Proceedings of the Annual Meeting of the American Statistical Association, August 5-9, 2001

GENERALIZATIONS OF BIASED REDUCED LINEARIZATION

Daniel F. McCaffrey, RAND
Robert M. Bell, AT&T Labs Research
Carsten H. Botts, Iowa State University

1. INTRODUCTION

Linearization (Skinner 1989) is a nonparametric method for estimating the standard errors of design-based statistics such as means and ratios as well as coefficients from linear and nonlinear regression models. Although the traditional linearization estimator for standard errors is consistent as the number of primary sampling units (PSUs) grows, 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 2000; Kott 1994; Murray *et al*. 1998).

Bell and McCaffrey (2000) developed biased reduced linearization (BRL) to eliminate or reduce this bias for linear regression models with unweighted data from nonstratified two-stage samples. Reduction in bias is achieved by replacing the ordinary residuals used in the standard linearization estimator by residuals adjusted to better approximate the joint distribution of the true errors.

In this paper, we extend the BRL method to weighted regression analyses. The method handles a variety of different types of weights, including:

- design weights equal to the inverse of the sample selection probability;
- weights that account for post-stratification, nonresponse, and other weighting adjustments (e.g., for multiplicity) provided the weights can be treated as known;
- diagonal or nondiagonal precisions weights to account for heteroskedastic or correlated errors;
- logistic regression and other generalized linear models that can be fit by iteratively reweighted least squares; and
- generalized estimating equations.

We discuss four alternative BRL specifications and investigate the performance (bias and variance) of these estimators and commonly used alternative via simulation. We also present an application of logistic regression used to estimate the treatment effect in a cluster-randomized experiment. The application demonstrates a natural extension of BRL to models where parameters are estimated by iteratively reweighted least squares.

2. BIAS REDUCED LINEARIZATION FOR WEIGHTED LEAST SQUARES

2.1 General Method

Throughout the paper we restrict attention to two-stage nonstratified samples. Let n equal the number of PSUs and m_i equal the number of final sampling units from the i-th PSU, $i = 1,\dots, n$. The overall sample size is $M = \sum_i m_i$. We assume $y_{ij} = \beta' x_{ij} + \varepsilon_{ij}$, where the vector of errors, ε, 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—i.e., that $\mathbf{V}$ is block diagonal, with $m_i \times m_i$ blocks $\mathbf{V}_i$ for $i = 1,\dots, n$. Furthermore we assume there exists a known diagonal or block diagonal matrix of weights $\mathbf{W}$.

When weights are not constant, the coefficients can be estimated using the weighted least squares estimator $\hat{\beta} = \mathbf{QX}'\mathbf{WY}$, where $\mathbf{Q} = (\mathbf{X}'\mathbf{WX})^{-1}$. To simplify presentation, we generally discuss a linear combination of the regression coefficients, $l'\hat{\beta}$, for an arbitrary column vector l. If the errors are uncorrelated across PSUs, then the variance of $l'\hat{\beta}$ is

$$Var(l'\hat{\beta}) = l'\mathbf{Q}\left(\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{W}_i\mathbf{V}_i\mathbf{W}_i\mathbf{X}_i\right)\mathbf{Q}l . \qquad (1)$$

The standard linearization estimator (Skinner 1989, Kott 1994) is

$$v_{LIN} = l'\mathbf{Q}\left(c\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{W}_i\mathbf{r}_i\mathbf{r}_i'\mathbf{W}_i\mathbf{X}_i\right)\mathbf{Q}l . \qquad (2)$$

where $\mathbf{r}_i = \mathbf{y}_i - \mathbf{X}_i\hat{\beta}$ and $c = n/(n-1)$. The ratio of v_{LIN} to the $Var(l'\hat{\beta})$ converges in probability to 1 as the number of PSU's grows large under some general assumptions about the design matrix $\mathbf{X}$, the weights and the errors. However, as noted by Kott (1994), the estimates can be biased when the number of PSUs is small because $E\mathbf{r}_i\mathbf{r}_i'$ does not equal $(1/c)\mathbf{V}_i$.

