Published in final edited form as:

J R Stat Soc Ser C Appl Stat. 2016 November ; 65(5): 731–753. doi:10.1111/rssc.12144.

Multipollutant measurement error in air pollution epidemiology studies arising from predicting exposures with penalized regression splines

Silas Bergen,

Winona State University, Winona, MN

Lianne Sheppard,

University of Washington, Seattle, WA

Joel D. Kaufman, and

University of Washington, Seattle, WA

Adam A. Szpiro

University of Washington, Seattle, WA

Summary

Air pollution epidemiology studies are trending towards a multi-pollutant approach. In these studies, exposures at subject locations are unobserved and must be predicted using observed exposures at misaligned monitoring locations. This induces measurement error, which can bias the estimated health effects and affect standard error estimates. We characterize this measurement error and develop an analytic bias correction when using penalized regression splines to predict exposure. Our simulations show bias from multi-pollutant measurement error can be severe, and in opposite directions or simultaneously positive or negative. Our analytic bias correction combined with a non-parametric bootstrap yields accurate coverage of 95% confidence intervals. We apply our methodology to analyze the association of systolic blood pressure with $PM_{2.5}$ and NO_2 in the NIEHS Sister Study. We find that NO_2 confounds the association of systolic blood pressure with $PM_{2.5}$ and vice versa. Elevated systolic blood pressure was significantly associated with increased $PM_{2.5}$ and decreased NO_2 . Correcting for measurement error bias strengthened these associations and widened 95% confidence intervals.

1. Introduction

Air pollution epidemiology is trending towards a multi-pollutant approach (Dominici et al., 2010; Billonnet et al., 2012; Vedal and Kaufman, 2011). Dominici et al. (2010) argue that members of long-term health cohorts are exposed to a complex mixture of pollutants that together affect health outcomes. Understanding how health outcomes are associated with pollutant mixtures is necessary to inform policy decisions aimed at managing multiple pollutant levels (Vedal and Kaufman, 2011).

One of the challenges of the multi-pollutant approach in long-term cohort studies is obtaining multivariate pollutant predictions at health cohort locations, as the true values are usually unobserved. To this end a first-stage exposure model is often built using exposures

observed at misaligned monitoring locations. A subsequent challenge is characterizing and correcting for measurement error when the predicted rather than the true exposures are used to estimate health effects in a second-stage health model. Zeger et al. (2000) outline some general conclusions based on multivariate classical measurement error. They suggest that in general the more poorly a pollutant is measured, the more attenuated its effect estimate will be. Exceptions to this may occur when the measurement errors are negatively correlated, in which case the attenuation of a poorly-measured pollutant's health effect estimate may induce upward bias of a well-measured pollutant's effect estimate. Schwartz and Coull (2003) develop multi-pollutant regression calibration to obtain unbiased health effect estimates under classical multivariate measurement error, which Zeka and Schwartz (2004) applied to the National Morbidity and Mortality Air Pollution Study (NMMAPS) and discovered a previously unobserved effect of carbon monoxide on daily death adjusted for PM_{10} . Strand et al. (2014) extended regression calibration (Carroll, 2006) to estimate the association in asthmatic children of an inflammation biomarker with an interactive form of smoking and $PM_{2.5}$. However, these methodologies do not address the spatial measurement error that dominates in air pollution epidemiology studies.

Much has been developed characterizing spatial measurement error for a single-pollutant (Madsen et al., 2008; Gryparis et al., 2009; Szpiro et al., 2011; Bergen et al., 2013; Lopiano et al., 2013, 2014; Szpiro and Paciorek, 2013; Bergen and Szpiro, 2015). Such approaches characterize the impact of measurement error in the second-stage health model by analyzing the statistical characteristics of the first-stage exposure model. Szpiro et al. (2011) develop the parametric and parameter bootstraps which Bergen et al. (2013) apply in a case study estimating the health effects on carotid intima-media thickness of four $PM_{2.5}$ components. This approach is valid assuming the exposure surface follows a universal kriging model, and could be extended to the multi-pollutant setting. However one would need to correctly specify a multivariate spatial random effect. This would require specifying cross-correlations in addition to the individual spatial correlations, a daunting prospect. Furthermore Szpiro and Paciorek (2013) reasoned that viewing the exposure surface as fixed and the monitoring and subject locations as random is a more realistic paradigm, motivating modeling spatial structure with spatially referenced geographic covariates and basis functions rather than spatially correlated random effects. Bergen and Szpiro (2015) described how in this context universal kriging can be viewed as a full-rank penalized spline, which can be approximated with fixed-rank penalized regression splines (Ruppert et al., 2003; Wakefield, 2013). They developed an analytic bias calculation useful for obtaining unbiased health effect estimation when modeling exposure with penalized regression splines.

We extend the methods of Bergen and Szpiro (2015) to the multi-pollutant case, characterizing multi-pollutant measurement error when using penalized regression splines to predict exposures at health subject locations. Although our methods apply to a general class of penalized regression models, our treatment focuses on fitting separate penalized regression models to each pollutant. We develop our methodology in the context of analyzing the association of systolic blood pressure (SBP) with particulate matter $2.5 \mu g/m^3$ in diameter ($PM_{2.5}$) and nitrogen dioxide (NO_2) in the Sister Study cohort (NIEHS, 2013). Chan et al. (Provisionally accepted) previously analyzed this association, but did not account for measurement error.

Our paper is organized as follows. In Section 2 we describe the Sister Study cohort used for our case study. In Section 3 we describe assumptions and modeling strategies. In Section 4 we decompose the error into Berkson- and classical-like components and derive an analytic bias correction that accounts for both. In Section 5 we discuss bias estimation approaches, exposure model selection and variance estimation. In Section 6 we perform simulations to investigate the impact of measurement error under a number of different scenarios and to demonstrate the effectiveness of our bias correction in achieving well-calibrated inference. In Section 7 we apply our methodology to analyze the joint association of $\text{PM}_{2.5}$ and NO_2 with SBP in the Sister Study cohort (NIEHS, 2013). We conclude with a discussion of our results in Section 8.

2. Case study: SBP, $\text{PM}_{2.5}$ and NO_2 in the NIEHS Sister Study

We develop our methodology in the context of analyzing the association of SBP with $\text{PM}_{2.5}$ and NO_2 in the Sister Study of the National Institute of Environmental Health Sciences (NIEHS, 2013). The Sister Study is a large nationwide (including Puerto Rico) prospective cohort study of women between the ages of 35 and 74 enrolled between 2003 and 2009, where each participating woman was the sister of a woman with breast cancer. The intent of the Sister Study is to identify genetic and environmental risk factors of breast cancer and other diseases.

Previously, Chan et al. (Provisionally accepted) analyzed the association between baseline SBP and predicted annual 2006 average $\text{PM}_{2.5}$ and NO_2 in 43,629 Sister Study participants in the continental United States. $\text{PM}_{2.5}$ predictions were derived from a regionalized universal kriging model (Sampson et al., 2013) and NO_2 predictions were derived from a national satellite-based land-use regression model (Novotny et al., 2011). Fitting separate health models they estimated a 1.4 mmHg change in SBP for a $10\text{-}\mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ (95% CI: 0.6, 2.3; $p < 0.001$) and a 0.2 mmHg change in SBP for a 10-ppb increase in NO_2 (95% CI: 0.0, 0.5; $p = 0.10$). These health models controlled for demographics (age and race); socioeconomic status (SES) (household income, education, marital status, working more than 20 hours per week outside the home, perceived stress score, SES z-score as described by Diez Roux et al. (2001)); large-scale spatial structure (urban rural continuum code and a 10-degree-of-freedom thin-plate spline); cardiovascular disease risk factors (body mass index, waist-to-hip ratio, smoking status, alcohol use, self-reported history of diabetes, and self-reported history of hypercholesterolemia); and use of blood pressure medication. However, the correlation between $\text{PM}_{2.5}$ and NO_2 in this analysis was 0.37. Modeling SBP on $\text{PM}_{2.5}$ without adjusting for NO_2 and vice versa could result in bias from residual confounding in the estimated health effects. When modeling both exposures together as main effects in the health model, Chan et al. (Provisionally accepted) estimated a 1.6 mmHg change in SBP for a $10\text{-}\mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ holding NO_2 constant (95% CI: 0.5, 2.6; $p < 0.001$) and a -0.1 mmHg change in SBP for a 10-ppb increase in NO_2 holding $\text{PM}_{2.5}$ constant (95% CI: -0.4 , 0.3; $p = 0.65$). These analyses did not account for measurement error, and as a result may suffer from bias and invalid standard error estimates.

3. Analytic framework

3.1. Data generating mechanisms

Our assumed data generating mechanisms follow those described in Szpiro and Paciorek (2013) and Bergen and Szpiro (2015). We describe our methodology in the context of our case study, emphasizing that extension to higher-dimension pollutant vectors is easily accomplished at the cost of notational simplicity. Let $\mathbf{s}_1, \dots, \mathbf{s}_n$ denote n subject locations drawn independently from an unknown spatial distribution function $G(\cdot)$. In the context of our case study, $\mathbf{s}_k$ denotes the residence of the k^{th} Sister Study participant, and $n = 43,629$. Given subject location, the true unobserved exposures are $\mathbf{x}_k \equiv (x_{k1}, x_{k2})^T$, where x_{k1} and x_{k2} are the $\text{PM}_{2.5}$ and NO_2 exposures, respectively, at the k^{th} participant's residence. We assume that the unobserved exposure vector follows:

$$\mathbf{x}_k = \Phi(\mathbf{s}_k) + \boldsymbol{\eta}_k.$$

Here $\Phi(\mathbf{s}_k) \equiv \{\Phi_1(\mathbf{s}_k), \Phi_2(\mathbf{s}_k)\}^T$ is a fixed function of space that denotes all characteristics of $\text{PM}_{2.5}$ and NO_2 that can be modeled with geographic covariates (such as distances to road, population density, and land use) as well as spatially referenced basis functions. As $\text{PM}_{2.5}$ and NO_2 are correlated across space, we assume $\text{Cor}(\Phi_1(\mathbf{s}), \Phi_2(\mathbf{s})) \in (-1, 1)$. The $\boldsymbol{\eta}_k^T \equiv \{\eta_{k1}, \eta_{k2}\}$ are not spatially dependent, but rather represent unexplainable “white noise” in the pollutants.

