  and  $C_I^c$  for the second exposure and interaction can be constructed using  $C_2, T_2$  and  $C_I, T_I$  respectively, in a similar way.

The likelihood corresponding to the above specification is given by

$$Y|\beta_1, \beta_2, \gamma \sim \text{Poisson}\left(e^{X_1\beta_1 + X_2\beta_2 + X_I\gamma}\right)$$

where  $Y = (y_1, \dots, y_T)^\top$ ,  $X_1 = (X_{11}, \dots, X_{1T})^\top$ ,  $X_2 = (X_{21}, \dots, X_{2T})^\top$ , and  $X_I = (X_{I1}, \dots, X_{IT})^\top$ . The prior specifications corresponding to the BCDLIM parameters are

$$\beta_1 = C_1\theta_1 + C_1^c\theta_1^c, \quad \beta_2 = C_2\theta_2 + C_2^c\theta_2^c, \quad \gamma = C_I\theta_I + C_I^c\theta_I^c$$

$$\theta_1 \sim N(\mathbf{0}, 100^2\mathbf{I}), \quad \theta_2 \sim N(\mathbf{0}, 100^2\mathbf{I}), \quad \theta_I \sim N(\mathbf{0}, 100^2\mathbf{I})$$

$$\theta_1^c \sim N(\mathbf{0}, \sigma_1^2\mathbf{I}), \quad \theta_2^c \sim N(\mathbf{0}, \sigma_2^2\mathbf{I}), \quad \theta_I^c \sim N(\mathbf{0}, \sigma_I^2\mathbf{I})$$

where  $\theta_1, \theta_2$ , and  $\theta_I$  are the coefficients without shrinkage and  $\theta_1^c, \theta_2^c$ , and  $\theta_I^c$  are the coefficients to be shrunk toward 0. In other words,  $\beta_1, \beta_2$ , and  $\gamma$ , are shrunk toward  $C_1\theta_1, C_2\theta_2$ , and  $C_I\theta_I$ , respectively. To complete the model specification, we assign hyper-priors on the variance parameters as

$$\sigma_1^2 \sim IG(a_0, b_0), \quad \sigma_2^2 \sim IG(a_0, b_0), \quad \sigma_I^2 \sim IG(a_0, b_0).$$

We fix  $a_0 = b_0 = 0.001$  to assume a noninformative hyper-prior (Gelman et al., 2006). Metropolis Hastings algorithm within a Gibbs sampler can be used to approximate the posterior distribution of the model parameters. The full conditional distributions are provided in Supplementary Appendix A.3. The hyper-priors of BCDLIM can alternatively be viewed as penalty terms in penalized likelihood. The dual representation is presented in Supplementary Appendix A.4.

### 3. Simulation Study

We conducted a simulation study to compare the estimation performance of the five methods introduced in Section 2 under different settings. We implemented the three frequentist methods using the built-in R function `glm` and the two Bayesian methods by calling the software Just Another Gibbs Sampler (JAGS) using R package `rjags` (Lunn et al., 2009). The

average computation times for 1000 data sets under each method are provided in Supplementary Appendix A.5 Table 1. All simulations were performed in R version 3.3.1.

### 3.1. Simulation Settings

We generated two separate exposure time series ( $i = 1, 2$ ) of length 1000 days with mean 3 and first-order autocorrelation equal to 0.5 from the model  $x_{it} = 0.5x_{it-1} + \epsilon_{it}$  where  $\epsilon_{it} \sim \text{i.i.d. } \mathcal{N}(0, 0.75)$  for  $i = 1, 2$  and  $t = 1, \dots, 1000$ . We set  $L_1 = L_2 = 9$  for both data generation and model fitting. The outcome  $y_t$  is generated from a Poisson distribution with mean  $\exp(\beta_0 + X_{1t}^\top \beta_1 + X_{2t}^\top \beta_2 + X_{It}^\top \gamma)$  for  $t = 1, \dots, 1000$  where  $X_{1t}$ ,  $X_{2t}$ , and  $X_{It}$  are defined as in Section 2. Let  $\beta_0 = 3$  and consider two DL functions for the main-effect coefficients  $\beta_1$  and  $\beta_2$  - (a) cubic and (b) a function with departure from cubic. We consider five different underlying true interaction structures for  $\gamma$  - (1) No interaction, (2) Tukey's style interaction, (3) Kronecker product interaction, (4) Sparse interaction, and (5) Unstructured interaction. The exact specifications are available in Supplementary Appendix A.6. In total, nine simulation scenarios, including all combinations of the two main-effect coefficients (a-b) and five interaction-effect coefficients (1-5) except the combination of (b) and (3), are considered. Exclusion of the combination of (b) and (3) is due to the fact that the Kronecker product interaction cannot be constructed when the corresponding main effects are not fully parametric as their underlying basis functions are undefined. In all simulations, we assume the lag structure of CDLIM, TDLIM, BTDLIM, and BCDLIM to be a cubic polynomial in the lags for all model fitting purposes.