Following our previous work with unweighted least squares (Bell and McCaffrey 2000), we consider the class of bias reduced linearization (BRL) variance estimators.

Theorem 1. For a specified block-diagonal target or working covariance matrix $\mathbf{U}$ and a specified weight matrix $\mathbf{W}$, consider the class of estimators

$$v_{WL} = l'\mathbf{Q}\left(\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{A}_i\mathbf{r}_i\mathbf{r}_i'\mathbf{A}_i'\mathbf{X}_i\right)\mathbf{Q}l \qquad (3)$$

where $\mathbf{A}_i$, satisfies

$$\mathbf{A}_i[(\mathbf{I} - \mathbf{G})_i\mathbf{U}(\mathbf{I} - \mathbf{G})_i']\mathbf{A}_i' = \mathbf{U}_i \qquad (4)$$

for $i = 1,\ldots, n$ and $\mathbf{G} = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})\mathbf{X}'\mathbf{W}$ and $(\mathbf{I} - \mathbf{G})_i$ denoting the rows of $(\mathbf{I} - \mathbf{G})$ corresponding to the observations from the i-th PSU. If $\mathbf{V} = k\mathbf{U}$ for some scalar k, then $E(v_{WL}) = Var(l'\hat{\beta})$.

Proof. If $\mathbf{V} = k\mathbf{U}$, then $E\mathbf{r}_i\mathbf{r}_i' = k(\mathbf{I} - \mathbf{G})_i\mathbf{U}(\mathbf{I} - \mathbf{G})_i'$ and the result follows.

For $m_i > 1$, $\mathbf{A}_i$ is not unique. If $\mathbf{A}_i$ satisfies (4), then so does $\mathbf{U}_i^{1/2}\mathbf{O}\mathbf{U}_i^{-1/2}\mathbf{A}_i$ where $\mathbf{U}_i^{1/2}\mathbf{U}_i^{1/2\prime} = \mathbf{U}_i$ and $\mathbf{O}$ is any $m_i \times m_i$ orthogonal matrix. If $\mathbf{V} = k\mathbf{U}$, the choice of $\mathbf{A}_i$ is unimportant because any solution to (4) will produce an unbiased variance estimator. However, the resulting estimators are biased when $\mathbf{V} \neq k\mathbf{U}$, and the bias can vary greatly with the choice of $\mathbf{A}_i$. Bell and McCaffrey (2000) found that for OLS estimates, the symmetric solution minimized squared error between the adjusted residuals and the true errors when the errors were *i.i.d*, and that the symmetric solution also tended to yield the smallest bias when the errors were not independent. Thus symmetric solutions to (4) are of particular interest.

We obtain a symmetric solution to (4) as $\mathbf{A}_i = \mathbf{U}_i^{1/2}\mathbf{P}\Lambda^{-1/2}\mathbf{P}'\mathbf{U}_i^{1/2}$ where $\mathbf{P}\Lambda\mathbf{P}'$ is the eigen decomposition of $[\mathbf{U}_i^{1/2}(\mathbf{I} - \mathbf{G})_i\mathbf{U}(\mathbf{I} - \mathbf{G})_i'\mathbf{U}_i^{1/2\prime}]$. The solution is invariant to choice of $\mathbf{U}_i^{1/2}$, $\mathbf{U}_i^{1/2}\mathbf{U}_i^{1/2\prime} = \mathbf{U}_i$. We refer to the estimator that obtains from (3) with the symmetric solution to (4) as $v_{BRL,S1}$. We will also explore the properties of the estimators that use the asymmetric solution to (4) $\mathbf{A}_i = \mathbf{U}_i^{1/2}[(\mathbf{I} - \mathbf{G})_i\mathbf{U}(\mathbf{I} - \mathbf{G})_i']^{-1/2}$ where the Cholesky roots are used for $\mathbf{U}_i$ and $(\mathbf{I} - \mathbf{G})_i\mathbf{U}(\mathbf{I} - \mathbf{G})_i'$. We refer to the resulting estimator as $v_{BRL,A1}$.

