Journal of Modern Applied Statistical Methods

Volume 9 | Issue 1 Article 19

5-1-2010

Nonlinear Parameterization in Bi-Criteria Sample Balancing

Stan Lipovetsky
GfK Custom Research North America, Minneapolis, MN, stan.lipovetsky@gfk.com

Journal of Modern Applied Statistical Methods
May 2010, Vol. 9, No. 1, 198-208

1538 – 9472/10/$95.00

Nonlinear Parameterization in Bi-Criteria Sample Balancing

Stan Lipovetsky
GfK Custom Research North America

Sample balancing is widely used in applied research to adjust a sample data to achieve better correspondence to Census statistics. The classic Deming-Stephan iterative proportional approach finds the weights of observations by fitting the cross-tables of sample counts to known margins. This work considers a bi-criteria objective for finding weights with maximum possible effective base size. This approach is presented as a ridge regression with the exponential nonlinear parameterization that produces nonnegative weights for sample balancing.

Key words: Sample balance, ridge regression, nonlinear parameterization.

Introduction

Sample balance method was introduced by Deming and Stephan (1940). It is also known in terms of raking or post-stratification, and it is widely used in applied research to adjust sample data to the known proportions in the population. Chi-squared criterion is applied to adjust the counts' contingency table to the needed margins (Stephan, 1942; Deming, 1964), which yields the weights for observations. The classic method has been developed in numerous approaches (Ireland & Kullback, 1968; Darroch & Ratcliff, 1972; Holt & Smith, 1979; Feinberg & Meyer, 1983; Little & Wu, 1991; Conklin & Lipovetsky, 2001; Bosch & Wildner, 2003; Kozak & Verma, 2006). The original technique has been further extended, particularly, in calibration and generalized regression (GREG) estimations (Deville & Sarndal, 1992; Sarndal, et al., 1992; Deville, et al., 1993; Sarndal, 1996; Chambers, 1996; Yung & Rao, 2000; Zhang, 2000; Singh, 2003).

Making a sample closer to the required margins, the weighting simultaneously reduces the effective base size of the data. The farther the sample cross-table subtotals are from the margins, the smaller is the effective base in comparison with the original sample size. Decreased effective base produces worse statistical test values and wider confidence intervals around the estimates which can be incorrectly identified as being insignificant. A problem of simultaneous sample balancing with maximization of the effective base was considered in Lipovetsky (2007a), and the solution was obtained in a ridge regression approach (Hoerl & Kennard, 1970, 1988; Lipovetsky, 2006, 2010). Changing the profile ridge parameter yields a better fit of the margins, or a higher effective base, and the trade-off between them is needed: For small ridge parameters corresponding to a better margins fit, some weights could get negative values which are hardly acceptable for applied research.

This article shows how to improve the weights estimation and how to obtain always positive values via nonlinear parameterization for the weights. This approach is presented in the nonlinear optimizing technique for a complex objective and can be reduced to iteratively re-weighted Newton-Raphson procedure (Becker & Le Cun, 1988; Arminger, et al., 1995; Hastie & Tibshirani, 1997; McCullagh & Nelder, 1997; Bender, 2000; Lipovetsky, 2006, 2007b, 2009a,b). The exponential, quadratic and logit parameterizations of the weights are tried. The exponential function is the most convenient for obtaining always nonnegative weights.

Stan Lipovetsky is Senior Research Director at GfK Custom Research North America in Minneapolis, MN. Email him at: stan.lipovetsky@gfk.com.

Methodology

Sample Balancing and Maximizing Effective Base

Let the data be presented in a matrix X of N by n order with elements x_{ij} for an i^{th} observation (i = 1, 2, …, N – number of observations) and a j^{th} variable x_j (j = 1, 2, …, n – number of variables). Besides the design matrix X, the required margins are given (census or other totals). Consider k_j bins of given margins for each variable x_j, so all the margins can be presented in a vector y of m^{th} order, where

$$m = \sum_{j=1}^{n} k_j \ .$$

Let the variable x_j be measured in the k_j point scale, or the values of x_j are segmented into k_j bins corresponding to the given margins. Each x_j can be categorized by k_j levels, and presented by a set of k_j binary variables. The whole set of these variables can be incorporated into a matrix Z of N by m order. The columns of Z present binary variables z_p with 0-1 values of the elements z_{ip} (p = 1, 2, …, m). The matrix Z is singular, because the rank of a matrix of categorized binary variables is not higher than m-n.

