8-28-2006

Adjustment Uncertainty in Effect Estimation

Ciprian M. Crainiceanu
Johns Hokins Bloomberg School of Public Health, Department of Biostatistics, ccrainic@jhsph.edu

Francesca Dominici
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics, fdominic@jhsph.edu

Giovanni Parmigiani
The Sydney Kimmel Comprehensive Cancer Center, Johns Hopkins University & Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, gp@jimmy.harvard.edu

Adjustment Uncertainty in Effect Estimation

Ciprian M. Crainiceanu* Francesca Dominici †

Giovanni Parmigiani ‡

August 26, 2006

Abstract

The selection of confounders and their functional relationship with the outcome affects exposure effect estimates. In practice, there is often substantial uncertainty about this selection, which we define here as "adjustment uncertainty". We address the problem of estimating the effect of exposure on an outcome with focus on quantifying the effect of unknown confounders from a large set of potential confounders. We propose a general statistical framework for handling adjustment uncertainty in exposure effect estimation, a specific implementation called "Structured Estimation under Adjustment Uncertainty (STEADy)", and associated visualization tools. Theoretical results and simulation studies show that STEADy consistently estimates the exposure of interest and its associated variability. An important by-product of our methodology is that it reveals that the standard version of Bayesian Model Averaging (BMA) can fail to estimate the effect of scientific interest and can over or underestimate statistical variability of the exposure effect estimate. This is essentially due to the fact that BMA averages parameter estimates, only a subset of which can

*Assistant Professor, Department of Biostatistics, Johns Hopkins University, 615 N. Wolfe St. Baltimore, MD 21205 USA. E-mail: ccrainic@jhsph.edu

†Associate Professor, Department of Biostatistics, Johns Hopkins University, 615 N. Wolfe St. Baltimore, MD 21205 USA. E-mail: fdominic@jhsph.edu

‡Professor, Departments of Oncology, Biostatistics, and Pathology Johns Hopkins University, 550 North Broadway, Baltimore, MD 21205 USA. E-mail: gp@jhu.edu

actually be interpreted as being the adjusted effect of interest. While this has been previously acknowledged, our methodology provides the theoretical platform for performance analysis of BMA estimation. We compare our approach (STEADy) to BMA on time series data on levels of fine particulate matter (PM10, PM2.5) and mortality and hospitalization counts.

Keywords: Adjustment uncertainty; Bayesian Model Averaging; Air pollution

1 Introduction

Estimating health risks associated with an exposure X and properly characterizing their uncertainty is one of the most common goals in epidemiology. Regression models are generally used to estimate the effect of X on a response Y, while controlling for unknown confounders from a set of M potential confounders $\boldsymbol{U}$, where M can be very large. Examples include both controlled and observational studies in nearly all fields of scientific investigation. We refer to the uncertainty in the selection of confounders and their functional relationship with the outcome as "adjustment uncertainty". There are important conceptual differences between adjustment uncertainty and model uncertainty, as commonly implemented. In adjustment uncertainty, the estimation of the adjusted exposure effect (e.g. the regression coefficient of X in a model for Y) is the goal of the analysis. Additional covariates $\boldsymbol{U}$ are included to account for potential confounding. In model uncertainty, all predictors $(X, \boldsymbol{U})$ are equally important, and their inclusion into the regression model is generally evaluated based on measures that reflect prediction performance, rather than adequacy in controlling for confounders of X.

We propose a general statistical framework for adjustment uncertainty in exposure effect estimation, a specific implementation called "Structured Estimation under Adjustment Uncertainty (STEADy)", and associated visualization tools. STEADy has two-stages. In the first stage we find a sequence of exposure models, $\alpha_{m,X}$, containing the best set of m predictors, $\boldsymbol{U}_{\alpha_{m,X}}$, of exposure X among the potential confounders

$\boldsymbol{U}$ for every $m = 1, \ldots, M$. We plot the exposure effect estimate for the model with Y as outcome, X as exposure, and $\boldsymbol{U}_{\alpha_{m,X}}$ as confounders for each m. We also plot the deviance difference between models $\alpha_{m,X}$ and $\alpha_{m-1,X}$ to visualize the exposure prediction change. The set of confounders is identified using stabilization of the exposure effect estimates and deviance differences. The second stage includes the set of confounders identified in the first stage, together with X, into the regression models for Y. The outcome model space is then explored using the same technique to identify additional covariates that are good predictors of Y.

A nice feature of STEADy is its practical visualization. For example, we produce plots of point estimates and confidence intervals for the exposure effect as a function of different levels of confounding adjustment. The set of models identified can be used to assist model selection based on expert opinion. Moreover, our methodology provides a sound statistical platform for identifying unknown confounders from a large set of potential confounders and quantifying their effect on exposure effect estimates.

The methodological development of this paper is motivated by time series studies of health effects of air pollution exposure. Time series studies of air pollution and health carried out around the world provide important epidemiological evidence for regulatory purposes [1, 2, 6, 8, 12, 13, 23, 24, 25, 28] and are at the center of an intense national debate in the U.S. [11, 19, 20, 21]. Therefore properly accounting for adjustment uncertainty is of fundamental importance in this context. Air pollution studies estimate whether day-to-day changes in ambient concentrations of air pollution (X) are associated with day-to-day changes in the daily number of deaths or hospital admissions for different diseases (Y) after accounting for time-varying confounders, such as weather and seasonality ($\boldsymbol{U}$). Time series data on pollution and mortality are generally modeled using Poisson regression for over–dispersed counts [4, 8, 14, 26, 28]. The daily number of deaths is the outcome with the, possibly lagged, daily level of pollution being the linear predictor. Air pollution effect estimates on mortality/ morbidity could be affected by observed and unobserved time-dependent confounders (such as weather variables, other pollutants, season, and influenza epidemics) that

vary in a similar manner as the air pollution and mortality/morbidity time series. To account for these, smooth functions of weather, calendar time, and other factors are also included into the semi-parametric Poisson regression model.

There remains substantial controversy in the scientific community about whether current statistical approaches for air pollution effect estimation properly account for the inherent uncertainty in confounding adjustment. For example, the selection of the number of degrees of freedom in the smooth functions of time and temperature (df), and whether to include other important potential confounders in the model such as co–pollutants, can have a large impact on the magnitude and statistical uncertainty of the mortality/morbidity relative rate estimates. In the absence of strong biological hypotheses, these choices have been based on expert judgment [8, 14], or on optimality criteria [4, 28]. Dominici et al. 2004, [7], focus on uncertainty associated with the selection of the number of degrees of freedom in the smooth function of time, and calculate the asymptotic bias and variance of the air pollution risk estimates. One important result is that selecting enough df to best predict air pollution provides more efficient estimates than methods based on selecting df to best predict the mortality outcome. In a simulation study Peng et al., [22], confirmed these results by comparing them with statistical methods commonly used for confounding adjustment in semiparametric regression.

Bayesian Model Averaging (BMA) is a general framework for addressing model uncertainty, by assuming that the true model is an unknown random variable [9, 16]. BMA predictions work out to be weighted averages of the individual models' predictions, using the posterior model probabilities as weights. In prediction, BMA can be justified from a decision theoretic standpoint [3] and performs well compared to model selection [17]. In air pollution research, weighted average of model-specific coefficients using the posterior probabilities as weights has been advocated as a way of handling adjustment uncertainty in the estimation of effects [5, 15].

Our theoretical results and simulations show that STEADy consistently estimates the exposure of interest and its associated variability. They also reveal that the

standard version of Bayesian Model Averaging (BMA) can fail to estimate the effect of scientific interest and over or underestimate the statistical variance of the exposure effect estimate. This is due to the fact that BMA averages parameter estimates only a subset of which can actually be interpreted as being the adjusted effect of interest.

The paper is organized as follows. In section 2 we introduce our general framework for adjustment uncertainty and describe STEADy. In section 3 we present theoretical results. In section 4 we carry out a simulation study to compare adjustment uncertainty versus model uncertainty. In Section 5 we apply our methods to time series data on particulate matter and both mortality and hospital admission counts for COPD. Finally, in section 6 we discuss technical details of the new methodology and addresses practical problems related to air pollution health research and its impact on decision making. The open source R package STEADy implementing this methodology for Generalized Linear Models (GLMs) is publicly available at
`www.biostat.jhsph.edu/~ccrainic/webpage/software/STEADy_1.2.tar.gz`

2 Statistical approaches for Adjustment Uncertainty

2.1 Likelihood models and estimands

In this section we outline our general framework. Suppose that we are interested in the relationship between exposure (X) and outcome (Y), while correcting for the effect of potential confounders $\boldsymbol{U} = (U_1, \ldots, U_M)$. We allow a set of additional covariates $\boldsymbol{Z} = (Z_1, \ldots, Z_K)$ to always enter into the model due to their scientific importance.

A critical point is to correctly identify the estimand. Consider, for example, a model space including models of the form

$$Y_i = \beta^{\alpha} X_i + \sum_{k=1}^{K} \gamma_k^{\alpha} Z_{ik} + \sum_{m=1}^{M} \alpha_m \delta_m^{\alpha} U_{im} + \epsilon_i^{\alpha} \ ,$$

where $i = 1, \ldots, n$, n is the sample size and M is the number of potential confounders. Here $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M)^T \in \{0,1\}^M$ with $\alpha_i = 1$ if and only if the ith covariate is included in the model. We refer to the vector $\boldsymbol{\alpha}$ as the model and use the notation

$\boldsymbol{\alpha} \subseteq \boldsymbol{\alpha}'$ if the model $\boldsymbol{\alpha}$ is nested within model $\boldsymbol{\alpha}'$. The meaning of the coefficient β^{α} may vary with $\boldsymbol{\alpha}$.

We assume that the M potential confounders include all the relevant confounders necessary for identifying the true exposure effect. In our notation this means that the true model $\boldsymbol{\alpha}^*$ belongs to the model space and we denote by $\beta^* = \beta^{\alpha^*}$ the corresponding true effect. In addition we assume that all models nesting $\boldsymbol{\alpha}^*$ estimate the same true effect, so $\beta^{\alpha} = \beta^*$ whenever $\boldsymbol{\alpha}^* \subseteq \boldsymbol{\alpha}$.

In this setting, adjustment uncertainty in effect estimation is estimation of β^* when $\boldsymbol{\alpha}^*$ is unknown. Identifying $\boldsymbol{\alpha}^*$ is a hard problem, but identifying a reasonable $\boldsymbol{\alpha}_0$ such that $\boldsymbol{\alpha}* \subseteq \boldsymbol{\alpha}_0$ is easier and will be the focus of the STEADy procedure described in the next section. Note that if $\boldsymbol{\alpha}^* \subseteq \boldsymbol{\alpha}$ then we can provide unbiased estimators $\hat{\beta}^{\alpha}$ of β^*. Typically $\mathrm{Var}(\hat{\beta}^{\alpha}) \geq \mathrm{Var}(\hat{\beta}^*)$ and confidence intervals based on model $\boldsymbol{\alpha}$ will tend to be conservative.