We also consider an alternative class of estimators

$$v_{WL^*} = l'\mathbf{Q}\times \left(\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{W}_i^{1/2}\mathbf{B}_i\mathbf{W}_i^{1/2}\mathbf{r}_i\mathbf{r}'\mathbf{W}_i^{1/2}\mathbf{B}_i'\mathbf{W}_i^{1/2}\mathbf{X}_i\right)\times\mathbf{Q}l \qquad (5)$$

where $\mathbf{B}_i$, satisfies

$$\mathbf{B}_i[\mathbf{W}_i^{1/2}(\mathbf{I} - \mathbf{G})_i\mathbf{U}(\mathbf{I} - \mathbf{G})_i'\mathbf{W}_i^{1/2}]\mathbf{B}_i' = \mathbf{W}_i^{1/2}\mathbf{U}_i\mathbf{W}_i^{1/2} \qquad (6)$$

Again, if $\mathbf{V} = k\mathbf{U}$, then $E(v_{WL^*}) = Var(l'\hat{\beta})$. If $\mathbf{U}_i^{1/2}\mathbf{W}_i^{1/2}[\mathbf{W}_i^{1/2}(\mathbf{I} - \mathbf{Q})_i\mathbf{U}(\mathbf{I} - \mathbf{Q})_i'\mathbf{W}_i^{1/2\prime}]\mathbf{W}_i^{1/2\prime}\mathbf{U}_i^{1/2\prime} = \mathbf{P}\Lambda\mathbf{P}'$, then $\mathbf{B}_i = \mathbf{W}_i^{1/2}\mathbf{U}_i^{1/2}\mathbf{P}\Lambda^{-1/2}\mathbf{P}'\mathbf{U}_i^{1/2\prime}\mathbf{W}_i^{1/2\prime}$ is a symmetric solution to (6) for any roots $\mathbf{W}_i^{1/2}$ and $\mathbf{U}_i^{1/2}$. The estimator derived from (5) with the symmetric solution to (6) is $v_{BRL,S2}$. An asymmetric solution is $\mathbf{W}_i^{1/2}\mathbf{U}_i^{1/2}[\mathbf{W}_i^{1/2}(\mathbf{I} - \mathbf{Q})_i\mathbf{U}(\mathbf{I} - \mathbf{Q})_i'\mathbf{W}_i^{1/2}]^{-1/2}$ where Cholesky roots are used for all matrices. We denote the estimator that results from using this asymmetric solution to (6) in (5) as $v_{BRL,A2}$.

2.2 Applications

The results of Section 2.1 are general and hold for any block diagonal weight matrix. However, several interesting special cases exist. When $\mathbf{W} = \mathbf{I}$ and $\mathbf{U} = \mathbf{I}$, we return to the case of OLS regression with the identity as the working covariance as considered by Bell and McCaffrey. Equations (3) and (5) are identical, so that $v_{BRL,S1} = v_{BRL,S2}$ and $v_{BRL,A1} = v_{BRL,A2}$. If $\mathbf{W} = \mathbf{I}$, but $\mathbf{U}$ is a specified, nonidentity, covariance matrix, then equations (3) and (5) are again identical and the symmetric solution to (4) generalizes the estimator of Bell and McCaffrey (2000) to general working covariance matrices.

