American Journal of Theoretical and Applied Statistics
2021; 10(4): 202-207
http://www.sciencepublishinggroup.com/j/ajtas
doi: 10.11648/j.ajtas.20211004.14
ISSN: 2326-8999 (Print); ISSN: 2326-9006 (Online)

Comparison of the New Estimators: The Semi-Parametric Likelihood Estimator, SPW, and the Conditional Weighted Pseudo Likelihood Estimator, WPCE

Samuel Joel Kamun[1], Richard Simwa[2], Stanley Sewe[1]

[1]Department of Mathematics and Computing Studies, Faculty of Science, Catholic University of Eastern Africa, Nairobi, Kenya

[2]Department of Mathematics, Faculty of Science, University of Nairobi, Nairobi, Kenya

Email address:

samuelkamun@gmail.com (S. J. Kamun), richardsimwa@uonbi.ac.ke (R. Simwa), ssewew@cuea.edu (S. Sewe)

*Corresponding author

To cite this article:

Samuel Joel Kamun, Richard Simwa, Stanley Sewe. Comparison of the New Estimators: The Semi-Parametric Likelihood Estimator, SPW, and the Conditional Weighted Pseudo Likelihood Estimator, WPCE. *American Journal of Theoretical and Applied Statistics*. Vol. 10, No. 4, 2021, pp. 202-207. doi: 10.11648/j.ajtas.20211004.14

Received: August 6, 2021; **Accepted**: August 21, 2021; **Published**: August 31, 2021

Abstract: The analysis of sample-based studies involving sampling designs for small sample size, is challenging because the sample selection probabilities (as well as the sample weights) is dependent on the response variable and covariates. The study has focused on using systems of weighted regression estimating equations, using different modified weights, to estimate the coefficients of Weighted Likelihood Estimators. Usually, the design-consistent (Weighted) estimators are obtained by solving (sample) weighted estimating equations. They are then used to construct estimates which have better relative efficiencies and smaller finite small sample bias than the estimates from the Horvitz-Thompson Weighted Estimator with unmodified weight, option A. The purpose of our study is to compare derived Estimators of the weighted regression estimating equations for estimating the coefficients of Weighted Likelihood Estimators, the Semi-Parametric Weighted Likelihood Estimator, SPW and the Weighted Conditional Pseudo Likelihood Estimator, WCPE with the conventional Horvitz-Thompson Weighted Likelihood Estimator, using relative efficiency, sample bias and Standard Error for small sample size. The constructed estimates from the system of weighted regression estimating equations, using different modified weights, are actually the Weighted Likelihood Estimators. The study compared the two new estimators, the Semi-parametric weighted estimator, SPW and the Weighted Conditional Pseudo Likelihood estimator, WCPE, for both the unmodified and modified Weights, which were found to have better relative efficiency and smaller finite small sample bias than the estimates from conventional Horvitz-Thompson Weighted Estimator, for both generated and for real data. The outcome of the tests show strong similarity in performance to those obtained using the simulated data. Estimates were constructed which have better relative efficiencies and smaller finite small sample bias than the estimates from the Horvitz-Thompson Weighted Estimator with unmodified weight, option A.

Keywords: Semi Parametric, Imputation, Estimating Error, Small Samples, Estimators, Relative Efficiency, Sample Bias

1. Introduction

This study is a follow up on the main reference found in Kamun et al [22]. The study has focused on using systems of weighted regression estimating equations [14], using different modified weights, to estimate the coefficients of Weighted Likelihood Estimators [1, 3-5]. We use them to construct estimates which have better relative efficiencies and smaller finite small sample bias than the estimates from the Horvitz-Thompson Weighted Estimator with unmodified weight, option A. Further inference on Stratified samples can be found in [15].

