Proceedings of Statistics Canada Symposium 2002
Modelling Survey Data for Social and Economic Research

BIAS REDUCTION IN STANDARD ERRORS FOR LINEAR AND GENERALIZED LINEAR MODELS WITH MULTI-STAGE SAMPLES

Daniel F. McCaffrey[1] and Robert M. Bell[2]

ABSTRACT

Linearization and the jackknife are widely used to estimate standard errors for the coefficients of linear regression models fit to multi-stage samples. For some designs, linearization estimators can have large negative bias, while the jackknife has a correspondingly large positive bias. We propose an alternative estimator, bias reduced linearization (BRL), based on residuals adjusted to better approximate the covariance of the true errors. When errors are iid, the BRL estimator is unbiased. The BRL method applies to samples with nonconstant selection weights and to generalized linear models such as logistic regression. We also discuss BRL standard error estimators for generalized estimating equation models that explicitly model the dependence among observations from the same PSU. Simulation study results show that BRL standard errors combined with the Satterthwaite approximation to determine the reference distribution yield tests with Type I error rates near nominal values.

KEY WORDS: complex samples, linearization, jackknife, Satterthwaite approximation, degrees of freedom, GEE

1. INTRODUCTION

Linearization is a widely used nonparametric method for estimating the standard errors of coefficients from linear and generalized linear regression models (Binder, 1983; Skinner, 1989). Although the traditional linearization estimator for standard errors performs well for samples with large numbers of primary sampling units (PSUs), the estimator can be biased, in particular biased low, when the number of PSUs is small or when the predictor variables are unbalanced across the PSUs (Bell and McCaffrey, 2002; Kott, 1994; Mancl and DeRouen, 2001). For example, Bell and McCaffrey (2002) show that the traditional linearization standard error estimator for ordinary least squares is biased low except under very restrictive assumption on the distribution of the explanatory variables. Bell and McCaffrey show that under similar conditions the jackknife estimator is biased high.

Kott (1996) proposed a method for reducing the bias in the linearization estimator for linear least squares regression. Mancl and DeRouen (2001) developed a different alternative in the context of generalized estimating equations (GEE). Both approaches suggest modifying residual vectors used in the traditional linearization estimator. Details are provided below. In Bell and McCaffrey (2002), we suggest an alternative method for adjusting residuals called biased reduced linearization (BRL).

In this paper, we review our results for ordinary least squares and discuss extensions of the BRL method to: 1) weighted least squares; 2) generalized least squares; 3) generalized linear models; 4) generalized estimating equations. We conclude with an application of logistic regression used to estimate the treatment effect in a cluster-randomized experiment.

2. ORDINARY LEAST SQUARES

We use ordinary least squares on a two-stage sample to develop the BRL estimator. We highlight the key steps in finding the estimator for least squares and these key steps suggest natural extensions for generalized linear models, GEE and weighted analyses.

[1] RAND, 201 North Craig Street, Suite 102, Pittsburgh, PA, USA, 15090
[2] AT&T Labs-Research, Room C211, 180 Park Avenue, Florham Park, NJ, USA, 07932

2.1 Linearization and the Jackknife

Let n equal the number of PSUs and m_i equal the number of final sampling units from the i-th PSU, for $i = 1,\ldots, n$. The overall sample size is $M = \sum_i m_i$. We assume that $y_{ij} = \beta' x_{ij} + \varepsilon_{ij}$, where $\boldsymbol{\varepsilon}$ has mean 0 and covariance matrix $\mathbf{V}$, and where y_{ij}, x_{ij}, and ε_{ij} all refer to the j-th observation from the i-th PSU. We drop the standard OLS assumption of *i.i.d.* errors, assuming only that errors from distinct PSUs are uncorrelated. Specifically, we assume that $\mathbf{V}$ is block diagonal, with $m_i \times m_i$ blocks $\mathbf{V}_i$ for $i = 1,\ldots, n$. In addition to the notation of this model, throughout the paper, we let $\mathbf{I}$ denote an $M \times M$ identity matrix and $\mathbf{I}_i$ equal a $m_i \times m_i$ identity matrix.

Let $\hat{\beta}$ denote the estimated coefficients of the linear regression model. To simplify presentation, we generally discuss a linear combination of the regression coefficients, $l'\hat{\beta}$, for an arbitrary column vector l. For the special case where one element of l is 1 and the rest are 0, $l'\hat{\beta}$ equals a single estimated coefficient. If errors are uncorrelated across PSUs, the variance of $l'\hat{\beta}$ is

$$Var(\, l'\hat{\beta}\,) = l'(\mathbf{X}'\mathbf{X})^{-1}\left(\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{V}_i\mathbf{X}_i\right)(\mathbf{X}'\mathbf{X})^{-1} l \,, \tag{2.1}$$

where $\mathbf{X}$ and $\mathbf{X}_i$ are the design matrices for the entire sample and for PSU i, respectively.

The standard linearization estimator of the variance of $l'\hat{\beta}$ is given by:

$$v_L = l'(\mathbf{X}'\mathbf{X})^{-1}\left(c\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{r}_i\mathbf{r}_i'\mathbf{X}_i\right)(\mathbf{X}'\mathbf{X})^{-1} l \tag{2.2}$$

where $\mathbf{r}_i$ is the vector of residuals for the i-th PSU. The unknown matrices $\mathbf{V}_i$ are estimated by $c\mathbf{r}_i\mathbf{r}_i'$, and c typically equals $n/(n-1)$.

The jackknife is sometimes used as an alternative to linearization (Rust and Rao, 1996). Let { $\tilde{\beta}_{[i]}$ } be a set pseudo values or estimates of β from data that exclude the i-th PSU. A jackknife estimator for multi-stage samples is:

$$v_{JK} = [(n-1)/n]\ \sum_i l'\left(\tilde{\beta}_{[i]} - \hat{\beta}\right)\left(\tilde{\beta}_{[i]} - \hat{\beta}\right)' l \tag{2.3}$$

In Bell and McCaffrey (2002) we show that the jackknife can be written as $v_{JK} = c\ l'(\mathbf{X}'\mathbf{X})^{-1}\ \{ \sum_{i=1}^{n}\ \mathbf{X}_i'\ (\mathbf{I}_i - \mathbf{H}_{ii})^{-1} \mathbf{r}_i\mathbf{r}_i'(\mathbf{I}_i - \mathbf{H}_{ii})^{-1}\mathbf{X}_i\} l$, with $c=(n-1)/n$, where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})\mathbf{X}'$, and $\mathbf{X}_i$ and $\mathbf{H}_{ii}$ denote the submatrices of $\mathbf{X}$ and $\mathbf{H}$ corresponding to the ith PSU. Thus, the jackknife is similar to the linearization estimator, but with the outer product of the raw residuals replaced by the outer product of adjusted residual. Theorems 1 and 2 of Bell and McCaffrey (2002) show that when $\mathbf{V} = \mathbf{I}$, linearization will be biased low except under restrictive conditions while the jackknife over corrects and ends up biased high.

