INCLUDING STRUCTURAL MEASUREMENT ERRORS IN THE NONLINEAR REGRESSION ANALYSIS OF CLUSTERED DATA

96

by

J.V. Zidek, N.D. Le, H. Wong
and R.T. Burnett

Technical Report #170

December 1996

WS/PR:ddp

Including Structural Measurement Errors in the Nonlinear Regression Analysis of Clustered Data*

JV Zidek[1], ND Le[2], H Wong[3] and RT Burnett[4]
[1]University of British Columbia [2]BC Cancer Agency [3] Statistics Canada [4]Health Canada

December 18, 1996

Abstract

This paper extends methods for nonlinear regression analysis that have been developed for the analysis of clustered data. Its novelty lies in its dual incorporation of random cluster effects and structural error in the measurement of the explanatory variables. Moments up to second order are assumed to have been specified for the latter to enable a generalised estimating equations approach to be used for fitting and testing nonlinear models linking response to these explanatory variables and random effects. Taylor expansion methods are used and a difficulty with earlier approaches overcome. Finally we describe an application of this methodology to indicate how it can be used. That application concerns the degree of association of hospital admissions for acute respiratory health problems and air pollution.

1 Introduction and Summary

In this paper we suggest a method of accounting for structural measurement errors in nonlinear regression analysis of clustered data. The novelty of the method lies in our simultaneous incorporation of random measurement errors and random cluster effects. In the process of developing that method we refine a method of Lindstrom

*The research reported in this paper was supported by a contract from Health Canada and a grant from the Natural Science and Engineering Research Council of Canada.

and Bates (1990) as adapted in Burnett and Krewski (1994, hereafter BK-94). We begin by describing how the need for such a method arises. Then we summarise the results of this paper.

Regression methods are commonly used in analysing clustered data. Statistical epidemiology provides examples such as the study of Burnett et al (1994, hereafter BET) who examine series of daily counts of hospital admissions for respiratory morbidity. Each series forms a cluster associated with one of a number of hospitals in southern Ontario. Their study concerns the effect of air pollution on respiratory health.

Data may be clustered to avoid the confounding of factors. For example in the just-cited study the number of daily hospital admissions and the level of air pollution will simultaneously increase with population size. Thus the unclustered data would indicate strong association between hospital admissions and air pollution. To avoid the confounding of the factors of pollution and population size, one might well disaggregate the count data by subregional clusters and analyse the time varying pollution and admission levels cluster-by-cluster to control for population size.

BET take exactly that approach. In their regression analysis of association between admission and pollution levels, they form (Y, X) pairs through natural temporal linkages, Y being daily admissions and X, lagged average daily concentration of air pollutants. Their study seeks to determine if 'blips' of short duration in Y tend to follow those in X. A positive finding gives evidence of a causal linkage between Y and X even with observational data, the short time-intervals of such blips tending to make confounding with blips in other factors seem implausible.

However, the approach encounters an obstacle: the X's needed for their analysis have not been measured in all clusters. This obstacle may be expected in any study involving a dataset formed by joining others compiled for a variety of different purposes. This obstacle forces BET to use as surrogates for their unmeasured X's others obtained from often distant ambient pollution monitors.

We regard X as random (having a distribution conditional on the values of the surrogates). Thus we face an errors-in-variables problem of the *structural* rather than *functional* type. [Even in the latter case, Pierce et al (1992) argue that X should be formally treated as random Their paper is but one from an extensive literature on structural measurement error that we cannot feasibly review in detail. Fuller (1987) does survey that literature in his comprehensive work on the theory of measurement error. More recent work (see Carroll, Ruppert and Stefanski 1995 for a review) has addressed that theory for nonlinear regression.]

Intuition suggests the use of the cluster means of the X-distributions instead of surrogates like those in BET to impute the unmeasured X's. However, while this approach may reduce the bias expected from structural measurement error, it does not get around the problem (see Zidek 1996) introduced in nonlinear regression by lack of precision in an imputation procedure like that decribed above. If ignored, that difficulty would call inferential findings into question.

Zidek (1996) discusses some of the deleterious effects of measurement errors. In a report of particular relevance here, Zidek et al (1994) consider a nonlinear regression analysis with two factors, one causative, the other not, but associated with the first. The variables representing these factors are measured with error. If the relative size of this error for the first compared to the second is large, the first may be found non-significant and the second significant, even when their association is only moderate.

We will not discuss here methods for finding the X-distribution but the value of our method is enhanced by the progress that has been made in finding that distribution when X represents certain kinds of air pollutants with Gaussian distributions (after suitable transformation). (Le and Zidek 1992; Brown, Le and Zidek 1994; Le, Sun and Zidek 1994; Sun 1994; Sun 1995; Sun, Le, Zidek and Burnett 1995). Diggle et al (1995) obtain results for the non-Gaussian case and Diggle et al (1996) demonstrate this methodology in an application.

With or without measurement error, the cluster-based analysis may lack the power needed to find a small positive association between 'blips' within clusters. How can we gain the power needed to find them?

The answer commonly adopted in statistical epidemiology consists of synthesising the cluster-based analyses to gain that power.

If data from different clusters were independent, even naive reasoning shows that a pattern of inidividually non-significant but positive correlations, are significant in the aggregate. More refined reasoning using an intrinsically Bayesian approach assigns random effects to clusters. The data remain conditionally independent given these effects. But their unconditional distribution integrates the data into a single overall analysis. That analyis preserves the cluster-structure and thereby avoids the problem of confounding described earlier. Breslow and Clayton (1993) unify and review various approaches of the sort we have just described.

