High Order Penalty Functions in Calibration Estimators

Pius Nderitu Kihara
Department of Statistics and Actuarial Sciences, Technical University of Kenya
Email: piuskihara@yahoo.com

Abstract

Use of penalty functions in calibration estimators has severally been considered by this author. A calibration problem is transformed to an unconstrained optimization problem by constructing a penalty function. To guarantee convergence in the minimization of the penalty function by the Newton method, the order of the penalty function is usually restricted to 2. In this paper, we consider use of more flexible higher order penalty functions by applying the variable metric method. We report on the results of the resulting population total estimator for a cubic penalty function.

Keywords: variable metric method**,** calibration, penalty function, model calibration, unconstrained optimization.

1. Introduction

Suppose $U = \{1,2,..., N\}$ is the set of labels for the finite population. Let (y_i, x_i) be the respective values of the study variable y and the auxiliary variable x attached to the ith unit. Let $s = \{1,2,..., n\}$ be the set of sampled units under a general sampling design p, and let $\pi_i = p(i \in s)$ be the first order inclusion probabilities and $d_i = \pi^{-1}$. Wu and Sitter [9] considered the conventional calibration estimator $\hat{y}_t = \sum_{i=1}^{n} w_i y_i$ for the population total where $w_i' s$ are design weights obtained by minimizing the chi-square distance measure below

$$\Phi = \sum_{i \in s} \frac{(w_i - d_i)^2}{q_i d_i} \tag{1}$$

Subject to the Deville and Sarnda l[2] constraint below

$$\sum_{i=1}^{n} w_i x_i = \sum_{i=1}^{N} x_i \tag{2}$$

and yet another constraint

$$\sum_{i=1}^{n} w_i = N \tag{3}$$

Kihara [4] considered the above calibration problem as a nonlinear optimization problem which he solved by converting it into unconstrained optimization problem by use of penalty function method. He obtained the penalty function below.

$$\phi\ (w, r_k, x) = \sum_{i=1}^{n} \frac{(w_i - d_i)^2}{q_i d_i} + r_k \left[\sum_{i=1}^{n} w_i x_i - \sum_{i=1}^{N} x_i \right]^q + r_k \left[\sum_{i=1}^{n} w_i - N \right]^q \tag{4}$$

Frank and Jorge [3] have discussed flexible ways of choosing the penalty. Also, the rationale of the penalty terms is described by Ozgur [6].

In particular, Kihara [4] considered a quadratic penalty function where $q = 2$. He used the Newton method of unconstrained optimization to obtain the optimal weights $w_i' s$. While quite efficient for penalty functions of degree $q = 2$, the Newton method does not guarantee convergence of the penalty functions of higher degrees $q > 2$ as illustrated in Rao[7]. In fact the Newton method may lead to divergence of the penalty function. It is with this in mind that Kihara [4] restricted the penalty function to degrees $q = 2$. In this paper, we look at the variable metric method as a way of handling higher order degree penalty functions.

2. Higher order penalty functions

In calculus, to obtain w_i that minimizes (1) we differentiate (4) partially with respect to w_i to obtain

$$\phi'(w_i, r_k, x) = \frac{2(w_i - d_i)}{q_i d_i} + q r_k x_i \left[\sum_{j=1}^{n} w_j x_j - \sum_{j=1}^{N} x_j \right]^{q-1} + q r_k \left[\sum_{i=1}^{n} w_i - N \right]^{q-1} \tag{5}$$

Equating (5) to zero and solving for w_i would give optimal w_i and the estimator of population total becomes

$$\hat{y}_t = \sum_{i=1}^{n} w_i y_i \tag{6}$$

Cleary, it's quite tedious to obtain w_i from equation (5) for high degrees q. With the penalty approach, to obtain the weights w_i, $(i = 1,2,..., n)$, we solve the penalty function (4) as unconstrained minimization problem in which case we only require to start with some initial guess for w_i and r_k and then iteratively improve on the initial values until we have optimal values. Since the constraints (2) and (3) are equality constraints, we need not start with a feasible guess for w_i. In this paper, we adopt the variable metric method discussed in Rao [7] to estimation problems to allow flexibility in the choice of the degree of the penalty function.

