An Report of A Measurement Error Model for Time-series Studies of Air Pollution and Mortality

Mi Zhou

September 22, 2011
1 Background

The adverse health effect of air pollution on our respiratory and cardiovascular condition has received lots of attentions. Many studies have been done on the association between measurement of ambient concentrations of particulate matter and non-accidental daily mortality counts Dominici et al. (2000a), and it shows that daily rates of mortality increases with levels of particulate air pollution. However, what we actually concern is the increase in risk of mortality per unit increase in personal exposure to particulates which is rarely obtained and is instead substituted by measured ambient concentrations from a few monitors where aggregate rates of morbidity and mortality are accessed. When exposures are measured with error, the power of analysis is reduced Carroll et al. (2006). The paper Dominici et al. (2000b) is to develop a measurement error model for evaluating the effects of exposure measurement error on estimates of effects of particulate air pollution on mortality in time-series studies.

2 Data and methods

The study is mainly based on the data of Baltimore, MD. The Baltimore data includes daily mortality, temperature, dew-point and particulate pollution concentrations for the period 1987-1994. Besides, data from five other studies of personal exposure and ambient concentrations of \( PM_{10} \) are available. Let \( Y_t \) denote the number of deaths in Baltimore, \( X_t \) be the average personal exposure to particulate pollution, and \( Z_t \) be the measured ambient concentration of particulates. The authors propose a log-linear and poisson error model as follows,

\[
\log \mu_t = W_zt \theta_z + B_t \gamma_z, \tag{2.1}
\]
where \( \mu_t \) denotes \( E(Y_t) \), \( W_{zt} \) includes \( Z_t \) and some indicator variables for specific age and day, \( B_t \) represents the \( t_{th} \) row of the design matrix for the cubic splines which are added for the nonparametric function of some other confounders, in addition, \( \theta_z \) and \( \gamma_z \) are the corresponding vectors of coefficients. Also, this paper denotes the observed and missing predictors for Baltimore by \( Z_t^B \) and \( X_t^B \). The five additional data sources are used to estimate \( X_t^B \) and \( \beta_z \). Let \( S = \{1, 2, 3, 4, 5, B\} \) be the label-set of data sources. Thus the available data are: the mortality counts in Baltimore \( Y_t^B \); the ambient \( PM_{10} \) measures for all the locations \( Z^s_t, s \in S \); and the average personal \( PM_{10} \) exposures for all the locations \( X^s_t \) except for Baltimore.

The authors model the data in two parts. The model for \( Y_t^B \) given \( X_t^B \) is

\[
Y_t^B | \mu_t \sim \text{Poisson}(\mu_t), \quad t = 1, \ldots, T, \tag{2.2}
\]

\[
\log \mu_t = W_{zt} \theta_z + B_t \gamma_z. \tag{2.3}
\]

The hierarchical model for \( X^s_t \) given \( Z^s_t \) is

\[
X^s_t = \alpha^s_0 + \alpha^s_1 Z^s_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2_x), \quad s \in S, \tag{2.4}
\]

\[
\alpha^s_0 | \alpha_0 \sim N(\alpha_0, \tau^2_0), \quad s \in S, \tag{2.5}
\]

\[
\alpha^s_1 | \alpha_1 \sim N(\alpha_1, \tau^2_1), \quad s \in S, \tag{2.6}
\]

where the slope \( \alpha^s_1 \) measures the change in personal exposure per unit change in a measured ambient concentration at location \( s \), and the intercept \( \alpha^s_0 \) represents personal exposure to particles that derive from personal activities or un-measured micro-environments rather than external sources.
3 Analysis of the data

The authors use two approaches to estimate $\beta_x$. The first is a non-Bayesian non-hierarchical method based on a regression calibration approach, and the second is the Bayesian hierarchical model. For the regression calibration approach, the authors fit linear regression models to the five validation data sets, and then estimate the overall intercept and slope $\hat{\alpha}_0 = 53.18$ and $\hat{\alpha}_1 = 0.53$ by using a weighted average approach for a random effect model. Next, they estimate $\hat{X}_t^B$ by $\hat{X}_t^B = \hat{\alpha}_0 + \hat{\alpha}_1 Z_t^B$, and use $\hat{X}_t^B$ instead of $Z_t^B$ in (2.3).

In the hierarchical Bayesian method, the prior of the parameters are specified: for the overall regression parameters $\alpha_0, \alpha_1$, and for the vector of the log relative rates $\theta_x$, they use normal prior; and for the variance parameters $\tau_0^2, \tau_1^2$ and $\sigma_x^2$ they use inverse gamma prior. And a block Gibbs Sampler with Metropolis steps is applied for drawing samples from posterior distributions of parameters. It turns out the estimate of $\beta_x$ is slightly larger under the regression calibration model than under the hierarchical Bayesian model with a substantial over a substantial overlap of the two IQRs. In summary, the Bayesian model is a model conservative approach since it takes into account key sources of uncertainty in $\hat{\beta}_x$ which results from estimating $\hat{\alpha}_0, \hat{\alpha}_1$ and $X_t^B$.

This paper has some strengths in analysis: it extends the regression calibration, giving a more conservative result by Bayesian analysis; it easily provides a way to consider the variability across studies in the personal-ambient relationships; the results are not sensitive to other measurement error model assumptions and specifications of the prior distribution within a reasonable range.

One important limitation of the paper is the modeling of the association between ambient and average personal exposure. For example, a time-series model of $X_t$ on $Z_t$ would be more appropriate than the simple linear models used here. Besides, it may be necessary to get more pollutants involved.
References