Deming-Stephan sample balancing consists in fitting the counts n_l in the cross-table (indexed as l = 1, 2, …, L) of Z matrix by the theoretical counts v_l in Chi-squared criterion

$$\chi^2 = \sum_{l=1}^{L} \frac{(n_l - v_l)^2}{n_l}, \qquad (1)$$

restricted by the conditions of equivalence of the sample adjusted totals by each variable to the given margins. Adding these restrictions to the objective (1) and minimizing such a conditional objective by the theoretical parameters v_l yields a solution for the weights w_i which can be reached in the algorithm of iterative proportional fitting. Total of the weights equals the sample base, or the weights can be normalized by the relation:

$$\sum_{i=1}^{N} w_i = N \ . \qquad (2)$$

With the weights w_i obtained the Deming-Stephan sample balance procedure, the effective base size of the weighted sample is evaluated by the expression:

$$EB = \left(\sum_{i=1}^{N} w_i \right)^2 / \sum_{i=1}^{N} w_i^2 = N^2 / \sum_{i=1}^{N} w_i^2 \ , \qquad (3)$$

where the last equality holds only for the normalized weights (2). When the weights are distributed more evenly, closer to 1, the effective base is close to (but always below) the original sample size. Adding and subtracting the constant of the base size, the effective base for any set of weights can be represented as follows:

$$\begin{aligned} \mathrm{EB} &= \mathrm{N} + \frac{\left(\sum_{i=1}^{N} w_i \right)^2}{\sum_{i=1}^{N} w_i^2} - \mathrm{N} \\ &= \mathrm{N} - \mathrm{N} \frac{\sum_{i=1}^{N} w_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} w_i \right)^2}{\sum_{i=1}^{N} w_i^2} \qquad (4) \\ &= \mathrm{N} \left(1 - \frac{\sum_{i=1}^{N} (w_i - \bar{w})^2}{\sum_{i=1}^{N} w_i^2} \right) \end{aligned}$$

where $\bar{w}$ is the mean value of the weights. For all weights equal one their mean is $\bar{w} = 1$, so the effective base equals the sample size. Minimization of the weights deviation from their mean corresponds to finding the most effective base (4).

Sample Balancing with Maximum Effective Base

Based on Lipovetsky (2007a), the relation between the given vector of margins y and theoretical $\hat{y}$ vector of margins is presented in a simple linear model:

$$y = \hat{y} + \varepsilon = Z'w + \varepsilon \ . \tag{5}$$

The theoretical vector $\hat{y} = Z'w$ is estimated by the weighted binary variables (prime denotes transposition), where w is the N^{th} order vector-column of unknown weights w_i, and ε is a vector of deviations between the given and theoretical margins. The model (5) reminds an ordinary linear regression – however, with the number N of the unknown coefficients w_i significantly larger than the number m of the values by the dependent variable of margins y. Chi-squared criterion can be applied directly to minimizing the deviations ε in (5) by fitting the given margins with the weighted binary data:

$$\chi^2 = \sum_{p=1}^{m} \frac{\left(y_p - \hat{y}_p\right)^2}{\tilde{z}_p} = \sum_{p=1}^{m} \left(\frac{1}{\tilde{z}_p}\right)\left(y_p - \left(Z'w\right)_p\right)^2 . \tag{6}$$

The notation $\hat{y}_p$ is used for the elements of the theoretical vector $\hat{y}$ (5), and $\tilde{z}_p$ in the denominator (6) are the total counts of the binary variables in the columns of matrix Z, so they are the elements of the vector of m^{th} order $\tilde{z} = Z'1_N$, where 1_N denotes a uniform vector-column of size N.

Simultaneous minimization of the Chi-squared criterion (6) and the efficient variance of the weights in (4) can be achieved by the conditional objective:

$$\begin{aligned} F &= \chi^2 + q\,\mathrm{var}(w) \\ &= \sum_{p=1}^{m}\left(\frac{1}{\tilde{z}_p}\right)\left(y_p - \sum_{i=1}^{N} z_{ip} w_i\right)^2 + q\sum_{i=1}^{N}\left(w_i - 1\right)^2 \\ &= \left(y - Z'w\right)' D^{-1}\left(y - Z'w\right) + q\left(w - 1_N\right)'\left(w - 1_N\right), \end{aligned} \tag{7}$$

where q is Lagrange term, D and D^{-1} denote the m^{th} order diagonal matrix and its inversion defined via the total counts:

$$D = \mathrm{diag}(\tilde{z}) , \quad D^{-1} = \mathrm{diag}(1/\tilde{z}) . \tag{8}$$