In contrast, considering any model $\boldsymbol{\alpha}$ such that $\boldsymbol{\alpha}^* \nsubseteq \boldsymbol{\alpha}$ will result in an incorrect estimand of the exposure effect. Using such a model would provide invalid inferences from either a frequentist or a Bayesian standpoint. For example, a standard Bayesian Model Averaging estimate of β^* is obtained by specifying a prior $p(\boldsymbol{\alpha})p(\beta^{\alpha}, \boldsymbol{\gamma}^{\alpha}, \boldsymbol{\delta}^{\alpha}|\boldsymbol{\alpha})$ and forming the estimator

$$\tilde{\beta}^* = \sum_{\boldsymbol{\alpha}:\boldsymbol{\alpha}^* \subseteq \boldsymbol{\alpha}} E[\beta^{\alpha}|\boldsymbol{\alpha}, D]p(\boldsymbol{\alpha}|D) + \sum_{\boldsymbol{\alpha}:\boldsymbol{\alpha}^* \nsubseteq \boldsymbol{\alpha}} E[\beta^{\alpha}|\boldsymbol{\alpha}, D]p(\boldsymbol{\alpha}|D) \qquad (1)$$

where D denotes the data and $p(\boldsymbol{\alpha}|D)$ is the posterior probability of model $\boldsymbol{\alpha}$. In expression (1) assigning any weight to models $\boldsymbol{\alpha}$ such that $\boldsymbol{\alpha}^* \nsubseteq \boldsymbol{\alpha}$ will result in an additional term that includes estimates of the incorrect quantities. In Section 6 we illustrate how this can lead to seriously misleading results.

Motivated by these considerations we define and address two problems related to adjustment uncertainty. First, to identify a reasonable $\boldsymbol{\alpha}_0$ such that $\boldsymbol{\alpha}^* \subseteq \boldsymbol{\alpha}_0$. Second, to devise a computationally tractable framework that describes the relationship between inference about β^{α} and the model space structure. The former is a global problem that is often addressed by considering a reasonably large full model. While

this strategy is reasonable in many applications, it may fail when the number of observations is small and/or the number of potential confounders is large. Moreover, using only the full model would ignore important information about the effect of various confounders on the exposure effect inference. This information can be recovered using a structured search of the model space.

2.2 STEADy: Structured Estimation under Adjustment Uncertainty

We address the two problems described in the previous section by devising STEADy, an algorithm for exploration of the exposure and outcome model spaces. In particular, STEADy will identify a reasonable model $\boldsymbol{\alpha}_0$ such that $\boldsymbol{\alpha}^* \subseteq \boldsymbol{\alpha}_0$.

Conceptually, STEADy focuses on the joint model for outcome and exposure

$$[Y, X|\boldsymbol{Z}, \boldsymbol{U}] = [Y|X, \boldsymbol{Z}, \boldsymbol{U}][X|\boldsymbol{Z}, \boldsymbol{U}] \; . \tag{2}$$

and can be described in two steps. In the first step, we identify strong predictors of exposure and focus on the exposure model space $\mathcal{M}_X$. More precisely, in $\mathcal{M}_X$ the exposure X plays the role of the response, and potential predictors include covariates $\boldsymbol{Z}$ and any subset of $\boldsymbol{U} = (U_1, \ldots, U_M)$. We then identify the potential confounders $\boldsymbol{U}$ that are strongly predictive of X and we include them in the full model for Y.

In Section 2.1, we introduced $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M)^T \in \{0, 1\}^M$ with $\alpha_i = 1$ if and only if U_i is in the model. Here we also denote by $\boldsymbol{U}_\alpha$ the subset of potential confounders of $\boldsymbol{U}$ selected by the indicator vector $\boldsymbol{\alpha}$. We divide $\mathcal{M}_X$ into $M+1$ subsets, or orbits, corresponding to models with a fixed number of confounders $\boldsymbol{U}$. For example, the mth orbit is the set of all regression models $\boldsymbol{\alpha}$ with the property $\sum_{i=1}^{M} \alpha_i = m$.

Within orbit m we select the maximum likelihood model and denote by $\boldsymbol{\alpha}_{m,X}$ the model maximizing the likelihood. While computationally intensive, finding the maximum likelihood model within each orbit avoids the typical problems of greedy algorithms such as forward selection or projection–pursuit. Details on the likelihood maximization are provided in the technical Appendix A1. The set of $M+1$ regression models $\boldsymbol{\alpha}_{0,X}, \ldots, \boldsymbol{\alpha}_{M,X}$ will be called the dominant model class. Once the dominant

model class is determined, it is useful to display the deviance differences $D(\boldsymbol{\alpha}_{m,X}) - D(\boldsymbol{\alpha}_{m+1,X})$, which show the change in deviance between adjacent orbits and indicate what combinations of variables are good predictors of X and are likely confounders. The region where the deviance difference function becomes small identifies a range for the required dimensionality of the exposure model. Such regions characterize the right hand side of the graph and are usually easy to identify visually. We recommend exploring the entire dominant model class and understanding the sequence in which variables are included into the model. An important visualization is to show the sequence of point estimates and confidence intervals for the exposure effect in the outcome models $[Y|X, \boldsymbol{Z}, \boldsymbol{U}_{\alpha_{0,X}}], \ldots, [Y|X, \boldsymbol{Z}, \boldsymbol{U}_{\alpha_{M,X}}]$.

In the second step we identify strong predictors of outcome that are weak predictors of exposure. Thus, the second step of STEADy starts by adding the subset of $\boldsymbol{U}$'s identified at the previous step to $\boldsymbol{Z}$ and deleting it from $\boldsymbol{U}$. Using a procedure similar to the one for exposure models, the outcome models are partitioned into $L+1$ orbits. Here $L \leq M$ because the number of potential confounders that are considered for exclusion is usually reduced during the first phase of the procedure. Denote by $\boldsymbol{\alpha}_{l,Y}$ the model that maximizes the likelihood of the outcome models on the lth orbit and call $\boldsymbol{\alpha}_{0,Y}, \ldots, \boldsymbol{\alpha}_{L,Y}$ the dominant model class for the outcome models. The plot of deviance differences $D(\boldsymbol{\alpha}_{l,Y}) - D(\boldsymbol{\alpha}_{l+1,Y})$ indicates what covariates, in addition to X, $\boldsymbol{Z}$, are predictive of Y. As before one is interested in finding the region where the deviance difference function becomes small. We also display the point estimates and confidence intervals for the parameter of X for the outcome models $[Y|X, \boldsymbol{Z}, \boldsymbol{U}_{\alpha_{0,Y}}], \ldots, [Y|X, \boldsymbol{Z}, \boldsymbol{U}_{\alpha_{L,Y}}]$. This plot captures variations of the effect of X on Y when adjusting for $\boldsymbol{Z}$ and a subset of $\boldsymbol{U}$. The point estimate and confidence interval for the parameter of interest is obtained after both the difference deviance function for outcome models and parameter estimate have stabilized. Stabilization is evidence that $\boldsymbol{\alpha}^*$ was reached.

In summary, the STEADy algorithm includes the following steps:

1. Obtain the dominant model class $\text{DMC}_X = \boldsymbol{\alpha}_{0,X}, \ldots, \boldsymbol{\alpha}_{M,X}$ of exposure models

2. Plot deviance differences for the exposure models $\boldsymbol{\alpha}_{0,X}, \ldots, \boldsymbol{\alpha}_{M,X}$

3. Plot point estimates and confidence intervals for the exposure parameter in the sequence of outcome models $[Y|X, \boldsymbol{Z}, \boldsymbol{U}_{\alpha_{0,X}}], \ldots, [Y|X, \boldsymbol{Z}, \boldsymbol{U}_{\alpha_{M,X}}]$

4. Identify $\boldsymbol{U}_{R,X}$ such that $D(\boldsymbol{\alpha}_{m,X}) - D(\boldsymbol{\alpha}_{m+1,X})$ is small for every $m \geq R$

5. Define $\boldsymbol{Z} = \boldsymbol{Z} \bigcup \boldsymbol{U}_{R,X}$, $\boldsymbol{U} = \boldsymbol{U} \setminus \boldsymbol{U}_{R,X}$, $L = M - R$.

6. Obtain the dominant model class $\text{DMC}_Y = \boldsymbol{\alpha}_{0,Y}, \ldots, \boldsymbol{\alpha}_{L,Y}$ of outcome models

7. Plot deviance differences for the outcome models $\boldsymbol{\alpha}_{0,Y}, \ldots, \boldsymbol{\alpha}_{L,Y}$

8. Plot point estimates and confidence intervals for the exposure parameter in the sequence of models $\boldsymbol{\alpha}_{0,Y}, \ldots, \boldsymbol{\alpha}_{L,Y}$

9. Identify the region where $D(\boldsymbol{\alpha}_{l,Y})$-$D(\boldsymbol{\alpha}_{l+1,Y})$ becomes small

10. Among models that provide similar exposure effect estimates identify the one with smallest exposure effect variance

This algorithm is designed to assist scientists in exploring the model space in a structured way while keeping the focus on exposure effect estimation. Automatic use of the algorithm is possible, but may not always be successful, while active and critical scientific analyses assisted by STEADy should be preferred. Practical issues related to implementation of the STEADy algorithm are provided in Appendix A1.

3 Theoretical results for Adjustment Uncertainty

In this section we explore the theoretical properties of the STEADy estimator and compare it with weighted estimators. A particular case of weighted estimator is the Bayesian Model Averaging (BMA) estimator for which the weights are the posterior model probabilities. Even though BMA has been proposed to incorporate model uncertainty in prediction, its use for exposure effect estimation has been advocated [5, 10, 15] as a procedure that incorporates the uncertainty about model selection.

We first introduce some notations. Let $\boldsymbol{\alpha}$ be a general model, $\boldsymbol{\alpha}_E$, $\boldsymbol{\alpha}^*$, $\boldsymbol{\alpha}_F$ be the model with no additional covariates, the true model, and the full model respectively.

Let R be the set of confounders for $\boldsymbol{\alpha}^*$ and I be the rest of the confounders. Let $\widehat{\beta}_n^{\alpha}$ be the MLE of β^{α}, the exposure effect under model α and assume that, under $\boldsymbol{\alpha}^*$, $\widehat{\beta}_n^{\alpha} \to \beta^{\alpha}$. For a set of weights $\boldsymbol{W}_n = \{w_n(\alpha)\}_{\alpha}$ with $\sum_{\alpha} w_n(\alpha) = 1$ and $w_n(\alpha) \geq 0$ denote the weighted estimator $\widehat{\beta}_{W_n} = \sum_{\alpha} w_n(\boldsymbol{\alpha})\widehat{\beta}_n^{\alpha}$. Because $\boldsymbol{W}_n$ is in a compact set it contains at least one convergent subsequence and all $\widehat{\beta}_n^{\alpha}$ sequences are convergent. To ensure convergence of the weighted estimator $\widehat{\beta}_{W_n}$ we will assume that $\boldsymbol{W}_n$ itself converges to a limit, say $\boldsymbol{W} = \{w(\boldsymbol{\alpha})\}_{\alpha}$. Let $\widehat{\beta}_n^{\text{STEADy}} = \widehat{\beta}_n(\widehat{\boldsymbol{\alpha}}_n^{\text{STEADy}})$ where

$$\widehat{\boldsymbol{\alpha}}_n^{\text{STEADy}} = \text{Argmin}\{\widehat{v}_n(\boldsymbol{\alpha}) : |\widehat{\beta}_n^{\alpha} - \widehat{\beta}_n^{\alpha_F}| < \frac{\epsilon}{\sqrt{n}}\}, \tag{3}$$

$\widehat{v}_n(\boldsymbol{\alpha}) = \widehat{\text{Var}}_{\alpha}(\widehat{\beta}_n^{\alpha})$ and $\epsilon > 0$ is a constant.