### 3.2. Evaluation Metrics

The marginal lagged effects of the first exposure defined in (3) depends on the level at which the second exposure is fixed. One way to eliminate the effect of the second exposure is to integrate it out. We consider to use finite Riemann sum to numerically approximate the integral given by  $\beta_1^* = \int \beta_1^*(x_2) dx_2 \approx \frac{1}{S} \sum_{s=1}^S \beta_1^* \left( x_2^{\left[ \frac{q_{(s-0.5)/S} \right]} \right)$  where  $x_2^{\left[ \frac{q_{(s-0.5)/S} \right]}$  is the  $(s - 0.5)/S$ -th quantile of  $x_2$ . The empirical bias and empirical relative efficiency of the above quantity with  $S = 20$  are used to summarize the simulation results across different scenarios. The squared bias is computed as  $(\hat{\beta}_1^* - \beta_1^*)^\top (\hat{\beta}_1^* - \beta_1^*)$  where  $\hat{\beta}_1^*$  is the average of the estimates obtained from the 1000 simulated datasets. The empirical mean squared error (MSE) is computed as  $\frac{1}{1000} \sum_{j=1}^{1000} \|\hat{\beta}_{1j}^* - \beta_1^*\|_2^2$ . The relative efficiency is expressed with respect to the MSE of the UDLIM estimate, namely the MSE of UDLIM divided by the MSE of a certain method. We emphasize that the efficiency is defined through the MSE rather than the variance in this article. Because of the symmetry between  $x_1$  and  $x_2$ , we only present results for the marginal lagged effects of  $x_1$ .

### 3.3. Simulation Results

Results for the setting with main effects generated from a cubic DL function are summarized in Table 1. As we can observe in scenario (1), e.g. no interaction, all methods are more efficient than UDLIM with relative efficiency ranging from 6.27 to 19.24. The empirical

squared bias is minimal for UDLIM (0.02), CDLIM (0.00) and BCDLIM (0.00) and is moderately small for TDLIM (0.19) and BTDLIM (0.13). Null interaction is a special case of Tukey's model with  $\eta = 0$ . Because TDLIM correctly specify the main effects and interaction effects with a smaller number of parameters, it achieves the highest efficiency (19.24). In scenario (2) where the non-null interaction effects are of Tukey's form, all methods have similar, though slightly smaller, relative efficiency in comparison with scenario (1), ranging from 5.76 to 18.66. Again, TDLIM has the highest relative efficiency as expected. Scenario (3) represents the situation where the true interaction structure departs from Tukey's form. We can see now that TDLIM (3.45) is less efficient than CDLIM (6.68) due to the bias introduced in estimating the interaction surface. However, TDLIM is still more efficient than UDLIM (1.00) and BTDLIM (2.77). CDLIM correctly specifies both main effects and interaction effects in this scenario and attains the highest efficiency.

Across scenarios (1)-(2), we note that the squared bias and relative efficiency of BTDLIM always fall between CDLIM and TDLIM, suggesting that BTDLIM successfully performs shrinkage and achieves a better average performance. In addition, we can observe that the BCDLIM (relative efficiency = 6.27, 5.76, 6.17) is slightly less efficient than CDLIM (relative efficiency = 6.82, 6.14, 6.68) across the three scenarios. The difference is due to the flexibility of BCDLIM that accounts for possible departure from Kronecker product type of interaction structure. Scenarios (4) and (5) are situations where UDLIM is the only method that can unbiasedly estimate the interaction surface. As expected, both CDLIM and TDLIM suffer from serious bias and the efficiency gains from dimension reduction diminish substantially. The class of interaction surfaces that CDLIM and TDLIM can describe is restricted. Note that all methods jointly estimate the main effects and interaction effects and thus mis-specifying the interaction effects could possibly distort the estimation of the main effects as they are not orthogonal. BCDLIM is less biased and more efficient than BTDLIM across the two scenarios. Across all scenarios when the main-effects are correctly specified, BCDLIM has the best average performance in terms of estimation efficiency.

We summarize the results where the main effects deviate from a cubic DL function in Table 2. Both CDLIM and TDLIM are seriously biased, largely due to the mis-specification of the main-effect terms. These two methods are the least efficient. If we contrast scenarios (1) and (2), we can observe that misspecification of the main effects not only influences the estimation accuracy of the main-effect DL function, but also the interaction DL surface. BTDLIM is biased across the board as well, with squared bias ranging from 7.39 to 35.50, respectively. It is more efficient than UDLIM only in situations where there is no interaction. BCDLIM is slightly biased across different scenarios with squared bias ranging from 0.09 to 0.52. The BCDLIM leads to gains in efficiency with reduced bias. The relative efficiencies are 3.25, 1.35, 1.78, and 1.34 across the four scenarios. Summarizing the results in Tables 1 and 2, it is clear that the BCDLIM approach has desirable MSE properties across the scenarios, offering a robust and efficient solution to this problem.

## 4. Application

### 4.1. Data Overview and Modeling

We apply the five methods compared in Section 3 to the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) data. We jointly model daily time series of (1) PM<sub>10</sub> and (2) O<sub>3</sub> in association to all-cause nonaccidental mortality counts in Chicago, Illinois for the period between 1987 and 2000. Details with respect to data assembly are available at <http://www.ihapss.jhsph.edu/data/NMMAPS/>. Zanobetti et al. (2000) indicated that it is unlikely that lags beyond two weeks would have substantial effect. We therefore set  $L_1 = L_2 = 14$  for PM<sub>10</sub> and O<sub>3</sub>, respectively.