The condition for minimization yields a system of linear equations:

$$\frac{\partial F}{\partial w'} = -2ZD^{-1}\left(y - Z'w\right) + 2q\left(w - 1_N\right) = 0 \tag{9}$$

which is a matrix equation:

$$\left(ZD^{-1}Z' + qI_N\right)w = ZD^{-1}y + q1_N , \tag{10}$$

For q close to zero this system corresponds to margins fit objective, and with q growing the main input comes from the efficient base objective with the solution of uniform weights. The equation (10) corresponds to the ridge regression system of equations with the profile parameter q. The regularization item qI_N added to the diagonal of the matrix in the left-hand side (10) guarantees that it becomes non-singular and invertible.

Solution of the system (10) is given in the work (Lipovetsky, 2007a), and can be presented explicitly as follows:

$$w = 1_N + Z\left(Z'Z + q\,diag(\tilde{z})\right)^{-1}(y - \tilde{z}) \tag{11}$$

Due to (11), the weights are distributed around 1, and depend on the difference of the given margins y and counts $\tilde{z} = Z'1_N$ by the categorized variables. For $y - \tilde{z} = 0$ all the weights are $w_i = 1$. A unit change $\Delta y_p = 1$ in a p^{th} component of the vector of margins leads to the weights change equal the elements $Z\left(Z'Z + q\,diag(\tilde{z})\right)^{-1}$ of the p^{th} column of the transfer matrix, which shows the rate of relaxation of the closeness to the given margins.

Variation in the parameter q permits a trade-off between better correspondence to the given margins versus more efficient weights of the higher effective base. Dividing the expression (4) by N yields a quotient EB/N of the effective to sample base, which is defined as one minus the ratio of the centered and non-centered weights' second moments:

$$R^2_{EB} = \frac{EB}{N} = 1 - \frac{\left(\sum_{i=1}^{N} (w_i - \bar{w})^2\right)}{\left(\sum_{i=1}^{N} w_i^2\right)}, \quad (12)$$

The expression (12) has a form of the coefficient of determination R^2 known in regression analysis, and demonstrates similar properties. If the residual sum of squares in the numerator at the right-hand side (12) is close to zero, R^2 is close to one, and the effective base reaches the sample base. It is convenient to introduce another coefficient of determination for the margins fitting in Chi-squared objective (6) which also is a weighted least squares objective:

$$R^2_{mrg} = 1 - \frac{\chi^2}{\chi^2_{orig}} = 1 - \frac{\left(\sum_{p=1}^{m} \left(\frac{1}{\tilde{z}_p}\right)\left(y_p - (Z'w)_p\right)^2\right)}{\left(\sum_{p=1}^{m} \left(\frac{1}{\tilde{z}_p}\right)\left(y_p - \tilde{z}_p\right)^2\right)}, \quad (13)$$

where the original value of the objective χ^2_{orig} is taken using the sample counts $\tilde{z}$. Both coefficients R^2_{EB} and R^2_{mrg} can be profiled by the parameter q for finding an acceptable level of adjustment to margins at a sufficiently large effective base.

Nonlinear Parameterization for Finding Nonnegative Weights

In practice researchers often encounter with the sample total counts too different from the assigned Census margins. Such a discrepancy can easily produce weights with negative values. In these cases the linear ridge-regression solution (11) requires to increase the parameter q high enough to reach all the weights non-negative. In the ridge regression it is not a problem, but at a price of losing the needed level R^2_{mrg} of margins fitting. To obtain positive weights a special parameterization for the weights can be used. For example, the positive weights can be presented by the exponent

$$w_i = \exp(v_i), \quad (14)$$

or the non-negative weights can be given by the quadratic dependence

$$w_i = (v_i)^2, \quad (15)$$

where v_i are the unknown parameters. The logistic parameterization is:

$$w_i = w_{min} + \Delta w \frac{1}{1 + \exp(-v_i)}, \quad \Delta w = w_{max} - w_{min} \quad (16)$$

where w_{min} and w_{max} are the given constants of the minimum and maximum values of the desired weights. For any v_i, the weights w_i always belong to the range from w_{min} to w_{max}.