Other authors have proposed adjustments to reduce the bias of linearization. Kott (1996) suggests calculating the ratio of $Var(\, l'\hat{\beta}\,)$ to $E(v_L)$ under the assumption that $\mathbf{V} = \mathbf{I}$ and adjusting v_L by the ratio. If $\mathbf{V} = \sigma^2\mathbf{I}$ then the resulting estimator will be unbiased. In the context of generalized estimating equations, Mancel and DeRouen (2001) suggest adjusting the residuals from each PSU to reduce the bias in $\mathbf{r}_i\mathbf{r}_i'$ as an estimator of $\mathbf{V}_i$. For an unweighted linear model, their method corresponds to approximating $E(\mathbf{r}_i\mathbf{r}_i')$ by $(\mathbf{I}_i - \mathbf{H}_{ii})\mathbf{V}_i(\mathbf{I}_i - \mathbf{H}_{ii})$ and replacing $\mathbf{r}_i$ by $(\mathbf{I}_i - \mathbf{H}_{ii})^{-1}\mathbf{r}_i$ in equation (2). Thus as shown in Bell and McCaffrey (2002) the Mancl and DeRouen estimator equals $[n/(n-1)]v_{JK}$ for unweighted linear models.

2.2 Bias Reduced Linearization Standard Errors

In Bell and McCaffrey we propose a compromise between the linearization and the jackknife estimators that we call bias reduced linearzation. Like the jackknife, BRL uses the outer product of adjusted residuals. However, the adjustment derives from the $E(\mathbf{r}_i\mathbf{r}_i') = (\mathbf{I} - \mathbf{H})_i\mathbf{V}(\mathbf{I} - \mathbf{H})_i'$. If we knew $\mathbf{V}$ then we could determine the matrices $\mathbf{A}_i$ such that $\mathbf{A}_i[(\mathbf{I} - \mathbf{H})_i\mathbf{V}(\mathbf{I} - \mathbf{H})_i']\mathbf{A}_i' = \mathbf{V}_i$. Because $\mathbf{V}$ is unknown we use a working covariance matrix in its place to derive our estimator. In particular, we suggest using a working covariance matrix of the form $\mathbf{U} = \sigma^2\mathbf{I}$, which simplifies the condition on $\mathbf{A}_i$ to $\mathbf{A}_i(\mathbf{I}_i - \mathbf{H}_{ii})\mathbf{A}_i' = \mathbf{I}$ or $\mathbf{A}_i = (\mathbf{I}_i - \mathbf{H}_{ii})^{-1/2}$. Theorem 3 of Bell and McCaffrey shows that if $\mathbf{V} = \sigma^2\mathbf{I}$, then v_{BRL} is unbiased. Section 3 considers alternative working covariance matrices.

For $m_i > 1$, $\mathbf{A}_i$ is not unique. If $\mathbf{V} = \sigma^2\mathbf{I}$, the choice of $\mathbf{A}_i$ is unimportant because any solution to (2.4) will produce an unbiased variance estimator. However, the resulting estimators are biased when $\mathbf{V} \neq \sigma^2\mathbf{I}$, and the bias can vary greatly with the choice of $\mathbf{A}_i$. We found (Bell and McCaffrey, 2002) that the symmetric square root of $(\mathbf{I}_i - \mathbf{H}_{ii})^{-1}$ worked best among the alternatives tried, and we refer to the estimator using this root as the biased reduced linearization estimator, $v_{BRL} = l'(\mathbf{X}'\mathbf{X})^{-1} \sum_i \mathbf{X}_i'\mathbf{A}_i\mathbf{r}_i\mathbf{r}_i'\mathbf{A}_i\mathbf{X}_i (\mathbf{X}'\mathbf{X})^{-1} l$.

2.3 Variation of Variance Estimators

Bell and McCaffrey (2002) show that v_{BRL} equals the weighted sum of n independent χ_1^2 random variables where the weights are the eigenvalues of the $n \times n$ matrix $\mathbf{G} = \{\mathbf{g}_i'\mathbf{V}\mathbf{g}_j\}$, for $\mathbf{g}_i = (\mathbf{I} - \mathbf{H})_i'\mathbf{A}_i\mathbf{X}_i(\mathbf{X}'\mathbf{X})^{-1}l$. In that paper we also show that, v_L and v_{JK} have similar distributions with $\mathbf{G}$ defined by $\mathbf{g}_{Li} = [n/(n-1)]^{1/2}(\mathbf{I} - \mathbf{H})_i'\mathbf{X}_i(\mathbf{X}'\mathbf{X})^{-1}l$ and $\mathbf{g}_{JKi} = [(n-1)/n]^{1/2}(\mathbf{I} - \mathbf{H})_i'(\mathbf{I} - \mathbf{H}_{ii})^{-1}\mathbf{X}_i(\mathbf{X}'\mathbf{X})^{-1}l$, respectively. If $\mathbf{V} = \sigma^2\mathbf{I}$ and $\mathbf{X}_i'\mathbf{X}_i(\mathbf{X}'\mathbf{X})^{-1}l$ for $i = 1, ..., n$ are constant, then av_L, av_{JK}, and av_{BRL} are all distributed χ_{n-1}^2 for $a = (n - 1)/Var(l'\hat{\beta})$ (Bell and McCaffrey, 2002). However, in general, the $\mathbf{X}_i'\mathbf{X}_i(\mathbf{X}'\mathbf{X})^{-1}l$ will not be constant and the squared coefficient of variation will exceed $2/(n - 1)$, the corresponding statistic for a χ_{n-1}^2 random variable. The coefficient of variation for any of the nonparametric variance estimators can be very large for certain designs. High variability occurs under the same conditions that v_L and v_{JK} are most biased—when residuals from only a few PSUs effectively determine the final variance estimate.