Previous studies showed that it is crucial to account for meteorologic variables as potential confounders in the analysis of air pollution effects (Welty and Zeger, 2005). Dominici et al. (2005) and Dominici et al. (2007) highlight the need to carefully adjust for a broad set of confounders and explore their functional forms. We specify the adjustment covariates in the same way as Dominici et al. (2005) and focus on choice of the lag structure in our application. We acknowledge that there may be more optimal adjustment models when we introduce interaction effects. Let  $x_{1tk}$ ,  $x_{2tk}$ ,  $y_{tk}$ , and  $z_{tk}$  denote PM<sub>10</sub> level, O<sub>3</sub> level, mortality count, and vector of time-varying covariates, measured on day  $t$  for age group  $k$  for  $t = 1, \dots, 5114$  and  $k = 1, 2, 3$ , respectively. The three age categories are “greater or equal to 75 years old”, “between 65 and 74 years old”, and “less than 65 years old”. PM<sub>10</sub> and O<sub>3</sub> were shared exposures across the three age groups so we have  $x_{\ell k} \equiv x_{\ell}$  for  $\ell = 1, 2$ . For each group  $k$ , we assume that given PM<sub>10</sub>, O<sub>3</sub>, and other time-varying confounders, the mortality count in Chicago on day  $t$  is a Poisson random variable  $Y_{tk}$  with mean  $\mu_{tk}$  such that

$$\begin{aligned} \log(\mu_{tk}) &= X_{1t}^T \beta_1 + X_{2t}^T \beta_2 + X_{It}^T \gamma + z_{tk}^T \alpha \\ &= X_{1t}^T \beta_1 + X_{2t}^T \beta_2 + X_{It}^T \gamma + \alpha_0 + \sum_{j=1}^2 \alpha_{1j} I(k=j) \\ &\quad + \sum_{j=1}^6 \alpha_{2j} I(\text{dow}_t=j) + \text{ns}(\text{temp}_t; 6 \text{ df}, \alpha_3) \\ &\quad + \text{ns}(\overline{\text{temp}}_t^{(3)}; 6 \text{ df}, \alpha_4) + \text{ns}(\text{dpt}_t; 3 \text{ df}, \alpha_5) + \text{ns}(\overline{\text{dpt}}_t^{(3)}; 3 \text{ df}, \alpha_6) \\ &\quad + \text{ns}(t; 98 \text{ df}, \alpha_7) + \text{ns}(t; 14 \text{ df}, \alpha_8) I(k=1) + \text{ns}(t; 14 \text{ df}, \alpha_9) I(k=2) \end{aligned}$$

where  $X_{1t} = (x_{1t}, \dots, x_{1,t-14})^T$ ,  $X_{2t} = (x_{2t}, \dots, x_{2,t-14})^T$ ,  $X_{It} = X_{1t} \otimes X_{2t}$ , and  $\text{ns}(\cdot)$  denotes the natural spline with specified degrees of freedom (df). Predictors  $\text{dow}_t$ ,  $\text{temp}_t$ ,  $\overline{\text{temp}}_t$ ,  $\text{dpt}_t$ , and  $\overline{\text{dpt}}_t$  represent the day of the week, current day's temperature, adjusted average lag 1–3 temperature, current day's dewpoint temperature, and adjusted average lag 1–3 dewpoint temperatures for day  $t$ . The indicator variables allow different baseline mortality rates within each age group and within each day of the week. The smooth term for time ( $t$ ) is to adjust

for long-term trends and seasonality and the choice of 98 df corresponds to 7 df per year over the 14-year time horizon. The last two product terms separate smooth functions of time with 2 df per year for each age group contrast. The primary goal is to estimate the coefficients  $\beta_1$ ,  $\beta_2$ , and  $\gamma$ , while  $\alpha$  is the set of covariate parameters. A four-degree polynomial DL function is applied to both  $\beta_1$  and  $\beta_2$  for CDLIM, TDLIM, BTDLIM, and BCDLIM. The analysis is performed in R version 3.3.1 and the source code is available at [https://github.com/yinhsiuc/NMMAPS\\_DLIM](https://github.com/yinhsiuc/NMMAPS_DLIM). The computational times are provided in Supplementary Appendix A.4 Table 2 and the summary statistics corresponding to PM<sub>10</sub> and O<sub>3</sub> are provided in Supplementary Appendix A.7.

#### 4.2. Estimating Marginal Distributed Lag Function

The quantity  $100 \left\{ \exp[10(\beta_{1i} + x_2^* \sum_{n=0}^{L_2} \gamma_{in})] \right\}$  represents the percentage change in daily mortality associated with a  $10 \mu\text{g}/\text{m}^3$  increase in PM<sub>10</sub> at lag  $i$  when O<sub>3</sub> is at  $x_2^*$  ppb.

Similarly, the quantity  $100 \left\{ \exp[10(\beta_{2j} + x_1^* \sum_{m=0}^{L_1} \gamma_{mj})] \right\}$  represents the percentage change in daily mortality associated with 10 ppb increase in O<sub>3</sub> at lag  $j$  when PM<sub>10</sub> is set at  $x_1^* \mu\text{g}/\text{m}^3$ .

We present the marginal lagged effects of PM<sub>10</sub> and O<sub>3</sub> in Figures 1 and 2. If we look across the panels in Figure 1, we can observe that the fits of UDLIM is under-smoothed and fits of CDLIM and TDLIM are over-smoothed, while those of BTDLIM and BCDLIM are in between. When O<sub>3</sub> is at the summer average level, the over-smoothing of CDLIM and TDLIM results in underestimation of the PM<sub>10</sub> effect at lag 3. For instance, the estimated percentage increases in mortality associated with a  $10 \mu\text{g}/\text{m}^3$  increase in PM<sub>10</sub> at lag 3 when O<sub>3</sub> is at average summer level are 0.53%, 0.14%, 0.03%, 0.23%, and 0.36% for UDLIM, CDLIM, TDLIM, BTDLIM, and BCDLIM, respectively. The lower bounds of 95% confidence/credible intervals for the methods except TDLIM are appreciably above zero. In this situation, shrinkage methods are more desirable since CDLIM and TDLIM mis-specify the DL function and potentially underestimate the relative lag effects. Similarly, we observe slight over-smoothness of CDLIM and TDLIM on O<sub>3</sub> effect in Figure 2. However, the degree of underestimation of O<sub>3</sub> effect at early lags is smaller. More similar DL functions across all methods except UDLIM indicates that the potential misspecification of the DL function by using CDLIM and TDLIM is minimal.