Theorem 1 *Under typical regularity assumptions*

i. $\widehat{\beta}_n^{\text{STEADy}}$ *is asymptotically unbiased*

ii. The asymptotic bias of $\widehat{\beta}_{W_n}$ *is* $\text{Bias}\{\widehat{\beta}_{W_n}\} = \sum_{\boldsymbol{\alpha}\in I} w(\boldsymbol{\alpha})\beta^{\alpha} - \left\{\sum_{\boldsymbol{\alpha}\in I} w(\boldsymbol{\alpha})\right\}\beta^*$

iii. If $v(\boldsymbol{\alpha}) = \lim_{n\to\infty}\{n\widehat{v}_n(\boldsymbol{\alpha})\}$ *then* $\overline{\lim}_{n\to\infty}\{n\widehat{v}_n(\widehat{\boldsymbol{\alpha}}_n^{\text{STEADy}})\} \leq v(\boldsymbol{\alpha}_F)$*. If, in addition, there exists an* $\boldsymbol{\alpha}_0$ *such that* $\boldsymbol{\alpha}^* \subset \boldsymbol{\alpha}_0$ *with* $v(\boldsymbol{\alpha}_0) < v(\boldsymbol{\alpha}_F)$ *then the above inequality is strict.*

The proof of the theorem is in Appendix A2. Note that if $w(\boldsymbol{\alpha}) = 0$ for every $\alpha \in I$ the weighted estimator will be asymptotically unbiased. However, if this is not true the asymptotic unbiasedness could only be the result of a lucky choice of weights. Indeed, the asymptotic bias is zero iff

$$\beta^* = \sum_{\boldsymbol{\alpha}\in I} \frac{w(\boldsymbol{\alpha})}{\sum_{\boldsymbol{\alpha}\in I} w(\boldsymbol{\alpha})}\beta^{\alpha},$$

which can only happen by accident when β^* and the set I are unknown. Moreover, if $\beta^{\alpha} > \beta^*$ for every $\alpha \in I$ then all weighted estimators with positive weights (including BMA) will be biased.

While the minimum variance of the STEADy estimator is ensured by definition in finite samples, the previous result shows that any subsequence of STEADy estimators for which the limit of $n\widehat{v}_n(\alpha)$ exists is asymptotically unbiased with asymptotic

variance smaller than the asymptotic variance of the full model. We also provide a sufficient condition under which the dominance is strict.

A popular choice is to use the Bayesian Information Criterion (BIC) weights

$$w_n(\boldsymbol{\alpha}) = \frac{\exp\{-\mathrm{BIC}_n(\boldsymbol{\alpha})/2\}}{\sum_{\boldsymbol{\alpha'}} \exp\{-\mathrm{BIC}_n(\boldsymbol{\alpha'})/2\}},$$

where $\mathrm{BIC}_n(\boldsymbol{\alpha}) = 2\log\{\rho_n(\boldsymbol{\alpha}) + p_\alpha \log(n)\}$, and $\rho_n(\boldsymbol{\alpha})$ and p_α are the maximum likelihood and the number of parameters of model α. The BIC is not only popular as an approximation to Bayesian posterior probabilities, but also as a model selection criterion in its own right. We examine now a property of BIC in this broader context.

Theorem 2 *Assume the usual regularity conditions (e.g. Cramèr, 1999) for existence, consistency and asymptotic normality of the MLEs and $w_n(\boldsymbol{\alpha})$ are the BIC weights. If there exists a $c > 0$ such that $|\beta^* - \beta^\alpha| \leq c/\sqrt{n}$ for every model $\boldsymbol{\alpha}$ then the asymptotic bias of $\widehat{\beta}_{W_n}$ is $\beta^{\alpha_E} - \beta^*$, or the bias induced by the model without confounders.*

The proof is in Appendix A2. This result shows that if parameters are all within $c/\sqrt{n}$ distance of β^* then the BIC based weighted estimator converges to the parameter of the model with no confounders rather than to the parameter of interest. Note that if the model that does not contain exposure is considered as part of the model then the BIC based weighted estimator converges to *zero*, or no exposure effect, regardless of the size of the signal. The condition on the parameters is the local asymptotics assumption, which is widely used for confidence intervals inference.

The finite sample behavior of the BIC when the sample size is large and the signal is small to moderate can be illustrated using a simple example. Suppose that only two nested models are available $\boldsymbol{\alpha}_0 \subset \boldsymbol{\alpha}_1$, with $\boldsymbol{\alpha}_1$ having just one extra parameter. BIC selects model M_0 if and only if $\mathrm{BIC}(\boldsymbol{\alpha}_0) < \mathrm{BIC}(\boldsymbol{\alpha}_1)$ which is equivalent to $2\log\{\rho(\boldsymbol{\alpha}_1)/\rho(\boldsymbol{\alpha}_0)\} > \log(n)$. The left hand side of this equation is the likelihood ratio statistic for testing $\boldsymbol{\alpha}_0$ versus $\boldsymbol{\alpha}_1$. In this case BIC is equivalent to using likelihood ratio test with a critical value equal to $\log(n)$. This may be reasonable when n is small or moderate. For example, $\log(n)$ varies between 3.9 and 6.2 when n varies

between 50 and 500. These values are comparable to the 0.95 and 0.99 quantiles of a χ^2_1 distribution, which are 3.84 and 6.63 respectively. Even in this case when the sample size, n, increases, BIC will change the level of the test from roughly $\alpha = 0.05$ to $\alpha = 0.01$. More dramatic effects occur when the sample size is large with the level of the test dropping to $\alpha = 0.002$ for $n = 15,000$ and $\alpha = 0.0005$ for $n = 150,000$. Thus, using BIC in large data sets might ignore even very strong signals and lead to wrong conclusions. This discussion is of particular interest for studies of air pollution effects on mortality, which routinely use over $10,000$ observations per city.

4 Simulation Study

In this section we describe a simulation study to compare adjustment uncertainty implemented with STEADy versus model uncertainty implemented with BMA. We consider a relatively simple data generating mechanism to allow a transparent comparison across methods, and yet capture important features of effect estimation in air pollution research and other areas. We generate data from the following model:

$$\begin{cases} Y_i & \sim \text{Poisson}(\mu_i) \\ \log(\mu_i) & = \beta_0 + \beta_1 X_{1i} + \beta_2 U_{2i} + \beta_3 U_{3i} \end{cases} \tag{4}$$

where $i = 1, \ldots, 1000$ and (X_{1i}, U_{2i}, U_{3i}) are independent normal vectors with mean zero, variance 1 and all covariances zero except for $\text{Cov}(X_{1i}, U_{2i}) = \rho$. Here X_1 is the exposure variable, U_2 is a confounder that predicts both X_1 and Y while U_3 predicts Y but is independent of the exposure. To define a correctly adjusted effect of X_1, the confounder U_2 must be included into the model. The set $\boldsymbol{U}$ of potential confounders includes U_2, U_3 as well as 50 additional independent $N(0,1)$ random variables. In this example there are no $\boldsymbol{Z}$ variables. The model space $\mathcal{M}$ has 2^{52} models if we include the exposure variable by default, and 2^{53} if we allow for it to be included or excluded.

We explore two simulation scenarios: the first draws a new set of Xs, Us and Ys for each new data set while the second draws a new set of Ys only and keeps

the same set of Xs and Us. We simulated 100 data sets from model (4) under each scenario. We set the parameter ρ to 0, corresponding to no need for adjustment, or to 0.7 corresponding to a moderate effect adjustment. The coefficients are set at $\beta_1 = \beta_2 = \beta_3 = 0.1$ and $\beta_0 = 0$.

This example includes several important characteristics common in air pollution effect estimation where Y_i and X_{1i} can be viewed as the daily mortality counts and lagged pollution level and U_{2i} can be viewed as an important confounder when $\rho \neq 0$. The additional 50 covariates represent noise that makes the estimation problem difficult. Parameters values were chosen to mimic the significance levels of typical effects in air pollution studies, given the simulation sample size.

4.1 Results when the exposure is always in the model

For each simulated data set, we estimated β_1 using STEADy, and using BMA over the dominant model class, with posterior model probabilities approximated using three methods: BIC (BIC-DMC), "Likelihood Ratio Test" (LRT-DMC) and AIC (AIC-DMC). These criteria correspond respectively to a penalty of $\log(1000) = 6.9$, 4 and 2 for each additional variable. For BIC we also report results based on a full stochastic exploration of the model space (BIC-SE). Because BMA approaches are based on prediction and therefore on outcome models, we first implement STEADy on the outcome's dominant model class only (STEADy-ODMC), thus allowing for a more direct comparison between the two approaches. The preferred application of STEADy is two-stage and involves a preliminary search for confounders in the exposure model.

When $\rho = 0$, that is in absence of confounding, all methods produced very similar estimates of β_1 (results not reported). Table 1 summarizes the average, $\mathrm{M}(\widehat{\beta}_1)$, and standard error, $\mathrm{SE}(\widehat{\beta}_1)$, of the β_1's estimates over 100 data sets for $\rho = 0.7$ and for the two simulation scenarios described above (X and U fixed and X and U random). Under STEADy-ODMC we report estimates of β_1 after they stabilize, which occurs by the 40-th orbit. In both scenarios, STEADy-ODMC and BMA produce pronouncedly

different results, and irrespective of the penalty used, STEADy-ODMC dominates BMA both in terms of bias and variance of estimation. The bias of BMA with BIC-DMC and BIC-SE is such that the (1-10^{-6})% approximate confidence interval based on 100 simulations does not contain the true value of the parameter. When BMA with AIC-DMC is used, although the penalty for including new variables is reduced, the 95% confidence interval still does not contain the true value of the parameter.

Table 1: Mean and standard error of β_1s estimates over 100 data sets for $\rho = 0.7$. The true parameter is $\beta_1 = 0.1$ STEADy here denotes the implementation of our algorithm to the outcome's dominant model class only (STEADy-ODMC). Results for random $\boldsymbol{U}$s correspond to the scenario where both covariates and outcomes were simulated. Results for fixed $\boldsymbol{U}$s correspond to the scenario where only the outcome is simulated.

Scenario		STEADy-ODMC	BIC-DMC	BIC-SE	LRT-DMC	AIC-DMC
Random	M($\widehat{\beta}_1$)	0.1020	0.1330	0.1310	0.1230	0.1130
Random	SE($\widehat{\beta}_1$)	0.0048	0.0057	0.0055	0.0059	0.0055
Fixed	M($\widehat{\beta}_1$)	0.1020	0.1350	0.1240	0.1210	0.1110
Fixed	SE($\widehat{\beta}_1$)	0.0040	0.0048	0.0054	0.0040	0.0045

Figure 1 illustrates the mechanism that determines such different performance in the two approaches. The top right plot shows the estimates of β_1 obtained under STEADy-ODMC plotted against each of the 52 models considered, for each of 3 data sets and for $\rho = 0.7$. The horizontal line is placed at the true value of $\beta_1 = 0.1$. The noticeable drop in point estimates occurs when U_2 enters the model. STEADy chooses the parameter estimate after stabilization of the deviance difference function (see columns 1 and 2) and after stabilization of the β_1 estimates (in column 3).

