Calibration Estimation via a Smoothing Newton Method

WEIZHE GU
Department of Mathematics
School of Science, Tianjin University
Tianjin 300072
P. R. China
guweizhe@tju.edu.cn

JUN ZHAI
Department of Mathematics
Division of the Physical Sciences
The University of Chicago
Chicago IL 60615, USA
junzhai0516@uchicago.edu

WEI WANG
School of Public Administration
Tianjin University of Commerce
Tianjin 300134
P. R. China
wowei@tju.edu.cn

Abstract: Calibration estimation is currently the most popular method of estimation using auxiliary information. Its major idea is to use auxiliary information to structure calibration weights, attaching them to survey data, in order to improve the accuracy of the gross or mean estimation. Calibration estimation problem with box constraints is equivalently to solve a nonlinear equations system. Mnnich at el proposed a semismooth Newton method for solving this system in [Calibration of estimator-weights via semismooth Newton Method. J. Glob. Optim. 52 (2012):471-485]. In this paper, we give more specific analysis about the semismooth Newton method and some numerical experiments have been reported. On the basis of that, a smoothing Newton method has been proposed and proved to be globally convergent without any assumptions and locally superlinearly convergent under certain assumptions. Numerical results show that both semismooth Newton method and smoothing Newton method are effect for solving the calibration estimation problem.

Key–Words: Calibration estimation; Auxiliary information; Sample weights; Semismooth Newton method; Smoothing Newton method; Global convergence

1 Introduction

Calibration estimation is an estimation method combining survey sampling theory and auxiliary information. Meanwhile, it has an important role in research dealing with the problem of non-response, sample rotation and so on. At present, there are many statistical softwares to calculate calibration weights with both advantages and disadvantages. In the practice of statistical agencies at all levels, it increasingly reflects that large-scale statistical estimation problems, which can not be resolved timely and effectively. The core content of calibration estimation method is to solve the calibration weights, which is an optimization process. This provides an opportunity to introduce optimize methods to efficiently and effectively for solving the calibration estimation problem.

Weighting is a common methodology in survey statistics to estimate the population value according to data from sampling [1]. This thought origins from the research of Hansen and Hurwitz (1943) [2], Horvitz and Thompson (1952) [3]. They use the reciprocal of the inclusion probabilities from a survey as weights to estimate the population value. Gradually, many weighting methods and estimators emerge, such as Post-stratification Estimator, Raking, Generalized Regression Estimator. Calibration estimation is a development of this weighting thought, which is originally proposed by Deville and Särndal(1992) [5, 6]. Its main idea is to formulate calibration equations according to different auxiliary information. Under the constraints of those equations, given a specific distance function, the classic Horvitz-Thompson estimator is optimized in order to derive the calibration estimator. This method is also called the minimum distance method. Nowadays scholars pay great attention to this method, which has been widely used in some national statistical agencies at all levels [8, 9, 10, 11, 12].

Recently, a lot of smoothing Newton algorithms have been proposed for solving various optimization problems [15, 16, 17, 18]. The idea of smoothing Newton methods is to reformulate the problem concerned as a system of smooth equations by using reformulation function, where some smoothing function is used. Instead of solving the original problem, one solves the reformulated problem so that a solution of the original problem can be found.

In this paper, we use a smoothing Newton method to solve the calibration estimation problem with box constraints. Münnich, Sachs and Wagner (2011) [13] reformulated the basic calibration optimal model and derived an equation system using the projection function and KKT conditions. They show that solving the calibration problem is equivalent to find a solution of the equation system. This reformulation greatly reduces the dimension compared to the original prob-

lem. Therefore, the computing effort of different algorithms to solve this problem has been significantly reduced. In this paper, we continue to apply their model to study the calibration problem.

In section 2, we briefly introduce the basic calibration optimal model and the reformulated optimal model with box constraints. In section 3, on the basis of results in [13], we give more specific analysis about the semismooth Newton method and some numerical experiments have been reported. In section 4, we propose a smoothing Newton method to solve the equation system derived by the calibration problem. Then, we make a comparison among the R algorithm 'calib' created by the group of Yves Tillé [12], the semismooth Newton method and the smoothing Newton method.

