Spatial measurement error methods in air pollution epidemiology

Silas Bergen

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Adam Szpiro, Chair

Lianne Sheppard

Ali Shojaie

Program Authorized to Offer Degree:
School of Public Health–Biostatistics

University of Washington

Abstract

Spatial measurement error methods in
air pollution epidemiology

Silas Bergen

Chair of the Supervisory Committee:
Adam Szpiro
Biostatistics

Air pollution epidemiology cohort studies often implement a two-stage approach to estimating associations of continuous health outcomes with one or more exposures. An inherent problem in these studies is that the exposures of interest are usually unobserved. Instead observations are available at misaligned monitoring locations. The first stage entails building exposure models with the monitoring data and predicting at subject locations; the second stage uses the predictions to estimate health effects. This induces measurement error that can induce bias and affect the standard error of resulting estimates. Berkson-like error arises from smoothing the exposure surface, while classical-like error comes from estimating the exposure model parameters. Accurately characterizing and correcting for both types of measurement error depends on assumptions made about the spatial surface and exposure model used to derive predictions. This dissertation addresses spatial measurement error in air pollution epidemiology. We first describe and apply parametric measurement error methodology when assuming the exposure surface is a stochastic Gaussian process. We extend these parametric approaches by deriving P-SIMEX, which yields more flexible bias correction. We then motivate a semi-parametric framework wherein the exposure surface is viewed as fixed and modeled with penalized regression splines. We discuss the resulting measurement error, describe how the exposure model penalty regulates measurement error, and derive an analytic bias correction. Finally we extend the semi-parametric methodology

to the multi-pollutant setting. We show the direction of the biases are unpredictable, and the magnitude of the biases are much larger than those in single-pollutant studies. We derive a multi-pollutant bias correction that can be combined with a simple non-parametric bootstrap to achieve accurate 95% confidence interval coverage. Throughout we apply our methods to analyzing associations of continuous health outcomes with predicted exposures in the Multi-Ethnic Study of Atherosclerosis and the Sister Study of the National Institute of Environmental Health Sciences.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The research described by this dissertation 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, 5P50ES015915 and 5R01ES020871, and T32 ES015459. Additional support was provided by an award to the University of Washington under the National Particle Component Toxicity (NPACT) initiative of the Health Effects Institute (HEI) and by the Environmental Protection Agency, Assistance Agreement RD-83479601-0 (Clean Air Research Centers).

I would like to thank Adam Szpiro for his invaluable patience and mentorship throughout this research process, for making the mountain feel more climbable and creating a safe space to ask dumb questions while helping me to ask better ones. Thank you to Lianne Sheppard for inviting me aboard the Biostatistics, Epidemiologic and Bioinformatic Training in Environmental Health (BEBTEH) Training Grant and taking a strong interest in my professional training. Thank you to Adam, Lianne, Ali Shojaie, Jon Wakefield and Peter Guttorp for serving on my dissertation committee and for their helpful feedback. Thank you to Joel Kaufman and Sverre Vedal for being inspiring and energizing collaborators.

A shout out to the Biostatistics class of 2009 for being the best classmates of all time. Working together in the T-wing during our first few years made the coursework load and exams so much more bearable and created fond memories. You're a brilliant, kind group of people and I'm lucky to know you.

Thank you to classmates, friends and family for putting up with and supporting me when I was discouraged with my research and helping me to see the bigger picture. Especially to Lydia, the love of my life and my best friend. I'm so glad I got to take this grad school journey with you. Thanks for being anticonservative with your pep talks. This dissertation would not exist without you; I should probably add you as a co-author.

Chapter 1

INTRODUCTION

Air pollution epidemiology cohort studies often implement a two-stage approach to estimating associations of continuous health outcomes with one or more exposures. An inherent problem in these studies is that the exposures of interest are usually unobserved. Instead observations are available at misaligned monitoring locations. The first stage entails building exposure models with the monitoring data and predicting at subject locations; the second stage uses the predictions to estimate health effects. This induces spatial measurement error that can bias and affect the standard error of resulting estimates. Berkson-like error arises from smoothing the exposure surface, while classical-like error comes from estimating the exposure model parameters. Accurately characterizing and correcting for this measurement error depends on assumptions made about the spatial surface and exposure model used to derive predictions. This dissertation addresses spatial measurement error in air pollution epidemiology.

Chapter 2 applies and extends parametric measurement error methods for a single pollutant. We develop a national prediction model for elemental carbon (EC), organic carbon (OC), silicon (Si) and sulfur (S) in the Multi-Ethnic Study of Atherosclerosis (MESA) cohort. Our prediction models use partial least squares as the mean of universal kriging models. Universal kriging assumes the exposure surface is a realization of a parametric Gaussian process. In this setting the Berkson-like error does not induce bias while classical-like error can, and both error types affect the standard error. We review the parametric and parameter bootstrap methods for correcting for bias and accurately estimating standard errors under this assumption. We generalize the parameter bootstrap by developing parameter simulation extrapolation (P-SIMEX), a more flexible bias correction method analogous to standard simulation extrapolation (SIMEX). We apply these methods to single-pollutant

association studies of carotid intima media thickness with EC, OC, Si and S in the MESA cohort. In the accompanying Appendix A we compare the parametric bootstrap approaches to joint maximum-likelihood and Bayesian approaches that account for measurement error by jointly fitting the health and exposure models.

Chapter 3 develops semi-parametric methods for a single pollutant. We motivate viewing the exposure surface as fixed and unknown rather than a realization of a Gaussian process, and using fixed- but high-rank penalized regression splines to approximate universal kriging in this setting. We decompose the measurement error into Berkson-like and classical-like components, both of which can induce bias in this setting. We analyze how the exposure model penalty parameter regulates a trade-off between these two types of error and propose a method for selecting the penalty parameter to mitigate bias from measurement error. We develop a post-hoc measurement error correction methodology that combines an analytic bias correction with a design-based nonparametric bootstrap of the exposure and health data. We demonstrate these methods using simulation studies and apply them to an association study of systolic blood pressure with annual average $PM_{2.5}$ in the Sister Study.

Chapter 4 extends the methodology developed in Chapter 3 to simultaneously estimating associations of health with multiple pollutants. We again decompose the multivariate measurement error into a Berkson-like and classical-like component and develop an analytic bias correction that corrects for both types of error. We examine the extent of measurement error and the efficacy of the bias correction in multi-pollutant studies through simulations and in estimating the joint association of systolic blood pressure with $PM_{2.5}$ and NO_2 in the Sister Study.

Chapter 2

A NATIONAL PREDICTION MODEL FOR COMPONENTS OF $PM_{2.5}$ AND MEASUREMENT ERROR CORRECTED HEALTH EFFECT INFERENCE

2.1 Summary

Studies estimating health effects of long-term air pollution exposure often use a two-stage approach, building exposure models to assign individual-level exposures which are then used in regression analyses. This requires accurate exposure modeling and careful treatment of exposure measurement error. To illustrate the importance of carefully accounting for exposure model characteristics in two-stage air pollution studies, we consider a case study based on data from the Multi-Ethnic Study of Atherosclerosis (MESA). We present national spatial exposure models that use partial least squares and universal kriging to estimate annual average concentrations of four $PM_{2.5}$ components: elemental carbon (EC), organic carbon (OC), sulfur (S), and silicon (Si). Our models perform well, with cross-validated R^2s ranging from 0.62 to 0.95. We predict $PM_{2.5}$ component exposures for the MESA cohort and estimate cross-sectional associations with carotid intima-media thickness (CIMT), adjusting for subject-specific covariates. In naïve analyses that do not account for measurement error, we find statistically significant associations between CIMT and increased exposure to OC, S, and Si. We correct for measurement error using recently developed parametric methods that account for the spatial structure of predicted exposures. OC exhibits little spatial correlation, and the corrected inference is unchanged from the naïve analysis. The S and Si exposure surfaces display notable spatial correlation, resulting in corrected confidence intervals (CIs) that are 50% wider than the naïve CIs, but that are still statistically significant. The impact on health effect inference is concordant with the degree of spatial correlation in the exposure surfaces.

2.2 Introduction

The relationship between air pollution and adverse health outcomes has been well-documented (Samet et al., 2000; Pope et al., 2002). Many studies focus on particulate matter, specifically particulate matter less than 2.5 μm in aerodynamic diameter ($PM_{2.5}$) (Miller et al., 2007; Kim et al., 2009). Health effects of $PM_{2.5}$ could depend on characteristics of the particles, including shape, solubility, pH, or chemical composition (Vedal et al., 2012), and a deeper understanding of these differential effects could help inform policy. One of the challenges in assessing the impact of different chemical components of $PM_{2.5}$ in an epidemiological study is the need to assign exposures to study participants based on monitoring data at different locations (i.e., spatially misaligned data). When doing this for many components, the prediction procedure needs to be streamlined in order to be practical. Whatever the prediction algorithm, using the estimated rather than true exposures induces measurement error in the subsequent epidemiologic analysis. This paper describes a flexible and efficient prediction model that can be applied on a national scale to estimate long-term exposure levels for multiple pollutants and implements existing methods of correcting for measurement error in the health model.

Current methods for assigning exposures include land-use regression (LUR) with Geographic Information System (GIS) covariates (Hoek et al., 2008) and universal kriging (UK) that also exploits residual spatial structure (Kim et al., 2009; Mercer et al., 2011). Often hundreds of candidate correlated GIS covariates are available necessitating a dimension reduction procedure. Variable selection methods that have been considered in the literature include exhaustive search, stepwise selection, and shrinkage by the "lasso"(Tibshirani, 1996; Mercer et al., 2011). However, stepwise variable selection methods tend to be computationally intensive, feasible perhaps when considering a single pollutant but quickly becoming impractical when developing predictions for multiple pollutants. A more streamlined alternative is partial least squares (PLS) (Sampson et al., 2009), which finds a small number of linear combinations of the GIS covariates that most efficiently account for variability in the

measured concentrations. These linear combinations reduce the covariate space to a much smaller dimension and can then be used as the mean structure in a LUR or UK model in place of individual GIS covariates. This provides the advantages of using all available GIS covariates and eliminating potentially time-consuming variable selection processes.

Using exposures predicted from spatially misaligned data rather than true exposures in health models introduces measurement error that may have implications for $\hat{\beta}$, the estimated health model coefficient of interest (Szpiro et al., 2011b). Berkson-like error that arises from smoothing the true exposure surface may inflate the standard error of $\hat{\beta}$. Classical-like error results from estimating the prediction model parameters, and may bias $\hat{\beta}$ in addition to inflating its standard error. Bootstrap methods to adjust for the effects of measurement error have been discussed by Szpiro et al. (2011b). We present a case study to illustrate a holistic approach to two-stage air pollution epidemiology modeling, which includes exposure modeling in the first stage and health modeling that incorporates measurement error correction in the second stage. We build national exposure models using PLS and UK, and employ them to estimate long-term average concentrations of four chemical species of $PM_{2.5}$: elemental carbon (EC), organic carbon (OC), silicon (Si) and sulfur (S), selected to reflect a variety of different $PM_{2.5}$ sources and formation processes and to have differing chemical composition (Vedal et al., 2012). After developing the exposure models we derive predictions for the Multi-Ethnic Study of Atherosclerosis (MESA) cohort. These predictions are used as the covariates of interest in health analyses assessing associations between carotid intima-media thickness (CIMT), a subclinical measure of atherosclerosis, and exposure to $PM_{2.5}$ components. We apply measurement error correction methods to account for the fact that predicted rather than true exposures are being used in these health models. We discuss our results and the insight it provides into how exposure surfaces impact health inference.

2.3 Data

2.3.1 Monitoring data

Data on EC, OC, Si and S were collected to build the national models. These data consisted of annual averages from 2009-2010 as measured by the EPA's Interagency Monitoring for Protected Visual Environments (IMPROVE) and Chemical Speciation Network (CSN) (EPA 2009). The IMPROVE monitors form a nation-wide network located mostly in remote areas. The CSN monitors are in more urban areas. These two networks provide data that are evenly dispersed throughout the lower 48 states (Figure 2.2).

All CSN and IMPROVE monitors that had at least 10 data points per quarter and a maximum of 45 days between measurements were included in our analyses. For Si and S, averages were over 01/01/2009-12/31-2009. The EC/OC data set consisted of 204 IMPROVE and CSN monitors averaged over 01/01/2009-12/31-2009, and 51 CSN monitors averaged over 05/01/2009-04/30/2010. The latter period was used since prior to 05/01/2009 these monitors used a protocol that was incompatible with the IMPROVE network. Comparing averages over 05/01/2009-04/30/2010 to those which used comparable protocol over 01/01/2009-12/31-2009 indicated little difference between the time periods.

2.3.2 Geographic covariates

For all monitor and subject locations, approximately 600 LUR covariates were available. These included distances to A1, A2, and A3 roads (Census Feature Class Codes (CFCC)); land use within a given buffer; population density within a given buffer; and normalized difference vegetation index (NDVI) which measures the level of vegetation in a monitor's vicinity. CFCC A1 roads are limited access highways; A2 and A3 roads are other major roads such as county and state highways without limited access (Mercer et al., 2011). For NDVI a series of 23 monitor-specific, 16-day composite satellite images were obtained, and the pixels within a given buffer averaged for each image. PLS incorporated the 25th, 50th and 75th percentile of these 23 averages. The median of "high-vegetation season" image

averages (defined as April 1-September 30) and "low-vegetation season" averages (October 1-March 31) were also included.

The geographic covariates were pre-processed to eliminate LUR covariates that were too homogeneous or outlier-prone to be of use. We eliminated variables with $> 85\%$ identical values, and those with the most extreme standardized outlier > 7. We log-transformed and truncated all distance variables at 10 km and computed additional "compiled" distance variables such as minimum distance to major roads, distance to any port, etc. These compiled variables were then subject to the same inclusion criteria. All selected covariates were mean-centered and scaled by their respective standard deviations.

2.3.3 MESA Cohort

The Multi-Ethnic Study of Atherosclerosis (MESA) is a population-based study that began in 2000, with a cohort consisting of 6,814 participants from six U.S. cities: Los Angeles, CA; St. Paul, MN; Chicago, IL; Winston-Salem, NC; New York, NY; and Baltimore, MD. Four ethnic/racial groups were targeted: white, African American, Hispanic, and Chinese American. All participants were free of clinical cardiovascular disease at time of entrance.

As the health outcome for our case study we selected the common carotid intima-media thickness (CIMT) endpoint in MESA. CIMT, a subclinical measure of atherosclerosis, was measured by B-mode ultrasound using a GE Logiq scanner, and the endpoint was quantified as the right far wall CIMT measures conducted during MESA exam 1, which took place for 2000-2002 (Vedal et al., 2012). We considered the 5,501 MESA participants who had CIMT measures during exam 1; our analysis was based on the 5,298 MESA participants who had CIMT measures during exam 1 and complete values of confounding variables.

2.4 Methods

The following sections describes our methods. Section 2.4.1 describes the first stage of the two-stage approach, building the exposure models which use PLS scores to model the mean in UK models. We describe the cross-validation we implemented to select the number of

PLS scores, to determine how reliable predictions from each exposure model were, and to assess the extent to which spatial structure was present for each pollutant. Section 2.4.2 describes the second, health modeling stage. We describe the health models we fit, and the measurement error correction methods we employed. Readers interested in a more detailed technical exposition are referred to Bergen et al. (2012).

2.4.1 Spatial prediction models

Notation

To describe our exposure models, we introduce some notation. Let $\mathbf{X}^*$ denote the $n^* \times 1$ vector of observed concentrations at monitor locations; $\mathbf{R}^*$ the $n^* \times p$ matrix of geographic covariates at monitor locations; $\mathbf{X}$ the $n \times 1$ vector of unknown concentrations at the unobserved subject locations; and $\mathbf{R}$ the $n \times p$ matrix of geographic covariates at subject locations. Note that for our exposure models, $\mathbf{X}^*$ and $\mathbf{X}$ play the role of the dependent variables, while $\mathbf{R}^*$ and $\mathbf{R}$ play the role of the independent variables. PLS was used to decompose $\mathbf{R}^*$ into a set of linear combinations of much smaller dimension than $\mathbf{R}^*$. Specifically,

$$\mathbf{R}^*\mathbf{H} = \mathbf{P}^*.$$

Here, $\mathbf{H}$ is a $p \times k$ matrix of weights for the geographic covariates, and $\mathbf{P}^*$ is an $n^* \times k$ matrix of PLS components or scores. These scores are intelligently-found linear combinations of the geographic covariates that maximize the covariance between $\mathbf{X}^*$ and all possible linear combinations of $\mathbf{R}^*$. One might notice similarities between PLS and principal components analysis (PCA). Although the two methods are similar in that they are both dimension reduction methods, the scores from PLS maximize the covariance *between* $\mathbf{X}^*$ and all other possible linear combinations of $\mathbf{R}^*$, whereas the scores from PCA are found only to explain as much as possible the covariance of $\mathbf{R}^*$. Thus, though both methods aim to reduce the dimension of $\mathbf{R}^*$, PLS more efficiently captures any information about the association between the geographic covariates and the observed exposures. For more details see Sampson et al.

(2013). PLS scores at unobserved locations are then derived as $\mathbf{P} = \mathbf{RH}$.

Once the PLS scores $\mathbf{P}$ and $\mathbf{P}^*$ were obtained for the unobserved and monitoring locations, the following joint model was assumed to motivate predictions,

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{X}^* \end{pmatrix} = \begin{pmatrix} \mathbf{P} \\ \mathbf{P}^* \end{pmatrix} \boldsymbol{\alpha} + \begin{pmatrix} \boldsymbol{\eta} \\ \boldsymbol{\eta}^* \end{pmatrix}. \tag{2.1}$$

Here $\boldsymbol{\alpha}$ is a vector of regression coefficients for the PLS scores and $\boldsymbol{\eta}$, $\boldsymbol{\eta}^*$ are $n \times 1$ and $n^* \times 1$ vectors of errors. Our primary exposure models assumed that the error terms exhibited spatial correlation that could be modeled with a kriging variogram (Cressie, 1993). Specifically,

$$\begin{pmatrix} \boldsymbol{\eta} \\ \boldsymbol{\eta}^* \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{11}(\boldsymbol{\theta}) & \Sigma_{12}(\boldsymbol{\theta}) \\ \Sigma_{21}(\boldsymbol{\theta}) & \Sigma_{22}(\boldsymbol{\theta}) \end{pmatrix} \right). \tag{2.2}$$

Here each corner of the covariance matrix is a kriging covariance matrix parameterized by a common vector of parameters $\boldsymbol{\theta} = (\tau^2, \sigma^2, \phi)$. τ^2 is the nugget, interpretable as the amount of variability in the pollution exposures that is unexplainable by spatial structure; σ^2_η is the partial sill, interpretable as the amount of variability that is explainable by spatial structure; and ϕ is the range, interpretable as the maximum distance between two locations beyond which they may no longer be considered spatially correlated. We estimated these parameters and the regression coefficients $\boldsymbol{\alpha}$ via profile maximum likelihood, using exponential variograms for each pollutant. Having obtained parameter estimates, predictions $\mathbf{W}$ were defined as

$$\mathbf{W} = \mathbf{P}\hat{\boldsymbol{\alpha}} + \Sigma_{12}(\hat{\boldsymbol{\theta}})\Sigma_{22}(\hat{\boldsymbol{\theta}})^{-1}(\mathbf{X}^* - \mathbf{P}^*\hat{\boldsymbol{\alpha}}).$$

As a comparison to our primary kriging models we also derived predictions from PLS alone without fitting a kriging variogram. This is analogous to a pure land-use regression model using the PLS scores instead of actual geographic covariates. For this analysis $\boldsymbol{\eta}$ and $\boldsymbol{\eta}^*$ were assumed to be independent, and $\boldsymbol{\alpha}$ was estimated using a least-squares fit to

regression of $\mathbf{X}^*$ on $\mathbf{P}^*$. PLS-only predictions at the unobserved locations were then derived as $\mathbf{W} = \mathbf{P}\hat{\boldsymbol{\alpha}}$.

Cross-validation and Model Selection

10-fold cross-validation (CV) (Hastie et al., 2001) was used to assess the models' prediction accuracy, to select the number of PLS components to use in the final prediction models, and to compare predictions generated using PLS only to our primary models which used both PLS and UK. Data were randomly assigned to one of ten groups. One group (a "test set") was omitted, and the remaining groups (a "training set") were used to fit the model and generate test set predictions. Each group played the role of test set until predictions were obtained for the entire data set. At each iteration, the following steps were taken to cross-validate our primary models; similar steps were followed to derive cross-validated predictions that used PLS only:

1. PLS was fit using the training set, and K scores were computed for the test set, for $K = 1, ..., 10$.

2. UK parameters $\boldsymbol{\theta}$ and coefficients $\boldsymbol{\alpha}$ were estimated via profile maximum likelihood using the training set. The first K PLS scores played the role of $\mathbf{P}^*$ in Equation 2.1, for $K = 1, ..., 10$.

3. Predictions were derived using the first K PLS components and the corresponding UK, using kriging parameters estimated from the training set.

The `R` package `pls` was used to fit the PLS. UK was done using the `R` package `geoR`. The best-performing models were selected out of those that used both PLS and kriging based on their cross-validated root mean squared prediction error (RMSEP) and corresponding R^2. For a data set with n^* observations and corresponding predictions, the formulae for these performance metrics are given by

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n^*}(Obs_i - Pred_i)^2}{n^*}},$$
$$R^2 = \max\left(0, 1 - \frac{\text{RMSEP}^2}{\text{Var}(Obs)}\right).$$

These metrics are sensitive to scale; accordingly they are useful for evaluating model performance for a given pollutant, but not for comparing models across pollutants.

2.4.2 Health modeling

Disease model

Multivariable linear regression models were used to estimate the effects of $PM_{2.5}$ component exposure on CIMT. Each model included a single $PM_{2.5}$ component along with a vector of subject-specific covariates. Let $\mathbf{Y}$ be the 5298×1 vector of health outcomes, $\mathbf{X}$ the 5298×1 vector of unobserved exposures, $\mathbf{W}$ the 5298×1 vector of predictions, and $\mathbf{Z}$ a matrix of potential confounders. We assume a truly linear relationship between $\mathbf{Y}$, $\mathbf{X}$ and $\mathbf{Z}$, as follows:

$$\mathbf{Y} = \beta_0 + \mathbf{X}\beta + \mathbf{Z}\boldsymbol{\beta}_Z + \boldsymbol{\epsilon}. \tag{2.3}$$

Here $(\epsilon_1, ..., \epsilon_N)$ are i.i.d. $N(0, \sigma_\epsilon^2)$ random variables, and we estimate β by ordinary least squares (OLS) using $\mathbf{W}$ instead of $\mathbf{X}$.

Measurement Error Correction

The model in Equation 2.3 was fit using the predicted exposures $\mathbf{W}$ instead of the true exposures as the covariate of interest. Using predictions rather than true exposures in health modeling introduces two sources of measurement error that potentially influence the behavior of $\hat{\beta}$. Berkson-like error arises from smoothing the true exposure surface and could inflate the standard error of $\hat{\beta}$. Classical-like error arises from estimating the exposure model

parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$. The classical-like error potentially inflates the standard error of $\hat{\beta}$ and could also bias the estimate. We implemented the parametric and parameter bootstraps to assess and correct for the effects of measurement error. See Szpiro et al. (2011b) for additional background and details.

We describe the parametric bootstrap in the context of predictions that use both PLS and UK; the approach would be very similar if PLS alone was used (though we did not implement that correction here).

1. Estimate exposure model parameters $\boldsymbol{\alpha}$ and $\log(\boldsymbol{\theta})$ by nonlinear optimization using $\mathbf{X}^*$ and $\mathbf{P}^*$, exploiting the likelihood defined by Equations 2.1 and 2.2.

2. Derive $\mathbf{W}$ and use in place of $\mathbf{X}$ along with $\mathbf{Z}$ to obtain $\hat{\beta}_0$, $\hat{\beta}$, $\hat{\boldsymbol{\beta}}_Z$ and $\hat{\sigma}^2_\epsilon$.

3. For $j = 1, ..., B$ bootstrap samples:

   (a) Simulate $\mathbf{X}^*_j$ (and $\mathbf{X}_j$) from Equation 2.1 and $\mathbf{Y}_j$ from Equation 2.3, using $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\theta}}$, $\hat{\beta}_0$, $\hat{\beta}$, $\hat{\boldsymbol{\beta}}_Z$ and $\hat{\sigma}^2_\epsilon$ in place of unknown parameters

   (b) Estimate new exposure parameters $\hat{\boldsymbol{\alpha}}_j$ and $\log(\hat{\boldsymbol{\theta}}_j)$ by nonlinear optimization using $\mathbf{X}^*_j$

   (c) Use $\hat{\boldsymbol{\alpha}}_j$, $\hat{\boldsymbol{\theta}}_j$ and $\mathbf{X}^*_j$ to derive $\mathbf{W}_j$

   (d) Calculate $\hat{\beta}_j$ using $\mathbf{W}_j$ by OLS.

4. Let $\widehat{Bias}_p(\hat{\beta}) = \left(E_p(\hat{\beta}^B) - E_0(\hat{\beta}^B)\right)$, were $E_p(\hat{\beta}^B)$ denotes the empirical mean of the $\hat{\beta}_j$. We define $E_0(\hat{\beta}^B)$ when we define the parameter bootstrap; for now it can be thought of as a bootstrap mean that only incorporates Berkson-like error. The corresponding corrected effect estimate is $\hat{\beta}^C_{X,p} = \hat{\beta} - \widehat{Bias}_p(\hat{\beta})$.

5. Estimate the bootstrap standard error as

$$\widehat{SE}(\hat{\beta}) = \sqrt{\frac{\sum_{j=1}^{B}\left(\hat{\beta}_j - \frac{1}{B}\sum_{j=1}^{B}(\hat{\beta}_j)\right)^2}{B}}.$$

For the parametric bootstrap we set $B = 15,000$. Note that using Step 5 to estimate the standard error of $\hat{\beta}_p^C$ will give an underestimate, as it does not account for the additional variability introduced by $\widehat{Bias}_p(\hat{\beta})$. To fully account for this variability we would need to perform a nested bootstrap within each original bootstrap sample; however as discussed in Szpiro et al. (2011b) (and as exemplified in our results) the estimated bias is so small that this underestimation is ignorable in practice.

An undesirable trait of the parametric bootstrap is the computational time required to implement it, as it requires B non-linear optimizations. The parameter bootstrap is a much more efficient alternative, which approximates the parametric bootstrap by adding a Step 1(a) and altering Step 3(b) of the algorithm described above:

1. Estimate exposure model parameters $\boldsymbol{\alpha}$ and $\log(\boldsymbol{\theta})$ by nonlinear optimization using $\mathbf{X}^*$ and $\mathbf{P}^*$, exploiting the likelihood defined by Equation 2.1.

   (a) Estimate a sampling density for $\hat{\boldsymbol{\alpha}}$ and $\log(\hat{\boldsymbol{\theta}})$ with a multivariate normal distribution.

2. Derive $\mathbf{W}$ and use it in place of $\mathbf{X}$ along with $\mathbf{Z}$ to obtain $\hat{\beta}_0$, $\hat{\beta}$, $\hat{\boldsymbol{\beta}}_Z$ and $\hat{\sigma}_\epsilon^2$.

3. For $j = 1, ..., B$ bootstrap samples:

   (a) Simulate $\mathbf{X}_j^*$ (and $\mathbf{X}_j$) from Equation 2.1 and $\mathbf{Y}_j$ from Equation 2.3, using $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\theta}}$, $\hat{\beta}_0$, $\hat{\beta}$, $\hat{\boldsymbol{\beta}}_Z$ and $\hat{\sigma}_\epsilon^2$ in place of unknown parameters

   (b) Simulate new exposure model parameters $\hat{\boldsymbol{\alpha}}_j$ and $\log(\hat{\boldsymbol{\theta}}_j)$ from the sampling density estimated in Step 1(a), using a constant covariance matrix multiplied by a scalar $\delta \geq 0$. δ controls the variability of $(\hat{\boldsymbol{\alpha}}_j, \log(\hat{\boldsymbol{\theta}}_j))$: the larger δ is the greater the variability of $(\hat{\boldsymbol{\alpha}}_j, \log(\hat{\boldsymbol{\theta}}_j))$.

   (c) Use $\hat{\boldsymbol{\alpha}}_j$, $\hat{\boldsymbol{\theta}}_j$ and $\mathbf{X}_j^*$ to derive $\mathbf{W}_j$

   (d) Calculate $\hat{\beta}_j$ using $\mathbf{W}_j$ by OLS.

4. Let $E_\delta(\hat{\beta}^B)$ denote the empirical mean of the $\hat{\beta}_j$. The estimated bias is defined as $\widehat{Bias}_\delta(\hat{\beta}) = \left(E_\delta(\hat{\beta}^B) - E_0(\hat{\beta}^B)\right)$, with corresponding bias-corrected effect estimate $\hat{\beta}^C_{X,\delta} = \hat{\beta} - \widehat{Bias}_\delta(\hat{\beta})$.

5. Estimate the bootstrap standard error as

$$\widehat{SE}_\delta(\hat{\beta}) = \sqrt{\frac{\sum_{j=1}^{B}\left(\hat{\beta}_j - E_\delta(\hat{\beta}^B)\right)^2}{B}}.$$

For Step 1(a) we used a multivariate normal distribution centered at $(\hat{\boldsymbol{\alpha}}, \log(\boldsymbol{\theta}))$ with covariance equal to the inverse Hessian of the negative log-likelihood evaluated at $(\hat{\boldsymbol{\alpha}}, \log(\boldsymbol{\theta}))$ and implemented the parameter bootstrap with $15,000$ bootstrap samples and $\delta = 1$. We used a multivariate normal so that bootstrapped $\{\hat{\boldsymbol{\alpha}}_j, \hat{\boldsymbol{\theta}}_j\}$ would be centered at $\{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}}\}$ and have a variability we could control independently by regulating δ.

The intent of both bootstrap approaches is to approximate the sampling properties of the measurement error-impacted $\hat{\beta}$ we would see if we performed our two-stage analysis with many actual realizations of monitoring observations and subject health data sets. We illustrate this by focusing on the parameter bootstrap algorithm outlined above. Step 3(a) gives us B new "realizations" of our data. For $\delta = 1$, step 3(b) accounts for the classical-like error by re-sampling the exposure model parameters. Step 3(c) accounts for the Berkson-like error by smoothing the true exposure surface. Step 3(d) then calculates B new $\hat{\beta}_j$'s, the sampling properties of which have incorporated all sources of measurement error. Comparing these to the mean of bootstrapped $\hat{\beta}_j$ derived using fixed exposure model parameters (i.e., $\delta = 0$) gives us an approximation of the bias induced by the classical-like error (Step 4), and the empirical standard deviation approximates the standard error that accounts for both sources of measurement error (Step 5).

Note that we also implement the parameter bootstrap for $\delta = 0$. This is equivalent to the "partial parametric bootstrap" described in Szpiro et al. (2011b), which corrects

for the Berkson-like error only since the exposure surface is still smoothed but with fixed parameters.

A desirable trait of the parameter bootstrap is the ability to "tune" the amount of the classical-like error by varying δ, which allows us to investigate how variability in the sampling distribution of $(\hat{\boldsymbol{\alpha}}, \log(\hat{\boldsymbol{\theta}}))$ affects the bias of $\hat{\beta}$. This can be useful in refining our bootstrap bias estimates and bears similarly to simulation extrapolation (SIMEX) (Stefanski and Cook, 1995). Consider first the parameter bootstrap with $\delta = 0$. Although with $\delta = 0$ we use the originally estimated parameters $(\hat{\boldsymbol{\alpha}}, \log(\hat{\boldsymbol{\theta}}))$ in all bootstrap samples, these original estimates are themselves a realization from a sampling distribution with what we heuristically term "one unit of variance." The bias estimate from the parameter bootstrap with $\delta = 1$ (corresponding to "two units of variance") assumes that the bias induced by classical-like error going from $\delta = 0$ to $\delta = 1$ is the same as the bias induced by using $(\hat{\boldsymbol{\alpha}}, \log(\hat{\boldsymbol{\theta}}))$ instead of $(\boldsymbol{\alpha}, \log(\boldsymbol{\theta}))$. In other words, the bias is treated as linear in δ. However if we perform the parameter bootstrap using different values of δ and estimate the bias for each one, we can get a more flexible representation of how the bias varies as a function of δ. Plotting realized $\widehat{Bias}_{\delta}(\hat{\beta}_X)$ versus δ for several values of δ and extrapolating to $\widehat{Bias}_{-1}(\hat{\beta}_X)$ gives an alternative estimate of the bias. This is analogous to performing a SIMEX analysis in the context of pure classical measurement error to extrapolate to the hypothetical setting where the variance of the error is zero (Stefanski and Cook, 1995). However, it differs from SIMEX in that the measurement error is not purely classical; accordingly we term this approach P-SIMEX. We performed P-SIMEX using sample sizes of 15,000, sampling $(\hat{\boldsymbol{\alpha}}_j, \log(\hat{\boldsymbol{\theta}}_j))$ from the inverse Hessian inflated by factors of $\delta \in \{0, 0.5, 1, 1.5, 2\}$ and plotted the corresponding $\widehat{Bias}_{\delta}(\hat{\beta}_X)$ against these values of δ. We then performed both linear and quadratic extrapolation to $\widehat{Bias}_{-1}(\hat{\beta}_X)$. The P-SIMEX-corrected estimate of $\hat{\beta}_X$ is defined as:

$$\hat{\beta}^{C}_{X,PS} = \hat{\beta}_X + \widehat{Bias}_{-1}(\hat{\beta}_X).$$

2.5 Results

2.5.1 Data

Monitoring data

Concentrations of the four pollutants by monitoring network are shown in Table 2.1. Table 2.1 indicates that the EC and OC concentrations measured by CSN monitors tended to be higher than those measured by IMPROVE monitors. Average Si and S concentrations measured by CSN monitors were also higher than the IMPROVE averages, but relative to their standard deviations the differences between CSN and IMPROVE monitors in Si and S concentrations were not as great as the EC and OC concentrations.

Geographic Covariates

The geographic variables that were selected as a result of the pre-processing procedure discussed in Section 2.3.2 are shown in Table 2.2. Table 2.1 shows the distributions of select geographic covariates, by monitoring network and at MESA locations. The summaries in Table 2.1 reflect the urban versus rural difference in placement between the IMPROVE and CSN monitors. Although relatively few monitors belonging to either IMPROVE or CSN were within 150 m of an A1 road there was a larger proportion of CSN monitors within 150 m of an A3 road (44%) than IMPROVE (19%). The median distance to commercial and service centers was much smaller for CSN monitors (127 m) than it was for IMPROVE monitors (4696 m). The median population density was much larger for CSN monitors (805 people/mi^2) than for IMPROVE monitors (only 3 people/mi^2). The median summer NDVI values within 250 m were slightly smaller for CSN monitors than for IMPROVE monitors, indicating IMPROVE monitors were located in greener areas. Table 2.1 also shows that MESA participant locations had covariate distributions that more closely mirrored the CSN monitors, as is especially evident for the number of sites less than 150 m from an A3 road and median population density.

Additionally we examined density plots of the geographic covariates for monitoring and subject locations, and saw significant overlap for all geographic covariates. This reassured us that the difference in geographic covariates in Table 2.1 was due to concentration of MESA subjects in urban locations, not extrapolation beyond our data.

MESA cohort

Summary statistics for the MESA cohort are in Table 2.3 Mean CIMT was 0.68 mm. The other variables summarized are the ones that were included as covariates in the health model. The mean age was 62 years, and the cohort was 52% female. 39% were white, 27% African-American, 22% Hispanic, and 12% Chinese. 44% had hypertension and 15% used a statin drug. The highest percentage of participants resided in Los Angeles (19.7%), but the distribution across the 6 cities was quite homogeneous. Only the 5,298 participants that had complete values of all the variables listed in Table 2.3 were included in the analysis.

2.5.2 Spatial prediction models

Model evaluation

The selected models corresponding to lowest cross-validated R^2 all used PLS and UK. For all four $PM_{2.5}$ components and for all numbers of PLS scores, using kriging improved prediction accuracy. Table 2.4 shows the number of PLS components used and the R^2 for the selected prediction models corresponding to the best-performing models. The CV statistics for the PLS only models are shown to illustrate the extent to which UK improved prediction accuracy. EC and OC were minimally improved; there was more improvement evident for Si and substantial improvement for the S predictions. The ratio of the nugget to the sill given in Table 2.4 also indicates the importance of spatial smoothing. For a fixed range, smaller values of this ratio indicate that there is more information about concentration variability at nearby locations and thus UK predictions draw heavily from nearby monitors.

Interpretation of PLS

Figure 2.1 can be used to examine which of the geographic covariates were most important for explaining pollutant variability. Specifically, Figure 2.1 summarizes the $p \times 1$ vector $\mathbf{m}$, the vector such that $\mathbf{Rm}$ equals the PLS-only predicted exposures. Each element of $\mathbf{m}$ is a weight for a corresponding geographic covariate. Positive elements in $\mathbf{m}$ indicate that increases in its corresponding geographic covariate were associated with higher predicted exposure; the larger the absolute value of an element in $\mathbf{m}$, the more the corresponding geographic covariate contributed to exposure prediction.

We see that increases in population density were associated with larger predicted values of all pollutants, but most strongly for EC, OC and S. Industrial land use within the smallest buffer was very predictive of EC, and evergreen forest land within a given buffer was strongly predictive of *decreases* in S. The NDVI, intense land use, emissions, and line-length variables were all positively associated with exposure, while the distance to source variables were negatively associated. The NDVI variables were more heavily exploited for OC and S than they were for EC. For Si, the NDVI and intense land use variables appeared to be the most informative and were mostly negatively associated with Si exposure. Proximity to features appeared to be informative for all four pollutants.

Exposure predictions

Figure 2.2 shows the predicted national concentrations, with finer detail illustrated for St. Paul, MN. The EC and OC predictions were much higher in the middle of urban areas, and quickly dissipated further from urban centers. S predictions were high across the midwestern and eastern states and in the Los Angeles area, and lower in the plains and mountains. Si predictions were low in most urban areas, and high in desert states.

Table 2.1 summarizes the predicted exposures for the MESA participants. Mean predicted EC and OC exposure concentration were 0.74 and 2.17 $\mu g/m^3$, respectively. Mean predicted Si and S exposure concentration was 0.09 ng/m^3 and 0.78 $\mu g/m^3$, respectively.

2.5.3 Health models

The results from the naïve health model that did not include any measurement error correction, as well as the results from the health modeling that included bootstrap-corrected point estimates and standard errors of $\hat{\beta}$, are displayed in Table 2.5. The naïve analysis found significant associations between OC, Si, and S and elevated CIMT. There was also evidence of association with EC, but this was not statistically significant. The point estimates and standard errors for the EC and OC health effects were virtually unchanged when measurement error correction was implemented, while the bootstrap-corrected standard errors for Si and S were about 50% larger than their respective naïve estimates. The estimated biases resulting from the classical-like measurement error were so small as to be uninteresting from an epidemiologic perspective.

Figure 2.3 shows the results of the P-SIMEX implementation of the parameter bootstrap using linear and quadratic extrapolation. We see that the choice of extrapolating function slightly affected the P-SIMEX estimate of the bias for all four of the pollutants, though the bias was so small that the differences between the extrapolating function were trivial. Overall, while the SIMEX bias corrections did not suggest any meaningful bias for any of the pollutants, all of these plots suggest that the bias from classical-like measurement error is away from the null, similar to previously published simulation results (Szpiro et al. 2011). This is different from the usual bias toward the null from classical measurement error, confirming that additional caution is needed in the air pollution setting since we cannot always assume that ignoring measurement error results in conservative inference.

2.6 Discussion

2.6.1 Summary

We have presented a comprehensive two-stage approach to analyzing long-term effects of air pollution exposure, and have applied our framework in a case study of four components of $PM_{2.5}$ and measurement error corrected inference of association between these components

and CIMT in the MESA cohort. Our approach includes a national prediction model for individual components and correction for measurement error in the epidemiologic analysis using a methodology that accounts for differing amounts of spatial structure in the exposure surfaces. Corrected standard errors corresponding to pollutants that exhibited significant spatial structure were 50% larger than naïve estimates.

2.6.2 National exposure models

We find that a national approach to exposure modeling is reasonable and performs well in terms of prediction accuracy. Our primary models resulted in cross-validated R^2 that ranged as high as 0.96 and no lower than 0.50 for any of the $PM_{2.5}$ components. Use of kriging improved the cross-validated R^2 for all four pollutants over comparison models that used PLS only, although the improvement was not equal across all four pollutants. These results are useful in terms of understanding the spatial nature of our exposure surfaces. For EC and OC, the R^2 only improved by at most 0.07 when kriging was used compared to when PLS alone was used, indicating little large-scale spatial structure in these pollutants. For Si, the R^2 improved from 0.32 to 0.50, and from 0.62 to 0.96 for S. This indicates that S (and to a lesser extent Si) had significant large-scale spatial structure that kriging was able to exploit. For all models using kriging only improved R^2, indicating nothing was lost (and quite a bit stood to be gained, when spatial structure was present) by using kriging.

One might question why we advocate two-stage modeling instead of joint modeling of exposure and health. The parametric assumptions that justify the bootstrap approaches also justify fitting a joint health and exposure model. Such an approach could result in efficiency gain in estimating β if the parametric assumptions hold. In Appendix A we estimate β using joint maximum-likelihood and Bayesian approaches, compare these results to the two-stage approach, and discuss strengths and weaknesses of each.

2.6.3 Epidemiologic case study

In this case study we focused on four $PM_{2.5}$ components. These were selected to gain insight into the sources or features of $PM_{2.5}$ that might contribute to the effects of $PM_{2.5}$ on cardiovascular disease. Elemental carbon and organic carbon were chosen as markers of primary emissions from combustion processes, with OC in addition including contributions from organic aerosols formed secondarily from atmospheric chemical reactions; silicon was chosen as a marker of crustal dust; and sulfur was chosen as a marker of sulfate, an inorganic aerosol formed secondarily from atmospheric chemical reactions (Vedal et al., 2012). The mechanisms whereby exposure to $PM_{2.5}$ or $PM_{2.5}$ components produce cardiovascular effects such as atherosclerosis are not well understood, although several mechanisms have been proposed (Brook et al., 2010). For discussion of other studies examining the effects of these pollutants, see Vedal et al. (2012).

Because nearest-monitor interpolation performed so poorly for EC, OC and Si one would not trust epidemiologic inference that used predictions derived from that method. For S, the only pollutant for which our models and nearest-monitor interpolation performed comparably, the health effect estimate using nearest-monitor interpolation was 0.074 (SE: 0.018), comparable to the naïve inference made using predictions from our exposure models. However, there is no way to correct for measurement error using this method, which is another significant advantage of our models.

Our naïve health analyses which used predictions from our national models showed a significant association with CIMT and OC, Si, and S, but not with EC. Using the parameter bootstrap to account and correct for measurement error led to noticeably larger standard errors and wider confidence intervals for Si and S, but the results still indicated a significant association between exposure to OC, Si and S and elevated CIMT regardless of whether or not measurement error was corrected for.