1.1. The Model

Suppose data, Lawless et al. [18], is produced according to a function

$$f(y \mid x; \theta)g(x) \tag{1}$$

where y is a response variable which is multivariate and x is a continuous or discrete vector of covariate variables and

$$f(y \mid x; \theta) \tag{2}$$

is the regression part of the function. The marginal distribution of x is denoted by g(x), which for this study we have used Gaussian density to represent, is as shown below.

$$k(u) = \frac{1}{\sqrt{(2\pi)}} e^{-\left(\frac{u^2}{2}\right)} \tag{3}$$

where $u = \frac{x_1 - mean(x_1)}{s\tan.dev(x_1)}$

We describe the conditional distribution of y given x_1 as θ. The likelihood is given by

$$\prod f(y|x;\theta) \tag{4}$$

1.2. System of Estimating Equations

Let the function below represents a regression equation

$$y = f(y_j|x_j;\theta) g(x_j) \tag{5}$$

multiplying equation (5) by x_j

$$x_j y = x_j f(y_j|x_j;\theta) g(x_j) \tag{6}$$

and summing from j=1 to N equation (6)

$$\sum_{j=1}^{N} x_j y = \sum_{j=1}^{N} x_j f(y_j|x_j;\theta) g(x_j) \tag{7}$$

We can rewrite the equation as

$$\sum_{j=1}^{N} x_j y - \sum_{j=1}^{N} x_j f(y_j|x_j;\theta) g(x_j) = 0 \tag{8}$$

or as

$$\sum_{j=1}^{N} x_j \left(y - f(y|x_j,\theta) g(x_j)\right) = 0 \tag{9}$$

and when we multiply the equation by the weight w_j it gives us the sample weighted estimating equations

$$\sum_{j=1}^{N} w_j x_j \left(y - f(y|x_j,\theta) g(x_j)\right) = 0 \tag{10}$$

the sample estimating equations [14] and sample weighted estimating equations have been used as equations (9) and (10) respectively and are used to find solutions to regression equations [13].

2. Weighted Likelihood

In our study, we propose two novel weighted system of regression estimating equation for estimating coefficients for small sample studies: the small sample pseudo likelihood weighted (SPW) model and the small sample semi parametric weighted conditional (WCPE) model [17]. By fitting a regression model to the establish sample weights against the sample variables and or response variable [7], which is the requirement for the two estimates to be obtained we end up estimating the conditional expectation of the weights.

We modified the original sample weights, for WCPE, where the estimated conditional expectation of the weights is used. For our study, both response variables and predictor variables were used to find estimated weights. The SPW and WCPE models are an improvement of the design-based weighted estimates with improvement achieved by using suitable modification on the original sample weights before the estimation (e.g., [8]).

2.1. Horvitz-Thompson Weighted Estimator

Consider a finite population Ω of N individuals. Let y denote the response variables and v(x', z')' denote the vector of all measured covariates, where x are covariates associated with the outcome and z are the sample variables used in the process of sample selection. x and z may or may not have common variables.

Let n be the size of the observed sample S. The probability that individual i, i = 1, 2, 3, …, n, is included in the sample is denoted by π_i. The base sample weight w_i is defined as a reciprocal of the sample inclusion probability πi, so πi = 1/w_i. We refer to the final sample weight as the sample weight of individual i and denoted by w_i.

We have assumed in this study that the observed data are the data for the simulated units, for individuals i ε S, i = 1, 2, …, n., as we observe (y_i, v_i', w_i')'. We have assumed also the availability of sample proportions and means for the variables in z_i', which can be used to calibrate the sample weights.

In the population, the outcomes Y_i conditional on covariates X_i have distribution

$$f(Y_i|X_i,\theta) g(X_i)$$

For our study we use

$$f(Y_i|X_i,\theta) = \beta_0 + \beta_i X_i^{'}$$

Where $\beta_i = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$.

By fitting model (1) to the entire finite population Ω, we can estimate β_i consistently by solving a system of estimating equations with respect to b:

$$G(b) = \sum_{i=1}^{N} x_i \left(y_i - f(y_i|x_i,\theta) g(x_i)\right) = 0 \tag{11}$$

where

$$f(y_i|x_i,\theta) g(x_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$$

and $g(x_i)$ is the marginal distribution of x_i obtained non parametrically. In equation (5), G(b) above, are the maximum likelihood estimating equations based on all the values (y_i, x_i) ε Ω if Ω is a stratified random sample from a super population generated under model (1)(cluster sampling could equally be used [2]). Denote the solution of (5) by B.