A few words about our notation. $\Re^n$ denotes the space of n-dimensional real column vectors and $\Re_+$ (respectively, $\Re_{++}$) denotes the non-negative (respectively, positive) orthant in $\Re$. For an index set A, Σ_A is a shorthand for $\Sigma_{k \in A}$, e.g., $\Sigma_A y_k = \Sigma_{k \in A} y_k$. For a function $G : \Re^n \to \Re^m$, let $G'(x, h)$ be the direction derivative of G at x in the direction h; $G'(x)$ be Jacobian matrix and $\nabla G(x)$ be the transpose Jacobian matrix of G at $x \in \Re^n$ where G is F-derivative. Denote ∂G be the generalized Jacobian of G, i.e.

$$\partial G(x) = \text{conv} \left(\partial_B G(x) \right)$$

where $\text{conv}(S)$ denotes the convex hull of the set S, and

$$\partial_B G(x) = \left\{ \lim_{x^j \to x; x^j \in D_G} G'(x^j) \right\}.$$

If $\{\alpha_K\}$ and $\{\beta_K\}$ are two sequences in $\Re$ with $\beta_K \neq 0$, for all $K = 1, 2, \ldots$, $\alpha_K = O(\beta_K)$ means

$$\limsup_{k \to +\infty} \alpha_K / \beta_K = C$$

with $C \neq 0$; and $\alpha_K = o(\beta_K)$ means

$$\limsup_{k \to +\infty} \alpha_K / \beta_K = 0.$$

2　Optimal model of calibration estimation

Consider a finite population $U = \{1, \ldots, k, \ldots, n\}$, from which a probability sample $s(s \subseteq U)$ is drawn with a given sampling design, $p(\cdot)$. The inclusion probabilities

$$\pi_k := \text{Pr}(k \in s) = \sum_{s:k \in s} p(s),$$
$$\pi_{kl} := \text{Pr}(k \& l \in s)$$

are assumed to be strictly positive. y_k is the value of the variable of interest y, for the kth population element and an auxiliary vector value associating with y is

$$\mathbf{x}_k = (x_{k1}, \ldots, x_{kj}, \ldots, x_{kp})^T.$$

The population total of $\mathbf{x}$, $\mathbf{t}_x = \sum_U \mathbf{x}_k$ is assumed to be accurately known. This knowledge may come from one or more sources, such as census data, administrative data files, and others.

One objective is to estimate the population total $t_y = \sum_U y_k$, by the values of the variable of interest y, and auxiliary vectors $\mathbf{x}$ from sampling, together with the population total of $\mathbf{x}$. By modifying the basic sampling design weight $d_k = 1/\pi_k$, that appear in the Horvitz-Thompson estimator of t_y, which is from the idea of [3, 4], that is,

$$\hat{t}_{y\pi} = \sum_s y_k / \pi_k = \sum_s d_k y_k.$$

Calibration estimation method [5, 6] uses a new series of weights $w_k, (k = 1, \ldots, n)$ to derive a new estimator of t_y, $\hat{t}_y = \sum_s w_k y_k$. The choice of w_k should follow two rules. First, calibration estimator w_k should exactly estimate the population total or the mean value of $\mathbf{x}$. Namely, it should satisfy the calibration equation $\sum_s w_k \mathbf{x}_k = \mathbf{t}_x$. Second, w_k should be as close as possible to d_k in an average sense for a given metric $F_k(w, d)$. Mathematically, we use the expectation of $F_k(w, d)$ with respect to the sampling design to measure the average distance. For all s, minimizing this equation is equivalent to minimize the sum of $F_k(w, d)$, $\sum_s F_k(w, d)$, for any given s. Here we use an optimal model to illustrate the above statement

$$
\begin{aligned}
\min \quad & \sum_s F_k(w_k, d_k) \\
s.t. \quad & \sum_s w_k \mathbf{x}_k = \mathbf{t}_x
\end{aligned}
\tag{1}
$$

In this way, w_k is called calibration weights and $\hat{t}_y = \sum_s w_k y_k$ is the calibration estimator. Therefore, the main job of calibration estimation is to solve problem (1). This method to solve calibration weight w_k is also called the minimum distance method. To sum up, we give the definition of calibration estimation as follows, which is from [7].