In summary, the possibility of confounding factors may force an investigator to cluster response data in observational studies and do conditional analyses within clusters. However in some clusters, data on explanatory variables may not be available, leading to structural measurement error. Such error needs to be incorporated into inferential analysis to increase the plausibility of the investigators conclusions. Finally, to gain the power needed to detect subtle associations common to the clusters, random effects can be introduced to enable information to flow between clusters.

While the resulting sub-structure seems inevitable it comes at a high cost: high levels of uncertainty about the models we need to build on it to perform conventional inference. That uncertainty could overwhelm any gains we may have achieved through the creation of the sub-structure. The generalised estimating equations technique offers a way around this difficulty (*cf* Zeger 1688; Liang and Zeger 1986; Zeger and Liang 1986; Zeger and Karim 1991; Zeger, Liang and Albert 1988). The investigator need only specify moments up to second order thereby gaining robustness against model misspecification and reducing the levels of uncertainty infused through modelling. The technique whose justification comes from the large sample paradigm, yields estimates of the parameters reflecting the association under investigation and associated tests.

Our particular generalised estimating equations technique is an adaption and refinement of that of BK-94 and Burnett, Ross, and Krewski (1993). That methodology builds on the work of Lindstrom and Bates (1990) which in turn depends on that of Laird and Ware (1982). In Section 2 we develop approximations to the first and second order moments needed for our generalised estimating equations technique in Section 3. Since our methodology extends that of *BK-94*, we can readily adapt their results for inference about model parameters in Section 3. We assume parameters of the covariate distribution known in this paper and comment on that assumption in Section 5.

The illustrative example in Section 4 derives from a re-analysis of the hospital admissions data studied in BET and *BK-94*. Our analysis differs from its earlier counterparts in a number of ways, notably: (i) pollutant levels are interpolated using the methodology of Le and Zidek (1992) as extended in papers cited above; (ii) uncertainty in the interpolated pollution levels is incorporated into the model through the methodology developed in this paper.

2 Moment Structure for Errors-in-Covariates Models

Let $\mathbf{Y}$ denote the vector of observable responses for all K clusters and T time-points and $\mathbf{Y}_k$, the subvector of observations corresponding to the k-th cluster. We further denote by Y_{kt} the observation for the k-th cluster on time-point t.

Let $\mathbf{X}_{kt}^{(1)}$ denote the covariate-vector corresponding to Y_{kt}. We include in $\mathbf{X}_{kt}^{(1)}$ both the covariates observed with error and those without. We assume first and second moments known: $\mathbf{E}(\mathbf{X}_{kt}) = \mathbf{z}_{kt}$; $\mathbf{Cov}(\mathbf{X}_{kt_1}, \mathbf{X}_{kt_2}) = \mathbf{G}_{kt_1t_2}$. We set to zero, any elements of $\mathbf{G}_{kt_1t_2}$ corresponding to covariates observed without error. In our theory we assume $\mathbf{G}_{kt_1t_2} = \mathbf{0}$ when $t_1 \neq t_2$.

The covariates and responses are related through $E(Y_{kt} \mid \mathbf{a}_k, \mathbf{X}_{kt}) = \zeta_{kt}$ and

$$\mathbf{Cov}(Y_{kt_1}, Y_{kt_2} \mid \mathbf{a}_k, \mathbf{X}_{kt_1}, \mathbf{X}_{kt_2}) \;=\; \phi\zeta_{kt_1}\delta_{t_1t_2},$$

where $\delta_{uv} = 1$ or 0 according as $u = v$ or not. We have let ϕ be an unknown scalar dispersion parameter. Finally, $\zeta_{kt} = \zeta(\mathbf{a}_k^T\mathbf{X}_{kt})$, ζ being positive and thrice continuously differentiable. We also require that ζ be log convex. That latter requirement insures that $0 < \zeta(u) + 2\zeta'(u)v + \zeta''(u)v^2$ for all u, v, a fact needed to insure below the positive definiteness of the approximate unconditional covariance matrix we adopt.

We assume the coefficient vector $\mathbf{a}_k$ [some of whose elements can be non-random] has mean α_k and covariance matrix $\mathbf{D}$. Elements of the latter corresponding to non-random co-ordinates of α_k are zero.

The estimating equations approach requires just the first two moments of the response variable distributions. Nevertheless we are forced to approximate these two moments through Taylor expansions of the mean response functions to make second order moments computable.

To simplify exposition we use in both stochastic and nonstochastic equations, $\approx$ to mean "approximately equal to" after dropping terms of order higher than two. For added simplicity we temporarily fix and drop the subscript k so that $Y_i = Y_{kt_i}, \mathbf{X}_i = \mathbf{X}_{kt_i}, \mathbf{G}_{12} = \mathbf{G}_{kt_1t_2}$ and so on. We return to our original notation when we summarise our results below.