For any fixed value of B, G(B) is a vector of finite sample totals, and hence, it can be estimated from the sample S by

$$\hat{G}_w(b) = \sum_{i\varepsilon S} w_i x_i \left(y_i - f\left(y_i | x_i, \theta\right) g\left(x_i\right)\right) \quad (12)$$

where y_i is the observed response for the unit i. We denote the solution of the system of the equations $\hat{G}_w(b) = 0$ by $\hat{\beta}_w$ which is referred to as the weighted estimator (the Horvitz-Thompson estimator).

2.2. Weighted Conditional Likelihood: Pseudo Maximum Likelihood Estimator

Let $f_p\left(w_i | y_i, x_i, \theta\right)$, which is the conditional pdf for the weight in the population [10], be the weight of the population distribution where the observed weight stems from the sample distribution f_s, obtained by Bayes rule as

$$f_p\left(w_i | y_i, v_i, \theta\right) = f_p\left(w_i | y_i, z_i, i\varepsilon S, \theta\right) = \frac{\Pr\left(i\varepsilon S | y_i, v_i\right) f_p\left(w_i | y_i, x_i, \theta\right)}{\Pr\left(i\varepsilon S\right)} \quad (13)$$

We assume from equation (13) and in this study that

$$f_p\left(w_i | y_i, v_i, \theta\right) = f_p\left(w_i | y_i, x_i, \theta\right)$$

for each value of i. This is an example of informative sampling that should be accounted for in the inferential process.

Equation (13) can be written equivalently by

$$f_s\left(w_i | y_i, v_i, \theta, \gamma, \beta\right) = \frac{E_s\left(w_i | \theta, \gamma, \beta\right) f_p\left(w_i | y_i, x_i, \theta, \beta\right)}{E_s\left(w_i | y_i, v_i, \gamma\right)} \quad (14)$$

Where $E_s\left(w_i | .\right)$ refers to the conditional expectation of the weights w_i with respect to their sample distribution. Also γ is an unknown vector of the unknown regression coefficients of the regression model on the design variables and the outcome for sample weights [11].

$$G_S\left(w_i, \gamma, \beta\right) = \sum_{i\varepsilon S} \frac{\partial E_s\left(w_i | y_i, v_i, \theta, \gamma, \beta\right)}{\partial \beta}$$

is the sample log-likelihood score equation in relation to β.

It follows then from equation (10), if it can be assumed that (1) is a population distribution of y_i given x_i, that these equations can equivalently be written as

$$G_S\left(w_i, \gamma, \beta\right) = \sum_{i\varepsilon S} x_i \left(y_i - f\left(y_i | x_i, \theta\right) g\left(x_i\right)\right) * \sum_{i\varepsilon S} \frac{\partial E_s\left(w_i | y_i, v_i, \theta, \gamma, \beta\right)}{\partial \beta} \quad (15)$$

or

$$G_S\left(w_i, \gamma, \beta\right) = \frac{\sum_{i\varepsilon S} x_i \left(y_i - f\left(y_i | x_i, \theta\right) g\left(x_i\right)\right)}{\sum_{i\varepsilon S} \frac{\partial E_s\left(w_i | y_i, v_i, \theta, \gamma, \beta\right)}{\partial \beta}} \quad (16)$$

Where $E_s\left(w_i | v_i, \theta, \beta, \gamma\right)$, can be differentiated with respect to β for any fixed γ.