Definition 1 *(Calibration Estimation) The calibration approach to estimation for finite populations consists of*

(1) a computation of weights that incorporate specified auxiliary information and are restrained by calibration equation,

(2) the use of these weights to compute linearly weighted estimates of totals and other finite population parameters: weight times variable value, summed over a set of observed units

(3) an objective to obtain nearly design unbiased estimates as long as non-response and other non-sampling errors are absent.

The essence of calibration estimation is to solve the optimal problem (1). The objective function $F_k(w, d)$ of the problem is a specific distance function. The choice of the distance function $F_k(w, d)$ plays an important role in solving this problem, since it may to some extend affect the CPU time of calculating the resulting calibration weights and their accuracy.

From [5, 6], we can find that for a proper distance function $F_k(w, d)$, it should have the following properties:

(1) for every fixed $d > 0$, $F_k(\cdot, d)$ is defined on an interval $D_k(d)$ containing d, and such that $F_k(d, d) = 0$;

(2) for every fixed $d > 0$, $F_k(\cdot, d)$ is nonnegative, differentiable with respect to w, strictly convex;

(3) $f_k(w, d) = \partial F_k(w, d)/\partial w$ is continuous and maps $D_k(d)$ onto an interval $Im_k(d)$ in a one-to-one fashion.

In most of our applications, for every k we choose the same distance function $F(w, d)$, where $f(w, d) = f(w/d)$. Let $g = w/d$, then f is a function of the single argument g, strictly increasing, and such that $f(1) = 0, f'(1) = 1$. Meanwhile, we can let the kth term have an individual, known positive weight $1/q_k$, unrelated to d_k, to adjust every term, i.e. $F_k(w, d) = F(w, d)/q_k$. Table 1 shows two examples of distance functions $F_k(w, d)$. For more distance functions, please refer to [5, 6].

Table 1: Examples of Distance Functions $F_k(w, d)$

Case	$F(w_k, d_k) =$ $q_k F_k(w_k, d_k)$	$f(g_k) =$ $q_k f_k(w_k, d_k)$
1	$d_k \frac{(g_k-1)^2}{2}$	$g_k - 1$
2	$d_k(g_k \cdot \ln(g_k) - g_k + 1)$	$\ln(g_k)$

Due to the occurrence of unrealistic or extreme weights w_k when choosing different distance functions, we define the convex, closed set of box constraints

$$\bar{U} = \left\{ \mathbf{g} = (g_1, \cdots, g_n)^T \in \Re^n : m_k \leq g_k \leq M_k \right\} \quad (2)$$

where we assume $0 \leq m_k \leq 1 \leq M_k$ and $m_k < M_k$ and g_k are called the calibration factors.

Then we give the optimal model of calibration estimation with box constraints, and write it as matrix-vector form, which can be also seen in [13].

Let $0 < p < n < \infty$,

$$\mathbf{x}_k = (x_{k1}, \cdots, x_{kp})^T \in \Re^p,$$

for $k = 1, \ldots, n$. Let x_{ki} be the value of the kth calibration variable for the ith sample element. Denote

$$\mathbf{d} = (d_1, \cdots, d_n)^T \in \Re^n$$

be the vector of the design weights. Meanwhile, there are calibration benchmarks $t_i(i = 1, \ldots, p)$, and

$$\mathbf{t}_x = (t_1, \cdots, t_p)^T \in \Re^p,$$

is the vector of the calibration total.

$$\mathbf{m} = (m_1, \cdots, m_n)^T \in \Re^n,$$
$$\mathbf{M} = (M_1, \cdots, M_n)^T \in \Re^n$$

are constraints vectors. We define the following matrices:

$$X^T := (x_1, \cdots, x_n) = \begin{pmatrix} x_{11} & \cdots & x_{n1} \\ \vdots & \ddots & \vdots \\ x_{1p} & \cdots & x_{np} \end{pmatrix}$$

$$D := \text{diag}(d_1, \cdots, d_n)$$

$$\bar{X}^T := (\xi_1, \cdots, \xi_n) = \begin{pmatrix} \xi_{11} & \cdots & \xi_{n1} \\ \vdots & \ddots & \vdots \\ \xi_{1p} & \cdots & \xi_{np} \end{pmatrix}$$
$$= \begin{pmatrix} x_{11}d_1 & \cdots & x_{n1}d_n \\ \vdots & \ddots & \vdots \\ x_{1p}d_1 & \cdots & x_{np}d_n \end{pmatrix}$$

where $\bar{X}^T$ is called the design matrix.