The left column of Figure 1 shows deviance differences plotted against models under STEADy-ODMC. The horizontal lines are placed at the BIC and AIC penalties which are equal to log(1000) = 6.91 and 2, respectively. Results are shown for 3 data sets simulated from model (4) with $\rho = 0.7$ (top row) and $\rho = 0$ (bottom row).

These plots highlight the severity of the BIC penalty. Except for very few initial models, deviance differences are always smaller than the penalty and therefore they favor the more parsimonious models. To see this, it is useful to write

down the following: $P(\boldsymbol{\alpha}_{k+1})/P(\boldsymbol{\alpha}_k) = \exp\left[\{D(\boldsymbol{\alpha}_k) - D(\boldsymbol{\alpha}_{k+1}) - 6.91\}/2\right]$, where $\boldsymbol{\alpha}_k$ and $\boldsymbol{\alpha}_{k+1}$ are two nested models in adjacent orbits of the dominant class, $P(\boldsymbol{\alpha}_{k+1})$ and $P(\boldsymbol{\alpha}_k)$ are the corresponding posterior model probabilities, and $D(\boldsymbol{\alpha}_{k+1})$ and $D(\boldsymbol{\alpha}_k)$ are the corresponding deviances. Therefore if $D(\boldsymbol{\alpha}_k) - D(\boldsymbol{\alpha}_{k+1}) < 6.91$ then $P(\boldsymbol{\alpha}_k) > P(\boldsymbol{\alpha}_{k+1})$. For example, if the deviance difference $D(\boldsymbol{\alpha}_k) - D(\boldsymbol{\alpha}_{k+1}) = 4$ then $P(\boldsymbol{\alpha}_k) = 4.28P(\boldsymbol{\alpha}_{k+1})$. In this case the likelihood ratio test would reject the model $\boldsymbol{\alpha}_k$ in favor of $\boldsymbol{\alpha}_{k+1}$ at the $\alpha = 0.05$ level when, at the same time, the BIC based model probabilities would assign more than 4 times as much probability mass to the model $\boldsymbol{\alpha}_k$ than to the model $\boldsymbol{\alpha}_{k+1}$.

By inspecting the list of variables that enter in every model we noted that when $\rho = 0.7$, the first model includes only U_3 in addition to X_1 but does not include U_2. This is because the deviance gain of including U_2 is typically in the range $[3, 8]$ and therefore BIC assigns higher probability mass to the model including only U_3. In this model, the estimate of β_1 is artificially inflated because U_2 is correlated with X_1 and not included into the model.

A limitation of BMA with BIC based posterior model probabilities is that this procedure averages β_1 estimates that have different interpretation across models: one is appropriately adjusted because U_2 is included in the model, while the other is a biased estimate of β_1 because U_2 is excluded. BMA with posterior model probabilities based on information criteria with smaller penalties also suffer from the same limitation, though in this example they tend to do slightly better, as the penalty is small irrespective of the sample size.

Next we consider the complete, two-stage, STEADy procedure which first explores the exposure model's dominant class and then the outcome model's dominant class (see Section 2.2). The second column of Figure 1 shows deviance differences plotted against models obtained by applying STEADy to the exposure's dominant class (STEADy-EDMC). The horizontal lines are placed at the BIC and AIC penalties which are equal to $\log(1000) = 6.91$ and 2, respectively. Results are shown for 3 data sets simulated from model (4) with $\rho = 0.7$ (top row) and $\rho = 0$ (bottom row). When

$\rho = 0.7$, we found that STEADy-EDMC identifies U_2 as an important predictor of X_1. We then force U_2 into the outcome model and apply STEADy-ODMC: here we found that the β_1s estimates stabilize at the 5th orbit roughly, much earlier than STEADy-ODMC. The two step STEADy procedure, STEADY-EDMC followed by STEADy-ODMC, produced slightly better results than STEADy-ODMC alone for $\rho = 0.7$ and similar results for $\rho = 0$.

Figure 2 compares the standard errors of the estimates of β_1 for 100 independent data sets obtained using STEADy-ODMC (X–axis) versus BMA with BIC-DMC (Y–axis), when $\rho = 0.7$ (top) and $\rho = 0$ (bottom). When $\rho = 0.7$, BMA can produce up to 35% wider or narrower confidence intervals than STEADy-ODMC. Very small standard errors for BMA correspond to models that do not include the key confounder U_2. In these models, part of the signal generated by U_2 is captured by the parameter estimate of X_1 and since X_1 and U_2 are positively correlated, $\widehat{\beta}_1$ is biased upwards and its statistical uncertainty is understated.

A different mechanism operates when BMA provides larger standard errors than STEADy. Because of the large BIC penalty, BMA assigns sizeable probability both to models that contain and do not contain U_2. In these cases, the $\hat{\beta}_1$s standard error obtained from BMA is artificially inflated because β_1 is estimated by averaging model-specific estimates that include and do not include U_2. STEADy avoids these problems because it relies on stabilization of point estimates and confidence intervals to ensure that all potential confounders are included into the model.

Finally, when $\rho = 0$, BMA and STEADy produce almost identical $\hat{\beta}_1$s standard errors. The standard errors produced by STEADy are potentially sensitive to the final orbit chosen, but are highly stable within a wide range of plausible choices.

In this example, we conclude that in the presence of confounding, BMA produces a biased estimate of the parameter of scientific interest (the effect of X_1 on Y), and that the 95% confidence interval of the parameter's estimate may be as much as 35% wider or narrower than the interval produced by STEADy, which has approximately correct coverage.

4.2 Results when the exposure is not always in the model

An important difference between STEADy and BMA is that STEADy requires that the covariate of interest, say X, is always included in the model whereas BMA does not. Therefore under BMA we could average estimates of the regression coefficient of X when X is included into the model with 0 when X is not. Therefore, in this approach, it is useful to estimate and monitor the posterior probability that X is included into the model.

As in Section 4.1, we simulated 100 independent data sets from model (4) with $\rho = 0.7$ and applied BMA implemented by BIC-DMC without requiring that X_1 be included in the model. Figure 3 (top panel) shows the sorted posterior probability that X_1 is included into the model against the 100 simulated data sets. The bottom panel shows the histogram of the averages of the β_1 estimates across the 100 data sets $\mathrm{M}(\widehat{\beta}_1)$ under the different models. The true parameter is $\beta_1 = 0.1$ and the t–test for $\beta_1 = 0$ has p–values ranging from 0.1 to 0.01. For 25 out of the 100 data sets, the posterior probability that X_1 is included into the model is less than 0.1 and for roughly 40 out of 100 data sets such posterior probability is less than 0.2. Moreover $\mathrm{M}(\widehat{\beta}_1)$ decreases by more than 30%, from 0.133 when X_1 is forced into the model, to 0.089 when X_1 is not. The standard error $\mathrm{SE}(\widehat{\beta}_1)$ increases 46% from 0.0057 to 0.0083. The bottom plot in Figure 3 demonstrates that while the central value of the histogram is equal to 0.089 and is close to the true effect $\beta_1 = 0.1$, the shape of the distribution of the $\mathrm{M}(\widehat{\beta}_1)$ is multimodal with little probability mass around 0.1. Note that there is roughly a 50% chance of underestimating the true effect by more than 50% and a 32% chance of overestimating the true effect by more than 50%.

5 Estimating Health Effects of Air Pollution

Studies of health effect of air pollution pose several methodological challenges related to detecting and quantifying weak signals in large data sets. Commonly used Poisson regression models for time series analyses of air pollution and health ([4, 7, 8, 14, 26, 28]) can easily include hundreds of covariates and are generally applied to very large

data sets with several thousands of observations.

In this section we illustrate two applications of STEADy to time series data on air pollution and health. In the first application, we use time series data for Detroit for the period 1987 to 2000 to estimate the % increase in mortality associated with a 10 μ_g/m^3 increase in PM_{10} (relative rate) accounting for adjustment uncertainty. This data set is part of the ongoing National Morbidity Mortality Air Pollution Study (NMMAPS) which includes time series data for the period 1987-2000 for the 100 largest cities [6, 22]. R-software and the NMMAPS data are available at `www.ihapss.jhsph.edu`.

In the second application, we apply and compare STEADy and BIC-based BMA to time series data for five large US counties (population larger than 200,000) to estimate the % increase in hospital admission rates associated with 10 μ_g/m^3 in $PM_{2.5}$. This second data set is part of a recently started multi-site time series study, called the National Medicare Study. This study is aimed at estimating national, regional, and county-specific relative rates for hospital admissions for several diseases associated with exposure to $PM_{2.5}$ for 205 US counties and for the period 1999-2002 ([2]). Hospital admission rates from the Medicare cannot be made publicly available.

5.1 PM_{10} and all-cause mortality: Detroit 1987-2000

In this section we apply STEADy to daily time series data in Detroit for the period 1987-2000. We consider the following model specification:

$$\begin{cases} Y_t & \sim \text{Poisson}(\mu_t, v_t),\ v_t = \phi\mu_t \\ \log(\mu_t) & = \text{PM}_{10t-1} + \text{Dow} + \text{Age}_2 + \text{Age}_3 + \text{O}_{3t-1} \\ & \quad + ns(\text{Temp}_t, \text{df}_{\text{Temp}}) + ns(\text{Temp}_{t1-3}, \text{df}_{\text{Temp}}) + ns(\text{Dew}_t, \text{df}_{\text{Dew}}) + ns(\text{Dew}_{t1-3}, \text{df}_{\text{Dew}}) + \\ & \quad + ns(t, \text{df}_t) + ns(t, 4) \times \text{Age2} + ns(t, 4) \times \text{Age3} \\ \log(\mu_t) & = \text{PM}_{10t-1} + \text{confounders} \end{cases} \tag{5}$$

where $\mu_t = E[Y_t|\mu_t, \phi]$ is the expected number of deaths, ϕ is the overdispersion parameter, PM_{10t-1} denotes particulate matter of a diameter smaller than 10 microns on day $t-1$, DOW is a categorical variable indicating the day of the week, Age_2 and

Age_3 are indicator variables of the age groups $(65-74)$ and (≥ 75), and O_{3t-1} is the ambient ozone level on day $t-1$. In addition $\text{Temp}_{1-3} = (\text{Temp}_t + \text{Temp}_{t-1} + \text{Temp}_{t-2})/3$ and $\text{Dew}_{1-3} = (\text{Dew}_t + \text{Dew}_{t-1} + \text{Dew}_{t-2})/3$ denote three-day averages of past temperature and dew point temperature levels, respectively. We include $ns(\text{Temp}_t, \text{df}_{\text{Temp}}), ns(\text{Temp}_{t1-3}, \text{df}_{\text{Temp}}), ns(\text{Dew}_t, \text{df}_{\text{Dew}})$ and $ns(\text{Dew}_{t1-3}, \text{df}_{\text{Dew}})$ to adjust for the potential non linear confounding effects of temperature and dew point temperature, where $ns(\cdot, df)$ denotes a natural cubic spline with df degrees of freedom. We also include $ns(t, \text{df}_\text{t})$ to adjust for seasonal variations in mortality rates due to unmeasured confounders such as influenza epidemics. Finally, we add the interaction terms $ns(t,4) \times \text{Age2}$ and $ns(t,4) \times \text{Age3}$ to allow these seasonal variations to be different across age groups.