We estimate γ separately, using individual-level data w_i, y_i, v_i observed for sample units, i ε S. We then express the sample expectations $E_S(w_i | v_i)$ as functions of β and the estimated value of γ. We then obtain the system of equations from (15) or (16) by substituting $\hat{\gamma}$ in place of γ, and solve it iteratively with respect to β,

$$G_S\left(\hat{\gamma}, \beta\right) = 0$$

We denote this solution out of (9 or 10) by $\hat{\beta}_{WCPE}$ and refer to it as an WCPE estimator according to [12].

2.3. Semi-parametric Weighted Estimator

We obtain the modified sample weights using four steps as follows:

i. We obtain sample weights from two samples, n and g, where w_g and w_n are worked out as regression of the sample weights w_i on y_i, z, x_1, x_2, x_3 and x_4, for w_g and w_i on x_1, x_2, x_3 and x_4 for w_n.
ii. We then obtain sample weights from the two samples, n and g, where w_{reg1} and w_{reg2} are worked out as regression using least squares of sample weights w_i on y_i, z, x_1, x_2, x_3 and x_4 for w_{reg1} and w_i on x_1, x_2, x_3 and x_4 for w_{reg2}.
iii. We define the weights that have been rescaled as $\tilde{w}_i$ by

$$\tilde{w}_i = \left(\frac{w_g + w_n}{2}\right)$$

iv. We then fit the regression model to the weight which have been rescaled, against the covariates v_i and compute

$$E_S\left(\tilde{w}_i | v_i\right) = E_S\left(\tilde{w}_i | z_i, \psi\right)$$

Where ψ refers to the estimated coefficients of the regression model which relates $\tilde{w}_i$ to the covariates v_i. We define the model with adjusted weights by

$$\tilde{w}_i^m = \frac{w_{reg1}}{E\left(w_{reg1}\middle|x_i, y_i, z\right)} * \frac{E\left(w_{reg2}\middle|x_i\right)}{w_{reg2}} * \frac{w_i}{E\left(w_i\middle|x_i\right)} \quad (17)$$

The system of weighted estimating equations

$$\hat{G}_{SPW} = \sum_{i\varepsilon S} \tilde{w}_i^m x_i \left(y_i - f\left(y_i\middle|x_i,\theta\right) g\left(x_i\right)\right) = 0 \quad (18)$$

is solved with regards to $\hat{\beta}_{SPW}$, and its solution, is the semi-parametric weighted estimator as proposed in [10].

3. Results

We repeated the simulations 10,000 times for sample plan A. We computed the simulated bias for the regression coefficients as the average of the difference of the estimated coefficient minus the true coefficient. We computed the simulated standard deviation for a regression coefficient by dividing the standard error by the number of simulations. We have used software packages designed for analysis of the weighted estimators using R [16].

Table 1. *Summary of Comparison and Analysis of Estimates for Generated Data.*

Estimators	Plan	Coefficient of Determination. R^2	Adjusted Coefficient of Determination. R^2	Bias	Standard Error	Relative Efficiency (Var1/Var2)	Mean	Stan.Dev
WCPE (MOM)	A	0.99999999997974	0.99999999997	2.5881519e-12	1.2250786e-11	0.982396	45.79445	0.51578
	B	0.99999999997379	0.99999999996	4.5261572e-12	1.4955370e-11	1.045718	45.74961	0.49992
	C	0.99999999998251	0.99999999998	2.8067548e-12	9.623465e-12	1.136823	45.70296	0.47947
WCPE (MASS)	A	0.99999999998446	0.999999999978	4.2077453e-12	6.4815375e-12	1.260005	45.76692	0.45543
	B	0.99999999996874	0.999999999956	8.6316509e-12	1.8434238e-11	1.042005	45.74961	0.50081
	C	0.99999999999274	0.99999999999	1.6465718e-12	4.3478758e-12	1.071496	45.77053	0.49387
WCPE (STATS4)	A	0.99999999997703	0.999999999968	4.3728354e-12	1.2832394e-11	0.993460	45.76174	0.51290
	B	0.99999999998055	0.999999999973	2.4388269e-12	1.1426680e-11	0.983769	45.76172	0.51542
	C	0.99999999999922	0.999999999999	1.2534418e-13	4.7164760e-13	0.034135	45.62707	2.76699
SPW	A	0.99999999994496	0.999999999923	8.9096508e-12	4.4634743e-11	2.695988	45.85448	0.31135
	B	0.99999999996414	0.99999999995	8.5930152e-12	1.5639812e-11	0.993661	45.74964	0.51286
	C	0.999999999985982	0.99999999998	3.2606140e-12	7.6167599e-12	1.152102	45.71676	0.47628
HTWE	A	0.9999999999585073	0.999999999942	1.0464185e-11	1.8758009e-11	1.000000	45.77449	0.51122
	B	0.999999999983550	0.999999999977	1.1735057e-12	1.2060731e-11	1.051093	45.74964	0.49864
	C	0.999999999981308	0.999999999974	3.9779291e-12	1.0356865e-11	1.118182	45.72881	0.48345
HTWE	w=1	0.9999999999844992	0.999999999978	3.1787906e-12	9.2209949e-12	1.046095	45.74960	0.49983
Monte Carlo		0.9999999999790133	0.999999999971	6.2848615e-12	8.9509268e-12	1.043087	45.74961	0.50055