Let F be a nonnegative, strictly convex, twice continuously differentiable function

$$F : \Re_+ \to \Re_+,$$

which satisfies

$$F(1) = 0, F'(1) = 0, \text{ and } F''(1) = 1.$$

We define

$$\mathbf{F} : \Re_+^n \to \Re_+^n,$$
$$\mathbf{F}(\mathbf{g}) = (F(g_1), \cdots, F(g_n)).$$

Then we can easily verify that

$$F(w_k, d_k) = d_k F(g_k), k = 1, \ldots, n$$

satisfy the properties of a distance function, and all functions in Table 1 can be written as this form. In order to simplify the problem, let $q_k = 1 (k = 1, \ldots, n)$. Then

$$F_k(w_k, d_k) = F(w_k, d_k) = d_k F(g_k),$$
$$(k = 1, \ldots, n).$$

In the following, we assume that $s = U$. To sum up, the optimal model with box constraints can be written as

$$\begin{aligned} \min \quad & \mathbf{d}^T \mathbf{F}(\mathbf{g}) \\ \text{s.t.} \quad & h(\mathbf{g}) := \bar{X}^T \mathbf{g} - \mathbf{t}_x = 0 \\ & u(\mathbf{g}) := \mathbf{g} - \mathbf{M} \le 0 \\ & v(\mathbf{g}) := \mathbf{m} - \mathbf{g} \le 0 \end{aligned} \tag{3}$$

In the following, we consider the above general calibration problem (3).

We define a map

$$\mathbf{g} : \Re^p \to \Re^n_+$$

componentwise as

$$\begin{aligned} & g_k(\lambda) \\ = \; & \mathrm{Proj}_{[m_k, M_k]} \left(F'^{-1} \left(-\frac{\xi_k^T \lambda}{d_k} \right) \right) \\ = \; & \begin{cases} M_k, & \text{if } F'^{-1} \left(-\frac{\xi_k^T \lambda}{d_k} \right) \ge M_k; \\ F'^{-1} \left(-\frac{\xi_k^T \lambda}{d_k} \right), & \text{if } m_k < F'^{-1} \left(-\frac{\xi_k^T \lambda}{d_k} \right) < M_k; \\ m_k, & \text{if } F'^{-1} \left(-\frac{\xi_k^T \lambda}{d_k} \right) \le m_k, \end{cases} \end{aligned}$$

where $\mathrm{Proj}_{[a,b]}(c)$ is the function that projects c into the interval $[a, b]$ and m_k, M_k are defined in (2). Since

$$\begin{aligned} & \mathrm{Proj}_{[a,b]}(c) \\ = \; & \mathrm{mid}\{a, b, c\} \\ = \; & \tfrac{1}{2}(a + b - \sqrt{(b-c)^2} + \sqrt{(a-c)^2}) \end{aligned}$$

where $\mathrm{mid}\{a, b, c\}$ denotes the middle number between a, b, c, which is defined in [15]. Then we can rewrite g_k in this way,

$$\begin{aligned} & g_k(\lambda) = \\ & \tfrac{1}{2} \left[m_k + M_k - \sqrt{\left(M_k - F'^{-1} \left(-\frac{\xi_k^T \lambda}{d_k} \right) \right)^2} \right. \\ & \left. + \sqrt{\left(m_k - F'^{-1} \left(-\frac{\xi_k^T \lambda}{d_k} \right) \right)^2} \right], \end{aligned}$$

for $k = 1, \ldots, p$.

The next theorem gives that we can get the solution of (3) by solving other equations.