In order to model exposure, we use $n^* = n_0^* + n_1^* + n_2^*$ monitoring locations where at least one pollutant is observed. Let $\mathbf{s}_1^*, \dots, \mathbf{s}_{n^*}^*$ denote these locations, which are drawn independently of each other and of the subject locations. It is possible for monitoring locations to exist at some subject locations, and in fact this is the case for some studies (Cohen et al., 2009). However even in Cohen et al. (2009) a great majority of subject locations are not collocated with a monitor. In the Sister Study analysis, the monitoring locations are completely misaligned from the subject locations. Here n_j^* of the locations belong to set $\mathcal{S}_j^*$ for $j \in \{0, 1, 2\}$. $\mathcal{S}_0^*$ denotes the set of monitoring locations where both pollutants are observed, while for $j \in \{1, 2\}$ $\mathcal{S}_j^*$ denotes the set of monitoring locations where only pollutant j is observed. To model $\text{PM}_{2.5}$ and NO_2 in our case study, we used $n_0^* = 178$ monitoring locations where both $\text{PM}_{2.5}$ and NO_2 were observed; $n_1^* = 859$ locations where only $\text{PM}_{2.5}$ was observed; and $n_2^* = 180$ locations where NO_2 alone was observed. Let π_j^* denote the probability a randomly drawn monitoring location belongs to set $\mathcal{S}_j^*$, and $H_j(\cdot)$ the unknown distribution of monitoring locations in $\mathcal{S}_j^*$. In this paper we will assume $H_j(\cdot) = G(\cdot)$ for all j . This is a strong assumption but we employ it for simplicity and to focus on the measurement error methodology, discussing implications of violating this assumption in Section 8.

Our methodology requires a nonzero number of locations where we observed both pollutants. Accordingly we assume $\pi_0^* > 0$, while π_j^* may equal 0 for $j = 1$ and/or 2; this implies $0 < n_0^* \leq n^*$. As we describe in Section 5.1 this assumption is necessary to ensure we can estimate bias from measurement error. We denote $\mathbf{x}_i^* = \Phi(\mathbf{s}_i^*) + \boldsymbol{\eta}_i^*$ as the complete (but

not necessarily completely observed) vector of exposures at monitoring location $\mathbf{s}_i^*$, for $i = 1, \dots, n^*$.

In order to model and predict exposure we have spatially-referenced geographic covariates and basis functions at each monitor and subject location. The following notation admits different covariates and basis functions for modeling a specific pollutant. Let $\mathbf{p}_j(\mathbf{s}_i^*)$ denote a $p_j \times 1$ vector of geographic covariates such as distance to road or land use features, and $\mathbf{q}_j(\mathbf{s}_i^*)^T$ a $q_j \times 1$ vector of spatial basis functions for $j \in \{1, 2\}$. Let

$\mathbf{r}_j(\mathbf{s}_i^*)^T = \{1(j=1)\mathbf{p}_1(\mathbf{s}_i^*)^T, 1(j=1)\mathbf{q}_1(\mathbf{s}_i^*)^T, 1(j=2)\mathbf{p}_2(\mathbf{s}_i^*)^T, 1(j=2)\mathbf{q}_2(\mathbf{s}_i^*)^T\}$, i.e. $\mathbf{r}_j(\mathbf{s}_i^*)$ is a vector of length $r = p_1 + q_1 + p_2 + q_2$. Thus the last $p_2 + q_2$ elements of $\mathbf{r}_1$ are all 0, and the first $p_1 + q_1$ elements of $\mathbf{r}_2$ are all 0. The vector $\mathbf{r}_j(\mathbf{s}^*)$ is admittedly cumbersome, but facilitates more concise representation of the exposure model defined in Section 3.2.1.

Finally, we define the $r \times 2$ location-specific model matrix $R(\mathbf{s}_i^*) = \{\mathbf{r}_1(\mathbf{s}_i^*), \mathbf{r}_2(\mathbf{s}_i^*)\}$. Analogously we define $\mathbf{r}_j(\mathbf{s}_k)$ and $R(\mathbf{s}_k)$ at subject locations; these are the geographic covariates and spatial basis functions we use to predict at subject locations after building the exposure model using monitoring locations.

Given true exposure at subject locations, we assume SBP follows the linear health model

$$y_k = \beta_0 + \boldsymbol{\beta}^T \mathbf{x}_k + \boldsymbol{\beta}_z^T \mathbf{z}_k + \varepsilon_k.$$

The ε_k are independent but not necessarily identically distributed random variables with mean zero, and are also independent of the $\mathbf{x}_k$ and $\mathbf{z}_k$. The ε_k could be heteroscedastic, and are not assumed to be normally distributed. As in Szpiro and Paciorek (2013) and Bergen and Szpiro (2015), the subject-specific covariates $\mathbf{z}_k$ are defined as

$$\mathbf{z}_k = \boldsymbol{\Theta}(\mathbf{s}_k) + \boldsymbol{\zeta}_k,$$

where $\boldsymbol{\Theta}(\mathbf{s}_k) = (\Theta_{k1}, \dots, \Theta_{km})^T$ represents the spatially structured components of the subject-specific covariates (e.g. body mass index, socioeconomic status, or thin plate splines used to adjust for unmeasured spatial confounding), and the $\boldsymbol{\zeta}_k = (\zeta_{k1}, \dots, \zeta_{km})$ are random m -vectors, independent between subjects and independent of $\boldsymbol{\eta}_k$, that do not depend on space. Interest lies in estimating the vector of pollutant health effects $\boldsymbol{\beta}$; as we do not directly observe exposure this requires building a prediction model using the $\mathbf{x}_i^*$ and $R(\mathbf{s}_i^*)$ and predicting at locations $\mathbf{s}_k$ using $R(\mathbf{s}_k)$. The following section describes this exposure model.

3.2. Penalized regression exposure model

In our simulations and data analysis we fit separate penalized regression models for each pollutant. These models are straightforward to fit and allow us to focus on measurement error instead of exposure modeling. However, our methodology applies to a more general class of penalized regression models. In what follows we define the general class of penalized regression models for predicting multiple pollutants and describe how separate penalized regression models fall into this class. In reality more sophisticated models that

better exploit correlation between pollutants may be desirable, but we defer treatment of such models to future work.

3.2.1. General form—Given a monitoring data set consisting of observed pollutant exposures and covariates we estimate penalized regression coefficients by minimizing the multi-pollutant analogue of the penalized sum-of-squares. Specifically, given $r \times r$ penalty matrix Λ and 2×2 weighting matrix Π , we obtain $\hat{\boldsymbol{\gamma}}$ where:

$$\hat{\boldsymbol{\gamma}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \frac{1}{n^*} \sum_{i=1}^{n^*} (\mathbf{x}_i^* - R(\mathbf{s}_i^*)^T \boldsymbol{\theta})^T W(\mathbf{s}_i^*) \Pi W(\mathbf{s}_i^*) (\mathbf{x}_i^* - R(\mathbf{s}_i^*)^T \boldsymbol{\theta}) + \boldsymbol{\theta}^T \Lambda \boldsymbol{\theta}. \quad (1)$$

Conventionally, Equation 1 is written without the $1/n^*$ term, which implies the penalty term is $O(1/n^*)$. However, we express the penalized sum-of-squares this way following Yu and Ruppert (2002), in order to maintain the penalty as $n^* \rightarrow \infty$. This facilitates defining asymptotic penalized coefficients. Here,

$W(\mathbf{s}_i^*) = \begin{pmatrix} 1(\mathbf{s}_i^* \in \mathcal{S}_0^* \cup \mathcal{S}_1^*) & 0 \\ 0 & 1(\mathbf{s}_i^* \in \mathcal{S}_0^* \cup \mathcal{S}_2^*) \end{pmatrix}$. The effect of $W(\mathbf{s}_i^*)$ is to remove contributions to the penalized sum-of-squares when only one pollutant is observed at any given location. The penalty matrix Λ penalizes roughness of the exposure model. It may be zero everywhere in which case (1) reduces to a weighted least squares equation. The weighting matrix Π is fixed and may be diagonal. We emphasize that Π and Λ are general weighting and penalty matrices, respectively, with forms that may differ according to the specifications of the exposure model being fitted. We describe below the explicit form of Λ and Π when the exposure model is separate penalized regression splines.

By the weak law of large numbers, as $n^* \rightarrow \infty$, $\hat{\boldsymbol{\gamma}} \rightarrow_p \boldsymbol{\gamma}$ where

$$\boldsymbol{\gamma} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{j=0}^2 \pi_j^* \int (\boldsymbol{\Phi}(\mathbf{s}) - R(\mathbf{s})^T \boldsymbol{\theta})^T W(\mathbf{s}) \Pi W(\mathbf{s}) (\boldsymbol{\Phi}(\mathbf{s}) - R(\mathbf{s})^T \boldsymbol{\theta}) dH_j(\mathbf{s}) + \boldsymbol{\theta}^T \Lambda \boldsymbol{\theta}.$$

We note that this convergence is conditional on a given Λ , which in practice must first be estimated from the data. Having obtained penalized regression coefficients, we define predictions at subject locations as $\hat{\mathbf{w}}(\mathbf{s}_k) = R(\mathbf{s}_k)^T \hat{\boldsymbol{\gamma}}$. Analogously, let $\mathbf{w}(\mathbf{s}_k) = R(\mathbf{s}_k)^T \boldsymbol{\gamma}$ denote the predictions we would make if we had infinite monitoring data to fit the exposure model, and $\hat{\boldsymbol{\beta}}$ the estimate of $\boldsymbol{\beta}$ using $\hat{\mathbf{w}}(\mathbf{s}_k)$ and $\mathbf{z}_k$.