2.6.4 Measurement error correction

To interpret our measurement error results, it is helpful to compare the importance of the PLS and kriging aspects of the four prediction models. For EC and OC, using PLS alone was sufficient to make accurate predictions, whereas the spatial smoothing from UK was much more important in improving prediction accuracy for Si and S. It is accordingly no coincidence that the bootstrap-corrected standard error estimates for EC and OC were unchanged from the naïve estimates, while the corrected SE estimates for Si and S were about 50% larger (and the resulting 95% confidence intervals 50% wider) than their respective naïve estimates. The fact that the EC and OC exposure predictions were derived mostly from the PLS components only with independent residuals implies that the Berkson-like error was almost pure Berkson error (i.e., independent across location), which is correctly accounted for by naïve standard error estimates. On the other hand, much more smoothing took place for S and Si which induced spatial correlation in the residual difference between true and predicted exposure. Accordingly, standard errors that correctly account for the Berkson-like error in these two pollutants are inflated because the correlated errors in the predictions translate into correlated residuals in the disease model that are not accounted for by naïve standard error estimates (Szpiro et al., 2011b) . The fact that the standard error estimates from the parameter bootstrap using $\delta = 1$ (which accounts for both Berkson-like and classical-like error) and using $\delta = 0$ (which accounts only for Berkson-like error) were so similar further indicates that the larger corrected SE estimates were most likely a result of the Berkson-like error. None of our measurement error analyses indicated that any important bias was induced by the classical-like error.

2.6.5 Limitations and model considerations

Although our exposure models perform well there is still room for improvement, especially for the Si, EC and OC models which had high but not outstanding cross-validated R^2. For these models it is possible that inclusion of additional geographic covariates in the PLS would

help improve model performance. Examples include wood burning sources within a given buffer for EC and OC concentrations, or dust and sand sources for Si. These covariates are currently not available in our databases. Furthermore, while it is possible to interpret the individual covariates in PLS components as in Section 2.5.2, such interpretations need to be regarded with caution because inclusion of many correlated covariates can lead to apparent associations that are counter-intuitive and opposite what might be expected scientifically. Finally, PLS does not consider interactions or nonlinear combinations of the geographic covariates, which could improve model performance.

2.6.6 Implications and future directions

Our results show that careful investigation of the exposure model characteristics can help to clarify the implications for the subsequent epidemiologic analyses that use the predicted exposures. As is pointed out in Szpiro et al. (2011a), such an overarching framework that considers the end goal of health modeling is a more reasonable and scientifically valid approach than treating exposure models as if they exist for their own sake. This analysis serves as an example that will inform ongoing efforts by our group and others to construct and utilize exposure prediction models that are most suitable for epidemiologic studies.

Our epidemiologic inference was based on one health model per pollutant. One might reasonably be interested in how multiple pollutants jointly affect health. A limitation of the approaches applied here is the reliance on a correctly-specified, spatially-correlated Gaussian process exposure surface. To extend these approaches to multiple pollutants one faces the difficult task of correctly specifying a multi-pollutant stochastic spatial surface. This motivates moving away from the constraints of parametric assumptions and toward semi-parametric exposure modeling and appropriate measurement error treatment. In the following chapter we develop spatial measurement error methods under semi-parametric exposure modeling for a single pollutant.

Location	# Sites	EC ($\mu g/m^3$)	OC ($\mu g/m^3$)	Si (ng/m^3)	S ($\mu g/m^3$)	#Sites <150m to A1 (%)	# Sites <150m to A3 (%)	Med dist to Comm[a]	Med Pop dens[b]	NDVI[c]
IMPROVE	190	0.19 (0.18)	0.93 (0.55)	0.16 (0.12)	0.41 (0.27)	4 (2)	36 (19)	4696	3	150
CSN	98	0.66 (0.24)	2.23 (0.71)	0.10 (0.09)	0.69 (0.25)	3 (3)	43 (44)	127	805	140
All monitors	288	0.37 (0.30)	1.43 (0.88)	0.14 (0.11)	0.51 (0.29)	7 (2)	79 (27)	1235	20	146
MESA Air	5501	0.74 (0.18)	2.17 (0.36)	0.09 (0.03)	0.78 (0.15)	349 (6)	2763 (50)	302	3496	137

[a] Median distance to commercial or service centers, in meters
[b] People/mi^2 for census block/block group monitor/subject belongs to
[c] Median value of summer NDVI medians within 250m buffer

Table 2.1: Mean (SD) concentration at IMPROVE and CSN monitoring networks and over both networks taken together; and predicted concentrations for the MESA Air cohort. Also shown are summary statistics of selected land-use regression covariates.

Figure 2.1 abbreviation	Variable description	Buffer sizes
	Distance to features, in km[a]	
distance to features	A1 road	NA
	Nearest road	NA
	Airport	NA
	Large airport	NA
	Port	NA
	Coastline‡	NA
	Commercial or service center	NA
	Railroad	NA
	Railyard	NA
	Emissions[b]	
so2	SO_2	30km
pm25	$PM^{\dagger}_{2.5}$	30km
pm10	$PM^{\dagger}_{10}$	30km
nox	NO_X	30km
	Population	
population	$\log_{10}$ population density	500m, 1km, 1.5km, 2km, 2.5km, 3km, 5km, 10km, 15km
	NDVI	
ndvi.winter	Median winter	250m, 500m, 1km, 2.5km, 5km, 7.5km, 10km
ndvi.summer	Median summer	250m, 500m, 1km, 2.5km, 5km, 7.5km, 10km
ndvi.q75	75^{th} %ile	250m, 500m, 1km, 2.5km, 5km, 7.5km, 10km
ndvi.q50	50^{th} %ile	250m, 500m, 1km, 2.5km, 5km, 7.5km, 10km
ndvi.q25	25^{th} %ile	250m, 500m, 1km, 2.5km, 5km, 7.5km, 10km
	Land use	
transport	Transportation, communities and utilities	750m, 3km, 5km, 10km, 15km
transition	Transitional areas	15km
stream	Streams and canals	3km†, 5km, 10km, 15km
shrub	Shrub and brush rangeland	1.5km, 3km, 5km, 10km, 15km
resi	Residential	400m, 500m, 750m, 1km, 1.5km, 3km, 5km, 10km, 15km
oth.urban	Other urban or built-up	400m†, 500m, 1.5km, 3km, 5km, 10km, 15km
mix.range	Mixed rangeland	3km, 5km, 10km, 15km
mix.forest	Mixed forest land	750m, 1km, 1.5km, 3km, 5km, 10km, 15km
lakes	Lakes †	10 km
industrial	Industrial	1km*, 1.5km*, 3km, 5km, 10km, 15km
industcomm	Industrial and commercial complexes †	15km
herb.range	Herbaceous rangeland	3km†, 5km, 10km
green	Evergreen forest land	400m, 500m, 750m, 1km, 1.5km, 3km, 5km, 10km, 15km
forest	Deciduous forest land	750m, 1km, 1.5km, 3km, 5km, 10km, 15km
crop	Cropland and pasture	400m, 500m, 750m, 1km, 1.5km, 3km, 5km, 10km, 15km
comm	Commercial and services	500m, 750m, 1km, 1.5km, 3km, 5km, 10km, 15km
	Line lengths	
a23	Total dist of A2 and A3 roads within buffer	100m, 150m, 300m, 400m, 500m, 750m, 1km, 1.5km, 3km, 5km
a1	Total dist of A1 roads within buffer	1km, 1.5km, 3km, 5km

[a] Truncated at 25km and $\log_{10}$ transformed
[b] Tons per year of emissions from tall stacks
† Variable used for modelling Si, S only
* Variable used for modelling EC, OC only
‡ $\log_{10}$ and untransformed values both included

Table 2.2: Land-use regression covariates and (where applicable) covariate buffer sizes that made it through pre-processing and were considered by PLS. Most variables were used in each of the four $PM_{2.5}$ component models; however the pre-processing procedure selected some variables for EC and OC that were not selected for Si and S, and vice versa. This is due to the fact that the monitors used to measure EC and OC were not all identical to the ones used to measure Si and S.

Variable	N	Mean (SD) or %
CIMT	5501	0.68 (0.19)
Age	5501	61.9 (10.1)
Weight (lb)	5501	173.0 (37.5)
Height (cm)	5501	166.6 (10.0)
Waist (cm)	5500	97.8 (14.1)
Body surface area	5501	1.9 (0.2)
BMI (kg/m^2)	5501	28.2 (5.3)
DBP	5499	71.8 (10.3)
Gender		
Female	2872	52.2
Male	2629	47.8
Race		
White, caucasian	2168	39.4
Chinese American	675	12.3
Black, African-American	1459	26.5
Hispanic	1199	21.8
Site		
New York	867	15.8
Baltimore	776	14.1
St. Paul & Minneapolis	899	16.3
Chicago	998	18.1
Los Angeles	1083	19.7
Education		
Complete high school	991	18.0
Some college	1571	28.6
Complete college	2010	36.5
Missing	13	0.2
Income		
$< \$12,000$	566	10.3
\$12,000-24,999	1022	18.6
\$25000-49999	1543	28.0
\$50000-74999	901	16.4
$> \$75000$	1271	23.1
Missing	198	3.6
Hypertension		
No	3106	56.5
Yes	2395	43.5
Statin use		
No	4681	85.1
Yes	817	14.9
Missing	3	0.1

Table 2.3: Summary of characteristics of the MESA cohort. Only variables that were used in the health modeling are summarized here.

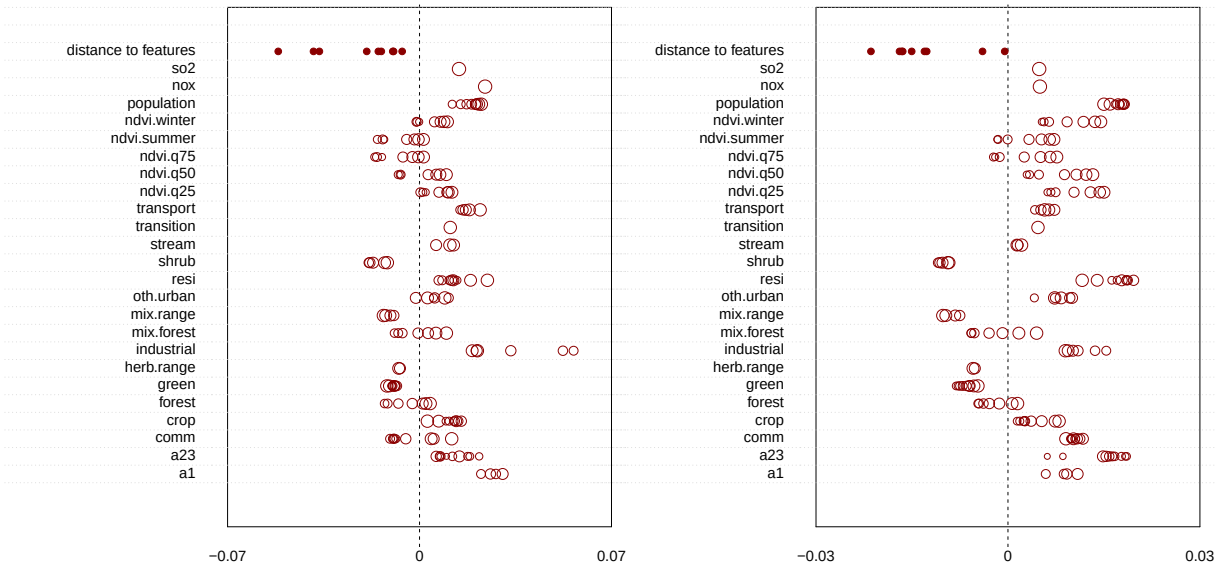
		R^2		Est. UK pars			
Pollutant	# Scores	PLS	PLS+UK	$(\tau^2)^a$	$(\sigma^2_\eta)^b$	$(\phi)^c$	τ^2/σ^2_η
EC	3	0.76	0.78	0.0153	0.0027	535	5.67
OC	2	0.54	0.61	0.1920	0.1467	241	1.31
Si	2	0.32	0.50	0.0044	0.0083	1698	0.53
S	2	0.62	0.96	0.0009	0.0510	1949	0.02

[a] Nugget used in kriging
[b] Partial sill used in kriging
[c] Range used in kriging

Table 2.4: Cross-validated R^2 for each component of $PM_{2.5}$, for both primary models and comparison PLS only models. The estimated kriging parameters from the likelihood fit on the entire data set for each pollutant is also shown.

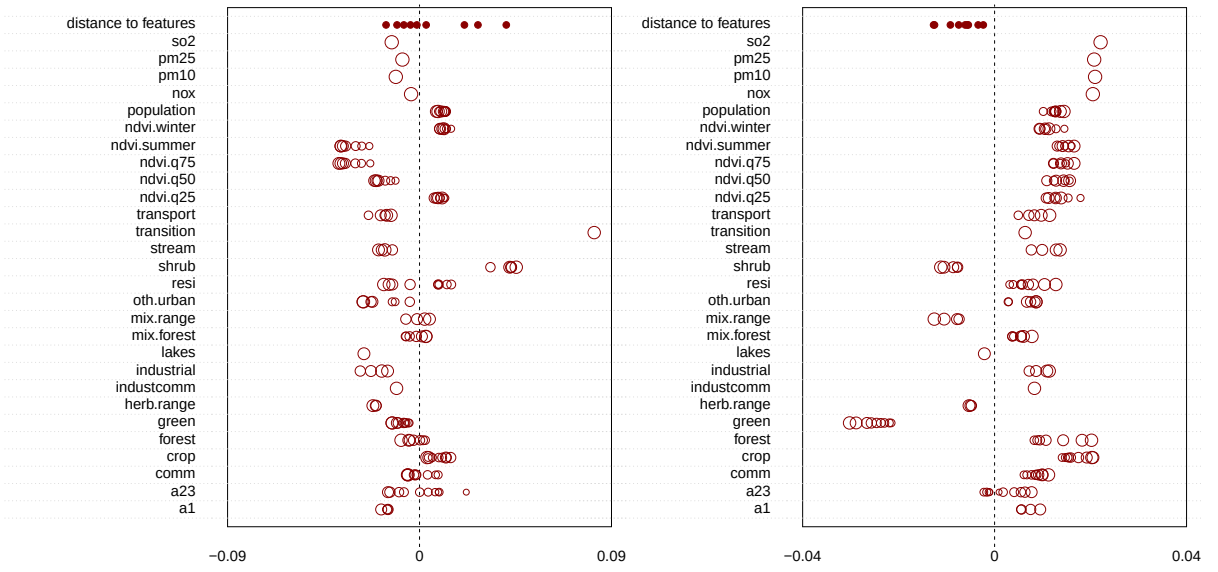
	$\hat{\beta}$ (SE)	95% CI	$\hat{\beta}$ (SE)	95% CI
	EC		OC	
Naïve	-0.011 (0.015)	(-0.04, 0.02)	0.025 (0.008)	(0.01, 0.04)
Parametric	-0.011 (0.015)	(-0.04, 0.02)	0.026 (0.009)	(0.01, 0.04)
PB, $\delta = 0$	-0.011 (0.015)	(-0.04, 0.02)	0.025 (0.009)	(0.01, 0.04)
PB, $\delta = 1$	-0.011 (0.015)	(-0.04, 0.02)	0.025 (0.009)	(0.01, 0.04)
	Si		S	
Naïve	0.285 (0.068)	(0.15, 0.42)	0.057 (0.018)	(0.02, 0.09)
Parametric	0.287 (0.103)	(0.09, 0.49)	0.056 (0.025)	(0.01, 0.11)
PB, $\delta = 0$	0.285 (0.092)	(0.11, 0.47)	0.057 (0.025)	(0.01, 0.11)
PB, $\delta = 1$	0.284 (0.093)	(0.10, 0.47)	0.057 (0.025)	(0.01, 0.11)

Table 2.5: Point estimates (standard errors) for the different pollutants, using naïve analysis and with parametric and parameter bootstrap correction for measurement error in covariate of interest. "PB" refers to results from parameter bootstrap implemented with given value of δ. For the parametric bootstrap and parameter bootstrap with $\delta = 1$, $\hat{\beta}$ refers to the estimate corrected for any bias from classical-like error.



(a) EC

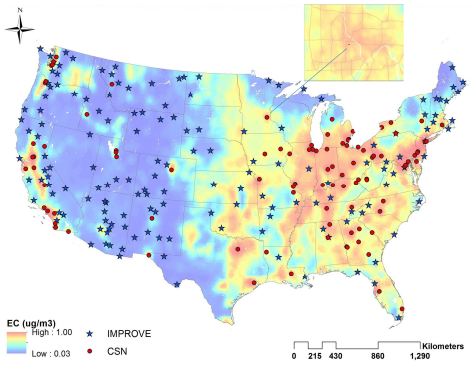
(b) OC



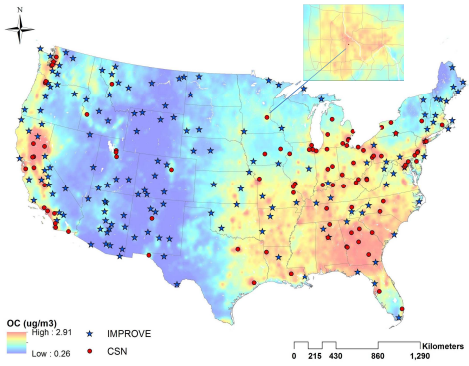
(c) Si

(d) S

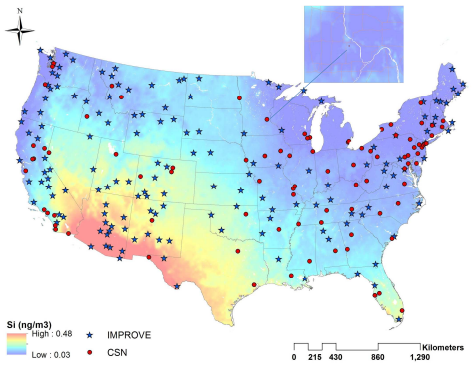
Figure 2.1: Coefficients of the PLS fit, by geographic covariate type. The size of each circle represents the buffer size, with larger circles indicating larger buffers. Explanation of variable abbreviations are given in Table 2.2.



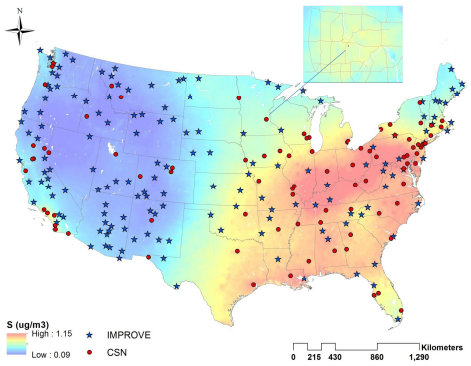
(a) EC



(b) OC



(c) Si



(d) S

Figure 2.2: Locations of IMPROVE and CSN monitors and predicted national average $PM_{2.5}$ component concentrations from final predictions models. Predictions are also shown for St. Paul, MN.

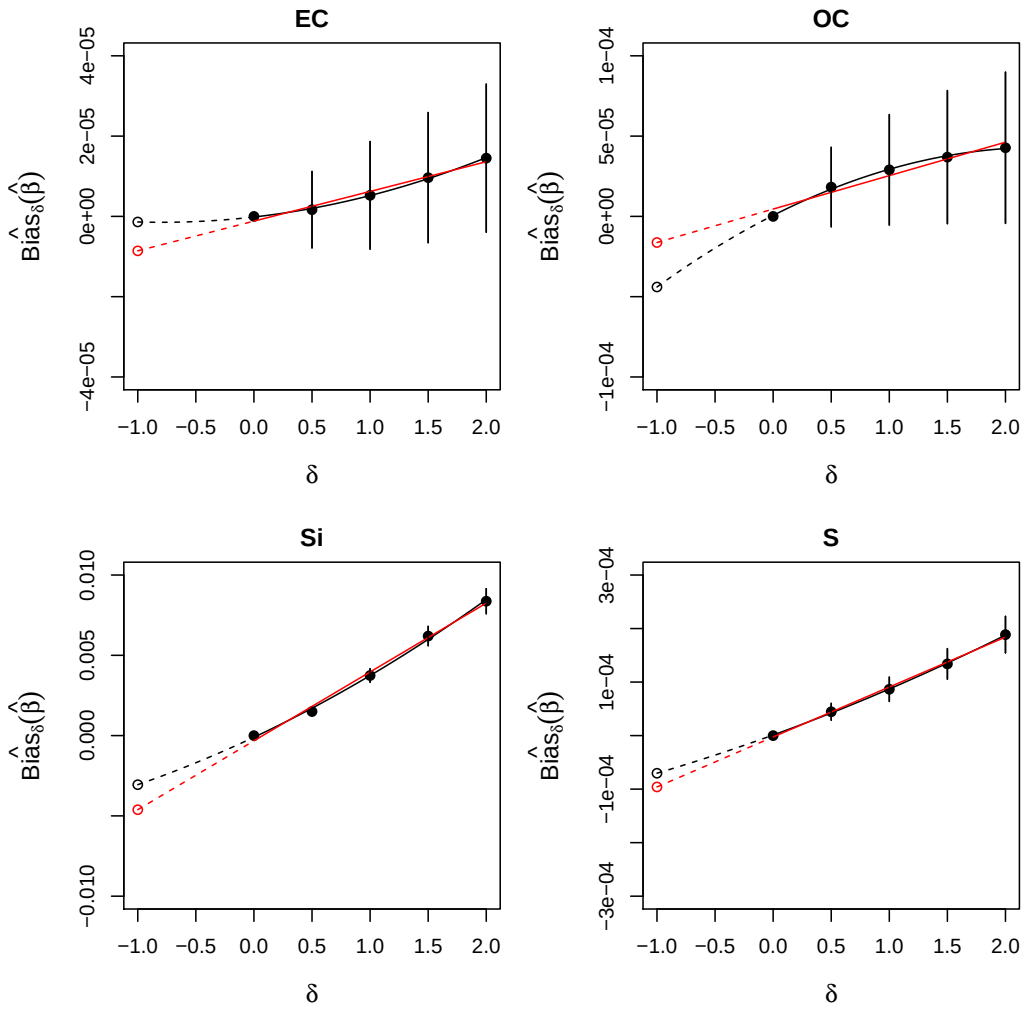


Figure 2.3: P-SIMEX bias estimates. 15,000 parameter bootstrap samples were drawn using values of $\delta \in \{0, 0.5, 1, 1.5, 2\}$, and the estimated bootstrap biases plotted as a function of these values. A linear or quadratic extrapolation was used to estimate $E_{-1}(\hat{\beta}_X^B)$. Confidence intervals from a t-test testing zero bias are also shown.

Chapter 3

MINIMIZING THE IMPACT OF MEASUREMENT ERROR WHEN USING PENALIZED REGRESSION TO MODEL EXPOSURE IN TWO-STAGE AIR POLLUTION EPIDEMIOLOGY STUDIES

3.1 Summary

The previous chapter used universal kriging to model the first-stage exposure model, which is commonly motivated by assuming the exposure surface is a spatially correlated random effect. In this chapter we motivate viewing the exposure surface as fixed and the subject and monitoring locations as random. Corresponding measurement error methods exist only when modeling exposure with simple, low-rank, unpenalized regression splines. We develop a comprehensive treatment of measurement error when modeling exposure with high-but-fixed-rank penalized regression splines. If sufficiently rich, these models well-approximate full-rank splines while remaining asymptotically tractable, and require a penalty parameter to ensure model regularity. We describe the implications of this penalty for measurement error, motivate choosing the penalty to optimize health effect inference, derive an analytic bias correction, and provide a simple non-parametric bootstrap to account for all sources of variability. We find that highly parameterizing the exposure model results in severely biased and inefficient health effect inference if no penalty is used. Choosing the penalty to mitigate measurement error yields much less bias and better efficiency, and can lead to better confidence interval coverage than other common penalty selection methods. Combining the bias correction with the non-parametric bootstrap yields accurate coverage of nominal 95% confidence intervals.

3.2 Introduction

Air pollution cohort studies relating long-term exposure to a single air pollutant to a health outcome often take a two-stage approach. The true subject-specific exposures are often unobservable and must be predicted from a first-stage exposure model fit to observed monitoring data at misaligned locations. The predicted exposures are then substituted for the true exposures in a subsequent second-stage health model (Miller et al., 2007; Kim et al., 2009). Universal kriging (Cressie, 1993) is a commonly used way of fitting the first-stage exposure model. Is is usually motivated from a parametric perspective, assuming the monitor and subject locations are fixed and the exposures are realizations of a stochastic, spatially correlated Gaussian process. Parametric methods of correcting for measurement error under these assumptions have been described elsewhere (Gryparis et al., 2009; Szpiro et al., 2011b; Bergen et al., 2013; Lopiano et al., 2013).

Szpiro and Paciorek (2013) propose an alternative analytic framework, in which the exposure surface is regarded as fixed but unknown and the subject and monitoring locations are sampled at random. They argue this framework more accurately depicts hypothetical repeated cohort studies, which would involve different subjects and monitor locations but a fixed exposure surface determined by topography, long-term meterologic trends, and other natural and anthropogenic features of the geography. In this framework, Szpiro and Paciorek (2013) develop measurement error correction methods, but their methods are limited to relatively simple first-stage exposure models such as unpenalized low-rank regression splines. Universal kriging is generally preferred for modeling air polluton exposure since it accommodates a richer class of spatial surfaces which can lead to more accurate predictions. Accordingly, there is a need to develop measurement error correction methods for flexible exposure models such as universal kriging, but under Szpiro and Paciorek (2013)'s assumptions of a fixed but unknown exposure surface, rather than the spatial random effect assumptions that traditionally motivate universal kriging. In the fixed exposure surface framework, we can regard universal kriging as a type of full-rank penalized spline with

approximately as many basis functions as monitoring locations (Ruppert et al., 2003).

A comprehensive treatment of measurement error in this setting should provide (i) an optimal method for selecting the penalty parameter (since it is no longer directly tied to an assumed random effect in the data-generating mechanism), (ii) a bias correction for the estimated second-stage health effect parameter, and (iii) a valid standard error estimate that accounts for all of the variability from fitting the first- and second-stage models. Optimal penalty parameter selection and bias correction require asymptotic results, and it is not clear how to characterize the asymptotics of full-rank regression models in our setting. As a pragmatic alternative, we replace full-rank splines with penalized regression splines that have a fixed, finite, but large basis such as thin-plate regression splines (Wood, 2003) or low-rank kriging (Kammann and Wand, 2003). As with universal kriging these models require a penalty on the basis function coefficients in order to ensure model regularity (Ruppert et al., 2003; Hastie et al., 2001), but they are asymptotically tractable due to their fixed-rank nature.

The remainder of this paper is organized as follows. Section 3.3 describes our analytic framework, including the assumed data generating mechanism and a review of how penalized regression splines approximate full-rank exposure models such as universal kriging. Section 3.4 decomposes the measurement error into Berkson-like and classical-like components, discusses how the choice of penalty parameter regulates the balance between these two components, proposes a novel approach to penalty parameter selection, and provides an analytic bias correction and bootstrap standard error estimate that ensure valid health effect inference. Section 3.5 illustrates our methods using simulation studies. Section 3.6 presents the application of our methods to estimating the association between systolic blood pressure and $PM_{2.5}$ in the NIEHS Sister Study. Section 3.7 summarizes our findings and discusses limitations and considerations for future work.

3.3 Analytic framework

3.3.1 Data generating mechanisms

Our notation generally follows that used by Szpiro and Paciorek (2013). Let $\mathbf{s}_1, ...\mathbf{s}_n$ denote i.i.d. subject locations and $\mathbf{s}_1^*, ..., \mathbf{s}_{n^*}^*$ i.i.d. monitoring locations in R^2, both drawn independently of each other from an unknown density $G(\cdot)$. We comment on this assumption in Section 3.7. At monitoring location $\mathbf{s}_i^*$ the true, observed exposure is

$$x_i^* = \Phi(\mathbf{s}_i^*) + \eta_i^*,$$

and similarly we define true (but unobserved) exposure $x_j = \Phi(\mathbf{s}_j) + \eta_j$ at subject j's location. Here $\Phi(\cdot) : \mathbb{R}^2 \to \mathbb{R}$ is a fixed function of space that denotes all aspects of the underlying exposure surface potentially explainable with spatially referenced covariates, including geographic or land-use features such as distance to traffic or populaton density and basis functions for a spatial smoother such as kriging or thin-plate splines (described in more detail below). The $\eta_1, ..., \eta_{n^*}$ are i.i.d random error terms with mean 0 and are uncorrelated with the $\mathbf{s}_i^*$; similarly the η_j are i.i.d. and uncorrelated with the $\mathbf{s}_i$ or the η_i^*.

Let $\mathbf{r}(\mathbf{s}_i^*)^T \equiv \{\mathbf{p}(\mathbf{s}_i^*)^T, \mathbf{q}(\mathbf{s}_i^*)^T\}$ map 2-dimensional geographic space to a $(p+q)$-dimensional covariate vector. Here $\mathbf{p}(\mathbf{s}_i^*)$ denotes a vector of p geographic covariates, while $\mathbf{q}(\mathbf{s}_i^*)$ denotes a vector of q spatial basis functions. As opposed to Szpiro and Paciorek (2013), the size of q may be such that $p + q$ is large relative to n^*, too large to comfortably estimate their coefficients using ordinary least squares. Let $\mathbf{R}(\mathbf{s}^*) \equiv \{\mathbf{P}(\mathbf{s}^*), \mathbf{Q}(\mathbf{s}^*)\}$ be the $n^* \times (p + q)$ model matrix created by stacking $\mathbf{r}(\mathbf{s}_1^*)^T, ..., \mathbf{r}(\mathbf{s}_{n^*}^*)$. Analogously, we define $\mathbf{r}(\mathbf{s}_i)$ and $\mathbf{R}(\mathbf{s})$ at subject locations.

If we observed the true exposures x_i at subject locations, we would use them to estimate associations with a continuous health outcome y_i. We assume the health model follows

$$y_i = \beta_0 + \beta x_i + \boldsymbol{\beta}_{\mathbf{z}}^T \mathbf{z}_i + \epsilon_i, \tag{3.1}$$

for $i = 1, ..., n$. Here β is the health effect of interest, $\mathbf{z}_i$ denotes a vector of additional subject-specific covariates such as age or socio-economic status that are potential confounders or precision variables, and the ϵ_i are independent but not necessarily identically distributed random variables with mean zero, independent of the x_i and $\mathbf{z}_i$. We assume that the subject-specific covariates satisfy

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

where $\Theta(\mathbf{s}_i)$ is a vector-valued function representing the spatial component of the subject-specific covariates, and the $\boldsymbol{\zeta}_i$ are random vectors independent between subjects and independent of η_i, with mean zero, but with individual elements not necessarily independent of each other.

A typical data set in this context consists of x_i^* and $\mathbf{R}(\mathbf{s}^*)$ at monitoring locations (used to build exposure models), $\mathbf{R}(\mathbf{s})$ at subject locations (used to predict exposures), and y_i and $\mathbf{z}_i$ at subject locations (used to estimate health effects, given exposure predictions).

3.3.2 Penalized regression exposure model

General form

The first stage in the two-stage approach is building the exposure model using observations x_i^* and modeling covariates $\mathbf{R}(\mathbf{s}^*)$. We want to include enough spatial basis functions in $\mathbf{R}(\mathbf{s}^*)$ to be able to flexibly estimate $\Phi(\cdot)$ of any form while not over-fitting the exposure model given available monitoring data. To this end we regularize some or all of the exposure model coefficients by means of an L_2-type penalty. For penalty parameter $\lambda \geq 0$, define

$$\hat{\boldsymbol{\gamma}}_\lambda = argmin_\theta \frac{1}{n^*} \sum_{i=1}^{n^*} \left(x_i^* - \mathbf{r}(\mathbf{s}_i^*)^T\theta\right)^2 + \lambda\theta^T\mathbf{D}\theta. \tag{3.2}$$

Here $\mathbf{D}$ is a $(p+q) \times (p+q)$ positive semi-definite matrix where often only the bottom-right $q \times q$ block is nonzero, implying we only penalize the coefficients of $\mathbf{q}(\mathbf{s}_i^*)$. Low-rank

kriging and thin-plate regression splines are specific forms of (3.2), and are described below. It is easy to see that $\hat{\boldsymbol{\gamma}}_\lambda = (\mathbf{R}(\mathbf{s}^*)^T\mathbf{R}(\mathbf{s}^*) + n^*\lambda\mathbf{D})^{-1}\mathbf{R}(\mathbf{s}^*)\mathbf{x}^*$. If $\lambda = 0$, then $\hat{\boldsymbol{\gamma}}_\lambda$ is just the ordinary least squares (OLS) fit to the data. As $\lambda \to \infty$, the coefficients of $\mathbf{q}(\mathbf{s}_i^*)$ are shrunk towards 0. This generally decreases the variability of the coefficients (and hence predictions), but at the cost of prediction bias compared to the OLS solution. Thus λ balances fidelity to the observed data and model flexibility with coefficient size and model regularity. Note that as $n^* \to \infty$, $\hat{\boldsymbol{\gamma}}_\lambda \to \boldsymbol{\gamma}_\lambda$, where

$$\boldsymbol{\gamma}_\lambda = argmin_\theta \int (\Phi(\mathbf{s}) - \mathbf{r}(\mathbf{s})^T\theta)^2 dG(\mathbf{s}) + \lambda\theta^T\mathbf{D}\theta. \tag{3.3}$$

That is, $\boldsymbol{\gamma}_\lambda$ are the penalized regression coefficients given infinite monitoring data, for fixed λ.

Let $\hat{w}_\lambda(\mathbf{s}_i) = \mathbf{r}(\mathbf{s}_i)^T\hat{\boldsymbol{\gamma}}_\lambda$ denote predictions at subject locations, and let $w_\lambda(\mathbf{s}_i) = \mathbf{r}(\mathbf{s}_i)^T\boldsymbol{\gamma}_\lambda$ denote predictions at subject locations if we had infinite monitoring data to fit the exposure model.

Low-rank kriging

To motivate low-rank kriging (LRK) as described by Kammann and Wand (2003), we begin by summarizing the usual spatial random effect context of universal kriging (Cressie, 1993). Then we review how kriging can alternatively be viewed as a penalized full-rank regression on spatial basis functions which can be approximated by a penalized regression on a fixed number of basis functions.

Conventionally, universal kriging describes a data generating mechanism under which exposures are realizations of a spatially correlated random effect. Marginally, the exposures at monitoring locations are assumed to follow:

$$\mathbf{x}^* = \mathbf{P}(\mathbf{s}^*)\boldsymbol{\alpha} + \tilde{\boldsymbol{\delta}} + \boldsymbol{\nu}. \tag{3.4}$$

Here $\tilde{\boldsymbol{\delta}} \sim N(0, \sigma^2\Sigma)$; $\boldsymbol{\nu} \sim N(0, \tau^2\mathbf{I})$; $\Sigma_{ij} = C(||\mathbf{s}_i^* - \mathbf{s}_j^*||)$; and $C(r) = exp(-|r|/\phi)$ if using an exponential model, for range parameter ϕ. The parameters τ^2 and σ^2 are often referred to as the nugget and partial sill, respectively. This can be reparameterized as follows:

$$\mathbf{x}^* = \mathbf{P}(\mathbf{s}^*)\boldsymbol{\alpha} + \mathbf{Q}(\mathbf{s}^*)\boldsymbol{\delta} + \boldsymbol{\nu}, \tag{3.5}$$

where $\mathbf{Q}(\mathbf{s}^*) = \Sigma$ and $\boldsymbol{\delta} \sim N(0, \sigma^2\Sigma^{-1})$. Given a choice of ϕ, $\mathbf{Q}(\mathbf{s}^*)$ can now be viewed as a set of radial basis functions with correlated random coefficients $\boldsymbol{\delta}$. Assuming for now that τ^2 and σ^2 are known, minimizing the penalized least squares criterion

$$\frac{1}{\tau^2}\left\{||\mathbf{x}^* - \mathbf{P}(\mathbf{s}^*)\boldsymbol{\alpha} - \mathbf{Q}(\mathbf{s}^*)\boldsymbol{\delta}||^2 + \tilde{\lambda}\boldsymbol{\delta}^T\Sigma\boldsymbol{\delta}\right\} \tag{3.6}$$

with $\tilde{\lambda} = \tau^2/\sigma^2$ yields the maximum likelihood estimate (MLE) of $\boldsymbol{\alpha}$ and the best linear unbiased predictor (BLUP) of $\boldsymbol{\delta}$. Restricted maximum likelihood (REML) is a standard approach to obtaining estimates $\hat{\tau}^2$ and $\hat{\sigma}^2$, and hence a corresponding choice of $\tilde{\lambda}$. The range parameter ϕ is typically also estimated using REML, although for practical reasons this parameter is selected in advance in the low-rank version of kriging. We note that minimizing (3.6) with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$ is equivalent to solving (3.2) with $\hat{\boldsymbol{\gamma}}_\lambda^T = \left\{\hat{\boldsymbol{\alpha}}^T, \hat{\boldsymbol{\delta}}^T\right\}$, $\lambda = \tilde{\lambda}/n^*$, and $\mathbf{D} = \begin{pmatrix} 0_{p\times p} & 0_{p\times q} \\ 0_{q\times p} & \Sigma \end{pmatrix}$.

We emphasize that, in our paradigm, Equations (3.4) and (3.5) do not reflect the true data generating mechanism. Rather, (3.5) provides a starting point from which to derive a penalized regression model and motivates a way of choosing λ by translating between the variance components in the mixed model formulation and λ in the penalized regression formulation (Hodges, 2013; Wakefield, 2013). From here we can see that if $\hat{\sigma}^2$ is large relative to $\hat{\tau}^2$, a greater proportion of the estimated variability in x_i^* is due to spatial structure than random noise, suggesting that a smaller penalty on $\hat{\boldsymbol{\delta}}$ is appropriate. On the other hand if $\hat{\tau}^2$ is large relative to $\hat{\sigma}^2$, the amount of spatial information is small relative to the variability

from noise, indicating that $\hat{\boldsymbol{\delta}}$ should be shrunk towards zero.

Kammann and Wand (2003) motivate using LRK to approximate full-rank kriging by fixing the dimension of $\boldsymbol{\delta}$. This feature of LRK facilitates our asymptotic analysis and allows us to derive estimates of bias from measurement error. Given a set of spatial knots $\boldsymbol{\kappa}_1, ..., \boldsymbol{\kappa}_q$ with fixed $q < n^*$, LRK radial basis functions $\mathbf{q}(\mathbf{s}_i^*)$ are defined as $C(||\mathbf{s}_i^* - \boldsymbol{\kappa}_j||)$ for $j = 1, ..., q$. Let $\boldsymbol{\Omega}$ denote a $q \times q$ matrix such that $\boldsymbol{\Omega}_{ij} = C(||\boldsymbol{\kappa}_i - \boldsymbol{\kappa}_j||)$. Then LRK coefficients are found by minimizing

$$||\mathbf{x}^* - \mathbf{P}(\mathbf{s}^*)\boldsymbol{\alpha} - \mathbf{Q}(\mathbf{s}^*)\boldsymbol{\delta}||^2 + \tilde{\lambda}\boldsymbol{\delta}^T\boldsymbol{\Omega}\boldsymbol{\delta}, \tag{3.7}$$

which (as in the full-rank version) is equivalent to finding the MLE and BLUP for a mixed effects model with $\boldsymbol{\delta} \sim N(0, \sigma^2\boldsymbol{\Omega}^{-1})$. Kammann and Wand (2003) recommend using REML to obtain estimates $\hat{\tau}^2$ and $\hat{\sigma}^2$ and fixing the range parameter ϕ at approximately the maximum distance between observations. The length of $\boldsymbol{\delta}$ is now $q < n^*$, and again minimizing (3.7) is equivalent to minimizing (3.2) with $\hat{\boldsymbol{\gamma}}_\lambda^T = \left\{\hat{\boldsymbol{\alpha}}^T, \hat{\boldsymbol{\delta}}^T\right\}$, $\lambda = \tilde{\lambda}/n^*$ and $\mathbf{D} = \begin{pmatrix} 0_{p\times p} & 0_{p\times q} \\ 0_{q\times p} & \boldsymbol{\Omega} \end{pmatrix}$.

We use the term "low-rank kriging" to remain consistent with Kammann and Wand (2003), but emphasize that the size of q, while fixed and less than n^*, is not necessarily "low". Indeed "fixed-rank kriging" may be more appropriate nomenclature but has already been used by Cressie and Johannesson (2008) to refer to a different model than what we have described.

Thin-plate regression splines

One disadvantage of LRK is the need to specify a range parameter ϕ. Full-rank thin-plate splines do not require specification of a range parameter, and can be approximated by fixed-rank thin-plate regression splines (TPRS). A complete derivation can be found elsewhere (Green and Silverman, 1994; Wood, 2003), which we briefly summarize here.

Let $\mathbf{E}$ be an $n^* \times n^*$ matrix with $\mathbf{E}_{ij} = C(||\mathbf{s}_i^* - \mathbf{s}_j^*||)$, where $C(r) = (8\pi)^{-1}r^2log(r)$. Let

$\mathbf{T}$ be the $n^* \times 3$ matrix with rows $\{\xi_1(\mathbf{s}_i^*), \xi_2(\mathbf{s}_i^*), \xi_3(\mathbf{s}_i^*)\}$; here the ξ_j are linearly independent polynomials that span the space of linear polynomials in $\mathbb{R}^2$. Full-rank thin-plate splines minimize

$$\frac{1}{n^*}||\mathbf{x}^* - \mathbf{P}(\mathbf{s}^*)\boldsymbol{\alpha} - \mathbf{T}\boldsymbol{\xi} - \mathbf{E}\boldsymbol{\delta}||^2 + \lambda\boldsymbol{\delta}^T\mathbf{E}\boldsymbol{\delta}, \tag{3.8}$$

subject to the constraint $\mathbf{T}^T\boldsymbol{\delta} = 0$. It is clear that the estimated coefficients will be full-rank, since $\boldsymbol{\delta}$ has length n^*. Wood (2003) presents a fixed-rank approximation that is derived from the singular value decomposition (SVD) of $\mathbf{E}$.

Let $\mathbf{USU}^T$ be the SVD of $\mathbf{E}$. Let $\mathbf{U}_q$ denote the first q columns of $\mathbf{U}$, $\mathbf{S}_q$ the first q rows and colums of $\mathbf{S}$, and $\mathbf{W}_q$ a $q \times (q-3)$ matrix such that $\mathbf{T}^T\mathbf{U}_q\mathbf{W}_q = \mathbf{0}$. TPRS coefficients are found by minimizing

$$\frac{1}{n^*}||\mathbf{x}^* - \mathbf{P}(\mathbf{s}^*)^T\boldsymbol{\alpha} - \mathbf{T}\boldsymbol{\xi} - (\mathbf{U}_q\mathbf{S}_q\mathbf{W}_q)\boldsymbol{\delta}||^2 + \lambda\boldsymbol{\delta}^T\boldsymbol{\Omega}\boldsymbol{\delta}, \tag{3.9}$$

where $\boldsymbol{\Omega} = \mathbf{W}_q^T\mathbf{S}_q\mathbf{W}_q$. From here it is apparent that with $\hat{\boldsymbol{\gamma}}_\lambda^T = \left\{\hat{\boldsymbol{\alpha}}^T, \hat{\boldsymbol{\xi}}^T, \hat{\boldsymbol{\delta}}^T\right\}$, $\mathbf{Q}(\mathbf{s}^*) = \{\mathbf{T}, \mathbf{U}_q\mathbf{S}_q\mathbf{W}_q\}$, and $\mathbf{D}$ the $(p+q)\times(p+q)$ matrix with $\boldsymbol{\Omega}$ in the lower $(q-3)\times(q-3)$ block and zeros elsewhere, minimizing (3.9) is equivalent to minimizing (3.2). Further, minimizing (3.9) yields the MLE and BLUP for a mixed effects model with $\boldsymbol{\delta} \sim N(0, \sigma^2\boldsymbol{\Omega}^{-1})$. As in LRK, we may use REML to estimate the variance parameters from the mixed effects model and hence obtain an estimate of λ. A more standard approach for TPRS is to bypass the mixed model formulation and directly select λ by generalized cross-validation (GCV) Wood (2003).