Numerical minimization of the objective (7) by the parameters v_i of the positive weights can be efficiently performed by Newton-Raphson optimizing technique. Consider the Newton-Raphson algorithm for the objective (7) which can be approximated as:

$$F(v) \approx F(v^{(0)}) + \frac{\partial F}{\partial v}(v - v^{(0)}), \quad (17)$$

where $v^{(0)}$ is an initial approximation for the vector v which consists of the unknown parameters v_i. An extreme value of a function can be found from the condition of the first derivative equals zero, thus taking the derivative of (17) yields:

$$\frac{dF}{dv} = \frac{\partial^2 F}{\partial v^2}(v - v^{(0)}) + \frac{\partial F}{\partial v} = 0. \quad (18)$$

Solution of the equation (18) for the vector v is:

$$v = v^{(0)} - \left(\frac{\partial^2 F}{\partial v^2}\right)^{-1} \left(\frac{\partial F}{\partial v}\right) = v^{(0)} - H^{-1}\nabla F \;, \tag{19}$$

where a matrix of the second derivatives, or Hessian, is denoted as H, so H^{-1} is the inverted Hessian, and the vector of the first derivatives is the gradient ∇F. The obtained expression (19) is used in the iterations for finding each (t+1)-st approximation for the vector $v^{(t+1)}$ via the previous vector $v^{(t)}$ at the t^{th} step.

The first derivative of (7) by each parameter v_k is:

$$\frac{\partial F}{\partial v_k} = \left\{ \begin{array}{l} -2\sum_{p=1}^{m} \tilde{z}_p^{-1}\left(y_p - \sum_{i=1}^{N} z_{ip} w_i \right) z_{kp} \\ +2q(w_k - 1) \end{array} \right\} \frac{dw_k}{dv_k} \;, \tag{20}$$

which corresponds to the derivative in matrix form (9) multiplied by the derivative of each weight by its parameter. The second derivative by any two parameters (r and k, running by the observations i = 1, 2, …, N) is as follows:

$$\begin{aligned} \frac{\partial^2 F}{\partial v_r \partial v_k} &= 2\left(\sum_{p=1}^{m} \tilde{z}_p^{-1} z_{rp} z_{kp} + q\delta_{rk} \right) \frac{dw_r}{dv_r}\frac{dw_k}{dv_k} \\ &+ \left\{ \begin{array}{l} -2\sum_{p=1}^{m} \tilde{z}_p^{-1}\left(y_p - \sum_{i=1}^{N} z_{ip} w_i \right) z_{kp} \\ +2q(w_k - 1) \end{array} \right\} \frac{d^2 w_k}{dv_k^2} \delta_{rk} \;, \end{aligned} \tag{21}$$

where δ_{rk} is Kronecker delta. Hessian (21) in the braces contains an expression coinciding with that in braces of the first derivatives (20). The first derivative reaches zero at the optimum, therefore Hessian can be reduced to the first part (21) which in matrix notation is:

$$\begin{aligned} &H = 2G\,(ZD^{-1}Z' + qI_N)\,G \;, \\ &G = \mathrm{diag}\left(dw_i / dv_i \right) \end{aligned} . \tag{22}$$

All the notations in (22) are the same as in (5), (8)-(10), and G denotes the N^{th} order diagonal matrix of the weight derivatives by the parameters. Vector of the first derivatives (20) can be also represented in matrix notation as:

$$\nabla F = (-2)G\left(ZD^{-1}(y - Z'w) - q(w - 1_N) \right) . \tag{23}$$

Substituting the expressions (22)-(23) into (19) yields the expression for minimization the objective (7):

$$\begin{aligned} v &= v^{(0)} + G^{-1}\left(ZD^{-1}Z' + qI_N \right)^{-1} \begin{pmatrix} ZD^{-1}(y - Z'w) \\ -q(w - 1_N \end{pmatrix} \\ &= v^{(0)} + G^{-1}\left(ZD^{-1}Z' + qI_N \right)^{-1} \begin{pmatrix} (ZD^{-1}y + q1_N) \\ -(ZD^{-1}Z' + qI_N)w \end{pmatrix} \\ &= v^{(0)} + G^{-1}\left(ZD^{-1}Z' + qI_N \right)^{-1}\left(ZD^{-1}y + q1_N \right) - G^{-1}w \;. \end{aligned} \tag{24}$$

The second item in (24) contains the expression coinciding with the solution of the system (10) which can be denoted as linear solution, w_{lin}, given in explicit form in (11). The recurrent equation (24) for a t^{th} and the next steps of approximation can be represented as:

$$v^{(t+1)} = v^{(t)} + G^{-1}(w_{lin} - w^{(t)}) \;. \tag{25}$$

Formula (25) presents the iteratively re-weighted Newton-Raphson procedure for minimizing the objective (7) in a nonlinear parameterization, and it usually quickly converges.

