Post-Stratification with Optimized Effective Base: Linear and Nonlinear Ridge Regression Approach

Stan Lipovetsky
stan.lipovetsky@gfk.com
GfK Custom Research North America, 8401 Golden Valley Rd., Minneapolis, MN 55427

Abstract

Post-stratification, or sample balancing, or raking, is widely utilized in survey research to weighting a sample data to Census or other known population quotas. Cross-tables of counts are used in Deming-Stephan iterative proportional fitting to find the weights for adjusting data to known margins. A bi-criteria objective for finding weights with minimum variance yields a solution with maximum effective sample size. This model can be expressed as a ridge regression, which is applied to the original data, without its collapsing to cross-tables. Linear and nonlinear parameterization models are studied. The explicit regression solution allows to study the weighting analytically, that helps to interpret and improve the sample balance results.

Keywords: Sample balance, post-stratification, raking, weighting, generalized regression, ridge regression, nonlinear optimization, iteratively re-weighted least squares.

"A false balance is abomination to LORD, but a just weight is his delight" - Proverbs, 11:1.

I. Introduction

In the words of the originator of sample balancing, W. Edwards Deming: “In social and economic surveys that are carried out by sampling, … the sample is to be *adjusted* to certain totals that are known from other sources” (Deming, 1964). Sample balance, also known as raking, or post-stratification, is necessary in social and economic surveys because the obtained composition of the respondents' characteristics cannot be totally controlled to match adequately the known proportions in the population. Sample balance is often employed in panel comparisons, to ensure that observed effects are not due to demographic differences. After adjusting the data can be used in any statistical evaluations.

The method of sample balancing was introduced in (Deming and Stephan, 1940; Stephan, 1942), where the Chi-squared criterion and iterative proportional fitting were applied to adjust the counts' contingency table to the given desired margins. The method has been developed in various approaches (Ireland and Kullback, 1968; Darroch and Ratcliff, 1972; Holt and Smith, 1979; Feinberg and Meyer, 1983; Little and Wu, 1991; Conklin and Lipovetsky, 2001; Bosch and Wildner, 2003). The original technique has been further extended, particularly, in model-based regression and propensity scores (Rubin, 1979; Rubin and Zanutto, 2002; Judkins et al., 2005), in calibration and generalized regression (GREG) estimations (Deville and Sarndal,1992; Sarndal et al., 1992; Deville et al., 1993; Sarndal, 1996; Yung and Rao, 2000; Zhang, 2000; Singh, 2003; Beaumont and Alavi, 2004; Andersson and Thorburn, 2005).

In practical applications besides adjusting to the given margins the main concern consists in getting the effective sample size close enough to the base size of data. The farther are the sample cross-table subtotals from the required margins, the higher is inflation of particular segments of the weighted responses, and the smaller is the effective base in comparison with the unweighted base. The effective base is a useful criterion for assessment the goodness of weighting. It serves to reduce the likelihood of an applied statistics to yield significant results only because of the weighting. Decreased effective base leads to wider confidence intervals around the estimated parameters, they could be incorrectly identified as being insignificant, so it causes a reduction of statistical power in the ability of statistical tests to detect true differences in the population.

I consider a bi-criteria objective for sample balancing performed simultaneously with minimization of the weights' variance, so with maximum of the effective base. Applying Chi-squared criteria to such an objective yields a ridge regression solution (Hoerl and Kennard, 1988; Lipovetsky, 2006, 2007). Varying parameter of the ridge regression suggests a trade-off between a better fitting to margins and a higher effective base. Researcher can decide which level of adjustment and effective base satisfies the requirements of the problem in each particular case. For small values of the ridge parameter some of the weights could get negative values, but with the parameter increase all the weights become positive. It is possible to get all the positive weights at any stage of the multi-objective profiling via the nonlinear parameterization for the weights. Estimation is presented in iteratively re-weighted Newton-Raphson procedure (Becker and Le Cun, 1988; Bender, 2000; Bishop, 2006; Lipovetsky, 2006). Among several others the logistic parameterization of the weights is the most convenient and it permits to obtain the weights in any desired range of their values.

II. Sample Balancing by Cross-Table Counts

Consider a data matrix X of N by n order with the elements x_{ij} presenting an i-th observation (i =1, 2,…, N – number of observations) by a j-th variable x_j (j =1, 2,…, n – number of variables). Besides the design matrix X, the values of the desired margins are given (census, or other totals for population characteristics). Suppose there are k_j bins of margins for each variable x_j, so all the margins can be stacked into one vector y of size m:

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

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. Total number of the binary variables equals m (1), and 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.