This model specification has been extensively used and discussed in previous NMMAPS analyses [8, 14]. Specifically, the NMMAPS basic model can be defined by the Equation (5) with $\text{df}_{\text{Temp}} = \text{df}_{\text{Dew}} = 6$, $\text{df}_\text{t} = 8$ per year for a total of 112 over 14 years. Model choice and sensitivity analyses with respect to the selection of the number of degrees of freedom in the smooth function of time are discussed in [7, 22] and with respect to the selection of the number of degrees of freedom in the smooth functions of temperature and dewpoint are discussed in Welty and Zeger 2005, [29].

To apply both STEADy and BMA we need to specify a list which is likely to include all the potential confounders. Specifically, we assume that the full model (that is the model that includes all potential confounders) has twice the number of degrees of freedom in the smooth functions of time and temperature than the basic NMMAPS model ($\text{df}_{\text{Temp}} = \text{df}_{\text{Dew}} = 12$, $\text{df}_\text{t} = 16$ per year).

The first step is to identify good predictors of PM_{10t-1} in the model space:

$$\text{PM}_{10t-1} = \text{confounders} + \epsilon_t \ . \tag{6}$$

The total number of confounders in this model is 289. Due to singularities in the design matrix we eliminated 55 spline basis. Thus, we use STEADy to explore the model space generated by the remaining 234 confounders. We then obtain the exposure's dominant model class $\text{DMC}_X = \boldsymbol{U_{0,X}}, \ldots, \boldsymbol{U_{234,X}}$ for the exposure (6)

using the stochastic search described in Appendix A1. We used 500 iterations per orbit for a total of $500 \times 234 = 117,000$ iterations with an average computation time 1.62 seconds/iteration on a PC (3.4GHz CPU, 3.4 Gb RAM). Note that fitting the full model with $n = 7,464$ observations takes approximately 7 seconds. In STEADy, the information for every model visited is recorded and computation time is saved each time a model is revisited by using the previously recorded information.

Figure 4 shows the deviance differences $D(X|\boldsymbol{Z}, \boldsymbol{U_{m,X}}) - D(X|\boldsymbol{Z}, \boldsymbol{U_{m+1,X}})$ (top panel) and the point estimates and 95% confidence intervals of β in $[Y|X, \boldsymbol{Z}, \boldsymbol{U_{m,X}}]$ (bottom panel) for $m = 0, \ldots, 234$. The horizontal red lines are placed at the BIC and AIC penalties $\log(7464) = 8.92$ and at 2. The deviance differences between neighboring orbits become negligible starting from orbit 130, which closely agrees with the stabilization of the exposure effect point estimate around orbit 132. The STEADy software also provides the full list of confounders in the dominant model corresponding to each orbit. For example, orbit 132 includes O_3 and DOW, but does not include the indicators for age categories. It also includes some, but not all bases of the natural cubic splines used in the model. Interestingly, O_3 had a p–value of 0.62 in the full outcome model and would be discarded by any model selection or averaging procedure focusing only on the outcome models.

At the second stage of STEADy, we first include in the outcome model the 132 covariates identified above and then we use STEADy to explore the model space generated by the remaining $234 - 132 = 102$ potential confounders. We apply STEADy to the outcome models using 500 iterations per orbit for a total of $500 \times 102 = 51000$ iterations with an average computation time of 4.5 seconds/iteration.

Figure 5 (top panel) considers the outcome model dominant class and shows the deviance differences for neighboring orbits. The horizontal red lines are placed at at the BIC and AIC penalties $\log(7464) = 8.92$ and 2. Figure 5 (bottom panel) shows the point estimates and 95% confidence intervals of β. The top plot shows that among the first 15 models, approximately, the deviance differences are large. This suggests that, in addition to the 132 variables forced in the model, there are 15 to

20 more variables that may be predictive of mortality. However, the bottom panel shows remarkable stability of point estimates across all models in the dominant model class suggesting that none of the 102 additional variables are necessary to obtain an unbiased parameter estimate. This indicates that selecting good predictors of the PM_{10} provides almost complete confounding adjustment.

Another interesting feature of these results is that the standard errors of the exposure effect estimates increase only 2% from orbit 0 to orbit 102. These models show that 10 $\mu g/m^3$ increases in PM_{10} at lag 1 is associated with a 0.64 percent increase in all cause mortality with 95% confidence interval equal to $(0.30, 0.98)$. The confidence intervals are based on the dominant model for orbit 40 in the outcome dominant model class. Under the NMMAPS basic model and the full model, the point estimates and 95% confidence interval of the PM_{10t-1} coefficient multiplied by 1000 (β) are equal to 0.66 (95% CI $0.30, 1.02$)) and 0.64 (95% CI $0.28, 1.00$), respectively. These estimates denote the percent increase in all cause mortality for $10\mu/m^3$ increase in previous day PM_{10t-1}.

5.2 $\mathrm{PM}_{2.5}$ and hospital admissions for COPD

In this section we carry out a data analysis where we apply and compare STEADy and BIC-based BMA to time series data on $PM_{2.5}$ levels and hospital admissions rates for Chronic Obstructive Pulmonary Disease (COPD) among Medicare enrollees. Data are for the period 1999-2002 and for five US counties with more than 200,000 people that are older than 65 and are part of the National Medicare Study.

We consider the following model:

$$\begin{cases} \log(\mu_t) = PM_{2.5t} + \mathrm{DOW} + \mathrm{Age2} + ns(\mathrm{Temp}, \mathrm{df}_{\mathrm{Temp}}) + ns(\mathrm{Temp}_{1-3}, \mathrm{df}_{\mathrm{Temp}}) + \\ \quad + ns(\mathrm{Dew}, \mathrm{df}_{\mathrm{Dew}}) + ns(\mathrm{Dew}_{1-3}, \mathrm{df}_{\mathrm{Dew}}) + ns(t, \mathrm{df}_t) + ns(t, 4) \times \mathrm{Age2} + \\ \quad + \mathrm{offset}\{\log(N_t)\} \end{cases} \tag{7}$$

where Age2 is an indicator for people older than 75, N_t is the number of people at risk on day t. In our ongoing Medicare study we are using the model defined in Equation (7) with $\mathrm{df}_{\mathrm{Temp}} = 6$, $\mathrm{df}_{\mathrm{Dew}} = 3$ and $\mathrm{df}_t = 8$ per year for a total of 32. We refer to this

as the Medicare basic model.

We define the full model as in Equation (7) with number of degrees of freedom in the smooth functions of temperature, dew point temperature and time equal to twice the number of degrees of freedom specified in the Medicare basic model (e.g. $\mathrm{df}_{\mathrm{Temp}} = 12$, $\mathrm{df}_{\mathrm{Dew}} = 6$ and $\mathrm{df}_t = 16$ per year). In the full model we include O_{3t-1} and year indicators as additional covariates with respect to the basic Medicare model.

We then estimate the short-term effects of $PM_{2.5}$ on hospital admission rates using the following 5 approaches: 1) the Medicare basic model (Basic); 2) the Medicare basic model with O_{3t-1} included in the model (Basic + O_3); 3) STEADy; 4) BIC-based BMA with $PM_{2.5}$ not forced in the model (BMA & $PM_{2.5}$ not-f); 5) BIC-based BMA with $PM_{2.5}$ forced in the model (BMA & $PM_{2.5}$f). In each case, the predictors DOW, Age2 and the year indicators were forced in all models while the rest of the variables where included or not according to the specific algorithm. Table 2 summarizes estimates of the $\mathrm{PM}_{2.5}$ effects and their standard errors.

The two baseline models (Basic and Basic + O_3) yield positive and statistically significant estimates of the $PM_{2.5}$ effect in Fresno and Sacramento, and positive, though not significant, estimates in the other three counties.

STEADy reaches an early effect stabilization in all cases: figures are not included, but are similar to those for the Detroit data set. STEADy estimates are roughly consistent with those of the full model. Compared to the baseline models, STEADy estimate are generally consistent in sign but smaller in magnitude. In data sets of such size, the inclusion of additional variables required to reach effect stabilization does not generally result in a large increase in the standard error of the estimates, though a slight increase is observed in Fresno and Sacramento.

In contrast, differences between BMA and the baseline models are less predictable. In Fresno both BMA approaches produce statistically significant effects that are slightly larger than the ones produced by the basic models. For Sacramento the two BMA approaches disagree though both give smaller coefficients than the baseline models and STEADy. This variability is the result of the fact that BIC-based BMA

penalizes the inclusion of confounders that are correlated with the exposure which leads to averaging over effects that have different biological meanings. In all counties except Fresno, BMA with $PM_{2.5}$ not forced in the model estimates the exposure effect to be zero. This is because when the exposure effect estimate is close or below the threshold of statistical significance, then BIC assigns a small weight to $PM_{2.5}$.

Table 2: Estimates of the percent increase in COPD admissions (multiplied by 10) associated with a $10\mu g/m^3$ increase in $PM_{2.5}$ for 5 US counties under five methods. 1) the Medicare basic model (Basic); 2) the Medicare basic model $+O_{3t-1}$ (Basic O_3) ; 3) STEADy; 4) BIC-based BMA with $PM_{2.5}$ forced in the model; 4) BIC-based BMA with $PM_{2.5}$ not forced in the model. Standard deviations are reported as index of the point estimate and 95% confidence intervals are reported below the point estimate and standard deviation.

	Basic	Basic+O_3	STEADy	BMA & $PM_{2.5}$ not-f	BMA & $PM_{2.5}$ f
Fresno	$7.28_{1.96}$ $(3.44, 11.12)$	$7.21_{2.02}$ $(3.25, 11.16)$	$5.39_{2.16}$ $(1.16, 9.62)$	$8.04_{2.37}$ $(3.39, 12.69)$	$8.14_{2.14}$ $(3.94, 12.34)$
Sacramento	$7.07_{2.74}$ $(1.71, 12.44)$	$7.35_{2.82}$ $(1.81, 12.88)$	$4.32_{3.05}$ $(-1.67, 10.30)$	$0.03_{0.40}$ $(-0.76, 0.82)$	$3.68_{2.27}$ $(-0.76, 8.12)$
Miami	$4.01_{2.73}$ $(-1.34, 9.36)$	$3.70_{2.78}$ $(-1.75, 9.14)$	$3.20_{2.78}$ $(-2.26, 8.65)$	$0.005_{0.18}$ $(-0.35, 0.36)$	$1.18_{2.52}$ $(-3.76, 6.12)$
Los Angeles	$0.59_{0.82}$ $(-1.01, 2.20)$	$0.67_{0.83}$ $(-0.96, 2.30)$	$0.29_{0.85}$ $(-1.38, 1.95)$	$0.002_{0.057}$ $(-0.11, 0.12)$	$0.62_{0.72}$ $(-.78, 2.02)$
Cook	$-1.99_{1.36}$ $(-4.65, 0.67)$	$-2.03_{1.36}$ $(-4.69, 0.64)$	$-1.28_{1.37}$ $(-3.97, 1.41)$	$0.001_{0.059}$ $(-0.11, 0.12)$	$0.67_{1.06}$ $(-1.41, 2.76)$

6 Discussion

Motivated by time series studies of air pollution and health, we introduced a new conceptual framework and practical approach (STEADy) for estimating an exposure effect while accounting for uncertainty in the selection of confounders used for adjustment. We consider the case in which the important confounders are available in the candidate set for analysis, so that the true model belongs to the model space. Then in the set of all models that includes the true model's variables the exposure parameter represents the same, correctly adjusted, exposure effect. Our approach to adjustment

uncertainty aims to identify such a set and to borrow strength across models within it for exposure effect estimation. We thus handle adjustment uncertainty by finding a set of models that agree on the effect estimate in both magnitude and interpretation. Because of the size and complexity of model spaces, we build this consensus by seeking stabilization of the estimated effect as the model dimensionality increases.