To obtain moments of $\mathbf{Y}$ with respect to the random covariates but conditional on $\mathbf{a}$, we expand $\zeta_i = \zeta_{kt_i}$ in a Taylor series with respect to $\mathbf{X}_i$ about $\mathbf{z}_i$ and retain only terms up to second order. The result:

$$\begin{aligned}\zeta_i \approx\ & \zeta\left(\mathbf{a}^T\mathbf{z}_i\right) + \zeta'\left(\mathbf{a}^T\mathbf{z}_i\right)\mathbf{a}^T\left(\mathbf{X}_i - \mathbf{z}_i\right) \\ & + \frac{1}{2}\zeta''\left(\mathbf{a}^T\mathbf{z}_i\right)\mathbf{a}^T\left(\mathbf{X}_i - \mathbf{z}_i\right)\left(\mathbf{X}_i - \mathbf{z}_i\right)^T\mathbf{a}.\end{aligned}$$

From this expansion we get

$$\mathbf{E}\left(Y_i \mid \mathbf{a}\right) \approx \zeta\left(\mathbf{a}^T\mathbf{z}_i\right) + \frac{1}{2}\zeta''\left(\mathbf{a}^T\mathbf{z}_i\right)\mathbf{a}^T\mathbf{G}_{ii}\mathbf{a}. \tag{1}$$

By retaining just the terms up to second order we get from the expansion given above for ζ_i and

$$\mathbf{Cov}\left(Y_1, Y_2 \mid \mathbf{a}\right) = \delta_{12}\phi E\left(Y_1 \mid \mathbf{a}\right) + \mathbf{Cov}\left(\zeta_1, \zeta_2 \mid \mathbf{a}\right)$$

the result

$$\mathbf{Cov}\left(Y_1, Y_2 \mid \mathbf{a}\right) \approx \Lambda_{12}(a), \tag{2}$$

where $\Lambda_{12}(a) = \delta_{12}\phi E\left(Y_1 \mid \mathbf{a}\right) + \zeta'\left(\mathbf{a}^T\mathbf{z}_1\right)\zeta'\left(\mathbf{a}^T\mathbf{z}_2\right)\mathbf{a}^T\mathbf{G}_{12}\mathbf{a}$.

Equations (1) and (2) give us models for the first and second conditional moments of the response. These models incorporate our uncertainty about the covariates measured with error. Motivated by the applications we make of these models (illustrated in Section 4), we simplify the conditional expectation model by adapting the approach of Lindstrom and Bates (1990) as modified by *BK-94*. In our adaptation of that approach we retain only terms up to second order in the Taylor expansion of the expectation about a point $a = a_k = \alpha_{ok} = \alpha_o$ specified in the next section. The result:

$$\mathbf{E}\left(Y_i \mid \mathbf{a}\right) \approx \eta_i(a), \tag{3}$$

where

$$\begin{aligned}\eta_i(a) &= \zeta\left(\alpha_o^T \mathbf{z}_i\right) + \hat{\mathbf{Z}}_i\left(\mathbf{a} - \alpha_o\right) \\ &\quad + \frac{1}{2}\zeta''\left(\alpha_o^T \mathbf{z}_i\right)\left[\mathbf{z}_i^T\left(\mathbf{a} - \alpha_o\right)\left(\mathbf{a} - \alpha_o\right)^T \mathbf{z}_i \right.\\ &\quad \left. + \alpha_o^T \mathbf{G}_{ii}\alpha_o\right]\end{aligned}$$

and $\hat{\mathbf{Z}}_i = \zeta'\left(\alpha_o^T \mathbf{z}_i\right)\mathbf{z}_i^T$.

Observe that the log convexity of ζ insures that $\eta_i > 0$. In turn this makes $\Lambda_{11} > 0$ when we approximate it by substituting η_1 for $E(Y_1 \mid \mathbf{a})$ in Equation 5 below.

We may now use our approximations for the conditional moments to get approximations for the unconditional moments of the responses. For the expectation we obtain

$$\mathbf{E}\left(Y_i\right) \approx \mu_i(\alpha_o), \tag{4}$$

where

$$\begin{aligned}\mu_i(\alpha_o) &= \zeta\left(\alpha_o^T \mathbf{z}_i\right) + \hat{\mathbf{Z}}_i\left(\alpha - \alpha_o\right) \\ &\quad + \frac{1}{2}\zeta''\left(\alpha_o^T \mathbf{z}_i\right)\left\{\mathbf{z}_i^T\left[\mathbf{D} + \left(\alpha - \alpha_o\right)\right]\left(\alpha - \alpha_o\right)^T \mathbf{z}_i\right] \\ &\quad + \alpha_o^T \mathbf{G}_{ii}\alpha_o\}.\end{aligned}$$

To get the unconditional reponse covariances we use the approximations:

$$\mathbf{E}\left(Cov(Y_1, Y_2 \mid \mathbf{a})\right) \approx \Lambda_{12}(\alpha_o)$$

and

$$E[\eta_1(a), \eta_2(a)] \approx \dot{\mathbf{Z}}_1 \mathbf{D} \dot{\mathbf{Z}}_2^T.$$

These last two approximations together give us our approximation to the unconditional covariance,

$$\mathbf{Cov}\,(Y_1, Y_2) \approx \Sigma_{12}(\alpha_o) \tag{5}$$

where

$$\Sigma_{12}(\alpha_o) = \Lambda_{12}(\alpha_o) + \dot{\mathbf{Z}}_1 \mathbf{D} \dot{\mathbf{Z}}_2^T.$$

Note that the positivity of $\mu_i(\alpha_o)$ assures that of $\Sigma(\alpha_o)$ obtained from this Equation 5.

