Biometrika (2002), **89,** 1, *pp.* 230–237

Printed in Great Britain

Using empirical likelihood methods to obtain range restricted weights in regression estimators for surveys

BY J. CHEN

Department of Statistics & Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1

jhchen@uwaterloo.ca

R. R. SITTER

Department of Statistics & Actuarial Science, Simon Fraser University, Burnaby, British Columbia, Canada, V5A 1S6

sitter@stat.sfu.ca

AND C. WU

Department of Statistics & Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1

cbwu@uwaterloo.ca

SUMMARY

Design weights in surveys are often adjusted to accommodate auxiliary information and to meet pre-specified range restrictions, typically via some ad hoc algorithmic adjustment to a generalised regression estimator. In this paper, we present a simple solution to this problem using empirical likelihood methods or generalised regression. We first develop algorithms for computing empirical likelihood estimators and model-calibrated empirical likelihood estimators. The first algorithm solves the computational problem of the empirical likelihood method in general, both in survey and non-survey settings, and theoretically guarantees its convergence. The second exploits properties of the model-calibration method and is particularly simple. The algorithms are adapted for handling benchmark constraints and pre-specified range restrictions on the weight adjustments.

Some key words: Benchmarking; Model calibration; Newton–Raphson.

1. INTRODUCTION

Design weights in surveys are often adjusted to accommodate auxiliary information and to meet pre-specified range restrictions. In particular, the resulting weights are often required to be nonnegative. Often, some ad hoc algorithms have to be used to apply a generalised regression estimator to meet these restrictions. In this paper, we present a simple algorithm for handling this problem using empirical likelihood methods or generalised regression. Our algorithms are simple and convergences are guaranteed. The pseudo-empirical maximum likelihood estimator (Chen & Sitter, 1999; Zhong & Rao, 2000) and the model-calibrated pseudo-empirical maximum likelihood estimator (Wu & Sitter, 2001) are asymptotically equivalent to the generalised regression estimator but with intrinsically positive weights. The simplicity of the algorithms and the theoretical support for the convergences make these two estimators serious competitors of existing methods such as those discussed in Huang & Fuller (1978), Deville & Särndal (1992), Fuller et al. (1994), Singh & Mohl (1996) and Rao & Singh (1997).

In § 2 we introduce the pseudo-empirical maximum likelihood and model-calibrated empirical likelihood estimators. In § 3, we describe the algorithm which solves the computational problem of empirical likelihood in general, not just in the survey context we focus on here. Some theoretical results are obtained to guarantee the convergence of the algorithm. A particularly simple and efficient algorithm is also given for the model-calibrated empirical likelihood estimator. A new simple method for handling general range-restrictions on weights, and adaptations of the algorithms for applying the method, are given in § 4. The resulting algorithms maintain their original simplicity and are guaranteed to converge. Some empirical results are reported in § 5. We conclude with some discussion in § 6.

2. Empirical likelihood methods in surveys

Consider a finite population consisting of N identifiable units. Associated with the ith unit are the study variable, y_i, and a vector of auxiliary variables, x_i. Let s be the set of sampled units. The values of $\{y_i, x_i, i \in s\}$ and the population mean of the x_i, $\bar{X} = N^{-1}\sum_{i=1}^{N} x_i$, are known. Assume the inclusion probabilities $\pi_i = \mathrm{pr}(i \in s)$ are strictly positive. We will restrict attention to estimating the population mean $\bar{Y} = N^{-1}\sum_{i=1}^{N} y_i$.

Often in sampling, estimators of $\bar{Y}$ have the form $\tilde{Y} = N^{-1}\sum_{i \in s} w_i y_i$. For example, the usual Horvitz–Thompson estimator is obtained by using basic design weights $w_i = d_i = \pi_i^{-1}$. There are a number of methods for adjusting these basic design weights to incorporate known $\bar{X}$, either for efficiency or to make survey estimates consistent with these known benchmarks:

$$N^{-1}\sum_{i \in s} w_i x_i = \bar{X}. \tag{2·1}$$

The generalised regression estimator is obtained if we use

$$w_i = d_i\left[1 + (\bar{X} - \bar{X}_{\mathrm{HT}})^{\mathrm{T}}\left\{N^{-1}\sum_{i \in s} d_i(x_i - \bar{x})(x_i - \bar{x})^{\mathrm{T}}\right\}^{-1}(x_i - \bar{x})\right], \tag{2·2}$$

where $\bar{X}_{\mathrm{HT}} = N^{-1}\sum_{i \in s} d_i x_i$, $\bar{x} = \sum_{i \in s} d_i^* x_i$ and $d_i^* = d_i / \sum_{i \in s} d_i$. The pseudo-empirical maximum likelihood estimator of Chen & Sitter (1999) is obtained by using

$$w_i = N\hat{p}_i = \frac{N d_i^*}{1 + \lambda^{\mathrm{T}}(x_i - \bar{X})}, \tag{2·3}$$

where the vector Lagrange multiplier, λ, is the solution to

$$g_1(\lambda) = \sum_{i \in s} \frac{d_i^*(x_i - \bar{X})}{1 + \lambda^{\mathrm{T}}(x_i - \bar{X})} = 0.$$

The model-calibrated empirical likelihood estimator introduced by Wu & Sitter (2001) reduces this high-dimensional computation problem to a scalar one. Under a linear working model, it is obtained using $w_i = N\hat{p}_i = N d_i^*/\{1 + \lambda(x_i - \bar{X})^{\mathrm{T}}\hat{\theta}\}$, in which the scalar Lagrange multiplier, λ, is the solution to

$$g_2(\lambda) = \sum_{i \in s} \frac{d_i^*(x_i - \bar{X})^{\mathrm{T}}\hat{\theta}}{1 + \lambda(x_i - \bar{X})^{\mathrm{T}}\hat{\theta}} = 0, \tag{2·4}$$

where $\hat{\theta}^{\mathrm{T}} = \sum_{i \in s} d_i(x_i - \bar{x})^{\mathrm{T}} y_i \{\sum_{i \in s} d_i(x_i - \bar{x})(x_i - \bar{x})^{\mathrm{T}}\}^{-1}$.

Both the generalised regression and pseudo-empirical maximum likelihood estimators satisfy the benchmark constraints in (2·1), while the model-calibrated empirical likelihood estimator satisfies a benchmark constraint formed by the optimal linear combination $\theta^{\mathrm{T}}x$ of x. All three estimators of $\bar{Y}$ are asymptotically equivalent and more efficient than the Horvitz–Thompson estimator. The weights of the pseudo-empirical maximum likelihood and model-calibrated empirical likelihood estimators are positive by definition.

3. ALGORITHMS FOR OBTAINING WEIGHTS IN PSEUDO-EMPIRICAL LIKELIHOOD AND MODEL-CALIBRATED EMPIRICAL LIKELIHOOD ESTIMATORS

Numerically, the key to the pseudo-empirical likelihood method is to find the vector solution of $g_1(\lambda)=0$ within the range of λ such that the resulting $\hat{p}_i>0$ for all i. A necessary and sufficient condition for the existence of the solution is that the convex hull of $\{x_i: i\in s\}$ contains $\bar{X}$ as an interior point. This condition is satisfied with probability approaching one for most sampling designs and conceivable populations as the sample and population sizes increase; see Chen & Sitter (1999) for details.

For notational simplicity and without loss of generality, assume that $\bar{X}=0$. If not, replace x_i by $x_i-\bar{X}$ throughout. If we let $\tilde{l}(\lambda)=\sum_{i\in s} d_i^* \log(1+\lambda^{\mathrm{T}}x_i)$, then $\tilde{l}(\lambda)$ is a concave function and its maximum point λ satisfies the same equation, $g_1(\lambda)=0$, and $\hat{p}_i$ given by (2·3) will satisfy $\hat{p}_i>0$, $\sum_{i\in s}\hat{p}_i=1$ and (2·1); that is, maximising $\tilde{l}(\lambda)$ is a dual problem of maximising $\hat{l}(p)$ subject to (2·1).

We first present an algorithm for finding the solution to $g_1(\lambda)=0$ and then show that the algorithm will always converge. We use $\|.\|$ for the Euclidean norm.