Accounting for uncertainty in the selection of the variables that are needed to properly adjust for confounders in an observational study is a fundamentally different problem from accounting for uncertainty in the selection of variables for predicting a response. We developed this idea both theoretically and via simulations, and re-examined the adequacy of using BMA for accounting for model uncertainty in risk estimation. We found that failing to recognize this difference leads to the several limitations of the standard BMA approach in this context, and also that it is possible to devise alternatives (such as STEADy) that provide unbiased estimates while accounting for uncertainty. These findings are supported by theoretical results and simulations and illustrated in real data analyses. An open question, and work in progress, is how to generalize the BMA approach to handle the pitfalls described within a Bayesian framework.

In this section we review strengths and limitations of the STEADy algorithm, and then discuss the implications of our findings on the comparison between uncertainty adjustment in effect estimation and prediction.

STEADy is a practical approach to explore both the exposure's and outcome's model spaces, find important covariates that need to be included in the regression model, and identify a set of models that provide a consensus estimate of the effect. STEADy is intuitive and directly designed to answer the question "What is the effect of X on Y?", by dividing it into $M+1$ simpler questions of the type "What is the effect of X on Y when we control for the $m \in \{0, \ldots, M\}$ covariates, $\boldsymbol{U}$, that are most predictive of X?". By identifying important predictors of the exposure and forcing them in the outcome models instead of exploring the outcome space directly, STEADy leads to a fast stabilization of the estimate, and therefore to an efficient

estimation of the exposure effect.

STEADy introduces a transparent and efficient algorithm for sequentially exploring the exposure and the outcome spaces. STEADy partitions the model spaces into orbits that correspond to models with the same number of covariates. This improves efficiency in the search of the exposure's and outcome's model spaces. In every orbit the algorithm explores a relevant and manageable subset of models, currently 500 to 1000 models per orbit, and provides estimates of the likelihood and of the exposure effect under each model. Importantly, this method frees the search from the need to specify arbitrary penalties to compare models of different dimensionality, or priors on the number of necessary confounders. An important component of STEADy is the visual presentation of exposure effect estimators corresponding to the maximum likelihood models within each orbit. The plot of point estimators and confidence intervals for exposure effects combined with the plots of deviance differences between orbits provides a sensible sensitivity analysis for adjustment uncertainty. In addition, the R software that we have developed to implement STEADy can be used for all distribution families of the R function `glm` including, but not limited to, normal, binomial and poisson.

Scientific knowledge can be incorporated in the STEADy algorithm easily. For example, known confounders can be forced in all models or be re-validated during the exploration of the exposure model space. In many applications, confounders are partitioned into scientifically relevant groups, such as biological, environmental, or socio–economic factors. STEADy can be used in this context by focusing first on the biological risk factors, identifying and forcing the important confounders in all models and iterate the procedure with the other groups of risk factors.

STEADy relies upon the specification of a candidate set of confounders and on the assumption that all necessary confounders are included in the set. In some circumstances, such specification might be challenging. However, the choice of a set of potential predictors is a problem that STEADy shares with many methods that account for model uncertainty, including BMA. Scientific knowledge and data availability usually

guide this choice while external studies and additional data may validate this choice and suggest alternative sets. Alternative Bayesian approaces to model uncertainty that do not require the true model to be in the model space have been proposed using the so-called $\mathcal{M}$-open approach [3]. This may provide a fruitful direction for uncertainty adjustment in effect estimation as well.

A potential limitation of the STEADY algorithm is that stabilization may be reached at a model that may not be convincing from a scientific point of view and the corresponding effect exposure estimate might not have an easy interpretation. However in this circumstance, because the STEADy estimate is a consensus over a set of models, we can report our results under any scientifically meaningful model in the set.

BMA approaches are designed to account for model uncertainty in model selection and prediction and are also sometime used for adjustment uncertainty. From a theoretical standpoint and by carrying out simulation studies, we compared our approach with the application of BMA. In our simulations, we found that STEADy outperforms standard implementations of BMA for exposure effect estimation while involving similar computational demands. We also applied STEADy and BMA to daily time series studies of air pollution and mortality and morbidity outcomes. We found that STEADy and BMA can produce results with markedly different policy implications.

Our results highlight serious limitations of BMA for exposure effect estimation, especially in large data sets. Specifically, BMA for parameter estimation can assign a considerable probability mass to models that include only a subset of the necessary confounders, thus biasing the exposure effect estimation. In addition, BMA for parameter estimation can either over or underestimate the statistical variance of the exposure effect estimate. Overestimation of the uncertainty can occur when very different exposure effect estimates are averaged across models that perform different confounder adjustments. Underestimation of the uncertainty can occur when a considerable mass of posterior probability is assigned to models that do not include

important confounders and thus produce biased estimates of the exposure effect with small standard errors.

Finally, the implementation of BMA depends on the calculation of posterior model probabilities which in turn depend on the prior distribution. This dependence is marked in large model spaces with correlated variables. We found that diffuse priors favor the most parsimonious model and therefore informative priors need to be specified. However, in time series studies of air pollution and health, because of the large number of correlated covariates that needs to be included into the model, specifying informative and biologically meaningful prior model probabilities is very challenging.

Currently, accounting for model and adjustment uncertainty when estimating health risk associated with exposure to air pollution is at the center of a heated policy debate [15, 18, 25]. Recently, BMA has been applied in time series analysis of air pollution and health to account for model uncertainty. Koop and Tole (2004), [15] applied BMA to daily time series data for the period 1992–1997 using daily number of deaths as outcome and daily levels of particulate matter, ozone, and other gaseous pollutants as exposures for Toronto, Canada. They used a linear model and prior model probabilities calibrated so that the posterior model probabilities are similar to the ones obtained by use of the BIC approximation. They concluded that "standard deviations for air pollution–mortality impacts become very large when model uncertainty is incorporated into the analysis". The authors argue that BMA should be used instead of expert-based model selection.

Results of our paper indicate several reasons for strong skepticism about the Koop and Tole (2004) analyses. First, our simulation demonstrate that their approach produces biased estimates of the exposure effect. One source of bias is that exposure variables were not forced into the outcome models and therefore estimated exposure effects where averaged with zero. A second source of bias is that estimated exposure effects were averaged across models that did not include all the relevant confounders. Moreover, their priors on coefficients are so diffuse that even variables with strong effects are likely to be excluded with high probability (see definitions B4 and B5 in

Appendix B of Koop and Tole (2004)). Specifically, the authors used g–priors, with the degree of diffusion controlled by the constant $C = \max(M^2, n) \geq n$. It is easy to show that when $C = n$ this is essentially equivalent to using a BIC approximation of the posterior model probabilities. However, Koop and Tole (2004) used $M^2 >> n$ which can lead to ignoring even stronger signals in the data. Finally, the authors used linear models to describe the distribution of low count time series data, a choice that is likely to be inappropriate for describing the sampling variability.

6.1 Conclusion

Despite the evident need for uncertainty adjustment in effect estimation across many research areas, little attention has been given to this issue from the standpoint of statistical methodology. In practice, uncertainty adjustment is either ignored or handled with methods, such as BMA, that have been developed to solve problems that are methodologically different, and may fail to recognize critical aspects of the problem. We hope to have demonstrated that this a unique, important, and difficult problem, to have provided practitioners with an efficient solution to begin exploring adjustment uncertainty in effect estimation, and to have stimulated further much needed methodological discussion.

Acknowledgments: Funding for Ciprian Crainiceanu was provided by the Johns Hopkins Faculty Innovation Award. Funding for Francesca Dominici was provided by a grant from the National Institute for Environmental Health Sciences (ES012054-01) and by the by NIEHS Center in Urban Environmental Health (P30 ES 03819).

Appendix A1

Likelihood maximization for models in $\mathcal{M}$ is usually fast and robust for large data sets and the dominant model in the orbit m, α_m, can be obtained using the following stochastic search algorithm. Start with a model, say α_0, in the mth orbit O_m and select at random one covariate in $\boldsymbol{U}$ that is already in the model, say U_i, and one that is not, say U_j. Construct a new model, say α_1, by replacing U_i by U_j. The new model becomes the current model with probability $p_{0\to 1} = \min(1, L_1/L_0)$ where L_0 and L_1 are the maximum likelihoods of the models α_0 and α_1 respectively. Otherwise a new

pair of covariates is simulated and the procedure is iterated. This algorithm is fast because at every step it requires at most one likelihood calculation for each model proposed to be visited. By keeping records of the maximum likelihood for each model previously proposed, the likelihood of a new model is often known and the model does not need to be refit.

A technical detail that proved very useful was to use the information from one orbit to the next. More precisely, if D_m is the maximum likelihood model on orbit m among *visited* models, the starting point for the optimization algorithm in the $m+1$th orbit is obtained by adding one covariate to D_m. This uses accumulated information from all previous orbit explorations and ensures that the likelihood function is increasing from one orbit to the next.

Appendix A2

Proof of Theorem 1.

i. $|\widehat{\beta}_n^{STEADy} - \beta^*| \leq |\widehat{\beta}_n^{STEADy} - \widehat{\beta}_n^{\alpha_F}| + |\widehat{\beta}_n^{\alpha_F} - \beta^*| < \epsilon/h_n + |\widehat{\beta}_n^{\alpha_F} - \beta^*|$. This shows that $\widehat{\beta}_n^{STEADy}$ is consistent.

ii. By construction $\widehat{v}_n(\widehat{\alpha}_n^{\text{STEADy}}) \leq \widehat{v}_n(\alpha_F)$ which implies $\overline{\lim}_{n\to\infty}\{n\widehat{v}_n(\widehat{\alpha}_n^{\text{STEADy}})\} \leq \overline{\lim}_{n\to\infty}\{n\widehat{v}_n(\widehat{\alpha}_F)\} = v(\alpha_F)$. Under standard regularity assumptions $h_n(\widehat{\beta}_n^{\alpha_0} - \beta^*) \to 0$ almost surely. Let A^h be the convergence set and denote by A^{α_0} and A^{α_F} the convergence sets for $n\widehat{v}_n(\alpha_0)$ and $n\widehat{v}_n(\alpha_F)$ respectively.