This excess variability is of particular concern when approximating the distribution of $t = l'\hat{\beta}/\sqrt{v*}$ under the null hypothesis that $l'\beta = 0$. For v_L, Shah, Holt and Folsom (1977) suggested comparing t to a reference t-distribution with $n - 1$ degrees of freedom. However, because the variance of $(n - 1)v_L/E(v_L)$ tends to be greater than $2(n - 1)$, tests using a t-distribution with $n - 1$ degrees of freedom would tend to have Type I error rates that exceed the nominal value, even if v_L were unbiased. Satterthwaite (1946) provides an alternative approximation for the distribution of the variance estimators. By matching the first two moments with that of a χ^2 random variable, we approximate, up to a scaling constant, the distribution of v_L, v_{BRL} or v_{JK} by a χ_f^2 where $f = 2/cv^2 = \left(\sum_{i=1}^n \lambda_i\right)^2 \Big/ \sum_{i=1}^n \lambda_i^2$ and the λ_i are the eigenvalues of the corresponding matrix $\mathbf{G}$. Tests based on reference t-distributions with f degrees of freedom would be expected to provide better Type I error rates than tests based on n - 1 degrees of freedom. Pan and Wall (2001) and Kott (1994, 1996) suggest using the Satterthwaite approximation to estimate the degrees of freedom for tests based on standard linearization or Kott's alternatives to linearization. The Satterthwaite degrees of freedom f require specifying the unknown matrix $\mathbf{V}$. We set $\mathbf{V}$ identically equal to the identity matrix—i.e., assume independent, homoskedastic errors for purposes of determining degrees of freedom.

The distribution of v_{BRL} (and the other variance estimators) tends to be less skewed and have less mass in the lower tail than the distribution of a χ_f^2 where f equals the Satterthwaite degrees of freedom. Hence, reference t-distributions based on the Satterthwaite approximation tend to overestimate tail probabilities. For example, when data from a couple of PSUs nearly determine the value of a coefficient, the Satterthwaite degrees of freedom can be less than two, incorrectly implying a chi-square density that is infinite at zero. Consequently, the probability of very large t-statistics may not be as large as the Satterthwaite approximation would imply, especially when the Satterthwaite degrees of freedom are less than 4 or 5. In such settings, saddlepoint approximations (Huzurbazar, 1999) provide a promising alternative.

3. EXTENSIONS

Derivation of the BRL estimator for OLS involved four steps:

1. derive the $Var(l'\hat{\beta})$ as the sum of terms $\mathbf{b}_i'\mathbf{V}_i\mathbf{b}_i$;
2. derive the $E(\mathbf{r}_i\mathbf{r}_i') = \mathbf{Q}_i$ using a working variance-covariance matrix, $\mathbf{U}$, for the unknown $\mathbf{V}_i$;
3. find the symmetric solutions to $\mathbf{A}_i\mathbf{Q}_i\mathbf{A}_i = \mathbf{U}_i$;
4. v_{BRL} equals the sum of the of terms $\mathbf{b}_i'\mathbf{A}_i\mathbf{r}_i\mathbf{r}_i'\mathbf{A}_i\mathbf{b}_i$.

We consider alternative models and extend BRL to these models by using the OLS template and deriving formulas for the $\mathbf{b}_i$, $\mathbf{Q}_i$ and $\mathbf{A}_i$. For all the extensions, $\mathbf{Q}_i$ and $\mathbf{A}_i$ will be of a similar form. $\mathbf{Q}_i$ equals the working covariance-matrix pre and post multiplied by rows of the projection matrix defining the residuals and their transpose. The $\mathbf{A}_i$'s equal the roots of matrices involving products of the roots of the working variance-covariance matrix and $\mathbf{Q}_i$.

Properties of the extensions to linear models follow directly from results for OLS. In particular, the estimators are unbiased when the working covariance matrix is proportional to the true covariance matrix. The OLS results do not apply to estimators for generalized linear models and GEE and the small sample properties of these estimators must be studied via simulation.

3.1 Ordinary Least Squares BRL for a Working Covariance That Is Not the Identity

In equation (2.4) we define the adjustment matrices for the BRL estimator assuming a variance covariance matrix, $\mathbf{V} = k\mathbf{I}$, for an unspecified constant k. In some instances, we might want to use an alternative block-diagonal matrix as the working, covariance matrix, $\mathbf{U}$, when estimating our standard errors. In this case, we would continue to estimate the variance of $l'\hat{\beta}_{OLS}$ as given by (2.1) but now $\mathbf{Q}_i = (\mathbf{I} - \mathbf{H})_i\mathbf{U}(\mathbf{I} - \mathbf{H})_i$ and the adjustment matrices solve

$$\mathbf{A}_i(\mathbf{I} - \mathbf{H})_i\mathbf{U}(\mathbf{I} - \mathbf{H})_i'\mathbf{A}_i' = \mathbf{U}_i. \tag{3.1}$$

We let $\mathbf{Q}^{1/2}$ denote any matrix $\mathbf{Q}^{1/2\prime}\mathbf{Q}^{1/2} = \mathbf{Q}$ and $\mathbf{Q}^*$ denote the symmetric root of $\mathbf{Q}^{-1}$ provided it exits, i.e., $\mathbf{Q}^*\mathbf{Q}^* = \mathbf{Q}^{-1}$ and $\mathbf{Q}^*\mathbf{Q}\mathbf{Q}^* = \mathbf{I}$. The symmetric solution to (3.1) is

$$\mathbf{A}_i = \mathbf{U}_i^{1/2}(\mathbf{U}_i^{1/2}\mathbf{Q}_i\mathbf{U}_i^{1/2\prime})^*\mathbf{U}_i^{1/2}. \tag{3.2}$$

3.2 Weighted Least Squares

We consider the case where each observation has a case weight w_{ij} and let $\mathbf{W} = \text{diag}\{w_{ij}\}$. The weighted least squares estimator of the regression coefficients are $\hat{\beta}_W = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}$ and $var(l'\hat{\beta}_W) = l'(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\left[\sum_i \mathbf{X}_i'\mathbf{W}_i\mathbf{V}_i\mathbf{W}_i\mathbf{X}_i\right](\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}l$. Because $\mathbf{r}_i = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y} = \mathbf{G}_W\mathbf{y}$, we have $\mathbf{Q}_i = (\mathbf{I} - \mathbf{G}_W)_i\mathbf{U}(\mathbf{I} - \mathbf{G}_W)_i'$ and $\mathbf{A}_i = \mathbf{U}_i^{1/2}(\mathbf{U}_i^{1/2}\mathbf{Q}_i\mathbf{U}_i^{1/2})^*\mathbf{U}_i^{1/2}$.