In vector-matrix form using our original notation we now have:

$$\begin{aligned}
\mathbf{E}\,(\mathbf{Y}_k \mid \mathbf{a}_k) &\approx (\eta_{kt_1}(\mathbf{a}_k), \ldots, \eta_{kt_T}(\mathbf{a}_k))^T \equiv \eta_k(\mathbf{a}_k); \\
\mathbf{Cov}\,(\mathbf{Y}_k \mid \mathbf{a}_k) &\approx diag\,(\Lambda_{kt_1t_1}(\mathbf{a}_k), \ldots, \Lambda_{kt_Tt_T}(\mathbf{a}_k)) \equiv \mathbf{\Lambda}_k(\mathbf{a}_k); \\
\mathbf{E}\,(\mathbf{Y}_k) &\approx (\mu_{kt_1}(\alpha_{ok}), \ldots, \mu_{kt_T}(\alpha_{ok}))^T \equiv \mu_k(\alpha_{ok}); \\
\mathbf{Cov}\,(\mathbf{Y}_k) &\approx \Lambda_k(\alpha_{ok}) + \dot{\mathbf{Z}}_k \mathbf{D} \dot{\mathbf{Z}}_k^T \equiv \Sigma_k(\alpha_{ok}),
\end{aligned}$$

where $\dot{\mathbf{Z}}_k^T = \left(\dot{\mathbf{Z}}_{kt_1}^T, \ldots, \dot{\mathbf{Z}}_{kt_T}^T\right)$.

3 Fitting the Errors-in-Covariates Model

To bring our results into line with those of *BK-94* as well as Lindstrom and Bates (1990), we let $\mathbf{a}_k = \beta + \mathbf{b}_k$ and for cluster-specific analysis (Subsection 3.1), $\alpha_{ko} = \beta + \hat{\mathbf{b}}_k$, the $\{\mathbf{b}_k\}$ having expectations $\{0\}$. The estimates $\{\hat{\mathbf{b}}_k\}$ would be the current values obtained in the interative estimation of the $\{\mathbf{b}_k\}$. We regard these estimates as fixed for our errors-in-covariates model thereby obtaining a moment structure which formally resembles that of *BK-94*. Thus after observing the $\{\mathbf{Y}_k = \mathbf{y}_k\}$ and associated covariates, we may employ the methods of *BK-94* with minor modifications. Subsequently in Subsection 3.2 we let $\alpha_{ok} = \beta$.

3.1 Cluster Specific Model

For our model, the parameters we need to estimate are $\mathbf{b}_k$, β, $\mathbf{D}$, and ϕ. We interpret β as the change in the response that would be observed at a "typical" cluster resulting from a change in the covariate levels. Zeger, Liang and Albert (1988) (*hereafter referred to as ZLA-88*) call a model with such an interpretation *subject-specific*. The $\{\mathbf{b}_k\}$ show how the response rates differ among the individual clusters and we will therefore call this model cluster-specific' instead.

To estimate the random effects $\{\mathbf{b}_k\}$, we iteratively solve an estimating equation using the Fisher scoring algorithm. As noted in the last section, some of the coordinates of these random effect vectors and all corresponding elements of $\mathbf{D}$ will be identically zero. To simplify exposition, we augment $\mathbf{D}$ with diagonal elements as necessary to make it nonsingular. Recall that $\hat{Z}_{kt}$ is the gradient of $\eta_{kt}(\mathbf{a}_k)$ with respect to $\mathbf{a}_k$ evaluated at $\mathbf{a}_k = \beta + \hat{b}_k = \alpha_{k0}$ and $\hat{\mathbf{Z}}_k$ the corresponding vector. We let $\Lambda_k = \Lambda_k(\beta + \hat{b}_k)$ be our current estimate of $Cov(\mathbf{Y}_k \mid \mathbf{a}_k)$ The appropriate estimating equations then become

$$\mathcal{W}_k \equiv \hat{\mathbf{Z}}_k^T \Lambda_k^{-1} (\mathbf{y}_k - \eta_k(\beta + \mathbf{b}_k)) - \mathbf{D}^{-1}\mathbf{b}_k = 0.$$

To use the scoring algorithm we need the matrix obtained as the expectation of the (row) gradient with respect to $\mathbf{b}_k$ of the (column) vector $\mathcal{W}_k$. The result:

$$\mathbf{A} = -\hat{\mathbf{Z}}_k^T \Lambda_k^{-1} \hat{\mathbf{Z}}_k - \mathbf{D}^{-1}.$$

To get the next $\hat{\mathbf{b}}_k$, say $\hat{\mathbf{b}}_k^*$ we would then solve,

$$\mathbf{A}\hat{\mathbf{b}}_k^* = \mathbf{A}\hat{\mathbf{b}}_k - \mathcal{W}_k.$$

Using the well known matrix identity, $(P + Q^T RQ)^{-1} = P^{-1} - P^{-1}Q^T(R^{-1} + QP^{-1}Q^T)^{-1}QP^{-1}$ we get after some simplification

$$\hat{\mathbf{b}}_k^* = \mathbf{D}\hat{\mathbf{Z}}_k^T \Sigma_k^{-1} \tilde{\mathbf{r}}_k, \tag{6}$$

where $\tilde{\mathbf{r}}_k = \mathbf{y}_k - \eta_k(\beta + \hat{\mathbf{b}}_k) + \hat{Z}_k\hat{\mathbf{b}}_k$.

Now reset to 0, the diagonal elements added to make $\mathbf{D}$ nonsingular. We see that equation (6) returns a 0 for each of the elements of the random effects vector actually corresponding to fixed effects.