Let $W = \{w_i, w_2,..., w_n\}$ be the set of the weights. We need to obtain W^* such that

$$g(W^*) = [\phi'(w_1, r_k, x), \ldots, \phi'(w_n, r_k, x)]' = 0 \tag{7}$$

We first start with some initial penalty r_1, some initial approximation W_{1i} of W^* and a $n by n$ positive definite symmetric matrix H_{1i} which may be taken as the identity matrix I. Compute $\phi'(W_{1i})$ at the point W_{1i} and set

$$Z_{1i} = -H_{1i}\phi'(W_{1i}) \tag{8}$$

Find the optimal step length λ_{1i}^* in the direction Z_{1i} by differentiating $\phi(W_{1i} + \lambda_{1i} + Z_{1i})$ with respect to λ_{1i}, equating the derivative to zero and solving for λ_{1i}. Now, we set

$$W_{1(i+1)} = W_{1i} + \lambda_{1i}^* Z_{1i} \tag{9}$$

and test for the optimality of $W_{1(i+1)}$. If optimal, it becomes the estimator for W^* at the penalty value r_1. If not optimal, we adjust the H matrix as

$$H_{1(i+1)} = H_{1i} + \frac{\lambda_{1i}^* Z_{1i} Z_{1i}^T}{Z_{1i}^T Q_{1i}} - \frac{(H_{1i}Q_{1i})(H_{1i}Q_{1i})^T}{Q_{1i}^T H_{1i} Q_{1i}} \tag{10}$$

where $Q_{1i} = \phi'(W_{1(i+1)}) - \phi'(W_{1i})$

We repeat the process by adjusting equation (8) accordingly. The process is repeated until an optimal estimator say W_1 for W^* at that particular penalty value r_1 is obtained. We proceed further to obtain a sequence of penalty values $r_2, r_2, ..., r_k, ...$ and obtain a corresponding sequence of estimators $W_1, W_2, ..., W_k, ...$ until $W_k = W_{k+1}$ to a given accuracy level. The penalty values are set such that the starting point $r_1 > 0$ and $r_{k+1} = c r_k$, where $c > 1$. We can now generalize our estimator for the population total as

$$\hat{y}_t = \sum_{i=1}^{n} w_i y_i = W^{*T} Y_s \tag{11}$$

where $Y_s^T = (y_1, y_2, \ldots, y_n)$ is the sample from the population of y

3. Nonparametric Model Estimation

Considered is a super population regression model which is denoted by ξ and given as

$$y_i = \mu(x_i) + \varepsilon \tag{12}$$

where $\mu(x_i)$ is a smooth function. Given n pair of observations $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ from a population of size N. Let $\hat{\mu}(x_i)$ be the estimator of $\mu(x_i) = E_\xi(y / x)$. A nonparametric method like local polynomial or splines may be used for this estimation. In model calibration, the weights w_i's may be obtained by minimizing the distance measure (1) subject to the constraint (3) and the constraint (13) below defined on the fitted values

$$\sum_{i=1}^{n} w_i \hat{\mu}(x_i) = \sum_{i=1}^{N} \hat{\mu}(x_i) \tag{13}$$

We have the penalty function

$$\phi_{np}(w, r_k, \hat{\mu}(x)) = \sum_{i=1}^{n} \frac{(w_i - d_i)^2}{q_i d_i} + H(r_k)\left[\sum_{i=1}^{n} w_i \hat{\mu}(x_i) - \sum_{i=1}^{N} \hat{\mu}(x_i)\right]^q + H(r_k)\left[\sum_{i=1}^{n} w_i - N\right]^q \tag{14}$$

For $q = 2$, Kihara [5], used Newton method to obtain the design weights. For higher order penalty functions, the weights w_i's may now be obtained by the variable metric method described in section (2) above. If we let the resulting set of weights be $W_{np}^{*T} = (w_1, w_2, \ldots, w_n)$, then we have the estimator of the population total as

$$\hat{y}_{np} = \sum_{i=1}^{n} w_i y_i = W_{np}^{*T} Y_s \tag{15}$$

4. Results

We report on the performance of the estimators $\hat{y}_t$ and $\hat{y}_{np}$ defined in equations (11) and (15) respectively for cubic penalty functions. That is, for $q = 3$. Using R program, we simulated a population of independent and identically distributed variable x using uniform (0, 1). Using x as the auxiliary variable we generated the populations of size 300 for random variable y as a linear function $y = 2 + 5x$. For each of different sample sizes n, 5 samples were generated. Our initial penalty constant was set at $r_1 = 0.00010$. The convergence criteria considered was $W_k^* = W_{k+1}^*$ to six decimal places. The performances of the estimators $\hat{y}_t$ and $\hat{y}_{np}$ were compared to that of the Horvitz
Thompson design estimator $y_{ht} = \sum_{i=1}^{n} y_i d_i$ discussed in Thompson [8]. For the nonparametric population total estimator $\hat{y}_{np}$, we use local polynomial of degree 1to fit the y values so that

$$\hat{\mu}(x_i) = S_{si}^T Y_s \tag{16}$$

where, given $\varepsilon = (1,0,\ldots,0)^T$, $Y_s = (y_1, y_2, \ldots, y_n)^T$, $\varpi_{si} = (1/h) diag(K((x_i - x_i)/h), \ldots, K((x_i - x_i)/h))$, h is the bandwidth and X_{si} is a matrix with rows $[1, (x_j - x_i), \ldots, (x_j - x_i)^q]$, $j = 1,2,\ldots,n$, then $S_{si}^T = \varepsilon_1^T (X_{si}^T \varpi_{si} X_{si})^{-1} X_{si}^T \varpi_{si}$. See Breidt and Opsomer[1] for a discussion of local polynomial.

