WEIGHT TRIMMING IN A RANDOM EFFECTS MODEL FRAMEWORK

Michael R. Elliott, Roderick J.A. Little,
University of Michigan
Michael R. Elliott, Department of Biostatistics, University of Michigan,
1420 Washington Heights, Ann Arbor, MI 48109 (Email: mrelliot@umich.edu).

Key Words: sample survey inference, sampling weights, unit nonresponse adjustments, random-effects models, non-parametric regression.

Abstract:

In sample surveys with unequal probabilities of inclusion, units are often weighted by the inverse of the probability of inclusion to avoid biased estimates of population quantities such as means. Highly disproportional sample designs yield large weights, which can result in weighted estimates that have a high variance. Weight trimming reduces large weights to a fixed cutpoint value and adjusts weights below this value to maintain the untrimmed weight sum. This approach reduces variance at the cost of introducing some bias. An alternative approach uses random-effects models to induce shrinkage across weight strata. We compare these two approaches, and introduce extensions of each: a compound weight pooling model that allows Bayesian averaging over estimators based on different trimming points, and a weight smoothing model based on a non-parametric spline function for the underlying weight stratum means. The latter method performs well in simulations when compared with alternative estimators.

1 Weight Pooling Models

Many survey samples yields units that have unequal probabilities of inclusion. In these settings, estimators like sample means that assign the included units the same weight are biased when there is an association between the probability of inclusion and the values of the sampled data. Unit i is usually weighted by the inverse of the probability of inclusion to remove this bias. For example, the Horvitz-Thompson estimator of a population total $T = \sum_{i=1}^{N} y_i$ from a sample is given by $\hat{T} = \sum_{i \in s} w_i y_i$, where $w_i = 1/\pi_i$, π_i is the probability of inclusion and s is the subset of the population units sampled. Throughout this paper, we assume the primary quantity of interest is the finite population mean $\overline{Y}$. Our models also assume normally-distributed data, although extensions to non-normal distributions are possible.

Suppose the population can be divided into H weight strata by the set of *ordered* distinct values of the weights w_h. Let n_h be the number of included units and N_h the population size in weight stratum h, so that $w_h = N_h/n_h$ for $h = 1, \ldots, H$. We assume here that N_h is known, as when the weight strata come from a stratified or post-stratified random sample. The untrimmed (design-based) weighted mean estimator is then $\overline{y}_w = \frac{\sum_h \sum_i w_h y_{hi}}{\sum_h \sum_i w_h} = \sum_h (N_h/N_+)\overline{y}_h$.

The weighted mean estimator $\overline{y}_w$, while unbiased, has greater variance than the unweighted mean estimator $\overline{y}$. This increase can overwhelm the reduction in bias, so that the mean square error actually increases under a weighted analysis. This is particularly likely when the weights are highly variable, when the association between the probability of inclusion and the data is weak, or when the sample size is small. Perhaps the most common approach to dealing with this problem is *weight trimming* (Potter 1990, Kish 1992). Weight trimming typically proceeds by establishing an *a priori* cutpoint, say 3 for the normalized weights, and multiplying the remaining weights by a normalizing constant $\gamma = (n - \sum \kappa_i w_o)/\sum(1-\kappa_i)w_i$, where κ_i is an indicator variable for whether or not $w_i \geq w_0$. The trimmed mean estimator is thus given by

$$\overline{y}_{wt} = \sum_{h=1}^{l-1} \gamma N_h/N_+ \overline{y}_h + \sum_{h=l}^{H} w_0 n_h/N_+ \overline{y}_h = \quad (1)$$

$$\gamma \sum_{h=1}^{l-1} N_h/N_+ \overline{y}_h + \frac{w_0 \sum_{h=l}^{H} n_h}{N_+} \overline{y}^{(l)}$$

where $\gamma = (N_+ - w_0 \sum_{h=l}^{H} n_h)/(\sum_{h=1}^{l-1} N_h)$ and $\overline{y}^{(l)} = (1/\sum_{h=l}^{H} n_h) \sum_{h=l}^{H} n_h \overline{y}_h$. The choice of cut-

This research is supported by National Science Foundation Grant DMS-9803720, and by Bureau of the Census Contract Number 50-YABC-7-66020.

point w_0 is often ad-hoc. Potter (1990) discusses systematic methods for choosing w_0.