If we had information for an estimate of $\mathbf{V}$, we might reduce the bias in BRL by using that information for the working $\mathbf{U}$ rather than assuming the identity. Alternatively, we might want to use that information to improve the properties of the estimates of β as well as the estimated standard errors by conducting precision weighted least squares (generalized least squares) to account for known heteroskedasticity or correlation of errors within PSUs. In this case $\mathbf{W} = \mathbf{U}^{-1}$. The variance of linear combinations of the resulting coefficient estimates is still given by equation (4), which reduces to $l'(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}l$ when $\mathbf{W}^{-1} = \mathbf{V}$. Let $\mathbf{H}_{\mathrm{W}} = \mathbf{W}^{1/2}\mathbf{X}\mathbf{Q}\mathbf{X}'\mathbf{W}^{1/2}$. If $\mathbf{W} = \mathbf{U}^{-1}$, then $(\mathbf{I} - \mathbf{G})_i\mathbf{U}(\mathbf{I} - \mathbf{G})_i' = \mathbf{U}_i^{1/2}(\mathbf{I} - \mathbf{H}_{\mathrm{W}})_{ii}\mathbf{U}_i^{1/2}$ which simplifies equation (4). Similarly, $\mathbf{W}_i^{1/2}(\mathbf{I} - \mathbf{G})_i\mathbf{U}(\mathbf{I} - \mathbf{G})_i'\mathbf{W}_i^{1/2} = (\mathbf{I} - \mathbf{H}_{\mathrm{W}})_{ii}$ which greatly simplifies equation (6). If we let $\mathbf{X}^* = \mathbf{W}^{1/2}\mathbf{X}$ and $\mathbf{r}^* = \mathbf{W}^{1/2}\mathbf{r}$, then $v_{BRL,\mathrm{S2}}$ can be obtained by using the formulas for OLS given in Bell and McCaffrey by using $\mathbf{X}^*$ and $\mathbf{r}^*$. In fact, $v_{BRL,\mathrm{S2}}$ can be derived from formulas for OLS even when $\mathbf{W} \neq \mathbf{U}^{-1}$ provide we also replace $\mathbf{U}$ in the formulas with $\mathbf{U}^* = \mathbf{W}^{1/2}\mathbf{U}\mathbf{W}^{1/2}$ and use the appropriate matrix roots.

Weighted least squares that combine precision and design weighting will also be of interest. The estimators $v_{BRL,\mathrm{S1}}$, $v_{BRL,\mathrm{A1}}$, $v_{BRL,\mathrm{S2}}$, $v_{BRL,\mathrm{A2}}$ apply to this general case but equations (4) and (6) do not simplify.

3. SIMULATION METHODS

We use a Monte Carlo simulation to study the properties of alternative variance estimators for a balanced two-stage cluster sample with $n = 20$ PSUs and a constant $m = 10$ observations in each PSU. All simulation replications use the common design matrix $\mathbf{X}$ of Bell and McCaffrey (2000). The design matrix has four independent variables chosen to represent a range of difficulty for nonparametric variance estimators. The first two independent 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 (ICC) of 0.5 within PSU, while x_4 was generated from independent normal distributions. Observed intra-cluster correlations are 1.00, 1.00, 0.62 and -0.04, respectively. Observed correlations among the independent variables are all very small with the exception of $Corr(x_1, x_2) = 0.14$, $Corr(x_1, x_3) = 0.25$ and $Corr(x_1, x_4) = -0.11$.

The dependent variable was generated from the equation $y_{ij} = \beta' x_{ij} + \varepsilon_{ij}$, where $\beta = 0$ and the ε_i's are standard multivariate normals with ICC equal to ρ. We use four alternative values of $\rho = 0$, 1/9, 2/9, and 1/3, corresponding to design effects for the sample mean of $DEFF = 1$, 2, 3, and 4, ($DEFF=1+(m-1)\rho$). We generate a single design matrix for the simulation study.

We consider two types of weights. The first weights are unrelated to the variance of the errors and are analogous to design weights where the selection probability was not directly related to the variance of the dependent variable in the regression. For the results reported in this paper, the weights were uniform random variables between 1 and 3 and were independent of the covariates and the errors. Weights varied among units from the same PSU. We explored alternative weights such as weights that were constant within PSUs and weights that were correlated with either the errors or the covariates. The results were qualitatively similar and so we report only the results for the independent weights. For our design weighted estimates the working covariance matrix used in deriving the BRL estimator is the identity, i.e., $\mathbf{U} = \mathbf{I}$.

The second type of weights used in the simulation study were precision weights where $\mathbf{W} = \mathbf{U}^{-1}$, the inverse of the working covariance matrix for the errors. We consider cases where $\mathbf{U}$ equals the true covariance matrix, $\mathbf{V}$, and cases where it does not.

We used the simulation to study the bias and variance of the four alternative BRL estimators. We also study the properties of common alternatives to BRL. For the study of design weights, the alternative estimators we consider are the standard linearization estimator (v_{LIN}), the jackknife (Cochran 1977) and Kott's method (Kott 1994). For the precision weighted estimates, the common alternatives to BRL are the standard linearization estimator and the model based generalized least squares estimator.