ALGORITHM

Step 0. *Let* $\lambda_0=0$. *Set* $k=0$, $\gamma_0=1$ *and* $\varepsilon=10^{-8}$.

Step 1. *Calculate* $\Delta_1(\lambda_k)=\partial\tilde{l}/\partial\lambda$ *and* $\Delta_2(\lambda_k)=\{\partial^2\tilde{l}/(\partial\lambda\,\partial\lambda^{\mathrm{T}})\}^{-1}\Delta_1(\lambda_k)$, *that is*

$$\Delta_1(\lambda)=\sum_{i\in s}\frac{d_i^* x_i}{1+\lambda^{\mathrm{T}}x_i},\quad \Delta_2(\lambda)=\left\{-\sum_{i\in s}\frac{d_i^* x_i x_i^{\mathrm{T}}}{(1+\lambda^{\mathrm{T}}x_i)^2}\right\}^{-1}\Delta_1(\lambda). \tag{3·1}$$

If $\|\Delta_2(\lambda_k)\|<\varepsilon$, *stop the algorithm and report* λ_k; *otherwise go to Step* 2.

Step 2. *Calculate* $\delta_k=\gamma_k\Delta_2(\lambda_k)$. *If* $1+(\lambda_k-\delta_k)^{\mathrm{T}}x_i\leqslant 0$ *for some* i *or* $\tilde{l}(\lambda_k-\delta_k)<\tilde{l}(\lambda_k)$, *let* $\gamma_k=\gamma_k/2$ *and repeat Step* 2.

Step 3. *Set* $\lambda_{k+1}=\lambda_k-\delta_k$, $k=k+1$ *and* $\gamma_{k+1}=(k+1)^{-1/2}$. *Go to Step* 1.

The above algorithm is similar to the modified Newton's method described in Polyak (1987, p. 63). Such algorithms for minimising a convex function or maximising a concave function almost always converge. Theoretically, however, some mild conditions are needed to guarantee the convergence. These amount to boundedness of the first derivative and of the inverse of the second derivative of the objective function and that the second derivative satisfies the Lipschitz condition. When these conditions are satisfied, a sufficiently short step-size can be determined so that the norm of the first derivative is always reduced after each iteration. These properties are established in Lemmas 1 and 2 below. All the proofs are deferred to the Appendix.

LEMMA 1. *Assume that* 0 *is an inner point of the convex hull of* $\{x_i: i\in s\}$ *and that* $\sum_{i\in s} d_i^* x_i x_i^{\mathrm{T}}$ *is positive definite. Let* $A_1=\{\lambda: 1+\lambda^{\mathrm{T}}x_i>0 \text{ for } i\in s\}$. *There exist constants* C *and* M, *depending on the particular set of* x_i *but not on* λ, *such that, for* $\lambda\in A_1$, (a) $\det|-\partial^2\tilde{l}/\partial\lambda\,\partial\lambda^{\mathrm{T}}|\geqslant C>0$, *and* (b) Δ_2 *given by* (3·1) *satisfies* $\|\Delta_2\|\leqslant M$.

Unfortunately, the Lipschitz condition is not satisfied here in the entire range of feasible λ. However, our algorithm is designed in such a way that the condition is practically satisfied. This is detailed in Lemma 2.

LEMMA 2. *Assume that the conditions in Lemma* 1 *are satisfied. Let* $A_2=\{\lambda:\tilde{l}(\lambda)\geqslant\tilde{l}(0)=0\}$. *Then there exists an* $L>0$ *such that, for any* λ_a, $\lambda_b\in A_2$,

$$\|\partial^2\tilde{l}/(\partial\lambda\,\partial\lambda^{\mathrm{T}})|_{\lambda=\lambda_a}-\partial^2\tilde{l}/(\partial\lambda\,\partial\lambda^{\mathrm{T}})|_{\lambda=\lambda_b}\|_\infty\leqslant L\|\lambda_a-\lambda_b\|,$$

where $\|.\|_\infty$ *is the largest absolute value of all the elements in the matrix.*