3.4 Measurement error

Once we have found exposure model parameter estimates $\hat{\boldsymbol{\gamma}}_\lambda$, we can define the measurement error that results from using $\hat{w}_\lambda(\mathbf{s}_i)$ in place of x_i to estimate β. Similarly to Szpiro and Paciorek (2013), we decompose the measurement error into Berkson-like and classical-like components as follows

$$\begin{aligned} x_i - \hat{w}_\lambda(\mathbf{s}_i) &= (x_i - w_\lambda(\mathbf{s}_i)) + (w_\lambda(\mathbf{s}_i) - \hat{w}_\lambda(\mathbf{s}_i)) \\ &= u_\lambda^B(\mathbf{s}_i) + u_\lambda^C(\mathbf{s}_i). \end{aligned}$$

The Berkson-like component u_λ^B results from smoothing the true exposure surface using $w_\lambda(\mathbf{s}_i)$. In other words, it is error that is still present even if we had infinite monitoring data available to fit the exposure model. It bears similarity to pure Berkson error because using $w_\lambda(\mathbf{s}_i)$ instead of x_i misses exposure surface characteristics resulting in a less variable predicted surface, but differs from Berkson error since it can induce bias in the health effect in addition to affecting its standard error. This is in contrast to previous results where the exposure model is either estimated without a penalty or is assumed to follow a correctly specified spatial random effect model Szpiro and Paciorek (2013); Szpiro et al. (2011b) and Berkson-like error does not leas to bias. Heuristically, larger λ induces more Berkson-like error as a result of a smoother predicted surface.

The classical-like component u_λ^C results from estimating $\boldsymbol{\gamma}_\lambda$ with finite monitoring data. It induces variability in the predicted exposures that is not related to the health outcome, and can accordingly bias the estimated health effect and affect its standard error. In this way it is similar to pure classical error, but it differs since the bias from u_λ^C diminishes as $n^* \to \infty$. Heuristically, there is more classical-like error for small λ, since there is more variability in the exposure model coefficients.

Although in practice β is estimated with finite n and n^*, in what follows we isolate the effects of measurement error by considering estimation of β with infinite n. Let $\hat{\beta}_{n^*}$ denote an estimate of β using infinite subject data and finite monitoring data, while $\hat{\beta}_{n,n^*}$ denotes a practical estimate of β using finite subject and monitoring data.

3.4.1 Analyzing behavior of $\hat{\beta}_{n^*}$

We use a Taylor expansion centered around $\boldsymbol{\gamma}_\lambda$ to derive an expression for the bias induced by u_λ^B and u_λ^C. For our present purposes we discuss bias heuristically in Lemma 1, deferring a more rigorous asymptotic treatment Shao (2003) to Appendix B.

Lemma 1: Let $\mathbf{r}^\perp(\mathbf{s})$ contain elements $(r_k(\mathbf{s}) - \Theta(\mathbf{s})^T\varphi_k)$, where $\varphi_k = \text{argmin}_\omega \int (r_k(\mathbf{s}) - \Theta(\mathbf{s})^T\omega)^2 dG(\mathbf{s})$ for $k \in \{1, ..., p+q\}$. Let $w_\lambda^\perp(\mathbf{s}_i) = \mathbf{r}^\perp(\mathbf{s}_i)^T\boldsymbol{\gamma}_\lambda$ and $\hat{w}_\lambda^\perp(\mathbf{s}_i) = \mathbf{r}^\perp(\mathbf{s}_i)^T\hat{\boldsymbol{\gamma}}_\lambda$. Then, if we define $f(\hat{\boldsymbol{\gamma}}_\lambda) = \frac{\int w_\lambda^\perp(\mathbf{s})\hat{w}_\lambda^\perp(\mathbf{s})dG(\mathbf{s})}{\int \hat{w}_\lambda^\perp(\mathbf{s})^2 dG(\mathbf{s})} + \frac{\int u_\lambda^B(\mathbf{s})\hat{w}_\lambda^\perp(\mathbf{s})dG(\mathbf{s})}{\int \hat{w}_\lambda^\perp(\mathbf{s})^2 dG(\mathbf{s})}$, we can show that $\hat{\beta}_{n^*} = \beta f(\hat{\boldsymbol{\gamma}}_\lambda)$. Let $\mathbf{h}_\lambda$ and $\mathbf{H}_\lambda$ denote the gradient and the Hessian matrix, respectively, of $f(\hat{\boldsymbol{\gamma}}_\lambda)$ evaluated at $\boldsymbol{\gamma}_\lambda$.

(a) The bias attributable to Berkson-like and classical-like error is

$$E\left(\frac{\hat{\beta}_{n^*} - \beta}{\beta}\right) = \psi_\lambda^B + \psi_\lambda^C, \tag{3.10}$$

where $\psi_\lambda^B = \dfrac{\int u_\lambda^B(\mathbf{s}) w_\lambda^\perp(\mathbf{s}) dG(\mathbf{s})}{\int w_\lambda^\perp(\mathbf{s})^2 dG(\mathbf{s})}$ is bias from Berkson-like error, and
$\psi_\lambda^C = \frac{1}{n^*}\left\{\mathbf{h}_\lambda^T E(\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda) + tr\left(\mathbf{H}_\lambda Cov(\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda)\right)\right\}$ is bias from classical-like error.

(b) The variance attributable to classical-like error is

$$Var\left(\frac{\hat{\beta}_{n^*} - \beta}{\beta}\right) = \frac{1}{n^*}\left(\mathbf{h}_\lambda^T Cov(\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda)\mathbf{h}_\lambda\right). \tag{3.11}$$

See Appendix B for rigorous statement and proof of Lemma 1, and for details on estimating ψ_λ^B and ψ_λ^C with available data.

Part (a) of Lemma 1 characterizes the bias of $\hat{\beta}_{n^*}$ as a function of both Berkson- and classical-like error. If either of the following conditions are satisfied we can guarantee unbiased estimation of β even under "optimal circumstances" where we have as much flexibility

as possible in fitting our exposure model ($\lambda = 0$) and infinite monitoring data ($n^* = \infty$). In that case we can interpret the bias as coming from $\lambda > 0$ and/or $n^* < \infty$.

Condition 1: Enough spatial basis functions have been included in the exposure model such that for all $\mathbf{s}_i^*$, $\Phi(\mathbf{s}_i^*) = \mathbf{r}(\mathbf{s}_i^*)^T\boldsymbol{\gamma}$ for some $\boldsymbol{\gamma}$.

Condition 2: The elements of $\Theta(\mathbf{s})$ are contained in the span of $\mathbf{r}(\mathbf{s})$.

Another way to think of Condition 1 is that enough basis functions have been included to correctly specify a $\Phi(\mathbf{s})$ of unknown form, though penalization might be necessary to avoid overfitting with finite n^* Yu and Ruppert (2002).

To see the implication of Conditions 1 and 2, we re-write the health model as

$$y_i = \beta_0 + \beta w_0(\mathbf{s}_i) + \boldsymbol{\beta}^T\mathbf{z}_i + (\epsilon_i + \beta(x_i - w_0(\mathbf{s}_i))\,,$$

and note that $u_0^B(\mathbf{s}_i)$ becomes part of the residual in the health model. If Condition 1 is met, it follows that $\boldsymbol{\gamma}_0 = \boldsymbol{\gamma}$ and $(\Phi(\mathbf{s}_i) - \mathbf{r}(\mathbf{s}_i)^T\boldsymbol{\gamma}_0) = 0$ for all $\mathbf{s}_i$, implying $u_0^B(\mathbf{s}_i) = \eta_i$. In this case, $u_0^B(\mathbf{s}_i)$ is pure Berkson error and does not induce bias in a linear health model Carroll (2006). If Condition 2 is met, $u_0^B(\mathbf{s}_i)$ is uncorrelated with both $w_0(\mathbf{s}_i)$ and $\mathbf{z}_i$, ensuring no bias is induced by using $w_0(\mathbf{s}_i)$ in place of x_i Szpiro and Paciorek (2013); White (1980). Although our bias expression from (a) is still valid even if both conditions are violated, it is helpful to operate under the assumption that one or both conditions are met so as to emphasize the roles of λ and n^* in determining bias.

Assuming either Condition 1 or 2 is met, from (a) we can see that if $\lambda = 0$ ψ_0^B is zero since $w_0^{\perp}(\mathbf{s})$ is orthogonal to u_0^B, but if $\lambda > 0$ this orthogonality is not guaranteed implying $|\psi_\lambda^B|$ could be large. On the other hand, when $\lambda = 0$ $Cov(\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda)$ is maximized implying ψ_0^C is large, and decreases as λ increases. Thus λ trades-off bias from the two error types. Also note that as $n^* \to \infty$, the bias from classical-like error goes to zero at a rate of $1/n^*$ while the bias from Berkson-like error is independent of n^*.

Note that Lemma 1 was derived assuming infinite n. The bias estimate in (3.10) remains valid even for finite n since the size of the health data set does not affect bias, but (3.11) is not a valid finite-n variance estimate as it accounts only for variability from classical-like error. This can be readily seen by observing that if $Cov(\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda) = 0$ the variance in (3.11) is zero. This does not account for finite-n sources of variability in $\hat{\beta}_{n,n^*}$ such as ϵ_i and $u_\lambda^B(\mathbf{s}_i)$. Thus (3.11) is useful only as a tool for λ selection, and to get valid finite-n variance estimates we must use methods such as the non-parametric bootstrap Efron and Tibshirani (1993); Szpiro and Paciorek (2013).

3.4.2 Accounting for measurement error

Selecting the penalty parameter

In Sections 3.3.2 and 3.3.2 we motivated using REML or GCV to select λ. These are motivated purely from the exposure modeling framework, and do not consider that the predictions will be used in health modeling. Lemma 1 suggests an alternative approach. Combining Equations (3.10) and (3.11) yields an expression for the MSE of $(\hat{\beta}_{n^*} - \beta)/\beta$ that is a function of λ. Choosing λ to minimize this MSE aims to minimize bias from both error types while paying heed to variance from classical-like error. We refer to this as the MSE method of choosing λ, noting that we can use this criteria without estimating β.

Bias correction and standard error estimation

Regardless of the method used to select λ, the resulting health effect estimate may still be biased. Once we choose λ and estimate ψ_λ^B and ψ_λ^C, we define a bias-corrected health effect estimate by $\hat{\beta}_{n,n^*}^C = \hat{\beta}_{n,n^*} / \left(1 + \hat{\psi}_\lambda^B + \hat{\psi}_\lambda^C\right)$. We use the nonparametric bootstrap to obtain valid standard error estimates that account for all sources of variability, including the bias correction if applied Szpiro and Paciorek (2013). In each bootstrap sample, n subject locations and n^* monitoring locations are randomly drawn with replacement. We fit the exposure model to the re-sampled monitoring data, predict at the re-sampled subject

locations, and estimate the health effect. If we are estimating the standard error of $\hat{\beta}^{C}_{n,n^*}$ we also estimate and apply a bias correction. We use the appropriate empirical standard deviations of the bootstrap estimated health effects to estimate the standard errors of $\hat{\beta}_{n,n^*}$ and $\hat{\beta}^{C}_{n,n^*}$.

Sensitivity analysis to verify sufficient treatment of measurement error

In practice it is advisable to conduct sensitivity analyses with different exposure models Peng (2013), and any two exposure models are sure to give non-identical health effect estimates even after applying a bias correction. One must then decide if one estimate remains more biased than the other, or if the differences are due to variability from measurement error. We can assess this by using a modified non-parametric bootstrap for each exposure model, keeping subjects fixed and re-sampling only the monitoring data. The resulting health effect variability is purely due to measurement error from using different exposure models. Comparing the two effect estimates in conjunction with this variability puts into perspective whether discrepancies suggest residual bias or are simply a result of variability from measurement error. This modification to the full bootstrap (which re-samples subjects) is critical, as in practice the same study subjects are used when estimating health effects based on different exposure models.

3.5 Simulations

3.5.1 Description of simulation studies

We performed simulations to assess the impact of measurement error from penalized regression exposure models. Subject and monitoring locations were simulated independently and uniformly on a 4500×4500 grid. Given location, the exposure surface was defined as

$$\Phi(\mathbf{s}) = \Phi_s(\mathbf{s}) + \sum_{k=1}^{6} r_k(\mathbf{s}). \tag{3.12}$$

Each geographic covariate $r_k(\mathbf{s})$ was a realization of a $N(0, \sqrt{0.25})$ random variable, independent across locations. $\Phi_s(\mathbf{s})$ was a fixed realization from a simulated spectral approximation to a Gaussian process with variance 6 and range parameter 1500 Paciorek (2007). Thus the total variance of $\Phi(\mathbf{s})$ was 7.5, with 80% of this variability due to spatial structure. True exposures at both subject and monitoring locations were simulated as $x = \Phi(\mathbf{s}) + \eta$ with $\eta \sim N(0, \sqrt{3})$, implying approximately 70% of the true exposure variance was potentially explainable by covariates and spatial basis functions.

Given true exposures at subject locations, we considered two different data-generating mechanisms for the health outcome. In Scenario 1, there were no additional covariates, and health outcomes were generated according to

$$y_i = \beta_0 + \beta x_i + \epsilon, \tag{3.13}$$

with $\beta_0 = 74$, $\beta = 0.1$, and $\epsilon \sim N(0, 1)$. In Scenario 2, the health outcomes were generated according to

$$y_i = \beta_0 + \beta x_i + \boldsymbol{\beta}_Z^T \mathbf{z}_i + \epsilon, \tag{3.14}$$

again with $\beta_0 = 74$, $\beta = 0.1$, and $\epsilon \sim N(0, 1)$. Here $\mathbf{z}_i$ was an 8 degree-of-freedom thin-plate regression spline (Wood 2003), and the elements of $\boldsymbol{\beta}_Z$ were all equal to 1.

For our simulations we used $n^* = 100$ and $n = 1000$. We considered LRK and TPRS exposure models, including all six geographic covariates in each. The LRK models used a range parameter $\phi = 6363$, the maximum distance between grid locations. The fixed LRK knot locations were chosen with a space-filling algorithm. We considered $q = 10, 15, 20$ or 25 for both exposure models, implying the maximum value of $p + q$ was 32 including basis functions, geographic covariates and an intercept.

We studied behavior of $\hat{\beta}_{n,n^*}$ for fixed $\lambda \in \{0, 0.01, ..., 0.99, 1, 1.05, ..., 2.95, 3.00\} \times 10^{-2}$. The smaller λ values were similar to those often selected by REML or GCV. The larger λ values were chosen to give a general idea of how the bias of $\hat{\beta}_{n,n^*}$ behaved for uncommonly

large penalties.

In addition to fixing λ we also performed more detailed simulations using exposure models that were built with no penalty ($\lambda = 0$), or with λ chosen using REML, GCV or the MSE criteria. For these simulations we implemented the bootstrap to estimate standard errors and compared confidence interval coverage to those based on naïve sandwich standard errors. We used 1000 Monte Carlo realizations for each simulation scenario, and our bootstrap standard error estimates were based on 100 Monte Carlo bootstrap draws.

3.5.2 Simulation Results: Fixing λ

LRK exposure models

Observed relative biases and standard deviations of $\hat{\beta}_{n,n^*}$ from the LRK exposure models are shown in Figure 3.1. For small λ there was notable bias towards the null which worsened for larger q. This reflects more bias from classical-like error resulting from increased variability of $\hat{\gamma}_\lambda$. For Scenario 1, as λ increased the bias from Berkson-like error canceled the bias from classical-like error, leading to no bias for $\lambda \approx 0.3 \times 10^{-2}$. On the other hand, Scenario 2 had no value of λ that gave zero bias. For both scenarios the standard deviation of $\hat{\beta}_{n,n^*}$ steadily increased with λ. Of note is that for most values of λ the relative standard deviation was comparable in size to the relative bias, indicating that the bias was often non-ignorable.

TPRS exposure models

The simulation results using TPRS exposure models and fixing λ are shown in Figure 3.2, and were somewhat different from the LRK results. For Scenario 1, the bias was not monotone as λ increased, and there were multiple values of λ that led to no bias. Similar to the LRK results, Scenario 2 had no values of λ that eliminated the bias, and the standard deviations in both scenarios increased as λ increased.

3.5.3 Simulation results: Estimating λ

LRK exposure models

Figures 3.3 and 3.4 summarize the results using LRK exposure models with $\lambda = 0$ or selected using GCV, REML, or the MSE criteria. Although not shown, the mean out-of-sample R^2 was between 0.44 and 0.50 for the penalized models, and was as low as 0.36 for $\lambda = 0$ and $q = 25$. As q increased the mean out-of-sample R^2 from the penalized models monotone increased, though the increases were slight. The MSE and GCV criteria led to slightly worse R^2 than the REML criteria.

For Scenario 1, we see in Figure 3.3 that $\hat{\beta}_{n,n^*}$ estimated using predictions from unpenalized exposure models was heavily biased towards the null, with a relative bias greater than 20% for $q = 25$. Penalizing the exposure models drastically reduced the bias of $\hat{\beta}_{n,n^*}$, and using MSE to choose λ induced the least bias. For $\lambda > 0$, applying a bias correction reduced the bias even more regardless of the size of q. If $\lambda = 0$ the bias correction only worked when q was small; for large q the exposure model was over-parameterized which violated the asymptotics underlying the analytic bias expression. For all choices of λ, applying a bias correction increased the health effect standard error (significantly so for the unpenalized models) and sandwich standard error estimates drastically underestimated the true standard errors of both $\hat{\beta}_{n,n^*}$ and $\hat{\beta}^C_{n,n^*}$.

Using GCV and applying a bias correction led to overcoverage for large q, as the bootstrap overestimated the standard error of $\hat{\beta}^C_{n,n^*}$. This was because GCV tended to choose very small values of λ, which for large q led to highly variable bootstrap samples of $\hat{\beta}^C_{n,n^*}$. If REML was used applying a bias correction and using bootstrap standard errors gave accurate coverage. Using MSE led to accurate coverage even without a bias correction; in this case applying a bias correction only led to efficiency loss.

Figure 3.4 shows the results from Scenario 2. As in Scenario 1, we saw drastic relative biases towards the null when $\lambda = 0$ (more than 50% when $q = 25$), failure of the bias correction for $\lambda = 0$ and large q, reduced bias from penalization, further bias reduction if

both a penalty and bias correction were applied, increased standard errors from applying a bias correction, and drastic undercoverage from sandwich standard error estimates. The standard errors of $\hat{\beta}^C_{n,n^*}$ using REML or MSE were very similar to each other, and both were much smaller than GCV, especially as q increased. As opposed to Scenario 1, using the MSE criteria to choose λ was not enough to obtain accurate confidence interval coverage; we needed to apply a bias correction to achieve accurate coverage.

TPRS exposure models

The results from using TPRS exposure models and estimating λ are shown in Figures 3.5 and 3.6. Mean out-of-sample R^2 was between 0.43 and 0.51 for all models, and slightly increased with q for the penalized models. The GCV and REML models led to very similar R^2 while the the MSE criteria yielded worse R^2 than either.

We again see that in Scenario 1 using the MSE criteria alone was sufficient to get accurate confidence interval coverage, but that in Scenario 2 a bias correction was needed. A penalty was needed to ensure the bias correction was effective for all values of q.

3.6 Application

3.6.1 $PM_{2.5}$ and elevated blood pressure in the NIEHS Sister Study

We study the association between systolic blood pressure (SBP) and exposure to particulate matter $\leq 2.5\mu g/m^3$ ($PM_{2.5}$) in the NIEHS Sister Study, a nationwide (including Puerto Rico) prospective cohort study that is primarily focused on the impact of genetic and environmental risk factors on breast cancer in women who are sisters of a woman with breast cancer (NIEHS, 2013). Our research group previously estimated an increase of 1.4 mmHg in SBP per 10-$\mu g/m^3$ in year 2006 annual average $PM_{2.5}$ (95% CI: 0.6, 2.3) (Chan et al., Subm). This result is based on data from the 43,629 Sister Study participants whose residences in the year 2006 could be identified and geo-coded (all such participants resided in the lower 48 states), and exposure predictions from a regionalized universal kriging model

described by Sampson et al. (2013). Our objective is to re-analyze these data and to account for exposure measurement error.

3.6.2 Exposure models

For our re-analysis, we predict exposures at participant locations using both LRK and TPRS models, based on observed year 2006 annual average $PM_{2.5}$ at 1037 EPA air quality system (AQS) monitoring locations across the lower 48 states. The monitoring locations and Sister Study participant locations are shown in Figure 3.7. We considered $\lambda = 0$ and λ chosen using REML, GCV or the MSE criteria. Our geographic covariates were two partial least squares (PLS) components derived from 300 land-use variables (Abdi, 2003; Bergen et al., 2013). We considered $q = 50, 100$, and 150. The LRK knot locations were found using a space-filling algorithm on a grid defined by fixed 25×25 km cells and ϕ was set equal to 6,363 km, the maximum distance between any two grid cells. We also modeled exposure with a universal kriging model fit to the entire nation-wide monitoring network (although our measurement error correction methods are not applicable here). We did this to compare health results to those using exposures from fixed-rank penalized regression splines and the regionalized universal kriging model in Chan et al. (Subm).

3.6.3 Health model

In a multivariate linear health model, we regressed SBP on predicted $PM_{2.5}$ exposure, controlling for the same variables included in the primary model of Chan et al. (Subm). Both sandwich and bootstrap standard errors were calculated. We also applied the sensitivity analysis described in Section 3.4.2 to give perspective as to whether any discrepancies in effect estimates were within measurement error variability or whether we had failed to adequately correct for bias.

3.6.4 Results

10-fold cross-validated R^2 ranged from 0.76–0.79 for the penalized models. For the unpenalized LRK models, over-parameterization led to decreases in predictive accuracy with $R^2 = 0$ when $q = 150$.

Figures 3.8 and 3.9 show the results of the health analyses. The top two rows of each panel show the health results using predictions from regionalized and national universal kriging models, respectively. The other rows correspond to health effects using exposures predicted with fixed-rank exposure models. For these rows, the black circles and dashed lines correspond to a naïve analysis: point estimates with no bias correction and 95% confidence intervals derived from sandwich standard error estimates. The plus-symbol refers to the bias-corrected health effect estimate, and the dashed red lines are confidence intervals derived from the full bootstrap. The solid parts of the red lines are confidence intervals derived from the partial bootstrap described in Section 3.4.2.

The regionalized and national universal kriging models yielded very similar health effect estimates. For all of the fixed-rank exposure models, using only $q = 50$ degrees-of-freedom led to much weaker and non-significant associations of SBP with $PM_{2.5}$ than the universal kriging models. When we used 100 or 150 degrees-of-freedom, we estimated associations between 1.0 and 1.6 mmHg SBP for a 10-$\mu g/m^3$ increase in year 2006 $PM_{2.5}$. These were comparable to the association estimated by Chan et al. (Subm) and by the national universal kriging model. However, when we corrected for measurement error bias only the models penalized via REML or MSE remained stable. Applying a bias correction without penalization or penalizing via GCV led to wildly variable bias-corrected estimates, especially for the LRK exposure models.

For the stable REML and MSE models, there was very little estimated bias. Confidence intervals derived using bootstrap standard errors that accounted for all sources of variability (denoted by the red dashed lines) were very similar to or slightly wider than naïve confidence intervals. The most drastic difference was the $q = 50$ LRK model with λ chosen via MSE,

where the bootstrap standard error was 13% larger than the naïve estimate. Even though the $q = 50$ exposure models yielded qualitatively different results than the $q = 100$ or $q = 150$ models, all the solid red confidence intervals overlapped, indicating that this discrepancy could be attributed to variability from measurement error and bias correction rather than unaccounted-for bias.

3.7 Discussion

We have presented a measurement error framework when approximating full-rank splines with high-but-fixed-rank penalized regression splines to model exposure, shown how the measurement error depends on the penalty parameter λ, and derived an estimate of the bias useful for bias correction. We have developed a novel method for choosing λ to optimize health effect estimation, in contrast to traditional methods that focus on prediction accuracy or fit of the first-stage exposure model.

The ability of our measurement error methodology to admit larger-rank exposure models is an important advance over previously existing methods. Our simulations show that including many basis functions in the exposure model without penalizing their coefficients can induce drastic bias towards the null and heavy efficiency loss from classical-like error, and that the bias correction fails in these settings as the underlying asymptotics are violated. Thus, without methodology for penalized regression models one is restricted to very low-rank exposure models which may miss important exposure surface characteristics. Our bias expression shows that penalizing the exposure model reduces bias from classical-like error but introduces bias from Berkson-like error. In simulations penalizing the exposure model not only reduced the bias of the uncorrected health estimates, but sufficiently regularized the exposure models so that implementing the bias correction and non-parametric bootstrap led to small biases and accurate coverage of 95% confidence intervals.

In our data analysis, at least 100 exposure model degrees-of-freedom were needed to yield associations of SBP with $PM_{2.5}$ that were comparable to previous work that used full-rank splines to model exposure. In order to well-approximate results obtained using full-rank

splines, penalization was necessary to ensure model regularity. Using MSE or REML to choose λ yielded the most stable inference when correcting for bias.

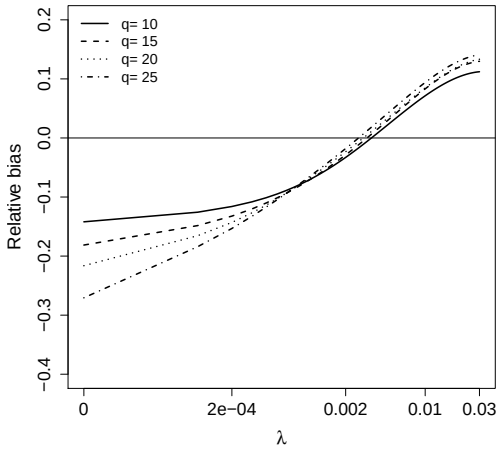
We have also proposed a novel method for selecting λ that balances the Berkson- and classical-like errors. Our simulations indicate that even applying the MSE criteria does not always sufficiently reduce the bias to give accurate confidence interval coverage. In those cases a bias correction is needed to get accurate coverage. Accordingly a bias correction should generally be used for penalized exposure models, since in general it is necessary for accurate coverage. When applying a correction, the REML or MSE criteria were the best methods of choosing λ, as they consistently led to better coverage and efficiency than the GCV criteria.In all simulations the non-parametric bootstrap was needed to account for all sources of variability.

We have motivated choosing λ to minimize the asymptotic MSE of $\hat{\beta}_{n^*}$. We use MSE rather than squared-bias as our criterion because we found that choosing λ to minimize the squared bias performed poorly when there were multiple values of λ that gave a minimum, such as Scenario 1 using TPRS. In these situations focusing only on bias often resulted in loss of efficiency since variance was ignored. The MSE criteria avoided this problem by finding λ that balanced bias from both error types and variance from classical-like error.

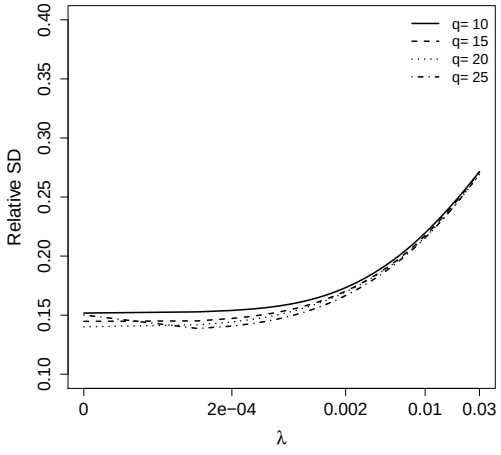
A strong assumption we have made throughout is that the subject and monitoring locations are both drawn from the same distribution G. In practice this is not likely to be the case. Two immediate consequences follow if this assumption is violated and $\mathbf{s}^*$ is drawn from some $H \neq G$. First, if Condition 1 is not met, ψ_0^B (the bias from Berkson-like error when $\lambda = 0$) is not guaranteed to be zero even if Condition 2 is met. This is because $u_0^B(\mathbf{s})$ is no longer guaranteed to be orthogonal to $w_0(\mathbf{s})$ under G, since $\boldsymbol{\gamma}_0$ now minimizes the least-squares criteria with respect to H (see Szpiro and Paciorek (2013) for more details). The second consequence is that we can not estimate the numerator of ψ_λ^B using monitoring data directly, since the empirical distribution of monitoring locations now estimates H whereas the integrals involved in ψ_λ^B are with respect to G. In this case more

care is needed in estimating ψ_λ^B. One plausible solution is to re-weight each monitoring location by an estimate of G/H and use these re-weighted observations to estimate ψ_λ^B. This would yield an estimate of the bias both from $\lambda > 0$ and the new source of bias from $H \neq G$. Although we assumed $H = G$ throughout our Sister Study analysis, in reality this assumption is likely violated. Furthermore, it is possible this violation is responsible for the non-significant associations of SBP with $PM_{2.5}$ using only 50 degrees-of-freedom to model exposure. With only 50 degrees-of-freedom it is possible that the exposure model is not sufficiently rich, resulting in violation of Condition 1. This violation coupled with $H \neq G$ could be inducing a new source of bias from Berkson-like error. Our estimate of ψ_λ^B may be missing this new source, since we estimated it by summing over the empirical distribution of monitoring locations (essentially treating $H = G$). This would explain why the bias-corrected estimates using 50 degrees-of-freedom were still so different from the bias-corrected 100- and 150-degree-of-freedom results. However, we can only speculate, as the sensitivity described in Section 3.4.2 and implemented in our application indicates the discrepancies may be purely attributable to variability. In future work, we will investigate the measurement error implications of $H \neq G$ and develop methods for estimating G/H to use in bias correction.

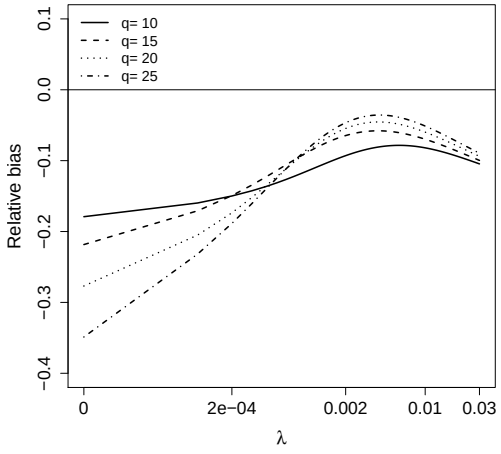
Our methods are readily generalizable to multi-pollutant health analyses. Our bias and variance expressions are simply functions of the moments of $(\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda)$, which in a multi-pollutant context may contain multiple "sub"-vectors of coefficients for each pollutant. As long as we can estimate these moments we can extend our methods to correct for measurement error in multi-pollutant studies. This is the topic of the next chapter.



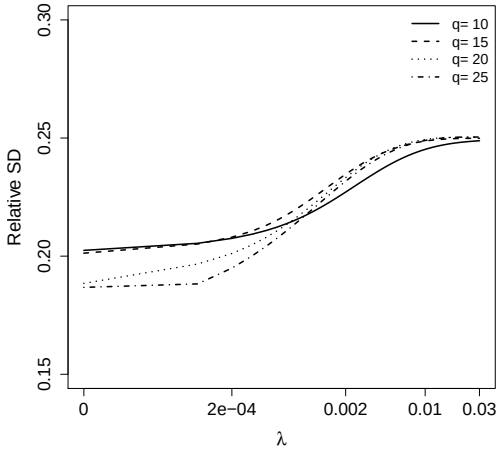
Relative bias: Scenario 1



Relative SD: Scenario 1

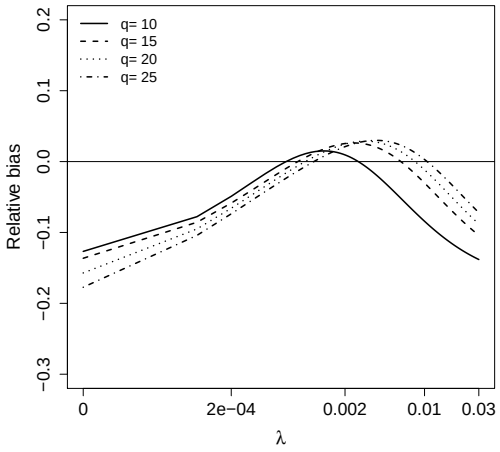


Relative bias: Scenario 2

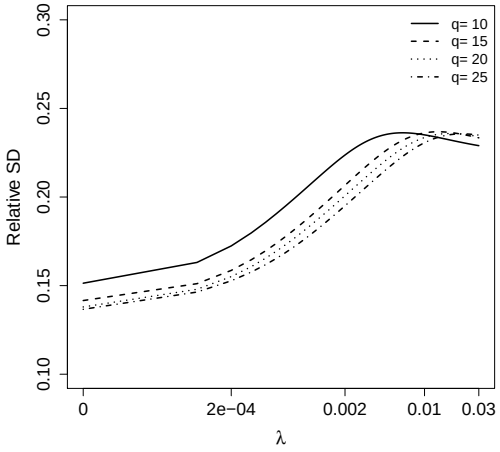


Relative SD: Scenario 2

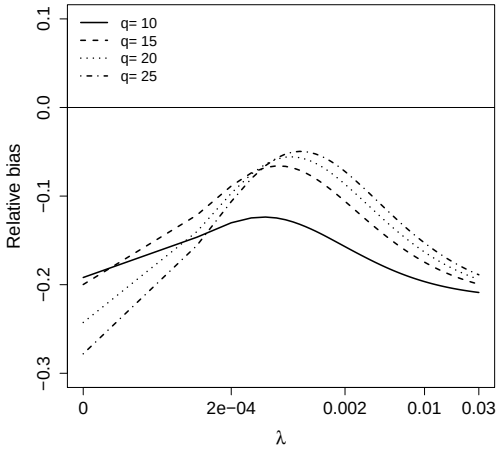
Figure 3.1: Observed relative bias and SE of $\hat{\beta}_{n,n^*}$ as a function of λ, using LRK to model exposure.



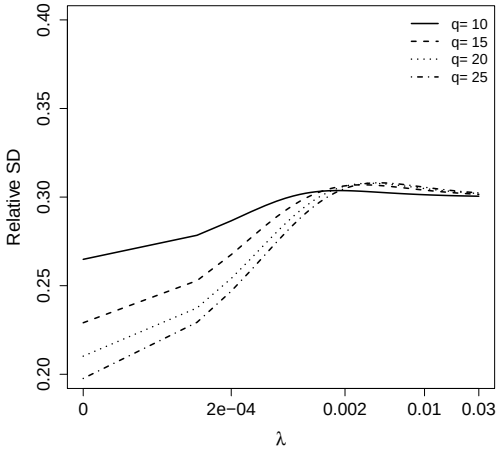
Relative bias: Scenario 1



Relative SD: Scenario 1



Relative bias: Scenario 2



Relative SD: Scenario 2

Figure 3.2: Observed relative bias and SE of $\hat{\beta}_{n,n^*}$ as a function of λ, using TPRS to model exposure.

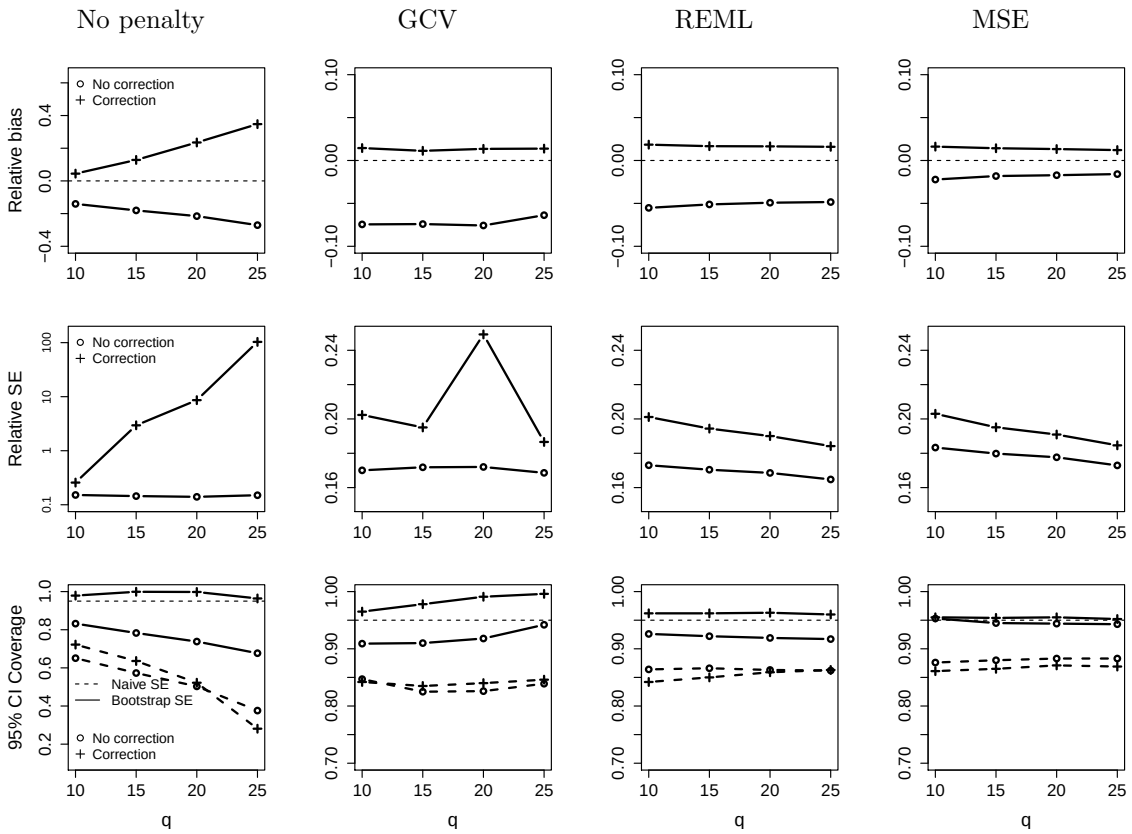


Figure 3.3: Scenario 1, using LRK to model exposure: Observed relative bias and SE of $\hat{\beta}_{n,n^*}$ and $\hat{\beta}^C_{n,n^*}$, and actual coverage of 95% confidence intervals with or without a bias correction, using naïve or bootstrap standard errors.

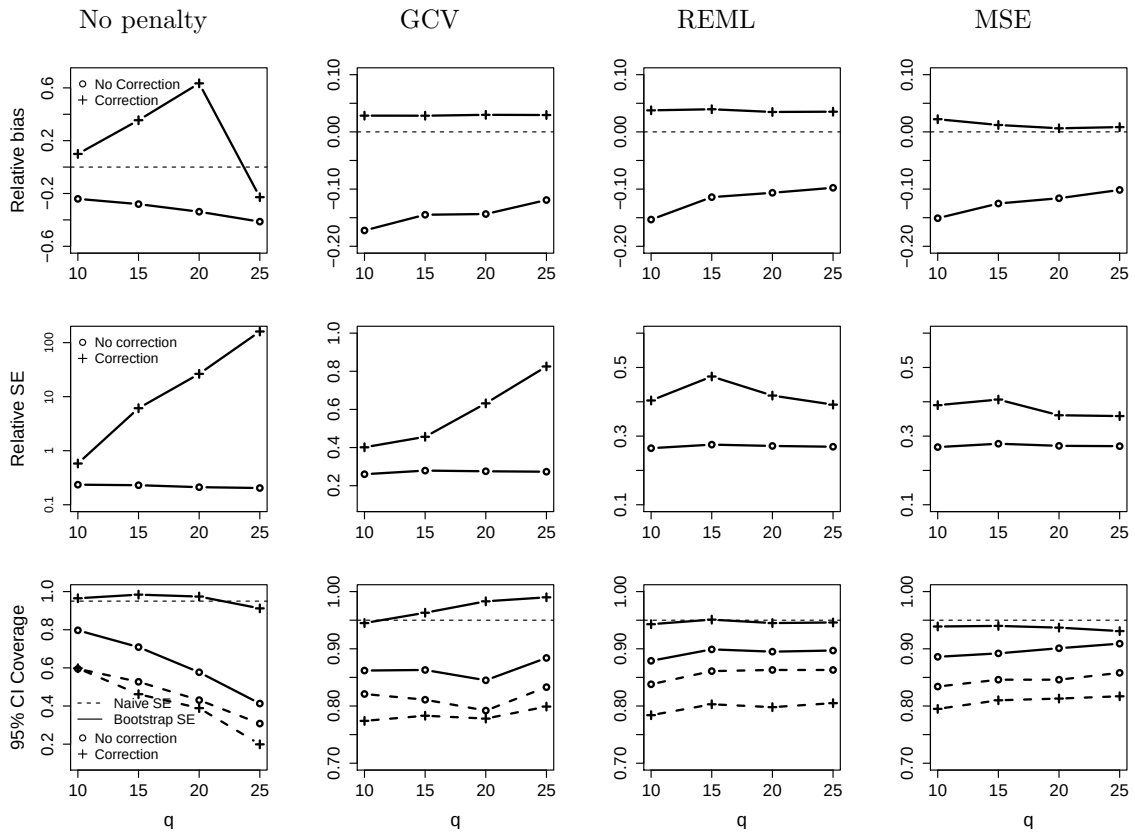


Figure 3.4: Scenario 2, using LRK to model exposure: Observed relative bias and SE of $\hat{\beta}_{n,n^*}$ and $\hat{\beta}^C_{n,n^*}$, and actual coverage of 95% confidence intervals with or without a bias correction, using naïve or bootstrap standard errors.

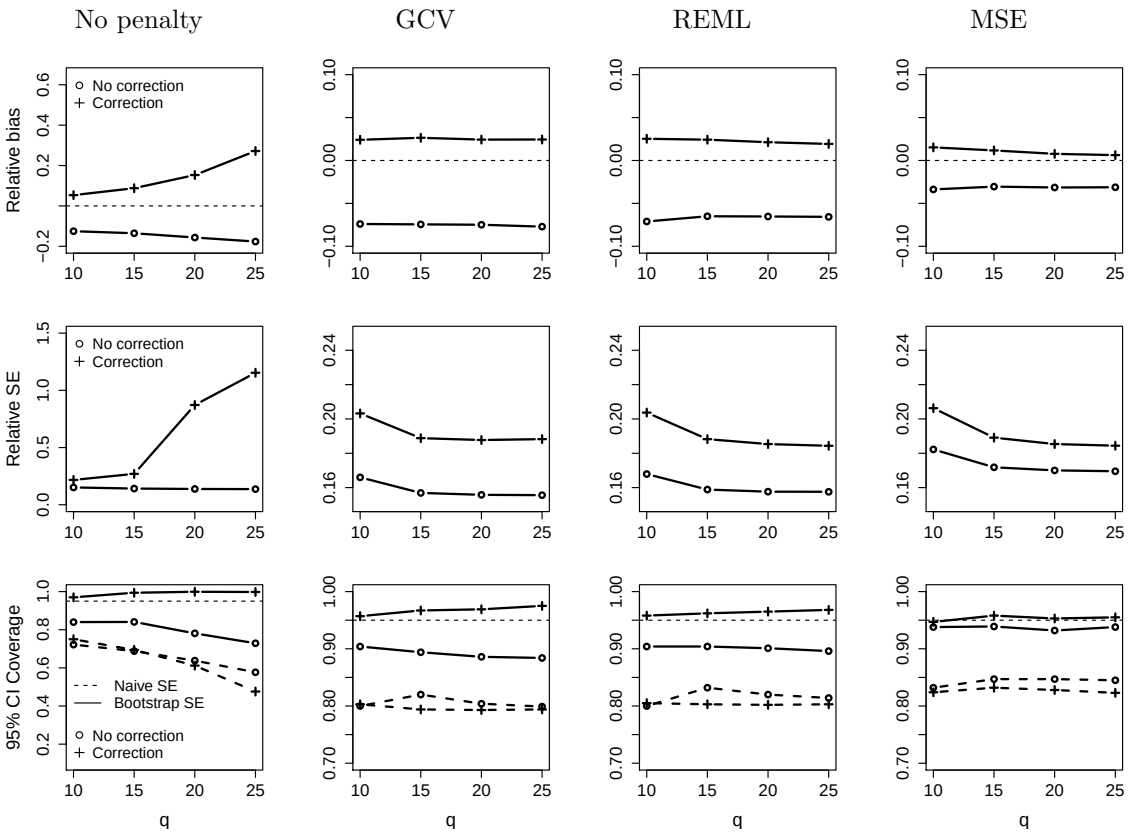


Figure 3.5: Scenario 1, using TPRS to model exposure: Observed relative bias and SE of $\hat{\beta}_{n,n^*}$ and $\hat{\beta}^C_{n,n^*}$, and actual coverage of 95% confidence intervals with or without a bias correction, using naïve or bootstrap standard errors.