3.2.2. Separate penalized regression splines (SPRS)—When fitting SPRS, finding $\hat{\boldsymbol{\gamma}}$ reduces to minimizing the sum of two individual penalized sums-of-squares. In this context Π is a 2×2 identity matrix, and Λ is block diagonal with elements $\lambda_j D_j$, where the D_j are $(p_j + q_j) \times (p_j + q_j)$ square matrices with penalty parameters λ_j penalizing the coefficients of $\{\mathbf{p}_j(\cdot), \mathbf{q}_j(\cdot)\}$ (see Bergen and Szpiro (2015) for more details). Then $\hat{\boldsymbol{\gamma}}$ can be expressed as

$$\begin{aligned}\hat{\boldsymbol{\gamma}} = & \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \left\{ \left(\sum_{i: \mathbf{s}_i^* \in \mathcal{S}_0^* \cup \mathcal{S}_1^*} (x_{i1} - \mathbf{r}_1(\mathbf{s}_i^*)^T \boldsymbol{\theta})^2 + \lambda_1 \boldsymbol{\theta}^T D_1 \boldsymbol{\theta} \right) \right. \\ & \left. + \left(\sum_{i: \mathbf{s}_i^* \in \mathcal{S}_0^* \cup \mathcal{S}_2^*} (x_{i2} - \mathbf{r}_2(\mathbf{s}_i^*)^T \boldsymbol{\theta})^2 + \lambda_2 \boldsymbol{\theta}^T D_2 \boldsymbol{\theta} \right) \right\} \\ = & \underset{\boldsymbol{\theta}}{\operatorname{argmin}} (SS_1(\boldsymbol{\theta}) + SS_2(\boldsymbol{\theta})).\end{aligned}$$

Because of the form of the $\mathbf{r}_j$, minimizing this equation with respect to $\boldsymbol{\theta}$ is equivalent to separately minimizing $SS_1(\boldsymbol{\theta})$ with respect to the first $(p_1 + q_1)$ elements of $\boldsymbol{\theta}$ and $SS_2(\boldsymbol{\theta})$ with respect to the last $(p_2 + q_2)$ elements of $\boldsymbol{\theta}$. The resulting coefficient vector $\hat{\boldsymbol{\gamma}}^T$ is equivalent to a vector $\{\hat{\boldsymbol{\gamma}}_1^T, \hat{\boldsymbol{\gamma}}_2^T\}$ of separately-fitted coefficients, and the j^{th} element of $\hat{\mathbf{w}}(\mathbf{s}_k) = \mathbf{r}(\mathbf{s}_k)^T \hat{\boldsymbol{\gamma}}$ is equivalent to $\{\mathbf{p}_j(\mathbf{s}_k)^T, \mathbf{q}_j(\mathbf{s}_k)^T\}^T \hat{\boldsymbol{\gamma}}_j$.

Low-rank kriging (LRK) (Kammann and Wand, 2003) and thin-plate regression splines (TPRS) (Wood, 2003) are two examples of penalized regression models for modeling a single pollutant. Bergen and Szpiro (2015) show how each model yields definitions of the spatial bases and the penalty matrices D_j , and show an explicit connection between penalized regression and mixed effects models. This translation motivates selecting the λ_j via restricted maximum likelihood (REML). A simple strategy when fitting SPRS is to use REML to separately choose the penalty parameters of each exposure model.

4. Measurement error

4.1. Decomposing the measurement error

Once (1) has been used to obtain $\hat{\boldsymbol{\gamma}}$ and predictions $\hat{\mathbf{w}}(\mathbf{s}_j)$ obtained at subject locations we use them to estimate $\boldsymbol{\beta}$. Doing so induces measurement error which we decompose as follows:

$$\mathbf{x}_k - \hat{\mathbf{w}}(\mathbf{s}_k) = (\mathbf{x}_k - \mathbf{w}(\mathbf{s}_k)) + (\mathbf{w}(\mathbf{s}_k) - \hat{\mathbf{w}}(\mathbf{s}_k)) = \mathbf{u}^B(\mathbf{s}_k) + \mathbf{u}^C(\mathbf{s}_k).$$

Here $\mathbf{u}^B(\mathbf{s}_k)$ is multivariate Berkson-like error that arises from smoothing the exposure surfaces even when fitting the exposure model with infinite monitoring data. We term it “Berkson-like” as each element of $\mathbf{w}(\mathbf{s}_k)$ is a smoothed version of its respective exposure surface, resulting in predictions that are less variable than the true exposures. As we will show it can bias the vector of health effect estimates and impact its covariance matrix. In this way it differs from pure Berkson error which does not induce bias. The multivariate classical-like error, $\mathbf{u}^C(\mathbf{s}_k)$, is error that arises from having finite monitoring data with which to estimate $\boldsymbol{\gamma}$. It is similar to classical error in that it introduces variability into the predicted exposures that is independent of the health outcome. Accordingly it can induce bias as well as impact the covariance matrix of the estimated health effects. It differs from pure classical error since its effect goes away with n^* , as we discuss below.

As in Bergen and Szpiro (2015), the penalty matrix Λ regulates the impact of each error type. For SPRS, Λ is regulated by two penalty parameters λ_1 and λ_2 . As the λ_j increase there is more Berkson-like error as the predicted exposure surfaces are smoother. Simultaneously, classical-like error is reduced since $\hat{\gamma}$ is less variable. On the other hand, if both λ_j are zero classical-like error is at a maximum while Berkson-like error is mitigated since the exposure model is as flexible as possible. See Bergen and Szpiro (2015) for more details.

4.2. Bias expression

In this section, we use a Taylor expansion to derive an analytic bias expression that accounts for bias from both multi-pollutant classical- and Berkson-like measurement error.

Lemma 1—Let $\mathbf{r}_j^\perp(\mathbf{s})$ contain elements $(r_{jk}(\mathbf{s}) - \boldsymbol{\Theta}(\mathbf{s})^T \boldsymbol{\phi}_k)$, where r_{jk} is the k^{th} element of $\mathbf{r}_j$ and $\boldsymbol{\phi}_k = \arg\min_{\boldsymbol{\omega}} \int (r_{jk}(\mathbf{s}) - \boldsymbol{\Theta}(\mathbf{s})^T \boldsymbol{\omega})^2 dG(\mathbf{s})$ for $k \in \{1, \dots, r\}$. Let $R^\perp(\mathbf{s})$ denote the corresponding $r \times 2$ matrix created by binding the $\mathbf{r}_j^\perp(\mathbf{s})$, let $\mathbf{w}^\perp(\mathbf{s}) = R^\perp(\mathbf{s})^T \boldsymbol{\gamma}$ and $\hat{\mathbf{w}}^\perp(\mathbf{s}) = R^\perp(\mathbf{s})^T \hat{\boldsymbol{\gamma}}$. Let $M(\hat{\boldsymbol{\gamma}}) = \int \hat{\mathbf{w}}^\perp(\mathbf{s}) \hat{\mathbf{w}}^\perp(\mathbf{s})^T dG(\mathbf{s})$ and $U(\hat{\boldsymbol{\gamma}}) = \int \hat{\mathbf{w}}^\perp(\mathbf{s}) \mathbf{u}^B(\mathbf{s})^T dG(\mathbf{s})$. Then with

$$f(\hat{\boldsymbol{\gamma}}) = M(\hat{\boldsymbol{\gamma}})^{-1} \left(\int \hat{\mathbf{w}}^\perp(\mathbf{s}) \mathbf{w}^\perp(\mathbf{s})^T dG(\mathbf{s}) + U(\hat{\boldsymbol{\gamma}}) \right),$$

we can show that $\hat{\boldsymbol{\beta}} = f(\hat{\boldsymbol{\gamma}})^T \boldsymbol{\beta}$. Let Δ_{ij} denote the 2×2 matrix created by taking the derivative of $f(\hat{\boldsymbol{\gamma}})$ with respect to $\hat{\gamma}_i$ and Δ_{ij}^2 the 2×2 matrix created by taking the derivative of Δ_{ij} with respect to $\hat{\gamma}_j$. Let $\boldsymbol{\delta}_{kl}$ denote the $r \times 1$ vector where δ_{ikl} equals the $\{k, l\}^{th}$ element of Δ_{ij} and $\Delta_{(kl)}^2$ the $r \times r$ matrix where the $\{i, j\}^{th}$ element equals the $\{k, l\}^{th}$ element of Δ_{ij}^2 .

Then the bias attributable to Berkson-like and classical-like error is

$$E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = (\Psi^B + \Psi^C) \boldsymbol{\beta} \quad (2)$$

where $\Psi^B = M(\boldsymbol{\gamma})^{-1} U(\boldsymbol{\gamma})$ is bias from Berkson-like error, and

$$\Psi^C = \frac{1}{n^*} \left\{ \begin{pmatrix} \delta_{11}^T E(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) & \delta_{12}^T E(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \\ \delta_{21}^T E(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) & \delta_{22}^T E(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \end{pmatrix} + \begin{pmatrix} \text{tr}(\Delta_{(11)}^2 \text{Cov}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})) & \text{tr}(\Delta_{(12)}^2 \text{Cov}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})) \\ \text{tr}(\Delta_{(21)}^2 \text{Cov}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})) & \text{tr}(\Delta_{(22)}^2 \text{Cov}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})) \end{pmatrix} \right\}$$

is bias from classical-like error.