We present the marginal DL functions of PM<sub>10</sub> and O<sub>3</sub> by integrating out the other pollutant in Figure 3. Similar to earlier findings, shrinkage is more needed for PM<sub>10</sub> as CDLIM and TDLIM tend to oversmooth the DL function in this situation. In addition, we observe that the DL function for PM<sub>10</sub> starts from negative, grows to zero and peaks at lag 3, while the DL function for O<sub>3</sub> is greater than zero at lag 0 and peaks at lag 2. The earlier peak for O<sub>3</sub> compared to PM<sub>10</sub> suggests a more acute effect of O<sub>3</sub> than PM<sub>10</sub> with an earlier window of susceptibility. We also observe that the UDLIM fits of O<sub>3</sub> fluctuate more drastically than the UDLIM fits of PM<sub>10</sub>. This is explained by the stronger autocorrelation of the O<sub>3</sub> time series and smoothing the DL function is certainly needed and preferred in this case. We can observe that some of the estimated lagged effects are negative at larger lags for PM<sub>10</sub>. This phenomenon is noted as mortality displacement (Zanobetti et al., 2000) and has been discovered in previous studies. Mortality displacement, also referred to as harvesting effect

(Zanobetti et al., 2002), is the temporal shift of mortality. Usually a higher mortality rate due to the deaths of frail individuals a couple of days after a high air pollution episode is followed by compensatory reduction in mortality rate due to the death of the more frail individuals.

### 4.3. Assessing Interaction Effects

Within each panel of Figures 1 and 2, we notice that the estimated DL functions of one pollutant vary with the level of the other pollutant, indicating that  $PM_{10}$  might moderate the effect of  $O_3$  and vice versa. For UDLIM, CDLIM, and TDLIM, we conducted likelihood ratio test to test for  $PM_{10}$ - $O_3$  interactions and the p-values are  $1.65 \times 10^{-11}$  (DF = 225),  $5.33 \times 10^{-9}$  (DF = 25), and  $< 10^{-4}$  (DF = 1), respectively. The precision of the p-value of TDLIM is only up to  $10^{-4}$  due to finite bootstrap samples. For the two shrinkage methods BTDLIM and BCDLIM, we computed the difference in deviance information criterion (DIC) (Spiegelhalter et al., 2002) between the models with and without interaction. The DIC differences are 25.56 and 68.35, respectively. It is difficult to determine a clear threshold of DIC difference for model selection (Plummer, 2008). However, models with smaller DIC are generally preferred when DIC differences are greater than 10. Coupled with the p-values obtained from the frequentist approaches, we conclude that the interaction between  $PM_{10}$  and  $O_3$  is evident.

From Figures 1 and 2, we can see that the summer curves are above winter curves suggesting that  $PM_{10}$  and  $O_3$  have synergistic effects on each other. Furthermore, we observe that the gaps between the curves of the three quartiles decrease beyond lag 6 and that happens across the board. The interaction between  $PM_{10}$  and  $O_3$  occurs at early lags. We added a dotted blue curve in each panel for the estimated DL function from a single-pollutant analysis (i.e. models with  $PM_{10}$  alone or  $O_3$  alone), representing the “average” DL effects if we disregard the interaction effect between the two pollutants. The evidence in favor of looking at  $PM_{10}$  and  $O_3$  jointly is compelling.

## 5. Discussion

In analyzing NMMAPS data, we demonstrated the importance of accounting for interaction between the  $PM_{10}$  and  $O_3$  time series when modeling the joint pollution effect on mortality. Two major pieces of evidence support the existence of pollutant-pollutant interaction - (1) the marginal DL function of one pollutant varies when the level of the other one changes, and (2) the small p-values from frequentist approaches and the large DICs from the Bayesian approaches suggest evidence in favor of  $PM_{10} \times O_3$  interaction. This adds to the finding of previous studies that supported the idea of a plausible synergism involving  $PM_{10}$  and  $O_3$  (Mauderly and Samet, 2009).

In this article, we presented five different strategies to model lagged effects of two pollutants in a joint model. We reviewed two existing frequentist methods UDLIM and CDLIM, and we proposed frequentist TDLIM using Tukey’s interaction structure, its Bayesian version, and a Bayesian approach to perform shrinkage between UDLIM and CDLIM. There are two major novelties. We adopted Tukey’s one-degree-of-freedom interaction structure to parsimoniously model two-way interactions. The estimation is efficient and the interaction



testing is powerful. We also introduced the Bayesian version of TDLIM (i.e. BTDLIM) and the Bayesian version of CDLIM (i.e. BCDLIM). These Bayesian models allow for departure from a pre-specified structure of DL function/surface, and have been shown to be robust to mis-specification. They are data-adaptive and able to achieve bias-variance trade-off.

Each of the five approaches has some limitations that we discuss below. UDLIM is unbiased but potentially less efficient, especially when the autocorrelation between serial pollution measurement is large. CDLIM imposes some structure to constrain the lag coefficients and can potentially achieve greater estimation precision. In practice, we recommend a DL structure no more complex than a cubic polynomial as the default choice since it is usually sufficient to capture the observed non-linear patterns as a function of the lags. Nonetheless, when the DL structure is misspecified, the model-dependent CDLIM estimator can be seriously biased. Tukey's type interaction is mostly used for hypothesis testing rather than estimation in previous research. Expressing interaction effects as a scaled product of the corresponding main effects implies that the interaction effects can be non-zero only when the main effects are non-zero. This hierarchical feature results in lack of identifiability for the scaled parameter in Tukey's model when the main effects are not present. In addition, Tukey's model is not invariant to location shifts. Different centering schemes lead to different estimates of the scaled parameter  $\eta$  and no universal remedy exists.