3.3 Linear Generalized Least Squares

We consider generalized least squares estimation of the coefficients using the working covariance matrix $\mathbf{U}$. However, rather than use the model-based standard error, we use the linearization standard error estimator to protect inference against mis-specification in the working covariance matrix. This is a common practice used in the analysis of longitudinal data (see for example Liang and Zeger, 1986).

The weighted least squares estimator of the regression coefficients are $\hat{\beta}_{GLS} = (\mathbf{X}'\mathbf{U}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}^{-1}\mathbf{y}$ and $var(l'\hat{\beta}_{GLS}) = l'(\mathbf{X}'\mathbf{U}^{-1}\mathbf{X})^{-1}\left[\sum_i \mathbf{X}_i'\mathbf{U}_i^{-1}\mathbf{V}_i\mathbf{U}_i^{-1}\mathbf{X}_i\right](\mathbf{X}'\mathbf{U}^{-1}\mathbf{X})^{-1}l$. The projection matrix is $G_{GLS} = \mathbf{X}(\mathbf{X}'\mathbf{U}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}^{-1}$ and $\mathbf{Q}_i = (\mathbf{I} - G_{GLS})_i\mathbf{U}(\mathbf{I} - G_{GLS})_i'$. $\mathbf{A}_i$ equals $\mathbf{U}_i^{1/2}(\mathbf{U}_i^{1/2}\mathbf{Q}_i\mathbf{U}_i^{1/2})^*\mathbf{U}_i^{1/2}$. However, $\mathbf{Q}_i = \mathbf{U}_i^{1/2}(\mathbf{I} - H_{GLS})_i\mathbf{U}^{-1/2}\mathbf{U}\mathbf{U}^{-1/2}(\mathbf{I} - H_{GLS})_i'\mathbf{U}i^{1/2} = \mathbf{U}_i^{1/2}(\mathbf{I} - H_{GLS,ii})\mathbf{U}i^{1/2}$, so that

$$\mathbf{A}_i = \mathbf{U}_i^{1/2}[\mathbf{U}_i(\mathbf{I}_i - \mathrm{H}_{GLS,ii})\mathbf{U}_i]^*\mathbf{U}_i^{1/2} \tag{3.7}$$

where $\mathrm{H}_{GLS} = \mathbf{U}^{-1/2}\mathbf{X}(\mathbf{X}'\mathbf{U}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}^{-1/2}$.

3.4 Generalized Linear Models

We consider the generalized linear model where the density function for individual response Y_{ij} is assumed to be

$$f_Y(y_{ij}) = exp\{(y_{ij}\theta_{ij} - b(\theta_{ij}))/a(\phi) + c(y_{ij}, \phi)\} \tag{3.8}$$

where $\theta_{ij} = h(\eta_{ij})$ and $\eta_{ij} = \boldsymbol{x}_{ij}'\beta$. The mean and variance are given by $\mu_{ij} = E(y_{ij}) = \dot{b}(\theta_{ij})$ and $v = \mathrm{E}(y) = \ddot{b}(\theta_{ij})a(\phi)$. For estimating the coefficients, the observations are assumed to be independent so that the maximum likelihood estimates of the coefficients are found as the solution to the estimating equation:

$$\textstyle\sum_i \mathbf{X}_i'\Delta_i(\mathbf{y}_i - \boldsymbol{\mu}_i) = 0 \tag{3.9}$$

where $\Delta_i = \mathrm{diag}\{d\theta_{ij}/d\eta_{ij}\}$. Solutions to (3.9) are found via iteratively reweighted least squares where at the final iteration

$$\hat{\beta}_{MLE} = (\mathbf{X}'\mathbf{U}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}^{-1}\mathbf{z} \tag{3.10}$$

where $z_{ij} = \boldsymbol{x}_{ij}'\hat{\beta}_{MLE} + (y_{ij} - \hat{\mu}_{ij}) / \{\ddot{b}(\hat{\theta}_{ij})\dot{h}(\hat{\eta}_{ij})\}$, $\mathbf{U}^{-1} = \mathrm{diag}\{\ddot{b}(\hat{\theta}_{ij})\dot{h}(\hat{\eta}_{ij})\}$, $\hat{\eta}_{ij} = \boldsymbol{x}_{ij}'\hat{\beta}_{MLE}$, $\hat{\theta}_{ij} = h(\hat{\eta}_{ij})$ and $\hat{\mu}_{ij} = \dot{b}(\hat{\theta}_{ij})$. Under the working assumption of independent observations, the variance of $\mathbf{z}_i$ is approximately $\mathbf{U}_i$, up to a scaling term. Thus, generalized linear models are analogous to generalized least squares for linear models and we can derive a BRL estimator for GLM using the formulas for GLS.

First we need an estimate of $var(l'\hat{\beta}_{MLE})$ under the less restrictive assumption that $var(\mathbf{z})$ is block diagonal, $\mathbf{V}_i$. $l'\hat{\beta}_{MLE}$ is approximately normally distributed with $l'(\mathbf{X}'\mathbf{U}^{-1}\mathbf{X})^{-1}\left[\sum_i \mathbf{X}_i'\mathbf{U}_i^{-1}\mathbf{V}_i\mathbf{U}_i^{-1}\mathbf{X}_i\right](\mathbf{X}'\mathbf{U}^{-1}\mathbf{X})^{-1}l$.