To estimate the hyperparameters, which to this point have been taken as fixed, we must turn to the marginal distribution of the $\{\mathbf{Y}_k\}$ evaluated at their realisations to get the quasi-log-likelihood

$$\mathcal{Q} = \sum_{k=1}^{K} \left(\ln \mid \Sigma_k \mid + \mathbf{r}_k^T \Sigma_k^{-1} \mathbf{r}_k \right), \tag{7}$$

where $\mathbf{r}_k = \mathbf{y}_k - \mu_k(\beta + \hat{\mathbf{b}}_k)$ It gives us the generalized estimating equations needed for estimating the hyperparameters. Those estimating equations for β are readily found to be

$$\sum_{k=1}^{K} \hat{\mathbf{X}}_k^T \Sigma_k^{-1} \mathbf{r}_k \;=\; \mathbf{0},$$

where $\hat{\mathbf{X}}_k$ is the $T \times p$ matrix of derivatives of μ_k with respect to β. We can solve these equations iteratively by the Fisher scoring algorithm, the leading term in $\hat{\mathbf{X}}_k$ is just $\hat{\mathbf{Z}}_k$. The updated estimate of β, say $\hat{\beta}^*$ would be given by

$$\hat{\beta}^* \;=\; \beta + \mathbf{H}_k \sum_{k=1}^{K} \hat{\mathbf{X}}_k \Sigma_k^{-1} \mathbf{r}_k, \tag{8}$$

where $\mathbf{H}_k = \left(\sum_{k=1}^{K} \hat{\mathbf{X}}_k^T \Sigma_k^{-1} \hat{\mathbf{X}}_k \right)^{-1}$ and parameters on the right hand sides of these equations are evaluated at their current values.

We adopt these updating equations as well as the *BK-94* iterative estimators for $\mathbf{D}$ and ϕ given by:

$$\hat{\mathbf{D}} = \mathbf{D} + \mathbf{D} \left(K^{-1} \sum_{k=1}^{K} \hat{\mathbf{Z}}_k^T \Sigma_k^{-1} \left(\mathbf{r}_k \mathbf{r}_k^T - \Sigma_k \right) \Sigma_k^{-1} \mathbf{Z}_k \right) \mathbf{D}; \tag{9}$$

$$\hat{\phi} \;=\; \phi (KT)^{-1} \sum_{k=1}^{K} \left(\mathbf{r}_k^T \Sigma_k^{-1} \mathbf{r}_k \right), \tag{10}$$

the right side of these equations being evaluated at the current parameter estimates in both cases. These equations may be obtained by using the EM algorithm to maximise the quasi-log-likelihood in Equation 7 which, being formally that of a normal distribution, admits the use of the approach of Laird and Ware (1982). To set up the algorithm we formally represent the responses by a linear model,

$$Y_{kt} = \mu_{kt} + \hat{\mathbf{Z}}_{kt}^T \mathbf{b}_k + \mathbf{U}_{kt} + \epsilon_{kt},$$

where the $\{\mathbf{b}_k\}$, $\{\mathbf{U}_{kt}\}$ and $\{\epsilon_{kt}\}$ are mutually independent and normally distributed with 0 means and variances/covariances $D = Cov(\mathbf{b}_k)$, $Var(\mathbf{U}_{kt}) = [\zeta'(\alpha_{ak}^T z_{kt})]^2 \alpha_{ak}^T \mathbf{G}_{kt t} \alpha_{ak}$ and $Var(\epsilon_{kt}) = \phi \mu_{kt}(\alpha_{ak})$, respectively. For the M-step we suppose we have observed the $\{\mathbf{b}_k\}$ and the residuals say $\{\epsilon_{kt}\}$. Then we would have natural estimates of $\mathbf{D}$ and ϕ:

$$\begin{aligned}
\hat{\mathbf{D}} &= K^{-1} \sum_{k=1}^{K} \mathbf{b}_k \mathbf{b}_k^T; \\
\hat{\phi} &= (KT)^{-1} \sum_{k=1}^{K} \sum_{t=1}^{T} (\epsilon_{kt})^2 \mu_{kt}^{-1}(\alpha_{ak}).
\end{aligned}$$

Since the $\{\mathbf{b}_k\}$ and $\{\epsilon_{kt}\}$ are not observed we need the E-step to calculate the conditional expectations of the $\{\mathbf{b}_k \mathbf{b}_k^T\}$ and the $(\epsilon_{kt})^2 \zeta_{kt}^{-1}$ given the $\{Y_{kt}\}$. The iterative estimate for $\mathbf{D}$ obtains directly from this calculation using standard distribution theory for the multivariate normal. That for ϕ does not stem directly from the analagous calculation which gives instead,

$$\begin{aligned}
\hat{\phi} = {} & \phi (KT)^{-1} \sum_{k=1}^{K} \sum_{t=1}^{T} \left(\mathbf{r}_k^T \Sigma_k^{-1} \mathbf{r}_k \right) \\
& + \phi (KT)^{-1} \sum_{k=1}^{K} \sum_{t=1}^{T} [I - \Sigma_k^{-1} \mathbf{r}_k \mathbf{r}_k^T] \mathbf{M}^*,
\end{aligned}$$

for a certain matrix $\mathbf{M}^*$. The second term in this last expression has expectation 0 so the first is 'unbiased' for ϕ leading to the iterative estimator stated in Equation 10.

Estimation of the parameters for the cluster specific model proceeds by iterating Equations (6)-(10) until the maximum relative change in each of the components of β is less than a specified tolerance. Alternatively we could continue until we see negligible relative change in the quasi-log-likelihood.