For the exponential function (14), the inverted matrix of derivatives (22) is:

$$G^{-1} = diag\left(\exp(-v_i^{(t)})\right) = diag(1 / w_i^{(t)}) \;, \tag{26}$$

and for the quadratic function (15) it is:

$$G^{-1} = diag\left(1/(2v_i^{(t)})\right) = diag\left(1/(2\sqrt{w_i^{(t)}})\right). \tag{27}$$

For the logistic function (16) its diagonal matrix of the inverted derivatives is:

$$
\begin{aligned}
G^{-1} &= \mathrm{diag}^{-1}\left(\Delta w \frac{\exp(-v_i^{(t)})}{\left(1+\exp(-v_i^{(t)})\right)^2} \right) \\
&= \mathrm{diag}^{-1}\left(\Delta w \frac{w_i^{(t)} - w_{\min}}{\Delta w} \left(1 - \frac{w_i^{(t)} - w_{\min}}{\Delta w} \right) \right) \\
&= \mathrm{diag}\left(\frac{w_{\max} - w_{\min}}{(w_{\max} - w_i^{(t)})(w_i^{(t)} - w_{\min})} \right),
\end{aligned} \tag{28}
$$

where the constants $w_{\min}$ and $w_{\max}$ define the range Δw of the desired weights. With the initial parameters $v_i^{(0)} = 1/N$, finding the initial weights by the formulae (14)-(16), and the related G^{-1} matrix by the corresponding formulae (26)-(28), and applying them in (25), it is easy to obtain the next approximation for the parameters, then the nonnegative weights, and to continue the process until it converges.

Numerical Example

Data from a marketing research project of six hundred observations contains variables of gender (two values), income (three levels), age group (three levels), and region (four levels) – these categories are given in the first columns of Table 1. The next two columns in Table 1 present the margins observed in the data and required by Census. Within each variable, a total of the observed or the required margins equal one. For example, the gender splits to 35% and 65%, while it should contain 40% and 60% of males and females, respectively. The largest difference of the sample and population values can be observed by the age groups of 18-34 and 54-65 years old respondents, and by Midwest and West regions.

The next column in Table 1 presents the results of the Deming-Stephan iterative proportional fitting (corresponds to the ridge parameter q = 0). All proportions are reached, thus, the fitted margins coincide with the required ones in Table 1 and the coefficient of determination R^2_{mrg} (13) equals one. However, the coefficient of determination R^2_{EB} (12) for the effective sample size equals 0.15, so the effective base is reduced by 85% from the sample of 600 observations to the effective base of only 90 observations, which is somewhat low. Descriptive statistics for the obtained weights are given in the last three rows of this column: they show that the weights vary (around mean value equal one) in the wide range from the minimum (min = −1.91) to the maximum (max = 18.29), with the standard deviation (std = 2.42). These results are poor and having negative weights is inconvenient in applied research (most of statistical software modules require the weights to be nonnegative).

Several other columns in Table 1 present the results of the linear ridge regression solutions (11) with the parameter q running by step 0.25 up to 2.25. Increasing q results in a loss on the margins adjustment, but a win on the effective sample size. Beginning from q = 0.75, all the weights become positive and distributed in the narrower range (the standard error reduces twice), and the effective base grows to $R^2_{EB} = 0.38$, so it becomes more than twice as large in comparison with the results of q = 0. Further increasing q to 1.75, the coefficient of determination for margins and for effective sample size becomes equal to 0.60.

Table 2 presents the results of the exponential parameterization (25)-(26) for the nonnegative weights (14). In difference to linear estimation, the nonlinear approach yields only nonnegative weights with similar characteristics of the quality of margins fit and effective base. The other nonnegative parameterizations (15)-(16) produce similar results to the exponential fitting. The outcomes in the considered example are typical for sample balance with maximizing effective size and nonnegative parameterization for weights.

As mentioned for the formulae (12)-(13), the coefficients of determination R^2_{EB} and R^2_{mrg} can be profiled by the growing parameter q for finding a point of intersection between the declining curve of margins adjustment R^2_{mrg} and the rising curve R^2_{EB} of the sufficiently effective base (see Figure 1). Comparison of the coefficients of determination - R^2_{EB} and R^2_{mrg} - in Tables 1 and 2 and in Figure 1 show that the feasible solutions can be found in the range of q from 0.75 to 1.75.