Next we need to derive $\mathbf{Q}_i$. We let $\mathbf{G}_{GLM} = \mathbf{X}(\mathbf{X}'\mathbf{U}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}^{-1}$ and $\mathbf{r}_i = (\mathbf{z}_i - \hat{\mathbf{z}}_i)$, and first order approximations yield

$$E(\mathbf{r}_i\mathbf{r}_i') = (\mathbf{I} - \mathbf{G}_{GLM})_i\mathbf{V}_i(\mathbf{I} - \mathbf{G}_{GLM})_i'$$

Thus, just as in GLS, $\mathbf{Q}_i = (\mathbf{I} - \mathbf{G}_{GLM})_i\mathbf{U}_i(\mathbf{I} - \mathbf{G}_{GLM})_i'$. Finally we derive adjustment matrices as solutions to

$$\mathbf{A}_i(\mathbf{I} - \mathrm{G}_{GLM})_i\mathrm{U}(\mathbf{I} - \mathrm{G}_{GLM})_i\mathbf{A}_i' = \mathbf{U}_i,$$

which according to the GLS derivations are given by

$$\mathbf{A}_i = \mathbf{U}_i^{1/2}(\mathbf{U}_i^{1/2}\mathbf{Q}_i\mathbf{U}_i^{1/2})^*\mathbf{U}_i^{1/2} = \mathbf{U}_i^{1/2}[\mathbf{U}_i(\mathbf{I}_i - \mathrm{H}_{GLM,ii})\mathbf{U}_i]^*\mathbf{U}_i^{1/2}$$

where $\mathrm{H}_{GLM} = \mathbf{U}^{-1/2}\mathbf{X}(\mathbf{X}'\mathbf{U}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}^{-1/2}$. The BRL estimator for GLM is

$$v_{BRL,GLM} = l'(\mathbf{X}'\mathbf{U}^{-1}\mathbf{X})^{-1}\left[\sum_i \mathbf{X}_i'\mathbf{U}_i^{-1}\mathbf{A}_i\mathbf{r}_i\mathbf{r}_i'\mathbf{A}_i'\mathbf{U}_i^{-1}\mathbf{X}_i\right](\mathbf{X}'\mathbf{U}^{-1}\mathbf{X})^{-1}l$$

3.5 Generalized Estimating Equations

Generalized estimating equations extend generalized linear models to allow for correlation among observations from the same unit. The working covariance matrix is given by a block diagonal matrix $\mathbf{U}$ such that the $a(\phi)var(\mathbf{y}_{ij}) = \mathbf{U}_i = \boldsymbol{\Omega}_i^{1/2}\mathbf{R}_i\boldsymbol{\Omega}_i^{1/2}$. The regression coefficients are estimated as the solution to the estimating equations

$$\sum_i \mathbf{D}_i' \mathbf{U}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i) = 0\,, \tag{3.11}$$

where $\mathbf{D}_i = d\boldsymbol{\mu}_i/d\beta = \boldsymbol{\Omega}_i\boldsymbol{\Delta}_i\mathbf{X}_i$. As discussed in Liang and Zeger (1986), the coefficient estimates are found by iterating between a modified Fisher scoring to estimate β conditional on the current values of parameters of $\mathbf{R}$ and ϕ and method of moment estimation of these correlation and scaling parameters. The modified Fisher scoring algorithm is equivalent to iteratively reweighted least squares. We let z_{ij} be defined as above and $\tilde{\boldsymbol{\Omega}}$ and $\tilde{\mathbf{R}}$ denote the matrices $\boldsymbol{\Omega}$ and $\mathbf{R}$ evaluated at the current estimated values of their parameters, so that a the working covariance matrix for $\mathbf{z}_i$ is $\mathbf{U}_i = \tilde{\boldsymbol{\Omega}}_i^{-1/2}\tilde{\mathbf{R}}_i\tilde{\boldsymbol{\Omega}}_i^{-1/2}$. The GEE estimator of the coefficients is then given by

$$\hat{\beta}_{GEE} = (\mathbf{X}'\tilde{\boldsymbol{\Omega}}^{1/2}\tilde{\mathbf{R}}^{-1}\tilde{\boldsymbol{\Omega}}^{1/2}\mathbf{X})^{-1}\mathbf{X}'\tilde{\boldsymbol{\Omega}}^{1/2}\tilde{\mathbf{R}}^{-1}\tilde{\boldsymbol{\Omega}}^{1/2}\mathbf{z} = (\mathbf{X}'\mathbf{U}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}^{-1}\mathbf{z}$$

Again following the analogy of GLS, $(\mathbf{X}'\mathbf{U}^{-1}\mathbf{X})^{-1}\left\{\sum_i \mathbf{X}_i'\mathbf{U}_i^{-1}\mathbf{V}_i\mathbf{U}_i^{-1}\mathbf{X}_i\right\}(\mathbf{X}'\mathbf{U}^{-1}\mathbf{X})^{-1}$ approximates the variance of the asymptotic normal distribution of $l'\hat{\beta}_{GEE}$ and $E(\mathbf{r}_i\mathbf{r}_i') \approx (\mathbf{I} - \mathbf{G}_{GEE})_i\mathbf{V}_i(\mathbf{I} - \mathbf{G}_{GEE})_i'$, $\mathbf{G}_{GEE} = \mathbf{G}_{GEE} = \mathbf{X}(\mathbf{X}'\mathbf{U}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}^{-1} = \mathbf{X}(\mathbf{X}'\tilde{\boldsymbol{\Omega}}^{1/2}\tilde{\mathbf{R}}^{-1}\tilde{\boldsymbol{\Omega}}^{1/2}\mathbf{X})^{-1}\mathbf{X}'\tilde{\boldsymbol{\Omega}}^{1/2}\tilde{\mathbf{R}}^{-1}\tilde{\boldsymbol{\Omega}}^{1/2}$. We let $\mathbf{Q}_i = (\mathbf{I} - \mathbf{G}_{GEE})_i\mathbf{U}_i(\mathbf{I} - \mathbf{G}_{GEE})_i' = (\mathbf{I} - \mathbf{G}_{GEE})_i'\,(\tilde{\boldsymbol{\Omega}}_i^{-1/2}\tilde{\mathbf{R}}_i\tilde{\boldsymbol{\Omega}}_i^{-1/2})(\mathbf{I} - \mathbf{G}_{GEE})_i'$ and