See Appendix A for a more rigorous statement and proof of Lemma 1. From here we can readily see that the bias from classical-like error vanishes as $n^* \rightarrow \infty$, since Ψ^C is of order $1/n^*$. Since Ψ^C also depends on $\text{Cov}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$, we see that Ψ^C can be lessened by increasing penalization of the regression coefficients. Conversely, note that Ψ^B induces bias from penalization, since $U(\boldsymbol{\gamma})$ is a zero-matrix only when the predictions are uncorrelated with the

Berkson-like error, which occurs if no penalty is used. Thus the bias from Berkson-like error increases as Λ increases, while the bias from classical-like error decreases.

Lemma 1 shows explicitly that the bias of $\hat{\beta}_j$ depends on the entire vector β . Accordingly, effect sizes of different magnitudes are more likely to lead to severe relative biases of the smaller effect than if the effect sizes are similar. This follows since $E(\hat{\beta}_j - \beta_j) = c_{1j}\beta_1 + c_{2j}\beta_2$, where c_{1j} and c_{2j} comprise the j^{th} row of $\Psi^B + \Psi^C$. If the c_{ij} (which depend only the exposure model) are of the same magnitude, the bias of $\hat{\beta}_j$ could be aggravated if β_j is of larger magnitude than β_j .

5. Methods

5.1. Bias correction

Lemma 1 implies $E(\hat{\beta}) = \beta + \Psi^B \beta + \Psi^C \beta$. Once we obtain estimates $\hat{\Psi}^B$ and $\hat{\Psi}^C$, it follows that the bias-corrected estimate of β is $\hat{\beta}^+ = (I_2 + \hat{\Psi}^B + \hat{\Psi}^C)^{-1} \hat{\beta}$. Appendix B details how to estimate Ψ^B and Ψ^C , and hence how to obtain an estimate of β that is corrected for bias from measurement error.

5.2. Standard error estimation

As described in Section 4.1, both $\mathbf{u}^B$ and $\mathbf{u}^C$ can affect the variance matrix of $\hat{\beta}$. Correcting for bias may have further impact on the standard errors of $\hat{\beta}^+$. We can account for all sources of variability using a non-parametric bootstrap, sampling n_j^* monitoring locations with replacement from the $\mathbf{s}_i^* \in \mathcal{S}_j^*$ and n subject locations with replacement from the $\mathbf{s}_k$. For B bootstrap samples we estimate the λ_j and fit the exposure model, predict at the bootstrapped subject locations, estimate a bias correction, and obtain uncorrected bootstrap estimates $\hat{\beta}^{Boot}$ and bias-corrected estimates $\hat{\beta}^{+,Boot}$. This accounts for all sources of variability: random selection of monitoring locations; subsequent observed exposures, predicted surfaces, and bias correction; and random selection of health locations. See Szpiro and Paciorek (2013) and Bergen and Szpiro (2015) for more details. We can then use the empirical standard deviations of the bootstrap estimates to form 95% confidence intervals that account for variability both from random sampling and from measurement error.

5.3. Model selection

When fitting SPRS one must choose both λ_1 and λ_2 . In single-pollutant studies Bergen and Szpiro (2015) motivated choosing the penalty in order to minimize the mean-squared error of the estimated health effect. It is not straightforward to extend this to the multi-pollutant context, as the exposure model can not be optimally chosen to mitigate relative bias. This follows since $E(\hat{\beta}_j - \beta_j) = c_{1j}\beta_1 + c_{2j}\beta_2$ as described in Section 4.2. Although one can trade-off Ψ^B and Ψ^C by choosing the λ_j , it is impossible to determine whether to minimize c_{1j} or c_{2j} since β is unknown. Bergen and Szpiro (2015) also found that formulating the penalized regression model as a mixed effects model and choosing the penalty parameter via REML performed better than minimizing the generalized cross-validation (GCV) criteria. Reiss and Ogden (2009) also found that using GCV could result in multiple local minima of the criteria. We can easily apply REML in the multi-pollutant case to select the λ_j individually.

An alternative to REML is to choose the λ_j to explicitly control the total effective degrees of freedom (EDF) in the exposure model. Equation 2 and our estimate of $Cov(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$ are based on asymptotics and are most valid when the EDF-to-monitor sample size ratio is not too large. We note that only the n_0^* locations in $\mathcal{S}_0^*$ contribute to estimating the off-block diagonal matrices corresponding to $E(\hat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1)(\hat{\boldsymbol{\gamma}}_2 - \boldsymbol{\gamma}_2)^T$ and that we have two correlated observations at each of these locations. This motivates controlling the ratio of EDF to $2n_0^*/(1 + \text{Cor}(x_{1i}^*, x_{2i}^*))$; we can heuristically think of $2n_0^*/(1 + \text{Cor}(x_{1i}^*, x_{2i}^*))$ as an “effective sample size” for estimating $E(\hat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1)(\hat{\boldsymbol{\gamma}}_2 - \boldsymbol{\gamma}_2)^T$. If $\text{Cor}(x_{1i}^*, x_{2i}^*) = 0$ the effective sample size is $2n_0^*$. If $\text{Cor}(x_{1i}^*, x_{2i}^*) = 1$ the effective sample size is n_0^* , implying we do not gain any extra information to estimate $E(\hat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1)(\hat{\boldsymbol{\gamma}}_2 - \boldsymbol{\gamma}_2)^T$ by having two observations at each $s_i^* \in \mathcal{S}_0^*$. Weighting each observation at each location by $1 + \text{Cor}(x_{1i}^*, x_{2i}^*)$ is motivated in Hanley et al. (2003) and corresponds to the weighting that minimizes the variance of a sample mean of correlated data. We refer to selecting the λ_j with REML and incrementally increasing them until the total EDF are 10% of the effective sample size as the “EDF” method of choosing the λ_j .

6. Simulations

6.1. Primary scenario

We performed simulations to investigate the impact of multi-pollutant measurement error and assess the efficacy of our bias correction under a number of different scenarios. We considered a primary simulation scenario in detail, and a number of sensitivity scenarios in less detail. In all scenarios the monitoring and subject locations were sampled uniformly from a 4500×4500 grid. For all simulation scenarios, the true exposure surface was

$$\Phi_j(\mathbf{s}) = \Phi_j^{NS}(\mathbf{s}) + c_{1j}\Phi_1^S(\mathbf{s}) + c_{2j}\Phi_2^S(\mathbf{s}). \quad (3)$$

The $\Phi_j^S(\mathbf{s})$ were the spatial components of the exposure surface, and were fixed realizations of a spatially correlated Gaussian process generated using the `spectralGP` package in R (Paciorek, 2007) with ranges 5000 and 500, respectively, and variances both equal to 1. The different ranges meant that $\Phi_1(\mathbf{s})$ was much smoother than $\Phi_2(\mathbf{s})$. The non-spatial part of the surface $\Phi_j^{NS}(\mathbf{s})$ was generated as $\mathbf{p}(\mathbf{s})^T \boldsymbol{\gamma}_{pj}$ where at each $\mathbf{s}$ the $\mathbf{p}(\mathbf{s})$ were three fixed realizations of $N(0, 1)$ random variables.

For our primary scenario, we modified the parameters of this exposure surface so as to create a 2-dimensional exposure surface wherein: (1) the two pollutants had the same overall variability; (2) the two surfaces were non-trivially correlated; (3) a large proportion of the overall variability for each pollutant necessitated modeling by spatial basis functions; and (4) one surface was smoother and thus easier to predict. Of these four characteristics, the second and third are shared by our application. We wanted (1) to make sure that any bias from measurement error was not unduly affected simply by a difference in pollutant magnitude. We wanted (4) to investigate the measurement error implications of having one surface more poorly predicted than the other, in the spirit of Zeger et al. (2000).

Accordingly, for our primary surface we set

$\gamma_{p,1}^T = \{0.578, 0.578, 0.578\}$; $\gamma_{p,2}^T = \{-0.528, -0.528, 0.660\}$; $\{c_{11}, c_{21}\} = \{3, 0\}$; and $\{c_{12}, c_{22}\} = \{2.3, 1.9\}$. Thus $\text{Var}(\Phi_j(\mathbf{s})) = 10$ and $\text{Cor}(\Phi_1(\mathbf{s}), \Phi_2(\mathbf{s})) = 0.70$ with $\approx 10\%$ of each surface's variability attributable to non-spatial structure and $\approx 90\%$ attributable to weighted average of shared spatially-structured surfaces. In each simulation we sampled 1000 subject locations and 200 monitoring locations from $\mathcal{S}_0^*$ uniformly across the grid. The exposures at monitoring locations were generated $x_{ij}^* = \Phi_j(\mathbf{s}_i^*) + \eta_{ij}^*$ with the $\eta_{i1}^*, \eta_{i2}^* \stackrel{iid}{\sim} N(0, \sqrt{3})$ while the unobserved exposures at subject locations were generated $x_{ij} = \Phi_j(\mathbf{s}_i) + \eta_{ij}$ with $\eta_{i1}, \eta_{i2} \stackrel{iid}{\sim} N(0, 1)$. Given true exposures the health outcomes were generated as

$$y_k = \beta_0 + \beta^T \mathbf{x}_k + \varepsilon_k, \quad (4)$$

with $\beta_0 = 0$ and $\varepsilon_k \sim N(0, 1)$. We considered $\beta^T = \{0.1, 0.5\}$ and $\beta^T = \{0.5, 0.1\}$, so as to be able to investigate measurement error implications when the effect sizes are of different magnitudes. We used LRK to separately model each pollutant, and considered exposure models with 4, 8, 12, or 16 spatial basis functions and no penalization as well as models with 10, 20, 30, 40, 60, 80, or 100 basis functions with penalization. The knots defining the LRK basis functions were chosen using a space-filling algorithm over the 4500×4500 grid (Kammann and Wand, 2003). For the penalized models we chose the λ_j using REML or the EDF criteria described in Section 5.3. Each model also included the $\mathbf{p}(\mathbf{s})$ as covariates to model the non-spatial part of the surface. In each simulation we assessed out-of-sample R^2 , correlation between the predictions, relative biases and standard deviations of $\hat{\beta}$ and $\hat{\beta}^+$, and actual coverage of 95% Wald confidence intervals using naïve sandwich standard errors or standard errors derived from 100 bootstrap samples.