HTWE = Horvitz-Thompson Weighted Estimator (HTWE); SPW = Semi parametric Weighted Estimator,
WCPE = Weighted Conditional semi parametric Estimator., HTWE (w = 1) = Un-weighted HTWE Estimator,
Var1 = Variance of HTWE Option A, Var2 = Variance of another Estimator.

The results in Table 1 show that all the Estimators for options A, B and C, except WCPE (STATS4) options A, B and C, with lower relative efficiency, have higher relative efficiencies and coefficients of determination than HTWE for option A including Monte Carlo Simulation, and hence are more efficient for Simulated Data for n = 15.

Table 2. *Summary of Comparison and Analysis of Estimates for Real Data.*

Estimators	Plan	Coefficient of Determination. R2	Bias	Standard Error	Relative Efficiency (Var1/Var2)	Mean	Stan.Dev
WCPE (MOM)	A	0.999999999910297	2.0048e-11	4.786e-11	0.8759	116.2194	2.6517
	B	0.9999999996825778	7.6514e-11	1.2443e-10	2.3213	116.2732	1.6288
	C	0.9999999999590754	1.0063e-11	1.65680e-11	0.4113	115.8890	3.8695
WCPE (MASS)	A	0.99999999984530186	3.7533e-11	5.6396e-11	0.8467	115.9398	2.6970
	B	0.999999999806429063	3.3201e-11	1.1911e-10	2.3218	116.2732	1.6287
	C	0.9999999999082826	2.7677e-11	3.8922e-11	0.5858	117.5860	3.2425
WCPE (STATS4)	A	0.99999999992502486	2.3693e-11	2.5764e-11	0.8518	116.0661	2.6888
	B	0.999999999833797504	3.5832e-11	9.3068e-11	2.3215	116.2732	1.6288
	C	0.999999999951708	1.1129e-11	2.1533e-11	0.8518	116.3275	2.5979
SPW	A	0.9999999999999848	3.9968e-15	6.2258e-15	1.0834	116.2212	2.3842
	B	0.9999999999999820	3.9968e-15	8.0151e-15	2.3215	116.2732	1.6288
	C	0.999999999999959	1.1102e-15	1.9884e-15	0.7106	116.6525	2.9439
HTWE	A	0.9999999999433389	1.4085e-11	2.3787e-11	1.0000	116.1443	2.4816
	B	0.99999999999999467	1.11022e-15	2.7760e-15	0.3582	116.2732	4.1462
	C	0.999999999999991	2.1094e-15	4.4433e-15	0.7051	116.3878	2.9554
HTWE (w=1)		0.999999999629764602	9.2224e-11	1.4834e-10	2.3233	116.2733	1.6281

HTWE = Horvitz-Thompson Weighted Estimator (HTWE); SPW = Semi parametric Weighted Estimator,
WCPE = Weighted Conditional semi parametric Estimator., HTWE (w = 1) = Un-weighted HTWE Estimator,
Var1 = Variance of HTWE Option A, Var2 = Variance of another Estimator.