$$\mathbf{A}_i = \mathbf{U}_i^{1/2}(\mathbf{U}_i^{1/2}\mathbf{Q}_i\mathbf{U}_i^{1/2})^*\mathbf{U}_i^{1/2}$$

$$= \tilde{\boldsymbol{\Omega}}_i^{-1/2}\tilde{\mathbf{R}}_i^{1/2}\{\,\tilde{\mathbf{R}}^{1/2}\tilde{\boldsymbol{\Omega}}^{-1}\tilde{\mathbf{R}}^{1/2}(\mathbf{I}_i - \mathbf{H}_{GEEii})\,\tilde{\mathbf{R}}^{1/2}\tilde{\boldsymbol{\Omega}}^{-1}\tilde{\mathbf{R}}^{1/2}\,\}\,\tilde{\mathbf{R}}_i^{1/2}\tilde{\boldsymbol{\Omega}}_i^{-1/2} \tag{3.12}$$

where $\mathbf{H}_{GEE} = \tilde{\mathbf{R}}^{-1/2}\tilde{\boldsymbol{\Omega}}^{1/2}\mathbf{X}(\mathbf{X}'\tilde{\boldsymbol{\Omega}}^{1/2}\tilde{\mathbf{R}}^{-1}\tilde{\boldsymbol{\Omega}}^{1/2}\mathbf{X})^{-1}\mathbf{X}'\tilde{\boldsymbol{\Omega}}^{1/2}\tilde{\mathbf{R}}^{-1/2}$. Equation (3.12) holds because $(\mathbf{I} - \mathbf{G}_{GEE}) = \tilde{\boldsymbol{\Omega}}_i^{-1/2}\tilde{\mathbf{R}}_i^{1/2}(\mathbf{I} - \mathbf{H}_{GEE})\,\tilde{\mathbf{R}}_i^{-1/2}\tilde{\boldsymbol{\Omega}}_i^{1/2}$.

The BRL estimator of the variance of $l'\hat{\beta}_{GEE}$ is

$$V_{BRL,GEE} = (\mathbf{X}'\tilde{\boldsymbol{\Omega}}^{1/2}\tilde{\mathbf{R}}^{-1}\tilde{\boldsymbol{\Omega}}^{1/2}\mathbf{X})^{-1}\left\{\sum_i \mathbf{X}_i'\tilde{\boldsymbol{\Omega}}_i\tilde{\mathbf{R}}_i^{-1}\tilde{\boldsymbol{\Omega}}_i\mathbf{A}_i\mathbf{r}_i\mathbf{r}_i'\mathbf{A}_i'\tilde{\boldsymbol{\Omega}}_i\tilde{\mathbf{R}}_i^{-1}\tilde{\boldsymbol{\Omega}}_i\mathbf{X}_i\right\}(\mathbf{X}'\tilde{\boldsymbol{\Omega}}^{1/2}\tilde{\mathbf{R}}^{-1}\tilde{\boldsymbol{\Omega}}^{1/2}\mathbf{X})^{-1}.$$

4. EMPIRICAL RESULTS

4.1 Monte Carlo Study for Ordinary Least Squares

In Bell and McCaffrey (2002), we report the results of a Monte Carlo simulation to study the properties of alternative variance estimators and tests for OLS and a balanced two-stage cluster sample with $n = 20$ PSUs and a constant $m = 10$ observations in each PSU. We provide a review of those results here.

4.1.1 Simulation Study Design

In that study, all simulation replications use a common design matrix $\mathbf{X}$ with four explanatory variables chosen to represent a range of difficulty for nonparametric variance estimators. The first two explanatory variables, x_1 and x_2, are dichotomous (0 or 1) and constant within PSU. The variable x_1 is 1 in half the clusters: 1, 3,...,19, while x_2 is 1 in just three clusters: 9, 10, and 11. Both x_3 and x_4 were generated from standard normal distributions. They differ in that x_3 was generated from a multivariate normal with intra-cluster correlation of 0.5 within PSU, while x_4 was generated from independent normal distributions.

The dependent variable $y_{ij} = \beta' x_{ij} + \varepsilon_{ij}$, where $\beta = 0$ and the $\boldsymbol{\varepsilon}_i$'s are standard multivariate normal random variables with intra-cluster correlation ρ. We use two alternative values of $\rho = 0$, and 1/3, corresponding to design effects for the sample mean of $DEFF$ = 1 and 4, respectively ($DEFF$=1+(m-1)ρ). Monte Carlo results are based on 100,000 replications of **y** for our fixed **X**. Results for ρ = 1/9 are presented in Bell and McCaffrey (2002).

We evaluated the ordinary least squares (OLS) variance estimator, $s^2 l'(\mathbf{X}'\,\mathbf{X})^{-1} l$, and four nonparametric variance estimators: the standard linearization estimator given in equation (2.2) with $c = n/(n-1)$; the jackknife estimator given in (2.3); bias reduced linearization; and Kott's 1996 method. BRL and the Kott adjustments are based on working intra-cluster correlations of ρ = 0.

We estimated Type I error rates for eight alternative test procedures based on 100,000 replications from the null hypothesis where each β_k = 0, for k = 0 to 4. Each procedure compares a "t-statistic" against a reference t-distribution. For the t's based on linearization, the jackknife, and BRL, we use critical values from t-distributions with both (n – 1) = 19 degrees of freedom and the corresponding Satterthwaite approximation. For Kott's method, we use his proposed degrees of freedom. All computations were implemented in SAS.

4.1.2 Simulation Study Results

Table 1 shows the bias of several variance estimators for the five regression coefficients (including the intercept) for ρ = 0 and 1/3. The OLS variances are unbiased for ρ = 0, but they are badly biased for ρ = 1/3. For PSU-level variables (including the intercept), the OLS variances are too small by roughly a factor of 1/$DEFF$. Similarly, the bias is smaller, but still substantial for x_3, the individual-level variable with large intra-cluster correlation. The positive bias for the OLS variance of $\hat{\beta}_4$ results from a slight negative intra-cluster correlation for x_4.

Table 1. Bias of Variance Estimators (as a Percentage of the True Variance).