6.2. Sensitivity scenarios

In addition to the primary scenario we considered three sensitivity scenarios that changed different aspects of the data generating mechanisms. Other than the described change all other aspects of each sensitivity scenario were the same as the primary scenario.

- a. We increased the non-spatial proportion of the surface variability to $\approx 50\%$ which decreased correlation between pollutants to 0.30, while keeping the variability of each pollutant still equal at 10. For this scenario, $\gamma_{p,1}^T = \{1.29, 1.29, 1.29\}$; $\gamma_{p,2}^T = \{-1.18, -1.18, 1.47\}$; $\{c_{11}, c_{21}\} = \{2.24, 0\}$; and $\{c_{12}, c_{22}\} = \{1.73, 1.41\}$.
- b. We induced negative correlation between the pollutants by setting $\{c_{12}, c_{22}\} = \{-2.3, -1.9\}$. For this scenario, $\text{Cor}(\Phi_1(\mathbf{s}), \Phi_2(\mathbf{s})) = -0.75$.
- c. Some monitors observed only one pollutant. In addition to $n_0^* = 200$ we separately sampled $n_1^* = 300$ and $n_2^* = 100$ monitors uniformly from the grid.

For these scenarios we only considered unpenalized LRK models with 12 basis functions or penalized LRK models with 30 basis functions. We assessed predictive accuracy, pollutant

correlation, bias, standard errors, and actual coverage of nominal 95% confidence intervals, as we did for the primary scenario.

6.3. Simulation results: primary scenario

Figures 1 and 2 show the simulation results when $\beta^T = \{0.5, 0.1\}$ and $\beta^T = \{0.1, 0.5\}$, respectively. For all exposure models mean out-of-sample R^2 s were between 0.70 and 0.85 for x_{11} while they were never higher than 0.70 and as low as 0.50 for x_{12} . The predictions were highly correlated; between 0.70 and 0.80 for all exposure models.

For the unpenalized and REML exposure models the biases were in opposite directions. Figure 1 shows slight downward bias of $\hat{\beta}_2$ led to upward relative biases near 40% for $\hat{\beta}_1$ due to β_2 being of higher magnitude than β_1 . The bias correction greatly reduced the bias and increased both estimates' standard errors, notably so for the highly parameterized unpenalized exposure models. Using naïve sandwich standard errors led to drastic under-coverage of 95% confidence intervals. Applying the bootstrap without the bias correction was also not enough to achieve accurate coverage. Fully accounting for measurement error by applying a bias correction and using the bootstrap to estimate standard errors yielded accurate 95% confidence interval coverage for both effect estimates. Using the EDF criteria to choose the λ_j led to strong upward biases for both effect estimates, which the bias correction greatly reduced. Fully accounting for measurement error led to accurate 95% confidence interval coverage of β_1 but not β_2 . This was not due to poor standard error estimation, as the mean of the bootstrap standard errors was identical to the actual standard error of $\hat{\beta}_2$. Density plots of the simulated z-scores for $\hat{\beta}_2$ revealed skewness in their distribution, leading to under-coverage of Wald confidence intervals. We also investigated confidence intervals based on percentiles of the bootstrap samples, but they did not perform any better than Wald intervals.

When $\beta_1 > \beta_2$, Figure 2 shows the unpenalized or REML exposure models led to upward bias of the poorer-measured pollutant's effect and downward bias of the better-measured pollutant's effect. Applying the bias correction eliminated this bias and when used in conjunction with the bootstrap led to accurate 95% confidence interval coverage. Using EDF to penalize the exposure models again led to drastic upward bias of both effect estimates. Applying the bias correction and using the bootstrap led to accurate 95% confidence interval coverage of both β_1 and β_2 .

Our results indicate that increasing predictive accuracy does not necessarily reduce bias. This is most clearly seen for the unpenalized and REML models in Figure 2, where increasing the number of exposure model basis functions improves out-of-sample R^2 but worsens the biases of both health effect estimates. Our results also advocate the use of REML to penalize the exposure model. The REML models yield the best efficiency and 95% confidence interval coverage even when using many basis functions to model exposure.

6.4. Simulation results: sensitivity scenarios

Results of the sensitivity scenarios described in Section 6.2 are shown in Tables 1–3. In general these results were similar to the primary scenario: strong biases in opposite

directions using unpenalized and REML exposure models; reduced bias after applying a bias correction; and strongly anti-conservative 95% confidence interval coverage if we do not apply both the bias correction and the bootstrap. We see the same tendency for the EDF criteria to lead to under-coverage even when applying both bias correction and bootstrap, most notably in Table 2.

However, there are some interesting differences between the primary and sensitivity results. When 50% of the exposure surface was explained by the non-spatial component Table 1 shows the correlation between the predictions was much lower than in the primary scenario. This appeared to mitigate the biases of the uncorrected health effect estimates to the extent that most of the exposure models achieved accurate 95% confidence coverage even without the bias correction. Table 2 shows that when the pollutants were negatively correlated the unpenalized and REML models yielded downward biases of both effect estimates while the EDF models yielded biases that were in opposite directions. Of all the REML modeling scenarios this one yielded the worst biases if no correction was used (as high as 61%) and poorest 95% confidence interval coverage (as low as 68%) if we did not fully account for measurement error. The results in Table 3 are similar to the primary results, and illustrate the effectiveness of our bias correction when both pollutants are not observed at all monitoring locations.

7. Measurement error correction in the Sister Study analysis

7.1. Exposure models

For our analysis we modeled annual 2006 average exposure as measured by 859 monitors that measured only PM_{2.5}, 180 monitors that measured only NO₂, and 178 monitors that measured both implying we had 1037 total PM_{2.5} observations and 358 total NO₂ observations. 155 of the PM_{2.5}-only monitors were of the Interagency Monitoring for Protected Visual Environments (IMPROVE) network, located mostly in rural areas. All other monitors belonged to the EPA's Air Quality System (AQS).

As a secondary analysis, we also analyzed the association of SBP with PM_{2.5} and NO₂ restricted to the nine northeastern states of Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New Jersey, New York and Pennsylvania. We wanted to see if the results of the national analysis held in this region, where spatial confounding was less of a concern due to reduced spatial structure in SBP. In these nine states we had 106 PM_{2.5}-only monitors, 20 NO₂-only monitors, and 39 monitors that measured both pollutants. Figure 3 shows the monitoring locations across the continental US and the northeastern regions.

We modeled each pollutant separately using national LRK models with 100 basis functions and penalty parameters chosen by REML. In the northeastern region, we modeled each pollutant with 15 basis functions. These levels of parameterization well-modeled the exposure surfaces without over-parameterizing. We selected spatial knots using a space-filling algorithm over a grid of 25km×25km cells, and LRK basis functions used a range parameter of 6363km for the national models and 1285km in the northeastern region. These corresponded to the maximum distance between any 25km×25km grid cell over the

respective region. Our exposure models also included 2 partial least squares (PLS) components to efficiently capture information from over 300 geographic covariates such as distances to road, population density and land-use variables (Abdi, 2003; Sampson et al., 2013; Bergen et al., 2013). We assessed prediction accuracy of these models via 10-fold cross-validation (Hastie et al., 2001).

7.2. Health models

Once we obtained predicted $PM_{2.5}$ and NO_2 exposures at subject residences, we used them to estimate health effects on SBP. We used the same 43,629 participants as Chan et al. (Provisionally accepted) in the primary national analysis and 7427 participants residing in the nine northeastern states for the sensitivity analysis. To adjust for large-scale spatial structure we used a 10-degree-of-freedom thin-plate spline, as (Chan et al., Provisionally accepted) did. Spatial confounding is less of a concern in the northeastern region due to greater spatial homogeneity of SBP so we did not include thin plate regression splines in the sensitivity analyses. The other adjustment variables listed in Section 2 were included in all models.

We modeled univariate and jointly additive associations of SBP with $PM_{2.5}$ and NO_2 . We corrected for bias from measurement error in all these models and compared 95% confidence intervals from naïve sandwich standard errors to 95% confidence intervals derived using median absolute deviation of bootstrap samples. We used median absolute deviation instead of standard deviation to estimate standard errors as the standard deviation was highly sensitive to outliers in the bootstrap samples.

We also investigated the interactive association of SBP with $PM_{2.5}$ and NO_2 . We fit this interaction to investigate the counterintuitive direction of the NO_2 effect (discussed below), rather than for inferential purposes. Thus we did not perform measurement error correction on the interaction term.

7.3. Sensitivity analyses

As sensitivity analyses, we also considered national exposure models with 50 and 150 basis functions, and northeastern exposure models with 10 and 20 basis functions. We also considered health models with 5 and 15 basis functions to adjust for spatial confounding on the national scale. The results of these analyses are described in Appendix C.