For every $\omega \in A^h$ $\exists N_{\omega,\epsilon}$ so that for every $n \geq N_{\omega,\epsilon}$ we have $h_n|\widehat{\beta}_n^{\alpha_0} - \beta^*| < \epsilon$. Thus for every $n \geq N_{\omega,\epsilon}$ α_0 is one of the models in the set over which the minimum is taken in equation (3). Thus $n\widehat{v}_n(\widehat{\alpha}_n^{STEADy}) \leq n\widehat{v}_n(\alpha_0)$ and for every $\omega \in A^h \cap A^{\alpha_0} \cap A^{\alpha_F}$

$$\overline{\lim}_{n\to\infty}\{n\widehat{v}_n(\widehat{\alpha}_n^{\text{STEADy}})\} \leq v(\alpha_0) < v(\alpha_F)$$

iii. From definition $\widehat{\beta}_{W_n} \to \sum_\alpha w(\alpha)\beta^\alpha = \sum_{\alpha\in I} w(\alpha)\beta^\alpha + \sum_{\alpha\in I^C} w(\alpha)\beta^\alpha$ almost surely under model α^*. Writing $\beta^* = \sum_{\alpha\in I} w(\alpha)\beta^* + \sum_{\alpha\in I^C} w(\alpha)\beta^*$ the asymptotic bias of $\widehat{\beta}_{W_n}$ is $\sum_{\alpha\in I} w(\alpha)\beta^\alpha + \{\sum_{\alpha\in I} w(\alpha)\}\beta^*$.

Proof of Theorem 2. It is sufficient to show that for any pair of nested models the BIC based weighting scheme will favor asymptotically the smallest of the two models. Without loss of generality, assume that $y_1, \ldots, y_n$ are i.i.d. with probability density function $\rho(y|\theta)$, with $\theta \in \Theta \subset \mathbb{R}$. Let $\theta_0 \in \text{int}(\Theta)$ and assume that the regularity conditions for the existence, consistency and asymptotic normality of the MLE are satisfied. Consider the following testing framework $H_0 : \theta = \theta_0$ versus $H_A : \theta \neq \theta_0$. Denote by α_0, α_1 the models corresponding to H_0 and H_A respectively and by $\widehat{\theta}_n$ the MLE of θ under α_1. Define the likelihood ratio statistic $\Lambda_n^0(\boldsymbol{Y}) = \rho(\boldsymbol{Y}|\theta_0)/\rho(\boldsymbol{Y}|\widehat{\theta}_n)$. It is known that

if $\theta = \theta_0$ then $-2\log\Lambda_n(\boldsymbol{Y}) \Rightarrow \chi^2_1$, where "$\Rightarrow$" denotes weak convergence. Since $\mathrm{BIC}(\alpha_0) \leq \mathrm{BIC}(\alpha_1)$ is equivalent to $-2\log\Lambda_n(\boldsymbol{Y}) \leq \log(n)$. It follows that $\lim_{n\to\infty}[P\{\mathrm{BIC}(\alpha_0) \leq \mathrm{BIC}(\alpha_1)\}] = 1$ under the null α_0.

Assume now that α_1 is the true model and θ_1 is the true value of the parameter. Also assume that $\theta_0 = \theta_1 + c/\sqrt{n}$. Note that $\{\mathrm{BIC}(\alpha_0) \geq \mathrm{BIC}(\alpha_1)\}$ is equivalent to

$$\left[-2\log\Lambda^1_n(\boldsymbol{Y}) + 2n\left\{\frac{1}{n}\sum_{i=1}^{n}\log\rho(y_i|\theta_1) - \frac{1}{n}\sum_{i=1}^{n}\log\rho(y_i|\theta_1 + c/\sqrt{n})\right\} \leq \log(n)\right]$$

It is sufficient to prove that, almost surely,

$$v_n = \frac{1}{n}\sum_{i=1}^{n}\log\rho(y_i|\theta_1) - \frac{1}{n}\sum_{i=1}^{n}\log\rho(y_i|\theta_1 + c/\sqrt{n})$$

converges in distribution to a random variable. Using a second order Taylor expansion around the MLE $\widehat{\theta}_n$ one obtains

$$\log\rho(\boldsymbol{Y}|\theta) = \log\rho(\boldsymbol{Y}|\widehat{\theta}_n) - \frac{\left(\theta - \widehat{\theta}_n\right)^2}{2} \times \frac{\partial^2}{\partial\theta^2}\log\rho(\boldsymbol{Y}|\widehat{\theta}_n) + R(\theta, \widehat{\theta}_n)\ ,$$

where there exists a constant M such that $|R(\theta, \widehat{\theta}_n)| \leq M(\theta - \widehat{\theta}_n)^3$. It follows immediately that

$$v_n = \left\{2c\sqrt{n}(\theta_1 - \widehat{\theta}_n) - c^2\right\}\frac{\partial^2}{\partial\theta^2}\rho(\boldsymbol{Y}|\widehat{\theta}_n) + 2nR(\theta_1, \widehat{\theta}_n) + 2nR(\theta_1 + c/\sqrt{n}, \widehat{\theta}_n)\ .$$

It is easy to show that, in probability, $2nR(\theta_1, \widehat{\theta}_n) + 2nR(\theta_1 + c/\sqrt{n}, \widehat{\theta}_n) \to 0$ and

$$\left\{2c\sqrt{n}(\theta_1 - \widehat{\theta}_n) - c^2\right\}\frac{\partial^2}{\partial\theta^2}\rho(\boldsymbol{Y}|\widehat{\theta}_n) \Rightarrow [-2cX + c^2]I_{\theta_1}\ ,$$

where I_θ is the Fisher information at θ and X denotes a random variable with $N(0, I^{-1}_{\theta_1})$ distribution, which shows that $\lim_{n\to\infty} P[\{\mathrm{BIC}(\alpha_0) \geq \mathrm{BIC}(\alpha_1)\}] = 0$ and ends the proof.

References

[1] E. Aga, E. Samoli, G. Touloumi, H.R. Anderson, E. Cadum, B. Forsberg, P. Goodman, A. Goren, F. Kotesovec, B. Kriz, M. Macarol-Hiti, S. Medina, A. Paldy, C. Schindler, J. Sunyer, P. Tittanen, B. Wojtyniak, D. Zmirou, J. Schwartz, and K. Katsouyanni. Short-term effects of ambient particles on mortality in the elderly: results from 28 cities in the APHEA2 project. *European Respiratory Journal Supplement*, 40:28–33, 2003.

[2] M. Bell, A. McDermott, S.L. Zeger, J.M. Samet, and F. Dominici. Ozone and mortality in 95 u.s. cities from 1987 to 2000. *Journal of American Medical Association*, 292:2372–2378, 2004.

[3] J.M. Bernardo and A.F.M. Smith. *Bayesian Theory*. John Wiley & Sons, Chicester, UK, 2000.

[4] R. Burnett, R. Ma, M. Jerrett, M. Goldberg, S. Cakmak, A. Pope, and D. Krewski. The spatial association between community air pollution and mortality: a new method of analyzing correlated geographic cohort data. *Environmental Health Perspectives*, 109:375–380, 2001.

[5] M. Clyde. Model uncertainty and health effects studies for particulate matter. *Environmetrics*, 11:745–763, 2000.

[6] F. Dominici, A. McDermott, M. Daniels, S. L. Zeger, and J. M. Samet. *A Special Report to the Health Effects Institute on the Revised Analyses of the NMMAPS II Data*. The Health Effects Institute, Cambridge, MA., 2003.

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

[8] F. Dominici, J. M. Samet, and S. L. Zeger. Combining evidence on air pollution and daily mortality from the twenty largest US cities: A hierarchical modeling strategy (with discussion). *Royal Statistical Society, Series A*, 163:263–302, 2000.

[9] D. Draper. Assessment and propagation of model uncertainty. *Journal of Royal Statistical Society – Series B*, 57:45–97, 1995.

[10] J. Hoeting, D. Madigan, A. Raftery, and C. Volinsky. Bayesian model averaging: a tutorial. *Statistical Science*, 14:382–417, 1999.

[11] J. Kaiser. Mounting evidence indicts fine-particles pollution. *Science*, 307:1858–1861, 2005.

[12] K. Katsouyanni, G. Toulomi, E. Samoli, A. Gryparis, A. LeTertre, Y. Monopolis, G. Rossi, D. Zmirou, F. Ballester, A. Boumghar, and H. R. Anderson. Confounding and effect modification in the short-term effects of ambient particles on total mortality: Results from 29 european cities within the APHEA2 project. *Epidemiology*, 12:521–513, 2001.

[13] K. Katsouyanni, G. Toulomi, C. Spix, F. Balducci, S. Medina, G. Rossi, B. Wojtyniak, J. Sunyer, L. Bacharova, J.P. Schouten, A. Ponka, and H. R. Anderson. Short term effects of ambient sulphur dioxide and particulate matter on mortality in 12 european cities: results from time series data from the APHEA project. *British Medical Journal*, 314:1658–1663, 1997.

[14] J.E. Kelsall, J. M. Samet, and S. L. Zeger. Air pollution, and mortality in Philadelphia, 1974-1988. *American Journal of Epidemiology*, 146:750–762, 1997.

[15] G. Koop and L. Tole. Measuring the health effects of air pollution: to what extent can we really say that people are dying of bad air. *Journal of Environmental Economics and Management*, 47:30–54, 2004.

[16] D. Madigan and A.E. Raftery. Model selection and accounting for model uncertainty in graphical models using occam's window. *Journal of the American Statistical Association*, 89:1535–46, 1994.

[17] D. Madigan, A.E. Raftery, C.T. Volinsky J.A., and Hoeting. Bayesian model averaging. *Integrating Multiple Learned Models, (IMLM-96), P. Chan, S. Stolfo, and D. Wolpert (Eds.)*, pages 77–83, 1996.

[18] S.H. Moolgavkar. A review and critique of the epa's rationale for a fine particle standard. *Regulatory Toxicology and Pharmacology*, 42:123–144, 2005.

[19] National Research Council. Research priorities for airborne particulate matter. part i. immediate priorities and a long-range research portfolio. *Washington, DC: National Academy Press.*, 1998.

[20] National Research Council. Research priorities for airborne particulate matter. part ii. evaluating research progress and updating the portfolio. *National Academy Press. Washington, DC*, 1999.

[21] National Research Council. Research priorities for airborne particulate matter. part iii. early research progress. *National Academy Press. Washington, DC*, 2001.

[22] R. Peng, F. Dominici, and T. Louis. Model choice in multi-site time series studies of air pollution and mortality. *Journal of the Royal Statistical Society – Series C*, to appear, 2005.

[23] J. M. Samet, F. Dominici, F.C. Curriero, I. Coursac, and S. L. Zeger. Fine particulate air pollution and mortality in 20 U.S. cities: 1987-1994 (with discussion). *New England Journal of Medicine*, 343(24):1742–1757, 2000.

[24] J. M. Samet, F. Dominici, S. L. Zeger, J. Schwartz, and D. W. Dockery. *The National Morbidity, Mortality, and Air Pollution Study: Methods and Methodological Issues.* 94 Pt I. Health Effects Institute, Boston MA, 2000.

[25] J. M. Samet, S. L. Zeger, F. Dominici, F. Curriero, I. Coursac, D. Dockery, J. Schwartz, and A. Zanobetti. *The National Morbidity, Mortality, and Air Pollution Study (HEI Project No. 96-7): Morbidity and Mortality from Air Pollution in the United States.* Health Effects Institute, Cambridge, MA., 2000.