3.2 Population Average Model

Often an analysis focuses on the average effect over all clusters; then our marginal expectation model becomes

$$E\left(Y_{kt} \mid \mathbf{X}_{kt}\right) = \zeta^{*}\left(\beta^{*T}\mathbf{X}_{kt}\right). \qquad (11)$$

Here β^{*} represents the 'population-average regression parameter' where in the cluster-specific model of the last Subsection, β was the regression parameter for a 'typical' cluster. The coefficients in our population average model determine the expected change in the number of responses for the entire population due to a change in the mean levels of the covariates. Hence this model has been called a *population-average* model (ZLA-88).

The interpretation of such nonlinear models requires care. Suppose we select β^{*} so that $\zeta^{*}\left(\beta^{*T}\mathbf{X}_{kt}\right) = \zeta\left(\beta^{T}\mathbf{X}_{kt}\right)$. Then we see that the population average impact of $\mathbf{X}_{kt}$ equals that at a typical cluster (where $\mathbf{b}_k = 0$). However, as noted in *BK-94*, $\beta^{*} \neq \beta$ in general.

Observe that β^{*}'s coordinates depend on the degree of heterogeneity in the response rates among the clusters (through the integration with respect to the distribution of $\mathbf{b}_k$). If that heterogeneity is not explicitly modelled, the covariance structure of the unconditional $\mathbf{Y}_k$ is not analytically defined. However, by evaluating the covariance matrix of the cluster-specific model at $\mathbf{b}_k = 0$, for all clusters, we obtain below a plausible 'working covariance' for the population average model. We substitute Γ for $\mathbf{D}$ in that covariance to emphasize the distinction between them and the difference in their interpretation in these different contexts. Our overall objective now becomes the maximisation of the quasi-log-likelihood as objective function with the working covariance matrix so-determined.

We will in the sequel let $\beta^{*} = \beta$ and $\zeta^{*} = \zeta$ with little risk of confusion and gains in expository simplicity. Moreover we borrow results from the last subsection by viewing (11) as a degenerate random effects model with $\mathbf{b}_k \equiv 0$ and hence $\mathbf{D} = \mathbf{0}$. Thus we have

$$E[Y_{kt}] \approx \nu_{kt} \equiv \mu_{kt}(\beta) = \zeta(\beta^{T}x_{kt}) + \frac{1}{2}\zeta''(\beta^{T}x_{kt})\beta^{T}\mathbf{G}_{ktt}\beta.$$

To state the objective function and working covariance explicitly let:

$$
\begin{aligned}
\nu_k &= \left(\nu_{kt_1}, \ldots, \nu_{kt_T}\right)^T; \\
\mathbf{s}_k &= \mathbf{y}_k - \nu_k; \\
M_k &= \left[\zeta'\left(\beta^T \mathbf{z}_{kt_1}\right) \mathbf{z}_{kt_1}, \ldots, \zeta'\left(\beta^T \mathbf{z}_{kt_T}\right) \mathbf{z}_{kt_T}\right]^T; \\
\mathbf{V}_{kt} &= \tau \nu_{kt} + [\zeta'\left(\beta^T \mathbf{z}_{kt}\right)]^2 \mathbf{z}_{kt}^T G_{ktt} \mathbf{z}_{kt}; \\
\mathbf{V}_k &= diag\left(\nu_{kt_1}, \ldots, \nu_{kt_T}\right); \\
\mathbf{W}_k &= \mathbf{V}_k + M_k \Gamma M_k^T.
\end{aligned}
$$

The last of these equations gives the working covariance. Observe that it can formally be obtained by positing a random effects model like that considered in Section 2 (and setting $\alpha = \alpha_* = \beta$ in the Taylor expansion approximations developed there). In particular, we can formally represent the responses using the linear model:

$$Y_{kt} = \nu_{kt} + \mathbf{M}_{kt}\mathbf{b}_k + \mathbf{U}_{kt} + \epsilon_{kt}, \tag{12}$$

where the $\mathbf{M}_{kt}$ denotes the t-th row of $\mathbf{M}_k$ while $\{\mathbf{b}_k\}$, $\{\mathbf{U}_{kt}\}$ and $\{\epsilon_{kt}\}$ are mutually independent and normally distributed with 0 means and variances/covariances $\Gamma = Cov(\mathbf{b}_k)$,

$$Var(\mathbf{U}_{kt}) = [\zeta'(\beta^T z_{kt})]^2 \beta^T \mathbf{G}_{ktt} \beta$$

and $Var(\epsilon_{kt}) = \tau \nu_{kt}$, respectively.

To maximise the quasi-log-likelihood

$$\mathcal{P} = \sum_{k=1}^{n} \left(\ln \mid \mathbf{W}_k \mid + \mathbf{s}_k' \mathbf{W}_k^{-1} \mathbf{s}_k\right). \tag{13}$$

we can invoke the representation of the responses given in Equation 12 and formally appeal to the results of Subsection 3.1. We thereby obtain the recursive relations needed for iteratively fitting the proposed working covariance matrix