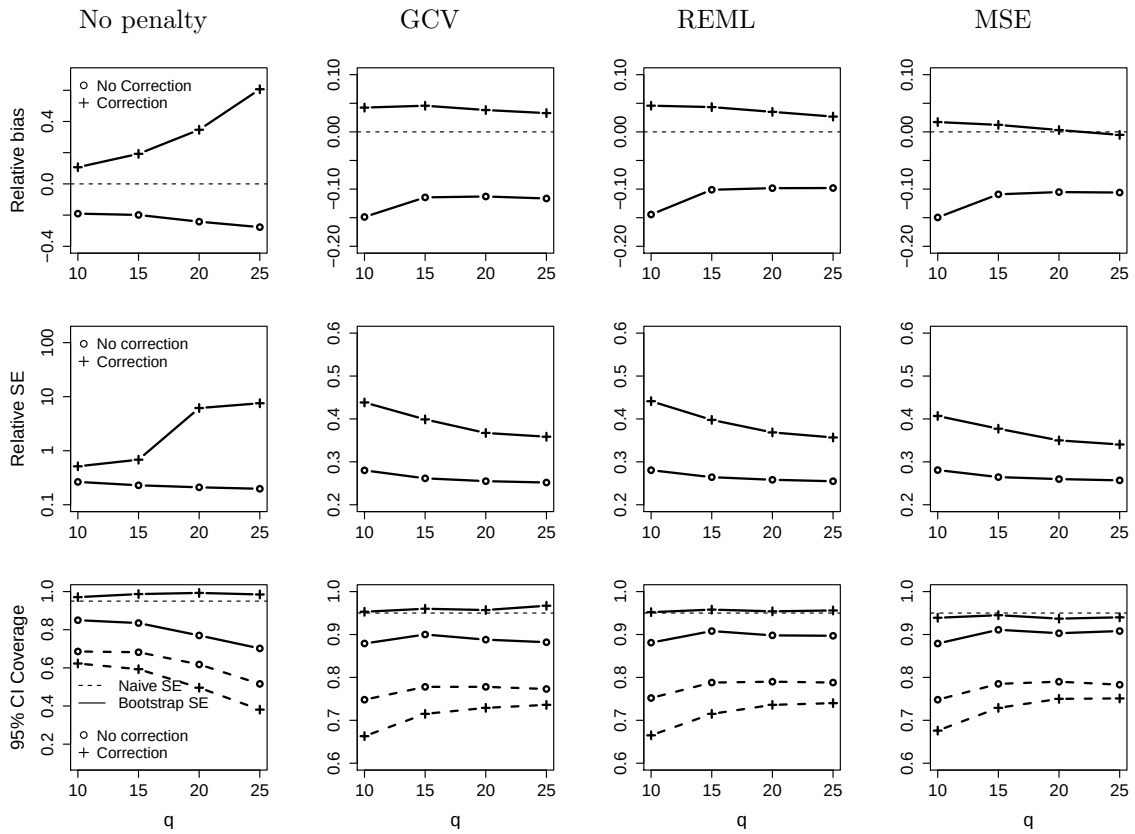


Figure 3.6: Scenario 2, using TPRS to model exposure: Observed relative bias and SE of $\hat{\beta}_{n,n^*}$ and $\hat{\beta}^C_{n,n^*}$, and actual coverage of 95% confidence intervals with or without a bias correction, using naïve or bootstrap standard errors.

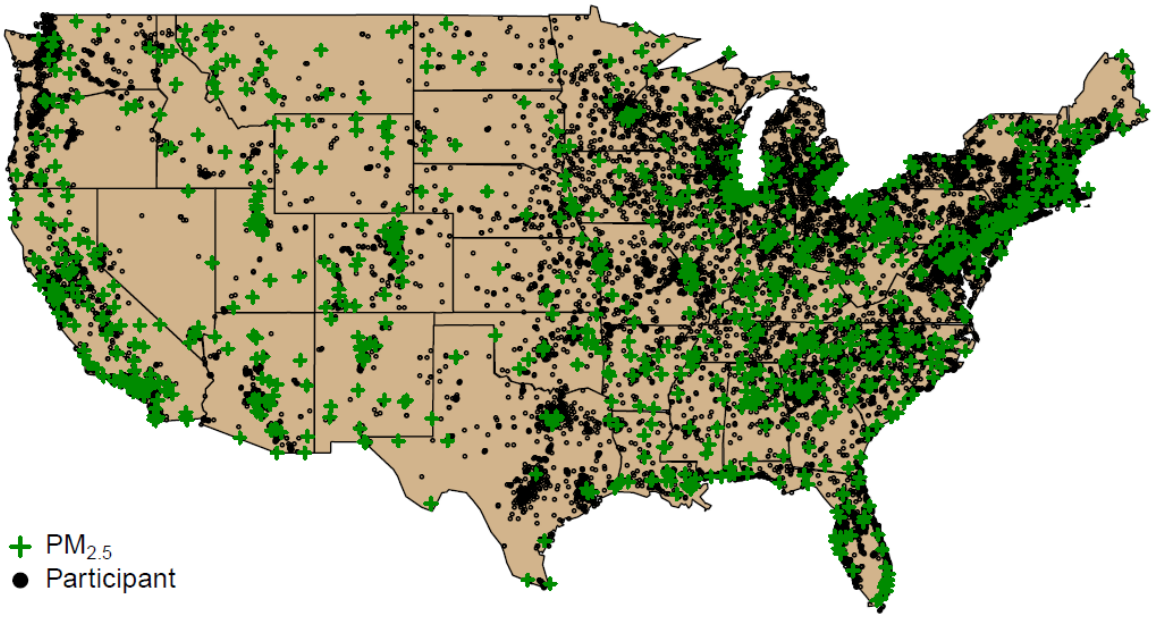
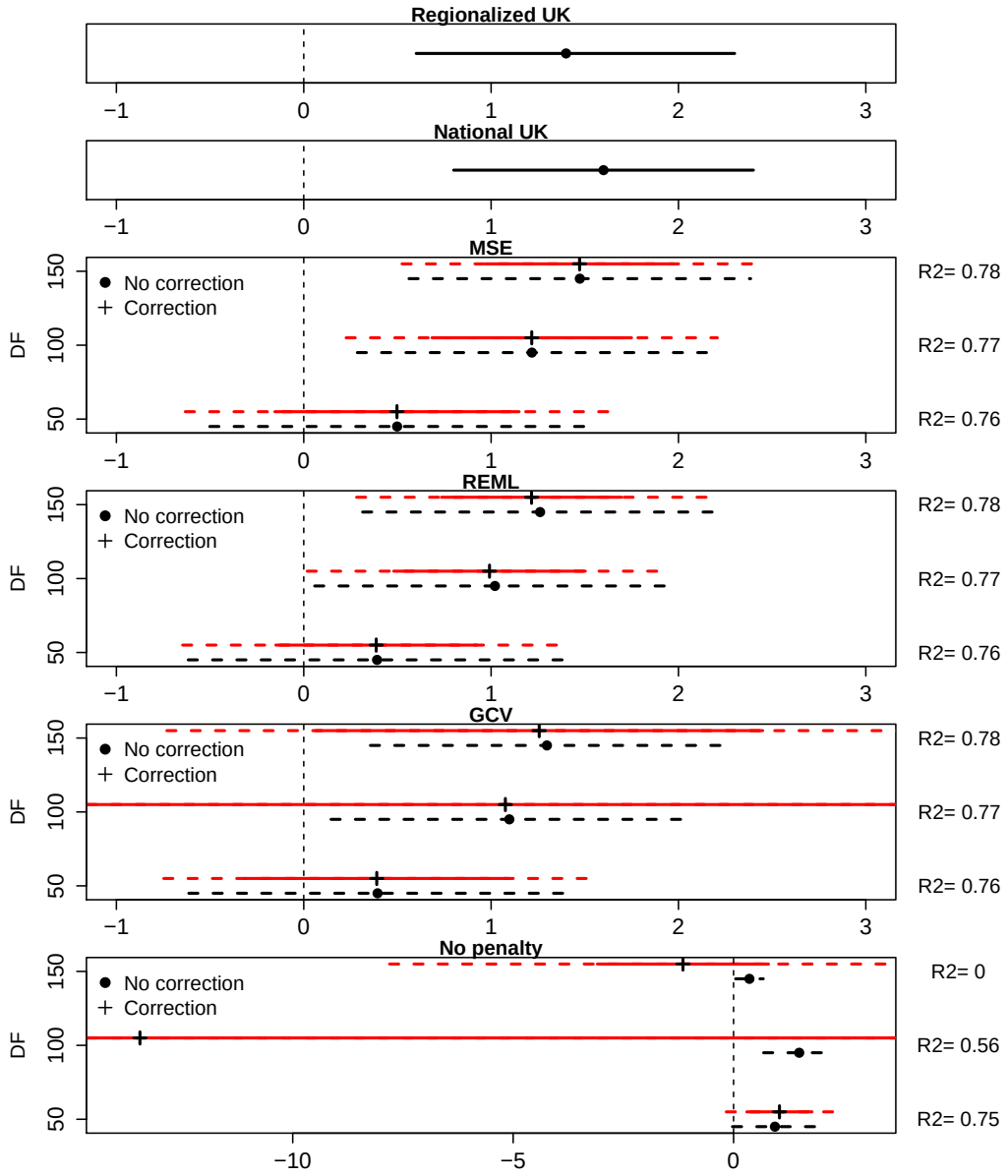
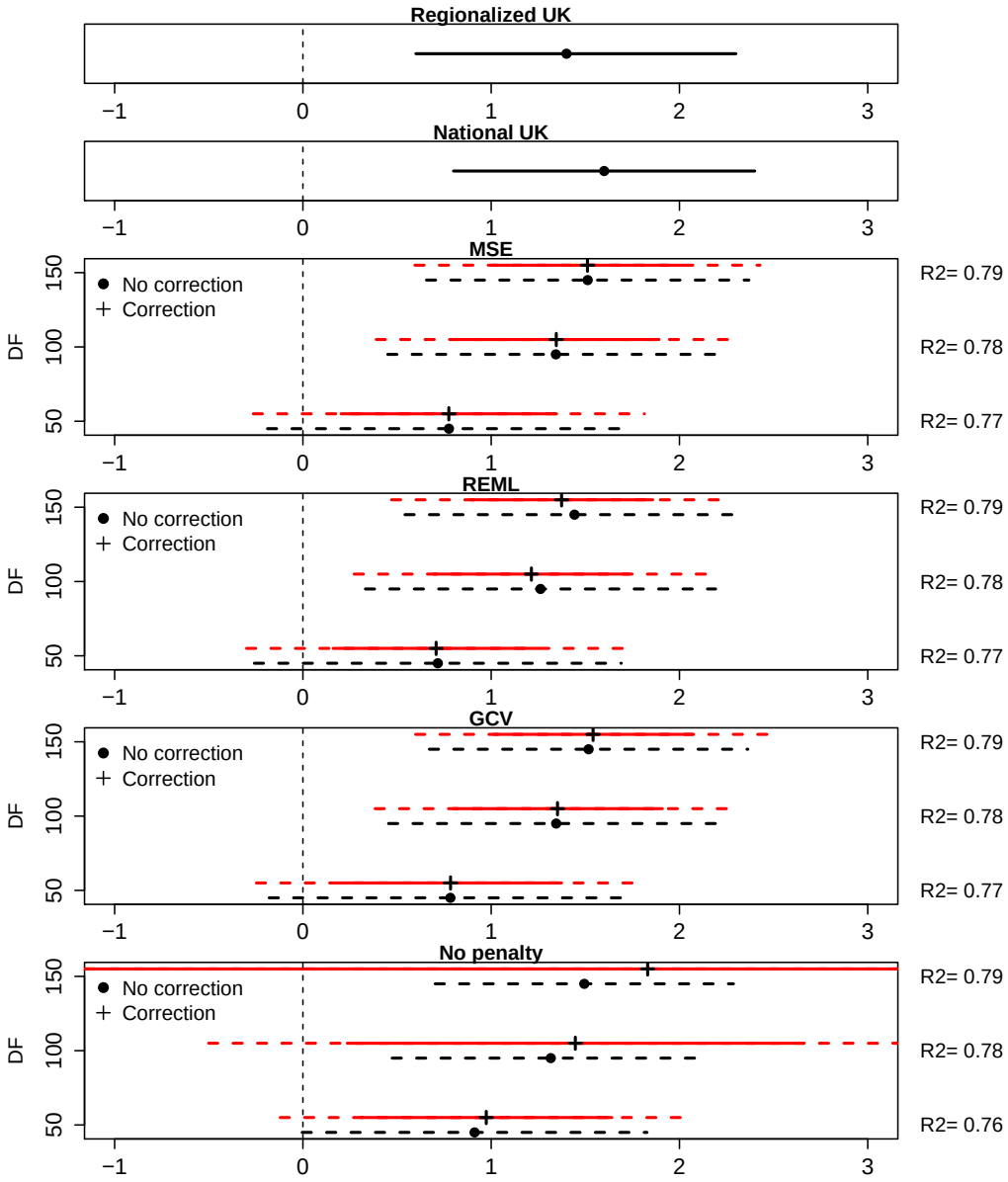


Figure 3.7: Sister Study participant locations and $PM_{2.5}$ monitoring locations used for data analysis.



Estimated change in SBP (in mmHg) associated with a 10-μg/m^3 increase in year 2006 annual average $PM_{2.5}$. Fixed-rank models: low-rank kriging.

Figure 3.8: Estimated change in SBP (in mmHg) associated with a 10-μg/m^3 increase in year 2006 annual average $PM_{2.5}$ predicted by various exposure models. The top two rows of each panel shows the health effect estimated using predictions from regionalized or national universal kriging models. The other rows show naïve and bias-corrected health effects using low-rank kriging to model exposure with λ selected via REML, GCV, MSE or set to be zero. Various 95% confidence intervals are also shown. The black dashed confidence interval were derived from naïve SE estimates. The red solid confidence interval were derived from bootstrapped standard errors that keep subject locations fixed and re-sample only monitoring locations, hence accounting only for measurement error and the bias correction. The red dashed confidence intervals were derived from bootstrap standard errors that took all sources of variability into account. 10-fold cross-validated R^2 for the fixed-rank models are also given.



Estimated change in SBP (in mmHg) associated with a 10-μg/m^3 increase in year 2006 annual average $PM_{2.5}$. Fixed-rank models: thin-plate regression splines.

Figure 3.9: Estimated change in SBP (in mmHg) associated with a 10-μg/m^3 increase in year 2006 annual average $PM_{2.5}$ predicted by various exposure models. The top two rows of each panel shows the health effect estimated using predictions from regionalized or national universal kriging models. The other rows show naïve and bias-corrected health effects using thin-plate regression splines to model exposure with λ selected via REML, GCV, MSE or set to be zero. Various 95% confidence intervals are also shown. The black dashed confidence interval were derived from naïve SE estimates. The red solid confidence interval were derived from bootstrapped standard errors that keep subject locations fixed and re-sample only monitoring locations, hence accounting only for measurement error and the bias correction. The red dashed confidence intervals were derived from bootstrap standard errors that took all sources of variability into account. 10-fold cross-validated R^2 for the fixed-rank models are also given.

Chapter 4

MULTI-POLLUTANT MEASUREMENT ERROR IN AIR POLLUTION EPIDEMIOLOGY STUDIES ARISING FROM PREDICTING EXPOSURES WITH PENALIZED REGRESSION SPLINES

4.1 Summary

Air pollution epidemiology studies are trending towards a multi-pollutant approach, as health outcomes are more realistically influenced by a mixture of pollutants rather than any single pollutant. There is a corresponding need to characterize and correct for spatial measurement error. We extend the methods of the previous chapter, characterizing this measurement error when the exposure models are penalized regression splines, and develop an analytic bias correction. Conventional wisdom suggests measurement error will attenuate all estimated health effects, or that attenuation in the estimated health effect of a poorly measured pollutant translates to upward bias of a well-measured pollutant's effect. We perform simulations that show the biases can be in opposite directions or simultaneously away from or toward the null. Our analytic bias correction combined with a simple non-parametric bootstrap yields accurate coverage of 95% confidence intervals. We illustrate our methodology by analyzing the association between systolic blood pressure and of $PM_{2.5}$ and NO_2 in the NIEHS Sisters Study. We find evidence of NO_2 confounding the association of systolic blood pressure with $PM_{2.5}$ and vice versa. Modeling both exposures together yielded significant positive associations of systolic blood pressure with $PM_{2.5}$ and negative associations with NO_2. Correcting for bias strengthened these associations and widened 95% confidence intervals.

4.2 Introduction

Air pollution epidemiology is trending towards a multi-pollutant approach (Dominici et al., 2010; Billionnet et al., 2012; Vedal and Kaufman, 2011). Dominici et al. (2010) argue that members of health cohorts are exposed to a complex mixture of pollutants that are likely 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).

An inherent challenge of a multi-pollutant approach is obtaining multivariate pollutant predictions at health cohort locations and characterizing and correcting for subsequent measurement error. 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 more complex spatial measurement error that arises from two-stage studies.

In previous chapters we have discussed at length spatial measurement error methods for a single-pollutant. In Chapter 2 we presented parametric methods which are difficult to extend to multi-pollutant setting. In Chapter 3 we developed semi-parametric methods with the intent to extend to the multi-pollutant context. In this chapter we extend those

semi-parametric methods, 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 show that the multi-pollutant measurement error can be decomposed into two components as in Bergen and Szpiro (sub). A Berkson-like component arises from smoothing the exposure surfaces, while a classical-like component arises from estimating the penalized regression spline coefficients. Both components can bias the health effects and affect their standard errors.

This chapter is organized as follows. In Section 4.3 we describe assumptions and modeling strategies. In Section 4.4 we decompose the error into Berkson- and classical-like components and derive an analytic bias correction that accounts for both. In Section 4.5 we discuss bias estimation approaches, exposure model selection and variance estimation. In Section 4.6 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 4.7 we apply our methodology to analyze the joint association of $PM_{2.5}$ and NO_2 with systolic blood pressure (SBP) in the Sister Study cohort (NIEHS, 2013). We conclude with a discussion of our results in Section 4.8.

4.3 Analytic framework

4.3.1 Data generating mechanisms

Our assumed data generating mechanisms follow those described in Szpiro and Paciorek (2013) and Bergen and Szpiro (sub). We formulate our methodology in the context of two pollutants, emphasizing that extension to higher-dimension pollutants is easily accomplished at the cost of notational simplicity. Let $\mathbf{s}_1, .., \mathbf{s}_n$ be n subject locations drawn independently from an unknown distribution function $G(\cdot)$. Given these subject locations, the true

unobserved exposures $\mathbf{x}_i \equiv (x_{i1}, x_{i2})^T$ follow:

$$\mathbf{x}_i = \mathbf{\Phi}(\mathbf{s}_i) + \boldsymbol{\eta}_i,$$

for $i = 1, ..., n$. Here $\mathbf{\Phi}(\mathbf{s}_i) \equiv \{\Phi_1(\mathbf{s}_i), \Phi_2(\mathbf{s}_i)\}^T$ is a fixed function from $\mathbb{R}^2 \to \mathbb{R}^2$ mapping space to a fixed two-dimensional exposure surface with $Cor(\Phi_1(\mathbf{s}), \Phi_2(\mathbf{s})) = \rho_\Phi \in (-1, 1)$. Each Φ_j is potentially explainable with spatially-referenced covariates and basis functions. The $\boldsymbol{\eta}_i^T \equiv \{\eta_{i1}, \eta_{i2}\}$ are random vectors drawn independently of each other where $Cor(\eta_{i1}, \eta_{i2}) = \rho_\eta$.

In addition to the subject locations we observe n^* monitoring locations $\mathbf{s}_1^*, ..., \mathbf{s}_{n^*}^*$ drawn independently of each other and of the subject locations, where n_j^* of the $\mathbf{s}_i^*$ 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. Let π_j^* denote the probability that a randomly drawn monitoring location belongs to set $\mathcal{S}_j^*$. Let $H_j(\cdot)$ denote the unknown distribution of monitoring locations in $\mathcal{S}_j^*$, and we 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 4.8.

We want to ensure we observe both pollutants at some locations. Accordingly we restrict π_0^* to be positive while π_j^* may equal 0 for $j = 1$ and/or 2; this implies $0 < n_0^* \leq n^*$. As we describe in Section 4.5.1 this assumption is necessary to ensure we can estimate the bias from measurement error. We denote $\mathbf{x}_i^* = \mathbf{\Phi}(\mathbf{s}_i^*) + \boldsymbol{\eta}_i^*$ as the complete vector of exposures at monitor location $\mathbf{s}_i^*$. We do not observe the complete vector $\mathbf{x}_i^*$ at all locations, instead we observe $\mathbf{W}(\mathbf{s}_i^*)\mathbf{x}_i^*$ where $\mathbf{W}(\mathbf{s}_i^*) = \begin{pmatrix} 1(\mathbf{s}_i^* \in \mathcal{S}_0^* \cup \mathcal{S}_1^*) & 0 \\ 0 & 1(\mathbf{s}_i^* \in \mathcal{S}_0^* \cup \mathcal{S}_2^*) \end{pmatrix}$.

We emphasize that this paradigm implies $\mathbf{x}_i^*$ and $\mathbf{x}_{i'}^*$ are independent of each other for $i \neq i'$ since the monitoring locations are independent, but that at any given $\mathbf{s}_i^*$ the elements of $\mathbf{x}_i^*$ are correlated due to correlation between elements of $\mathbf{\Phi}(\mathbf{s}_i^*)$ and $\boldsymbol{\eta}_i^*$. This is analogous

to viewing locations as i.i.d. "clusters" in a GEE setting with multiple measures in a cluster (Liang and Zeger, 1986), which as we will describe has implications for estimating the covariance of penalized regression coefficients.

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 allows using different covariates and basis functions to model the two surfaces. Let $\mathbf{p}_j(\mathbf{s}_i^*)$ denote a function mapping 2-dimensional space to 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 function mapping 2-dimensional space to a $q_j \times 1$ vector of spatial basis functions for $j \in \{1, 2\}$. Let $\mathbf{r}_j(\mathbf{s}_i^*)$ be a function of the $\mathbf{p}_j(\mathbf{s}_i^*)$ and $\mathbf{q}_j(\mathbf{s}_i^*)$ such that $\mathbf{r}_1(\mathbf{s}_i^*)$ and $\mathbf{r}_2(\mathbf{s}_i^*)$ both have r elements. For example, if fitting a separate regression model for each pollutant $\mathbf{r}_j(\mathbf{s}_i^*)^T = \{1(j=1)\mathbf{p}_1(\mathbf{s}_i^*)^T, 1(j=1)\mathbf{q}_1(\mathbf{s}_i^*)^T, 1(j=2)\mathbf{p}_2(\mathbf{s}_i^*)^T, 1(j=2)\mathbf{q}_2(\mathbf{s}_i^*)^T\}$. We define the $r \times 2$ location-specific model matrix $\mathbf{R}(\mathbf{s}_i^*) = \{\mathbf{r}_1(\mathbf{s}_i^*), \mathbf{r}_2(\mathbf{s}_i^*)\}$, and the $n^* \times r$ pollutant-specific model matrix $\mathbf{R}_j^*$ created by stacking the $\mathbf{r}_j(\mathbf{s}_i^*)^T$ for $i = 1, ..., n^*$. Analogously we define $\mathbf{r}_j(\mathbf{s}_i)$, $\mathbf{R}(\mathbf{s}_i)$ and $\mathbf{R}_j$ at subject locations.

Given true exposure at subject locations, the health outcomes follow the linear model

$$y_i = \beta_0 + \boldsymbol{\beta}^T \mathbf{x}_i + \boldsymbol{\beta}_Z^T \mathbf{z}_i + \epsilon_i.$$

The ϵ_i are independent but not necessarily identically distributed random variables with mean zero, and are also independent of the $\mathbf{x}_i$ and $\mathbf{z}_i$. As in Szpiro and Paciorek (2013) and Bergen and Szpiro (sub), the subject-specific covariates $\mathbf{z}_i$ are defined as

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

where $\Theta(\mathbf{s}_i) = (\theta_{i1}, ..., \theta_{im})^T$ is an $m-$dimensional vector-valued function representing the spatial component of the subject-specific covariates, and the $\boldsymbol{\zeta}_i = (\zeta_{i,1}, ..., \zeta_{i,m})$ are random $m-$vectors independent between subjects and indepdendent of $\boldsymbol{\eta}_i$, with mean zero, but

where the individual components of $\boldsymbol{\zeta}_i$ are not necessarily independent of each other. We are interested in estimating $\boldsymbol{\beta}$; as we do not observe the $\mathbf{x}_i$ this requires building a prediction model using the $\mathbf{x}_i^*$ and $\mathbf{R}(\mathbf{s}_i^*)$ and predicting at locations $\mathbf{s}_i$ using $\mathbf{R}(\mathbf{s}_i)$. The following section describes the exposure model.

4.3.2 Penalized regression exposure model

In our simulations and data analysis we fit separate penalized regression models for each pollutant. These models are straightforward and easy to implement 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. In Section C.1 we describe such a model and how it fits into this general class, though we do not implement it in this paper.

General form

Given a monitoring data set consisting of observed pollutant exposures and spatially referenced 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 $\boldsymbol{\Lambda}$ and 2×2 weighting matrix $\boldsymbol{\Pi}$, we obtain $\hat{\boldsymbol{\gamma}}$ where:

$$\hat{\boldsymbol{\gamma}} = \underset{\theta}{\operatorname{argmin}} \frac{1}{n^*} \sum_{i=1}^{n^*} (\mathbf{x}_i^* - \mathbf{R}(\mathbf{s}_i^*)^T\theta)^T \mathbf{W}(\mathbf{s}_i^*) \boldsymbol{\Pi} \mathbf{W}(\mathbf{s}_i^*) (\mathbf{x}_i^* - \mathbf{R}(\mathbf{s}_i^*)^T\theta) + \theta^T \boldsymbol{\Lambda} \theta. \tag{4.1}$$

Note that including $\mathbf{W}(\mathbf{s}_i^*)$ allows for observing one or both pollutants at any location. The penalty matrix $\boldsymbol{\Lambda}$ penalizes roughness of the exposure model. It may be zero everywhere in which case (4.1) reduces to a weighted least squares equation. However not all exposure models necessarily admit all-zero $\boldsymbol{\Lambda}$, for example if $\mathbf{r}_j(\mathbf{s}_i^*)$ contains multiple copies of the

same basis functions. One example of such a model is the low-rank analogue of the common component model, which we describe in more detail in Section C.1. The weighting matrix $\mathbf{\Pi}$ is fixed and may be diagonal. We give specific examples of $\mathbf{\Lambda}$ and $\mathbf{\Pi}$ under different modeling scenarios below and in Section C.1.

As $n^* \to \infty$, $\hat{\boldsymbol{\gamma}} \to \boldsymbol{\gamma}$ where

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

Having obtained penalized regression coefficients, we define predictions at subject locations as $\hat{\mathbf{w}}(\mathbf{s}_i) = \mathbf{R}(\mathbf{s}_i)^T\hat{\boldsymbol{\gamma}}$. Analogously, let $\mathbf{w}(\mathbf{s}_i) = \mathbf{R}(\mathbf{s}_i)^T\boldsymbol{\gamma}$ denote the predictions we would make if we had infinite monitoring data to fit the exposure model. Let $\hat{\boldsymbol{\beta}}$ denote the estimate of $\boldsymbol{\beta}$ using $\hat{\mathbf{w}}(\mathbf{s}_i)$ and $\mathbf{z}_i$.

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 $\mathbf{\Pi}$ is a 2×2 identity matrix, $\mathbf{r}_j(\mathbf{s}_i^*)^T = \{1(j = 1)\mathbf{p}_1(\mathbf{s}_i^*)^T, 1(j = 1)\mathbf{q}_1(\mathbf{s}_i^*)^T, 1(j = 2)\mathbf{p}_2(\mathbf{s}_i^*)^T, 1(j = 2)\mathbf{q}_2(\mathbf{s}_i^*)^T\}$, and $\mathbf{\Lambda}$ is block diagonal with elements $\lambda_j\mathbf{D}_j$, where the $\mathbf{D}_j$ 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 (sub) for more details). Then $\hat{\boldsymbol{\gamma}}$ can be expressed as

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

$$= \underset{\theta}{\operatorname{argmin}}(SS_1(\theta) + SS_2(\theta)).$$

Because of the form of the $\mathbf{r}_j$, minimizing this equation with respect to θ is equivalent to separately minimizing $SS_1(\theta)$ with respect to the first (p_1+q_1) elements of θ and $SS_2(\theta)$ with respect to the last $(p_2 + q_2)$ elements of θ. 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-fit coefficients, and the j^{th} element of $\hat{\mathbf{w}}(\mathbf{s}_i) = \mathbf{r}(\mathbf{s}_i)^T\hat{\boldsymbol{\gamma}}$ is equivalent to $\left\{\mathbf{p}_j(\mathbf{s}_i)^T, \mathbf{q}_j(\mathbf{s}_i)^T\right\}^T \hat{\boldsymbol{\gamma}}_j$.

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 (sub) show how each model yields definitions of the spatial bases and the penalty matrices $\mathbf{D}_j$, 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.

4.4 Measurement error

4.4.1 Decomposing the measurement error

Once (4.1) has been used to obtain $\hat{\boldsymbol{\gamma}}$ and predictions $\hat{\mathbf{w}}(\mathbf{s}_i)$ obtained at subject locations we use them to estimate $\boldsymbol{\beta}$. Doing so induces measurement error which we decompose as follows:

$$\begin{aligned}\mathbf{x}_i - \hat{\mathbf{w}}(\mathbf{s}_i) &= (\mathbf{x}_i - \mathbf{w}(\mathbf{s}_i)) + (\mathbf{w}(\mathbf{s}_i) - \hat{\mathbf{w}}(\mathbf{s}_i)) \\ &= \mathbf{u}^B(\mathbf{s}_i) + \mathbf{u}^C(\mathbf{s}_i)\end{aligned}$$

Here $\mathbf{u}^B(\mathbf{s}_i)$ 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 $\mathbf{w}(\mathbf{s}_i)$ is a smoothed version of is 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, $\mathbf{u}^C(\mathbf{s}_i)$, 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 (sub), the penalty matrix $\mathbf{\Lambda}$ regulates the impact of each error type. For SPRS,$\mathbf{\Lambda}$ 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 $\hat{\boldsymbol{\gamma}}$ 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 (sub) for more details.

We will derive analytic expressions for the bias arising from each error type. First we discuss conditions under which the Berkson-like error does not induce any bias under optimal conditions, namely having infinite monitoring data and maximum flexibility in the exposure model.

4.4.2 Infinite n^* bias

In what follows we operate under the infinite-n^* limit so as to isolate the impact of the Berkson-like error. With infinite monitoring data we can let $\mathbf{\Lambda}$ be zero everywhere since its purpose is to ensure model regularity under finite n^*. This maximizes exposure model flexibility since there is no restriction on the elements of $\boldsymbol{\gamma}$. Even with infinite monitoring data, if neither of the following spatial compatibility conditions are satisfied, $\mathbf{u}^B$ may induce bias when the resulting predictions are used to estimate $\boldsymbol{\beta}$.

1. Sufficient basis functions are included in the exposure models such that for all $\mathbf{s}$, $\mathbf{\Phi}_j(\mathbf{s}) = \mathbf{r}_j(\mathbf{s})^T\boldsymbol{\gamma}$ for some $\boldsymbol{\gamma}$.

2. Both of the following are met:

   (a) The elements of $\mathbf{r}_j(\mathbf{s})$ are spanned by $\mathbf{r}_{j'}(\mathbf{s})$ for $j \neq j'$; and

   (b) the elements of $\Theta(\mathbf{s})$ are contained in the span of $\mathbf{r}_j(\mathbf{s})$ for $j \in \{1, 2\}$.

To see the implications of these conditions when fitting SPRS with both $\lambda_j = 0$, we re-write the health model as:

$$y_i = \beta_0 + \beta_1 \left(\mathbf{r}_1(\mathbf{s}_i)^T\boldsymbol{\gamma}\right) + \beta_2 \left(\mathbf{r}_2(\mathbf{s}_i)^T\boldsymbol{\gamma}\right) + \boldsymbol{\beta}_Z^T\Theta(\mathbf{s}_i) + \left\{\beta_1 \left(\Phi_1(\mathbf{s}_i) - \mathbf{r}_1(\mathbf{s}_i)^T\boldsymbol{\gamma}\right) + \beta_2 \left(\Phi_2(\mathbf{s}_i) - \mathbf{r}_2(\mathbf{s}_i)^T\boldsymbol{\gamma}\right) + \boldsymbol{\beta}_Z^T\boldsymbol{\zeta}_i + \boldsymbol{\beta}^T\boldsymbol{\eta}_i + \epsilon_i\right\}.$$

From here we note that the elements of $\mathbf{u}^B$ become part of the residual in the health model, and unbiasedness of $\boldsymbol{\beta}$ is only guaranteed if each residual term is orthogonal to all health model covariates under $G(\cdot)$ (White, 1980). As $\boldsymbol{\zeta}_i$, $\boldsymbol{\eta}_i$ and ϵ_i are independent of everything in the health model, this implies we need orthogonality of $\left(\Phi_j(\mathbf{s}_i) - \mathbf{r}_j(\mathbf{s}_i)^T\boldsymbol{\gamma}\right)$ to the health model covariates.

If Condition 1 is met, the $\left(\Phi_j(\mathbf{s}_i) - \mathbf{r}_j(\mathbf{s}_i)^T\boldsymbol{\gamma}\right) = 0$ for all $\mathbf{s}_i$ and no bias is induced by $\mathbf{u}^B$. If Condition 1 is not met the $\left(\Phi_j(\mathbf{s}_i) - \mathbf{r}_j(\mathbf{s}_i)^T\boldsymbol{\gamma}\right)$ are not guaranteed to always be zero and we must rely on Condition 2 to ensure orthogonality under $G(\cdot)$. Under our assumption that $H_j(\cdot) = G(\cdot)$ for all j, it follows that $\mathbf{r}_j(\mathbf{s})^T\boldsymbol{\gamma}$ is the projection of $\Phi_j(\mathbf{s})$ onto the $(p_j + q_j)$-dimensional subspace spanned by $\mathbf{r}_j(\mathbf{s})$. The subspace is of dimension p_j+q_j since only these elements of $\mathbf{r}_j(\mathbf{s})$ are non-zero. Along with 2(a), Condition 2(b) ensures $\left(\Phi_j(\mathbf{s}_i) - \mathbf{r}_j(\mathbf{s}_i)^T\boldsymbol{\gamma}\right)$ is orthogonal to $\mathbf{r}_{j'}(\mathbf{s})^T\boldsymbol{\gamma}$ under $G(\cdot)$ even if $j \neq j'$. This follows since each element of $\mathbf{r}_{j'}(\mathbf{s})$ is a linear combination of elements of $\mathbf{r}_j(\mathbf{s})$ which are in turn orthogonal to $\left(\Phi_j(\mathbf{s}) - \mathbf{r}_j(\mathbf{s})^T\boldsymbol{\gamma}\right)$. If Condition 2(b) is satisfied these terms are also orthogonal to the elements of $\Theta(\mathbf{s}_i)$ since they are spanned by the $\mathbf{r}_j(\mathbf{s})$.

Note that unbiasedness is guaranteed if either Condition 1 or 2 is met. It is apparent that Condition 1 is the easiest to satisfy by including many spatial basis functions in the exposure model. In this case penalization will likely be necessary to ensure model regularity given finite monitoring data. Condition 2(b) is most easily satisfied by using the same set of geographic covariates and spatial basis functions to model each pollutant. However, it may be difficult to ensure that Condition 2(b) is met since subject-specific covariates such

as socio-economic status may not be defined at monitoring locations.

We emphasize that $\mathbf{\Lambda} = 0$ is required to guarantee $\mathbf{u}^B$ does not induce bias, but that not all exposure models admit $\mathbf{\Lambda} = 0$. Accordingly for some exposure models such as the one described in Section C.1 it may be impossible to guarantee unbiased health effect estimation even with many monitoring locations.

4.4.3 Finite n^* bias

In practice we have finite monitoring data to model exposure and will likely need non-zero $\mathbf{\Lambda}$ to ensure model regularity. Finite monitoring data induces bias from classical-like error and non-zero $\mathbf{\Lambda}$ induces bias from Berkson-like error even if Condition 1 or 2 are met. We derive an analytic bias expression using a Taylor expansion that accounts for both biases.

Lemma 2: Let $\mathbf{r}_j^{\perp}(\mathbf{s})$ contain elements $(r_{jk}(\mathbf{s}) - \Theta(\mathbf{s})^T\varphi_k)$, where $\varphi_k = \text{argmin}_\omega \int (r_{jk}(\mathbf{s}) - \Theta(\mathbf{s})^T\omega)^2 dG(\mathbf{s})$ for $k \in \{1, ..., r\}$. Let $\mathbf{R}^{\perp}(\mathbf{s})$ denote the corresponding $r \times 2$ matrix created by binding the $r_j^{\perp}(\mathbf{s})$, let $\mathbf{w}^{\perp}(\mathbf{s}) = \mathbf{R}^{\perp}(\mathbf{s})^T\boldsymbol{\gamma}$ and $\hat{\mathbf{w}}^{\perp}(\mathbf{s}) = \mathbf{R}^{\perp}(\mathbf{s})^T\hat{\boldsymbol{\gamma}}$. Let $\mathbf{M}(\hat{\boldsymbol{\gamma}}) = \int \hat{\mathbf{w}}^{\perp}(\mathbf{s})\hat{\mathbf{w}}^{\perp}(\mathbf{s})^T dG(\mathbf{s})$ and $\mathbf{U}(\hat{\boldsymbol{\gamma}}) = \int \hat{\mathbf{w}}^{\perp}(\mathbf{s})\mathbf{u}^B(\mathbf{s})^T dG(\mathbf{s})$. Then with

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

we can show that $\hat{\boldsymbol{\beta}} = f(\hat{\boldsymbol{\gamma}})^T\boldsymbol{\beta}$. Let $\mathbf{G}_i(\hat{\boldsymbol{\gamma}})$ denote the 2×2 matrix created by taking the derivative of $f(\hat{\boldsymbol{\gamma}})$ with respect to $\hat{\gamma}_i$ and $\mathbf{H}_{ij}$ the 2×2 matrix created by taking the derivative of $\mathbf{G}_i(\hat{\boldsymbol{\gamma}})$ with respect to $\hat{\gamma}_j$. Let $\mathbf{g}_{kl}(\hat{\boldsymbol{\gamma}})$ denote the $r \times 1$ vector where $g_{ikl}(\hat{\boldsymbol{\gamma}})$ equals the $\{k, l\}^{th}$ element of $\mathbf{G}_i(\boldsymbol{\gamma})$, and $\mathbf{h}_{kl}(\hat{\boldsymbol{\gamma}})$ the $r \times r$ matrix where $h_{ijkl}(\hat{\boldsymbol{\gamma}})$ equals the $\{k, l\}^{th}$ element of $\mathbf{H}_{ij}(\hat{\boldsymbol{\gamma}})$.

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

$$E\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) = \left(\Psi^B + \Psi^C\right)\boldsymbol{\beta} \tag{4.2}$$

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

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

is bias from classical-like error.

See Appendix C for a more rigorous statement and proof of Lemma 2. From here we can readily see that the bias from classical-like error vanishes as $n^* \to \infty$, since Ψ^C is of order $1/n^*$. Since Ψ^C also depends on $Cov(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})$, we see that Ψ^C can be lessened by increasing penalization of the regression coefficients. Conversely the bias from Berkson-like error increases as $\boldsymbol{\Lambda}$ increases, as the elements of $\mathbf{w}$ and $\mathbf{u}^B$ are more correlated.

Lemma 2 shows explicitly that the bias of $\hat{\beta}_j$ depends on the entire vector $\boldsymbol{\beta}$. 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 $\beta_{j'}$ is of larger magnitude than β_j.

4.5 Methods

4.5.1 Bias estimation and correction

In order to correct for bias, we must estimate Ψ^B and Ψ^C. Estimating Ψ^B requires estimates of $\mathbf{M}(\boldsymbol{\gamma})$ and $\mathbf{U}(\boldsymbol{\gamma})$. We can estimate $\mathbf{M}(\boldsymbol{\gamma})$ by replacing $\boldsymbol{\gamma}$ with its consistent estimate $\hat{\boldsymbol{\gamma}}$, and $\int \mathbf{R}^{\perp}(\mathbf{s})^T \mathbf{R}^{\perp}(\mathbf{s}) dG(\mathbf{s})$ directly by summing over the empirical distribution of the subject locations. We must use monitoring data to estimate $\mathbf{U}(\boldsymbol{\gamma})$, since only at monitoring locations do we have estimates of $\mathbf{u}^B$ (obtained using the difference between the observed and fitted exposures). The diagonal entries of $\mathbf{U}(\boldsymbol{\gamma})$ can be estimated by summing over the empirical distribution of monitoring locations in $\mathcal{S}_0 \cup \mathcal{S}_1$ and $\mathcal{S}_0 \cup \mathcal{S}_2$ respectively, while the

off-diagonal entries can only be estimated by summing over the empirical distribution of monitoring locations in $\mathcal{S}_0$. Note that this is the reason we assumed $\pi_0^* > 0$, as without this assumption there is no guarantee we can estimate the off-diagonal elements of $\mathbf{U}(\boldsymbol{\gamma})$.

Estimating Ψ^C requires estimating the $\mathbf{g}_{ij}$, $\mathbf{h}_{ij}$, and moments of $(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$. $\mathbf{g}_{ij}$ and $\mathbf{h}_{ij}$ are estimable with subject locations or with re-weighted monitoring locations; we describe these quantities in more detail in Section C.2. To estimate $E(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$ we use the approach described by Bergen and Szpiro (sub), since $(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$ is equivalent to two vectors of separately-fit penalized spline coefficients. We estimate $Cov(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$ using a sandwich covariance, noting that Equation 4.1 can be formulated using GEE. Each location can be considered an independent cluster with potentially multiple observations in each cluster, and we can apply standard sandwich covariance calculations to obtain an estimate of $Cov(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$. The corresponding estimate incorporates reduction in $Cov(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$ from non-zero $\boldsymbol{\Lambda}$.