Weight trimming effectively pools units with high weights by assigning them a common, trimmed weight. The choice of $w_0 = \frac{\sum_{h=l}^{H} N_h}{\sum_{h=l}^{H} n_h}$ yields $\gamma = 1$ and $\overline{y}_{wt} = \sum_{h=1}^{l-1}(N_h/N_+)\overline{y}_h + (\sum_{h=l}^{H} N_h)\overline{y}^{(l)}/N_+$, which corresponds to the estimate for a model that assumes distinct stratum means for the smaller weight strata and a common mean for the larger weight strata, that is:

$$y_{hi} \mid \mu_h \sim N(\mu_h, \sigma^2) \quad h < l$$
$$y_{hi} \mid \mu_l \sim N(\mu_l, \sigma^2) \quad h \geq l \tag{2}$$
$$\mu_h, \mu_l \propto \text{const.}$$

We call (2) the simple weight pooling model.

Equation (2) can be extended by treating the pooling level l as a realization of the random variable L with support $(1, \ldots, H)$. Assuming that the location of pooling level L is *a priori* equally likely across the H weight strata, we obtain the compound weight pooling model:

$$y_{hi} \mid \mu_h \sim N(\mu_h, \sigma^2) \quad h < l$$
$$y_{hi} \mid \mu_l \sim N(\mu_l, \sigma^2) \quad h \geq l \tag{3}$$
$$p(L = l) = 1/H$$

Then

$$E(\overline{Y} \mid \boldsymbol{y}) = E(E(\overline{Y} \mid \boldsymbol{y}, l)) =$$
$$\sum_{l=1}^{H} \left(\sum_{h=1}^{l-1} \frac{N_h}{N_+}\overline{y}_h + \frac{\sum_{h=l}^{H} N_h}{N_+}\overline{y}^{(l)} \right) p(L = l \mid \boldsymbol{y}) \tag{4}$$

That is, pooling is conducted at every possible level and the weighted average computed, where the weighting is based on the posterior probability for the model that pools from the lth stratum onward.

To derive $p(L = l \mid \boldsymbol{y})$, we note that (3) is a special case of a Bayesian variable selection problem (Halpern 1973, Atkinson 1978) with $y \mid \beta_l, l, \sigma^2 \sim N(X_l\beta_l, \sigma^2 I)$, where X_l is an $n \times l$ matrix consisting of an intercept and dummy variables for each of the first $l-1$ weight strata. Utilizing priors of the form $p(\sigma^2 \mid l) = (1/\sigma^2)^{l/2+1}$ (Dempster et al. 1977) and $p(\boldsymbol{\beta}_l \mid l) = (2\pi)^{-l}$ (Halpern 1973) yields

$$p(L = l \mid \boldsymbol{y}) = \frac{\left[\prod_{h=1}^{l-1} n_h (\sum_{h=l}^{H} n_h)\right]^{-1/2} Q_l^{-n}}{\sum_{l=1}^{H} \left[\left[\prod_{h=1}^{l-1} n_h (\sum_{h=l}^{H} n_h)\right]^{-1/2} Q_l^{-n}\right]} \tag{5}$$

2 Weight Smoothing Models

Instead of mimicking the idea of weight trimming, we can simply model the weight-stratum means directly as random effects. The general form of the weight smoothing models we consider is

$$y_{hi} \mid \mu_h \overset{ind}{\sim} N(\mu_h, \sigma^2) \tag{6}$$
$$\boldsymbol{\mu} \sim N_H(\boldsymbol{\phi}, D)$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_H)$, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_H)$, $\boldsymbol{\phi}$, D, and σ^2 all have non-informative priors, and h indexes the "weight strata", with constant inclusion probabilities. Unlike the weight pooling models, there is no need for the weight strata to be ordered by inclusion probability; a more natural ordering may be used if available, e.g., if the weight strata represent a disproportionately stratified sample by age. Under the model (6),

$$E(\overline{Y} \mid \boldsymbol{y}) = \sum_h [n_h\overline{y}_h + (N_h - n_h)\hat{\phi}_h]/N_+ \tag{7}$$

where $\hat{\phi}_h = E(\overline{Y}_h \mid \boldsymbol{y}) = E(\phi_h \mid \boldsymbol{y})$.