Estimator	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
			$\rho = 0$		
OLS	0.0	0.0	0.0	0.0	0.0
Linearization	-9.6	-13.2	-32.5	-13.3	-1.8
Jackknife	11.7	17.2	51.2	17.6	2.1
Kott `96	0.0	0.0	0.0	0.0	0.0
BRL	0.0	0.0	0.0	0.0	0.0
			$\rho = 1/3$		
OLS	-75.8	-75.5	-76.2	-65.3	13.8
Linearization	-10.7	-14.8	-33.5	-19.9	-4.1
Jackknife	10.7	15.9	49.5	21.4	5.9
Kott `96	-1.2	-1.9	-1.5	-7.7	-2.3
BRL	-1.0	-1.5	-1.3	-2.1	0.4

Source: Bell and McCaffrey (2002)

Linearization and the jackknife each suffer from large biases, relatively independent of ρ, but the biases point in opposite directions. For each estimator, the magnitude of the bias varies greatly among the coefficients. The largest biases (in absolute value) occur for $\hat{\beta}_2$, which depends mainly on the data from only three PSUs. The next greatest biases occur for $\hat{\beta}_3$, followed closely by $\hat{\beta}_1$ and $\hat{\beta}_0$.

By design, Kott `96 and BRL eliminate the bias for $\rho$ = 0. Both methods reduce the magnitude of bias dramatically relative to linearization for $\rho$ = 1/9 and 1/3. Although the two methods are practically indistinguishable for PSU-level variables, Kott `96 performs substantially worse for $\hat{\beta}_3$ and $\hat{\beta}_4$ with relative biases of -7.7 and –2.3 percent compared to –2.1 and 0.4 for BRL.

Table 2 shows that Type I error rates for the standard linearization method with (n-1) degrees of freedom consistently exceed 5 percent for both values of ρ. Type I errors are most common for $\hat{\beta}_2$, where they reach as high as 16 percent, but they also occur much too frequently for $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_3$, ranging from 7.0 to 8.8 percent. The magnitude of this problem correlates closely with the size of the bias of the linearization estimator (see Table 1). Type I error rates are much lower, 5.7 to 6.4 percent, for tests based on the Satterthwaite degrees of freedom. Thus using the alternative degrees of freedom improved the Type I error rates by about 30 to 88 percent.

Table 2. Type I Error Rates for Tests of the Null Hypothesis that β = 0

Estimator	Df	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
				$\rho = 0$		
Linearization	n-1	7.54	7.00	15.99	7.35	5.38
Linearization	Satt	5.75	6.45	6.33	6.28	5.18
Jackknife	n-1	5.01	3.92	7.58	4.52	5.02
Jackknife	Satt	3.80	3.43	1.41	3.26	4.77
Kott `96	Kott	5.11	5.08	4.85	4.76	5.07
BRL	n-1	6.28	5.37	11.25	5.90	5.21
BRL	Satt	4.73	4.86	3.12	4.72	5.00
				$\rho = 1/3$		
Linearization	n-1	8.10	7.28	16.39	8.79	5.66
Linearization	Satt	6.30	6.78	6.62	7.53	5.44
Jackknife	n-1	5.45	4.11	7.76	4.56	4.67
Jackknife	Satt	4.13	3.61	1.51	3.35	4.46
Kott `96	Kott	5.59	5.44	5.14	5.88	5.31
BRL	n-1	6.76	5.63	11.55	6.45	5.19
BRL	Satt	5.18	5.14	3.30	5.26	4.98

NOTE: Entries with a true value of 5.00 percent have standard errors of 0.07 percent.
Source: Bell and McCaffrey (2002)

There is a less consistent pattern for the Type I error probabilities for the jackknife. The jackknife with (n-1) degrees of freedom tends to be conservative for $\hat{\beta}_1$ and $\hat{\beta}_3$, in accord with the positive bias in the jackknife variance. In contrast, the probability of Type I error is much too large for $\hat{\beta}_2$, and a bit too large for the intercept $\hat{\beta}_0$ when $\rho = 1/3$. The apparent explanation is that the choice of (n-1) as the degrees of freedom for the reference t-distribution sometimes counteracts the bias in the jackknife variance. This conclusion is supported by the very low Type I error rates for the jackknife with Satterthwaite degrees of freedom; smaller degrees of freedom combined with large positive biases result in very conservative tests.

BRL with (n-1) degrees of freedom improves substantially on linearization with the same degrees of freedom. Because BRL is unbiased when $\rho = 0$, comparing the fifth row of the table against the first demonstrates the reduction in Type I errors that results from removing the bias of linearization. Excluding $\hat{\beta}_4$, BRL reduces Type I error rates by about 45 to 88 percent. However, BRL with (n-1) degrees of freedom remains consistently liberal, especially for $\hat{\beta}_2$. Comparison of rows 2 and 6 of each section shows the relative impact of bias reduction and the Satterthwaite adjustment. For $\hat{\beta}_0$ and $\hat{\beta}_2$, degrees of freedom are more important, while bias matters more for $\hat{\beta}_1$ and $\hat{\beta}_3$. Performance for BRL with the Satterthwaite approximation is very good, except for $\hat{\beta}_2$, where the Type I error falls to about 3 percent.

Tests based on Kott's 1996 estimator also perform well. For almost all the coefficients and both values of ρ the Type I error rate is close to 5 percent. The exception is the test for $\hat{\beta}_3$ when $\rho = 1/3$, which has an error rate of 5.88 percent as a result of the moderate bias in the variance estimator.

We have also conducted simulation studies for the extensions of BRL and the desirable properties obtained in OLS appear to transfer to these other models. In McCaffrey, Bell and Botts (2001) we use the same design matrix as Bell and McCaffrey (2002) to study the properties of BRL estimators for weighted and generalized least squares. BRL standard errors for WLS and GLS have very small bias when the working covariance matrix deviates from the true covariance matrix of the error terms. For generalized linear models (logistic regression), preliminary simulation study results suggest that inference based on the BRL estimates and reference distributions using Satterthwaite-like approximate degrees of freedom, tend to have near nominal Type 1 errors over a range of true values for the regression coefficients and the intra-cluster correlation. However additional simulation study is required to test the generality of these preliminary findings.