Lemma 2 implies $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + \Psi^B\boldsymbol{\beta} + \Psi^C\boldsymbol{\beta}$. One we obtain estimates $\hat{\Psi}^B$ and $\hat{\Psi}^C$, it follows that the bias-corrected estimate of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}}^C = \left(\mathbf{I}_2 + \hat{\Psi}^B + \hat{\Psi}^C\right)^{-1} \hat{\boldsymbol{\beta}}$.

4.5.2 Standard error estimation

As described in Section 4.4.1, both $\mathbf{u}^B$ and $\mathbf{u}^C$ can affect the variance matrix of $\hat{\boldsymbol{\beta}}$. Correcting for bias may have further impact on the standard errors of $\hat{\boldsymbol{\beta}}^C$. We can account for all sources of variability using a simple 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 $\mathbf{s}_i$. 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 $\hat{\boldsymbol{\beta}}^B$ and bias-corrected estimates $\hat{\boldsymbol{\beta}}^{C,B}$. We can then use the empirical standard deviations of the bootstrap estimates to form 95% confidence intervals that account for all sources of variability.

4.5.3 Model selection

As described in Section 4.4.2 there is rationale for highly parameterizing the exposure model, which requires penalization for finite monitoring data to ensure regularity. For SPRS this necessitates choosing both λ_1 and λ_2. In single-pollutant studies Bergen and Szpiro (sub) 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 4.4.3. 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 $\boldsymbol{\beta}$ is unknown. Bergen and Szpiro (sub) 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 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 4.2 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 $2n_0^*/(1 + Cor(x_{1i}, x_{2i}))$; we can heuristically think of $2n_0^*/(1 + Cor(x_{1i}, x_{2i}))$ as an "effective sample size" for estimating $E(\hat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1)(\hat{\boldsymbol{\gamma}}_2 - \boldsymbol{\gamma}_2)^T$. If $Cor(x_{1i}, x_{2i}) = 0$ the effective sample size is $2n_0^*$. If $Cor(x_{1i}, x_{2i}) = 1$ 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 $\mathbf{s}_i^* \in \mathcal{S}_0^*$. Weighting each observation at each location by $1 + Cor(x_{1i}, x_{2i})$ 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 $\leq 10\%$ of the effective sample size as the "EDF"

method of choosing the λ_j.

4.6 Simulations

4.6.1 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 scenario and a number of sensitivity scenarios in less detail. In all scenarios the monitoring and subject locations were sampled from a 4500×4500 grid. Under our primary scenario the true exposure surface was

$$\Phi_j(\mathbf{s}) = \Phi_j^{NS}(\mathbf{s}) + c_{1j}\Phi_1^S(\mathbf{s}) + c_{2j}\Phi_2^S(\mathbf{s}). \tag{4.3}$$

The non-spatial part of the surface $\Phi_j^{NS}(\mathbf{s})$ was generated as $\mathbf{p}(\mathbf{s})^T\boldsymbol{\gamma}_{p,j}$ where at each $\mathbf{s}$ the $\mathbf{p}(\mathbf{s})$ were three fixed realizations of $N(0,1)$ random variables with $\boldsymbol{\gamma}_{p,1}^T = \{0.578, 0.578, 0.578\}$ and $\boldsymbol{\gamma}_{p,1}^T = \{-0.528, -0.528, 0.660\}$. The $\Phi_j^S(\mathbf{s})$ were fixed realizations of a spatially correlated Gaussian process generated using the `spectralGP` package in `R` (**?**) with ranges 5000 and 500, respectively, and $Cor\left(\Phi_1^S(\mathbf{s}), \Phi_2^S(\mathbf{s})\right) = 0.12$. The variance of each $\Phi_j^S(\mathbf{s}) = 1$ and we set $\{c_{11}, c_{21}\} = \{3, 0\}$ and $\{c_{12}, c_{22}\} = \{2.3, 1.9\}$. Thus $Var(\Phi_j(\mathbf{s})) = 10$ and $Cor(\Phi_1(\mathbf{s}), \Phi_2(\mathbf{s})) = 0.70$ 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 $\eta_{i1}^*, \eta_{i2}^* \overset{iid}{\sim} N(0, \sqrt{3})$ while the unobserved exposures at subject locations were generated $x_{ij} = \Phi_j(\mathbf{s}_i) + \eta_{ij}$ with $\eta_{i1}, \eta_{i2} \overset{iid}{\sim} N(0,1)$. Given true exposures the health outcomes were generated as

$$y_i = \beta_0 + \boldsymbol{\beta}^T\mathbf{x}_i + \epsilon_i, \tag{4.4}$$

with $\beta_0 = 0$ and $\epsilon_i \sim N(0,1)$. We considered $\boldsymbol{\beta}^T = \{0.1, 0.5\}$ and $\boldsymbol{\beta}^T = \{0.5, 0.1\}$. We used LRK to separately model each pollutant. We considered models with $4, 8, 12$, or 16 spatial degrees of freedom and no penalization as well as models with $10, 20, 30, 40, 60, 80$, or 100 degrees of freedom with penalization. For the penalized models we chose the λ_j using REML or the EDF criteria described in Section 4.5.3. Each model also included the $\mathbf{p}(\mathbf{s})$ as covariates to model the non-spatial part of the surface. The knots defining the LRK basis functions were chosen using a space-filling from the 4500×4500 grid. In each simulation we assessed out-of-sample R^2, correlation between the predictions, relative biases and standard deviations of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^C$, and actual coverage of 95% Wald confidence intervals using naïve sandwich standard errors or standard errors derived from 100 bootstrap samples.

4.6.2 Sensitivity scenarios

In addition to the primary scenario we considered three 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.

1. We increased the non-spatial proportion of the surface variability to $\approx 50\%$ which decreased correlation between pollutants to 0.30. As in the primary scenario, each $Var(\Phi_j(\mathbf{s})) = 10$ but with $\boldsymbol{\gamma}_{p,1}^T = \{1.29, 1.29, 1.29\}$, $\boldsymbol{\gamma}_{p,2}^T = \{-1.18, -1.18, 1.47\}$, $\{c_{11}, c_{21}\} = \{2.24, 0\}$ and $\{c_{12}, c_{22}\} = \{1.73, 1.41\}$.

2. We induced negative correlation between the pollutants by setting $\{c_{12}, c_{22}\} = \{-2.3, -1.9\}$, implying $Cor(\Phi_1(\mathbf{s}), \Phi_2(\mathbf{s})) = -0.75$.

3. We allowed some monitors to observe only one of the pollutants. In addition to $n_0^* = 200$ we sampled $n_1^* = 300$ and $n_2^* = 100$ monitors uniformly from the grid.

For these scenarios we only considered unpenalized LRK models with 12 degrees of freedom or penalized LRK models with 30 degrees of freedom. We assessed predictive accuracy,

pollutant correlation, bias, standard errors, and 95% confidence coverage as for the primary scenario.

4.6.3 Simulation results: primary scenario

Figures 4.1 and 4.2 show the simulation results when $\boldsymbol{\beta}^T = \{0.5, 0.1\}$ and $\boldsymbol{\beta}^T = \{0.1, 0.5\}$, respectively. Table 4.1 shows a subset of the results using 12 DF in each unpenalized model and 30 DF in each penalized model. For all exposure models mean out-of-sample R^2s were between 0.70 and 0.85 for x_{i1} while they were never higher than 0.70 and as low as 0.50 for x_{i2}. 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 4.1 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 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. Table 4.1 shows that 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.

When $\beta_1 > \beta_2$, Figure 4.2 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 4.2, where increasing the exposure model degrees of freedom improves out-of-sample R^2 but worsens the biases of both health effect estimates. Our results also advocate for use of REML to penalize the exposure model. The REML models yield the best efficiency and 95% confidence interval coverage even when using many degrees of freedom to model exposure.

4.6.4 Simulation results: sensitivity scenarios

Results of the sensitivity scenarios described in Section 4.6.2 are shown in Tables 4.2–4.4. 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 4.3.

There are however some interesting differences between the primary and sensitivity results. When 50% of the exposure surface was explained by the non-spatial component Table 4.2 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 4.3 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 4.4 are similar to the primary results, and illustrate the effectiveness of our bias correction when both pollutants are not observed at all monitoring locations.

4.7 Associations of SBP with $PM_{2.5}$ and NO_2 in the Sister Study

We applied our measurement error methodology in analyzing the association of SBP with $PM_{2.5}$ 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.

4.7.1 Previous analysis

Chan et al. (Subm) analyzed the association between baseline SBP and predicted annual 2006 average $PM_{2.5}$ and NO_2 in 43,629 Sister Study participants in the continental United States. $PM_{2.5}$ predictions were derived from a regionalized universal kriging model (Sampson et al., 2013) and NO_2 predictions were derived from a national sattelite-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-μg/m^3 increase in $PM_{2.5}$ (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; $p = 0.10$). When modeling both exposures in the health model as main effects they estimated a 1.6 mmHg change in SBP for a 10-μg/m^3 increase in $PM_{2.5}$ 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 $PM_{2.5}$ constant (95% CI: -0.4, 0.3; $p = 0.65$). 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 (DF) 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. These analyses did not account for measurement error.

4.7.2 Measurement error analysis: 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 NO_2, and 178 monitors that measured both implying we had 1037 total $PM_{2.5}$ observations and 358 total NO_2 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).

We also considered a sensitivity analysis restricted to the nine northeastern states of Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New Jersey, New York and Pennsylvania. We restricted to this region to eliminate large-scale spatial structure in systolic blood pressure. In these nine states we had 106 $PM_{2.5}$-only monitors, 20 NO_2-only monitors, and 39 monitors that measured both pollutants. Figure 4.3 shows the monitoring locations across the continental US and the northeastern regions.

We modeled each pollutant separately using LRK models with 50, 75, or 100 DF and penalty parameters chosen by REML. Exposure models in the northeastern region used 10, 15, or 20 DF. We used the same DF to model each pollutant, and selected knots using a space-filling algorithm over a grid of 25km×25km cells. We used a range parameter of 6363km for the national models 1285km in the northeastern region. These corresponded to the maximum distance between any 25km×25km grid cell over the region. Our exposure models also included 2 partial least squares (PLS) components to efficiently capture inform-

ation 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 via 10-fold cross-validation (Hastie et al., 2001).

4.7.3 Measurement error analysis: Health models

We used the predicted exposures to estimate pollutant associations with SBP. We used the same 43,629 participants as Chan et al. (Subm) in the primary national analyses and 7427 participants residing in the nine northeastern states for the sensitivity analyses. To adjust for large-scale spatial structure in the national models we considered thin-plate regression splines with 5, 10, and 15 DF. 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 to adjust for spatial confounding in the sensitivity analyses. The other adjustment variables listed in Section 4.7.1 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 on the national scale, though we did not apply any measurement error correction.

4.7.4 Results

Figure 4.4 shows smoothed associations between SBP and exposure in the national analysis controlling for all other health model variables (including the other exposure for the joint models), using 100 DF in the LRK exposure models and a 10-DF thin plate regression spline to control for spatial confounding. When modeled by itself the association between SBP

and $PM_{2.5}$ was notably weaker and less linear than when NO_2 was included in the health model. When controlled for $PM_{2.5}$ the association between SBP and NO_2 appeared more negative than when NO_2 was modeled individually.

Figure 4.5 shows complete national results, including cross-validated R^2 and correlation between predicted pollutants. The models performed equally well for both pollutant, with R^2 between 0.75 and 0.78. The predictions were moderately correlated, between 0.42 and 0.44. Correlations between our LRK $PM_{2.5}$ predictions and those used by Chan et al. (Subm) were between 0.94 and 0.96, while correlations between our LRK NO_2 predictions and those used by Chan et al. (Subm) were between 0.86 and 0.88. When modeled individually only the 100- and 150-DF exposure models yielded significant estimated associations between SBP and $PM_{2.5}$, and this only when the health models used 10 DF to adjust for spatial confounding. We estimated a significant negative univariate association between SBP and NO_2 only when using 15 DF to control for spatial confounding; all other models yielded essentially null effect estimates. In all univariate analyses there was no meaningful estimated bias from measurement error and the bootstrap standard error estimates were very similar to the naïve standard error estimates.

We saw very different results when the pollutants were modeled jointly. There were strong positive associations between SBP and $PM_{2.5}$ controlling for NO_2, and strong negative associations between SBP and NO_2 controlling for $PM_{2.5}$. All associations were much stronger than those described by Chan et al. (Subm). The 10-DF thin plate regression spline appeared to be the best adjustment for spatial confounding. There appeared to be some bias toward the null from residual spatial confounding in both effect estimates using only 5 DF while 15 DF led to similar effect estimates as the 10-DF models and inflated standard errors. 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 standard error estimates or bootstrap standard error estimates without the bias correction.

Figure 4.6 shows results from the sensitivity analysis restricted to the nine northeastern states. Prediction accuracy was higher than the national models with 10-fold cross-validated R^2 between 0.82 and 0.89. With a correlation of 0.72 the predictions were also more correlated than on the national scale. The univariate estimated associations between SBP and $PM_{2.5}$ were all 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. The joint associations were qualitatively very similar to the national models. We again saw strong positive associations of SBP with $PM_{2.5}$ adjusted for NO_2 and strong negative associations of SBP with NO_2 adjusted for $PM_{2.5}$. As opposed to the national analysis the estimated biases were quite small and slightly away from the null.

Figure 4.7 shows the national interactive association of SBP with $PM_{2.5}$ and NO_2 using the 100-DF exposure models and 10 DF to control for spatial confounding. Although the p-value for the interaction terms from all exposure models were between 0.01 and 0.02, Figure 4.7 represents an association that does not drastically differ from additive. This is most easily explored by traveling horizontally across Figure 4.7 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.

4.8 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 4.1.

As we saw in our 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.

Having a mechanism for bias correction is especially relevant in analyses such as our NIEHS Sister Study data analysis. We saw clear evidence of NO_2 confounding the association between systolic blood pressure and $PM_{2.5}$, and vice versa. To accurately estimate these associations we needed to adjust for both pollutants. In doing so we induced multi-pollutant measurement error which our simulations demonstrated could severely bias the health effect estimates. Indeed when adjusting for $PM_{2.5}$ and NO_2 together 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. 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 $PM_{2.5}$ 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 4.7.4. One possibility is that there is still unmeasured confounding responsible for our results. Another is that the results reflect reality, in which case more needs to be done to investigate possible mechanisms for these negative associations.

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 (sub) 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., 2011b; Bergen et al., 2013). In the multi-pollutant setting we can not rely on improving prediction accuracy to achieve unbiased estimation, as our simulations showed that increasing the number of exposure model degrees of freedom could improve out-of-sample R^2 while increasing bias.

We note that throughout we have assumed that $H_j(\cdot) = G(\cdot)$ for $j = 0, 1, 2$. In reality the $H_j(\cdot)$ are unlikely to be equal; for example monitors that measure traffic-related pollutants are more likely to be sited near roadways than monitors of general air quality. Additionally each $H_j(\cdot)$ likely differs from the distribution of subject locations. Two implications follow from letting $H_j(\cdot)$ differ from each other and from $G(\cdot)$. The first is that if Condition 1 from Section 4.4.2 is violated, Condition 2 cannot ensure unbiased estimation even with large n^* and $\mathbf{\Lambda} = 0$. This is because $\boldsymbol{\gamma}$ is no longer an ordinary least squares solution under $G(\cdot)$ implying $\left(\Phi_j(\mathbf{s}_i) - \mathbf{r}_j(\mathbf{s}_i)^T\boldsymbol{\gamma}\right)$ is not guaranteed to be orthogonal to $\Phi_j(\mathbf{s}_i)$ under $G(\cdot)$. The second implication is that we cannot estimate $\mathbf{U}(\boldsymbol{\gamma})$ by summing over the empirical

distributions of monitoring locations, 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 (Olives, In preparation). 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 $j = 0, 1, 2$. In future work we will expand our developed methodology to include analyzing the spatial patterns of $g(\cdot)/h_j(\cdot)$, their impact on measurement error, and methods for estimating this ratio in order to estimate bias.

	$\beta_1 = 0.1$				$\beta_2 = 0.5$			
	RB	SD	SE	Cov	RB	SD	SE	Cov
No penalty; $R^2(x_1) = 0.75$; $R^2(x_2) = 0.57$; $Cor(w_1, w_2) = 0.76$								
No correction	0.35	0.52	0.25	0.53	-0.11	0.11	0.11	0.41
Bias correction only	-0.09	0.90	0.25	0.54	0.03	0.23	0.23	0.51
Bootstrap SE only	0.35	0.52	0.54	0.89	-0.11	0.11	0.11	0.79
Bias correction + bootstrap	-0.09	0.90	0.89	0.95	0.03	0.23	0.23	0.96
REML; $R^2(x_1) = 0.81$; $R^2(x_2) = 0.62$; $Cor(w_1, w_2) = 0.75$								
No correction	0.35	0.49	0.24	0.54	-0.05	0.11	0.11	0.57
Bias correction only	-0.01	0.63	0.24	0.60	0.00	0.14	0.14	0.56
Bootstrap SE only	0.35	0.49	0.48	0.86	-0.05	0.11	0.11	0.87
Bias correction + bootstrap	-0.01	0.63	0.59	0.94	0.00	0.14	0.14	0.93
EDF; $R^2(x_1) = 0.75$; $R^2(x_2) = 0.55$; $Cor(w_1, w_2) = 0.73$								
No correction	0.41	0.70	0.29	0.48	0.14	0.17	0.17	0.45
Bias correction only	0.07	0.62	0.29	0.64	-0.02	0.14	0.14	0.64
Bootstrap SE only	0.41	0.70	0.70	0.87	0.14	0.17	0.17	0.90
Bias correction + bootstrap	0.07	0.62	0.60	0.93	-0.02	0.14	0.14	0.90
	$\beta_1 = 0.5$				$\beta_2 = 0.1$			
	RB	SD	SE	Cov	RB	SD	SE	Cov
No penalty; $R^2(x_1) = 0.75$; $R^2(x_2) = 0.57$; $Cor(w_1, w_2) = 0.76$								
No correction	-0.05	0.10	0.04	0.57	0.10	0.52	0.52	0.61
Bias correction only	0.01	0.13	0.04	0.51	-0.02	0.69	0.69	0.50
Bootstrap SE only	-0.05	0.10	0.10	0.92	0.10	0.52	0.52	0.95
Bias correction + bootstrap	0.01	0.13	0.14	0.97	-0.02	0.69	0.69	0.97
REML; $R^2(x_1) = 0.81$; $R^2(x_2) = 0.62$; $Cor(w_1, w_2) = 0.75$								
No correction	-0.03	0.09	0.04	0.61	0.27	0.48	0.48	0.57
Bias correction only	0.00	0.10	0.04	0.56	0.02	0.58	0.58	0.57
Bootstrap SE only	-0.03	0.09	0.09	0.92	0.27	0.48	0.48	0.88
Bias correction + bootstrap	0.00	0.10	0.10	0.94	0.02	0.58	0.58	0.93
EDF; $R^2(x_1) = 0.75$; $R^2(x_2) = 0.55$; $Cor(w_1, w_2) = 0.73$								
No correction	0.08	0.12	0.05	0.51	0.77	0.71	0.71	0.35
Bias correction only	-0.01	0.11	0.05	0.64	0.08	0.61	0.61	0.63
Bootstrap SE only	0.08	0.12	0.12	0.92	0.77	0.71	0.71	0.78
Bias correction + bootstrap	-0.01	0.11	0.10	0.93	0.08	0.61	0.61	0.94

Table 4.1: Subset of primary results. Detail of Figures 4.1 and 4.2 using 24 basis functions with no penalty and 60 knots with penalty. "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^2s for each pollutant and mean correlation between predictions is given.

	$\beta_1 = 0.1$				$\beta_2 = 0.5$			
	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	0.07	0.55
Bias correction only	0.01	0.33	0.15	0.63	0.01	0.07	0.07	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	0.07	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	0.07	0.61
Bias correction only	0.01	0.30	0.14	0.68	-0.01	0.07	0.07	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	0.07	0.58
Bias correction only	0.01	0.30	0.15	0.67	-0.01	0.07	0.07	0.62
Bootstrap SE only	0.20	0.31	0.32	0.90	0.03	0.07	0.07	0.92
Bias correction + bootstrap	0.01	0.30	0.31	0.94	-0.01	0.07	0.07	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	0.3	0.64
Bias correction only	0.00	0.06	0.03	0.62	0.01	0.32	0.32	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	0.32	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	0.29	0.64
Bias correction only	-0.01	0.05	0.03	0.64	0.01	0.29	0.29	0.65
Bootstrap SE only	0.00	0.05	0.05	0.95	0.10	0.29	0.29	0.92
Bias correction + bootstrap	-0.01	0.05	0.06	0.95	0.01	0.29	0.29	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	0.31	0.54
Bias correction only	-0.01	0.06	0.03	0.64	0.01	0.30	0.30	0.66
Bootstrap SE only	0.03	0.06	0.06	0.92	0.22	0.31	0.31	0.89
Bias correction + bootstrap	-0.01	0.06	0.05	0.95	0.01	0.30	0.30	0.94

Table 4.2: Sensitivity scenario; proportion of explainable variability due to geographic covariates equals 50%, implying $Cor(\Phi_1(\mathbf{s}), \Phi_2(\mathbf{s})) = 0.30$. "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^2s for each pollutant and mean correlation between predictions is given.

	$\beta_1 = 0.1$				$\beta_2 = 0.5$			
	RB	SD	SE	Cov	RB	SD	SE	Cov
No penalty; $R^2(x_1) = 0.75$; $R^2(x_2) = 0.57$; $Cor(w_1, w_2) = -0.81$								
No correction	-0.67	0.54	0.27	0.37	-0.17	0.12	0.12	0.28
Bias correction only	0.20	1.07	0.27	0.47	0.05	0.24	0.24	0.47
Bootstrap SE only	-0.67	0.54	0.53	0.71	-0.17	0.12	0.12	0.62
Bias correction + bootstrap	0.20	1.07	8.91	0.94	0.05	0.24	0.24	0.94
REML; $R^2(x_1) = 0.81$; $R^2(x_2) = 0.62$; $Cor(w_1, w_2) = -0.81$								
No correction	-0.61	0.51	0.26	0.38	-0.12	0.11	0.11	0.42
Bias correction only	0.01	0.81	0.26	0.50	0.01	0.18	0.18	0.49
Bootstrap SE only	-0.61	0.51	0.47	0.68	-0.12	0.11	0.11	0.73
Bias correction + bootstrap	0.01	0.81	0.67	0.90	0.01	0.18	0.18	0.90
EDF; $R^2(x_1) = 0.74$; $R^2(x_2) = 0.54$; $Cor(w_1, w_2) = -0.81$								
No correction	-0.24	0.80	0.33	0.55	0.10	0.19	0.19	0.58
Bias correction only	-0.14	0.73	0.33	0.61	-0.03	0.17	0.17	0.62
Bootstrap SE only	-0.24	0.80	0.74	0.89	0.10	0.19	0.19	0.93
Bias correction + bootstrap	-0.14	0.73	0.65	0.90	-0.03	0.17	0.17	0.89
	$\beta_1 = 0.5$				$\beta_2 = 0.1$			
	RB	SD	SE	Cov	RB	SD	SE	Cov
No penalty; $R^2(x_1) = 0.75$; $R^2(x_2) = 0.57$; $Cor(w_1, w_2) = -0.81$								
No correction	-0.11	0.10	0.05	0.41	-0.48	0.49	0.49	0.48
Bias correction only	0.03	0.17	0.05	0.51	0.14	0.87	0.87	0.49
Bootstrap SE only	-0.11	0.10	0.10	0.75	-0.48	0.49	0.49	0.80
Bias correction + bootstrap	0.03	0.17	0.32	0.94	0.14	0.87	0.87	0.95
REML; $R^2(x_1) = 0.81$; $R^2(x_2) = 0.62$; $Cor(w_1, w_2) = -0.81$								
No correction	-0.09	0.08	0.04	0.43	-0.54	0.44	0.44	0.42
Bias correction only	-0.01	0.12	0.04	0.54	-0.03	0.65	0.65	0.53
Bootstrap SE only	-0.09	0.08	0.08	0.75	-0.54	0.44	0.44	0.69
Bias correction + bootstrap	-0.01	0.12	0.11	0.92	-0.03	0.65	0.65	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	0.68	0.45
Bias correction only	-0.03	0.11	0.06	0.64	-0.16	0.62	0.62	0.65
Bootstrap SE only	0.04	0.12	0.12	0.95	-0.59	0.68	0.68	0.81
Bias correction + bootstrap	-0.03	0.11	0.11	0.90	-0.16	0.62	0.62	0.91

Table 4.3: Sensitivity scenario; $Cor(\Phi_1(\mathbf{s}), \Phi_2(\mathbf{s})) = -0.75$. "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^2s for each pollutant and mean correlation between predictions is given.

	$\beta_1 = 0.1$				$\beta_2 = 0.5$			
	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	0.10	0.53
Bias correction only	-0.10	0.72	0.26	0.52	0.02	0.15	0.15	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	0.15	0.96
REML; $R^2(x_1) = 0.84$; $R^2(x_2) = 0.64$; $Cor(w_1, w_2) = 0.76$								
No correction	0.29	0.42	0.24	0.62	-0.05	0.09	0.09	0.63
Bias correction only	-0.05	0.64	0.24	0.51	0.01	0.13	0.13	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	0.13	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	0.15	0.48
Bias correction only	0.01	0.73	0.29	0.56	-0.01	0.15	0.15	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	0.15	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	0.37	0.78
Bias correction only	0.01	0.12	0.05	0.59	-0.08	0.74	0.74	0.5
Bootstrap SE only	-0.02	0.07	0.07	0.94	0.01	0.37	0.37	0.95
Bias correction + bootstrap	0.01	0.12	0.18	0.96	-0.08	0.74	0.74	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	0.33	0.77
Bias correction only	0.00	0.10	0.04	0.61	-0.04	0.59	0.59	0.55
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	0.59	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	0.51	0.51
Bias correction only	0.00	0.12	0.05	0.65	-0.06	0.73	0.73	0.62
Bootstrap SE only	0.10	0.08	0.09	0.79	0.54	0.51	0.51	0.82
Bias correction + bootstrap	0.00	0.12	0.15	0.96	-0.06	0.73	0.73	0.96

Table 4.4: Sensitivity scenario; $n_C^* = 200$; $n_1^* = 300$; $n_2^* = 100$. "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^2s for each pollutant and mean correlation between predictions is given.

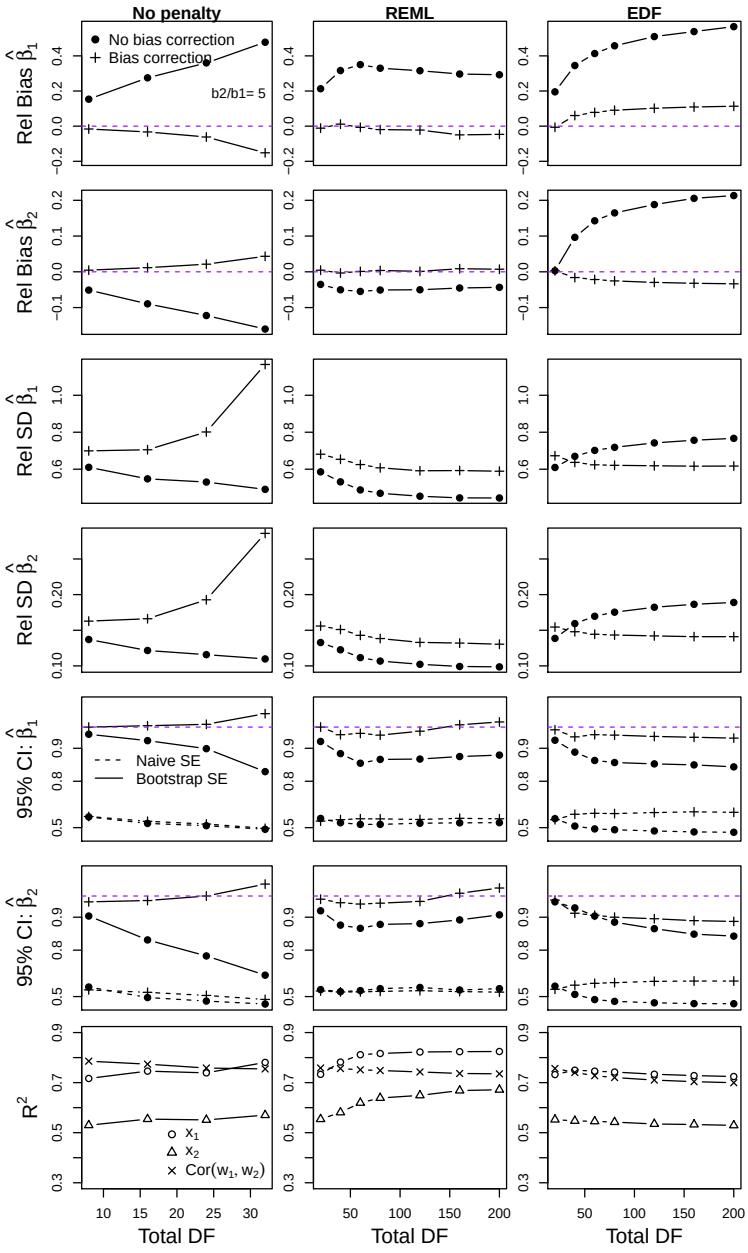


Figure 4.1: Primary simulation results, with $\beta_1 = 0.1$ and $\beta_2 = 0.5$. Relative biases and standard deviations are shown both with and without a bias correction. Also shown are actual coverages of 95% confidence intervals both with and without a bias correction, and with naïve or non-parametric bootstrap standard errors. The bottom row shows mean out-of-sample R^2 and prediction correlation.

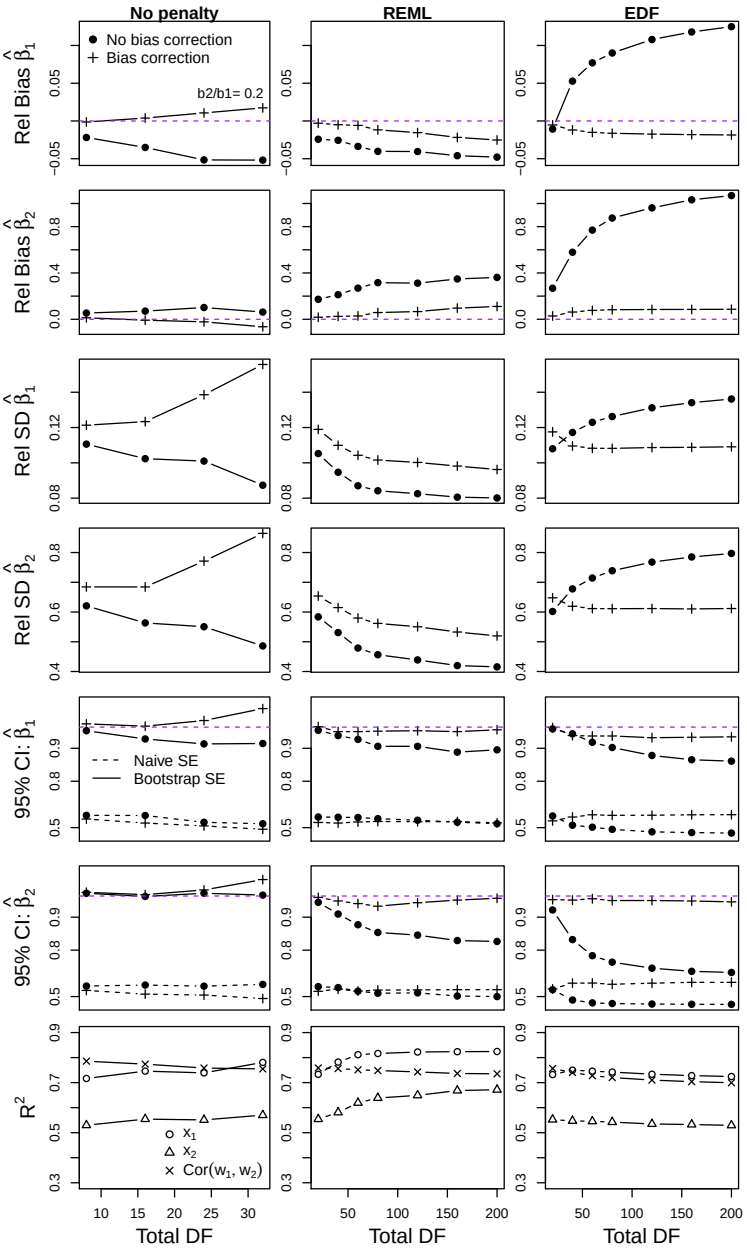


Figure 4.2: Primary simulation results, with $\beta_1 = 0.5$ and $\beta_2 = 0.1$. Relative biases and standard deviations are shown both with and without a bias correction. Also shown are actual coverages of 95% confidence intervals both with and without a bias correction, and with naïve or non-parametric bootstrap standard errors. The bottom row shows mean out-of-sample R^2 and prediction correlation.

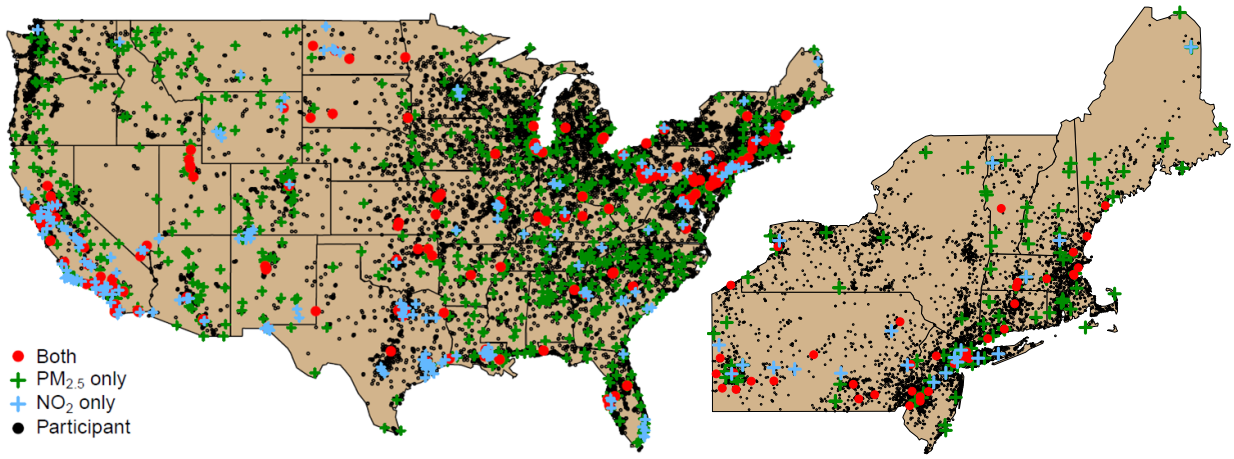


Figure 4.3: Map of monitoring and Sister Study participant locations, both for the entire nation and the nine upper northeastern states used in the sensitivity analysis.

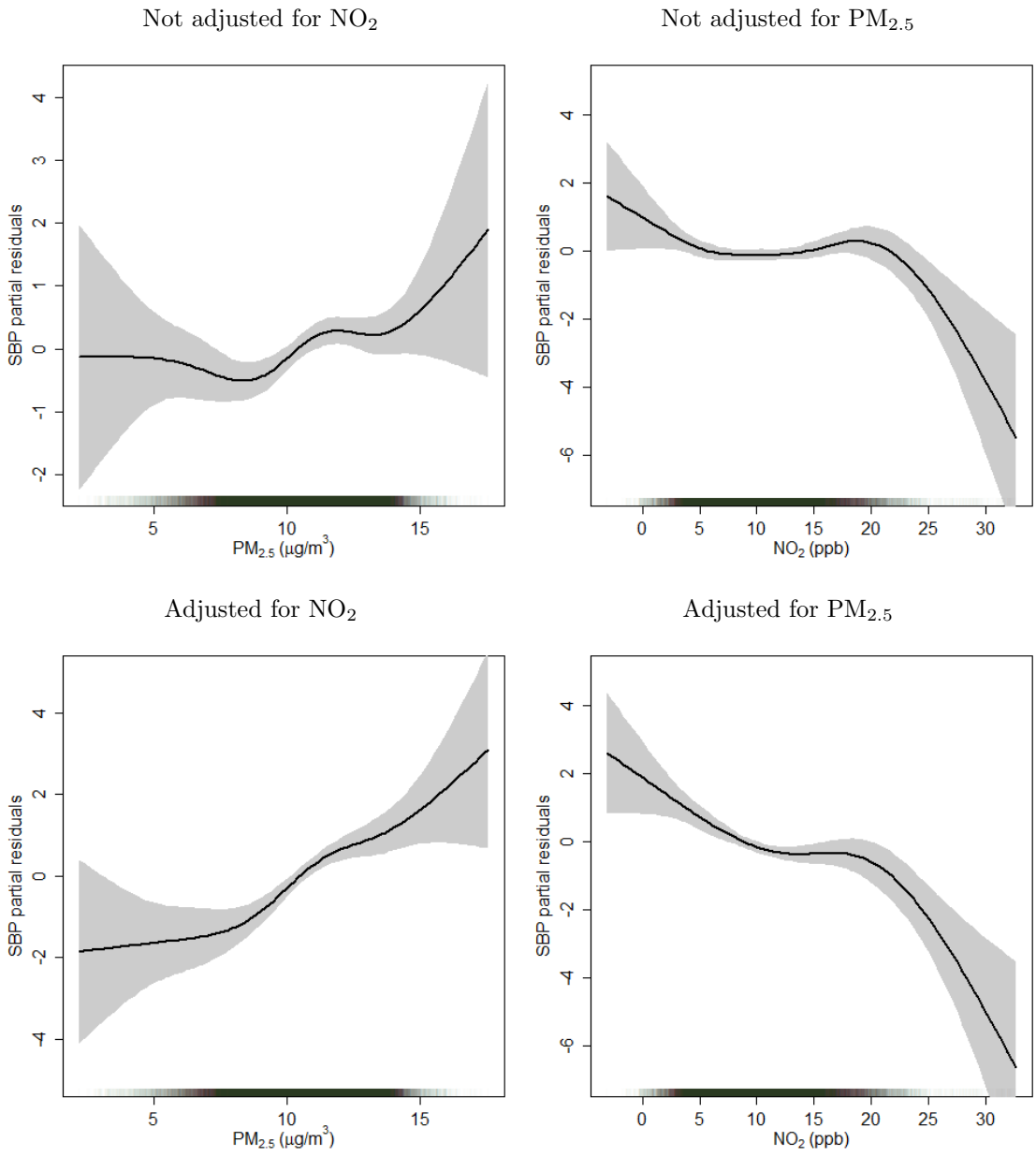


Figure 4.4: SBP partial residuals as smooth functions of predicted $PM_{2.5}$ or NO_2 (using a 5-DF thin-plate regression spline to flexibly model exposure), when SBP is modeled as a function of each pollutant separately or when adjusting for the other pollutant. These figures correspond to predictions derived from exposure models using 100 DF, and spatial confounding controlled by a 10 DF thin plate regression spline.

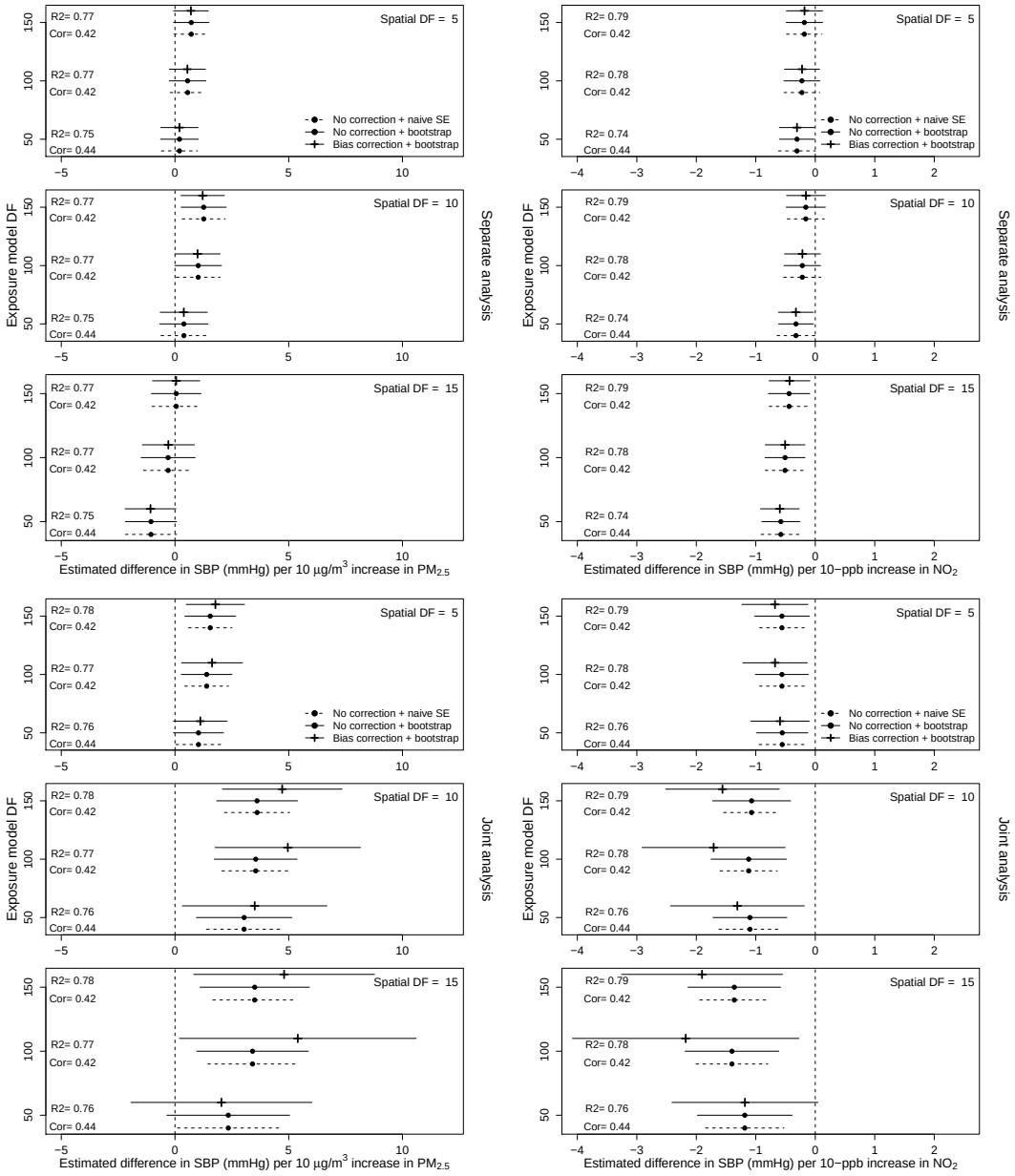


Figure 4.5: Estimated change in SBP (mmHg) for a 10-μg/m^3 increase in $PM_{2.5}$ holding NO_2 constant, and estimated change in SBP (mmHg) for a 10-ppb increase in NO_2 holding $PM_{2.5}$ constant. The first row corresponds to modeling SBP as a function of each pollutant separately, while the second row corresponds to regressing SBP on both pollutants. Both bias-corrected and uncorrected estimates are shown along with naïve sandwich and bootstrap standard errors. Also shown are cross-validated R^2 and correlation between predicted $PM_{2.5}$ and NO_2.