4.1 Results for Estimator $\hat{y}_t$

We let $y_t = \sum_{i=1}^{N} y_i$ be the actual population total, and $y_t - \hat{y}_t$ and $y_t - y_{ht}$ be the errors in the estimation

Table 1: Estimates and the Errors for $\hat{y}_t$ and y_{ht}

sample number	1	2	3	4	5
sample size n	100	100	100	100	100
y_t	1384.49498	1384.49498	1384.49498	1384.49498	1384.49498
$\hat{y}_t$	1399.01338	1404.88568	1397.04904	1322.13453	1331.37077
y_{ht}	1399.26412	1405.27219	1397..28737	1321.30981	1330.55318
$y_t - \hat{y}_t$	-14.51840	-20.390070	-12.55406	62.36045	53.12421
$y_t - y_{ht}$	-14.76914	-20.77721	-12.79239	63.18517	53.94189

In table (1), the errors in estimation for $\hat{y}_t$ are consistently smaller than the errors for y_{ht} .However; the difference in the errors is so small. From Fig (1), the ratio $\text{var}(\hat{y}_t)/\text{var}(y_{ht})$ tends to a constant, which is less than 1but very close to 1. This indicates that $\hat{y}_t$ has a smaller variance than y_{ht} .

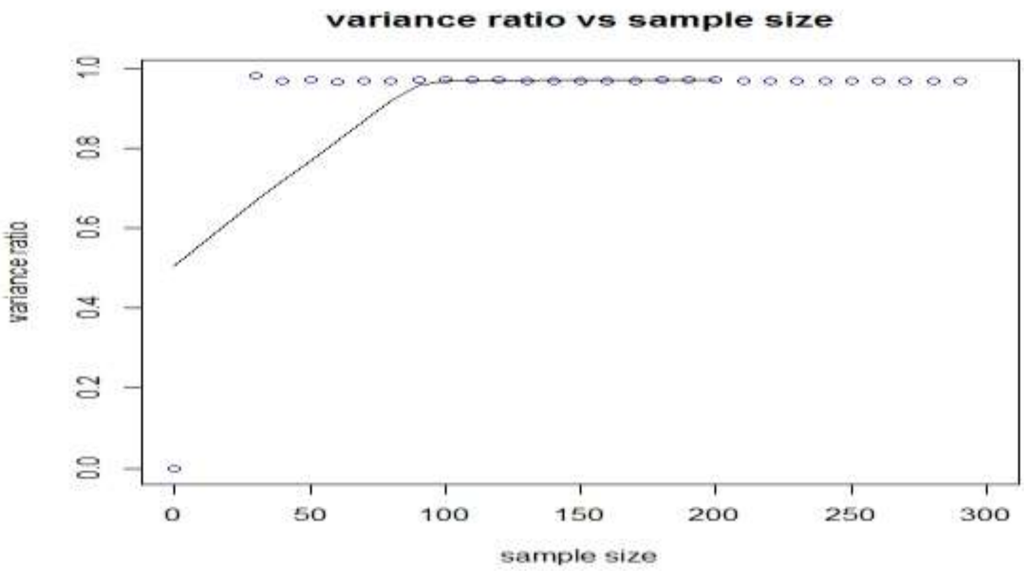

Fig 1: Variance Ratio $\text{var}(\hat{y}_t)/\text{var}(y_{ht})$

4.2. Results for Nonparametric Estimator $\hat{y}_{np}$

We let $y_t - \hat{y}_{np}$ and $y_t - y_{ht}$ be the errors in the estimation.

Table 2: Estimates and the errors for $\hat{y}_{np}$ and y_{ht}					
sample number	1	2	3	4	5
sample size n	100	100	100	100	100
y_t	1344.531793	1344.531793	1344.531793	1344.531793	1344.531793
$\hat{y}_{np}$	1344.875876	1342.336025	1332.540571	1347.019506	1347.815665
y_{ht}	1345.813667	1340.217041	1323.345573	1349.288456	1350.717785
$y_t - y_{np}$	-0.344083	2.195768	11.991222	-2.487713	-3.283872
$y_t - y_{ht}$	-1.281874	4.314752	21.186220	-4.756663	-6.185992

From table (2), both the estimators $\hat{y}_{np}$ and y_{ht} have small errors in estimation and consistently, $\hat{y}_{np}$ has smaller errors than y_{ht} . This is expected because the data is linear and $\hat{y}_{np}$ is obtained from a linear local polynomial model. That is, the nonparametric model is correctly specified for the data. From figure (2), as the sample size increases, the ratio $\text{var}(\hat{y}_{np})/\text{var}(y_{ht})$ stabilizes almost to a constant between 0.3 and 0.4, That is, $\hat{y}_{np}$ has a lower variance than y_{ht} as is expected since $\hat{y}_{np}$ is correctly specified for the data.