7.4. Results

10-fold cross-validated R^2 was 0.77 for $PM_{2.5}$ and 0.78 for NO_2 on the national scale; and 0.83 for $PM_{2.5}$ and 0.89 for NO_2 in the northeastern region. The predictions were moderately correlated on the national scale (0.42) and more correlated in the northeastern region (0.72). Correlations between our LRK $PM_{2.5}$ predictions and those used by Chan et al. (Provisionally accepted) were between 0.94 and 0.96, while correlations between our LRK NO_2 predictions and those used by Chan et al. (Provisionally accepted) were between 0.86 and 0.88.

Estimated associations of SBP with pollution exposure are shown in Figure 4. We focus first on the national results. When modeled separately, there was a slightly significant association of SBP with $PM_{2.5}$, and no significant association of SBP with NO_2 . There was no meaningful estimated bias from measurement error and the bootstrap standard error estimates were very similar to the naïve sandwich standard error estimates.

We saw very different results when the pollutants were modeled jointly. There was a much stronger association of SBP with $PM_{2.5}$ when controlling for NO_2 , and a strong and significant negative association between SBP and NO_2 controlling for $PM_{2.5}$. All associations were much stronger than those described by Chan et al. (Provisionally accepted). Unlike the univariate analyses, there was notable estimated downward bias from measurement error in both effect estimates and bias-corrected estimates were stronger than their uncorrected counterparts. 95% confidence intervals that accounted for all sources of variability, including the bias correction, were wider than confidence intervals derived from naïve sandwich standard error estimates or bootstrap standard error estimates without the bias correction.

The results from the northeastern analysis were very similar to the national analysis. The univariate estimated associations between SBP and $PM_{2.5}$ were positive and significant, while as in the national models the estimated associations between SBP and NO_2 were essentially null. As in the national models accounting for measurement error did not qualitatively alter the inference for the univariate models. The joint associations were qualitatively very similar to the national models, though the estimated biases from measurement error were not as large and were slightly away from the null.

Figure 5 shows the national interactive association of SBP with $PM_{2.5}$ and NO_2 . Although the p-value for the interaction was 0.01, Figure 5 suggests the association between SBP and pollution exposure does not drastically differ from additive. This is most easily explored by traveling horizontally across Figure 5 and noting that where most of the data are the horizontal distance between the contours does not drastically change depending on where one is with respect to the NO_2 axis. This implies the necessary change in $PM_{2.5}$ for a one-unit change in the SBP partial residuals is roughly the same regardless of NO_2 . To confirm this we estimated the association of SBP with $PM_{2.5}$ at the quartiles of NO_2 , and vice versa, and found no qualitative difference between associations that allow for interaction and those estimated using main effects.

8. Discussion

We have developed a holistic framework for multi-pollutant analyses in air pollution epidemiology. Measurement error from using predictions derived from misaligned monitoring locations is often an inevitability in these studies and needs to be accounted for. Penalized regression splines offer a viable way highly parameterizing the exposure models in order to accurately model each surface while ensuring regularity through penalization. Our methodology provides a way of characterizing and correcting for measurement error when using these flexible, commonly-used exposure models. Although we have focused on

the simple strategy of modeling each pollutant separately, our paradigm seamlessly admits more sophisticated models as long as they follow the general form specified by Equation 1.

As we saw in our simulations and data analysis, multi-pollutant measurement error can induce severe biases even when one or both pollutants are well-measured. Our simulations showed the directions and magnitudes of these biases are unpredictable and depend on the degree of model penalization, signs of the true health effects, pollutant correlation and relative sizes of the health effects. Not applying the bias correction can lead to undercoverage of 95% confidence intervals even if the bootstrap is employed, while if REML is used to choose the penalty parameters we achieve accurate coverage if we combine the bias correction and bootstrap.

The measurement error methodology we developed was essential in the Sister Study data analysis. We saw clear evidence of NO_2 confounding the association between systolic blood pressure and $\text{PM}_{2.5}$, and vice versa. To accurately estimate these associations we needed to adjust for both pollutants, which in turn induced multi-pollutant measurement error. Our simulations demonstrated this error can severely bias the health effect estimates, and we saw that in the joint health analysis we estimated much larger biases than those estimated in the univariate analyses, resulting in corrected point estimates that were stronger than the naïve estimates. Although 95% confidence intervals that accounted for bias correction were notably wider resulting in p-values that were similar to the naïve analysis, they are more likely to cover the true health effects as our simulations demonstrated.

The large biases and corresponding need for bias correction in our simulations and data analysis is rather unique to the multi-pollutant context. Bergen and Szpiro (2015) saw only slight undercoverage without bias correction in their univariate simulations to accompany relative biases that never exceeded 15%. The out-of-sample R^2 from their models needed to elicit these biases were between 0.40 and 0.50, much lower than our simulations. Parametric univariate methods also tend to yield negligible biases (Szpiro et al., 2011; Bergen et al., 2013). In the multi-pollutant setting we can not rely on prediction accuracy to achieve negligible bias. Our simulations showed that not only does drastic bias exist even when exposure models well-predict exposure, but improving prediction accuracy by increasing the number of exposure model basis functions can actually improve out-of-sample R^2 while increasing bias.

An unexpected result of our data analysis is the negative association of SBP with NO_2 . We performed a series of sensitivity analyses (not shown) to test the robustness of our results. First we restricted the monitoring data used to model exposure to the 178 locations that measured both $\text{PM}_{2.5}$ and NO_2 , in order to see if differences between monitoring data distributions was responsible for our results. Second we adjusted for time of blood pressure exam as a potential confounder, as time of exam is spatially structured and blood pressure exhibits seasonal patterns. Third we modeled NO_X instead of NO_2 . All three analyses yielded qualitatively identical results to those in Section 7.4. One possibility is that there is still unmeasured confounding responsible for our results. NO_2 is typically considered a traffic-related pollutant especially prominent in the near-road environment, but nearly all the NO_2 monitors were sited far from roadways by EPA direction. Progress is being made

toward locating more monitors closer to roads (EPA, 2015), but in the meantime traffic-sourced NO₂ may represent a “third” surface that, when omitted from the health model, induces residual confounding in our observed ambient NO₂ health effect.

Throughout this paper, we have assumed a linear health effect of exposure. Most air pollution epidemiology studies to date have utilized linear or generalized linear modeling frameworks. Violation of linear modeling assumptions may have important implications, especially for understanding the impacts of air pollution at relatively low doses, and this is a subject of active research (Burnett et al., 2014). While estimating nonlinear dose responses is beyond the scope of this paper, our multi-pollutant methods lay the groundwork for studying nonlinear dose responses using fixed-rank approximations such as regression splines. One possibility is to model each spline basis function as a separate “pollutant” and adapt the methods in this paper to estimate a measurement error corrected nonlinear dose response. In our application, we considered modeling SBP with quadratic PM_{2.5} and NO₂ health model terms (without applying measurement error correction) as a sensitivity analysis. There was no statistical evidence that SBP had a quadratic association with either of these pollutants.

We acknowledge two limitations of our methodology. First, we note that throughout we have assumed $H_j(\cdot) = G(\cdot)$ for $j = 0, 1, 2$. In reality the $H_j(\cdot)$ are unlikely to be equal, and furthermore each $H_j(\cdot)$ likely differs from the distribution of subject locations. If the monitoring and subject locations are drawn from different distributions, our bias correction may worsen or fail to completely eliminate bias. In Appendix D we present results from a simulation study when this assumption is violated, and find that this is indeed the case. It is possible to relax this assumption to follow Condition 1 from Szpiro and Paciorek (2013), namely that the $R(\mathbf{s})$ have the same distribution regardless of whether $\mathbf{s}$ is sampled from $H(\cdot)$ or $G(\cdot)$, although to simplify the exposition we have not undertaken this in the present paper. An additional implication is that we cannot estimate $U(\boldsymbol{\gamma})$ by summing over the empirical distributions of monitoring locations (as described in Appendix B), as they do not approximate $G(\cdot)$. Instead we must re-weight observations in $\mathcal{S}_j^*$ by an estimate of $g(\cdot)/h_j(\cdot)$ which we could do via e.g. propensity score methods. Note that this requires the ratio $g(\cdot)/h_j(\cdot)$ to be everywhere defined; hence we can weaken the assumption of identically distributed subject and monitoring locations by assuming that the support of $H_j(\cdot)$ contains the support of $G(\cdot)$ for $j = 0, 1, 2$. Second, a subtle assumption necessary for estimation of the bias from Berkson-like error, Ψ^B , is that $\boldsymbol{\Theta}(\cdot)$ (the spatially structured components of the subject-specific covariates included in the health model) be defined at monitoring locations. In practice, this assumption is likely to be violated. In Appendix B we discuss this assumption in further detail, its implications, and ideas for accounting for its violation. Future work will extend our methodology to scenarios where these assumptions are violated and will develop suitable methodology to account for such violations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the many participants and study staff comprising the Sister Study who contributed to making this study possible. This research was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01-ES044005). Additional support by the NIEHS was provided through R01-ES009411, P50-ES015915, R01-ES020871, K24-ES013195, P30-ES07033 and T32-ES015459. Although the research described in this presentation has been funded wholly or in part by the US Environmental Protection Agency through R831697 and RD-83479601 to the University of Washington, it has not been subjected to the agency's required peer and policy review and therefore does not necessarily reflect the views of the agency, and no official endorsement should be inferred.