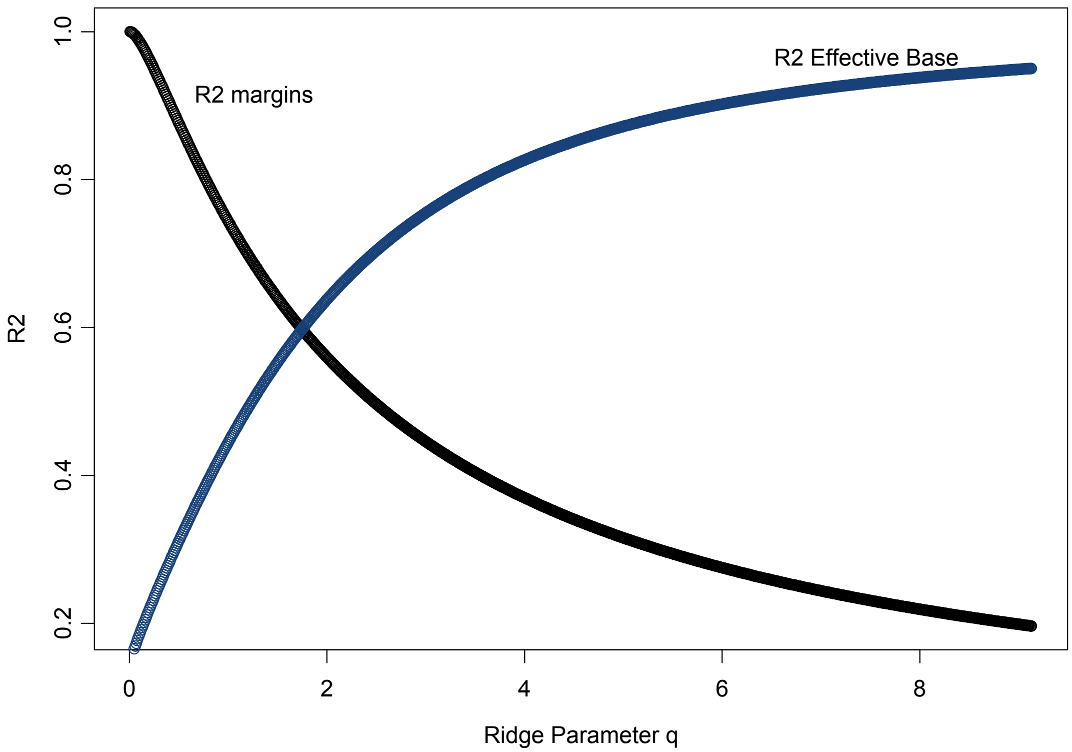
Table 1: Sample Balance with Maximum Effective Size: Linear Ridge Regression

Variable Category	Margins		Ridge Profile Parameter q									
	Observed	Census	0.00	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25
Gender Male	0.35	0.40	0.40	0.34	0.32	0.32	0.31	0.31	0.31	0.31	0.31	0.31
Gender Female	0.65	0.60	0.60	0.66	0.68	0.68	0.69	0.69	0.69	0.69	0.69	0.69
Income Low	0.44	0.48	0.48	0.40	0.37	0.36	0.35	0.35	0.34	0.34	0.34	0.34
Income Mid	0.49	0.43	0.43	0.52	0.55	0.57	0.58	0.59	0.59	0.60	0.60	0.61
Income High	0.07	0.09	0.09	0.09	0.08	0.08	0.07	0.07	0.06	0.06	0.06	0.06
Age 18-34	0.04	0.32	0.32	0.26	0.22	0.19	0.17	0.16	0.14	0.13	0.12	0.12
Age 35-54	0.41	0.40	0.40	0.39	0.40	0.40	0.40	0.40	0.41	0.41	0.41	0.41
Age 54-65	0.55	0.28	0.28	0.35	0.38	0.41	0.43	0.44	0.45	0.46	0.47	0.47
Region Midwest	0.19	0.34	0.34	0.31	0.28	0.26	0.25	0.24	0.23	0.22	0.22	0.21
Region West	0.29	0.13	0.13	0.22	0.27	0.30	0.32	0.34	0.36	0.37	0.38	0.39
Region South	0.35	0.33	0.33	0.27	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26
Region Northeast	0.17	0.20	0.20	0.20	0.19	0.18	0.17	0.16	0.16	0.15	0.15	0.14
Descriptive Statistics												
		R^2_{mrg}	1.00	0.95	0.88	0.81	0.74	0.69	0.64	0.60	0.56	0.53
		R^2_{EB}	0.15	0.23	0.31	0.38	0.44	0.50	0.55	0.60	0.64	0.67
		Min	-1.91	-0.60	-0.15	0.09	0.24	0.35	0.43	0.49	0.54	0.58
		Max	18.29	13.48	11.25	9.72	8.60	7.78	7.13	6.59	6.14	5.75
		Std	2.42	1.81	1.49	1.28	1.12	1.00	0.90	0.82	0.75	0.70

