Biometrics 000, 000–000 DOI: 000
000 0000

Web-based Supplementary Materials for "Estimating Acute Air Pollution Health Effects From Cohort Study Data" by A. A. Szpiro, L. Sheppard, S. D. Adar, and J. D. Kaufman

This paper has been submitted for consideration for publication in *Biometrics*

Web Appendix A

We give additional details on the calculations leading to (4.2). Define

$$\mathbf{x} = (x_1, \ldots, x_n)^\top = \mathbb{H}_{m_2} \boldsymbol{\alpha}_{m_2} + \boldsymbol{\eta},$$

$\bar{\mathbf{x}} = \bar{\mathbb{H}}_{m_2} \boldsymbol{\alpha}_{m_2} + \bar{\boldsymbol{\eta}}$, $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)^\top$, and $\bar{\mathbf{z}} = (\mathbf{w}(1), \ldots, \mathbf{w}(T))^\top$. We can now adapt an argument from Dominici et al. (2004).

As long as the design matrix is non-singular and regularity conditions similar to those given by White (1980) hold, it is easy to verify that fitting equation (4.1) by OLS with $m = m_2$ gives

$$\begin{aligned}
E\left(\begin{array}{c|c} \hat{\beta}_x - \beta_x & \\ \hat{\boldsymbol{\beta}}_z - \boldsymbol{\beta}_z & t_1, \ldots, t_n \\ \hat{\boldsymbol{\gamma}}_{m_2} - \boldsymbol{\gamma}_{m_2} & \end{array} \right) &= \left(\begin{pmatrix} \mathbf{x}^\top \\ \mathbf{z}^\top \\ \mathbb{H}_{m_2}^\top \end{pmatrix} \begin{pmatrix} \mathbf{x} & \mathbf{z} & \mathbb{H}_{m_2} \end{pmatrix} \right)^{-1} \begin{pmatrix} \mathbf{x}^\top \\ \mathbf{z}^\top \\ \mathbb{H}_{m_2}^\top \end{pmatrix} \mathbb{H}_{m_1/m_2} \boldsymbol{\gamma}_{m_1/m_2} \\
&\overset{\text{a.s.}}{\to} \left(\begin{pmatrix} \bar{\mathbf{x}}^\top \\ \bar{\mathbf{z}}^\top \\ \bar{\mathbb{H}}_{m_2}^\top \end{pmatrix} \begin{pmatrix} \bar{\mathbf{x}} & \bar{\mathbf{z}} & \bar{\mathbb{H}}_{m_2} \end{pmatrix} \right)^{-1} \begin{pmatrix} \bar{\mathbf{x}}^\top \\ \bar{\mathbf{z}}^\top \\ \bar{\mathbb{H}}_{m_2}^\top \end{pmatrix} \bar{\mathbb{H}}_{m_1/m_2} \boldsymbol{\gamma}_{m_1/m_2},
\end{aligned}$$

where the limit in the second line is for large n.

It follows that

$$\begin{aligned}
E(\hat{\beta}_x - \beta_x) &\overset{\text{a.s.}}{\to} \frac{\bar{\mathbf{x}}^\top \bar{\mathbb{H}}_{m_1/m_2} \boldsymbol{\gamma}_{m_1/m_2}}{\bar{\mathbf{x}}^\top \left(I - \bar{\mathbb{H}}_{m_2} \left(\bar{\mathbb{H}}_{m_2}^\top \bar{\mathbb{H}}_{m_2} \right)^{-1} \bar{\mathbb{H}}_{m_2}^\top \right) \bar{\mathbf{x}}} \\
&= \frac{\bar{\boldsymbol{\eta}}^\top \bar{\mathbb{H}}_{m_1/m_2} \boldsymbol{\gamma}_{m_1/m_2}}{\bar{\boldsymbol{\eta}}^\top \left(I - \bar{\mathbb{H}}_{m_2} \left(\bar{\mathbb{H}}_{m_2}^\top \bar{\mathbb{H}}_{m_2} \right)^{-1} \bar{\mathbb{H}}_{m_2}^\top \right) \bar{\boldsymbol{\eta}}}.
\end{aligned}$$

The first line uses orthogonality of the spline basis and $m_2 \geq m_3$; the denominator in the first line is the top left element of the partitioned inverse from the previous display; and the second line uses $\bar{\mathbf{x}} = \bar{\mathbb{H}}_{m_2} \boldsymbol{\alpha}_{m_2} + \bar{\boldsymbol{\eta}}$.

Web Appendix B

To calculate standard errors for the GLS-based estimates in Section 5.3, it is useful to define B as the upper triangular Cholesky factor of W^{-1}. then the GLS estimate $\hat{\beta}_x$ is also the OLS solution to

$$y_i' = \hat{\eta}'(t_i)\beta_x + \hat{g}'(t_i)\breve{\beta}_x + \mathbf{z}_i'\boldsymbol{\beta}_z + (f'(t_i) + \varepsilon_i'), \tag{0.1}$$

where the y_i' are obtained by left multiplying the outcome vector in equation (5.1) by B^{-1}, and similarly for the other quantities in this equation. We treat B^{-1} as fixed, a reasonable approximation that can be rigorously justified if the random effect variance estimates converge to an asymptotic limit sufficiently quickly (Liang and Zeger, 1986).

Multiplication by B^{-1} induces correlation in equation (0.1); this remains true even if the t_i are chosen independently. Fortunately, since the formal random effects are all multiplied by functions of t_i, the dominant entries in B^{-1} preserve a block diagonal covariance structure enabling us to calculate approximately correct standards for $\hat{\beta}_x$ by using GEE with clusters determined by common values of t_i. Further refinement is possible by allowing banded or similar correlation structures, but at the expense of requiring specialized software and a greater computational burden (Lumley and Heagerty, 1999). Also, the bias-variance tradeoff from estimating a covariance matrix that is closer to being dense implies that it may not lead to improved standard error estimates. The block diagonal approximation works well in our examples, so we do not pursue additional refinements.

References

F. Dominici, A. McDermott, and T. J. Hastie. Improved semiparametric time series models of air pollution and mortality. *Journal of the American Statistical Association*, 99(468): 938–948, 2004.

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

T. Lumley and P. Heagerty. Weighted empirical adaptive variance estimators for correlated data regression. *Journal of the Royal Statistical Society, Series B*, 61(Part 2):459–477, 1999.

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