References

- Abdi H. Partial least square regression (PLS regression). Encyclopedia for research methods for the social sciences. 2003;792–795.
- Bergen S, Sheppard L, Sampson P, Kim S, Richards M, Vedal S, Kaufman J, Szpiro A. A national prediction model for PM_{2.5} component exposures and measurement error-corrected health effect inference. Environmental Health Perspectives. 2013; 121:1017–1025. [PubMed: 23757600]
- Bergen S, Szpiro A. Mitigating the impact of measurement error when using penalized regression to model exposure in two-stage air pollution epidemiology studies. Environmental and Ecological Statistics. 2015; 22:601–631.
- Billionnet C, Sherrill D, Annesi-Maesano I. Estimating the health effects of exposure to multi-pollutant mixture. Annals of Epidemiology. 2012; 22:126–141. [PubMed: 22226033]
- Burnett R, Pope C, Ezzati M, Olives C, Lim S, Mehta S, Shin H, Singh G, Hubbell B, Brauer M, et al. An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. Environmental Health Perspectives. 2014; 122:397–403. [PubMed: 24518036]
- Carroll, R. Measurement error in nonlinear models: a modern perspective. CRC Press; 2006.
- Chan S, Van Hee V, Bergen S, Szpiro A, Oron A, DeRoo L, London S, Marshall J, Kaufman J, Sandler D. Long term air pollution exposure and blood pressure in the Sister Study. Environmental Health Perspectives. (Provisionally accepted).
- Cohen M, Adar S, Allen R, Avol E, Curl C, Gould T, Hardie D, Ho A, Kinney P, Larson T, et al. Approach to estimating participant pollutant exposures in the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). Environmental Science & Technology. 2009; 43:4687–4693. [PubMed: 19673252]
- Diez Roux A, Merkin S, Arnett D, Chambless L, Massing M, Nieto F, Sorlie P, Szklo M, Tyroler H, Watson R. Neighborhood of residence and incidence of coronary heart disease. New England Journal of Medicine. 2001; 345:99–106. [PubMed: 11450679]
- Dominici F, Peng R, Barr C, Bell M. Protecting human health from air pollution: Shifting from a single-pollutant to a multi-pollutant approach. Epidemiology. 2010; 21:187–194. [PubMed: 20160561]
- EPA. Integrated science assessment for oxides of nitrogen health criteria (second external review draft). Washington, DC: U.S. Environmental Protection Agency; 2015. EPA/600/R-14/006
- Gryparis A, Paciorek C, Zeka A, Schwartz J, Coull B. Measurement error caused by spatial misalignment in environmental epidemiology. Biostatistics. 2009; 10:258–274. [PubMed: 18927119]
- Hanley J, Negassa A, Forrester J. Statistical analysis of correlated data using generalized estimating equations: an orientation. American Journal of Epidemiology. 2003; 157:364–375. [PubMed: 12578807]
- Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning. Springer Series in Statistics; 2001.
- Kamman E, Wand M. Geoadditive models. Journal of the Royal Statistical Society: Series C (Applied Statistics). 2003; 52:1–18.
- Liang K, Zeger S. Longitudinal data analysis using generalized linear models. Biometrika. 1986; 73:13–22.

- Lopiano K, Young L, Gotway C. Estimated generalized least squares in spatially misaligned regression models with berkson error. *Biostatistics*. 2013
- Lopiano K, Young L, Gotway C. A pseudo-penalized quasi-likelihood approach to the spatial misalignment problem with non-normal data. *Biometrics*. 2014; 70:648–660. [PubMed: 24749487]
- Madsen L, Ruppert D, Altman N. Regression with spatially misaligned data. *Environmetrics*. 2008; 19:453–467.
- NIEHS. The Sister Study. 2013 <http://www.sisterstudy.org/>.
- Novotny E, Bechle M, Millet D, Marshall J. National satellite-based land-use regression: No2 in the united states. *Environmental Science & Technology*. 2011; 45:4407–4414. [PubMed: 21520942]
- Paciorek C. Bayesian smoothing with Gaussian processes using Fourier basis functions in the spectralGP library. *Journal of Statistical Software*. 2007; 19
- Reiss P, Ogden T. Smoothing parameter selection for a class of semiparametric linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2009; 71:505–523.
- Ruppert, D.; Wand, M.; Carroll, R. *Semiparametric regression*. Cambridge University Press; 2003.
- Sampson P, Richards M, Szpiro A, Bergen S, Sheppard L, Larson T, et al. A regionalized national universal kriging model using partial least squares regression for estimating annual PM_{2.5} concentrations in epidemiology. *Atmospheric Environment*. 2013; 75:383–392. [PubMed: 24015108]
- Schwartz J, Coull B. Control for confounding in the presence of measurement error in hierarchical models. *Biostatistics*. 2003; 4:539–553. [PubMed: 14557110]
- Shao, J. *Mathematical Statistics*, 2nd Edition. Springer; 2003.
- Strand M, Sillau S, Grunwald G, Rabinovitch N. Regression calibration for models with two predictor variables measured with error and their interaction, using instrumental variables and longitudinal data. *Statistics in Medicine*. 2014; 33:470–487. [PubMed: 23901041]
- Szpiro A, Paciorek C. Measurement error in two-stage analyses, with application to air pollution epidemiology. *Environmetrics*. 2013; 24:501–517. [PubMed: 24764691]
- Szpiro A, Sheppard L, Lumley T. Efficient measurement error correction with spatially misaligned data. *Biostatistics*. 2011; 12:610–623. [PubMed: 21252080]
- Vedal S, Kaufman J. What does multi-pollutant air pollution research mean? *American Journal of Respiratory and Critical Care Medicine*. 2011; 183:4–6. [PubMed: 21193783]
- Wakefield, J. *Bayesian and Frequentist Regression Methods*. Springer Series in Statistics; 2013.
- Wood S. Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2003; 65:95–114.
- Yu Y, Ruppert D. Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association*. 2002; 97:1042–1054.
- Zeger S, Thomas D, Dominici F, Samet J, Schwartz J, Dockery D, et al. Exposure measurement error in time-series studies of air pollution: concepts and consequences. *Environ Health Perspect*. 2000; 108:419–426. [PubMed: 10811568]
- Zeka A, Schwartz J. Estimating the independent effects of multiple pollutants in the presence of measurement error: An application of a measurement-error-resistant technique. *Environmental Health Perspectives*. 2004; 112:1686–1690. [PubMed: 15579414]

Figure 1.

Primary simulation results, with $\beta_1 = 0.1$ and $\beta_2 = 0.5$. Relative biases and standard deviations are shown both with and without a bias correction. Also shown are actual coverages of 95% confidence intervals both with and without a bias correction, and with naïve or non-parametric bootstrap standard errors. The bottom row shows mean out-of-sample R^2 and prediction correlation.

Figure 2.

Primary simulation results, with $\beta_1 = 0.5$ and $\beta_2 = 0.1$. Relative biases and standard deviations are shown both with and without a bias correction. Also shown are actual coverages of 95% confidence intervals both with and without a bias correction, and with naïve or non-parametric bootstrap standard errors. The bottom row shows mean out-of-sample R^2 and prediction correlation.

Figure 3.

Map of monitoring and Sister Study participant locations, both for the entire nation and the nine upper northeastern states used in the sensitivity analysis.

Figure 4.

Estimated difference in SBP (mmHg) for a 10- $\mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ and for a 10-ppb increase in NO_2 . “Separate” refers to modeling the association of SBP with an individual pollutant, while “joint” refers to modeling SBP jointly with $\text{PM}_{2.5}$ and NO_2 . Both bias-corrected and uncorrected estimates are shown along with naïve sandwich and bootstrap standard errors.

Figure 5. SBP partial residuals (contours) as a smooth function of $PM_{2.5}$ and NO_2 . Predicted exposures shown were using 100 degrees of freedom in the exposure model.

Table 1

Sensitivity scenario; proportion of explainable variability due to geographic covariates equals 50%, implying $Cor(\Phi_1(s), \Phi_2(s)) = 0.30$.

	$\beta_1 = 0.1$					$\beta_2 = 0.5$				
	RB	SD	SE	Cov		RB	SD	SE	Cov	
No penalty; $R^2(x_1) = 0.81$; $R^2(x_2) = 0.70$; $Cor(w_1, w_2) = 0.27$										
No correction	0.02	0.30	0.15	0.66		-0.04	0.07	0.03	0.55	
Bias correction only	0.01	0.33	0.15	0.63		0.01	0.07	0.03	0.59	
Bootstrap SE only	0.02	0.30	0.32	0.96		-0.04	0.07	0.07	0.88	
Bias correction + bootstrap	0.01	0.33	0.37	0.96		0.01	0.07	0.08	0.94	
REML; $R^2(x_1) = 0.84$; $R^2(x_2) = 0.73$; $Cor(w_1, w_2) = 0.25$										
No correction	0.11	0.30	0.14	0.64		0.00	0.07	0.03	0.61	
Bias correction only	0.01	0.30	0.14	0.68		-0.01	0.07	0.03	0.60	
Bootstrap SE only	0.11	0.30	0.30	0.93		0.00	0.07	0.07	0.93	
Bias correction + bootstrap	0.01	0.30	0.31	0.94		-0.01	0.07	0.07	0.93	
EDF; $R^2(x_1) = 0.83$; $R^2(x_2) = 0.72$; $Cor(w_1, w_2) = 0.21$										
No correction	0.20	0.31	0.15	0.57		0.03	0.07	0.03	0.58	
Bias correction only	0.01	0.30	0.15	0.67		-0.01	0.07	0.03	0.62	
Bootstrap SE only	0.20	0.31	0.32	0.9		0.03	0.07	0.07	0.92	
Bias correction + bootstrap	0.01	0.30	0.31	0.94		-0.01	0.07	0.06	0.92	
	$\beta_1 = 0.5$					$\beta_2 = 0.1$				
	RB	SD	SE	Cov		RB	SD	SE	Cov	
No penalty; $R^2(x_1) = 0.81$; $R^2(x_2) = 0.70$; $Cor(w_1, w_2) = 0.27$										
No correction	-0.03	0.06	0.03	0.59		-0.02	0.30	0.14	0.64	
Bias correction only	0.00	0.06	0.03	0.62		0.01	0.32	0.14	0.61	
Bootstrap SE only	-0.03	0.06	0.06	0.92		-0.02	0.30	0.30	0.94	
Bias correction + bootstrap	0.00	0.06	0.07	0.96		0.01	0.32	0.36	0.95	
REML; $R^2(x_1) = 0.84$; $R^2(x_2) = 0.73$; $Cor(w_1, w_2) = 0.25$										
No correction	0.00	0.05	0.03	0.64		0.10	0.29	0.14	0.64	
Bias correction only	-0.01	0.05	0.03	0.64		0.01	0.29	0.14	0.65	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Bootstrap SE only	0.00	0.05	0.05	0.95	0.10	0.29	0.28	0.92
Bias correction + bootstrap	-0.01	0.05	0.06	0.95	0.01	0.29	0.28	0.94
EDF; $R^2(x_1) = 0.83$; $R^2(x_2) = 0.72$; $Cor(w_1, w_2) = 0.21$								
No correction	0.03	0.06	0.03	0.61	0.22	0.31	0.14	0.54
Bias correction only	-0.01	0.06	0.03	0.64	0.01	0.30	0.14	0.66
Bootstrap SE only	0.03	0.06	0.06	0.92	0.22	0.31	0.30	0.89
Bias correction + bootstrap	-0.01	0.06	0.05	0.95	0.01	0.30	0.29	0.94