The variability of the estimators is described by twice the reciprocal of the square of the coefficient of variation. We call this quantity the degrees-of-freedom (DF) of the estimator because it equals the degrees-of-freedom for the Satterthwaite approximation to the distribution of each estimator (Bell and McCaffrey 2000). Given the design matrix and weights, the exact bias and DF can be obtained analytically for all estimators other than Kott's. For Kott's method we use 100,000 simulated replicates of the error to estimate bias and DF.

4. SIMULATION STUDY RESULTS

We first describe the simulation results for the design weighted estimates and then for the precision weighted estimates.

4.1 Simulated Design Weights

By design, all four BRL estimators are unbiased when $\rho = 0$. Figure 1, compares the relative biases of the three BRL estimators when $\rho = 1/3$. The two asymmetric estimators are the same for this study. Because the errors have equal correlation within PSU, the ratio of the relative biases of any two BRL estimates is invariant to the value of ρ for $\rho > 0$. Thus, for all $\rho > 0$, the relative size of the bars would be the same as in Figure 1, although the scale would change; the magnitude of the bias increase with ρ. The relative bias in $v_{BRL,S1}$ is closer to zero than the relative bias in $v_{BRL,S2}$ for all covariates with positive ICC (intercept, x_1, x_2, x_3). For these variables, the ratios of the relative biases range from 1.3 to 2.1. For x_4 which has a very small negative ICC, the ratio of relative biases of $v_{BRL,S2}$ to $v_{BRL,S1}$ is 0.7. The relative bias of $v_{BRL,A1}$ is always further from zero than the relative bias in $v_{BRL,S1}$. The ratios of the relative biases range from 1.1 to 2.0. For the intercept and x_1, x_2, and x_3 the relative bias of $v_{BRL,A1,}$ is smaller than the relative bias of $v_{BRL,S2}$. For x_4 the relative bias of $v_{BRL,S2}$ is closer to zero than that of $v_{BRL,A1}$.

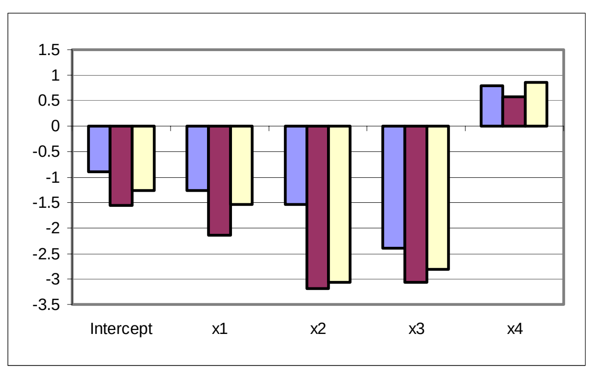

Figure 1. Relative biases of three BRL estimators, $v_{BRL,S1}$ (blue) $v_{BRL,S2}$ (red) and $v_{BRL,A1}$ (yellow), $\rho = 1/3$.

For each BRL estimator the DF vary among the covariates (for example for $v_{BRL,S2}$ the DF range from 2.8 for X_2 to 14.2 for X_4 for $\rho = 0$) and tend to decrease ρ increases (for example for $v_{BRL,S2}$ the DF for X_3 decrease from 11.9 for $\rho = 0$ to 9.3 for $\rho = 1/3$). Across the estimators the DF are very similar with the ratios of the DF ranging from 0.99 to 1.01.

While the bias in the BRL estimators is small (no more then 2.4 percent in absolute value) for all the covariates and for all values of ρ, the bias in the linearization estimators and the jackknife estimators can be large. The linearization estimators are always biased low and the bias ranges from –2.8% for X_4 to –33.0% for X_2 for $\rho = 0$. The bias is similar for other values of ρ. The bias in the jackknife estimators is always positive and similar in magnitude to the bias in the linearization estimators. Kott's methods produces estimator with small postive bias of less than five percent and often less than one percent. Kott's method also appears mostly invariant to the value of ρ. Figure 2a presents the relative bias for the various estimator when $\rho = 1/3$. The results for other values of ρ are similar with the exception that BRL is unbiased when $\rho = 0$. As shown in Figure 2b the values of DF are very similar across the estimators. These results are very similar to the OLS results of Bell and McCaffrey (2000).