The unweighted and fully weighted means are obtained as estimators of $E(\overline{Y} \mid \boldsymbol{y})$ as $D \to 0$ and $D \to \infty$, respectively. We consider four other special cases of the model:

Exchangeable random effects (XRE): (Holt and Smith 1979, Ghosh and Meeden 1986, Little 1991, Lazzaroni and Little 1998)

$$\phi_h = \mu \text{ for all } h, D = \tau^2 I_H \tag{8}$$

Autoregressive (AR1) (Lazzaroni and Little 1998)

$$\phi_h = \mu \text{ for all } h, D = \tau^2\{\rho^{|i-j|}\}, \tag{9}$$
$$i, j \in (1, \ldots, H)$$

Linear (LIN): (Lazzaroni and Little 1998)

$$\phi_h = \alpha + \beta h, D = \tau^2 I_H \tag{10}$$

Nonparametric (NPAR):

$$\phi_h = f(h), D = 0, \tag{11}$$

where $f(h)$ is a twice differentiable smooth function of h, where f and f' are absolutely continuous, $\int [f''(u)]^2 du < \infty$, and $f(h)$ minimizes the residual sum of squares plus a roughness penalty parameterized by λ:

$$\sum_h \sum_i (y_{hi} - f(h))^2 + \lambda \int (f^{(2)}(u))^2 du \tag{12}$$

(Wahba 1978, Hastie and Tibshirani 1990).

All of these models can be written in the mixed-effect form (Laird and Ware 1982)

$$\boldsymbol{y} = NX\boldsymbol{\beta} + NZ\boldsymbol{u} + \boldsymbol{\epsilon} \tag{13}$$

where N is an $H \times n$ "incidence" matrix relating the distinct weight strata to the data ($n_{jk} = 1$ if y_j is in stratum k and 0 otherwise), X is an $H \times p$ fixed-effect design matrix, $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed-effect parameters, Z is an $H \times q$ random-effect design matrix, $\boldsymbol{u} \sim N_q(0, G)$, and $\boldsymbol{\epsilon} \sim N(0, \sigma^2 I_n)$. The replacements for X, Z and G in (13) are obvious for (8), (9), and (10); for (11), the NPAR model, $X = (\boldsymbol{1}\ \boldsymbol{h}); Z_{H\times(H-1)}$ such that $ZZ' = \Omega$ where $\Omega_{hk} = \int_0^1((h-1)/(H-1)-t)_+((k-1)/(H-1)-t)_+dt$, $(x)_+ = x$ if $x \geq 0$ and $(x)_+ = 0$ if $x < 0$, $h, k = 1, \ldots, H$; and $G = (\sigma^2/H\lambda)I_{H-1}$ where λ is the penalty parameter in (12) (Speed, in discussion of Robinson 1991).

Under these formulations,

$$\hat{\boldsymbol{\phi}} = X\hat{\boldsymbol{\beta}} + Z\hat{\boldsymbol{u}}, \tag{14}$$

where $\hat{\boldsymbol{\beta}} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\overline{\boldsymbol{y}}$ and $\hat{\boldsymbol{u}} = \hat{G}Z'\hat{V}^{-1}(\overline{\boldsymbol{y}} - X\hat{\boldsymbol{\beta}})$. Here $V = ZGZ' + \sigma^2\Sigma$, where $\Sigma = diag(1/n_h)$, and $\overline{\boldsymbol{y}} = (\overline{y}_1, \ldots, \overline{y}_H)'$. In the case of XRE, AR1, and LIN, estimates of G and σ^2, and thus of $\boldsymbol{\beta}$ and $\boldsymbol{u}$, may be obtained by maximum likelihood (ML) or restricted maximum likelihood (REML) methods. In NPAR the REML likelihood given by model (13) and the likelihood given by (11) differ by only a constant when $\hat{f}(h) = \boldsymbol{X}'_h\hat{\boldsymbol{\beta}} + \boldsymbol{Z}'_h\hat{\boldsymbol{u}}$ is a natural cubic spline with knots at $(1, \ldots, H)$ (Wahba 1985, Green 1987); hence we utilize ML estimates for the XRE, AR1, and LIN models and REML estimates for the NPAR models to obtain mean and variance component estimates and thus $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{u}}$.