4.3 Application: Logistic Regression for Partners-In-Care Intervention

We illustrate the methods in this paper using data from Partners in Care, a longitudinal experiment assessing the effect of "quality improvement" programs on care for depression in managed care organizations (MCOs) (Wells *et al.* 2000). The experiment followed 1356 patients who screened positive for depression in 1996-1997 in 43 clinics of seven MCOs. In each of nine blocks, clinic sets of one to four clinics were assigned at random to one of three experimental cells: usual care, or a quality improvement program supplemented by either nurses for medication follow-up or access to psychotherapists. Six MCOs constituted single blocks, and one MCO was divided into three blocks based on ethnic mix of the clinics. Within blocks with more than three clinics, clinics were combined into sets matched as closely as possible on anticipated sample size and patient characteristics. See Wells *et al.* (2000) for additional details.

One outcome of particular interest was receipt of appropriate care during the six months preceding the first follow-up. Receipt of appropriate care was coded as a dichotomous variable equaling one if the patient received appropriate medication or therapy and zero otherwise (Wells *et al*. 2000). We present results from a logistic regression model for appropriate care for 1143 patients at 6-month follow-up. As in Wells *et al.* (2000), the independent variable of primary interest is an intervention indicator that estimates the combined effect of medication or therapy versus care as usual. Our regression differs from theirs because we do not use sampling weights or impute for missing values of the outcome variable, but the results for the intervention effect agree reasonably closely.

Because patients from the same clinics could have similar outcomes, logistic regression standard errors could easily be too low—especially for PSU-level variables like Intervention. We compare the linearization estimator to the BRL estimator given in Section 3, $v_{BRL,GLM}$ using the adjustment matrices given in equation 3.7.

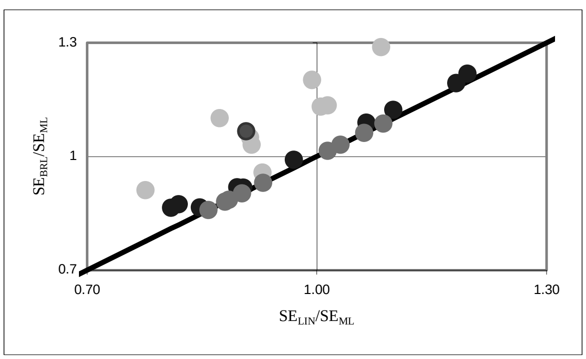

Figure 1. Ratio of SE_{BRL} to SE_{ML} vs. SE_{LIN} to SE_{ML} for coefficients of model for appropriate care, intervention (red), other cluster-level variables (pink), demographics (blue), and baseline health (brown).

Receipt of appropriate care was essentially uncorrelated for patients from the same clinic. Using the GEE method of Liang and Zeger (1986), we estimate the intra-clinic correlation of the errors as –0.0014, easily consistent with a true value of 0. Therefore the maximum likelihood (ML) based standard errors, which are precise for a sample of this size, should also be accurate and there is no reason to expect any of the correct standard errors to fall much below those obtained from logistic regression. However, the linearization standard errors are less than the ML standard errors for 18 of the 29 coefficients and 7 of the 10 coefficients for clusters-level variables. This can be seen in Figure 1. The horizontal axis plots the ratio of linearization to ML standard errors and many of the points are to the left of the vertical line at 1.00, where the ML and linearization estimates are equal. Also, there is considerable variability in the linearization estimators, which is apparent in the figure in the range of the ratios from about 0.8 to 1.2.

The BRL estimator performs much better than the traditional linearization estimator. The ratio of BRL to ML estimators is plotted along the vertical axis of Figure 1 and 8 of 10 of the pink dots are above the horizontal line at 1.00. All points are above the 45° indicating that the BRL estimates exceed their linearization counterparts for every coefficient.

REFERENCES

Bell R.M. and McCaffrey, D.F., (2002) "Bias Reduction in Standard Errors for Linear Regression with Multi-Stage Samples," *Survey Methodology*, forthcoming.

Binder, D. (1983), "On the Variance of Asymptotically Normal Estimators from Complex Surveys," *International Statistical Review*, 51, pp. 279-292.

Huzurbazar, S. (1999), "Practical Saddlepoint Approximations," *American Statistician*, 53, pp. 225-232.

Kott, P. S. (1994), "A Hypothesis Test of Linear Regression Coefficients with Survey Data," *Survey Methodology*, 20, pp. 159-64.

Kott, P. S. (1996), "Linear Regression in the Face of Specification Error: Model-Based Exploration of Randomization-Based Techniques," *Statistical Society of Canada Proceedings of the Survey Methods Section*, pp. 39-47.

Liang, K-,Y., and Zeger, S. L. (1986) "Longitudinal Data Analysis Using Generalized Linear Model," *Biometrika*, 73, pp. 13-22.

Mancl, L. A. and DeRouen T. A. (2001) "A Covariance Estimator for GEE with Improved Small-Sample Properties," *Biometrics*, 57, pp. 126-134.

McCaffrey, D. F., Bell R. M. and Botts, C. H. (2001), "Generalizations of Bias Reduced Linearization," *Proceeding of the Annual Meeting of the American Statistical Association, August 5-9, 2001*.

Pan, W. and Wall, M. (2001) "Small-sample Adjustments in Using the Sandwich Variance Estimator in Generalized Estimating Equations," *Statistics in Medicine*, 21, pp. 1429-1441.

Rust, K. F., and Rao, J. N. K. (1996), "Variance Estimation for Complex Surveys Using Replication Techniques," *Statistical Methods in Medical Research*, 5, pp. 283-310.

Satterthwaite, F. (1946), "An Approximate Distribution of Estimates of Variance Components," *Biometrics*, 2, pp. 110-114.

Shah, B. V., Holt, M. M., and Folsom, R. E., (1977), "Inference About Regression Models from Survey Data," *Bulletin of the International Statistical Institute*, 41, pp. 43-57.

Skinner, C. J., (1989), "Domain Means, Regression and Multivariate Analyses," in *Analysis of Complex Surveys*, eds. C. J. Skinner, D. Holt, and T. M. F. Smith, New York: Wiley, pp. 59-88.

Wells, K. B, Sherbourne, C., Schoenbaum, M., Duan, N., Meredith, L., Unutzer, J., Miranda, J., Carney, M., and Rubenstein, L. V., (2000), "Impact of Disseminating Quality Improvement Programs for Depression in Managed Primary Care: A Randomized Controlled Trial," *JAMA*, 283, pp. 212-220.