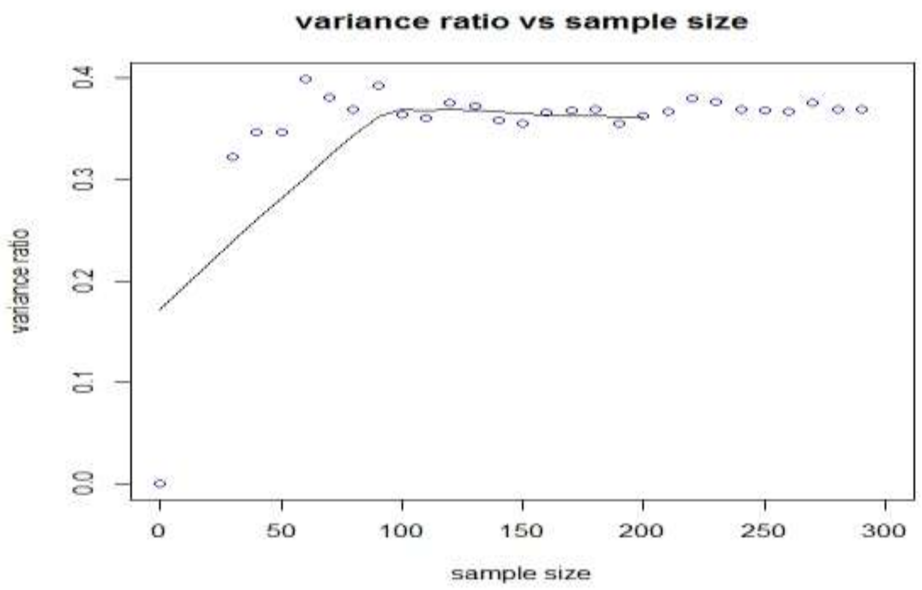

Figure 2: Variance Ratio $\text{var}(\hat{y}_{np})/\text{var}(y_{ht})$

5. Conclusion

We conclude that the estimator $\hat{y}_t$ is more accurate than the Horvitz Thompson design estimator y_{ht} since $\hat{y}_t$ yields smaller errors in estimation than does y_{ht}. Also, $\hat{y}_t$ has lesser variance than y_{ht}. Since y_{ht} is considered to be a very reliable design estimator, we conclude that $\hat{y}_t$ is also quite reliable.

We conclude that when the nonparametric model is correctly specified for the data, the nonparametric estimator $\hat{y}_{np}$ is quite accurate, more than the Horvitz Thompson design estimator y_{ht}. A comparison of the performance of both population total estimators $\hat{y}_t$ and $\hat{y}_{np}$ when penalty function is cubic and when the function is quadratic as shown in Kihara[4] and Kihara[5], shows that cubic penalty functions yields more efficient estimators than quadratic penalty functions.

References

[1] Breidt, F.J., & Opsomer, J.D. (2000). Local Polynomial Regression Estimation in Survey Sampling, *Annals of Statistics, Vol* 28, *No 4*, 1026-1053.
[2] Deville, J.C. & Sarndal C.E. (1992). Calibration Estimators in Survey Sampling, *Journal of the American Statistical Association, Vol* 87, 376-82.
[3] Frank E.C. Jorge N. (2007). Flexible Penalty Functions for Nonlinear Constrained Optimization, *IMA Journal of Numerical Analysis*
[4] Kihara, P.N. (2017). Calibration Estimators by Penalty Function Method. *Mathematical Theory and Modeling, Vol* 7, No 6, 22-32.
[5] Kihara, P.N. (2017). Robust Nonparametric and Semiparametric Model Calibration Estimators by Penalty Function Method. *Mathematical Theory and Modeling, Vol* 7, *No 8*, 22-39.
[6] Ozgur Y. (2005). Penalty Function Methods for Constrained Optimization with Generic Algorithms, *Mathematical and Computation Applications, Vol 10, No 1*, 45-56
[7] Rao S.S. (1984). Optimization Theory and Applications, *Wiley Eastern Limited*
[8] Thompson M.E. (1997). Theory of Sample Surveys, *Chapman Hall*, *London*
[9] Wu, C, & Sitter, R.R. (2001). A Model Calibration Approach to Using Complete Auxiliary Information from Survey Data, Journal *of American Statistical Association* , *Vol* 96, 185-93.