Classical Deming-Stephan sample balance consists in fitting the counts n_l in the cross-table bins (indexed as l= 1, 2, …, L) of X matrix variables by the theoretical adjusted counts v_l using Chi-squared criterion

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

restricted by the conditions of equivalence of the sample adjusted totals by each variable to the given margins. Adding these restrictions to the objective (2) and minimizing such a conditional objective by the parameters v_l yields a solution which can be presented in an explicit closed form – more detail are given in (Conklin and Lipovetsky, 2001). The weights can be estimated in the algorithm of iterative proportional fitting as well. Total of the weights equals the sample base, or the weights can be normalized by the relation:

$$\sum_{i=1}^{N} w_i = N \ . \qquad (3)$$

With obtained weights w_i the effective base is evaluated as:

$$EB = \left(\sum_{i=1}^{N} w_i \right)^2 / \sum_{i=1}^{N} w_i^2 = N^2 / \sum_{i=1}^{N} w_i^2 \ , \qquad (4)$$

where the last equality holds only for the normalized weights (3). If the effective base is noticeably less than the sample size, $EB<<N$, it is possible to add a simple heuristic weights tuning that increases EB. The procedure consists in powering the weights $\tilde{w}_i = w_i^q$ (and normalizing them by (3)), with the parameter q diminishing from 1 with a small step. When q is decreasing this transformation makes the new weights $\tilde{w}_i$ to be distributed more evenly, closer to 1, which gives a lift to the effective base. However, because of this tuning the equalities of the fitted totals by each variable to the given margins alter into the approximate relations. In practice, a trade-off between a better fitting to margins and a higher effective base allows to find an appropriate solution satisfying the needs of the research.

Adding and subtracting the constant of the base size, the effective base for any set of weights can be represented as follows:

$$EB = N + \frac{\left(\sum_{i=1}^{N} w_i \right)^2}{\sum_{i=1}^{N} w_i^2} - N = N \left(1 - \frac{\sum_{i=1}^{N} (w_i - \overline{w})^2}{\sum_{i=1}^{N} w_i^2} \right) , \qquad (5)$$

where the mean value of the weights is:

$$\overline{w} = \frac{1}{N} \sum_{i=1}^{N} w_i \ . \qquad (6)$$

Dividing (5) 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. This quotient EB/N has a form similar to the coefficient of determination R^2 well-known in regression analysis, and shows similar properties. Particularly, if the residual sum of squares in the numerator at the right-hand side (5) is close to zero, this R^2 is close to one, or the effective base is reaching the sample base. For a set of weights satisfying (3) their mean (6) equals one, so the residual sum of squares in the numerator (5) is taken from $\overline{w} = 1$. Minimization of this residual sum of squares corresponds to finding the minimum variance *var(w)* estimator for the sample weighting.

III. Raking by Bi-Criteria Objective

As it was described above, a data consists of a matrix Z of N by m order of the binary categorized variables and a corresponding vector-column y of size m of the given counts of margins by all these categories. A vector-column w of unknown weights w_i of the base size N is the aim of the estimation. Relation between the given y and theoretical $\hat{y}$ margins is presented in a simple linear form:

$$y = \hat{y} + \varepsilon = Z' w + \varepsilon \ , \qquad (7)$$

where $\hat{y} = Z' w$ is a theoretical vector of margins estimated by the weighted binary variables, prime denotes transposition, and ε is a vector of deviations between the given and theoretical margins. The model (7) reminds an ordinary linear regression – however, with the number N of the coefficients w_i significantly bigger than the number m of the variables z_p, and than the number m of the values by the dependent variable of margins y.

Consider a Chi-squared criterion, similar to (2) but applied directly to minimizing the deviations ε (7) by fitting the given margins with the weighted binary data:

$$\chi^2 = \sum_{p=1}^{m} \frac{\left(y_p - \hat{y}_p\right)^2}{\tilde{z}_p} = \sum_{p=1}^{m} \left(\frac{1}{\tilde{z}_p}\right)\left(y_p - (Z'w)_p\right)^2 \quad (8)$$

which is a weighted least squares objective. The notation $\hat{y}_p$ is used for the elements of a theoretical vector $\hat{y} = Z'w$ (7). In the denominator (8) $\tilde{z}_p$ 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:

$$\tilde{z} = Z'1_N , \quad (9)$$

where 1_N denotes a uniform vector-column of size N. In Chi-squared criterion (8) the choice of elements in the denominator can be different, but the metric (9) has special optimal features considered farther.