4.2 Nondiagonal Precision Weighting

For each of the six estimators included in the study of nondiagonal precision weighted estimation (the four BRL alternatives, the model based GLS estimator and the traditional linearization estimator), Figure 3 plots relative bias versus working ICC, ρ_W, by relative bias for each of the four values of the true ICC, ρ_T, used in the study (0, 1/9, 2/9 and 3/9). The figure contains the relative bias for the estimator of the variance of the coefficient for x_1. Plots for the intercept and $x2$ are similar except that the relative bias tends to be larger for x_2. For x_3 and x_4 the performance of $v_{BRL,A1}$ is still poor but the differences between this and the other BRL estimators are much less pronounced. In addition the bias for x_4 has the opposite sign because of the negative ICC for x_4. The performance of the linearization and GLS estimators for x_3 and x_4 is qualitatively similar to the performance in the plots although the relative bias tends to be smaller and the relationship is between ρ_W is not linear. The plots for the other coefficients are available from the authors.

By design the relative bias of the four BRL estimators and the GLS estimator equals zero when $\rho_W = \rho_T$. When the $\rho_W > \rho_T$, then the bias in each of these five estimators is positive. When $\rho_W < \rho_T$, then bias is negative. For a given value of the ρ_T, the expected values of $v_{BRL,S1}$, $v_{BRL,S2}$, $v_{BRL,A2}$ and v_{GLS} are nearly linear in ρ_W as demonstrated by the nearly linear relationship between relative bias and ρ_W for each estimator. However, the slope of the line decreases as the ρ_T increases. Thus, relative bias is not symmetric with respect to the difference between ρ_W and ρ_T. Over estimation of the ICC results in relative bias that is further from zero than does underestimation of the ICC; however, underestimation of the ICC results in negative bias while overestimation results in positive bias.

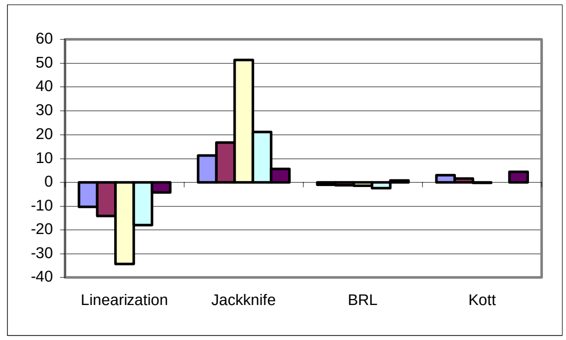

Figure 2a. Relative bias of alternative variance estimators for simulated design weights by covariate, ICC=1/3: intercept (dark blue), x_1 (red), x_2 (yellow), x_3 (light blue) and x_4 (black).

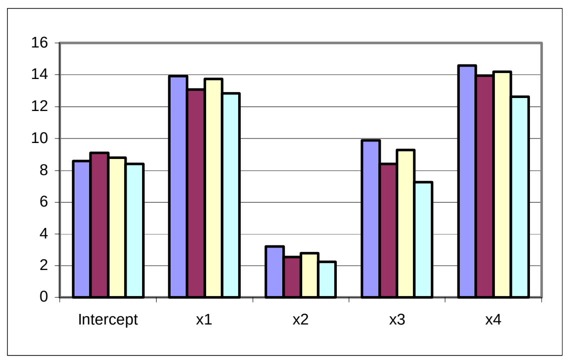

Figure 2b. DF for variance estimators for covariates for simulated design weights by estimator: linearizatioin (dark blue), jackknife (red), BRL,S1 (yellow), and Kott (light blue).

The relative bias of $v_{BRL,S1}$, $v_{BRL,S2}$ and $v_{BRL,A2}$ is small to moderate, less than 10 percent in absolute value for all values of ρ_W and ρ_T. The relative bias of $v_{BRL,S1}$ is nearly equal to that of $v_{BRL,S}$ and the relative bias of $v_{BRL,A2}$ is roughly 1.5 times as large when $\rho_W \neq \rho_T$. The relative bias in the v_{GLS} is very large in absolute value when the ρ_W deviates ρ_T by more than 1/9.