Table 2: Sample Balance with Maximum Effective Size: Linear Ridge Regression with Exponential Parameterization of the Coefficients

Variable Category	Margins		Ridge Profile Parameter q									
	Observed	Census	0.00	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25
Gender Male	0.35	0.40	0.36	0.33	0.32	0.32	0.31	0.31	0.31	0.31	0.31	0.31
Gender Female	0.65	0.60	0.64	0.67	0.68	0.68	0.69	0.69	0.69	0.69	0.69	0.69
Income Low	0.44	0.48	0.41	0.38	0.37	0.36	0.35	0.35	0.34	0.34	0.34	0.34
Income Mid	0.49	0.43	0.52	0.53	0.55	0.57	0.58	0.59	0.59	0.60	0.60	0.61
Income High	0.07	0.09	0.08	0.08	0.08	0.08	0.07	0.07	0.06	0.06	0.06	0.06
Age 18-34	0.04	0.32	0.27	0.25	0.22	0.19	0.17	0.16	0.14	0.13	0.12	0.12
Age 35-54	0.41	0.40	0.46	0.41	0.40	0.40	0.40	0.40	0.41	0.41	0.41	0.41
Age 54-65	0.55	0.28	0.27	0.34	0.38	0.41	0.43	0.44	0.45	0.46	0.47	0.47
Region Midwest	0.19	0.34	0.29	0.30	0.28	0.26	0.25	0.24	0.23	0.22	0.22	0.21
Region West	0.29	0.13	0.23	0.25	0.28	0.30	0.32	0.34	0.36	0.37	0.38	0.39
Region South	0.35	0.33	0.28	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26
Region Northeast	0.17	0.20	0.19	0.19	0.19	0.18	0.17	0.16	0.16	0.15	0.15	0.14
Descriptive Statistics												
		R^2_{mrg}	0.96	0.93	0.87	0.81	0.74	0.69	0.64	0.60	0.56	0.53
		R^2_{EB}	0.21	0.25	0.32	0.38	0.44	0.50	0.55	0.60	0.64	0.67
		Min	0.00	0.00	0.00	0.09	0.24	0.35	0.43	0.49	0.54	0.58
		Max	15.48	13.02	11.15	9.72	8.60	7.78	7.13	6.59	6.14	5.75
		Std	1.93	1.72	1.47	1.28	1.12	1.00	0.90	0.82	0.75	0.70

Figure 1: Profiling R2 for Margins and Effective Base Size



Conclusion

This article considers a sample balancing procedure with simultaneous maximization of the effective base size and the parameterization which guarantees the nonnegative weights. A multiple criteria objective is reduced to a ridge regression model (10). The analytical linear solution for the weights (11) is generalized to the nonlinear parameterization of the weights by exponential and other functions (14)-(16). To obtain always nonnegative weights, solution of the nonlinear system of equations is considered in the Newton-Raphson iteratively re-weighted procedure (17)-(28). The suggested weighting scheme is optimal for finding the best margins adjustment with the best effective base size. With growth of the ridge profile parameter *q*, the margins fit (13) decreases while the effective base (12) increases, thus a trade-off between them is used. The suggested approach can serve in solving various practical and theoretical problems involving sample balance for nonnegative weights. For example, the described method can be applied to solving calibration problem for data obtained by different sources or in international market research. The data gathered in several countries by various attributes measured in ordinal scales can be skewed to higher or lower levels due to the cultural norms and specifics dissimilar in different countries. To render the data samples comparable for statistical research one country can be taken as a basic pattern, Census likewise, and its counts of the response distribution can be found by the attributes levels. Fitting each other country distribution to the basic one can be performed exactly by the sample balance procedure which yields a solution for weighting the adjusting data with positive weights.

References

Arminger, G., Clogg, C. C., & Sobel, M. E., Eds. (1995). *Handbook of Statistical Modeling for the Social and Behavioral Sciences*. Plenum Press: New York.

Becker, S., & Le Cun, Y. (1988). Improving the convergence of back-propagation learning with second order methods. In: Touretzky, D. S., Hinton, G. E., & Sejnowski, T. J. (Eds.), *Proceedings of the 1988 Connectionist Models Summer School*, 29-37. San Mateo, CA: Morgan Kaufmann.