This observation emphasises the benefits we (following *BK-94*) gain in linking the population-average (or marginal) model to the cluster- specific model. While the interpretation of parameters in the resulting model differs from that of the cluster-specific model, the analysis proceeds along the same lines. We solve the estimating equations iteratively by the Fisher scoring algorithm. The updated estimate of β, say $\hat{\beta}^*$ would be given by

$$\hat{\beta}^* = \beta + \mathbf{A} \sum_{k=1}^{K} \mathbf{U}_k \mathbf{W}_k^{-1} \mathbf{s}_k, \tag{14}$$

where $\mathbf{U}_k = \nabla_\beta \nu_k$ and $\mathbf{A} = \left(\sum_{k=1}^{K} \mathbf{U}_k^T \mathbf{U}_k^{-1} \mathbf{M}_k \right)^{-1}$, all quantities being evaluated at current parameters estimates. To estimate Γ we get from the last subsection,

$$\hat{\Gamma} = \Gamma + \Gamma \left(K^{-1} \sum_{k=1}^{K} \mathbf{M}_k^T \mathbf{W}_k^{-1} \left(\mathbf{s}_k \mathbf{s}_k^T - \mathbf{W}_k \right) \mathbf{W}_k^{-1} \mathbf{M}_k \right) \Gamma, \tag{15}$$

where the right hand side of the equation is evaluated at the current parameter estimates. Estimating τ requires

$$\hat{\tau} = \tau (KT)^{-1} \sum_{k=1}^{K} \sum_{t=1}^{T} \mathbf{s}_k^T \mathbf{W}_k^{-1} \mathbf{s}_k. \tag{16}$$

As in Subsection 3.1 iteration of these estimates continues until the estimated regression coefficients converge or changes in the quasi-log-likelihood become negligible.

4 Applying the Theory

In this section we show how the methods described above can be used by briefly describing the study which led to their development. That study of Zidek et al (1996) used daily series of hospital admission counts in Ontario due to respiratory problems for the years 1983 through 1988. In the earlier work of BET and *DK-94* admissions were classified by hospital. Recognising that pollution effects would be confounded with population size the authors of those studies introduced hospital size into the analysis to control for population size.

Recognising that population size might not be fully accounted for in this way, it seemed desirable to check the original findings of the studies cited above by controlling for population size more directly. That led to the investigation described in this section, a reanalysis of the data with admissions classified by the census subdivision (CSD) in which the patient resided (Zidek et al 1996).

The approach taken in that study, led to the analysis of a series of daily admission counts for each of the 733 CSDs (clusters). The subseries comprised of the months May through August in each of the six years were extracted for modelling since prior work (*c.f.* *BK-94*) indicated that the strongest correlation between pollution and admissions occurred during those months.

The analysis sought the average effect of changes in the mean levels of pollution taken over all CSDs and hence used a population-average approach in the terminology of Subsection 3.2.

Daily hospital admissions series in Ontario exhibit seasonal variation as well as day-of-the-week effects. Given the need to adjust for these factor and the goal of estimating population-average effects, the expected admission count Y_{kt} given the pollution levels $\mathbf{X}_{kt}$ were modelled by:

$$
\begin{aligned}
\mathbb{E}(Y_{kt} \mid \mathbf{b}_k, \mathbf{X}_{kt}) &= \zeta(\beta^T \mathbf{X}_{kt}) \\
&\equiv m_{kt} \exp\left(\beta^T \mathbf{X}_{kt}\right),
\end{aligned}
$$

m_{kt} being a multiplier accounting for the effects of seasonality (trend), day-of-the-week, and the population size of the CSD. The first two effects were estimated from the data prior to fitting the model (see *BK-94* for details). The population counts were obtained from the 1986 Census. Hence, m_{kt} entered the model as a known constant. The conditional covariance of Y_{kt} was determined as in Section 2.

The effects of four pollutants, SO_4, O_3, SO_2 and NO_2 were investigated through their spatial interpolants. The climatic variables, maximum daily temperature and average daily humidity were interpolated into the analysis as well. All the explanatory variables were log-transformed (to improve the normal approximation of their joint distribution) and then passed through a high-pass filter so that only 'blips' remained in each series. AR(1) residuals were then extracted from the result to gain an approximately independent series on which the spatial interpolator could be applied. The AR(1) step was inverted on the interpolated series to get a spatial predictor for the high frequency interpolated explanatory vector series, one for each cluster.

The re-analysis found imprecision in the imputed cluster means to be effectively negligble. Regression coefficients were unchanged (to two significant figures) by the incorporation of the residual covariance of spatial prediction error. That is the role of spatial prediction turned out to be limited to bias correction.

Data were analysed on a summer-by-summer basis, so that changes in pollution effects over time could be assessed. In this way the analysis tried to assess the "model uncertainty" where standard errors merely reflect model-specific parameter uncertainty. Any deficiencies in the model would be compensated for by adjustments in fitted parameters. That in turn would lead to unstable parameter estimates, 'tracking' the deficiency of the model over summers. The summer-to-summer variability in these estimates could grossly dominate the variability estimated from the standard errors. In that case, the latter would if adopted uncritically yield a false sense of reliability in the model fit and ensuing substantive findings.

Partitioning the analysis as described above does lead to reduction in the apparent significance of any tests which might be performed. However, if the parameter estimates turn out to be fairly stable they can be combined albeit in a mildly suboptimal way to obtain overall conclusions and substantial power.

In fact, the study found the regression coefficient estimates to be generally stable from year-to-year and they were combined to get single overall tests of significance for the association of the various explanatory factors with hospital admission counts.