Combined minimization for both the Chi-squared criterion (8) and the minimum variance of the weights (5) can be presented in the multi-criteria objective:

$$F = \alpha\chi^2 + \beta\,\mathrm{var}(w)$$
$$= \alpha \sum_{p=1}^{m} \left(\frac{1}{\tilde{z}_p}\right)\left(y_p - \sum_{i=1}^{N} z_{ip} w_i\right)^2 + \beta \sum_{i=1}^{N} (w_i - 1)^2 . \quad (10)$$

Here α and β are share parameters of the fitting margins and minimum variance objectives, respectively, and w_i are the unknown weights. Dividing in (10) by α and using one quotient parameter $q = \beta / \alpha$, yields this objective in the matrix form:

$$F = (y - Z'w)' D^{-1} (y - Z'w) + q(w - 1_N)'(w - 1_N) \quad (11)$$

where 1_N is the uniform vector as in (9), D and D^{-1} denote the m-th order diagonal matrix and its inversion defined via the total counts (9):

$$D = diag(\tilde{z}) , \quad D^{-1} = diag(1/\tilde{z}) . \quad (12)$$

The condition for minimization yields a system of linear equations:

$$\partial F / \partial w' = -2 ZD^{-1}(y - Z'w) + 2q(w - 1_N) = 0$$
(13)

which is a matrix equation:

$$\left(ZD^{-1}Z' + qI_N\right)w = ZD^{-1}y + q1_N , \quad (14)$$

where I_N denotes a uniform diagonal matrix of the N-th order. For q close to zero this system corresponds to fitting margins objective, and with q growing the system expresses the main input from the effective base with the solution of all weights equal one.

Up to the last item at the right-hand side (14), this system can be recognized as a ridge regression normal system of equations with the ridge parameter q. In contrast to the ridge regression with a covariance matrix of the order defined by the number of predictors, the system (14) contains the matrix $ZD^{-1}Z'$ of the N-th order defined by the number of observations. The rank of such a matrix $ZD^{-1}Z'$ can not be higher than m-n (1), and it is usually significantly lower than the number of the observations in the matrix Z. However, the regularization item qI_N added to its diagonal guarantees that the matrix in the left-hand side (14) becomes non-singular and invertible.

To solve the system (14), a matrix analogue of the Sherman-Morrison formula, also known as Woodbury identity (Harville, 1997; Lipovetsky and Conklin, 2005a; Bishop, 2006) can be applied:

$$\left(AA' + qI_N\right)^{-1} = \frac{1}{q} I_N - \frac{1}{q} A\left(A'A + qI_m\right)^{-1} A' , \quad (15)$$

where for the matrix (14) $A = ZD^{-1/2}$ and $A' = D^{-1/2}Z'$. The matrix I_m denotes a uniform diagonal matrix of the m-th order. The formula (15) presents inversion of the N-th order matrix via the inversion of the smaller m-th order matrix. Using (15) in (14) yields the solution of the matrix equation in the explicit form:

$$w = 1_N + \frac{1}{q} ZD^{-1} y$$
$$- \frac{1}{q} Z\left(Z'Z + qD\right)^{-1} Z'ZD^{-1} y - Z\left(Z'Z + qD\right)^{-1} \tilde{z} \quad (16)$$

where $\tilde{z}$ is the vector of counts (9). This formula can be easily used for finding weights, but it can be simplified farther. In the third item of the final expression (16) it is possible to use the transformation:

$$\left(Z'Z + qD\right)^{-1} Z'Z = I_m - q\left(Z'Z + qD\right)^{-1} D . \quad (17)$$

Substituting (17) into (16) and using (12) yields:

$$w = 1_N + q^{-1} ZD^{-1} y - q^{-1} ZD^{-1} y$$
$$+ Z\left(Z'Z + qD\right)^{-1} DD^{-1} y - Z\left(Z'Z + qD\right)^{-1} \tilde{z} \quad (18)$$

which reduces to the following expression:

$$w = 1_N + Z\left(Z'Z + q\,diag(\tilde{z})\right)^{-1} (y - \tilde{z}) . \quad (19)$$

It is the explicit presentation for the weights found in minimization of the objective (10)-(11).