The results in Table 2 show that all estimators whose relative efficiency is greater than one are more efficient than HTWE (A) for real data for n = 15.

Table 3. S*ummary of the Performance of Estimators based on Bias and Standard Error for Real Data.*

Estimators	Plan	$\hat{\beta}_0$		$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\beta}_3$		$\hat{\beta}_4$	
		Bias	**Error**	**Bias**	**Error**	**Bias**	**Error**	**Bias**	**Error**	**Bias**	**Error**
WCPE (MOM)	A	-21.60857	1.4995e-07	0.21271	2.5080e-10	-0.01046	4.7754e-11	0.06382	1.5103e-10	-0.01963	1.6181e-07
	B	26.61126	7.0506e-08	0.08192	2.9014e-10	-0.01491	7.1071e-12	0.01473	8.0998e-11	-0.01984	2.5863e-12
	C	-22.97434	6.7705e-08	0.33323	3.4611e-10	0.11031	7.5084e-12	0.04374	6.7999e-11	-0.01901	2.5813e-12
WCPE (MASS)	A	-21.60916	8.3168e-08	0.18817	3.3363e-10	-0.01035	8.5216e-12	0.06680	8.76977e-11	-0.01892	2.7146e-12
	B	26.62450	6.0548e-08	0.08189	2.2084e-10	-0.01491	6.7780e-12	0.01472	6.5860e-11	-0.01984	1.9245e-12
	C	-21.60980	8.8651e-08	2.89294	2.8204e-10	-8.9177e-03	9.8664e-12	0.05403	1.0917e-10	-0.01943	3.4142e-12
WCPE (STATS4)	A	-21.60903	5.5451e-08	0.17072	1.6534e-10	-0.01002	6.7411e-12	0.06919	7.1569e-11	-0.01877	2.5745e-12
	B	26.62442	5.5440e-08	0.54149	2.0925e-10	-0.01491	5.6512e-12	0.01472	6.2208e-11	-0.01984	1.8769e-12
	C	-21.60879	4.6955e-08	0.20427	1.6957e-10	-0.01102	5.1655e-12	0.06645	5.7165e-11	-0.01955	1.7770e-12
SPW	A	-21.60885	7.7450e-10	0.14864	2.5578e-12	-0.01118	9.2353e-14	0.07652	9.3822e-13	-0.01929	2.9830e-14
	B	26.61728	4.9624e-10	0.08191	1.8309e-12	-0.01491	5.8573e-14	0.01473	6.8968e-13	-0.01984	2.3004e-14
	C	-21.60913	3.8547e-04	0.23908	1.1342e-06	-9.9980e-03	1.3266e-08	0.05820	4.5258e-07	-0.01946	2.0363e-14
HTWE	A	-68.33828	4.4372e-08	0.58190	1.6172e-10	-0.02928	5.2694e-12	0.07680	5.6618e-11	0.01894	1.9212e-12
HTWE (w=1)		26.75116	7.48091e-08	0.08164	3.0336e-10	-0.01492	7.7339e-12	0.01457	8.8192e-11	-0.01984	3.3560e-12

The simulated standard error and bias for most of the estimator are lower than for the Horvitz–Thompson weighted estimator option A, except for HTWE w = 1.