The hierarchical Bayesian model BCDLIM is robust to mis-specification of the DL structure. The data-adaptive shrinkage can be regarded as an automatic procedure to attain a balance between the more general UDLIM and the more constrained CDLIM. The full-rank transformation on UDLIM imposes smoothness on the shrinkage path and any *a priori* knowledge about the DL structure can be incorporated. It is important to note that BCDLIM can be extended to explore higher-order interaction and multiple-pollutant scenarios. We also tried to adapt HDDLM to two-pollutant scenarios. However, the unmodified predictive process interpolator (Banerjee et al., 2008), the major technique used in HDDLM for dimension reduction (Finley et al., 2009), leads to overly smooth DL functions/surfaces which result in seriously biased estimates. We therefore decided to not include HDDLM in this manuscript.

The two-pollutant DLIMs can be directly combined with DLNMs (Gasparrini et al., 2010) to flexibly capture non-linear exposure-outcome associations by replacing the linear terms in DLM specifications with some basis functions (e.g. B-spline). As indicated by He et al. (2015), failing to account for nonlinear main effects may lead to spurious detection of linear interaction terms. However, when the covariates are correlated as in our application, the signals from nonlinear main effects and linear interaction effects may be indistinguishable. In addition, some regularization may be needed in this high-dimensional situation to avoid overfitting. We consider this line of extension for future research.

The two-pollutant DLIM approaches introduced in this article can also be extended to multi-pollutant situations where up to two-way interactions are considered. If one would like to consider higher-order interactions and/or nonlinear interactions, extension of tree-based approaches such as CART and Bayesian kernel machine regression (BKMR) can be



promising. In some situations, choosing the most important pollutants among multiple candidates that are associated with a health outcome is the primary goal.

In real-world settings, it is usually difficult to validate the underlying assumptions of a model-based estimator. The notion of data-adaptive shrinkage is attractive when no single estimator is universally optimal. When facing uncertainty, robust models such as BCDLIM that possesses better average performance are more desirable. BCDLIM can potentially be extended to areas outside environmental epidemiology. We hope our work will lead to more attempts in developing two-dimensional and multi-dimensional DLIM in the future.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

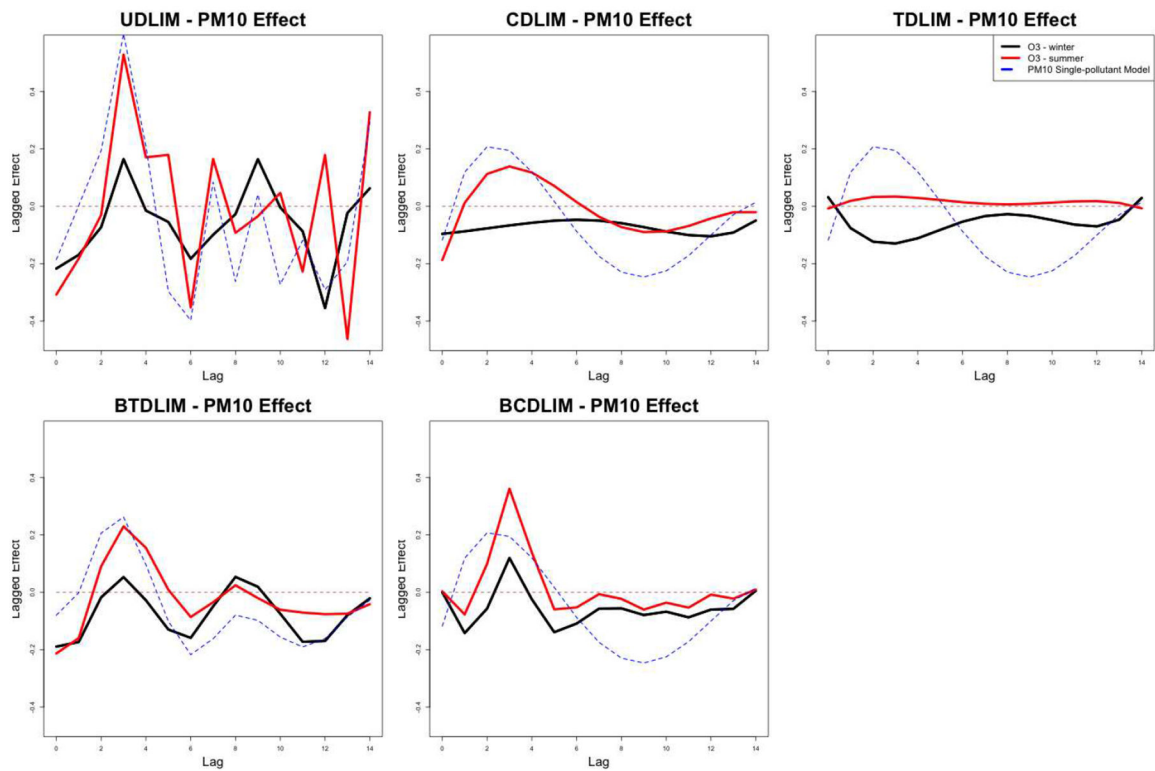
## Acknowledgements

The research is supported by NSF DMS 1406712 and NIH grant ES 20811.