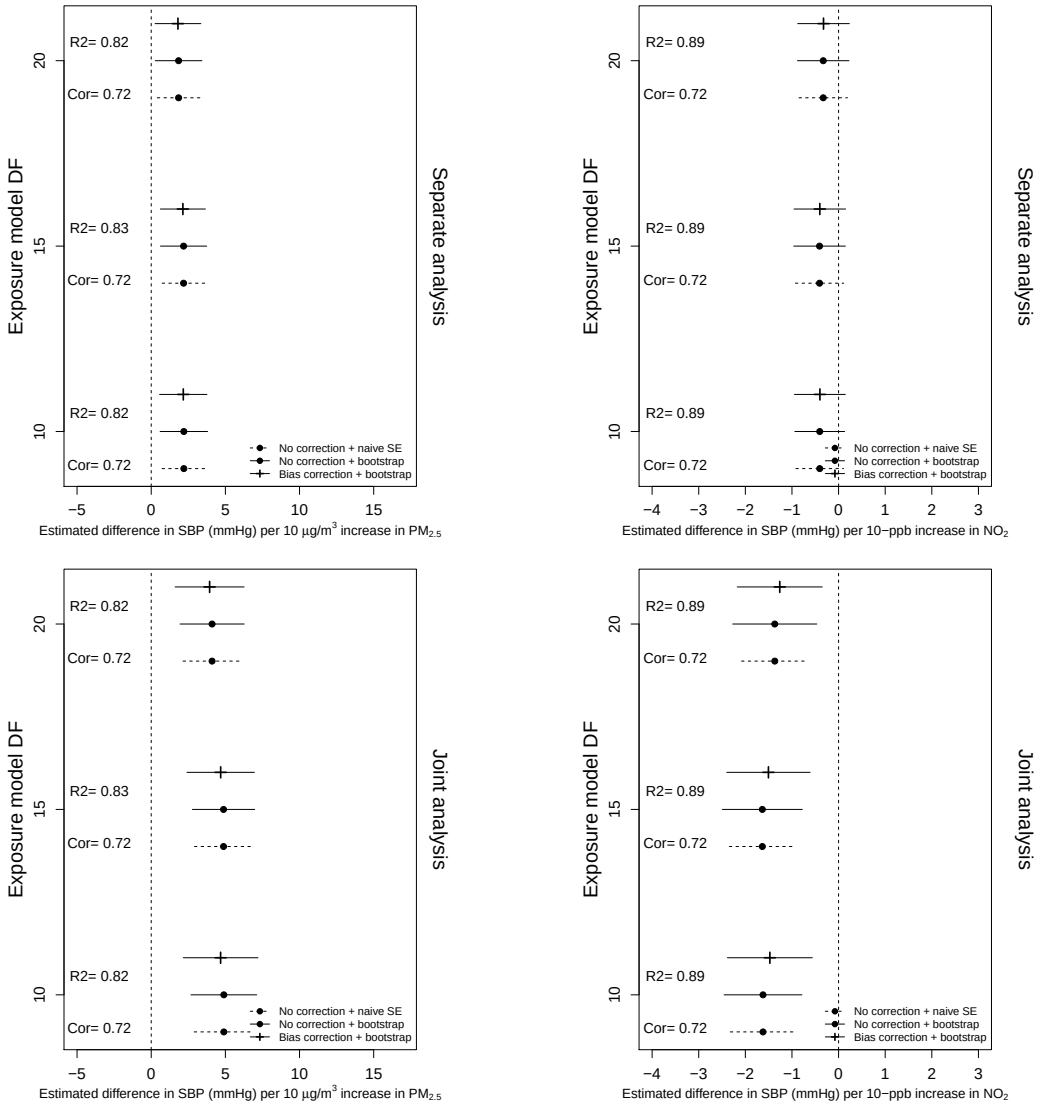


Figure 4.6: Analysis restricted to nine northeastern states. Estimated change in SBP (mmHg) for a 10-μg/m^3 increase in $PM_{2.5}$ holding NO_2 constant, and estimated change in SBP (mmHg) for a 10-ppb increase in NO_2 holding $PM_{2.5}$ constant. The first row corresponds to modeling SBP as a function of each pollutant separately, while the second row corresponds to modeling SBP on both pollutants. Both bias-corrected and uncorrected estimates are shown along with naïve sandwich and bootstrap standard errors. Also shown are cross-validated R^2 and correlation between predicted $PM_{2.5}$ and NO_2.

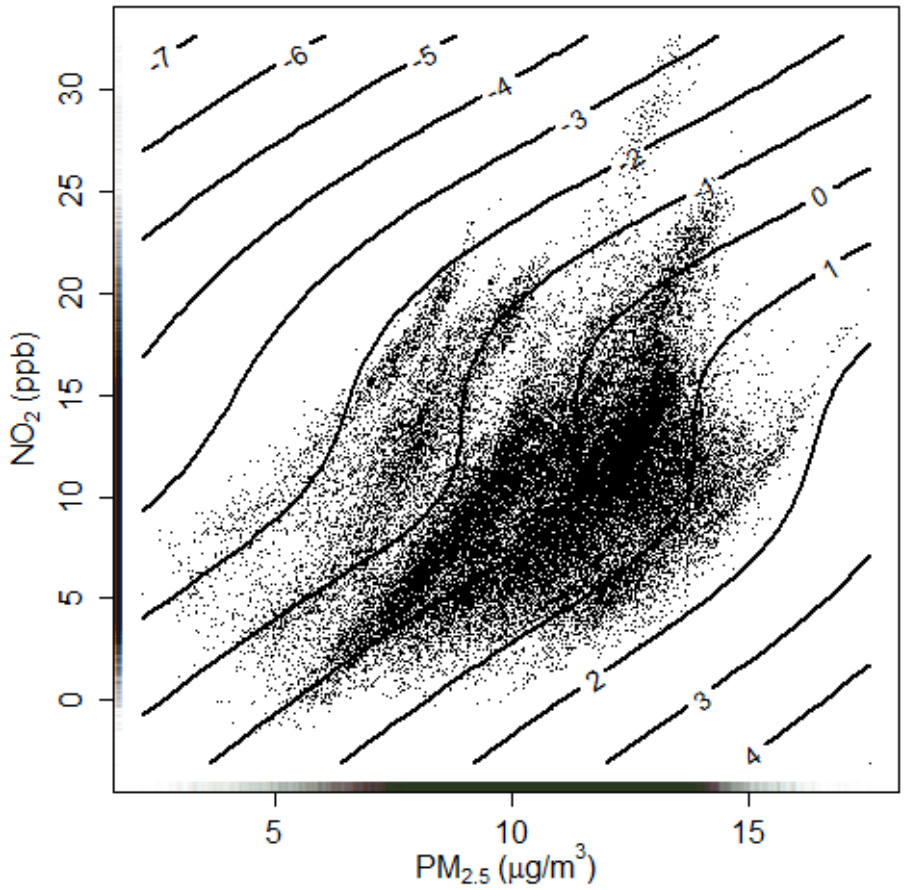


Figure 4.7: SBP partial residuals (contours) as a smooth function of $PM_{2.5}$ and NO_2. Predicted exposures shown were using 100 degrees of freedom in the exposure model.

Appendix A

APPENDIX FOR CHAPTER 2

A.1 Joint exposure and health modeling

The parametric and parameter bootstraps are justified by assuming the spatial surface is a spatially correlated Gaussian process. As the health outcomes are also assumed normal we could consider the implied joint likelihood and estimate the parameters of the health and exposure models jointly. The joint approach fully accounts for uncertainty in not observing $\mathbf{X}$ and allows information in the health data to influence exposure model estimation. Because of this we may stand to gain efficiency by fitting joint models. In what follows we outline two joint approaches to estimating β: maximizing the joint likelihood and summarizing the posterior obtained from a Bayesian model. We apply both methods to our case study analyzing associations of EC, OC, Si and S with CIMT in the MESA Air cohort and compare the results to the two-stage bootstrap approaches described in Chapter 2.

A.1.1 Joint maximum likelihood

The joint likelihood is implied by Equations 2.1 and 2.3, and a simplified version is outlined in Madsen et al. (2008). Jointly the observed data follow:

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{X}^* \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_0 + \beta \mathbf{P}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta}_Z \\ \mathbf{P}^*\boldsymbol{\alpha} \end{pmatrix}, \begin{pmatrix} \beta^2\Sigma_{\eta\eta} + \sigma_\epsilon^2\mathbf{I} & \beta\Sigma_{\eta\eta^*} \\ \beta\Sigma_{\eta\eta^*}^T & \Sigma_{\eta^*\eta^*} \end{pmatrix} \right). \tag{A.1}$$

Maximizing the implied likelihood with respect to β, $\boldsymbol{\theta}$, σ_ϵ^2, $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}_Z$ implicitly accounts for the variability inherent in not observing $\mathbf{X}$. We can obtain estimated standard errors of these parameter estimates using the inverse Hessian of the log-likelihood evaluated at the estimated parameters.

A.1.2 Bayesian model

An alternative joint approach is to fit a Bayesian model, and obtain parameter estimates as summaries of the Bayesian posterior. The posterior is:

$$p(\beta_0, \beta_1, \boldsymbol{\beta}_z, \tau_\epsilon, \boldsymbol{\alpha}, \boldsymbol{\theta'} | \mathbf{Y}, \mathbf{X}^*) \propto L(\mathbf{Y}, \mathbf{X}^* | \beta_0, \beta_1, \boldsymbol{\beta}_z, \tau_\epsilon, \boldsymbol{\alpha}, \boldsymbol{\theta'}) \times \pi(\beta_0, \beta_1, \boldsymbol{\beta}_z, \sigma^2_\epsilon, \boldsymbol{\alpha}, \boldsymbol{\theta'}), \quad \text{(A.2)}$$

where $L(\cdot)$ is the likelihood implied by (A.1), $\boldsymbol{\theta'}$ is a reparameterization of $\boldsymbol{\theta}$ described below, and $\pi(\cdot)$ denotes the prior distribution of the model parameters. We fit this model using the stochastic partial differential equation (SPDE) model formulation in R-INLA (Rue et al., 2009). Briefly, the SPDE model represents continuous space as a refined Delaunay triangulation, where each location is represented by a linear combination of edges of the grid triangle it resides in. See Lindgren et al. (2011) for more details. Rather than specifying priors for σ^2_ϵ and $\boldsymbol{\theta'}$, the internal representation of these parameters in R-INLA is τ_ϵ and $\boldsymbol{\theta'}$ where

$$\begin{aligned}
\tau_\epsilon &= -2\log(\sigma_\epsilon), \\
\theta'_1 &= \frac{1}{2}\log\left(\frac{\Gamma(\nu)}{\Gamma(\kappa)4\pi}\right) + \nu\log(\phi) - \log(\sigma_\eta), \\
\theta'_2 &= \frac{\log(8\nu)}{2} - \log(\phi), \\
\theta'_3 &= -\log(\tau).
\end{aligned}$$

Recall that ϕ and σ^2_η denote the range and partial sill of the Gaussian process while τ^2 is the nugget. The parameters κ and $\nu = \kappa - 1$ are fixed and specify variants of the Matérn covariance. We fit an exponential covariance to remain consistent with our two-stage and joint maximum-likelihood approaches, corresponding to $\kappa = 3/2$ and $\nu = 1/2$. Thus there is a one-to-one relationship between an R-INLA internal parameter and the nugget, range,

and σ_ϵ^2; while θ_1' is a function of both the range and the partial sill.

We used independent priors on the model parameters, as follows:

$$\begin{aligned}
\beta &\sim N(0, 0.001), \\
\tau_\epsilon &\sim loggamma(1, 0.00005), \\
\theta_1' &\sim N(4.9, 0.001), \\
\theta_2' &\sim N(-6.1, 0.001), \\
\theta_3' &\sim N(\mu_{\theta_3}, 0.001), \\
\beta_0 &\sim N(0, 0.001), \\
\boldsymbol{\beta}_Z &\sim N(0, 0.001\mathbf{I}_{12}), \\
\alpha_j &\overset{iid}{\sim} N(0, 0.001).
\end{aligned}$$

Here the normal distributions are parameterized by a mean and *precision*, or variance^{-1}. Accordingly, precisions of 0.001 translate to variances of 1000, or very wide priors. The prior for τ_ϵ is the default in R-INLA, as are the prior means for θ_1' and θ_2'. As a sensitivity analysis we perturbed μ_{θ_3}, the prior mean of θ_3', around $\log(10)/2$ by considering $\mu_{\theta_3} = \log(10)/2 + c$ for $c \in \{-2, -1, 0, 1, 2\}$. For each perturbation we obtained Bayesian estimates of β, $\boldsymbol{\theta}$ and σ_ϵ^2 by summarizing the marginal posteriors of appropriate transformations of $\{\beta, \boldsymbol{\theta}', \tau_\epsilon\}$.

A.1.3 Results

Parameter estimates

The health effect estimates $\hat{\beta}$, estimated standard errors, and 95% confidence or Bayesian credible intervals are shown in Table A.1. Figures A.1–A.4 show the marginal posteriors for β, $\boldsymbol{\theta}$ and σ_ϵ^2 ordered by each perturbation $c \in \{-2, -1, 0, 1, 2\}$. Superimposed on these figures are the corresponding estimates from the two-stage and joint MLE approaches.

We see different estimated health effects depending on the modeling approach used. Table A.1 shows that maximizing the joint likelihood led to estimated health effects that were noticeably stronger than the corresponding two-stage estimates, and smaller standard errors. The estimated association between CIMT and EC became statistically significant and negative when maximizing the joint likelihood, while joint MLE led to stronger positive associations for the other three pollutants.

Most striking is the sensitivity of the Bayesian results to the specified prior means of θ_3. This sensitivity was strongest for EC, where varying the prior mean of θ_3 led to drastically different qualitative conclusions. The estimated association between CIMT and EC was essentially null for $c = -2$, strong positive for $c =\in \{-1, 2\}$ and strong negative for $c \in \{0, 1\}$. The OC analysis was less sensitive qualitatively, though Figure A.2 shows that the shapes of the posterior densities of β were sensitive to c. The estimated association of CIMT with Si was drastically different for $c = 0$ than for any other c, while the estimated association with S appeared insensitive to c. Figures A.1–A.4 also show that the marginal posteriors of β were not the only ones to be affected by c; the marginal posteriors of $\boldsymbol{\theta}$ and σ^2_ϵ were also affected. Also of note is that although it was the prior mean of θ'_3 (corresponding to a transformation of the nugget) we perturbed, these perturbations did not always lead to differences in the marginal posteriors of the nugget. Furthermore the marginal posteriors of β and the nugget did not necessarily vary together, as can be seen clearly in Figure A.1 for $c \in \{-1, 0, 1, 2\}$, among other places.

Model checking

We carried out a number of model-checking procedures to investigate possible reasons for the drastic differences between the joint MLE and Bayesian approaches and the sensitivity of the Bayesian approaches to the prior specification of θ'_3. As a sanity check we specified priors centered at the joint MLEs with precisions equal to $100,000$, and returned the joint MLEs as expected. We also refined the Delaunay triangulation to specify a finer spatial

grid, but this did not noticeably change our results. Running R-INLA with 21 instead of 9 evaluation points also resulted in identical results to what is shown in Figures A.1–A.4.

Although all Hessians obtained via joint MLE were negative-definite, we sought further assurance that the joint MLEs truly maximized the likelihood. We evaluated the profile likelihood at all combinations of $\hat{\boldsymbol{\theta}}(c)$, $\hat{\sigma}^2_\epsilon(c)$, and $\hat{\beta}(c)$ for $c \in \{-2, -1, 0, 1, 2\}$, where $\hat{\boldsymbol{\theta}}(c)$ denotes the marginal posterior means of the $\boldsymbol{\theta}$ estimated with given perturbation c; analogously for $\hat{\sigma}^2_\epsilon(c)$ and $\hat{\beta}(c)$. Thus we evaluated the likelihood $5 \times 5 \times 5$ times, and we compared the likelihood at these values to the likelihood evaluated at the joint MLE. We also evaluated the prior $\times$ likelihood at these same parameter values, yielding a scaled version of the posterior. This allowed us to investigate if the scaled posterior differed noticeably from the likelihood in how it varied for different parameter values; if it did this might shed light on why the posteriors were so sensitive to choice of prior. These results are shown in Figures A.5–A.8. We can see that the joint MLE does indeed maximize the likelihood (at least among our considered parameter space), and that there is no discernible difference in pattern between the log-likelihood and log-likelihood + log-prior as we vary the parameter estimates. Accordingly, this gives us no insight into why the posteriors might be so sensitive to different prior choices.

A.1.4 Discussion

We carried out two different joint approaches to estimating the effect of $PM_{2.5}$ components on CIMT and compared them to the two-stage approach. Performing joint maximum likelihood led to stronger effect estimates and smaller standard errors than the two-stage approach, and were also different from the Bayesian approach. We are concerned about the validity of the Bayesian results and the sensitivity of the health effect estimates to prior specification of θ'_3. These results should not be sensitive to prior mean specification, especially since we specified extremely small prior precisions (wide variances) for θ'_3.

All of these analyses rely on a correctly specified exposure and health model. In reality

one or both are likely to be violated. The extent to which health effect estimates are sensitive to these violations remains to be studied. Szpiro et al. (2011b) investigated behavior of the joint MLE under exposure model misspecification and found it was not very sensitive. However, misspecification of the health model, especially in the presence of large health data sets (as we often have in air pollution epidemiology studies) may lead to harmful feedback when fitting a joint model. Simulation studies need to be done to investigate sampling properties of $\hat{\beta}$ when fitting joint versus two-stage models under different levels of model misspecification and relative sizes of monitoring and epidemiology cohort data. As we have shown in our case study, the different approaches may lead to drastically different qualitative results.

Our results also motivate developing methods to correct for measurement error when the exposure model is not correctly specified. This is especially true when considering multi-pollutant studies, where interest lies in estimating the joint association of multiple pollutants with a single health outcome. In these settings the multiple pollutants must each be predicted, and extending parametric measurement error approaches would involve correctly specifying a multi-pollutant exposure model. As we have shown with our case study, rigorously comparing joint and two-stage approaches for a single pollutant is difficult enough; the problem compounds when considering multiple pollutants. This is a motivating factor in deriving the semi-parametric approaches in Chapters 3 and 4.

	EC		OC	
	$\hat{\beta}$ (SE)	95% CI	$\hat{\beta}$ (SE)	95% CI
Naïve	-0.011 (0.015)	(-0.04, 0.02)	0.025 (0.008)	(0.01, 0.04)
Parametric	-0.011 (0.015)	(-0.04, 0.02)	0.026 (0.009)	(0.01, 0.04)
PB, $\delta = 0$	-0.011 (0.015)	(-0.04, 0.02)	0.025 (0.009)	(0.01, 0.04)
PB, $\delta = 1$	-0.011 (0.015)	(-0.04, 0.02)	0.025 (0.009)	(0.01, 0.04)
Joint MLE	-0.056 (0.013)	(-0.08, -0.03)	0.043 (0.005)	(0.03, 0.05)
Bayes, $c = -2$	0.090 (0.082)	(-0.05, 0.08)	0.054 (0.017)	(0.03, 0.05)
Bayes, $c = -1$	0.225 (0.050)	(0.14, 0.22)	0.076 (0.006)	(0.06, 0.08)
Bayes, $c = 0$	-0.257 (0.054)	(-0.37, -0.25)	0.085 (0.005)	(0.07, 0.08)
Bayes, $c = 1$	-0.258 (0.054)	(-0.37, -0.26)	0.045 (0.018)	(0.02, 0.04)
Bayes, $c = 2$	0.229 (0.050)	(0.14, 0.23)	0.085 (0.012)	(0.06, 0.08)
	Si		S	
	$\hat{\beta}$ (SE)	95% CI	$\hat{\beta}$ (SE)	95% CI
Naïve	0.285 (0.068)	(0.15, 0.42)	0.057 (0.018)	(0.02, 0.09)
Parametric	0.287 (0.103)	(0.09, 0.49)	0.056 (0.025)	(0.01, 0.11)
PB, $\delta = 0$	0.285 (0.092)	(0.11, 0.47)	0.057 (0.025)	(0.01, 0.11)
PB, $\delta = 1$	0.284 (0.093)	(0.10, 0.47)	0.057 (0.025)	(0.01, 0.11)
Joint MLE	0.361 (0.057)	(0.25, 0.47)	0.100 (0.013)	(0.07, 0.13)
Bayes, $c = -2$	0.486 (0.116)	(0.27, 0.48)	0.059 (0.040)	(-0.01, 0.06)
Bayes, $c = -1$	0.578 (0.140)	(0.34, 0.57)	0.067 (0.033)	(0.00, 0.07)
Bayes, $c = 0$	-0.748 (0.184)	(-1.14, -0.74)	0.066 (0.033)	(0,00, 0.07)
Bayes, $c = 1$	0.579 (0.139)	(0.34, 0.57)	0.061 (0.034)	(0.00, 0.06)
Bayes, $c = 2$	0.566 (0.140)	(0.33, 0.55)	0.071 (0.033)	(0.00, 0.07)

Table A.1: Point estimates (standard errors) for the different pollutants using a two-stage approach with no correction or with parametric and parameter bootstrap correction for measurement error, and joint approaches. 95% confidence intervals (credible intervals for the Bayesian analyses) are also shown. "PB" refers to results from parameter bootstrap implemented with given value of δ. "Joint MLE" refer to estimates of β obtained by maximizing the log-likelihood implied by Equation A.1, with standard error estimates obtained from the inverse Hessian evaluated at the estimated parameters. For the Bayesian models the means, standard deviations, and 95% credible intervals of the marginal posteriors of β are reported, for different perturbations c of the prior mean of θ_3'.

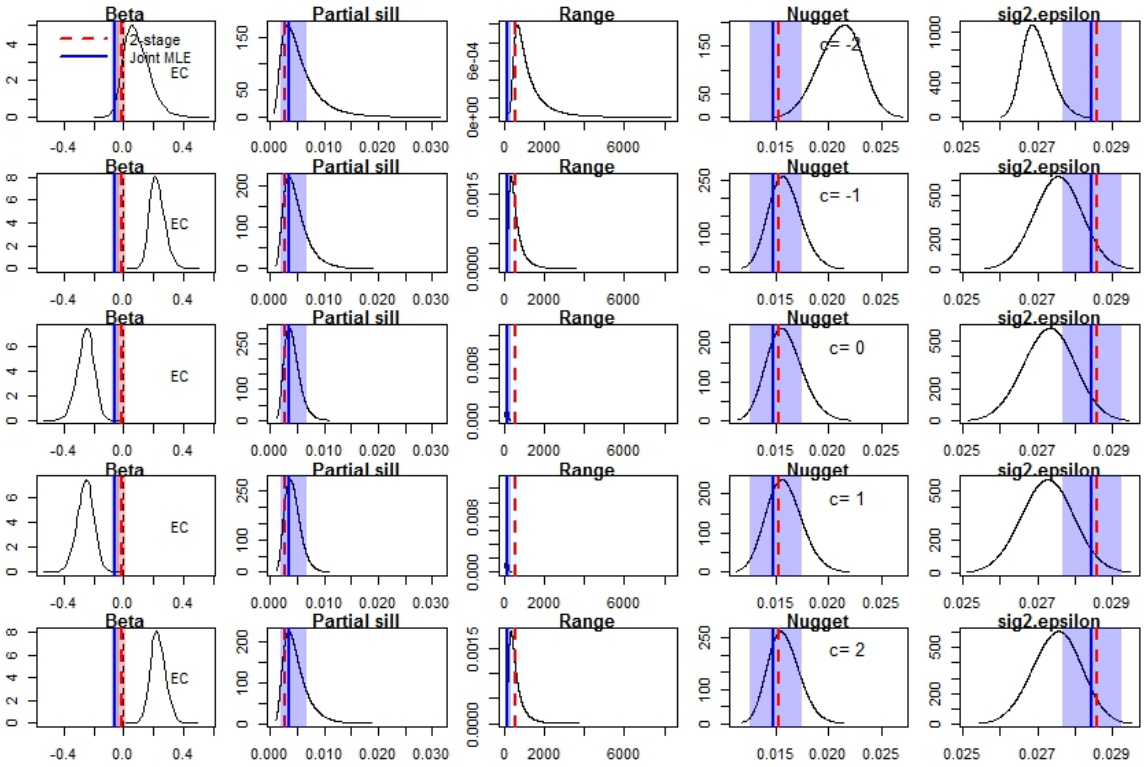


Figure A.1: EC: marginal posteriors of β, σ_ϵ^2 and kriging parameters $\boldsymbol{\theta}$. Analogous two-stage and joint MLE estimates are also shown. For $\hat{\beta}$ the red and blue shading correspond to 95% confidence intervals from the two-stage and joint MLE approaches, respectively. 95% confidence intervals from the joint MLE approach are shown for $\boldsymbol{\theta}$ and σ_ϵ^2. The rows are ordered according to different c in the prior mean $c + \log(10)/2$ of θ_3.

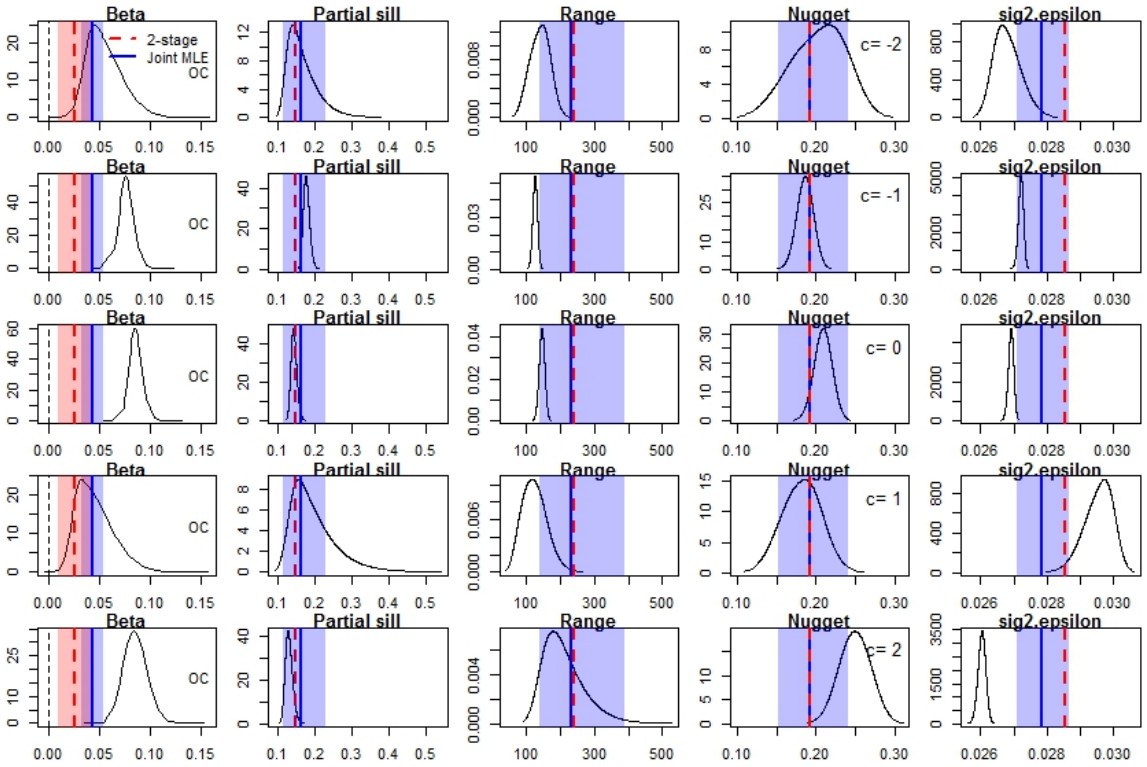


Figure A.2: OC: marginal posteriors of β, σ_ϵ^2 and kriging parameters $\boldsymbol{\theta}$. Analogous two-stage and joint MLE estimates are also shown. For $\hat{\beta}$ the red and blue shading correspond to 95% confidence intervals from the two-stage and joint MLE approaches, respectively. 95% confidence intervals from the joint MLE approach are shown for $\boldsymbol{\theta}$ and σ_ϵ^2. The rows are ordered according to different c in the prior mean $c + \log(10)/2$ of θ_3'.

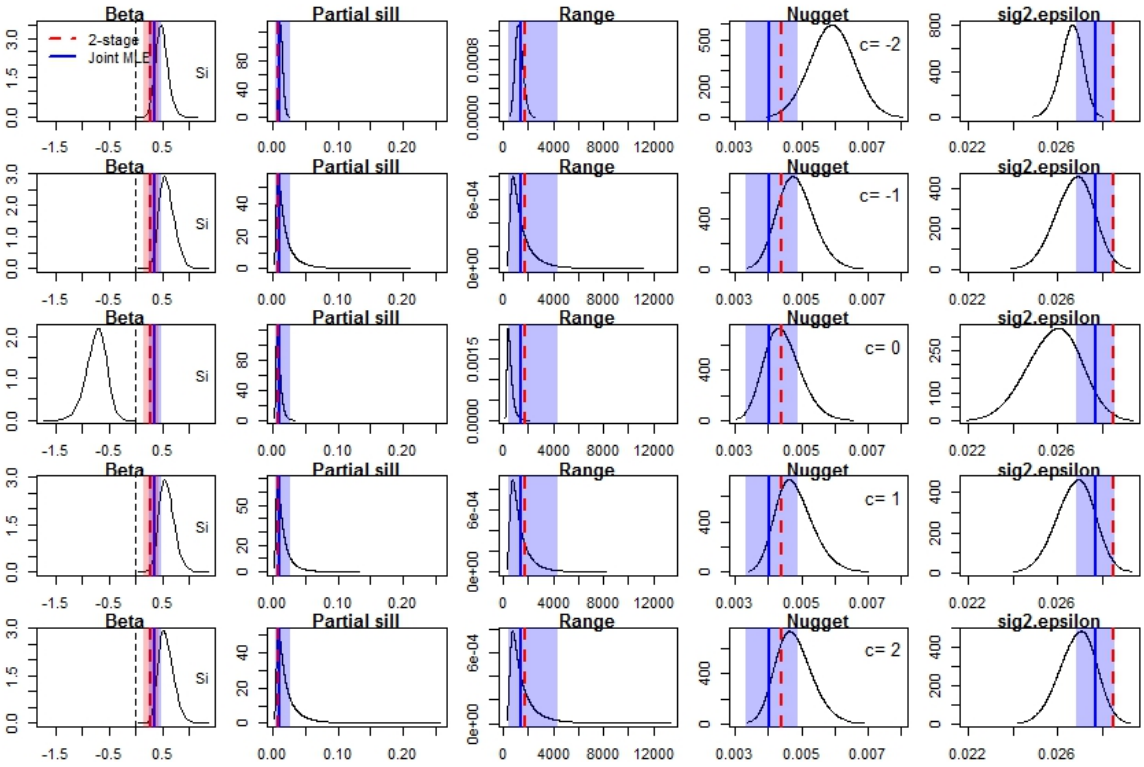


Figure A.3: Si: marginal posteriors of β, σ_ϵ^2 and kriging parameters $\boldsymbol{\theta}$. Analogous two-stage and joint MLE estimates are also shown. For $\hat{\beta}$ the red and blue shading correspond to 95% confidence intervals from the two-stage and joint MLE approaches, respectively. 95% confidence intervals from the joint MLE approach are shown for $\boldsymbol{\theta}$ and σ_ϵ^2. The rows are ordered according to different c in the prior mean $c + \log(10)/2$ of θ_3'.

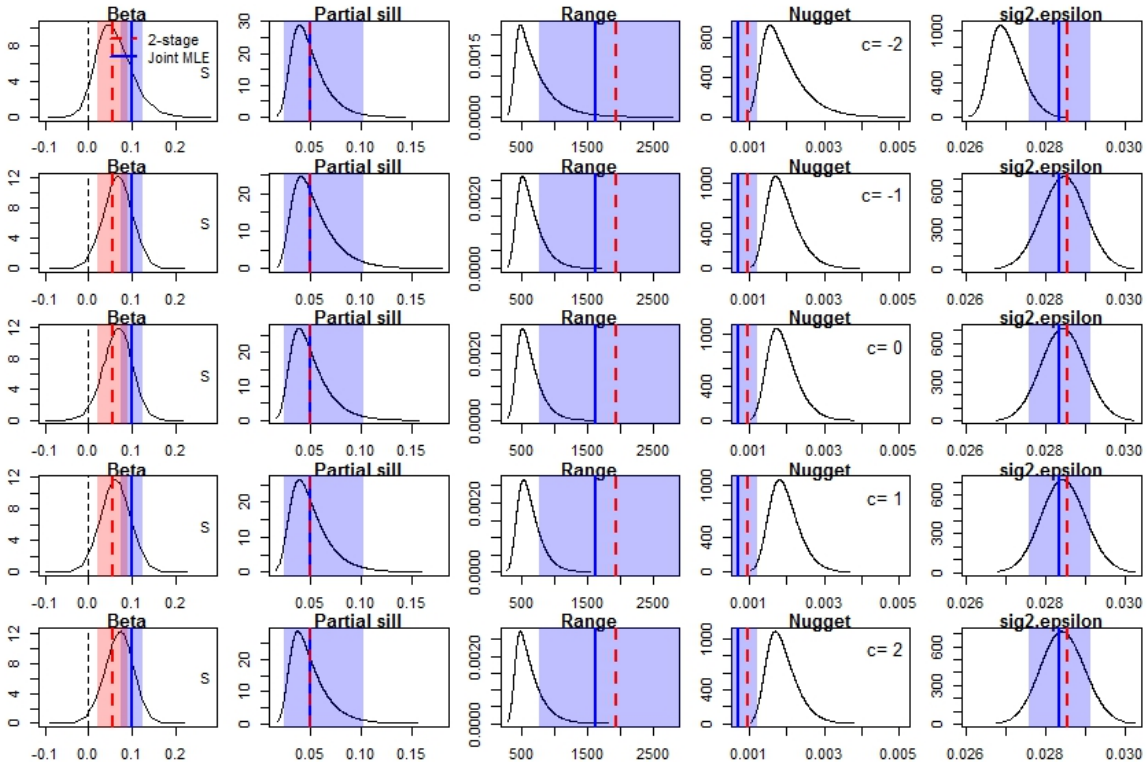


Figure A.4: S: marginal posteriors of β, σ_ϵ^2 and kriging parameters $\boldsymbol{\theta}$. Analogous two-stage and joint MLE estimates are also shown. For $\hat{\beta}$ the red and blue shading correspond to 95% confidence intervals from the two-stage and joint MLE approaches, respectively. 95% confidence intervals from the joint MLE approach are shown for $\boldsymbol{\theta}$ and σ_ϵ^2. The rows are ordered according to different c in the prior mean $c + \log(10)/2$ of θ_3'.

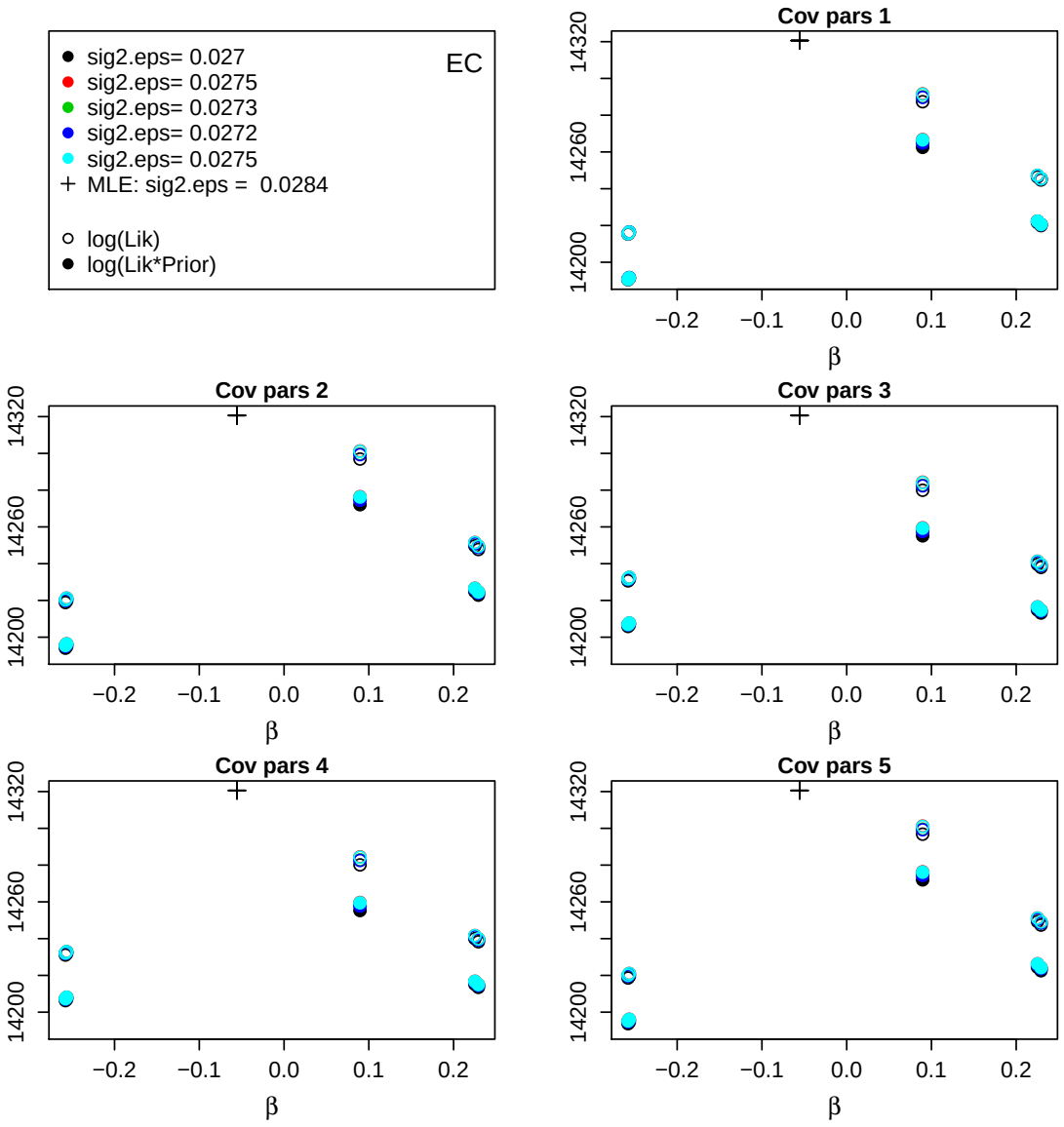


Figure A.5: EC: log joint likelihood (solid circles) and log joint likelihood × prior (empty circles) for different $\hat{\beta}$ (horizontal axes), $\hat{\sigma}^2_\epsilon$ (color), and $\hat{\boldsymbol{\theta}}$ (plot region).

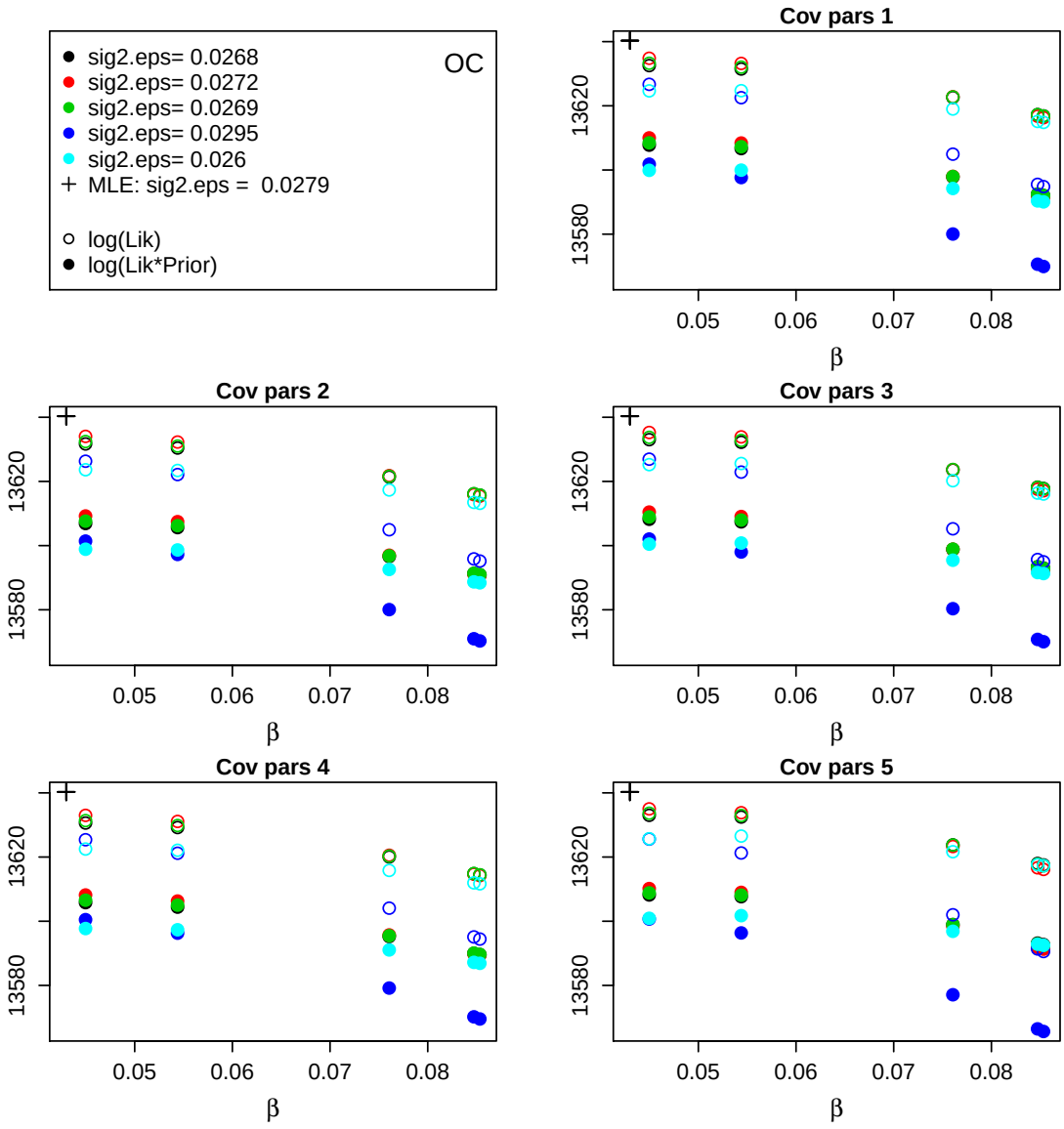


Figure A.6: OC: log joint likelihood (solid circles) and log joint likelihood × prior (empty circles) for different $\hat{\beta}$ (horizontal axes), $\hat{\sigma}^2_\epsilon$ (color), and $\hat{\boldsymbol{\theta}}$ (plot region).