These weight smoothing models allow compromises between weighted and unweighted estimates. As an example, note that, under the XRE model, $\widehat{\phi_h} = w_h\overline{y}_h + (1 - w_h)\tilde{y}$, where $w_h = \frac{\tau^2 n_h}{\tau^2 n_h + \sigma^2}$ and $\tilde{y}$ is an overall weighted mean given by $\left(\sum_h \frac{n_h}{n_h\tau^2+\sigma^2}\right)^{-1} \sum_h \frac{n_h}{n_h\tau^2+\sigma^2}\overline{y}_h$. As $\tau^2 \to \infty$, $w_h \to 1$ so that $\widehat{\phi_h} \to \overline{y}_h$. Thus a flat prior for μ_h recovers the fully-weighted estimator, which can then be viewed as a fixed effects ANOVA model. On the other hand, as $\tau^2 \to 0$, $w_h \to 0$ so that $\widehat{\phi_h} \to \tilde{y}\ |_{\tau^2=0} = \overline{y}$, which estimates the excluded units at the pooled mean since the model now assumes that y_{hi} are drawn from a common mean. The AR1 model extends the XRE model by adding autocorrelation in the variance structure, allowing more borrowing of strength from strata that are close than from strata that are distant. This feature should somewhat degrade performance when the means are truly exchangeable but provide increased protection again model failure. Similarly for NPAR, $\lambda \to 0$ implies $\overline{y}_{NPAR} \to \overline{y}_w$ and $\lambda \to \infty$ implies $\overline{y}_{NPAR} \to \overline{y}_{linear}\ |_{\tau^2=0}$, so NPAR should be somewhat less effective than LIN when the means are linear, but should have reduced bias when the mean structure is nonlinear.

3 Simulation Study

A population of $N = 36,000$ was constructed consisting of 10 strata, where N_h=(800, 1000, 1200, 1500, 2000, 3000, 4000, 5000, 7500, 10000). The values of Y_{hi} were generated as

$$y_{hi} = \delta_h + \epsilon_{hi}$$

where

$\boldsymbol{\delta}^C$ = (22.5, 14.4, 9.0, 4.8, 1.8, -1.2, -1.8, -2.16, -1.92, -1.8).

$\boldsymbol{\delta}^D$ = (-1.8, -1.92, -2.16, -1.8, -1.2, 1.8, 4.8, 9.0, 14.4, 22.5).

$\boldsymbol{\delta}^E$ = (10.88, 10.88, 10.88, 10.88, 10.88, 10.88, 10.88, 10.88, 10.88, 10.88).

$\boldsymbol{\delta}^L$ = (-12.09, -8.64, -5.20, -1.75, 1.70, 5.14, 8.59, 12.03, 15.48, 18.93).

and $\epsilon_{hi} \sim N(0, \sigma^2)$. Disproportional samples of size 500 (90, 80, 70, 60, 50, 50, 40, 30, 20, 10) were then drawn (maximum normalized weight=13.9).

The mean structures $\boldsymbol{\delta}^C$ and $\boldsymbol{\delta}^D$ are best- and worst-case scenarios for weight trimming, with C = close and D = distant means in the high-weight strata. The E=equal mean structure $\boldsymbol{\delta}^E$ provides the best-case scenario for the XRE and AR1 models, and the L = linear structure $\boldsymbol{\delta}^L$ provides the best-case scenario the LIN model; parameters are chosen so that $E(\overline{Y}|\boldsymbol{\delta}^D) = E(\overline{Y}|\boldsymbol{\delta}^E) = E(\overline{Y}|\boldsymbol{\delta}^L)$.

The two primary outcomes of interest are root mean squared error (RMSE) and coverage of nominal 90% confidence intervals. RMSE is estimated as $\sqrt{\frac{1}{200}\sum_{i=1}^{200}(\hat{\theta}_i - \theta)^2}$, where θ is the population mean and $\hat{\theta}_i$ is the estimate from the ith of the 200 samples. We leave out the details of variance calculations required for confidence interval coverage because of space limitations.