## References

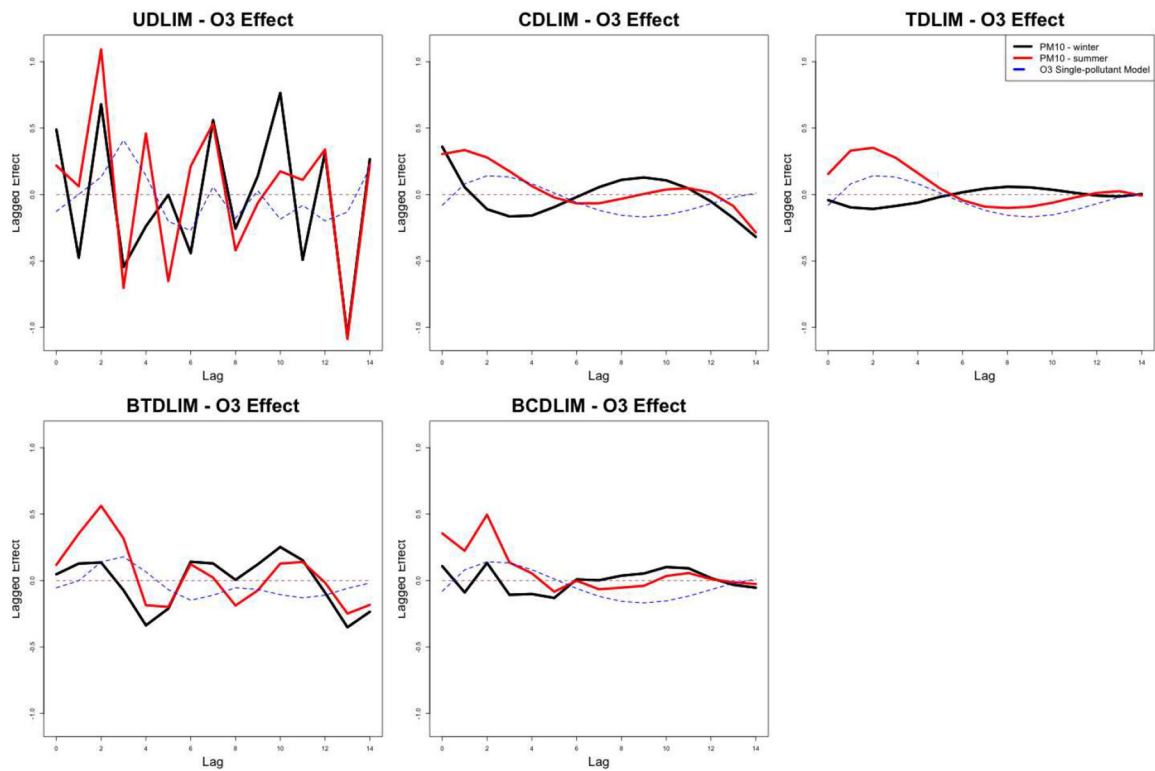
- Almon S (1965). The distributed lag between capital appropriations and expenditures. *Econometrica: Journal of the Econometric Society* pages 178–196.
- Banerjee S, Gelfand AE, Finley AO, and Sang H (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70, 825–848. [PubMed: 19750209]
- Billionnet C, Sherrill D, Annesi-Maesano I, et al. (2012). Estimating the health effects of exposure to multi-pollutant mixture. *Annals of epidemiology* 22, 126–141. [PubMed: 22226033]
- Bobb JF, Valeri L, Henn BC, Christiani DC, Wright RO, Mazumdar M, Godleski JJ, and Coull BA (2014). Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. *Biostatistics* page kxu058.
- Chatterjee N, Kalaylioglu Z, Moslehi R, Peters U, and Wacholder S (2006). Powerful multilocus tests of genetic association in the presence of gene-gene and gene-environment interactions. *The American Journal of Human Genetics* 79, 1002–1016. [PubMed: 17186459]
- Dominici F, McDermott A, Daniels M, Zeger SL, and Samet JM (2005). Revised analyses of the national morbidity, mortality, and air pollution study: mortality among residents of 90 cities. *Journal of Toxicology and Environmental Health, Part A* 68, 1071–1092. [PubMed: 16024489]
- Dominici F, Peng RD, Barr CD, and Bell ML (2010). Protecting human health from air pollution: shifting from a single-pollutant to a multi-pollutant approach. *Epidemiology (Cambridge, Mass.)* 21, 187.
- Dominici F, Peng RD, Bell ML, Pham L, McDermott A, Zeger SL, and Samet JM (2006). Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *Jama* 295, 1127–1134. [PubMed: 16522832]
- Dominici F, Peng RD, Ebisu K, Zeger SL, Samet JM, and Bell ML (2007). Does the effect of pm10 on mortality depend on pm nickel and vanadium content? a reanalysis of the nmmaps data. *Environmental health perspectives* 115, 1701. [PubMed: 18087586]
- Finley AO, Sang H, Banerjee S, and Gelfand AE (2009). Improving the performance of predictive process modeling for large datasets. *Computational statistics & data analysis* 53, 2873–2884. [PubMed: 20016667]
- Gasparrini A, Armstrong B, and Kenward MG (2010). Distributed lag non-linear models. *Statistics in medicine* 29, 2224–2234. [PubMed: 20812303]
- Gelman A et al. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian analysis* 1, 515–534.