Due to the solution (19) the weights are distributed around 1, and depend on the difference of the given margins y and counts $\tilde{z}$ by the categorized variables. For $y - \tilde{z} = 0$ all the weights are $w_i = 1$. For fitting only by the margins, without the effective variance, the parameter q=0, then a solution (19) exists if the matrix $Z'Z$ is non-singular, so the matrix Z of categorized binary variables should be arranged without redundant levels, to reduce its rank from m (1) to m-n. However, even with a small q close to zero, the solution (19) exists with a total Z matrix of all m

margins (1), and produces results practically coinciding with those obtained in the regular sample balancing (2).

As it was mentioned in relation to the Chi-squared criterion (8), the metric (9) has a special property that makes its choice to be very valuable – the solution (19) yields the total weights satisfying the relation (3). It guarantees that the mean weight (6) equals one, $\overline{w}=1$, so the effective variance in (10) is centered around the mean weight. Indeed, multiplying the transposed uniform vector $1'_N$ from the left by the normal system (14) yields:

$$1'_N ZD^{-1}Z'w + q1'_N w = 1'_N ZD^{-1}y + q1'_N 1_N \,. \quad (20)$$

Due to the definition (9), there is a vector-row $1'_N Z = \tilde{z}'$, so its product by the matrix D^{-1} (12) yields the uniform vector-row of the m-th order:

$$1'_N ZD^{-1} = \tilde{z}' diag(1/\tilde{z}) = 1'_m \,. \quad (21)$$

With (21) the relation (20) can be represented as:

$$1'_N w = \sum_{i-1}^{N} w_i = \frac{1}{q}\sum_{p=1}^{m}\left(y - Z'w\right)_p + N = N \,, \quad (22)$$

where the properties $1'_N 1_N = N$ and total $\varepsilon = y - Z'w$ of the deviations (7) equals zero (known in regression modeling) are used. Thus, the objective (10) does not need conditioning by (3), and its solution (19) does not need additional normalizing by (3). This feature is important because an additional normalizing by (3) needed for any other than (9) metric (for instance, of the uniform weights $\tilde{z}_p = 1$, when the Chi-squared (8) is reducing to the regular un-weighted least-squares objective) diminishes the effective base (5). By the solution (19), the centered second moment of the weights is:

$$\| w - 1 \|^2 = \| Z\left(Z'Z + q\, diag(\tilde{z})\right)^{-1}(y - \tilde{z}) \|^2 \,, \quad (23)$$

and variance of the weights equals this expression divided by number of observations.

The theoretical vector of margins (7) estimated by the weights (19) is:

$$\hat{y} = Z'w = Z'1_N + Z'Z\left(Z'Z + q\, diag(\tilde{z})\right)^{-1}(y - \tilde{z}) \,. \quad (24)$$

Using transformation (17) with definitions (9) and (12) the expression (24) can be reduces to:

$$\hat{y} = y - q diag(\tilde{z})\left(Z'Z + q\, diag(\tilde{z})\right)^{-1}(y - \tilde{z}) \,. \quad (25)$$

The vector of residuals (7) of the estimates from the given margins is defined by the second item (25), so the minimum of Chi-squared residual in the criterion (10)-(11) is:

$$\begin{aligned} \chi^2_{\min} &= \left(y - \hat{y}\right)' diag(\tilde{z}^{-1})\left(y - \hat{y}\right) \\ &= q^2\left(y - \tilde{z}\right)'\left(Z'Z + q diag(\tilde{z})\right)^{-1} \\ &\cdot diag(\tilde{z})\left(Z'Z + q diag(\tilde{z})\right)^{-1}\left(y - \tilde{z}\right). \end{aligned} \quad (26)$$

For the total objective (11), adding the Chi-squared (26) and second moment of the weights (23) with the factor q, yields the expression for the objective minimum:

$$\begin{aligned} F_{\min} &= \chi^2_{\min} + q \,\| w - 1 \|^2 \\ &= q\left(y - \tilde{z}\right)'\left(Z'Z + q diag(\tilde{z})\right)^{-1}\left(y - \tilde{z}\right). \end{aligned} \quad (27)$$