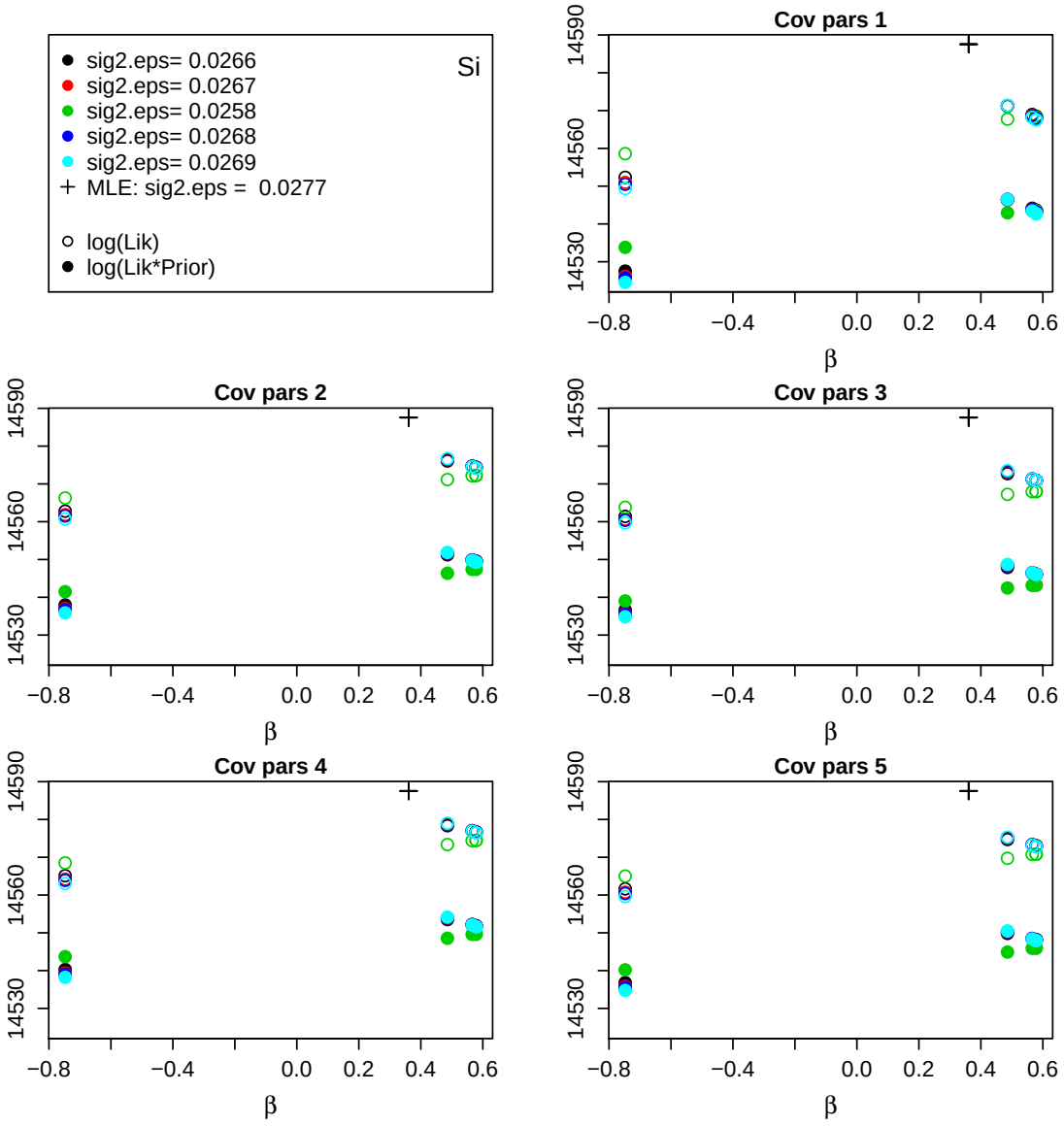


Figure A.7: Si: log joint likelihood (solid circles) and log joint likelihood × prior (empty circles) for different $\hat{\beta}$ (horizontal axes), $\hat{\sigma}^2_\epsilon$ (color), and $\hat{\boldsymbol{\theta}}$ (plot region).

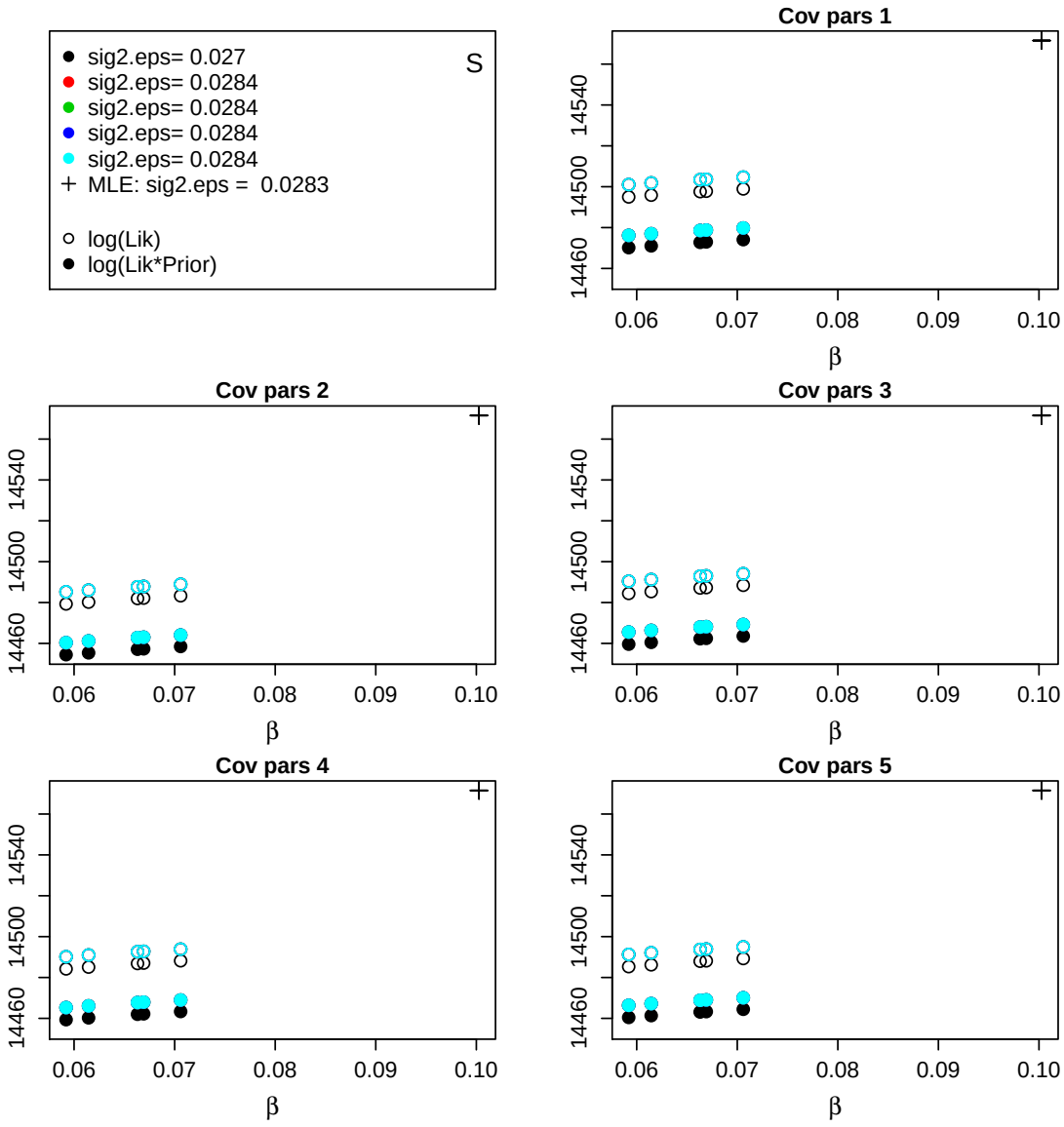


Figure A.8: S: log joint likelihood (solid circles) and log joint likelihood × prior (empty circles) for different $\hat{\beta}$, $\hat{\sigma}^2_\epsilon$, and $\hat{\boldsymbol{\theta}}$.

Appendix B

APPENDIX FOR CHAPTER 3

B.1 Asymptotic definitions of moments

The following definition of asymptotic moments was adapted by Szpiro and Paciorek (2013) from Shao (2003), for use in a setting very similar to ours.

Let v_1, v_2,... be a sequence of random variables and let a_1, a_2,... be a sequence of positive numbers such that $\lim_{n\to\infty} a_n = \infty$. Let ϑ be a real number.

Asymptotic mean. Suppose v is such that $E|v| < \infty$ and we can write $(v_n - \vartheta) = \tilde{v}_n + v'_n$ with $E(\tilde{v}_n) = 0$ and $a_n v'_n \to_d v$. Then we denote $E_{[a_n]}(v_n - \vartheta) = E(v)$ and call $E(v)/a_n$ be an order $1/a_n$ asymptotic mean of $(v_n - \vartheta)$.

Asymptotic variance. Suppose v is such that $Cov(v) < \infty$ and $\sqrt{a_n}(v_n - \vartheta) \to_d v$. Then we denote $Cov_{[a_n]}(v_n) = Cov(v)$ and call $Cov_{[a_n]}(v_n)/a_n$ be an order $1/a_n$ asymptotic covariance of v_n.

B.2 Statement and proof of Lemma 1.

Let $\mathbf{r}^{\perp}(\mathbf{s})$ contain elements $(r_k(\mathbf{s}) - \Theta(\mathbf{s})^T\varphi_k)$, where $\varphi_k = \text{argmin}_\omega \int (r_k(\mathbf{s}) - \Theta(\mathbf{s})^T\omega)^2 dG(\mathbf{s})$ for $k \in \{1, ..., p+q\}$. Let $w^{\perp}_{\lambda}(\mathbf{s}_i) = \mathbf{r}^{\perp}(\mathbf{s}_i)^T\boldsymbol{\gamma}_\lambda$ and $\hat{w}^{\perp}_{\lambda}(\mathbf{s}_i) = \mathbf{r}^{\perp}(\mathbf{s}_i)^T\hat{\boldsymbol{\gamma}}_\lambda$. Then, with:

$$f(\hat{\boldsymbol{\gamma}}_\lambda) = \frac{\int w^{\perp}_{\lambda}(\mathbf{s})\hat{w}^{\perp}_{\lambda}(\mathbf{s})dG(\mathbf{s})}{\int \hat{w}^{\perp}_{\lambda}(\mathbf{s})^2 dG(\mathbf{s})} + \frac{\int u^B_{\lambda}(\mathbf{s})\hat{w}^{\perp}_{\lambda}(\mathbf{s})dG(\mathbf{s})}{\int \hat{w}^{\perp}_{\lambda}(\mathbf{s})^2 dG(\mathbf{s})},$$

$\hat{\beta}_{n^*} = \beta f(\hat{\boldsymbol{\gamma}}_\lambda)$. Let $\mathbf{h}_\lambda$ and $\mathbf{H}_\lambda$ denote the gradient and the Hessian matrix, respectively, of $f(\hat{\boldsymbol{\gamma}}_\lambda)$ evaluated at $\boldsymbol{\gamma}_\lambda$. Let

$$\psi^B_{\lambda} = \frac{\int u^B_{\lambda}(\mathbf{s})w^{\perp}_{\lambda}(\mathbf{s})dG(\mathbf{s})}{\int w^{\perp}_{\lambda}(\mathbf{s})^2 dG(\mathbf{s})},$$

and

$$\psi_\lambda^C = \frac{1}{n^*}\left\{\mathbf{h}_\lambda^T E_{[n^*]}(\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda) + tr\left(\mathbf{H}_\lambda Cov_{[n^*]}(\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda)\right)\right\}.$$

(a)

$$\frac{1}{n^*}E_{[n^*]}\left(\frac{\hat{\beta}_{n^*} - \beta}{\beta} - \psi_\lambda^B\right) = \psi_\lambda^C$$

is an asymptotic expectation of $\left((\hat{\beta}_{n^*} - \beta)/\beta - \psi_\lambda^B\right)$.

(b)

$$\frac{1}{n^*}Var_{[n^*]}\left(\frac{\hat{\beta}_{n^*} - \beta}{\beta}\right) = \frac{1}{n^*}\left(\mathbf{h}_\lambda^T Cov_{[n^*]}(\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda)\mathbf{h}_\lambda\right)$$

is an asymptotic variance of $(\hat{\beta}_{n^*} - \beta)/\beta$.

Proof:

It is easy to see that

$$\hat{\beta}_{n,n^*} = \frac{\sum_{i=1}^n \hat{w}_\lambda^\perp(\mathbf{s}_i)y_i}{\sum_{i=1}^n \hat{w}_\lambda^\perp(\mathbf{s}_i)^2}.$$

Each health outcome measurement can be written

$$y_i = \beta w_\lambda^\perp(\mathbf{s}_i) + \beta(w_\lambda(\mathbf{s}_i) - w_\lambda^\perp(\mathbf{s}_i)) + \beta(x_i - w_\lambda(\mathbf{s}_i)) + \boldsymbol{\beta}_Z^T\mathbf{z}_i + \epsilon_i,$$

which implies

$$\begin{aligned}\hat{\beta}_{n,n^*} &= \beta\left(\frac{\sum_1^n \hat{w}_\lambda^\perp(\mathbf{s}_i)w_\lambda^\perp(\mathbf{s}_i)}{\sum_1^n \hat{w}_\lambda^\perp(\mathbf{s}_i)^2} + \frac{\sum_1^n \hat{w}_\lambda^\perp(\mathbf{s}_i)u_\lambda^B(\mathbf{s})}{\sum_1^n \hat{w}_\lambda^c(\mathbf{s}_i)^2} + \frac{\sum_1^n \hat{w}_\lambda^\perp(\mathbf{s}_i)(w_\lambda(\mathbf{s}_i) - w_\lambda^\perp(\mathbf{s}_i))}{\sum_1^n \hat{w}_\lambda^\perp(\mathbf{s}_i)^2}\right) \\ &\quad + \frac{\sum_1^n \hat{w}_\lambda^\perp(\mathbf{s}_i)\boldsymbol{\beta}_Z^T\mathbf{z}_i}{\sum_1^n \hat{w}_\lambda^\perp(\mathbf{s}_i)^2} + \frac{\sum_1^n \hat{w}_\lambda^\perp(\mathbf{s}_i)\epsilon_i}{\sum_1^n \hat{w}_\lambda^\perp(\mathbf{s}_i)^2}.\end{aligned}$$

The numerator in each of the final three terms on the right-hand side of this expression has mean zero, since every random variable being summed over has mean zero. The

numerator of the third term has mean zero because each element of $\mathbf{r}(\mathbf{s})$ is orthogonal to each element of $\mathbf{r}(\mathbf{s}) - \mathbf{r}^{\perp}(\mathbf{s})$ by construction. The fourth numerator has mean zero because $\mathbf{z}_i = \Theta(\mathbf{s}_i) + \boldsymbol{\zeta}_i$, the elements of $\mathbf{r}^{\perp}(\mathbf{s}_i)$ are orthogonal to $\Theta(\mathbf{s}_i)$ by construction, and each element of $\boldsymbol{\zeta}_i$ has mean zero and is independent of every element of $\mathbf{r}^{\perp}(\mathbf{s}_i)$. Finally, the fifth numerator has mean zero because every ϵ_i has mean zero and is independent of everything else. Letting $n \to \infty$, the weak law of large numbers implies

$$\begin{aligned} \hat{\beta}_{n^*} &= \beta \frac{\int w_\lambda^{\perp}(\mathbf{s}) \hat{w}_\lambda^{\perp}(\mathbf{s}) dG(\mathbf{s})}{\int \hat{w}_\lambda^{\perp}(\mathbf{s})^2 dG(\mathbf{s})} + \beta \frac{\int u_\lambda^{B}(\mathbf{s}) \hat{w}_\lambda^{\perp}(\mathbf{s}) dG(\mathbf{s})}{\int \hat{w}_\lambda^{\perp}(\mathbf{s})^2 dG(\mathbf{s})} \\ &= \beta f(\hat{\boldsymbol{\gamma}}_\lambda). \end{aligned}$$

If we write $\boldsymbol{b}_\lambda = \int u_\lambda^{B}(\mathbf{s}) \mathbf{r}^{\perp}(\mathbf{s}) dG(\mathbf{s})$ and $\mathbf{A} = \int \mathbf{r}^{\perp}(\mathbf{s}) \mathbf{r}^{\perp}(\mathbf{s})^T dG(\mathbf{s})$ we can express this as

$$f(\hat{\boldsymbol{\gamma}}_\lambda) = (\boldsymbol{\gamma}_\lambda^T \mathbf{A} \hat{\boldsymbol{\gamma}}_\lambda)(\hat{\boldsymbol{\gamma}}_\lambda^T \mathbf{A} \hat{\boldsymbol{\gamma}}_\lambda)^{-1} + (\boldsymbol{b}_\lambda^T \hat{\boldsymbol{\gamma}}_\lambda)(\hat{\boldsymbol{\gamma}}_\lambda^T \mathbf{A} \hat{\boldsymbol{\gamma}}_\lambda)^{-1},$$

and we are now prepared to perform a Taylor expansion of $f(\hat{\boldsymbol{\gamma}}_\lambda)$ around $\boldsymbol{\gamma}_\lambda$. The gradient of $f(\hat{\boldsymbol{\gamma}}_\lambda)$ is:

$$\begin{aligned} Df(\hat{\boldsymbol{\gamma}}_\lambda) &= -2(\boldsymbol{\gamma}_\lambda^T \mathbf{A} \hat{\boldsymbol{\gamma}}_\lambda)(\hat{\boldsymbol{\gamma}}_\lambda^T \mathbf{A} \hat{\boldsymbol{\gamma}}_\lambda)^{-2} \mathbf{A} \hat{\boldsymbol{\gamma}}_\lambda + (\hat{\boldsymbol{\gamma}}_\lambda^T \mathbf{A} \hat{\boldsymbol{\gamma}}_\lambda)^{-1} \mathbf{A} \boldsymbol{\gamma}_\lambda \\ &\quad -2(\boldsymbol{b}_\lambda^T \hat{\boldsymbol{\gamma}}_\lambda)(\hat{\boldsymbol{\gamma}}_\lambda^T \mathbf{A} \hat{\boldsymbol{\gamma}}_\lambda)^{-2} \mathbf{A} \hat{\boldsymbol{\gamma}}_\lambda + (\hat{\boldsymbol{\gamma}}_\lambda^T \mathbf{A} \hat{\boldsymbol{\gamma}}_\lambda)^{-1} \boldsymbol{b}_\lambda. \end{aligned}$$

The Hessian is:

$$\begin{aligned} D^2 f(\hat{\boldsymbol{\gamma}}_\lambda) &= -2(\boldsymbol{\gamma}_\lambda^T \mathbf{A} \hat{\boldsymbol{\gamma}}_\lambda)(\hat{\boldsymbol{\gamma}}_\lambda^T \mathbf{A} \hat{\boldsymbol{\gamma}}_\lambda)^{-2} \mathbf{A} + 8(\boldsymbol{\gamma}_\lambda^T \mathbf{A} \hat{\boldsymbol{\gamma}}_\lambda)(\hat{\boldsymbol{\gamma}}_\lambda^T \mathbf{A} \hat{\boldsymbol{\gamma}}_\lambda)^{-3} \mathbf{A} \hat{\boldsymbol{\gamma}}_\lambda \hat{\boldsymbol{\gamma}}_\lambda^T \mathbf{A} \\ &\quad -2(\hat{\boldsymbol{\gamma}}_\lambda^T \mathbf{A} \hat{\boldsymbol{\gamma}}_\lambda)^{-2} \mathbf{A} \hat{\boldsymbol{\gamma}}_\lambda \boldsymbol{\gamma}_\lambda^T \mathbf{A} - 2(\hat{\boldsymbol{\gamma}}_\lambda^T \mathbf{A} \hat{\boldsymbol{\gamma}}_\lambda)^{-2} \mathbf{A} \boldsymbol{\gamma}_\lambda \hat{\boldsymbol{\gamma}}_\lambda^T \mathbf{A} \\ &\quad -2(\boldsymbol{b}_\lambda^T \hat{\boldsymbol{\gamma}}_\lambda)(\hat{\boldsymbol{\gamma}}_\lambda^T \mathbf{A} \hat{\boldsymbol{\gamma}}_\lambda)^{-2} \mathbf{A} + 8(\boldsymbol{b}_\lambda^T \hat{\boldsymbol{\gamma}}_\lambda)(\hat{\boldsymbol{\gamma}}_\lambda^T \mathbf{A} \hat{\boldsymbol{\gamma}}_\lambda)^{-3} \mathbf{A} \hat{\boldsymbol{\gamma}}_\lambda \hat{\boldsymbol{\gamma}}_\lambda^T \mathbf{A} \\ &\quad -2(\hat{\boldsymbol{\gamma}}_\lambda^T \mathbf{A} \hat{\boldsymbol{\gamma}}_\lambda)^{-2} \boldsymbol{b}_\lambda \hat{\boldsymbol{\gamma}}_\lambda^T \mathbf{A} - 2(\hat{\boldsymbol{\gamma}}_\lambda^T \mathbf{A} \hat{\boldsymbol{\gamma}}_\lambda)^{-2} \mathbf{A} \hat{\boldsymbol{\gamma}}_\lambda \boldsymbol{b}_\lambda^T. \end{aligned}$$

Evaluating these quantities at $\hat{\boldsymbol{\gamma}}_\lambda = \boldsymbol{\gamma}_\lambda$ yields:

$$f(\boldsymbol{\gamma}_\lambda) = 1 + (\boldsymbol{b}_\lambda^T \boldsymbol{\gamma}_\lambda)(\boldsymbol{\gamma}_\lambda^T \mathbf{A} \boldsymbol{\gamma}_\lambda)^{-1}$$

$$\begin{aligned} Df(\hat{\boldsymbol{\gamma}}_\lambda)|_{\boldsymbol{\gamma}_\lambda} &= -(\boldsymbol{\gamma}_\lambda^T \mathbf{A} \boldsymbol{\gamma}_\lambda)^{-1} \mathbf{A} \boldsymbol{\gamma}_\lambda \\ &\quad - 2(\boldsymbol{b}_\lambda^T \boldsymbol{\gamma}_\lambda)(\boldsymbol{\gamma}_\lambda^T \mathbf{A} \boldsymbol{\gamma}_\lambda)^{-2} \mathbf{A} \boldsymbol{\gamma}_\lambda + (\boldsymbol{\gamma}_\lambda^T \mathbf{A} \boldsymbol{\gamma}_\lambda)^{-1} \boldsymbol{b}_\lambda \\ \frac{1}{2} D^2 f(\hat{\boldsymbol{\gamma}}_\lambda)|_{\boldsymbol{\gamma}_\lambda} &= -(\boldsymbol{\gamma}_\lambda^T \mathbf{A} \boldsymbol{\gamma}_\lambda)^{-1} \mathbf{A} + 2(\boldsymbol{\gamma}_\lambda^T \mathbf{A} \boldsymbol{\gamma}_\lambda)^{-2} \mathbf{A} \boldsymbol{\gamma}_\lambda \boldsymbol{\gamma}_\lambda^T \mathbf{A} \\ &\quad - (\boldsymbol{b}_\lambda^T \boldsymbol{\gamma}_\lambda)(\boldsymbol{\gamma}_\lambda^T \mathbf{A} \boldsymbol{\gamma}_\lambda)^{-2} \mathbf{A} + 4(\boldsymbol{b}_\lambda^T \boldsymbol{\gamma}_\lambda)(\boldsymbol{\gamma}_\lambda^T \mathbf{A} \boldsymbol{\gamma}_\lambda)^{-3} \mathbf{A} \boldsymbol{\gamma}_\lambda \boldsymbol{\gamma}_\lambda^T \mathbf{A} \\ &\quad - (\boldsymbol{\gamma}_\lambda^T \mathbf{A} \boldsymbol{\gamma}_\lambda)^{-2} \boldsymbol{b}_\lambda \boldsymbol{\gamma}_\lambda^T \mathbf{A} - (\boldsymbol{\gamma}_\lambda^T \mathbf{A} \boldsymbol{\gamma}_\lambda)^{-2} \mathbf{A} \boldsymbol{\gamma}_\lambda \boldsymbol{b}_\lambda^T. \end{aligned}$$

We let $\mathbf{h}_\lambda = Df(\hat{\boldsymbol{\gamma}}_\lambda)|_{\boldsymbol{\gamma}_\lambda}$ and $\mathbf{H}_\lambda = \frac{1}{2} D^2 f(\hat{\boldsymbol{\gamma}}_\lambda)|_{\boldsymbol{\gamma}_\lambda}$, and then Taylor expansion of $\hat{\beta}_{n^*}$ about $\boldsymbol{\gamma}_\lambda$ gives

$$\hat{\beta}_{n^*} = \beta f(\hat{\boldsymbol{\gamma}}_\lambda) = \beta + \beta(\boldsymbol{b}_\lambda^T \boldsymbol{\gamma}_\lambda)(\boldsymbol{\gamma}_\lambda^T \mathbf{A} \boldsymbol{\gamma}_\lambda)^{-1} + \beta \left\{ \mathbf{h}_\lambda^T (\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda) + (\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda)^T \mathbf{H}_\lambda (\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda) \right\} + o_p(1/n^*).$$

Noting that $(\boldsymbol{b}_\lambda^T \boldsymbol{\gamma}_\lambda)(\boldsymbol{\gamma}_\lambda^T \mathbf{A} \boldsymbol{\gamma}_\lambda)^{-1} = \psi_\lambda^B$, we have:

$$\frac{(\hat{\beta}_{n^*} - \beta)}{\beta} = \psi_\lambda^B + \mathbf{h}_\lambda^T (\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda) + (\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda)^T \mathbf{H}_\lambda (\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda) + o_p(1/n^*).$$

Thus,

$$\begin{aligned} E_{[n^*]} \left(\frac{(\hat{\beta}_{n^*} - \beta)}{\beta} - \psi_\lambda^B \right) &= \mathbf{h}_\lambda^T E_{[n^*]} (\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda) + tr(\mathbf{H}_\lambda Cov_{[n^*]} (\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda)); \\ Var_{[n^*]} \left(\frac{(\hat{\beta}_{n^*} - \beta)}{\beta} \right) &= \mathbf{h}_\lambda^T Cov_{[n^*]} (\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda) \mathbf{h}_\lambda; \end{aligned}$$

as desired.

B.3 Estimation of Lemma 1 quantities

In order to use Lemma 1 to estimate the bias and variance of $\hat{\beta}_{n^*}$, we need to estimate ψ_λ^B, $\mathbf{h}_\lambda$, $\mathbf{H}_\lambda$, and the moments of $(\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda)$ from available data.

The numerator of ψ_λ^B, $\int u_\lambda^B(\mathbf{s}) w_\lambda^\perp(\mathbf{s}) dG(\mathbf{s})$, can only be estimated using data from monitoring locations since that is where we can observe approximate values of u_λ^B and $w_\lambda^\perp$. Since u_λ^B is the difference between the observed and predicted exposure values, and the prediction model was also fit at monitoring locations, we were concerned that simply plugging in residuals from the penalized regression would lead to an understimate of the numerator of ψ_λ^B. Accordingly we also considered 10-fold and leave-one-out cross-validation to estimate u_λ^B, but found that this overestimated ψ_λ^B, especially for small values of λ. For small λ, the numerator of ψ_λ^B is close to zero since u_λ^B is nearly orthogonal to $w_\lambda(\mathbf{s})$, and using cross-validation broke this near-orthogonality. Therefore, we opted for the more straightforward approach of directly plugging in the residuals. The denominator of ψ_λ^B can be estimated using subject or monitoring locations. Because there are often many more subjects than monitors, it makes sense to use the subject locations to approximate this integral; this is what we did in our simulations and data analysis.

Following Yu and Ruppert (2002), we estimate $Cov(\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda)$ using a sandwich form that accounts for having $\lambda > 0$. In some applications, it is reasonable to simply ignore λ when constructing asymptotic standard errors because it converges to zero as the sample size increases. However, we have found that we obtain better finite-sample variance estimates of $Cov(\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda)$ if we explicitly account for λ. If we set

$$\mathbf{B} = \int \left\{ \left(\mathbf{r}(\mathbf{s}) \left(\Phi(\mathbf{s}) - \mathbf{r}(\mathbf{s})^T \boldsymbol{\gamma}_\lambda \right) - \lambda \mathbf{D} \boldsymbol{\gamma}_\lambda \right) \left(\mathbf{r}(\mathbf{s}) \left(\Phi(\mathbf{s}) - \mathbf{r}(\mathbf{s})^T \boldsymbol{\gamma}_\lambda \right) - \lambda \mathbf{D} \boldsymbol{\gamma}_\lambda \right)^T \right\} dG(\mathbf{s})$$

and $\mathbf{A} = \lambda \mathbf{D} + \int \mathbf{r}(\mathbf{s}) \mathbf{r}(\mathbf{s})^T dG(\mathbf{s})$, then $\sqrt{n^*}(\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda) \rightarrow_d N(0, \mathbf{V})$ where $\mathbf{V} = \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}$ where $\mathbf{V}$ is a model-robust estimate of the covariance of $\hat{\boldsymbol{\gamma}}_\lambda$ that incorporates reduced variance from $\lambda > 0$. We can estimate $\mathbf{V}$ by plugging in consistent estimates of all the

quantities: $\hat{\boldsymbol{\gamma}}_\lambda$ for $\boldsymbol{\gamma}_\lambda$ and $\mathbf{R}(\mathbf{s}^*)^T\mathbf{R}(\mathbf{s}^*)$ whenever an estimate of $\int \mathbf{r}(\mathbf{s})\mathbf{r}(\mathbf{s})^T dG(\mathbf{s})$ is needed. Note however that this covariance estimate, though effectively accounting for the reduction in variance induced by λ, maintains all the characteristics of a sandwich covariance estimate for unpenalized regression, and in particular it may be biased for small samples. Mancl and DeRouen (2001) showed that in the context of generalized estimating equations, assuming a correctly specified mean model and no penalty, inversely weighting each residual in the $\hat{\mathbf{B}}$ portion of the estimated covariance matrix by one minus the influence of each data point reduces the bias of the covariance estimate. Although we make fewer assumptions, we similarly found that weighting each residual term in $\hat{\mathbf{B}}$ by $1 - h_{ii}$ reduced the bias of our sandwich covariance estimates for smaller sample sizes, where $h_{ii} = \mathbf{r}(\mathbf{s}_i^*)^T(n^*\lambda\mathbf{D} + \mathbf{R}(\mathbf{s}^*)^T\mathbf{R}(\mathbf{s}^*))^{-1}\mathbf{r}(\mathbf{s}_i^*)$.

It remains to describe estimation of $\mathbf{h}_\lambda$, $\mathbf{H}_\lambda$, and $E_{[n^*]}(\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda)$. Estimation of $\mathbf{h}_\lambda$ and $\mathbf{H}_\lambda$ only requires estimation of $\boldsymbol{b}_\lambda$, $\mathbf{A}$ and $\boldsymbol{\gamma}_\lambda$. $\boldsymbol{b}_\lambda$ was estimated as described in the text, using residuals from the penalized regression fits to the monitoring data to estimate u_λ^B. $\mathbf{A}$ was estimated by summing over plug-ins of $\mathbf{r}^\perp(\mathbf{s}_i)\mathbf{r}^\perp(\mathbf{s}_i)^T$ at subject locations, and $\boldsymbol{\gamma}_\lambda$ was estimated with $\hat{\boldsymbol{\gamma}}_\lambda$.

To estimate $E_{[n^*]}(\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda)$ we used an approach very similar to the one outlined in Szpiro and Paciorek (2013). For arbitrary m_j let $\mathbf{M}$ denote the $n^* \times n^*$ diagonal matrix where the j^{th} diagonal element is m_j. If we let $m_j = 1/n^*$ for $j = 1, ..., n^*$ and Λ the $(p+q) \times (p+q)$ matrix $\lambda\mathbf{D}$, then we note that

$$\hat{\boldsymbol{\gamma}}_\lambda = \kappa(m_1, ..., m_{n^*}) = (\mathbf{R}(\mathbf{s}^*)^T\mathbf{M}\mathbf{R}(\mathbf{s}^*) + \Lambda)^{-1}\mathbf{R}(\mathbf{s}^*)^T\mathbf{M}\mathbf{x}^*.$$

From here we want to take the expectation of $\hat{\boldsymbol{\gamma}}_\lambda$ over repeated realizations m_j^* of the m_j, where with each realization the rows of $\mathbf{R}(\mathbf{s}^*)$ and elements of $\mathbf{x}^*$ are re-weighted. Heuristically, we are using $\{\mathbf{s}_1, ..., \mathbf{s}_{n^*}\}$ to approximate the support of $G(\cdot)$; the weights $1/n^*$ to approximate $G(\cdot)$, the probability distribution over this support; and considering repeated sampling with replacement of monitoring locations from $\{\mathbf{s}_1, ..., \mathbf{s}_{n^*}\}$ to approxim-

ate repeated sampling from $G(\cdot)$. This is equivalent to sampling new m_j^* from a multinomial distribution with probabilities $p_j = 1/n^*$. With each realization of m_j^* we obtain new coefficients $\hat{\boldsymbol{\gamma}}_\lambda^*$, and treat the asymptotic mean of $(\hat{\boldsymbol{\gamma}}_\lambda^* - \hat{\boldsymbol{\gamma}}_\lambda) \equiv (\kappa(m_1^*, ..., m_{n^*}^*) - \kappa(m_1, ..., m_{n^*}))$ as an approximation to what we are truly interested in, which is the asymptotic mean of $(\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda)$. This is justified more formally in van der Vaart 1998, Section 20.1. We find the asymptotic distribution (and hence mean) of $(\boldsymbol{\kappa}(m_1^*, ..., m_{n^*}^*) - \boldsymbol{\kappa}(1/n^*, ..., 1/n^*))$ using a multinomial Taylor expansion, perturbing $\{m_1^*, ..., m_{n^*}^*\}$ about their means $\{1/n^*, ..., 1/n^*\}$.

Following Szpiro and Paciorek (2013), we see that this implies

$$E_{[n^*]}(\hat{\boldsymbol{\gamma}}_\lambda - \boldsymbol{\gamma}_\lambda) \approx \frac{1}{2}\left(\frac{1}{n^*} - \frac{1}{(n^*)^2}\right)\sum_{j=1}^{n^*}\frac{\partial^2\kappa}{\partial m_j^2} - \frac{1}{2}\frac{1}{(n^*)^2}\sum_{j,k=1;j\neq k}^{n^*}\frac{\partial^2\kappa}{\partial m_j \partial m_k}$$

where, with $\mathbf{R}^* = \mathbf{R}(\mathbf{s}^*)$ for simplicity of notation;

$$
\begin{aligned}
\frac{\partial^2 \kappa}{\partial m_j \partial m_k} &= (\mathbf{R}^{*T}\mathbf{M}\mathbf{R}^* + \Lambda)^{-1}\mathbf{R}^{*T}\frac{\partial \mathbf{M}}{\partial m_k}\mathbf{R}^*(\mathbf{R}^{*T}\mathbf{M}\mathbf{R}^* + \Lambda)^{-1}\mathbf{R}^{*T}\frac{\partial \mathbf{M}}{\partial m_j}\mathbf{R}^*(\mathbf{R}^{*T}\mathbf{M}\mathbf{R}^* + \Lambda)^{-1}\mathbf{R}^{*T}\mathbf{M}\mathbf{x}^* \\
&- (\mathbf{R}^{*T}\mathbf{M}\mathbf{R}^* + \Lambda)^{-1}\mathbf{R}^{*T}\frac{\partial^2 \mathbf{M}}{\partial m_j \partial m_k}\mathbf{R}^*(\mathbf{R}^{*T}\mathbf{M}\mathbf{R}^* + \Lambda)^{-1}\mathbf{R}^{*T}\mathbf{M}\mathbf{x}^* \\
&+ (\mathbf{R}^{*T}\mathbf{M}\mathbf{R}^* + \Lambda)^{-1}\mathbf{R}^{*T}\frac{\partial \mathbf{M}}{\partial m_j}\mathbf{R}^*(\mathbf{R}^{*T}\mathbf{M}\mathbf{R}^* + \Lambda)^{-1}\mathbf{R}^{*T}\frac{\partial \mathbf{M}}{\partial m_k}\mathbf{R}^*(\mathbf{R}^{*T}\mathbf{M}\mathbf{R}^* + \Lambda)^{-1}\mathbf{R}^{*T}\mathbf{M}\mathbf{x}^* \\
&- (\mathbf{R}^{*T}\mathbf{M}\mathbf{R}^* + \Lambda)^{-1}\mathbf{R}^{*T}\frac{\partial \mathbf{M}}{\partial m_j}\mathbf{R}^*(\mathbf{R}^{*T}\mathbf{M}\mathbf{R}^* + \Lambda)^{-1}\mathbf{R}^{*T}\frac{\partial \mathbf{M}}{\partial m_k}\mathbf{x}^* \\
&- (\mathbf{R}^{*T}\mathbf{M}\mathbf{R}^* + \Lambda)^{-1}\mathbf{R}^{*T}\frac{\partial \mathbf{M}}{\partial m_k}\mathbf{R}^*(\mathbf{R}^{*T}\mathbf{M}\mathbf{R}^* + \Lambda)^{-1}\mathbf{R}^{*T}\frac{\partial \mathbf{M}}{\partial m_j}\mathbf{x}^* \\
&+ (\mathbf{R}^{*T}\mathbf{M}\mathbf{R}^* + \Lambda)^{-1}\mathbf{R}^{*T}\frac{\partial^2 \mathbf{M}}{\partial m_j \partial m_k}\mathbf{x}^*
\end{aligned}
$$

$$
\begin{aligned}
&= (\mathbf{R}^{*T}\mathbf{M}\mathbf{R}^* + \Lambda)^{-1}\mathbf{r}_k^*\mathbf{r}_k^{*T}(\mathbf{R}^{*T}\mathbf{M}\mathbf{R}^* + \Lambda)^{-1}\mathbf{r}_j^*\mathbf{r}_j^{T}(\mathbf{R}^{*T}\mathbf{M}\mathbf{R}^* + \Lambda)^{-1}\mathbf{R}^{*T}\mathbf{M}\mathbf{x}^* \\
&- 0 \\
&+ (\mathbf{R}^{*T}\mathbf{M}\mathbf{R}^* + \Lambda)^{-1}\mathbf{r}_j^*\mathbf{r}_j^{*T}(\mathbf{R}^{*T}\mathbf{M}\mathbf{R}^* + \Lambda)^{-1}\mathbf{r}_k^*\mathbf{r}_k^{T}(\mathbf{R}^{*T}\mathbf{M}\mathbf{R}^* + \Lambda)^{-1}\mathbf{R}^{*T}\mathbf{M}\mathbf{x}^* \\
&- (\mathbf{R}^{*T}\mathbf{M}\mathbf{R}^* + \Lambda)^{-1}\mathbf{r}_j^*\mathbf{r}_j^{*T}(\mathbf{R}^{*T}\mathbf{M}\mathbf{R}^* + \Lambda)^{-1}\mathbf{r}_k^* x_k^* \\
&- (\mathbf{R}^{*T}\mathbf{M}\mathbf{R}^* + \Lambda)^{-1}\mathbf{r}_k^*\mathbf{r}_k^{*T}(\mathbf{R}^{*T}\mathbf{M}\mathbf{R}^* + \Lambda)^{-1}\mathbf{r}_j^* x_j^* \\
&+ 0.
\end{aligned}
$$

Appendix C

APPENDIX FOR CHAPTER 4

C.1 Derivation of low-rank common component model (LRCCM)

Fitting separate penalized regression spline models such as described in Section 4.3.2 does nothing to exploit correlation between pollutants. The common component model of Fanshawe and Diggle (2012) exploits this correlation and is motivated using a spatial random effect. We describe it briefly here and develop its low-rank analogue as an example of a low-rank penalized regression model that accounts for between-pollutant correlation. For simplicity assume $n_0^* = n^*$ so we observe both pollutants at each location, and let $\mathbf{X}_j^*$ denote the $n^* \times 1$ vector of observations of the j^{th} pollutant. Let $\mathbf{P}_j^*$ denote the corresponding $n^* \times p_j$ matrix of geographic covariates for modeling the j^{th} pollutant. Then the common component model of Fanshawe and Diggle (2012) is:

$$\mathbf{X}_j^* = \mathbf{P}_j^* \boldsymbol{\alpha}_j + s_j \tilde{\boldsymbol{\delta}}_0 + \tilde{\boldsymbol{\delta}}_j + \boldsymbol{\nu}_j \tag{C.1}$$

where $\tilde{\boldsymbol{\delta}}_k \sim N(0, \sigma_k^2 \Sigma_k)$ for $k \in \{0, 1, 2\}$ and $\boldsymbol{\nu}_j \sim N(0, \tau_j^2 \mathbf{I}_{n^*})$ for $j \in \{1, 2\}$, with $Cor(\nu_{i1}, \nu_{j2}) = 1(\mathbf{s}_i^* = \mathbf{s}_j^*)\rho_\nu$. The random coefficients $\tilde{\boldsymbol{\delta}}_0$ denote a spatially correlated Gaussian process that is held in common by the two surfaces and is scaled by s_j; for identifiability we set $s_1 = 1$. The $\tilde{\boldsymbol{\delta}}_j$ denote two pollutant-specific Gaussian processes independent of each other and of $\tilde{\boldsymbol{\delta}}_0$. The correlation matrices Σ_k are parameterized by correlation functions $\rho_k(r)$, such that the $\{i, j\}^{th}$ element of Σ_k is $\rho_k(||\mathbf{s}_i - \mathbf{s}_j||)$. For example, $\rho_k(r)$ could be an exponential correlation function with $\rho_k(r) = \exp(-|r|/\phi_k)$ given range parameter ϕ_k. In practice, ϕ_k may be fixed (e.g., at the maximum distance between monitor locations, see Kammann and Wand (2003)). If we let $\mathbf{Q}_k^* = \Sigma_k^{1/2}$ we can reparameterize Equation C.1

as

$$\mathbf{X}_j^* = \mathbf{P}_j^* \boldsymbol{\alpha}_j + s_j \mathbf{Q}_0^* \boldsymbol{\delta}_0 + \mathbf{Q}_j^* \boldsymbol{\delta}_j + \boldsymbol{\nu}_j, \tag{C.2}$$

where $\boldsymbol{\delta}_k \sim N(0, \sigma_k^2 \mathbf{I}_{n^*})$. Let $\mathbf{R}_j^* = \{(1_{j=1})\mathbf{P}_1^*, (1_{j=2})\mathbf{P}_2^*, s_j\mathbf{Q}_0^*, (1_{j=1})\mathbf{Q}_1^*, (1_{j=2})\mathbf{Q}_2^*\}$ and $\mathbf{R}(\mathbf{s}_i^*)$ the $r \times 2$ matrix created by combining the i^{th} rows of $\mathbf{R}_1^*$ and $\mathbf{R}_2^*$. For fixed σ_k, τ_j and ρ_ν, it follows that the corresponding full data negative log-likelihood is:

$$\ell(\boldsymbol{\gamma}|\mathbf{X}_1^*, \mathbf{X}_2^*; s_1, s_2, \sigma_0, \sigma_1, \sigma_2, \tau_1, \tau_2, \rho_\nu) = \frac{1}{n^*} \sum_{i=1}^{n^*} (\mathbf{x}_i^* - \mathbf{R}(\mathbf{s}_i^*)^T \boldsymbol{\gamma})^T \boldsymbol{\Pi} (\mathbf{x}_i^* - \mathbf{R}(\mathbf{s}_i^*)^T \boldsymbol{\gamma}) + \boldsymbol{\gamma}^T \boldsymbol{\Lambda} \boldsymbol{\gamma}, \tag{C.3}$$

where $\boldsymbol{\gamma} = \left\{\boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_2^T, \boldsymbol{\delta}_0^T, \boldsymbol{\delta}_1^T, \boldsymbol{\delta}_2^T\right\}$, $\boldsymbol{\Pi} = \begin{pmatrix} \tau_1^2 & \rho_\nu \tau_1 \tau_2 \\ \rho_\nu \tau_1 \tau_2 & \tau_2^2 \end{pmatrix}^{-1}$ and

$$\boldsymbol{\Lambda} = \frac{1}{n^*} \begin{pmatrix} 0_{(p_1+p_2)\times(p_1+p_2)} & 0 & 0 & 0 \\ 0 & \frac{1}{\sigma_0^2}\mathbf{I}_{n^*} & 0 & 0 \\ 0 & 0 & \frac{1}{\sigma_1^2}\mathbf{I}_{n^*} & 0 \\ 0 & 0 & 0 & \frac{1}{\sigma_2^2}\mathbf{I}_{n^*} \end{pmatrix}.$$

It is apparent that minimizing (C.3) with respect to $\boldsymbol{\gamma}$ is equivalent to minimizing (4.1). We note that (C.3) is more like what Hodges (2013) terms "not-quite-a-likelihood" than a true log-likelihood, as we minimize (C.3) with respect to fixed parameters (the $\boldsymbol{\alpha}_j$) and random effects (the $\boldsymbol{\delta}_j$) to yield the penalized regression coefficients. The mixed-effects formulation motivates estimating the σ_k, τ_j, and ρ_ν with REML to obtain the matrices $\boldsymbol{\Pi}$ and $\boldsymbol{\Lambda}$. One could also use REML to estimate s_2, though as it defines the basis functions used for modeling the common component it may be preferable to fix it along with the ϕ_k. The model described by (C.3) is full-rank since $r = p_1 + p_2 + n^* + n^* + n^*$, which poses problems for defining the asymptotic quantity $\boldsymbol{\gamma}$. An alternative is to fix the dimension of this model by approximating (C.1) with $\tilde{\boldsymbol{\delta}}_k' \sim N(0, \sigma_k^2 \boldsymbol{\Omega}_k)$ where $\boldsymbol{\Omega}_k$ is a $c_k \times c_k$ matrix with the $\{i,j\}^{th}$ element equal to $\rho_k(||\boldsymbol{\kappa}_i - \boldsymbol{\kappa}_j||)$. Here the $\boldsymbol{\kappa}_i$ denote c_k fixed spatial knots chosen via a space filling algorithm (Kammann and Wand, 2003). Following the same line

of thought outlined above with these new random effects defines the low-rank version of the common component model, with r now fixed at $p_1 + p_2 + c_0 + c_1 + c_2$. An interesting result follows if we set $\rho_\nu = 0$. In this case we can show that

$$\hat{\boldsymbol{\delta}}_0 = \left(s_1^2 \mathbf{Q}_0^{*T}\mathbf{Q}_0^* + s_2^2 \mathbf{Q}_0^{*T}\mathbf{Q}_0^* + \tilde{\lambda}_0 \mathbf{I}_{c_0}\right)^{-1} \mathbf{Q}_0^{*T} \sum_{j=1}^{2} \frac{\lambda_{0j}^{-1}}{\lambda_{01}^{-1} + \lambda_{02}^{-1}} s_j \left(\mathbf{X}_j^* - \mathbf{P}_j^* \hat{\boldsymbol{\alpha}}_j - \mathbf{Q}_j^* \hat{\boldsymbol{\delta}}_j\right); \tag{C.4}$$

$$\hat{\boldsymbol{\delta}}_j = \left(\mathbf{Q}_j^{*T}\mathbf{Q}_j^* + \lambda_j \mathbf{I}_{c_j}\right)^{-1} \mathbf{Q}_j^{*T} \left(\mathbf{X}_j^* - \mathbf{P}_j^* \hat{\boldsymbol{\alpha}}_j - s_j \mathbf{Q}_0^* \hat{\boldsymbol{\delta}}_0\right); \tag{C.5}$$

$$\hat{\boldsymbol{\alpha}}_j = \left(\mathbf{P}_j^{*T}\mathbf{P}_j^*\right)^{-1} \left(\mathbf{X}_j^* - s_j \mathbf{Q}_0^* \hat{\boldsymbol{\delta}}_0 - \mathbf{Q}_j^* \hat{\boldsymbol{\delta}}_j\right); \tag{C.6}$$

where $\lambda_j = \tau_j^2/\sigma_j^2$, $\lambda_{0j} = \tau_j^2/\sigma_0^2$ and $\tilde{\lambda}_0 = \left(\lambda_{01}^{-1} + \lambda_{02}^{-1}\right)^{-1}$. The λ_j can be thought of as penalizing the roughness of the pollutant-specific surfaces, while λ_{0j} can be viewed as penalizing the contribution of the appropriately scaled $\mathbf{X}_j^*$ to the common surface. This formulation further demonstrates the translation between variance components of a spatial mixed effects model and penalties in a penalized regression model, as well as how a penalized regression model can be built to model correlation between pollutants. One characteristic of this model is the inability to consider "unpenalized" coefficients as one can when fitting the separate models of Section 4.3.2. This can be readily seen by noting that the only way to let $\tilde{\lambda}_0 \to 0$ in (C.4) is by letting at least one of the $\lambda_{0j} \to 0$, which in turn leads to undefined $\lambda_{0j}^{-1}/(\lambda_{01}^{-1} + \lambda_{02}^{-1})$. An alternative way of viewing this is assuming the same basis functions are used for modeling the common component and the two pollutant-specific components. In this setting $\mathbf{Q}_0^* = \mathbf{Q}_1^* = \mathbf{Q}_2^* \equiv \mathbf{Q}^*$ and (ignoring the geographic covariates) $\mathbf{r}_1(\mathbf{s}_i^*)^T = \{\mathbf{q}(\mathbf{s}_i^*)^T, 0_q^T, \mathbf{q}(\mathbf{s}_i^*)^T\}$ and $\mathbf{r}_2(\mathbf{s}_i^*)^T = \{s_2\mathbf{q}(\mathbf{s}_i^*)^T, \mathbf{q}(\mathbf{s}_i^*)^T, 0_q^T\}$. With $\boldsymbol{\Lambda}$ zero everywhere, this model is rank-deficient since $\mathbf{q}(\cdot)$ (or a scaled version thereof) appears twice in any row of the pollutant-specific model matrices $\mathbf{R}_j^*$. This is a setting such as we alluded to in Section 4.3.2 where we cannot let $\boldsymbol{\Lambda}$ be zero everywhere and still define each element of $\hat{\boldsymbol{\gamma}}$. This implies we cannot guarantee unbiased estimation of $\boldsymbol{\beta}$ even if we had infinite monitoring data available, as we described in Section 4.4.2.