3.1 Stratum Pooling Methods

For the mean structures $\boldsymbol{\delta}^C$ and $\boldsymbol{\delta}^D$, we considered estimates from five weighting schemes: unweighted

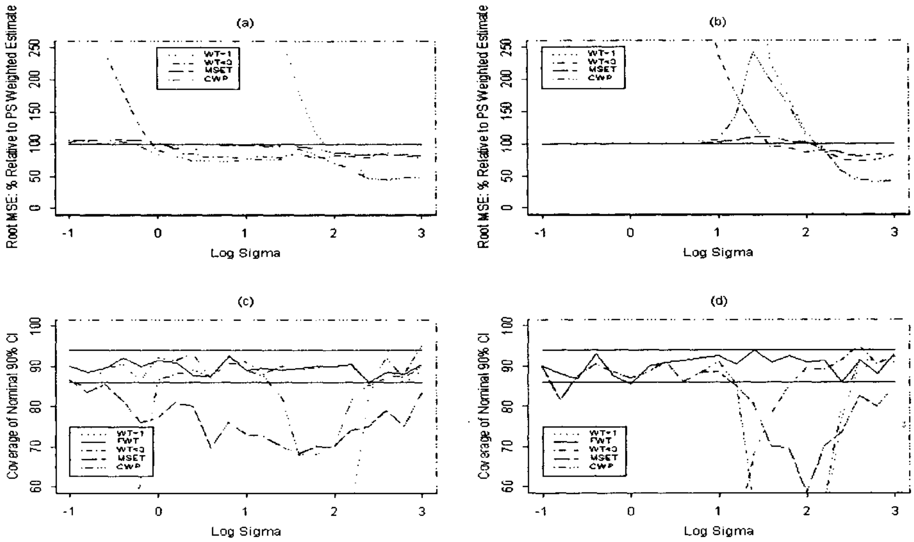

Figure 1: Average RSME (200 simulations) relative to fully weighted estimator of unweighted (WT=1), crude trimming (WT<3), minimum MSE (MSET), and compound weight pooling (CWP) estimators, and 90% nominal coverage of all estimators, when stratum means favor trimming [(a) and (c)], and when stratum means do <u>not</u> favor trimming [(b) and (d)]

(WT=1); fully weighted, $\bar{y}_w$ (FWT); "crude" trimming, $\bar{y}_{wt}$, where the maximum normalized weight is 3 (WT<3); the posterior mean from the compound weight pooling model given by (4) and (5) (CWP); and the MSE trimming method (MSET). The MSE trimming procedure (Potter 1990, Cox and McGrath 1981) estimates MSE at a variety of trimming levels and chooses the one at which $\widehat{MSE}$ is minimized. Here the trimming points are selected from all possible pooling values (in these simulations described, 10 distinct cutpoints are possible).

The left panels of Figure 1 compares the RMSE ratio of the unweighted and three other trimmed estimators to the fully weighted estimator [panel (a)] and nominal 90% coverage of all five estimators [panel (c)], for the stratum mean configuration $\boldsymbol{\delta}^C$ that favors pooling of the high-weight strata. The unweighted estimator (WT=1) is poor when $\log \sigma < 2$ and some form of weighting is need to counteract bias. "Crude" weight trimming (WT<3) is an improvement, but fails badly when $\log \sigma < 0$. The compound weight pooling estimator (CWP) works well in this setting: when σ is small and weighting is needed to counteract bias, CWP mimics the fully weighted estimator, and when σ is large and variance is of primary concern, CWP behaves more like an unweighted estimator. The coverage of the crude trimming estimator (WT<3) is very poor when $\log \sigma < 0$, and the method has unacceptable bias. CWP suffers coverage problems when the between- and within-variances are approximately equal. The coverage of MSET is markedly below the nominal level, because $Var(\bar{y}_{wt})$ does not account for the variability in estimating the cutpoint.

The right panels of Figure 1 gives the RMSE relative to the fully weighted estimator [panel (b)] and nominal 90% coverage [panel (d)], when the mean structure is given by $\boldsymbol{\delta}^D$. This is an unfavorable scenario for trimming, since the highest weight stratum has a mean substantially different from the other strata. The unweighted and crude trimming methods perform very poorly unless σ is large. The CWP estimator behaves well for small or large values of the variance. However, it is less satisfactory for intermediate values of σ, tending to overpool. The MSET estimator is more protective of overpooling, although it is less effective than CWP at reducing RMSE for large values of σ. The CWP interval estimates have below nominal coverage when σ is moderate. The MSET estimator displays coverage rates closer to nominal levels, but also has poorer coverage than the FWT estimator when variance is moderate.