In summary, the study confirmed in a qualitative way, the earlier findings of BET, that increased levels of ozone and sulfate, suitably lagged, are associated with increases in the numbers of hospital admissions for respiratory morbidity. For the most part, these findings appear to be reproducible from year-to-year in a stable fashion. Possible residual confounding of population size with pollution effects after accounting for hospital size in catchment areas seems negligible. Effects due to meteorological factors (daily temperature and humidity) seem relatively small and not confounded with those due to pollutants. Finally, unmeasured pollutant levels can be interpolated sufficiently accurately that for the purpose of the regression analysis, that imprecision can be ignored in inferences made.

5 Concluding Remarks

We have chosen the generalised estimating approach as a vehicle for our methodology pending the development of credible models for conventional inference. Our view of the inferential problem is Bayesian however; we do not believe the random effects substructure can be justified on any other grounds.

In the model given here, the parameters of the covariate distribution are assumed known although in practice they will be estimates. Our theory fails to incorporate additional uncertainty stemming from the resulting estimation error. In principle we could include model structure representing that uncertainty and such improvement might be worthwhile in future work. However we believe that source of uncertainty to be negligible in our applications of the theory.

The assumption underlying the generalised estimating equations method, of conditional independence for data from different clusters may not be realistic. It would not be realistic for example when these clusters represent contiguous sampling domains such as geographical subregions. We intend to incorporate between cluster dependence in developing alternatives to that method.

6 Acknowledgements

During the completion of this manuscript the one of the authors (James V Zidek) was on leave at the Institute of Mathematics and Statistics, University of Kent at Canterbury, which generously provided facilities. At the same time he was partially supported by a Fellowship from the Engineering and Physical Sciences Research Council of the United Kingdom.

References

Breslow, N and Clayton, DG (1993). Approximate inference in generalised linear mixed models. *Jour Amer Statist Assoc*, 88, 9-25.

Brown, PJ, Le, ND and Zidek, JV(1994). Multivariate spatial interpolation and exposure to air pollutants. *Canadian Journal of Statistics* 22: 489-509.

Burnett, RT, Dales, RE, Raizenne, MR, Krewski, D, Summers, PW, Roberts, GR, Raad-Young, M, Dann, T, Brook, J (1994). Effects of low ambient levels of ozone and sulfates on the frequency of respiratory admissions to Ontario hospitals. *Environmental Research*, **65**, 172-94.

Burnett, R. and Krewski, D. (1994). "Air pollution effects on hospital admission rates: a random effects modelling approach," *Can Jour Statist*, 22, 441-458.

Burnett, RT, Ross, WH and Krewski, D (1995). Nonlinear mixed regression models. *Environmetrics*, 8, pages 85-99.

Carroll , RJ, Ruppert, D and Stefanski, LA (1995). *Measurement error in nonlinear models*. London: Chapman and Hall.

Diggle , PJ, Moyeed, RA and Tawn, JA (1995). Non-Gaussian geostatistics. Unpublished manuscript.

Diggle , PJ, Harper, L and Simon, SL (1996). Geostatistical analysis of residual contamination from nuclear weapons testing. Unpublished manuscript.

Fuller , W. (1987). *Measurement Error Models*. New York: Wiley

Laird, NM and Ware, JH (1982). Random-effects models for longitudinal data. *Biometrics*, 38, 963-074.

Le, ND and Zidek, JV (1992). Interpolation with uncertain spatial covariances: a Bayesian alternative to kriging." *Journal of Multivariate Analysis*, **43**, 351-74.

Le, ND, Sun, W and Zidek, JV (1995). Bayesian Multivariate Spatial Interpolation with Data Missing-by-Design. Submitted.

Lindstrom, , MJ and Bates, DM (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, 46, 673-687.

Liang, K-Y, and Zeger, SL (1986). Longitudinal data analysis using generalised linear models. *Biometrika*, 73, 13-22.

Pierce, DA, Stram, DO, Vaeth, M and Schafer, DW (1992), The errors-in-variables problem: considerations provided by radiation dose-response analyses of the a-bomb survivor data. it Jour Amer Statist Assoc, 87, 351-359.

Sun, W (1994). Bayesian multivariate interpolation with missing data and its applications. Ph.D. Thesis, University of British Columbia.

Sun, W. (1995). Comparison of a co-Kriging method with a Bayesian alternative". Technical Report No. 147, Department of Statistics, University of British Columbia.

Sun, W., Le, N. D., Zidek, J.V. and Burnett, R. (1995). Bayesian multivariate spatial interpolation: application and assessment. Technical Report No. 146, Department of Statistics, University of British Columbia.

Zeger, SL (1988) A regression model for time series of counts. *Biometrika*, 75, 621-629.

Zeger, SL and Liang, KY (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrika*, 42, 121-130.

Zeger, SL, and Karim, MR (1991). Generalised linear models with random effects; a Gibb's sampling approach. *Jour Amer Statist Assoc*, 86, 79-86.

Zeger, SL, Liang, KY and Albert, PS (1988). Models for longitudinal data; a generalised estimating equation approach. *Biometrics*, 44 1049-1060.

Zidek, , JV (1996).Interpolating Air Pollution for Health Impact Assessment. Proceedings of SPRUCE III. Submitted

Zidek, JV, Wong, H, Le, ND and Burnett, RT (1995). Causality, measurement error and multicollinearity in epidemiology. Technical Report BRR-08, Biostatistical Research Reports, Department of Statistics, University of British Columbia.

Zidek, JV, Wong, H, Le, ND and Zidek, JV (1996). Imputing Unmeasured Explanatory Variables in Environmental Epidemiology With Application To Heath Impact Analysis of Air Pollution. Submitted