Table 4. *Summary of the Performance of Estimators based on Standardized (Beta) Regression Coefficients.*

Estimators	**Plan**	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
WCPE (MOM)	A	-4.2317e-15	1.0510e+00	8.3729e-01	6.5699e-01	5.2401e-01
	B	-1.6718e-15	8.2424e-01	1.2161e-01	-2.0638e-01	7.2585e-01
	C	1.6697e-16	1.0642e+00	3.9770e-01	2.3048e-01	5.1152e-01
WCPE (MASS)	A	1.5269e-15	9.3284e-01	8.4169e-01	6.9271e-01	7.6632e-01
	B	-3.3346e-15	8.2408e-01	1.2133e-01	-2.0675e-01	7.2581e-01
	C	-1.1667e-15	9.9878e-01	9.0068e-01	4.0943e-01	4.8786e-01
WCPE (STATS4)	A	4.4415e-16	8.6397e-01	9.0014e-01	7.3247e-01	8.2153e-01
	B	1.7106e-15	8.2407e-01	1.2134e-01	-2.0677e-01	7.2581e-01
	C	-2.8381e-15	1.0369e+00	7.5675e-01	7.1347e-01	5.7631e-01
SPW	A	6.1736e-16	8.7213e-01	7.9371e-01	9.5627e-01	7.1912e-01
	B	1.0540e-16	8.2416e-01	1.2149e-01	-2.0655e-01	7.2584e-01
	C	-2.9349e-15	1.0456e+00	8.2519e-01	5.1056e-01	5.2668e-01
HTWE	A	2.7255e-16	7.2067e-01	1.0512e+00	9.3430e-01	8.3798e-01
	B	-1.0039e-15	1.0630e+00	4.0809e-01	1.9436e-01	4.9729e-01
	C	-9.1776e-16	1.0718e+00	6.1968e-01	5.2272e-01	5.3370e-01
HTWE (w=1)		-5.2853e-15	8.2268e-01	1.1821e-01	-2.1060e-01	7.2512e-01

Simulated Standardized (Beta) Regression Coefficients for each coefficient for over 10,000 repeated simulations.
A standardized beta coefficient compares the strength of the effect of each individual independent variable to the dependent variable. The higher the absolute value of the beta coefficient, the stronger the effect, here we make comparison with HTWE option A.

Additional techniques of variance estimation can be found in [6].

4. Conclusion

In this paper, we propose two estimators for regression coefficients for analyses: an SPW estimator and a WCPE estimator. The estimators proposed provide an alternative to the conventionally used weighted (Horvitz–Thompson) estimator that can be very inefficient when applied to data with highly variable weights [4].

The two proposed estimators show significant improvement in efficiency and can be readily applied to the real data. The WCPE estimator and its standard errors can be computed using existing R software [16]. Therefore, its implementation is straightforward and requires only knowledge of the individual values of the weights and covariates observed in the sample.

The SPW estimator can be sensitive to the misspecification of the selection model, and hence it may not be appropriate for the analyses in which the selection model cannot be specified and estimated accurately. Also, unlike the WCPE estimator, computing the SPW estimator requires additional programming to incorporate the selection model into the estimating equations.

In survey research, regression and post stratification calibration of the sample weights are used in weighted estimation [3, 9] to reduce the variance and bias (such as from deficient coverage of the sample frame) of the weighted estimators. Both types of calibration require knowledge of population proportions or means of design variables that are often available from population census data. In this paper, the sample weights we used were calibrated to real data categories.

To summarize, we recommend using the WCPE estimator because our simulations show it to be nearly as efficient as the SPW estimator. We recommend that the choice of the adjusted sample weights provided for analyses should be made without regard to any specific analysis but according to a general criteria of efficiency such as the CV of the weights. This will avoid choosing adjusted sample weights that provide the ‘desired’ results for a particular analysis. Further empirical research would be useful to investigate the finite sample properties of these estimators under other sample designs and types of sample weighting. Also, it would be useful to further extend the SPW and WCPE estimations to other types of studies. It is now evident that weighted regression is a viable option for obtaining estimators [19-21].

References

[1] Scott A. J, Wild CJ. Fitting logistic models under case-control or choice based sampling. JRSS, B 1986; 48: 170–182.

[2] Graubard B. I, Fears TR, Gail MH. Effects of cluster sampling on epidemiologic analysis in population-based case-control studies. Biometrics 1989; 45: 1053–1071.

[3] Korn EL, Graubard BI. Analysis of Health Surveys. Wiley: New York, NY, 1999.