3.2 Weight Smoothing Methods

For weight smoothing models, mean structures $\boldsymbol{\delta}^E$, $\boldsymbol{\delta}^L$, and $\boldsymbol{\delta}^D$ were considered. Five weighting schemes were considered: the fully-weighted estimator $\overline{y}_w$ (FWT), and posterior estimate of $\overline{Y}$ from (7) where $\hat{\phi}_h$ is obtained under the XRE, AR1, LIN, and NPAR models.

Figure 2 compares the RMSE for $\overline{y}_{XRE}$, $\overline{y}_{AR1}$, $\overline{y}_{LIN}$, and $\overline{y}_{NPAR}$ relative to $\overline{y}_w$. As expected, the XRE and AR1 estimators do well when the means are equal. Both these estimators perform poorly relative to $\overline{y}_w$ when the means are unequal, although as expected the AR1 estimator is more robust than the XRE estimator. The LIN estimator works well for both equal and linear mean structures, although as expected it is less efficient than XRE or AR1 when the stratum means are equal. For the non-linear mean structure $\boldsymbol{\delta}^D$, LIN performs poorly relative FWT for moderate σ^2. Since $\overline{y}_{NPAR} \rightarrow \overline{y}_w$ as the roughness penalty $\lambda \rightarrow 0$ and $\overline{y}_{NPAR} \rightarrow \overline{y}_{LIN}$ as $\lambda \rightarrow \infty$, NPAR can be viewed as a compromise between LIN and FWT; the simulations suggest that this compromise works well. Specifically, NPAR performs nearly as well as LIN for equal and linear mean structures. When the stratum means are non-linear, NPAR mimics FWT for small to moderate values of σ^2, and mimics LIN as σ^2 increases and the RMSE of LIN is lower than the RMSE of FWT.

Figure 2 also shows the coverage of the various weight smoothing estimators for different mean structures and variances. All estimators have good coverage properties when the true superpopulation means are equal, since all models allow equal means. The XRE and AR1 models yield intervals with poor coverage when the means follow a linear trend and variance is moderate. The LIN model has moderate coverage problems when the trend is non-linear. The NPAR and FWT procedures are close to nominal levels for all mean structures and all values of σ considered.

4 Discussion

Survey weights are generally trimmed in an ad-hoc manner, with little attention given to the optimum degree of trimming. We have considered a number of methods that use the data to determine adjustments of the weights that involve appropriate bias-variance trade-offs. One approach is to obtain an estimate of root mean squared error and then choose a trimming point that minimizes this estimate. This method performed reasonably well in our simulations, although model-based methods were more efficient for some problems, and confidence intervals that fail to reflect uncertainty in the trimming point did not achieve nominal levels of coverage.

Our model-based procedures are divided into two classes, weight pooling models and weight smoothing models. The compound weight pooling model is proposed as a model-based analogue of weight trimming that allows Bayesian averaging over estimates based on different trimming points. Bayesian methodology also allows the uncertainty about the choice of trimming point to be included in the inference. In our empirical studies, this model did well in terms of RMSE when the mean configuration was favorable towards trimming, but tended to over-pool for some regions of the parameter space when the mean configuration was not favorable towards trimming. The over-pooling is even more problematic when confidence coverage, rather than MSE, is of interest, since the resulting bias results in intervals that are systematically shifted away from the population mean.

We also consider weight smoothing models that treat the unknown weight stratum means as random variables with their own mean and covariance structure. Choosing between models in this class involves trade-offs between robustness and efficiency. In particular, assuming exchangeable means and including between-stratum variance components to induce shrinkage, as in the XRE and AR1 models, yields estimators that have good properties when the sample design is highly disproportional and the data highly variable, but that are vulnerable to model misspecification when the between-stratum and within-stratum variances are approximately equal. In contrast, adding parameters to the mean structure, as in the LIN and NPAR models, reduces the problem of misspecification at some cost in efficiency. The NPAR model has the advantage of being more "believable" when the strata are nominal rather than ordinal, so there is less reason to believe a linear trend exists in the data. It yields estimates that behave somewhat like the MSET estimator, but with more efficiency and better confidence coverage properties. Indeed, this model was nearly as robust to alternative mean configurations as the fully-weighted estimator in simulations, yet approximates the efficiency of the LIN model estimator when variance overwhelms bias.

5 References

Atkinson, A.C. (1978). Posterior probabilities for choosing regression models. *Biometrika*, 65, 39-48.

Cox, B.G., and McGrath, D.S. (1981). An examination of the effect of sample weight truncation on the mean square error of survey estimates. Paper presented at the 1981 Biometric Society ENAR meeting, Richmond, VA.