We are now able to show that the properties described in Lemmas 1 and 2 ensure the convergence of the algorithm. For simplicity of presentation, we will proceed as if λ is a scalar. The layout of the algorithm clearly guarantees that $\lambda_k\in A_2$ for all $k=0, 1, 2, \ldots$. Let c be the upper bound of

$\tilde{l}'(\lambda)/\tilde{l}''(\lambda)$. Then c is finite by Lemma 1(b). For each k such that $\lambda_{k+1}-\lambda_k=-(k+1)^{-1/2}\tilde{l}'(\lambda_k)/\tilde{l}''(\lambda_k)$, by the mean value theorem, we have

$$\tilde{l}'(\lambda_{k+1})=\tilde{l}'(\lambda_k)+\tilde{l}''(\xi_k)(\lambda_{k+1}-\lambda_k)=\tilde{l}'(\lambda_k)\left\{1-\frac{\tilde{l}''(\xi_k)}{(k+1)^{1/2}\tilde{l}''(\lambda_k)}\right\},$$

where ξ_k is between λ_k and λ_{k+1}. Since $\tilde{l}(\lambda)$ is a concave function and the set A_2 is closed and compact, we must have $\xi_k\in A_2$. By the Lipschitz condition, see Lemma 2, $|\tilde{l}''(\xi_k)/\tilde{l}''(\lambda_k)-1|\leqslant L|(\xi_k-\lambda_k)/l''(\lambda_k)|$. By Lemma 1(b), $|\xi_k-\lambda_k|\leqslant|\lambda_{k+1}-\lambda_k|\leqslant M(k+1)^{-1/2}$. Combining the above with Lemma 1(a), we get $|\tilde{l}''(\xi_k)/\tilde{l}''(\lambda_k)-1|\leqslant L^*(k+1)^{-1/2}$ with $L^*=LM/C$. Therefore, for large k,

$$|\tilde{l}'(\lambda_{k+1})|\leqslant|\tilde{l}'(\lambda_k)|\{1-(k+1)^{-1/2}+(k+1)^{-1}L^*\}\leqslant|\tilde{l}'(\lambda_k)|\{1-(2k)^{-1/2}\}.$$

This implies that, when k is large, a reduction in $|\tilde{l}'(\lambda_{k+1})|$ is guaranteed. As a result of concavity, $\tilde{l}\{\lambda_k-(k+1)^{-1/2}\Delta_k\}\geqslant\tilde{l}(\lambda_k)$ and $1+\{\lambda_k-(k+1)^{-1/2}\Delta_k\}x_i>0$ will be satisfied for all i and larger k. Since $\prod_{k=1}^{\infty}\{1-(2k)^{-1/2}\}=0$, we must have $|\tilde{l}'(\lambda_{k+1})|\to0$ as $k\to\infty$. This implies that the algorithm converges and $\tilde{l}(\lambda_k)$ converges to the global maximum.

While the step-size reduction procedure of having $\gamma_k=(k+1)^{-1/2}$ is typically needed to prove the convergence of the algorithm, it seems that the specific nature of the empirical likelihood calculation renders it unnecessary. An algorithm without this procedure worked well for our computations in § 5.

As for the model-calibrated empirical likelihood estimator, there exists an extremely simple algorithm for computing the weights. Note that $\hat{p}_i>0$ imply that the solution to $g_2(\lambda)=0$ is within the range of $(-U^{-1},-L^{-1})$, where $U=\max\{(x_i-\bar{X})^{\mathrm{T}}\hat{\theta},\,i\in s\}$, and $L=\min\{(x_i-\bar{X})^{\mathrm{T}}\hat{\theta},\,i\in s\}$. Then $g_2(\lambda)$ is a monotone decreasing function of λ in this interval. The solution can then be found using a bisection method.