- He Z, Zhang M, Lee S, Smith JA, Guo X, Palmas W, Kardina SL, Roux AV, and Mukherjee B (2015). Set-based tests for genetic association in longitudinal studies. *Biometrics* 71, 606–615. [PubMed: 25854837]
- Heaton MJ and Peng RD (2014). Extending distributed lag models to higher degrees. *Biostatistics* 15, 398–412. [PubMed: 23990524]
- Hu W, Mengersen K, McMichael A, and Tong S (2008). Temperature, air pollution and total mortality during summers in Sydney, 1994–2004. *International journal of biometeorology* 52, 689–696. [PubMed: 18506490]
- Lunn D, Spiegelhalter D, Thomas A, and Best N (2009). The bugs project: Evolution, critique and future directions. *Statistics in medicine* 28, 3049–3067. [PubMed: 19630097]
- Maity A, Carroll RJ, Mammen E, and Chatterjee N (2009). Testing in semiparametric models with interaction, with applications to gene–environment interactions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71, 75–96. [PubMed: 19838317]
- Marx BD and Eilers PH (1998). Direct generalized additive modeling with penalized likelihood. *Computational Statistics & Data Analysis* 28, 193–209.
- Mauderly JL (1993). Toxicological approaches to complex mixtures. *Environmental health perspectives* 101, 155.
- Mauderly JL and Samet JM (2009). Is there evidence for synergy among air pollutants in causing health effects? *Environmental health perspectives* 117, 1. [PubMed: 19165380]
- Muggeo VM (2007). Bivariate distributed lag models for the analysis of temperature-by-pollutant interaction effect on mortality. *Environmetrics* 18, 231–243.
- Plummer M (2008). Penalized loss functions for Bayesian model comparison. *Biostatistics* 9, 523–539. [PubMed: 18209015]
- Pope CA and Dockery DW (2006). Health effects of fine particulate air pollution: lines that connect. *Journal of the air & waste management association* 56, 709–742. [PubMed: 16805397]
- Pope CA, Dockery DW, and Schwartz J (1995). Review of epidemiological evidence of health effects of particulate air pollution. *Inhalation toxicology* 7, 1–18.
- Roberts S (2005). An investigation of distributed lag models in the context of air pollution and mortality time series analysis. *Journal of the Air & Waste Management Association* 55, 273–282. [PubMed: 15828669]
- Spiegelhalter DJ, Best NG, Carlin BP, and Van Der Linde A (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64, 583–639.
- Sun Z, Tao Y, Li S, Ferguson KK, Meeker JD, Park SK, Batterman SA, and Mukherjee B (2013). Statistical strategies for constructing health risk models with multiple pollutants and their interactions: possible choices and comparisons. *Environ Health* 12, 85. [PubMed: 24093917]
- Tibshirani R (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* pages 267–288.
- Tukey JW (1949). One degree of freedom for non-additivity. *Biometrics* 5, 232–242.
- Welty LJ, Peng R, Zeger S, and Dominici F (2009). Bayesian distributed lag models: estimating effects of particulate matter air pollution on daily mortality. *Biometrics* 65, 282–291. [PubMed: 18422792]
- Welty LJ and Zeger SL (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. [PubMed: 15961590]
- Zanobetti A, Schwartz J, Samoli E, Gryparis A, Touloumi G, Atkinson R, Le Tertre A, Bobros J, Celko M, Goren A, et al. (2002). The temporal pattern of mortality responses to air pollution: a multicity assessment of mortality displacement. *Epidemiology* 13, 87–93. [PubMed: 11805591]
- Zanobetti A, Wand M, Schwartz J, and Ryan L (2000). Generalized additive distributed lag models: quantifying mortality displacement. *Biostatistics* 1, 279–292. [PubMed: 12933509]