“RB” denotes relative bias; “SD” denotes empirical relative standard deviations; “SE” denotes relative mean estimated standard errors; and “Cov” denotes actual coverage of nominal 95% Wald confidence intervals. For each penalization method, the mean out-of-sample R^2 s for each pollutant and mean correlation between predictions is given.

Table 2

Sensitivity scenario; $\text{Corr}(\Phi_1(\mathbf{s}), \Phi_2(\mathbf{s})) = -0.75$.

	$\beta_1 = 0.1$				$\beta_2 = 0.5$			
	RB	SD	SE	Cov	RB	SD	SE	Cov
No penalty; $R^2(x_1) = 0.75$; $R^2(x_2) = 0.57$; $\text{Corr}(w_1, w_2) = -0.81$								
No correction	-0.67	0.54	0.27	0.37	-0.17	0.12	0.06	0.28
Bias correction only	0.20	1.07	0.27	0.47	0.05	0.24	0.06	0.47
Bootstrap SE only	-0.67	0.54	0.53	0.71	-0.17	0.12	0.11	0.62
Bias correction + bootstrap	0.20	1.07	0.91	0.94	0.05	0.24	2.92	0.94
REML; $R^2(x_1) = 0.81$; $R^2(x_2) = 0.62$; $\text{Corr}(w_1, w_2) = -0.81$								
No correction	-0.61	0.51	0.26	0.38	-0.12	0.11	0.06	0.42
Bias correction only	0.01	0.81	0.26	0.5	0.01	0.18	0.06	0.49
Bootstrap SE only	-0.61	0.51	0.47	0.68	-0.12	0.11	0.10	0.73
Bias correction + bootstrap	0.01	0.81	0.67	0.9	0.01	0.18	0.15	0.9
EDF; $R^2(x_1) = 0.74$; $R^2(x_2) = 0.54$; $\text{Corr}(w_1, w_2) = -0.81$								
No correction	-0.24	0.8	0.33	0.55	0.10	0.19	0.08	0.58
Bias correction only	-0.14	0.73	0.33	0.61	-0.03	0.17	0.08	0.62
Bootstrap SE only	-0.24	0.8	0.74	0.89	0.10	0.19	0.17	0.93
Bias correction + bootstrap	-0.14	0.73	0.65	0.9	-0.03	0.17	0.15	0.89
	$\beta_1 = 0.5$				$\beta_2 = 0.1$			
	RB	SD	SE	Cov	RB	SD	SE	Cov
No penalty; $R^2(x_1) = 0.75$; $R^2(x_2) = 0.57$; $\text{Corr}(w_1, w_2) = -0.81$								
No correction	-0.11	0.10	0.05	0.41	-0.48	0.49	0.25	0.48
Bias correction only	0.03	0.17	0.05	0.51	0.14	0.87	0.25	0.49
Bootstrap SE only	-0.11	0.10	0.10	0.75	-0.48	0.49	0.51	0.8
Bias correction + bootstrap	0.03	0.17	0.32	0.94	0.14	0.87	2.01	0.95
REML; $R^2(x_1) = 0.81$; $R^2(x_2) = 0.62$; $\text{Corr}(w_1, w_2) = -0.81$								
No correction	-0.09	0.08	0.04	0.43	-0.54	0.44	0.24	0.42
Bias correction only	-0.01	0.12	0.04	0.54	-0.03	0.65	0.24	0.53

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Bootstrap SE only	−0.09	0.08	0.08	0.75	−0.54	0.44	0.42	0.69
Bias correction + bootstrap	−0.01	0.12	0.11	0.92	−0.03	0.65	0.58	0.92
EDF; $R^2(x_1) = 0.74$; $R^2(x_2) = 0.54$; $Cor(w_1, w_2) = -0.81$								
No correction	0.04	0.12	0.06	0.62	−0.59	0.68	0.32	0.45
Bias correction only	−0.03	0.11	0.06	0.64	−0.16	0.62	0.32	0.65
Bootstrap SE only	0.04	0.12	0.12	0.95	−0.59	0.68	0.67	0.81
Bias correction + bootstrap	−0.03	0.11	0.11	0.90	−0.16	0.62	0.59	0.91

“RB” denotes relative bias; “SD” denotes empirical relative standard deviations; “SE” denotes relative mean estimated standard errors; and “Cov” denotes actual coverage of nominal 95% Wald confidence intervals. For each penalization method, the mean out-of-sample R^2 s for each pollutant and mean correlation between predictions is given.

Table 3

Sensitivity scenario; $n_C^*=200; n_1^*=300; n_2^*=100$.

	$\beta_1 = 0.1$				$\beta_2 = 0.5$			
	RB	SD	SE	Cov	RB	SD	SE	Cov
No penalty; $R^2(x_1) = 0.77; R^2(x_2) = 0.58; Cor(w_1, w_2) = 0.77$								
No correction	0.28	0.45	0.26	0.64	-0.08	0.10	0.05	0.53
Bias correction only	-0.10	0.72	0.26	0.52	0.02	0.15	0.05	0.55
Bootstrap SE only	0.28	0.45	0.46	0.89	-0.08	0.10	0.10	0.82
Bias correction + bootstrap	-0.10	0.72	3.07	0.96	0.02	0.15	0.79	0.96
REML; $R^2(x_1) = 0.84; R^2(x_2) = 0.64; Cor(w_1, w_2) = 0.76$								
No correction	0.29	0.42	0.24	0.62	-0.05	0.09	0.05	0.63
Bias correction only	-0.05	0.64	0.24	0.51	0.01	0.13	0.05	0.57
Bootstrap SE only	0.29	0.42	0.40	0.85	-0.05	0.09	0.09	0.89
Bias correction + bootstrap	-0.05	0.64	0.58	0.93	0.01	0.13	0.12	0.94
EDF; $R^2(x_1) = 0.77; R^2(x_2) = 0.54; Cor(w_1, w_2) = 0.74$								
No correction	0.57	0.59	0.29	0.44	0.15	0.15	0.07	0.48
Bias correction only	0.01	0.73	0.29	0.56	-0.01	0.15	0.07	0.64
Bootstrap SE only	0.57	0.59	0.58	0.78	0.15	0.15	0.15	0.87
Bias correction + bootstrap	0.01	0.73	0.93	0.94	-0.01	0.15	0.20	0.94
	$\beta_1 = 0.5$				$\beta_2 = 0.1$			
	RB	SD	SE	Cov	RB	SD	SE	Cov
No penalty; $R^2(x_1) = 0.77; R^2(x_2) = 0.58; Cor(w_1, w_2) = 0.77$								
No correction	-0.02	0.07	0.05	0.79	0.01	0.37	0.24	0.78
Bias correction only	0.01	0.12	0.05	0.59	-0.08	0.74	0.24	0.50
Bootstrap SE only	-0.02	0.07	0.07	0.94	0.01	0.37	0.38	0.95
Bias correction + bootstrap	0.01	0.12	0.18	0.96	-0.08	0.74	1.15	0.96
REML; $R^2(x_1) = 0.84; R^2(x_2) = 0.64; Cor(w_1, w_2) = 0.76$								
No correction	-0.02	0.06	0.04	0.77	0.12	0.33	0.22	0.77
Bias correction only	0.00	0.10	0.04	0.61	-0.04	0.59	0.22	0.55

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Bootstrap SE only	−0.02	0.06	0.06	0.95	0.12	0.33	0.33	0.93
Bias correction + bootstrap	0.00	0.10	0.09	0.93	−0.04	0.59	0.53	0.93
EDF; $R^2(x_1) = 0.77$; $R^2(x_2) = 0.54$; $Cor(w_1, w_2) = 0.74$								
No correction	0.10	0.08	0.05	0.45	0.54	0.51	0.29	0.51
Bias correction only	0.00	0.12	0.05	0.65	−0.06	0.73	0.29	0.62
Bootstrap SE only	0.10	0.08	0.09	0.79	0.54	0.51	0.52	0.82
Bias correction + bootstrap	0.00	0.12	0.15	0.96	−0.06	0.73	0.89	0.96

“RB” denotes relative bias; “SD” denotes empirical relative standard deviations; “SE” denotes relative mean estimated standard errors; and “Cov” denotes actual coverage of nominal 95% Wald confidence intervals. For each penalization method, the mean out-of-sample R^2 s for each pollutant and mean correlation between predictions is given.