All the formulae (19)-(27) permit to consider analytically the sample balancing procedure and evaluate changes in a the characteristics due to the varying margins. For instance, differentiating weights (19) by the margins yields the estimate of the weights change due to a small change in the vector y:

$$\Delta w = Z\left(Z'Z + q\, diag(\tilde{z})\right)^{-1}\Delta y \,. \quad (28)$$

So a unit change $\Delta y_p = 1$ in a p-th component of the vector of margins inducts the weights change equal the elements of the p-th column of the transfer matrix $Z\left(Z'Z + q\, diag(\tilde{z})\right)^{-1}$. Also, using matrix spectral decomposition (Lipovetsky and Conklin, 2005b), it is easy to find the derivatives of the sample balancing characteristics due to the change in the parameter q. For instance, differentiating the weights (19) by q yields:

$$\Delta w = Z\left(Z'Z + q\, diag(\tilde{z})\right)^{-2} diag(\tilde{z})(y - \tilde{z})\Delta q. \quad (29)$$

The derivatives help to estimate the rate of relaxation of the closeness to the given margins with a simultaneous lift in the effectiveness of the weights variance, produced by the increasing parameter q. So q is a parameter of trade-off between better correspondence to the given margins versus more efficient weights of the higher effective base. As it was discussed above in relation to the effective base (5)-(6), the quotient EB/N of the effective to sample base can serve as a coefficient of determination:

$$R^2_{EB} = \frac{EB}{N} = 1 - \left(\sum_{i=1}^{N}(w_i - \overline{w})^2\right) / \left(\sum_{i=1}^{N} w_i^2\right), \quad (30)$$

which is reaching 1 for the most effective weights' variance. It is convenient to introduce another coefficient of determination for the margins fitting in Chi-squared objective (8) which also is a weighted least squares objective:

$$R^2_{mrg} = 1 - \frac{\chi^2}{\chi^2_{orig}} = 1 - \frac{\sum_{p=1}^{m}\left(\frac{1}{\tilde{z}_p}\right)\left(y_p - (Z'w)_p\right)^2}{\sum_{p=1}^{m}\left(\frac{1}{\tilde{z}_p}\right)\left(y_p - \tilde{z}_p\right)^2} \,, \quad (31)$$

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 effective base.

IV. Nonlinear Parameterization for Positive Weights

In practice researchers often encounter with the sample total counts too different from the assigned census margins. Such a discrepancy can easily produce some weights with negative values. In these cases the linear ridge-regression solution (19) 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 instance, the positive weights can be presented by the exponent:

$$w_i = \exp(v_i), \tag{32a}$$

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

$$w_i = (v_i)^2, \tag{32b}$$

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

$$w_i = w_{\min} + \Delta w \frac{1}{1+\exp(-v_i)}, \quad \Delta w = w_{\max} - w_{\min}, \tag{32c}$$

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 (10) by the parameters v_i of the positive weights can be efficiently achieved by the Newton-Raphson or another optimizing technique available in modern statistical packages. Consider the Newton-Raphson algorithm for the objective (10) which can be approximated as:

$$F(v) \approx F(v^{(0)}) + \frac{\partial F}{\partial v}(v - v^{(0)}), \tag{33}$$

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, so taking the derivative of (33) yields:

$$\frac{dF}{dv} = \frac{\partial^2 F}{\partial v^2}(v - v^{(0)}) + \frac{\partial F}{\partial v} = 0. \tag{34}$$

Solution of the equation (34) for the vector v is:

$$v = v^{(0)} - \left(\frac{\partial^2 F}{\partial v^2}\right)^{-1}\left(\frac{\partial F}{\partial v}\right) = v^{(0)} - H^{-1}\nabla F, \tag{35}$$

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 (35) 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 (10) by each parameter v_k is:

$$\frac{\partial F}{\partial v_k} = \Big\{ -2\sum_{p=1}^{m} \tilde{z}_p^{-1}(y_p - \sum_{i=1}^{N} z_{ip}w_i)z_{kp} + 2q(w_k - 1) \Big\} \frac{dw_k}{dv_k}, \tag{36}$$

that corresponds to the derivative in matrix form (13) 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:

$$\frac{\partial^2 F}{\partial v_r \partial v_k} = 2\left(\sum_{p=1}^{m} \tilde{z}_p^{-1} z_{rp} z_{kp} + q\delta_{rk}\right)\frac{dw_r}{dv_r}\frac{dw_k}{dv_k} + \left\{ \begin{matrix} -2\sum_{p=1}^{m} \tilde{z}_p^{-1}\left(y_p - \sum_{i=1}^{N} z_{ip}w_i\right)z_{kp} \\ +2q(w_k - 1) \end{matrix} \right\} \frac{d^2 w_k}{dv_k^2}\delta_{rk}, \tag{37}$$

where δ_{rk} is Kronecker delta. Hessian (37) in the figure parentheses contains an expression coinciding with that in figure parentheses of the first derivatives (36). The first derivative reaches zero at the optimum, so Hessian can be reduced to the first part (37) which in matrix notation is:

$$H = 2G(ZD^{-1}Z' + qI_N)G, \quad G = diag(dw_i / dv_i). \tag{38}$$

All the notations in (38) are the same as in (7), (12)-(14), and G denotes the N-th order diagonal matrix of the weight derivatives by the parameters. Vector of the first derivatives (36) can be also represented in matrix notation as:

$$\nabla F = (-2)G\left(ZD^{-1}(y - Z'w) - q(w - 1_N)\right). \tag{39}$$

Substituting the expressions (38)-(39) into (35) yields the expression for minimization the objective (10)-(11):

$$v = v^{(0)} + G^{-1}\left(ZD^{-1}Z' + qI_N\right)^{-1}\left(ZD^{-1}y + q1_N\right) - G^{-1}w. \tag{40}$$

The second item in (40) contains the expression coinciding with the solution of the system (14) that can be denoted as linear solution:

$$w_{lin} = \left(ZD^{-1}Z' + qI_N\right)^{-1}\left(ZD^{-1}y + q1_N\right). \tag{41}$$

In explicit form this solution is given in (19). Then the recurrent equation (40) for any t-th and the next steps of approximation can be represented as:

$$v^{(t+1)} = v^{(t)} + G^{-1}(w_{lin} - w^{(t)}), \qquad (42)$$

where the linear solution w_{lin} is known (41), and $w^{(t)}$ is a current vector of weights defined by one of the functions (32). Formula (42) presents the iteratively re-weighted Newton-Raphson procedure for minimizing the objective (10) in a nonlinear parameterization (32), and it usually quickly converges.

Consider application of the process (42) for a parameterization (32). For the exponential function (32a), the inverted matrix of derivatives (38) is:

$$G^{-1} = diag\left(\exp(-v_i^{(t)})\right) = diag(1/w_i^{(t)}), \qquad (43a)$$

and for the quadratic function (32b) it is:

$$G^{-1} = diag\left(1/(2v_i^{(t)})\right) = diag\left(1/(2\sqrt{w_i^{(t)}})\right). \qquad (43b)$$

For the logistic function (32c) its diagonal matrix of the inverted derivatives is:

$$G^{-1} = diag^{-1}\left(\Delta w \frac{\exp(-v_i^{(t)})}{\left(1+\exp(-v_i^{t)})\right)^2}\right) = diag\left(\frac{w_{\max} - w_{\min}}{(w_{\max} - w_i^{(t)})(w_i^{(t)} - w_{\min})}\right), \qquad (43c)$$

where the constants $w_{\min}$ and $w_{\max}$ define the range Δw of the desired weights. Beginning, for instance, with the parameters $v_i^{(0)} = 1/N$, finding the initial weights (32) and the related G^{-1} matrix (43), and applying them in (42) it is easy to obtain the next approximation for the parameters, then nonnegative weights, and to continue until the process converges.

V. Numerical example

A real marketing research project of 595 observations contains variables of gender (two values) and income (three values). This data is used in the cross-tabulation presented in Table 1. The sample row- and column- totals and their percents are presented too. In the last row and column of Table 1 the Census statistics for the demography of this study are given. The sample margins for gender (32% and 68%) should be made closer to the given 40% and 60%, while income sample proportions (33%, 64%, and 3%) should be pushed to the desired shares of 48%, 43%, and 9%, respectively. The first row of Table 2 presents results of the Deming-Stephan iterative proportional fitting (corresponds to q=0 in multi-criteria objective). The gender and income proportions are all adjusted (the fitted margins in Table 2 coincides with the given in Table 1), so the coefficient of determination R^2_{mrg} (31) equals one. However, the coefficient of determination R^2_{EB} (30) in the first row of Table 2 equals 0.66, or the effective base is 66% of the sample base, so the effective size is rather low. Descriptive statistics in the last columns of Table 2 show that obtained weights vary (around mean=1) in the wide range from *min*=0.39 to *max*=6.68, with the standard deviation *std* =0.71.