4. General range restrictions on weights

In most cases it is desirable to place restrictions on the adjusted weights so as not to allow them to be too different from the basic design weights d_i. Suppose we wish to restrict the weights to the range $\gamma_1 d_i^*\leqslant w_i\leqslant\gamma_2 d_i^*$ where $0\leqslant\gamma_1<1<\gamma_2$; the case $\gamma_1=0$ and $\gamma_2=\infty$ represents a nonnegative weight restriction. If $\gamma_1=\gamma_2=1$, the $\bar{y}=\sum_{i\in s}d_i^*y_i$ will be the only one that meets the range restrictions. In this section we present a simple adaptation of the algorithms of § 3 for handling this problem. It is more direct than the methods of Huang & Fuller (1978) and Rao & Singh (1997), which require ad hoc adjustments at each iteration. Our method uses a minimum relaxation of benchmark constraints to meet the pre-specified range restriction requirement. The algorithm remains simple and a solution is guaranteed.

We only consider the pseudo-empirical maximum likelihood estimator, but the same idea works for the model-calibrated empirical likelihood and the generalised regression estimators, as shown in a technical report available from the authors. Without loss of generality, we assume that $\bar{X}=0$ and a solution to the pseudo-empirical likelihood method exists. The benchmark constraints (2·1) become $\sum_{i\in s}p_ix_i=0$. If the weights in (2·1) do not meet the range restrictions or if the convex hull of $\{x_i:i\in s\}$ does not contain $\bar{X}$, so that the benchmark constraints are unattainable, we relax the benchmark constraints by using $\sum_{i\in s}p_ix_i=t$ for some t that differs from 0. Note that a solution always exists if we choose $t=\bar{x}=\sum_{i\in s}d_i^*x_i$, which amounts to removing the benchmark constraints, and the solution, which is given by $p_i=d_i^*$, will automatically meet the range restrictions. Thus we can choose $t=\delta\bar{x}$ with the smallest possible δ such that the resulting weights meet the range restrictions; that is, we can move t away from 0 in the direction of $\bar{x}$. Since $0\leqslant\delta\leqslant1$, the simple and stable bisection method can be used to search for this δ. Note that this algorithm will always yield a solution. Also, it will find a point which represents the smallest possible departure from $t=0$, in the direction of $\bar{x}$.

Sometimes, some components of (2·1) are more important than others (Rao & Singh, 1997).

One would prefer to relax these components less. This translates into letting $t = (t_1, \ldots, t_k)^{\mathrm{T}} = (\delta_1 \bar{x}_1, \ldots, \delta_k \bar{x}_k)$, where $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_k)^{\mathrm{T}}$, and moving each of the δ_j from 0 towards 1 at different speeds which reflect their different importance. To do this, consider a small step of size dt from t towards $\bar{x}$. Then $dt_1/dt_2 = d\delta_1/d\delta_2 = (1-\delta_1)/(1-\delta_2)$ will represent the relative distance that components 1 and 2 of t move, or the relative speed at which the components are being moved. If, however, we feel that the second constraint is v times as important as the first constraint we might wish to reduce its speed by a factor of v, that is, to force $dt_1/dt_2 = v(1-\delta_1)/(1-\delta_2)$. If we enforce this at each point as we move t from 0 toward $\bar{x}$, the solution to the resulting differential equation is to let $\delta_1 = 1-(1-\delta_2)^v$. This translates into letting $t = (h_1(\delta)\bar{x}_1, \ldots, h_k(\delta)\bar{x}_k)^{\mathrm{T}}$, where $h_j(\delta) = 1-(1-\delta)^{v_j}$, and letting δ range from 0 to 1. Note that $h_j(0) = 0$ and $h_j(1) = 1$ for all $j = 1, \ldots, k$.

More specifically, order the constraints from most important, $j = 1$, to least important, $j = k$, and suppose that we wish the first $m < k$ constraints to remain fixed. Let $v_j = 0$ for $j = 1, \ldots, m$ and let $v_{m+1} = 1$. Define v_j for $j = m+2, \ldots, k$ to be the importance of constraint j relative to constraint $m+1$; that is, if constraint j is twice as important as constraint $m+1$, let $v_j = 2$. With this rephrasing of the problem one can now apply the above algorithm by merely replacing t by this new definition.