[4] Scott AJ, Wild CJ. Population-based case-control studies. In Handbook of Statistics: Sample Surveys: Inference and Analysis, Vol. 29B, Pfefferman D, Rao CR (eds). Elsevier: Amsterdam, 2009; 431–453.

[5] Li Y, Graubard BI, DiGaetano R. Weighting methods for population-based case-control studies with complex sampling. JRSS (C) 2011; 60: 165–185.

[6] Rust KF, Rao JNK. Variance estimation for complex surveys using replication techniques. Statistical Methods in Medical Research 1996; 5: 283–310.

[7] Pfeffermann D, Sverchkov M. Inference under informative sampling. In Handbook of Statistics: Sample Surveys: Inference and Analysis, Vol. 29B, Pfefferman D, Rao CR (eds). Elsevier: Amsterdam, 2009; 455–487.

[8] Robins J. M, Rotnitzky A, Zhao L-P. Estimation of regression coefficients when some regressors are not always observed. Journal of American Statistical Association 1994; 89: 846–866.

[9] Lumley T, Shaw PA, Dai JY. Connections between survey calibration estimators and semiparametric models for incomplete data. International Statistical Review 2011; 97 (2): 200–220.

[10] Pfeffermann D, Sverchkov M. Parametric and semi-parametric estimation of regression models fitted to survey data. Sankhya, Series B 1999; 61: 166–186.

[11] Pfeffermann D, Krieger A, Rinott Y. Parametric distributions of complex survey data under informative probability sampling}. Statistica Sinica 1981; 8: 1087–1114.

[12] Gong G, Samaniego F. Pseudo maximum likelihood estimation: theory and applications. The Annals of Statistics 1981; 9: 861–869.

[13] Pfeffermann D, Sverchkov M. Fitting generalized linear models under informative sampling. In Analysis of Survey Data, Skinner C, Chambers R (eds). Wiley: New York, 2003; 175–195.

[14] Yuan K-H, Jennrich R. Estimating equations with nuisance parameters: theory and applications. Annals of the Institute of Statistical Mathematics 2000; 52: 343–350.

[15] Krewski D., Rao J. N. K. Inference from stratified samples: properties of linearization, jackknife and balanced repeated replication methods. The Annals of Statistics 1981; 9: 1010–1019.

[16] Lumley T. Complex Surveys: A Guide to Analysis Using R. Wiley: Hoboken NJ, 2010.

[17] Landsman, V. and Graudbard, B. I. Efficient analysis of case-control studies. Wiley on line library.com, DOI: 10.1002/SIM 5530, 2012.

[18] J. F. Lawless, J. D. Kalbfleisch and C. J. Wild. Semiparametric Methods for Response-Selective and Missing Data Problems in Regression, Wiley, Journal of the Royal Statistical Society. Series B, Vol. 61, No. 2 (1999), pp. 413-438 (26 pages).

[19] Edwin L. Bradley. The Equivalence of Maximum Likelihood and Weighted Least Squares Estimates in the Exponential class of Distribution, Wiley, Journal of the American Statistical Association; 68: 341, 199-200, DOI (1973).

[20] Robert J. Gray. Weighted Estimating Equations for Linear Regression Analysis of Clustered Failure Time Data, Lifetime Data 9 (2): 123-138, (2003), DOI: 10.1023/A:1022932117951.

[21] Adhidev Biswas, Suman Majumder, Pratim Guha Niyogi and Aye nendranath Basu. A Weighted Likelihood Approach to Problems in Survival Data, sankhaya (2019). http://doi.org/10.1007/513571-019-00214-W

[22] Samuel J. Kamun, Richard O. Simwa and Stanley Sewe. On Derivation of the Semi-Parametric Weighted Likelihood Estimator, SPW, and the Weighted Conditional Pseudo Likelihood Estimator, WPCE, Far East Journal of Theoretical Statistics © 2021 Pushpa Publishing House, Prayagraj, India, http://www.pphmj.com, http://dx.doi.org/10.17654/