Theorem 2 ([13, Theorem 3]) *A vector* $\mathbf{g}^* \in \Re^n$ *is the unique solution of the optimization problem (3) if and only if there exists a multiplier* $\lambda^* \in \Re^p$ *such that* $\mathbf{g}(\lambda^*)$ *satisfies*

$$h(\mathbf{g}(\lambda^*)) = 0 \tag{4}$$

In the general case, $p < n$ always holds. Therefore, solving (4) is much easier than solving (3). Theorem 2 also states that finding a solution of the calibration problem (3) is equivalent to solve the equation

$$\Psi(\lambda) = 0 \tag{5}$$

where

$$\Psi : \Re^p \to \Re^p, \lambda \mapsto \bar{X}^T \mathbf{g} - \mathbf{t}_x.$$

with

$$\begin{aligned} & \Psi_i(\lambda) = \\ & \tfrac{1}{2} \sum_{k=1}^{n} \xi_{ki} \left[m_k + M_k - \sqrt{\left(M_k - F'^{-1} \left(-\frac{\xi_k^T \lambda}{d_k} \right) \right)^2} \right. \\ & \left. + \sqrt{\left(m_k - F'^{-1} \left(-\frac{\xi_k^T \lambda}{d_k} \right) \right)^2} \right] - t_i. \end{aligned}$$

for $i = 1, \ldots, p$.

In order to solve the non-smooth equations (4), Münnich, Sachs and Wagner [13] applied the semismooth Newton method to the problem when distance functions are Case 1 and Case 2, since Ψ are semismooth when choosing those functions. However, in their paper, they did not give the specific steps of their semismooth Newton method as well as the proof of its convergence theorem. We will supplement and modify their work in the following section.

3 A semismooth Newton method

Münnich, Sachs and Wagner in [13] applied the semismooth Newton method to solve the non-smooth equations (4). Then the calibration weights w_k can be easily derived. However, they did not give the specific steps of their semismooth Newton method. For example, they did not give the choice of H_k from the generalized Jacobian $\partial \Psi(\lambda)$ in the Newton equation. Furthermore, they did not show the proof of the method's convergence, see [13]. In this section, we will show that their semismooth Newton method does not maintain convergence property locally and globally.

First, we give the definitions of semismoothness, which are can be got from [15].

Definition 3 *(Semismoothness) Let* $X \subseteq \Re^n$ *and* $G : X \to \Re^m$ *be a locally Lipschitzian function. Then* G *is called*

(i) *Semismooth in $x \in X$, if for any $V \in \partial G(x + h), h \to 0$,*

$$G(x + h) - G(x) - Vh = o(\|h\|),$$

(ii) *Strongly semismooth in $x \in X$, if for any $V \in \partial G(x + h), h \to 0$,*

$$G(x + h) - G(x) - Vh = O(\|h\|^2),$$

(iii) *(Strongly) semismooth on X, if G is (strongly) semismooth in every $x \in X$.*

By the definition of strongly semismoothness and Lemma 4 in [13], we get that $\Psi(\lambda)$ is strongly semismooth on $\Re^p$ for the distance functions Case 1 and 2.

At the non-differentiable points of $\Psi(\lambda)$, we need to choose an element H_k form $\partial\Psi(\lambda)$ when solving the Newton equation (8). To achieve that, we show the set $\partial\Psi(\lambda)$.

Proposition 4 *Let $\Psi(\lambda)$ be defined in (5), then*

$$\partial\Psi(\lambda) = (\partial\Psi_i(\lambda_j))_{p \times p},$$

with

$$\partial\Psi_i(\lambda_j) = -\sum_{k=1}^{n} \frac{\xi_{ki}\xi_{kj}}{d_k} \left[F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right) \right]'$$

in the case of that $F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right) \in (m_k, M_k)$, for all $k = 1, \ldots, n$;

$$\partial\Psi_i(\lambda_j) \in \left(-\sum_{k=1}^{n} \frac{\xi_{ki}\xi_{kj}}{d_k} \left[F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right) \right]', \right.$$
$$\left. -\sum_{k \notin S} \frac{\xi_{ki}\xi_{kj}}{d_k} \left[F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right) \right]' \right)$$

in the case of that $F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right) = m_k$ or M_k, for $k \in S \subseteq \{1, \ldots, n\}$.

Proof: For $i, j = 1, \ldots, p$, we give the proof of the desired results of $\partial\Psi_i(\lambda_j)$ by considering the following cases: (i) $F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right) \in (m_k, M_k)$, for all $k = 1, \ldots, n$,

and (ii) $F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right) = m_k$ or M_k, for $k \in S \subseteq \{1, \ldots, n\}$.

Case (i) Suppose that λ satisfies

$$F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right) \in (m_k, M_k).$$

for all $k = 1, \ldots, n$.