5. SOME EMPIRICAL RESULTS

We consider the 1996 Statistics Canada's Family Expenditure Survey for the province of Ontario. The data consist of 2396 sampled households and a variety of variables. As auxiliary variables we choose x_1, the number of children (age <15), x_2, the number of youths (age 15–24), x_3, the number of people in the household, and x_4, the total income after taxes. As response variables we use y_1, the annual expenditure on clothing, y_2, the annual expenditure on furnishings and equipment, and y_3, the total annual expenditure. We treat the dataset itself as the population in our study. This population is split into eight strata using the original design weights, and stratified random samples are taken from the resulting population.

For sample size $n = 40$, 80 and 160, under repeated sampling, about 35%, 13% and 4%, respectively, of the samples produce negative regression weights. The ranges of the weights for generalised regression and pseudo-empirical maximum likelihood vary dramatically from sample to sample. The g-weights, $g_i = w_i/d_i$, for generalised regression can be as low as $-3{\cdot}0$. The weights from pseudo-empirical maximum likelihood are intrinsically positive, but they sometimes contain a few very large values, exceeding 5·0, for $n = 40$. This may be undesirable as it places too much emphasis on a single observation. The range of the g-weights for the model-calibrated empirical likelihood estimator is extremely stable, typically between 0·5 and 2·0 and often in the range (0·8, 1·25).

Table 1 reports the relative bias as a percentage, $\%\mathrm{RB} = (B/\bar{Y}) \times 100$, and the relative efficiency to the Horvitz–Thompson estimator, $\mathrm{RE} = \mathrm{MSE}(\bar{Y}_{\mathrm{HT}})/\mathrm{MSE}$, where the bias, B, and the mean squared error, MSE, of each estimator are estimated from 1000 simulation runs with $n = 80$. Without range restrictions, $\%\mathrm{RB} < 2{\cdot}5$ for all three methods. In terms of RE, the methods all outperform $\bar{Y}_{\mathrm{HT}}$ similarly. When benchmark constraints are relaxed to meet range restrictions, that is $\gamma_1 = 0{\cdot}8$ and $\gamma_2 = 1{\cdot}25$, the relative biases remain small. The impacts on the generalised regression and pseudo-empirical maximum likelihood methods are much greater than on the model-calibrated empirical likelihood estimator.

We further examine the percentage relative relaxation of benchmark constraints, that is $(\sum_{i \in s} \hat{p}_i x_{ji} - \bar{X}_j)/\bar{X}_j$, for $j = 1, \ldots, k$. To do so, we considered a single simulated sample with two different range restrictions: (i) $\gamma_1 = 0{\cdot}5$ and $\gamma_2 = 2{\cdot}0$, and (ii) $\gamma_1 = 0{\cdot}8$ and $\gamma_2 = 1{\cdot}25$. The first case is milder and more typical, while the second is very restrictive. We assumed equal importance of the benchmark constraints for each of $x_1, \ldots, x_4$. Note that the more one relaxes the benchmark constraints the less efficient the estimator becomes as one moves towards ignoring the auxiliary information. Results are given in Table 2.

For both these cases the percent relative relaxation of the benchmark constraints is not extreme, with absolute value less than 13. For the mild constraints, the model-calibrated empirical likelihood

Table 1. *Statistics Canada data. Relative bias,* %RB, *and relative efficiency,* RE, *for* $n = 80$

	%RB			RE		
	GREG	PEML	MCPE	GREG	PEML	MCPE
	Without range restrictions					
y_1	−0·93	−1·25	−0·82	1·17	1·20	1·16
y_2	−2·43	−1·55	−1·91	1·11	1·06	1·09
y_3	0·00	0·04	0·02	1·68	1·68	1·65
	With range restrictions					
y_1	−0·86	−0·97	−1·13	1·10	1·10	1·16
y_2	−2·02	−2·06	−2·23	1·07	1·06	1·07
y_3	0·03	−0·04	−0·17	1·33	1·34	1·58

GREG, generalised regression estimator; PEML, pseudo-empirical maximum likelihood estimator; MCPE, model-calibrated pseudo-empirical maximum likelihood estimator.