Bender, E. A. (2000). *Mathematical Methods in Artificial Intelligence*. Los Alamitos, CA: IEEE Computer Society Press.

Bosch, V., & Wildner, R. (2003). Optimum allocation of stratified random samples designed for multiple mean estimates and multiple observed variables. *Communications in Statistics: Theory and Methods*, *32*, 1897-1909.

Chambers, R. L. (1996). Robust case-weighting for multipurpose establishment surveys. *Journal of Official Statistics*, *12*, 3- 32.

Conklin, W. M., & Lipovetsky, S. (2001). Sample balancing made easy in a closed form solution. *International Journal of Operations and Quantitative Management*, *7*, 133-141

Darroch, J. N., & Ratcliff, D. (1972). Generalized iterative scaling for log-linear models. *Annals of Mathematical Statistics*, *43*, 1470-1480.

Deming, W. E. (1964). *Statistical Adjustments of Data*. New York:Dover.

Deming, W. E., & Stephan, F. F. (1940). On a least square adjustment of a sample frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*, *11*, 427-444.

Deville, J. C., & Sarndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, *87*, 376-382.

Deville, J. C., Sarndal, C. E., & Sautory, O. (1993). Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, *88*, 1013-1020.

Feinberg, S. E., & Meyer, M. M. (1983). Iterative proportional fitting. In Kotz, S., & Johnson, N. L., Eds., *Encyclopedia of Statistical Sciences*, *4*, 275-279.

Hastie, T. J., & Tibshirani, R. J. (1997). *Generalized Additive Models*. London: Chapman and Hall.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for non-orthogonal problems. *Technometrics*, *12*, 55-67 (republished in *Technometrics*, 2000, *42*, 80-86).

Hoerl, A. E., & Kennard, R. W. (1988). Ridge Regression. In Kotz, S., & Johnson, N. L., Eds., *Encyclopedia of Statistical Sciences*, *8*, 129-136.

Holt, D., & Smith, T. M. F. (1979). Post stratification. *Journal of the Royal Statistical Society*, *A*, *142*, 33-46.

Ireland, C. T., & Kullback, S. (1968). Contingency tables with given marginals. *Biometrika*, *55*, 179-188.

Kozak, M., & Verma, R. M. (2006). Geometric versus optimization approach to stratification: a comparison of efficiency. *Survey Methodology*, *32*, 157-163.

Lipovetsky, S. (2006). Two-parameter ridge regression and its convergence to the eventual pairwise model. *Mathematical and Computer Modelling*, *44*, 304-318.

Lipovetsky, S. (2007a). Ridge regression approach to sample balancing with maximum effective base. *Model Assisted Statistics and Applications*, *2*, 17-26.

Lipovetsky, S. (2007b). Priority and Choice Probability Estimation by Ranking, Rating, and Combined Data. *Journal of Statistical Theory and Practice*, *1*, 265-278.

Lipovetsky, S. (2009a). Linear regression with special coefficient features attained via parameterization in exponential, logistic, and multinomial–logit forms. *Mathematical and Computer Modelling*, 49, 1427-1435.

Lipovetsky, S. (2009b). PCA and SVD with Nonnegative Loadings. *Pattern Recognition*, 42, 68-76.

Lipovetsky, S. (2010). Enhanced ridge regressions, *Mathematical and Computer Modelling*, 51, 338-348.

Little, R. J. A., & Wu, M. M. (1991). Models for contingency tables with known margins when target and sampled population differ. *Journal of the American Statistical Association*, *86*, 87-95.

McCullagh, P., & Nelder, J. A. (1997). *Generalized Linear Models*. London: Chapman & Hall.

Sarndal, C. E., Swensson, B, & Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.

Sarndal, C. E. (1996). Efficient estimators with sample variance in unequal probability sampling. *Journal of the American Statistical Association*, *91*, 1289-1300.

Singh, S. (2003). *Advanced sampling theory with applications: How Michael 'selected' Amy*. The Netherlands: Kluwer.

Stephan, F. F. (1942). An iterative method of adjusting sample frequency tables when expected marginal totals are known. *Annals of Mathematical Statistics*, *13*(2), 166-178.

Yung, W., & Rao, J. N. K. (2000). Jackknife variance estimation under imputation for estimators using poststratification information. *Journal of the American Statistical Association*, *95*, 903-915.

Zhang, L. C. (2000). Post-stratification and calibration – a synthesis. *The American Statistician*, *54*, 178-184.