Table 1. Data Cross-tabulation and Given Margins.

Vari-able	**incom low**	**income middle**	**income high**	**row total**	**row total %**	**census margin %**
gender male	13	172	3	188	32	40
gender female	183	211	13	407	68	60
column total	196	383	16	595	100	100
column total %	33	64	3	100		
census margin %	48	43	9	100		

Applying the multi-objective (10)-(11) of adjusting to margins with gaining the optimal effective base produces results shown in the second and lower rows of Table 2, with the parameter q running from 0.05 to 0.70. Even for small values of q up to 0.10 the margins are very close to those given in Table 1, but the increase for the effective base is quite significant – up to $R^2_{EB} = 0.75$, or 75% of the sample size. Weights become distributed in the narrower range (around mean=1), and the standard error reduces to *std*=0.57. With farther increase of q to 0.15 and above, the margins are fitted with still a high value of the coefficient of determination R^2_{mrg}, although the structure of the income proportions becomes different from the required – the first share of low income befalls lower than the next one of the middle income share, while due to the Table 1 the first given income margin should be the highest one. Thus, if structure of the margins should correspond to the structure given in Table 1, then a feasible solution is defined by the multi-criteria objective parameter q equals 0.10, when the effective base comprises 75% of the sample size.

However, a researcher might be satisfied with the approximate structure of margins fitted, and prefer to get a higher effective base. In this case, the two 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. Comparison of the coefficients of determination R^2_{EB} and R^2_{mrg} by Table 2 show that these curves approximately intersect at the level q=0.45, where the effective base is about 86% of the sample base, and the weights are clustered closely to their mean value 1. For a high value q=0.70 the effective base can reach 90% of the sample base, when the standard error of the weights equals *std*=0.33, or twice less than its value in the regular sample balance.

Table 2. Sample Balance with Maximum Effective Size: Linear Ridge Model.

	Margins Fitted					R^2		Weights descriptive statistics		
	Gender %		Income %							
q	mal	fem	low	mid	hig	mrg	EB	min	max	std
0	40	60	48	43	9	1.0	.66	.39	6.7	.71
.05	39	61	47	44	9	1.0	.73	.33	3.9	.61
.10	38	62	46	45	9	.99	.75	.38	3.7	.57
.15	38	62	45	47	8	.97	.78	.42	3.6	.54
.20	37	63	45	47	8	.96	.79	.46	3.5	.51
.30	36	64	43	49	7	.92	.82	.51	3.2	.46
.40	36	64	43	50	7	.88	.85	.56	3.1	.42
.50	35	65	42	51	7	.84	.87	.60	2.9	.39
.60	35	65	41	52	7	.81	.89	.63	2.8	.36
.70	34	66	41	53	6	.77	.90	.66	2.7	.33

Consider now the same data with the gender margins taken as 70% and 30% (in place of the previous 40% and 60%). These margins are far from the sample counts, and produce results shown in the upper section of Table 3, (the part corresponding to Linear model) which is arranged similarly to Table 2. As Table 3 shows, for small q the linear ridge solution (19) yields some weights with negative minimum values, but with q above 0.30 all the weights become positive. Note that although for q=0 the solution in Table 3 is given with all positive weights, it is an adjusted solution that doesn't permit the negative weights, even at a price of large standard deviation and huge maximum value for weights. The lower section of Table 3 presents the Logit parameterization (32c) of the weights with the values $w_{\min} = 0$ and $w_{\max} = 2$, so all the weights are positive, around 1, and within the span from 0 to 2. Standard error in the logit model is smaller than in the linear model, and both of them reach the effective base above 70% in the vicinity of the profile parameter q=0.60. The results in the considered examples are typical for sample balance with maximizing effective size.

Table 3. Sample Balance with Maximum Effective Size: Linear and Logit Models.