Table 2. *Statistics Canada data. Percent relative relaxation of benchmark constraints*

	$\gamma_1 = 0{\cdot}5, \gamma_2 = 2{\cdot}0$					$\gamma_1 = 0{\cdot}8, \gamma_2 = 1{\cdot}25$				
	δ	x_1	x_2	x_3	x_4	δ	x_1	x_2	x_3	x_4
GREG	0·52	7·6	−4·5	−0·6	4·1	0·81	11·8	−7·0	−1·0	6·4
PEML	0·20	2·9	−1·7	−0·2	1·6	0·76	11·0	−6·6	−0·9	6·0
MCPE_1	0·00	3·8	−1·0	−4·5	−1·1	0·41	8·2	−4·1	−3·1	2·6
MCPE_2	0·00	12·0	0·5	−2·3	1·1	0·31	12·7	−2·5	−2·0	3·2
MCPE_3	0·00	10·0	−8·5	−3·5	1·2	0·18	11·3	−8·5	−3·0	2·4

GREG, generalised regression estimator; PEML, pseudo-empirical maximum likelihood estimator; MCPE_k, model-calibrated pseudo-empirical maximum likelihood estimator using response variable y_k, for $k = 1, 2, 3$.

estimator weights require no adjustment and the percent absolute relative relaxation of the benchmark constraints are all less than 4·5 for x_3 and x_4, but vary from 0·5 to 12·0 for x_1 and x_2. It is necessary to relax the benchmark constraints for the pseudo-empirical maximum likelihood method and more so for the generalised regression method, but in some cases to lesser extent than is inherent in the model-calibrated empirical likelihood estimator. When we tighten to $\gamma_1 = 0{\cdot}8$ and $\gamma_2 = 1{\cdot}25$, we obtain $\delta = 0{\cdot}76$ and $\delta = 0{\cdot}81$ for the pseudo-empirical maximum likelihood and generalised regression methods, respectively, and the percent absolute relative relaxation of the benchmark constraints of both become comparable to those of model-calibrated empirical likelihood estimator.

6. Discussion

Usually, both benchmark constraints and range restrictions on weights are desired. In most situations all existing methods no longer match benchmark constraints exactly and require an algorithm. With this in mind, the empirical likelihood-based methods, combined with the algorithms presented here, become attractive competitors.

Acknowledgement

This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada.

APPENDIX

Proofs

Proof of Lemma 1(a). It is easy to see that $\partial^2\tilde{l}/\partial\lambda\,\partial\lambda^{\mathrm{T}}$ is negative definite. Let ξ be a vector of unit length. Then $\sum_{i\in s} d_i^*(\xi^{\mathrm{T}}x_i)^2 \geqslant \tau_{(1)} > 0$, where $\tau_{(1)}$ is the smallest eigenvalue of $\sum_{i\in s} d_i^* x_i x_i^{\mathrm{T}}$. Furthermore, we have

$$\xi^{\mathrm{T}}\{-\partial^2\tilde{l}/(\partial\lambda\,\partial\lambda^{\mathrm{T}})\}\xi \geqslant \frac{\tau_{(1)}}{\{1+\max_{i\in s}(\lambda^{\mathrm{T}}x_i)\}^2}. \tag{A·1}$$

If we can show that

$$\sup\left\{\max_{i\in s}(\lambda^{\mathrm{T}}x_i) : \lambda\in A_1\right\} \leqslant M < \infty, \tag{A·2}$$

then the conclusion of the lemma follows from (A·1).

Let $\rho = \|\lambda\|$ and $\tilde{\lambda} = \lambda/\rho$. Since $\lambda\in A_1$, $1+\lambda^{\mathrm{T}}x_i = 1+\rho\tilde{\lambda}^{\mathrm{T}}x_i > 0$ for all $i\in s$. Meanwhile, since 0 is assumed to be an interior point of the convex hull of $\{x_i : i\in s\}$, $\min_{i\in s}(\tilde{\lambda}^{\mathrm{T}}x_i) < 0$. Since the set $\{\tilde{\lambda} : \|\tilde{\lambda}\| = 1\}$ is closed and compact, it follows that $\sup\{\min_{i\in s}(\tilde{\lambda}^{\mathrm{T}}x_i)\} \leqslant -c < 0$, where $c > 0$ is a constant, independent of $\tilde{\lambda}$. Hence, for each λ, $\rho \leqslant \{\max_{i\in s}(-\tilde{\lambda}^{\mathrm{T}}x_i)\}^{-1}$, and consequently $\rho \leqslant \sup\{\max_{i\in s}(-\tilde{\lambda}^{\mathrm{T}}x_i)\}^{-1} \leqslant c^{-1}$, from which (A·2) immediately follows. □