Unlike the other BRL alternatives, the relative bias $v_{BRL,A1}$ grows exponentially when the working $\rho_W > \rho_T$ resulting in extremely large relative bias when the $\rho_T = 0$ and the ρ_W equals2/9 or 1/3.

The linearization estimator always under-estimates the true variance. This is true for all coefficients and regardless of the values of the ρ_W and ρ_T.

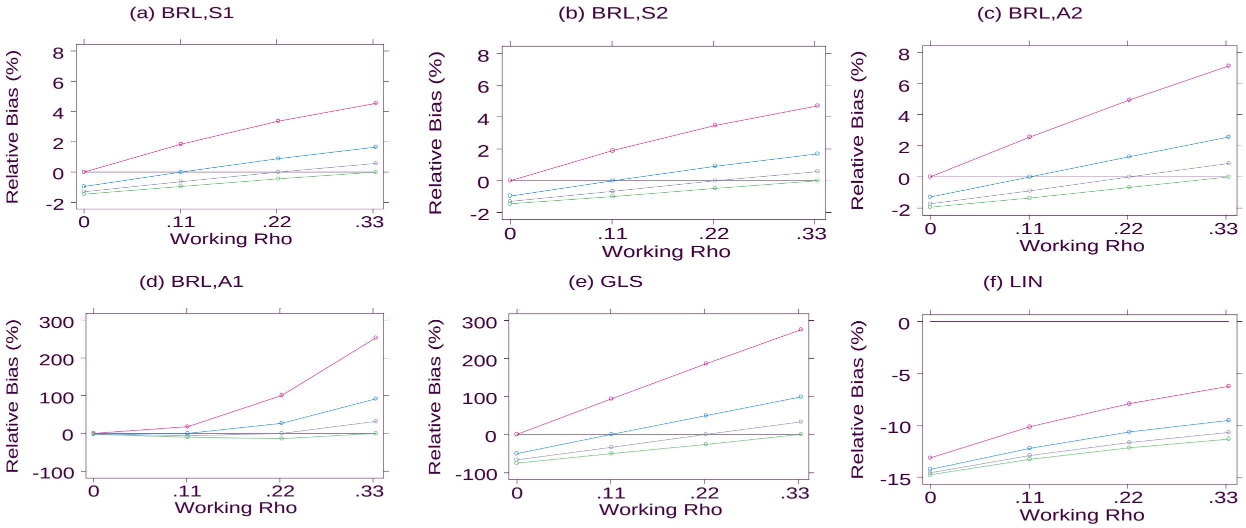

Figure 3. Relative bias versus values of ρ_W at four values of ρ_T: 0 (red), 1/9 (blue), 2/9 (purple) and 1/3 (green), for six estimator of the variance of the coefficient for x_1 estimated by GLS.

The DF are similar across the BRL estimators. For all the estimators the DF are smallest for x_2 (ranging from 2.8 to 4.4 across estimators and values of ρ_W and ρ_T) and largest for x_4 with values around for 16.4 for all conditions. The linearization estimator is relatively more variable for the PSU level estimators when the $\rho_W > 0$ and the DF are inversely related to ρ_W. For example when the $\rho_W = 1/3$ the DF for the linearization estimator for x_1 is only 20% as large as the DF for the BRL estimators. When $\rho_W = 1/9$ the DF for linearization is a little over 60% of the DF for the BRL estimators. Thus, for GLS, BRL not only reduces bias but it also increases the relative precision of the variance estimators.