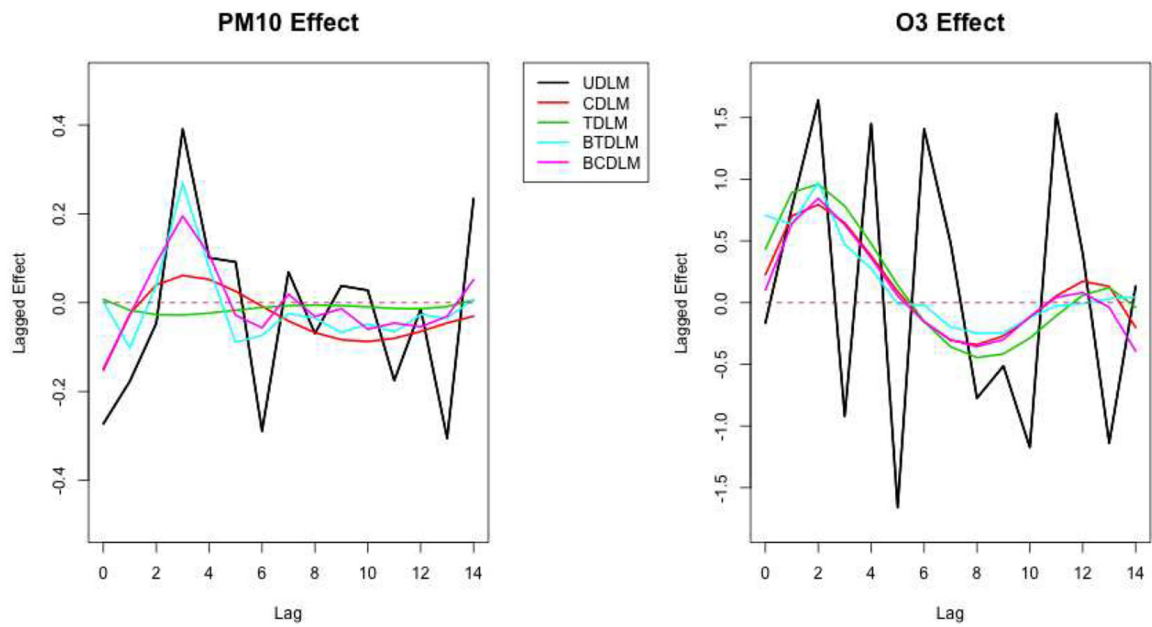
**Fig. 1.**

Estimated distributed lag functions up to 14 days for the effects of  $PM_{10}$  on mortality in Chicago, Illinois from 1987 to 2000 based on data from the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) under five estimation methods. In all panels,  $O_3$  levels are fixed at the average series levels in winter (black) and the average series levels in summer (red). The blue curve represents the estimated DL function relative to  $PM_{10}$  when  $O_3$  is disregarded in a single-pollutant model for  $PM_{10}$ . Lag effects are presented as the percentage change in mortality associated with an  $10 \mu g/m^3$  increase in  $PM_{10}$ . The five estimation methods are unconstrained distributed lag interaction model (UDLIM), constrained distributed lag interaction model (CDLIM), Tukey's distributed lag interaction model (TDLIM), Bayesian Tukey's distributed lag interaction model (BTDLIM), and Bayesian constrained distributed lag interaction model (BCDLIM).



**Fig. 2.**

Estimated distributed lag functions up to 14 days for the effects of  $O_3$  on mortality in Chicago, Illinois from 1987 to 2000 based on data from the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) under five estimation methods. In all panels,  $PM_{10}$  levels are fixed at the average series levels in winter (black) and the average series levels in summer (red). The blue curve represents the estimated DL function relative to  $O_3$  when  $PM_{10}$  is disregarded in a single-pollutant model for  $O_3$ . Lag effects are presented as the percentage change in mortality associated with with an 10 ppb increase in  $O_3$ . The five estimation methods are unconstrained distributed lag interaction model (UDLIM), constrained distributed lag interaction model (CDLIM), Tukey's distributed lag interaction model (TDLIM), Bayesian Tukey's distributed lag interaction model (BTDLIM), and Bayesian constrained distributed lag interaction model (BCDLIM).



**Fig. 3.** Estimated distributed lag functions up to 14 days for the effects of  $\text{PM}_{10}$  (left) and  $\text{O}_3$  (right) on mortality in Chicago, Illinois from 1987 to 2000 based on data from the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) under five estimation methods. The DL functions presented here are estimated by integrating out the other pollutant. Lag effects are presented as the percentage change in mortality associated with an  $10 \mu\text{g}/\text{m}^3$  increase in  $\text{PM}_{10}$  and a 10 ppb increase in  $\text{O}_3$ , respectively. The five estimation methods are unconstrained distributed lag interaction model (UDLIM), constrained distributed lag interaction model (CDLIM), Tukey's distributed lag interaction model (TDLIM), Bayesian Tukey's distributed lag interaction model (BTDLIM), Bayesian constrained distributed lag interaction model (BCDLIM).

**Table 1.**

Empirical squared bias and empirical relative efficiency (measured with respect to the mean squared error of UDLIM estimate) of marginal lagged effects across five different two-dimensional distributed lag interaction models based on 1000 simulation datasets. The lagged effects of the both exposures are generated from the same cubic DL function.

<b>Interaction Structure</b>	<b>Metric</b>	<b>UDLIM</b>	<b>CDLIM</b>	<b>TDLIM</b>	<b>BTDLIM</b>	<b>BCDLIM</b>
(1) No Interaction	Squared Bias	0.02	0.00	0.19	0.13	0.00
	Relative Efficiency	1.00	6.82	19.24	8.09	6.27
(2) Tukey's Structure	Squared Bias	0.01	0.00	0.01	0.01	0.00
	Relative Efficiency	1.00	6.14	18.66	6.71	5.76
(3) Kronecker Product	Squared Bias	0.02	0.00	1.05	0.90	0.00
	Relative Efficiency	1.00	6.68	3.45	2.77	6.17
(4) Sparse	Squared Bias	0.00	66.22	67.14	1.43	0.08
	Relative Efficiency	1.00	0.07	0.07	1.71	2.80
(5) Unstructured	Squared Bias	0.00	93.08	93.98	1.08	0.09
	Relative Efficiency	1.00	0.05	0.05	1.88	2.70

**Table 2.**

Empirical squared bias and empirical relative efficiency (measured with respect to the mean squared error of UDLIM estimate) of marginal lagged effects across five different two-dimensional distributed lag interaction models based on 1000 simulation datasets. The lagged effects of the both exposures are generated from the same cubic-like DL function (moderate departure from a cubic function).

<b>Interaction Structure</b>	<b>Metric</b>	<b>UDLIM</b>	<b>CDLIM</b>	<b>TDLIM</b>	<b>BTDLIM</b>	<b>BCDLIM</b>
(1) No Interaction	Squared Bias	0.02	69.51	70.03	7.39	0.10
	Relative Efficiency	1.00	0.24	0.25	1.59	3.25
(2) Tukey's Structure	Squared Bias	0.01	990.83	1023.84	35.50	0.09
	Relative Efficiency	1.00	0.00	0.00	0.05	1.35
(4) Sparse	Squared Bias	0.01	210.32	215.94	10.80	0.52
	Relative Efficiency	1.00	0.02	0.02	0.35	1.78
(5) Unstructured	Squared Bias	0.01	989.93	1019.06	31.83	0.10
	Relative Efficiency	1.00	0.00	0.00	0.04	1.34