	Margins Fitted					R^2		Weights descriptive statistics		
	Gender %		Income %							
q	mal	fem	low	mid	hi	mrg	EB	min	max	std
Lin										
0	70	30	48	43	9	1.0	.66	.10	13	1.8
.05	67	33	46	45	9	1.0	.45	-.34	5.1	1.1
.10	65	35	45	47	8	.98	.48	-.25	4.8	1.0
.15	63	37	44	48	8	.97	.51	-.17	4.6	.97
.20	62	38	43	49	8	.95	.54	-.10	4.5	.92
.30	59	41	41	51	7	.91	.60	.02	4.2	.82
.40	56	44	40	53	7	.86	.64	.12	3.9	.75
.50	55	45	39	54	7	.82	.68	.20	3.7	.68
.60	53	47	38	55	6	.78	.72	.26	3.5	.63
.70	51	49	38	56	6	.74	.75	.32	3.3	.58
Log										
.05	59	41	41	54	5	.85	.62	.00	1.9	.79
.10	59	41	41	54	5	.85	.62	.00	1.9	.79
.15	59	41	41	54	5	.86	.61	.00	1.9	.79
.20	59	41	41	53	5	.86	.61	.00	2.0	.79
.30	58	42	41	53	6	.86	.63	.02	2.0	.77
.40	56	44	40	55	6	.82	.67	.12	2.0	.70
.50	54	46	39	56	6	.78	.71	.20	2.0	.64
.60	52	48	38	56	5	.74	.74	.27	2.0	.60
.70	51	49	38	57	5	.71	.76	.33	2.0	.56

V. Summary

The work suggests a convenient sample balancing procedure with maximum effective sample size. The bi-criteria objective yields a ridge regression model (14). A simple analytical solution for the weights (19) is suggested. Solution is obtained by original data categorized to the binary variables corresponding to bins of the given margins. So the original variables can contain missing values, or any other non-relevant values to the margin bins – those levels in the categorized variables are redundant and skipped. The considered technique automatically yields the weights' mean equals one, if in the Chi-squared objective (8), or (10) the normalization by the totals in columns of the categorized variables is used. Other normalizing schemes produce a different from one mean value, that requires additional normalization, which in its turn decreases the effective size, and leads to a worse margins fitting for a needed effective base. The suggested weighting scheme is optimal for finding the best margins adjustment with the best effective base size. With growth of the profiling parameter q in the solution (19)-(29), the margins fit (31) is decreasing and the effective base (30) is increasing, so a trade-off between both criteria is utilized. To obtain always non-negative weights the non-linear parameterizations (32) are considered in Newton-Raphson iteratively re-weighted procedure (33)-(43). The logistic function (32c) has the best features, particularly, permitting to construct the weights within a desired range of the values. The considered approach can serve numerous practical applications as well as theoretical consideration of the sample balance problems.

References

Andersson P.G. and Thorburn D. (2005) An optimal calibration distance leading to the optimal regression estimator, *Survey Methodology*, 31, 95-99.

Beaumont J.F. and Alavi A. (2004) Robust generalized regression estimation, *Survey Methodology*, 30, 195-208.

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

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

Bishop C.M. (2006) *Pattern Recognition and Machine Learning*, Springer, New York.

Bosch V. and 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.

Conklin W.M. and 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. and 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*, Dover, New York

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

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

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

Harville D.A. (1997) *Matrix Algebra from a Statistician' Perspective*, Springer, New York.

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

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

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

Judkins D.R., Morganstein D., and Piesse A. (2005) Raking as a form of propensity scoring, *Proceedings of the Joint Statistical Meeting*, *JSM'05*, 2434-2441, Minneapolis, MN.

Lipovetsky S. and Conklin W.M. (2005a) Regression by data segments via discriminant analysis, *Journal of Modern Applied Statistical Methods*, 4, 63-74.

Lipovetsky S. and Conklin W.M. (2005b) Singular Value Decomposition in Additive, Multiplicative, and Logistic Forms, *Pattern Recognition*, 38, 1099-1110.

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

Lipovetsky S. (2007) Ridge Regression Approach to Sample Balancing with Maximum Effective Base, *Model Assisted Statistics and Applications*, 2, 17-26.

Little R.J.A. and 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.

Rubin D.B. (1979) Using multivariate matched sampling and regression adjustment to control bias in observational studies, *Journal of the American Statistical Association*, 74, 318-328.

Rubin D.B. and Zanutto E. (2002) Using matched substitutes to adjust for nonignorable nonresponse through multiple imputations. In: Groves R., Dillman D., Little R, and Eltinge J., eds., *Survey Nonresponse*, 389-402, Wiley, New York.

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

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*, Kluwer, The Netherlands.

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. and 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.