5. 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, clinics 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. Binder (1983) suggested linearization standard errors for logistic regression. The estimated coefficients for logistic regression satisfy the equation:

$$\hat{\beta} = (\mathbf{X'WX})^{-1}\mathbf{X'Wz}, \qquad (8)$$

where the ij-th element of the outcome vector $\mathbf{z}$ is $z_{ij} = \mathbf{x}_{ij}'\hat{\beta} + (y_{ij} - p_{ij})/\{p_{ij}(1-p_{ij})\}$, for $p_{ij} = 1/(1 + e^{-\mathbf{x}_{ij}'\hat{\beta}})$ and $\mathbf{W}$ a diagonal matrix with $w_{ij} = p_{ij}(1-p_{ij})$. Estimates are analogous to coefficients for a weighted linear regression in the last iteration of iteratively reweighted least squares. Equation 8 suggests that the BLR method can be extended to logistic regression using the techniques for weighted least squares and using the residuals $\mathbf{r} = \mathbf{z} - \mathbf{X}\hat{\beta}$. Properties of this natural extension of BRL are less obvious and are currently being studied.

Figure 4 plots the ratio of BRL standard errors (SE_{BLR}) to the standard large sample approximation (SE_{ML}) versus the ratio of linearization standard errors (SE_{LIN}) to SE_{ML}. We use clinic as the PSU because there is very little reason to expect correlations of errors across clinics after controlling for block.

If outcomes are correlated within clinic, then we can expect the ratios to be greater than one for variables with a positive ICC—e.g., clinic-level variables. The ratio for BRL is greater than one for eight of the nine clinic-level variables but the linearization ratio is greater than one for only three and often falls far below one. Using the GEE method of Zeger and Liang (1986), we estimate the intra-clinic correlation of the errors as –0.0014, easily consistent with a true value of 0. Nonetheless, there is no reason to expect any of the correct standard errors to fall much below those obtained from logistic regression. For example, the linearization standard error for Intervention is only 91 percent as large as the asymptotic standard error from logistic regression.

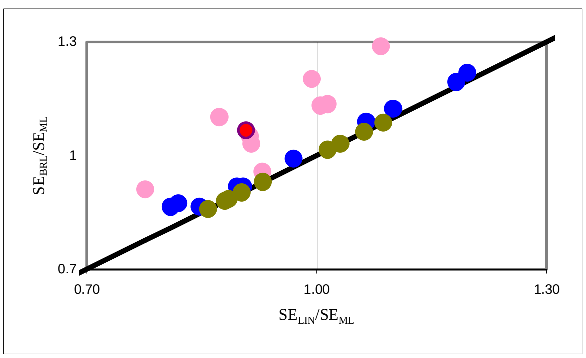

Figure 4. Ratio of SE_{BRL} to SE_{ML} vs. SE_{LIN} to SE_{ML} for coefficients of model for appropriate care, intervention (red), other cluster-level variables (pink), demographics (blue), and baseline health (brown).

6. SUMMARY

In this paper we extend the bias reduced linearization estimators of Bell and McCafrey (2000) for unweighted least squares to weighted estimators. We considered two classes of weighted estimators those where the weights equal the inverse of the working covariance matrix and those where the weights do not. The method is widely applicable, including generalized linear models. For both types of weights, we developed a class of estimators that is unbiased for estimating the variance of the coefficients if the working covariance matrix used in deriving the estimator equals the true covariance of the residual errors. Our simulation study results show that these estimators can have small relative bias even when the working covariance matrix is incorrect. The simulation study also shows that using a symmetric matrix to transform the residuals used in the variance estimator is necessary to obtain the reduced bias for the design matrix and error distributions included in our study.

REFERENCES

Bell R.M. and McCaffrey, D.F., (2000) *Bias Reduction in Standard Errors for Linear Regression with Multi-Stage Samples*, AT&T Technical Report TD-4S9H9T.

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

Cochran, W. G. (1977), *Sampling Techniques*, *Third Edition*, New York: Wiley.

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

Murray, D. M., Hannan, P. J., Wolfinger, R. D., Baker, W. L., and Dwyer, J. H. (1998), "Analysis of Data from Group-Randomized Trials with Repeat Observations on the Same Groups," *Statistics in Medicine*, 17, 1581-1600.

Skinner, C. J., (1989b), "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, 212-220.

Zeger, S. L., and Liang, K. Y. (1986), "Longitudinal Data Analysis for Discrete and Continuous Outcomes," *Biometrics*, 42, 121-130