C.2 Proof of Lemma 2

In what follows we rigorously state and prove Lemma 2 from Section 4.4.3. We isolate the impact of measurement error by letting $n \to \infty$ and consider estimating $\boldsymbol{\beta}$ with $n^* < \infty$. We let $\hat{\boldsymbol{\beta}}_{n^*}$ denote the resulting estimate, while $\hat{\boldsymbol{\beta}}_{n,n^*}$ denotes a pragmatic estimate of $\boldsymbol{\beta}$ using finite n^* and n. A rigorous statement of Lemma 2 requires definition of asymptotic moments following Shao (2003).

Let v_1, v_2,... be a sequence of vector-valued random variables and let a_1, a_2,... be a sequence of positive numbers such that $\lim_{n\to\infty} a_n = \infty$. Let ϑ be a vector of real numbers.

Asymptotic mean. Suppose v is such that $E|v| < \infty$ and we can write $(v_n - \vartheta) = \tilde{v}_n + v'_n$ with $E(\tilde{v}_n) = 0$ and $a_n v'_n \to_d v$. Then we denote $E_{[a_n]}(v_n - \vartheta) = E(v)$ and call $E(v)/a_n$ an order $1/a_n$ asymptotic mean of $(v_n - \vartheta)$.

Asymptotic variance. Suppose v is such that $E(vv^T)$ is positive definite and $\sqrt{a_n}(v_n - \vartheta) \to_d v$. Then we denote $Cov_{[a_n]}(v_n) = Cov(v)$ and call $Cov_{[a_n]}(v_n)/a_n$ an order $1/a_n$ asymptotic covariance of v_n.

Lemma 2 (rigorous statement): Let $\mathbf{r}_j^{\perp}(\mathbf{s})$ contain elements $(r_{jk}(\mathbf{s}) - \Theta(\mathbf{s})^T\varphi_k)$, where $\varphi_k = \text{argmin}_\omega \int (r_{jk}(\mathbf{s}) - \Theta(\mathbf{s})^T\omega)^2 dG(\mathbf{s})$ for $k \in \{1, ..., p+q\}$. Let $\mathbf{R}^{\perp}(\mathbf{s})$ denote the corresponding $r \times 2$ matrix created by binding the $r_j^{\perp}(\mathbf{s})$, let $\mathbf{w}^{\perp}(\mathbf{s}) = \mathbf{R}^{\perp}(\mathbf{s})^T\boldsymbol{\gamma}$ and $\hat{\mathbf{w}}^{\perp}(\mathbf{s}) = \mathbf{R}^{\perp}(\mathbf{s})^T\hat{\boldsymbol{\gamma}}$. Let $\mathbf{M}(\hat{\boldsymbol{\gamma}}) = \int \hat{\mathbf{w}}^{\perp}(\mathbf{s})\hat{\mathbf{w}}^{\perp}(\mathbf{s})^T dG(\mathbf{s})$ and $\mathbf{U}(\hat{\boldsymbol{\gamma}}) = \hat{\mathbf{w}}^{\perp}(\mathbf{s})\mathbf{u}^B(\mathbf{s})^T dG(\mathbf{s})$. Then with

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

we can show that $\hat{\boldsymbol{\beta}} = f(\hat{\boldsymbol{\gamma}})^T\boldsymbol{\beta}$. Let $\mathbf{G}_i(\hat{\boldsymbol{\gamma}})$ denote the 2×2 matrix created by taking the partial derivative of $f(\hat{\boldsymbol{\gamma}})$ with respect to $\hat{\gamma}_i$ and $\mathbf{H}_{ij}$ the 2×2 matrix created by taking the partial derivative of $\mathbf{G}_i(\hat{\boldsymbol{\gamma}})$ with respect to $\hat{\gamma}_j$. Let $\mathbf{g}_{kl}(\hat{\boldsymbol{\gamma}})$ denote the $r \times 1$ vector where $g_{ikl}(\hat{\boldsymbol{\gamma}})$ equals the $\{k,l\}^{th}$ element of $\mathbf{G}_i(\boldsymbol{\gamma})$, and $\mathbf{h}_{kl}(\hat{\boldsymbol{\gamma}})$ the $r \times r$ matrix where $h_{ijkl}(\hat{\boldsymbol{\gamma}})$ equals the $\{k,l\}^{th}$ element of $\mathbf{H}_{ij}(\hat{\boldsymbol{\gamma}})$.

Additionally let $\Psi^B = \mathbf{M}(\boldsymbol{\gamma})^{-1}\mathbf{U}(\boldsymbol{\gamma})$, and

$$\Psi^C = \begin{pmatrix} \mathbf{g}_{11}(\boldsymbol{\gamma})^T E_{[n^*]}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) & \mathbf{g}_{12}(\boldsymbol{\gamma})^T E_{[n^*]}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \\ \mathbf{g}_{21}(\boldsymbol{\gamma})^T E_{[n^*]}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) & \mathbf{g}_{22}(\boldsymbol{\gamma})^T E_{[n^*]}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \end{pmatrix} + \begin{pmatrix} tr\left(\mathbf{h}_{11}(\boldsymbol{\gamma}) Cov_{[n^*]}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})\right) & tr\left(\mathbf{h}_{12}(\boldsymbol{\gamma}) Cov_{[n^*]}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})\right) \\ tr\left(\mathbf{h}_{21}(\boldsymbol{\gamma}) Cov_{[n^*]}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})\right) & tr\left(\mathbf{h}_{22}(\boldsymbol{\gamma}) Cov_{[n^*]}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})\right) \end{pmatrix}.$$

Then the asymptotic mean of $\left(\hat{\boldsymbol{\beta}}_{n^*} - \boldsymbol{\beta} - \Psi^B\boldsymbol{\beta}\right)$ is:

$$\frac{1}{n^*} E_{[n^*]}\left(\hat{\boldsymbol{\beta}}_{n^*} - \boldsymbol{\beta} - \Psi^B\boldsymbol{\beta}\right) = \Psi^C\boldsymbol{\beta},$$

where Ψ^B is bias from Berkson-like error and $\frac{1}{n^*}\Psi^C$ is bias from classical-like error.

Proof.

After regressing out the subject-specific covariates it follows that

$$\hat{\boldsymbol{\beta}}_{n,n^*} = \left(\sum_{i=1}^{n} \mathbf{R}^{\perp}(\mathbf{s}_i)^T\hat{\boldsymbol{\gamma}}\hat{\boldsymbol{\gamma}}^T\mathbf{R}^{\perp}(\mathbf{s}_i)\right)^{-1}\left(\sum_{i=1}^{n}(\mathbf{R}^{\perp}(\mathbf{s}_i)^T\hat{\boldsymbol{\gamma}})y_i\right).$$

We can write each health outcome as

$$y_i = \boldsymbol{\gamma}^T\mathbf{R}^{\perp}(\mathbf{s}_i)\boldsymbol{\beta} + \left(\boldsymbol{\Phi}(\mathbf{s}_i) - \mathbf{R}(\mathbf{s}_i)^T\boldsymbol{\gamma}\right)^T\boldsymbol{\beta} + \boldsymbol{\gamma}^T(\mathbf{R}(\mathbf{s}_i) - \mathbf{R}^{\perp}(\mathbf{s}_i))^T\boldsymbol{\beta} + \boldsymbol{\eta}_i^T\boldsymbol{\beta} + \mathbf{z}_i^T\boldsymbol{\beta}_z + \epsilon_i,$$

from which it follows that

$$\left(\sum_{i=1}^{n}(\mathbf{R}^{\perp}(\mathbf{s}_i)^T\hat{\boldsymbol{\gamma}})y_i\right) = \left(\sum_{i=1}^{n}\mathbf{R}^{\perp}(\mathbf{s}_i)^T\hat{\boldsymbol{\gamma}}\boldsymbol{\gamma}^T\mathbf{R}^{\perp}(\mathbf{s}_i)\right)\boldsymbol{\beta} + \left(\sum_{i=1}^{n}\mathbf{R}^{\perp}(\mathbf{s}_i)^T\hat{\boldsymbol{\gamma}}(\boldsymbol{\Phi}(\mathbf{s}_i) - \mathbf{R}(\mathbf{s}_i)^T\boldsymbol{\gamma})^T\right)\boldsymbol{\beta}$$

$$+ \left(\sum_{i=1}^{n}\mathbf{R}^{\perp}(\mathbf{s}_i)^T\hat{\boldsymbol{\gamma}}\boldsymbol{\gamma}^T(\mathbf{R}(\mathbf{s}_i) - \mathbf{R}^{\perp}(\mathbf{s}_i))^T\right)\boldsymbol{\beta} + \left(\sum_{i=1}^{n}\mathbf{R}^{\perp}(\mathbf{s}_i)^T\hat{\boldsymbol{\gamma}}\boldsymbol{\eta}_i^T\right) + \left(\sum_{i=1}^{n}\mathbf{R}^{\perp}(\mathbf{s}_i)^T\hat{\boldsymbol{\gamma}}\mathbf{z}_i^T\right) + \left(\sum_{i=1}^{n}\mathbf{R}^{c}(\mathbf{s}_i)^T\hat{\boldsymbol{\gamma}}\epsilon_i\right).$$

By the weak law of large numbers, as $n \to \infty$ each summation converges in probability to an integral with respect to $G(\mathbf{s})$. The four summations in the bottom row converge to 0. This

follows for the first summation since each element of $\mathbf{R}^{\perp}(\mathbf{s})$ is orthogonal to $\left(\mathbf{R}(\mathbf{s}) - \mathbf{R}^{\perp}(\mathbf{s})\right)$ by construction; for the second summation since $\boldsymbol{\eta}_i$ has mean zero and is independent of $\mathbf{s}$; for the third summation since $\mathbf{z}_i = \Theta(\mathbf{s}_i) + \boldsymbol{\zeta}_i$, each element of $\mathbf{R}^{\perp}(\mathbf{s})$ is orthogonal to linear combinations of $\Theta(\mathbf{s})$ by construction, and $\boldsymbol{\zeta}_i$ has mean zero and is independent of $\mathbf{s}_i$; and for the fourth summation since ϵ_i has mean zero and is independent of $\mathbf{s}_i$. Thus as $n \to \infty$,

$$\hat{\boldsymbol{\beta}}_{n,n^*} \longrightarrow_p \left(\int \mathbf{R}^{\perp}(\mathbf{s})^T \hat{\boldsymbol{\gamma}}\hat{\boldsymbol{\gamma}}^T \mathbf{R}^{\perp}(\mathbf{s}) dG(\mathbf{s})\right)^{-1} \left(\int \mathbf{R}^{\perp}(\mathbf{s})^T \hat{\boldsymbol{\gamma}}\boldsymbol{\gamma}^T \mathbf{R}^{\perp}(\mathbf{s}) dG(\mathbf{s})\right) \boldsymbol{\beta}$$

$$+ \left(\int \mathbf{R}^{\perp}(\mathbf{s})^T \hat{\boldsymbol{\gamma}}\hat{\boldsymbol{\gamma}}^T \mathbf{R}^{\perp}(\mathbf{s}) dG(\mathbf{s})\right)^{-1} \left(\int \mathbf{R}^{\perp}(\mathbf{s})^T \hat{\boldsymbol{\gamma}} \left(\boldsymbol{\Phi}(\mathbf{s}) - \mathbf{R}(\mathbf{s})^T \boldsymbol{\gamma}\right)^T dG(\mathbf{s})\right) \boldsymbol{\beta}$$

$$\equiv f_1(\hat{\boldsymbol{\gamma}})\boldsymbol{\beta} + f_2(\hat{\boldsymbol{\gamma}})\boldsymbol{\beta} \equiv f(\hat{\boldsymbol{\gamma}}) = \hat{\boldsymbol{\beta}}_{n^*}.$$

From here let $\mathbf{M}(\hat{\boldsymbol{\gamma}}) = \int \mathbf{R}^{\perp}(\mathbf{s})^T \hat{\boldsymbol{\gamma}}\hat{\boldsymbol{\gamma}}^T \mathbf{R}^{\perp}(\mathbf{s}) dG(\mathbf{s})$. Let $\mathbf{u}^B(\mathbf{s}) \equiv \boldsymbol{\Phi}(\mathbf{s}) - \mathbf{R}(\mathbf{s})^T\boldsymbol{\gamma}$, and $\mathbf{U}(\hat{\boldsymbol{\gamma}}) = \int \mathbf{R}^{\perp}(\mathbf{s})^T \hat{\boldsymbol{\gamma}}\mathbf{u}(\mathbf{s})^T dG(\mathbf{s})$. It follows that we can re-write f_1 and f_2 as

$$f_1(\hat{\boldsymbol{\gamma}}) = \mathbf{M}(\hat{\boldsymbol{\gamma}})^{-1} \left(\int \mathbf{R}^{\perp}(\mathbf{s})^T \hat{\boldsymbol{\gamma}}\boldsymbol{\gamma}^T \mathbf{R}^{\perp}(\mathbf{s}) dG(\mathbf{s})\right)$$

$$f_2(\hat{\boldsymbol{\gamma}}) = \mathbf{M}(\hat{\boldsymbol{\gamma}})^{-1}\mathbf{U}(\hat{\boldsymbol{\gamma}})$$

Let $\mathbf{R}_j^{\perp}(\mathbf{s})$ denote the j^{th} row of $\mathbf{R}^{\perp}(\mathbf{s})$. Let $\mathbf{A}_j(\hat{\boldsymbol{\gamma}}) \equiv \int \mathbf{R}_j^{\perp}(\mathbf{s})\hat{\boldsymbol{\gamma}}^T \mathbf{R}^{\perp}(\mathbf{s}) dG(\mathbf{s})$, $\mathbf{V}_i = \int \mathbf{R}_i^{\perp}(\mathbf{s})\mathbf{u}(\mathbf{s})^T dG(\mathbf{s})$, and $\mathbf{B}_{ij} = \int \mathbf{R}_i^{\perp}(\mathbf{s})\mathbf{R}_j^{\perp}(\mathbf{s})^T dG(\mathbf{s})$. Then it follows that

$$\frac{\partial}{\partial \hat{\gamma}_i}\mathbf{M}(\hat{\boldsymbol{\gamma}}) = \mathbf{A}_i(\hat{\boldsymbol{\gamma}}) + \mathbf{A}_i(\hat{\boldsymbol{\gamma}})^T$$

$$\frac{\partial}{\partial \hat{\gamma}_i}\mathbf{U}(\hat{\boldsymbol{\gamma}}) = \mathbf{V}_i$$

$$\frac{\partial}{\partial \hat{\gamma}_j}\mathbf{A}_i(\hat{\boldsymbol{\gamma}}) = \mathbf{B}_{ij}$$

$$\frac{\partial^2}{\partial \hat{\gamma}_i \partial \hat{\gamma}_j}\mathbf{M}(\hat{\boldsymbol{\gamma}}) = \mathbf{B}_{ij} + \mathbf{B}_{ij}^T$$

Thus we have

$$\frac{\partial f_1}{\partial \hat{\gamma}_i} = \mathbf{M}(\hat{\gamma})^{-1}\mathbf{A}_i(\gamma) - \mathbf{M}(\hat{\gamma})^{-1}\left(\mathbf{A}_i(\hat{\gamma}) + \mathbf{A}_i(\hat{\gamma})^T\right)\mathbf{M}(\hat{\gamma})^{-1}\int \mathbf{R}^{\perp}(\mathbf{s})^T\hat{\gamma}\gamma^T\mathbf{R}^{\perp}(\mathbf{s})dG(\mathbf{s}),$$
$$\frac{\partial f_2}{\partial \hat{\gamma}_i} = \mathbf{M}(\hat{\gamma})^{-1}\mathbf{V}_i - \mathbf{M}(\hat{\gamma})^{-1}\left(\mathbf{A}_i(\hat{\gamma}) + \mathbf{A}_i(\hat{\gamma})^T\right)\mathbf{M}(\hat{\gamma})^{-1}\mathbf{U}(\hat{\gamma}),$$

$$\frac{\partial^2 f_1}{\partial \hat{\gamma}_i \gamma_j} = -\mathbf{M}(\hat{\gamma})^{-1}(\mathbf{A}_j(\hat{\gamma}) + \mathbf{A}_j(\hat{\gamma})^T)\mathbf{M}(\hat{\gamma})^{-1}\mathbf{A}_i(\gamma)$$

$$+\mathbf{M}(\hat{\gamma})^{-1}(\mathbf{A}_j(\hat{\gamma}) + \mathbf{A}_j(\hat{\gamma})^T)\mathbf{M}(\hat{\gamma})^{-1}(\mathbf{A}_i(\hat{\gamma}) + \mathbf{A}_i(\hat{\gamma})^T)\mathbf{M}(\hat{\gamma})^{-1}\int \mathbf{R}(\mathbf{s})^T\hat{\gamma}\gamma^T\mathbf{R}(\mathbf{s})dG(\mathbf{s})$$

$$-\mathbf{M}(\hat{\gamma})^{-1}\left(\mathbf{B}_{ij} + \mathbf{B}_{ij}^T\right)\mathbf{M}(\hat{\gamma})^{-1}\int \mathbf{R}(\mathbf{s})^T\hat{\gamma}\gamma^T\mathbf{R}(\mathbf{s})dG(\mathbf{s})$$

$$+\mathbf{M}(\hat{\gamma})^{-1}(\mathbf{A}_i(\hat{\gamma}) + \mathbf{A}_i(\hat{\gamma})^T)\mathbf{M}(\hat{\gamma})^{-1}(\mathbf{A}_j(\hat{\gamma}) + \mathbf{A}_j(\hat{\gamma})^T)\mathbf{M}(\hat{\gamma})^{-1}\int \mathbf{R}(\mathbf{s})^T\hat{\gamma}\gamma^T\mathbf{R}(\mathbf{s})dG(\mathbf{s})$$

$$-\mathbf{M}(\hat{\gamma})^{-1}(\mathbf{A}_i(\hat{\gamma}) + \hat{\mathbf{A}}_i(\hat{\gamma})^T)\mathbf{M}(\hat{\gamma})^{-1}\mathbf{A}_j(\gamma),$$

$$\frac{\partial^2 f_2}{\partial \hat{\gamma}_i \gamma_j} = -\mathbf{M}(\hat{\gamma})^{-1}(\mathbf{A}_j(\hat{\gamma}) + \mathbf{A}_j(\hat{\gamma})^T)\mathbf{M}(\hat{\gamma})^{-1}\mathbf{V}_i$$

$$+\mathbf{M}(\hat{\gamma})^{-1}(\mathbf{A}_j(\hat{\gamma}) + \mathbf{A}_j(\hat{\gamma})^T)\mathbf{M}(\hat{\gamma})^{-1}(\mathbf{A}_i(\hat{\gamma}) + \mathbf{A}_i(\hat{\gamma})^T)\mathbf{M}(\hat{\gamma})^{-1}\mathbf{U}(\hat{\gamma})$$

$$-\mathbf{M}(\hat{\gamma})^{-1}(\mathbf{B}_{ij}+\mathbf{B}_{ij}^T)\mathbf{M}(\hat{\gamma})^{-1}\mathbf{U}(\hat{\gamma})+\mathbf{M}(\hat{\gamma})^{-1}(\mathbf{A}_i(\hat{\gamma})+\mathbf{A}_i(\hat{\gamma})^T)\mathbf{M}(\hat{\gamma})^{-1}(\mathbf{A}_j(\hat{\gamma})+\mathbf{A}_j(\hat{\gamma})^T)\mathbf{M}(\hat{\gamma})^{-1}\mathbf{U}(\hat{\gamma})$$

$$-\mathbf{M}(\hat{\gamma})^{-1}(\mathbf{A}_i(\hat{\gamma}) + \mathbf{A}_i(\hat{\gamma})^T)\mathbf{M}(\hat{\gamma})^{-1}\mathbf{V}_j.$$

Evaluating these quantities at $\boldsymbol{\gamma}$ and combining yields

$$\frac{\partial f}{\partial \hat{\boldsymbol{\gamma}}}|\boldsymbol{\gamma} = \left(\frac{\partial f_1}{\partial \hat{\gamma}_i} + \frac{\partial f_2}{\partial \hat{\gamma}_i}\right)|\boldsymbol{\gamma} = \mathbf{M}(\boldsymbol{\gamma})^{-1}\mathbf{A}_i(\boldsymbol{\gamma}) - \mathbf{M}(\boldsymbol{\gamma})^{-1}(\mathbf{A}_i(\boldsymbol{\gamma}) + \mathbf{A}_i(\boldsymbol{\gamma})^T)$$

$$+\mathbf{M}(\boldsymbol{\gamma})^{-1}\mathbf{V}_i - \mathbf{M}(\boldsymbol{\gamma})^{-1}(\mathbf{A}_i(\boldsymbol{\gamma}) + \mathbf{A}_i(\boldsymbol{\gamma})^T)\mathbf{M}(\boldsymbol{\gamma})^{-1}\mathbf{U}(\boldsymbol{\gamma}) \equiv \mathbf{G}_i(\boldsymbol{\gamma})$$

$$\left(\frac{\partial^2 f}{\partial \hat{\gamma}_i \hat{\gamma}_j}\right)|\boldsymbol{\gamma} = \left(\frac{\partial^2 f_1}{\partial \hat{\gamma}_i \hat{\gamma}_j} + \frac{\partial^2 f_2}{\partial \hat{\gamma}_i \hat{\gamma}_j}\right)|\boldsymbol{\gamma} = -\mathbf{M}(\boldsymbol{\gamma})^{-1}(\mathbf{A}_j(\boldsymbol{\gamma}) + \mathbf{A}_j(\boldsymbol{\gamma})^T)\mathbf{M}(\boldsymbol{\gamma})^{-1}\mathbf{A}_i(\boldsymbol{\gamma})$$

$$+\mathbf{M}^{-1}(\boldsymbol{\gamma})(\mathbf{A}_j(\boldsymbol{\gamma}) + \mathbf{A}_j(\boldsymbol{\gamma})^T)\mathbf{M}(\boldsymbol{\gamma})^{-1}(\mathbf{A}_i(\boldsymbol{\gamma}) + \mathbf{A}_i(\boldsymbol{\gamma})^T) - \mathbf{M}(\boldsymbol{\gamma})^{-1}(\mathbf{B}_{ij} + \mathbf{B}_{ij}^T)$$

$$+\mathbf{M}(\boldsymbol{\gamma})^{-1}(\mathbf{A}_i(\boldsymbol{\gamma})+\mathbf{A}_i(\boldsymbol{\gamma})^T)\mathbf{M}(\boldsymbol{\gamma})^{-1}(\mathbf{A}_j(\boldsymbol{\gamma})+\mathbf{A}_j(\boldsymbol{\gamma})^T)-\mathbf{M}(\boldsymbol{\gamma})^{-1}(\mathbf{A}_i(\boldsymbol{\gamma})+\mathbf{A}_i(\boldsymbol{\gamma})^T)\mathbf{M}(\boldsymbol{\gamma})^{-1}\mathbf{A}_j(\boldsymbol{\gamma})$$

$$-\mathbf{M}(\boldsymbol{\gamma})^{-1}(\mathbf{A}_j(\boldsymbol{\gamma})+\mathbf{A}_j(\boldsymbol{\gamma})^T)\mathbf{M}(\boldsymbol{\gamma})^{-1}\mathbf{V}_i+\mathbf{M}(\boldsymbol{\gamma})^{-1}(\mathbf{A}_j(\boldsymbol{\gamma})+\mathbf{A}_j(\boldsymbol{\gamma})^T)\mathbf{M}(\boldsymbol{\gamma})^{-1}(\mathbf{A}_i(\boldsymbol{\gamma})+\mathbf{A}_i(\boldsymbol{\gamma})^T)\mathbf{M}(\boldsymbol{\gamma})^{-1}\mathbf{U}(\boldsymbol{\gamma})$$

$$-\mathbf{M}(\boldsymbol{\gamma})^{-1}(\mathbf{B}_{ij}+\mathbf{B}_{ij}^T)\mathbf{M}(\boldsymbol{\gamma})^{-1}\mathbf{U}(\boldsymbol{\gamma})+\mathbf{M}(\boldsymbol{\gamma})^{-1}(\mathbf{A}_i(\boldsymbol{\gamma})+\mathbf{A}_i(\boldsymbol{\gamma})^T)\mathbf{M}(\boldsymbol{\gamma})^{-1}(\mathbf{A}_j(\boldsymbol{\gamma})+\mathbf{A}_j(\boldsymbol{\gamma})^T)\mathbf{M}(\boldsymbol{\gamma})^{-1}\mathbf{U}(\boldsymbol{\gamma})$$

$$-\mathbf{M}(\boldsymbol{\gamma})^{-1}(\mathbf{A}_i(\boldsymbol{\gamma})+\mathbf{A}_i(\boldsymbol{\gamma})^T)\mathbf{M}(\boldsymbol{\gamma})^{-1}\mathbf{V}_j \equiv \mathbf{H}_{ij}(\boldsymbol{\gamma})$$

Let $\mathbf{g}_{kl}(\boldsymbol{\gamma})$ denote the $r \times 1$ vector where $g_{ikl}(\hat{\boldsymbol{\gamma}})$ equals the $\{k,l\}^{th}$ element of $\mathbf{G}_i(\boldsymbol{\gamma})$, and $\mathbf{h}_{kl}(\boldsymbol{\gamma})$ the $r \times r$ matrix where $h_{ijkl}(\boldsymbol{\gamma})$ equals the $\{k,l\}^{th}$ element of $\mathbf{H}_{ij}(\hat{\boldsymbol{\gamma}})$. Then by Taylor's Theorem:

$$\hat{\boldsymbol{\beta}}_{n^*} = \boldsymbol{\beta} + \mathbf{M}(\boldsymbol{\gamma})^{-1}\mathbf{U}(\boldsymbol{\gamma})\boldsymbol{\beta} + \begin{pmatrix} \mathbf{g}_{11}(\boldsymbol{\gamma})^T(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}) & \mathbf{g}_{12}(\boldsymbol{\gamma})^T(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}) \\ \mathbf{g}_{21}(\boldsymbol{\gamma})^T(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}) & \mathbf{g}_{22}(\boldsymbol{\gamma})(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}) \end{pmatrix} \boldsymbol{\beta}$$

$$+ \begin{pmatrix} (\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})^T\mathbf{h}_{11}(\boldsymbol{\gamma})(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})^T & (\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})^T\mathbf{h}_{12}(\boldsymbol{\gamma})(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})^T \\ (\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})^T\mathbf{h}_{21}(\boldsymbol{\gamma})(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})^T & (\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})^T\mathbf{h}_{22(\boldsymbol{\gamma})}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})^T \end{pmatrix} \boldsymbol{\beta} + o_p(1/n^*).$$

Accordingly from the definitions of asymptotic moments defined above,

$$\begin{aligned}
&\frac{1}{n^*}E_{[n^*]}\left(\hat{\boldsymbol{\beta}}_{n^*} - \boldsymbol{\beta} - \mathbf{M}(\boldsymbol{\gamma})^{-1}\mathbf{U}(\boldsymbol{\gamma})\boldsymbol{\beta}\right) = \\
&\frac{1}{n^*}\begin{pmatrix} \mathbf{g}_{11}(\boldsymbol{\gamma})^T E_{[n^*]}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}) & \mathbf{g}_{12}(\boldsymbol{\gamma})^T E_{[n^*]}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}) \\ \mathbf{g}_{21}(\boldsymbol{\gamma})^T E_{[n^*]}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}) & \mathbf{g}_{22}(\boldsymbol{\gamma})^T E_{[n^*]}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}) \end{pmatrix} \boldsymbol{\beta} \\
&+ \frac{1}{n^*}\begin{pmatrix} tr\left(\mathbf{h}_{11}(\boldsymbol{\gamma})Cov_{[n^*]}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})\right) & tr\left(\mathbf{h}_{12}(\boldsymbol{\gamma})Cov_{[n^*]}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})\right) \\ tr\left(\mathbf{h}_{21}(\boldsymbol{\gamma})Cov_{[n^*]}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})\right) & tr\left(\mathbf{h}_{22}(\boldsymbol{\gamma})Cov_{[n^*]}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})\right) \end{pmatrix} \boldsymbol{\beta} \\
&= \Psi^C\boldsymbol{\beta}.
\end{aligned}$$

Noting that $\mathbf{M}(\boldsymbol{\gamma})^{-1}\mathbf{U}(\boldsymbol{\gamma}) = \Psi^B$, Lemma 2 follows.

BIBLIOGRAPHY

Abdi H. 2003. Partial least square regression (PLS regression). *Encyclopedia for research methods for the social sciences*, pages 792–795.

Bergen S and Szpiro AA. sub. Minimizing the impact of measurement error when using penalized regression to model exposure in two-stage air pollution epidemiology studies.

Bergen S, Sheppard L, Sampson PD, Kim SY, Richards M, Vedal S, et al. 2012. A national model built with partial least squares and universal kriging and bootstrap-based measurement error correction techniques: An application to the Multi-Ethnic Study of Atherosclerosis. *UW Biostatistics Working Paper Series*, Working Paper 386.

Bergen S, Sheppard L, Sampson PD, Kim SY, Richards M, Vedal S, Kaufman JD, and Szpiro AA. 2013. A national prediction model for $PM_{2.5}$ component exposures and measurement error-corrected health effect inference. *Environmental Health Perspectives*, 121(9):1017–1025.

Billionnet C, Sherrill D, and Annesi-Maesano I. 2012. Estimating the health effects of exposure to multi-pollutant mixture. *Annals of Epidemiology*, 22(2):126–141.

Brook RD, Rajagopalan S, Pope CA, Brook JR, Bhatnagar A, Diez-Roux AV, et al. 2010. Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the american heart association. *Circulation*, 121(6):2331–2378.

Carroll RJ. *Measurement error in nonlinear models: a modern perspective*. CRC Press, 2006. ISBN 1584886331.

Chan SH, Van Hee VC, Bergen S, Szpiro AA, Oron AP, DeRoo LA, London SJ, Marshall

JD, Kaufman JD, and Sandler DP. Subm. Long term air pollution exposure and blood pressure in the Sister Study.

Cressie N. *Statistics for spatial data*. John Wiley and Sons, Inc, 1993.

Cressie N and Johannesson G. 2008. Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):209–226.

Diez Roux AV, Merkin SS, Arnett D, Chambless L, Massing M, Nieto FJ, Sorlie P, Szklo M, Tyroler HA, and Watson RL. 2001. Neighborhood of residence and incidence of coronary heart disease. *New England Journal of Medicine*, 345(2):99–106.

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*, 21(2):187–194.

Efron B and Tibshirani R. *An introduction to the bootstrap*, volume 57. Chapman & Hall/CRC, 1993.

EPA US. Integrated Science Assessment for Particulate Matter EPA/600/R-08/139F. Available at: `http://www.epa.gov/ncea/pdfs/partmatt/Dec2009/PM_ISA_full.pdf`.

Fanshawe TR and Diggle PJ. 2012. Bivariate geostatistical modelling: a review and an application to spatial variation in radon concentrations. *Environmental and Ecological Statistics*, 19(2):139–160.

Green PJ and Silverman BW. *Nonparametric regression and generalized linear models: a roughness penalty approach*. Chapman & Hall London, 1994.

Gryparis A, Paciorek CJ, Zeka A, Schwartz J, and Coull BA. 2009. Measurement error caused by spatial misalignment in environmental epidemiology. *Biostatistics*, 10(2):258–274.

Hanley JA, Negassa A, and Forrester JE. 2003. Statistical analysis of correlated data using generalized estimating equations: an orientation. *American Journal of Epidemiology*, 157 (4):364–375.

Hastie T, Tibshirani R, and Friedman J. *The Elements of Statistical Learning*. Springer Series in Statistics, 2001.

Hodges JS. *Richly Parameterized Linear Models: Additive, Spatial, and Time Series Models Using Random Effects*. Chapman & Hall/CRC Texts in Statistical Science, 2013.

Hoek G, Beelen R, de Hoogh K, Vienneau D, Gulliver J, Fischer P, et al. 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos Environ*, 42(33):7561–7578.

Kammann EE and Wand MP. 2003. Geoadditive models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(1):1–18.

Kim SY, Sheppard L, and Kim H. 2009. Health effects of long-term air pollution: influence of exposure prediction methods. *Epidemiology*, 20(3):442–450.

Liang Kung-Yee and Zeger Scott L. 1986. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

Lindgren F, Rue H, and Lindström J. 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498.

Lopiano KK, Young LJ, and Gotway CA. 2013. Estimated generalized least squares in spatially misaligned regression models with berkson error. *Biostatistics*, 14(4):737–751.

Madsen L, Ruppert D, and Altman NS. 2008. Regression with spatially misaligned data. *Environmetrics*, 19(5):453–467.

Mancl LA and DeRouen TA. 2001. A covariance estimator for gee with improved small-sample properties. *Biometrics*, 57(1):126–134.

Mercer LD, Szpiro AA, Sheppard L, Lindström J, Adar SD, Allen RW, et al. 2011. Comparing universal kriging and land-use regression for predicting concentrations of gaseous oxides of nitrogen (NO_X) for the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). *Atmos Environ*, 45(26):4412–4420.

Miller KA, Siscovick DS, Sheppard L, Shepherd K, Sullivan JH, Anderson GL, et al. 2007. Long-term exposure to air pollution and incidence of cardiovascular events in women. *N Engl J Med*, 356(5):447–458.

NIEHS. 2013. The Sister Study. *http://www.sisterstudy.org/*.

Novotny EV, Bechle MJ, Millet DB, and Marshall JD. 2011. National satellite-based land-use regression: No2 in the united states. *Environmental Science & Technology*, 45(10): 4407–4414.

Olives C. In preparation. Measurement error in the presence of spatial misalignment and non-transportability.

Paciorek CJ. 2007. Bayesian smoothing with Gaussian processes using Fourier basis functions in the spectralGP library. *Journal of Statistical Software*, 19(2).

Peng RD. 2013. Measurement error in air pollution epidemiology: guidance for uncertain times. *Environmetrics*, 24(8):529–530.

Pope CA, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, et al. 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA (J Am Med Assoc)*, 287(9):1132–1141.

Rue H, Martino S, and Chopin N. 2009. Approximate bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.

Ruppert D, Wand MP, and Carroll RJ. *Semiparametric regression*. Cambridge University Press, 2003.

Samet JM, Dominici F, Curriero FC, Coursac I, and Zeger SL. 2000. Fine particulate air pollution and mortality in 20 US cities, 1987–1994. *N Engl J Med*, 343(24):1742–1749.

Sampson PD, Szpiro AA, Sheppard L, Lindström J, and Kaufman JD. 2009. Pragmatic estimation of a spatio-temporal air quality model with irregular monitoring data. *Atmos Environ*, 45(36):6593–6606.

Sampson PD, Richards M, Szpiro AA, Bergen S, Sheppard L, Larson TV, et al. 2013. A regionalized national universal kriging model using partial least squares regression for estimating annual $PM_{2.5}$ concentrations in epidemiology. *Atmospheric Environment*, 75: 383–392.

Schwartz Joel and Coull Brent A. 2003. Control for confounding in the presence of measurement error in hierarchical models. *Biostatistics*, 4(4):539–553.

Shao J. *Mathematical Statistics, 2nd Edition*. Springer, 2003.

Stefanski LA and Cook JR. 1995. Simulation-extrapolation: the measurement error jackknife. *J Am Stat Assoc*, 90(432):1247–1256.

Strand Matthew, Sillau Stefan, Grunwald Gary K, and Rabinovitch Nathan. 2014. Regression calibration for models with two predictor variables measured with error and their interaction, using instrumental variables and longitudinal data. *Statistics in Medicine*, 33(3):470–487.

Szpiro AA and Paciorek CJ. 2013. Measurement error in two-stage analyses, with application to air pollution epidemiology. *Environmetrics*, 24(8):501–517.

Szpiro AA, Paciorek CJ, and Sheppard L. 2011a. Does more accurate exposure prediction necessarily improve health effect estimates? *Epidemiology*, 22(5):680–685.

Szpiro AA, Sheppard L, and Lumley T. 2011b. Efficient measurement error correction with spatially misaligned data. *Biostatistics*, 12(4):610–623.

Tibshirani R. 1996. Regression shrinkage and selection via the lasso. *J Royal Stat Soc. Series B (Meth)*, 58(1):267–288.

van der Vaart AW. *Asymptotic Statistics*. University of Cambridge Press, 1998.

Vedal S and Kaufman JD. 2011. What does multi-pollutant air pollution research mean? *American Journal of Respiratory and Critical Care Medicine*, 183(1):4–6.

Vedal S, Kaufman JD, Larson TV, Sampson PD, Sheppard L, Simpson CD, et al. 2012. University of Washington/Lovelace Respiratory Research Institute National Particle Component Toxicity (NPACT) Initiative: Integrated Epidemiological and Toxicological Cardiovascular Studies to Identify Toxic Components and Sources of Fine Particulate Matter (DRAFT). *Heath Effects Institute, Boston, MA*.

Wakefield J. *Bayesian and Frequentist Regression Methods*. Springer Series in Statistics, 2013.

White H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, 48(4): 817–838.

Wood SN. 2003. Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1):95–114.

Yu Y and Ruppert D. 2002. Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association*, 97(460):1042–1054.

Zeger SL, Thomas D, Dominici F, Samet JM, Schwartz J, Dockery D, et al. 2000. Exposure measurement error in time-series studies of air pollution: concepts and consequences. *Environ Health Perspect*, 108(5):419–426.

Zeka A and Schwartz J. 2004. Estimating the independent effects of multiple pollutants in the presence of measurement error: An application of a measurement-error–resistant technique. *Environmental Health Perspectives*, 112(17):1686–1690.