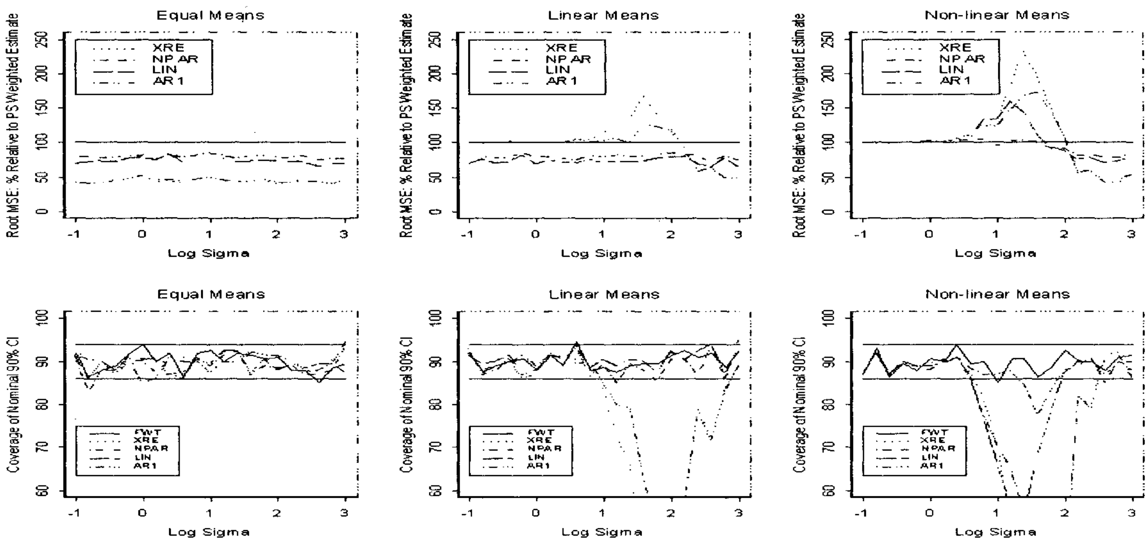

Figure 2: Average RSME (200 simulations) relative to unweighted estimator of exchangeable random effects (XRE), autoregressive (AR1), linear (LIN), and non-parametric (NPAR) weight smoothing estimators, and 90% nominal coverage of all estimators

Dempster, A.P., Schatzoff, M., Wermuth, N. (1977). A simulation study of alternatives to ordinary least squares (with discussion). *Journal of the American Statistical Association*, 72:77-106.

Ghosh, M., and Meeden, G. (1986). Empirical Bayes estimation of means from stratified samples. *Journal of the American Statistical Association*, 81:1058-1062.

Green, P.J. (1987). Penalized likelihood for general semi-parametric regression models. *International Statistical Review*, 55:245-260.

Halpern, E.F. (1973). Polynomial regression from a Bayesian approach. *Journal of the American Statistical Association*, 68:137-143.

Hastie, T.J., and Tibshirani, R.J. (1990). *Generalized Additive Models*, London: Chapman and Hall.

Holt, D., Smith, T.M.F. (1979). Poststratification. *Journal of the Royal Statistical Society*, A142:33-46.

Kish, L. (1992). Weighting for Unequal P_i. *Journal of Official Statistics*, 8:183-200.

Laird, N.M, Ware, J.H. (1982). Random effects models for longitudinal data, *Biometrics*, 38:963-974.

Lazzeroni, L.C. and Little, R.J.A. (1998). Random-effects models for smoothing post-stratification weights. *Journal of Official Statistics*, 14:61-78.

Little, R.J.A. (1991). Inference with survey weights. *Journal of Official Statistics*, 7:405-424.

Potter, F. (1990). A study of procedures to identify and trim extreme sample weights. Proceeds of the Survey Research Methods Section, American Statistical Association, 1990, 225-230.

Robinson, G.K. (1991). That BLUP is a good thing: the estimation of random effects (with discussion), *Statistical Science*, 6:15-51.

Wahba, G. (1978). Improper priors, spline smoothing, and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society*, B40:364-372.

Wahba, G. (1985). A comparison of GCV and GML for choosing the smoothing parameters in the generalized spline smoothing problem. *The Annals of Statistics*, 4:1378-1402.