[26] J. M. Samet, S. L. Zeger, J.E. Kelsall, J. Xu, and L.S. Kalkstein. *Air pollution, weather and mortality in Philadelphia, In Particulate Air Pollution and Daily Mortality: Analyses of the Effects of Weather and Multiple Air Pollutants. The Phase IB report of the Particle Epidemiology Evaluation Project.* Health Effects Institute, Cambridge, MA., 1997.

[27] P. Speckman. Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society – Series B*, 50:413–436, 1988.

[28] G. Toulomi, K. Katsouyanni, D. Zmirou, and J. Schwartz. Short-term effects of ambient oxidant exposure on mortality: A combined analysis within the APHEA project. *American Journal of Epidemiology*, 146:177–183, 1997.

[29] L Welty and S.L. Zeger. Are the acute effects of particulate matter on mortality in the national morbidity, mortality, and air pollution study the result of inadequate control for weather and season? a sensitivity analysis using flexible distributed lag models. *American Journal of Epidemiology*, 162:80–88, 2005.

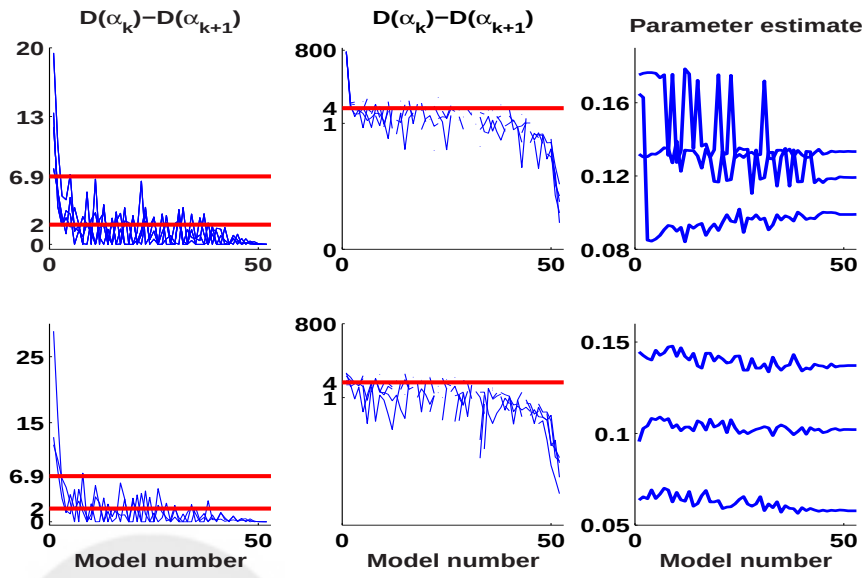

Figure 1: Results for 3 data sets simulated from model (4) with $\rho = 0.7$ (first row) and $\rho = 0$ (second row) with random X and U's. The first column shows the deviance differences between the outcome dominant models on orbit k and on orbit $k+1$, the second column shows the deviance differences (in log scale) between the exposure dominant model on orbit k and on orbit $k+1$, and the third column shows the estimates of β_1 for each of the 3 simulated data sets all plotted against model's number $1, \ldots, 53$. The horizontal red lines in the first two columns are placed at the BIC and AIC penalties $\log(1000) = 6.91$ and 2, respectively. The horizontal red line in the third column is placed at the true value of the parameter, $\beta = 0.1$.

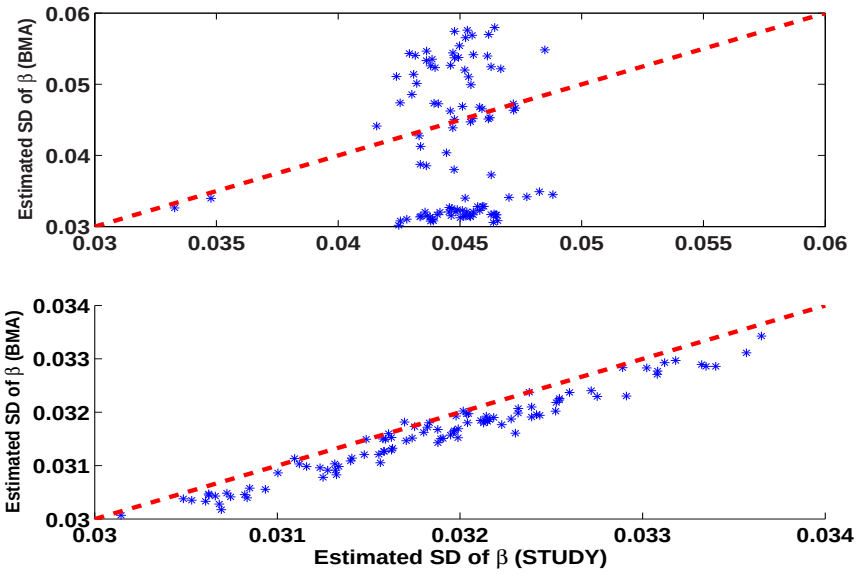

Figure 2: Estimated standard errors of $\widehat{\beta}_1$ for 100 simulated data sets using STEADy (X–axis) and BMA (Y-axis). Data was simulated from model (4) with $\rho = 0.7$ (top) and $\rho = 0$ (bottom) using random X and U's.

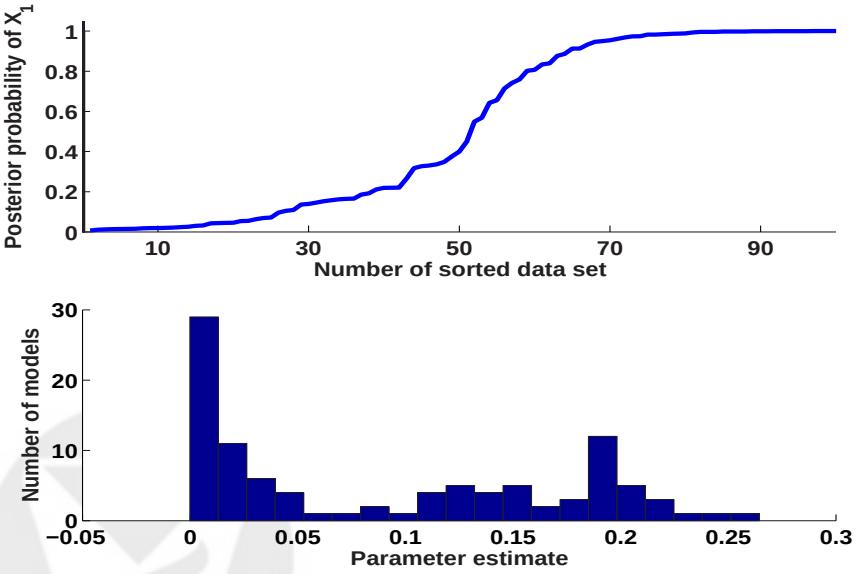

Figure 3: Top: sorted posterior probability that X_1 is included into the model for each of the 100 data sets. Posterior probabilities are calculated by use of the BIC approximation. Bottom: histogram of the averages of the β_1s estimates across the 100 datasets $\mathrm{M}(\widehat{\beta}_1)$ under the different models. Data were simulated for $\rho = 0.7$ with random X and U's and with X_1 was not forced into the model. True value of the parameter is $\beta = 0.1$.

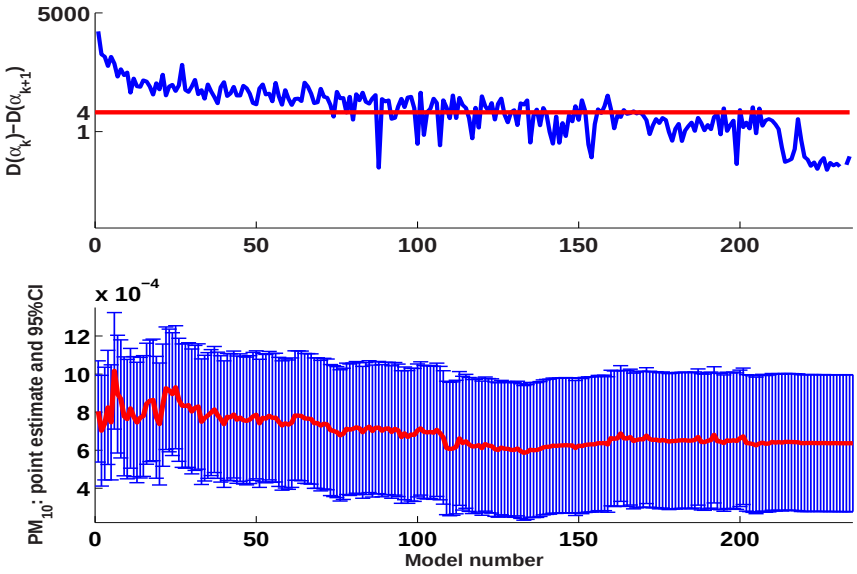

Figure 4: Results for Detroit 1987–2000. Top panel: deviance differences between the exposure dominant models on orbit k and on orbit $k+1$ plotted against model numbers with complexity increasing from left to right. The horizontal red lines are placed at the BIC and AIC penalties $\log(7464) = 8.92$ and 2, respectively. Bottom graph: Estimated percent increase in all cause mortality associated with a $10\mu g/m^3$ of PM_{10t-1} with 95% confidence intervals corresponding to each exposure dominant model.

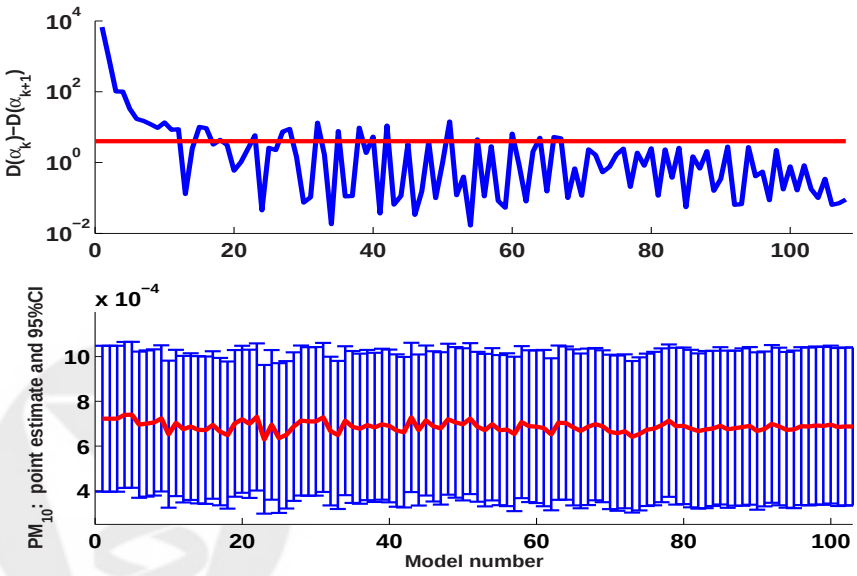

Figure 5: Detroit 1987–2000. Top: deviance differences between the outcome dominant models on orbit k and orbit $k+1$ vs. model numbers with complexity increasing from left to right. Horizontal red lines are placed at the BIC and AIC penalties $\log(7464) = 8.92$ and 2, respectively. Bottom: Estimated percent increase in all cause mortality associated with a $10\mu g/m^3$ of PM_{10t-1} with 95% confidence intervals.