Proof of Lemma 1(b). Let

$$A = -\partial^2\tilde{l}/(\partial\lambda\,\partial\lambda^{\mathrm{T}}) = \sum_{i\in s} d_i^*\frac{x_i x_i^{\mathrm{T}}}{(1+\lambda^{\mathrm{T}}x_i)^2}, \quad B = \partial\tilde{l}/(\partial\lambda) = \sum_{i\in s} d_i^*\frac{x_i}{1+\lambda^{\mathrm{T}}x_i}.$$

Since $\sum_{i\in s} d_i^* = 1$, we could define a random vector Z such that $A = E(ZZ^{\mathrm{T}})$ and $B = E(Z)$. Note that, since $A - BB^{\mathrm{T}} \geqslant 0$, it follows that $\mathrm{tr}(A^{-1} - A^{-2}BB^{\mathrm{T}}) = \mathrm{tr}\{A^{-1}(A-BB^{\mathrm{T}})A^{-1}\} \geqslant 0$. Hence

$$\|\Delta_2\|^2 = B^{\mathrm{T}}A^{-2}B = \mathrm{tr}(A^{-2}BB^{\mathrm{T}}) \leqslant \mathrm{tr}(A^{-1}).$$

By (a), $\mathrm{tr}(A^{-1}) \leqslant M$ for some constant that is independent of λ. □

Proof of Lemma 2. Note that, by the assumed conditions, 0 is an interior point of the convex hull of $\{x_i, i\in s\}$. Note also that A_1 forms a bounded set with $\lambda = 0$ an interior point; see the proof of Lemma 1 for the bound on $\rho = \|\lambda\|$. Consider further the set A_2 of λ such that $\tilde{l}(\lambda) \geqslant \tilde{l}(0) = 0$. Note that $A_2 \subset A_1$, it is a closed set, and hence is compact. The third derivative of $\tilde{l}(\lambda)$ is obviously continuous on A_2 and hence has a finite upper bound, component-wise. Thus, its second derivative must have the property as stated. □

REFERENCES

CHEN, J. & SITTER, R. R. (1999). A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. *Statist. Sinica* **9**, 385–406.

DEVILLE, J. C. & SÄRNDAL, C. E. (1992). Calibration estimators in survey sampling. *J. Am. Statist. Assoc.* **87**, 376–82.

FULLER, W. A., LOUGHIN, M. M. & BAKER, H. D. (1994). Regression weighting in the presence of nonresponse with application to the 1987–1988 nationwide food consumption survey. *Survey Methodol.* **20**, 75–85.

HUANG, E. T. & FULLER, W. A. (1978). Nonnegative regression estimation for sample survey data. In *Proc. Social Statist. Sect., Am. Statist. Assoc.*, pp. 300–5. Washington, DC: American Statistical Association.

POLYAK, B. T. (1987). *Introduction to Optimization.* New York: Optimization Software, Inc. Publications Division.

RAO, J. N. K. & SINGH, A. C. (1997). A ridge-shrinking method for range-restricted weight calibration in survey sampling. In *Proc. Survey Res. Meth. Sect., Am. Statist. Assoc.*, pp. 57–65. Washington, DC: American Statistical Association.

SINGH, A. C. & MOHL, C. A. (1996). Understanding calibration estimators in survey sampling. *Survey Methodol.* **22**, 107–15.

Wu, C. & Sitter, R. R. (2001). A model-calibration approach to using complete auxiliary information from survey data. *J. Am. Statist. Assoc.* **96**, 185–93.

Zhong, C. X. B. & Rao, J. N. K. (2000). Empirical likelihood inference under stratified random sampling using auxiliary information. *Biometrika* **87**, 929–38.

[*Received July* 2000. *Revised August* 2001]