Then $\Psi(\lambda)$ is differentiable at λ, we have

$$\nabla\Psi(\lambda) = (\partial\Psi_i(\lambda_j))_{p \times p} = \left(\frac{\partial\Psi_i}{\partial\lambda_j}\right)_{p \times p},$$

where

$$\frac{\partial\Psi_i}{\partial\lambda_j} = \frac{1}{2}\sum_{k=1}^{n} \frac{\xi_{ki}\xi_{kj}}{d_k} \left[F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right) \right]'$$

$$\left[\frac{m_k - F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right)}{\sqrt{\left(m_k - F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right)\right)^2}} - \frac{M_k - F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right)}{\sqrt{\left(M_k - F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right)\right)^2}} \right]$$

$$= -\sum_{k=1}^{n} \frac{\xi_{ki}\xi_{kj}}{d_k} \left[F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right) \right]'.$$

Case (ii) Suppose that λ satisfies

$$F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right) = m_k \text{ or } M_k,$$

for $k \in S \subseteq \{1, \ldots, n\}$.

For any $k \in S$, suppose that $F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right) = m_k$, then

$$\left[\frac{m_k - F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right)}{\sqrt{\left(m_k - F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right)\right)^2}} - \frac{M_k - F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right)}{\sqrt{\left(M_k - F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right)\right)^2}} \right]$$

$$\to [-2, 0],$$

further,

$$\partial\Psi_i(\lambda_j) \in \left(-\sum_{k=1}^{n} \frac{\xi_{ki}\xi_{kj}}{d_k} \left[F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right) \right]', \right.$$
$$\left. -\sum_{k \notin S} \frac{\xi_{ki}\xi_{kj}}{d_k} \left[F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right) \right]' \right).$$

Therefore, we complete the proof. □

For any $\lambda \in \Re^p$, let

$$H(\lambda) = (h_{ij}(\lambda))_{p \times p} \qquad (6)$$

with

$$h_{ij}(\lambda) = -\sum_{k=1}^{n} \frac{\xi_{ki}\xi_{kj}}{d_k} \left[F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right) \right]'. \qquad (7)$$

It is easily to show that $H(\lambda) \in \partial\Psi(\lambda)$. And $H(\lambda)$ can be written in the matrix form

$$H(\lambda) = \frac{1}{2}\bar{X}^T D_\lambda D^{-1}\bar{X},$$

where

$$D_\lambda = \text{diag}\left(-2\left[F'^{-1}\left(-\frac{\xi_1^T\lambda}{d_1}\right)\right]', \cdots, -2\left[F'^{-1}\left(-\frac{\xi_n^T\lambda}{d_n}\right)\right]'\right)$$

and $D = \text{diag}(\mathbf{d})$.

The next Theorem will prove that for any $\lambda \in \Re^p$, $H(\lambda)$ is nonsingular under the condition that $\bar{X}$ has full column rank. We choose $H(\lambda^K)$ as the matrix H_K in the Newton equation (8).

Theorem 5 *If $\bar{X}$ has full column rank, then for any $\lambda \in \Re^p$, $H(\lambda)$ is nonsingular.*

Proof: We give the proof by considering the two cases of F in Table 1.

(i) For Case 1,

$$F(g_k) = \frac{(g_k - 1)^2}{2}.$$

Then $F'^{-1}(g_k) = 1 + g_k$, and

$$(F'^{-1}(g_k))' = 1.$$

According to Formula (7), we have

$$h_{ij} = -\sum_{k=1}^{n} \frac{\xi_{ki}\xi_{kj}}{d_k}, (i, j = 1, 2, \ldots, p)'$$

we have that

$$H(\lambda) = -\bar{X}^T D^{-1}\bar{X}.$$

It is easy to get that $H(\lambda)$ is nonsingular if $\bar{X}$ has full column rank.

(ii) For Case 2,

$$F(g_k) = g_k \cdot \ln(g_k) - g_k + 1,$$

then

$$F'^{-1}(g_k) = (F'^{-1}(g_k))' = e^{g_k}.$$

According to Formula (7), we have

$$h_{ij} = -\sum_{k=1}^{n} \frac{\xi_{ki}\xi_{kj}}{d_k} e^{-\frac{\xi_k^T\lambda}{d_k}}, (i, j = 1, 2, \ldots, p),$$

and

$$H(\lambda) = \frac{1}{2}\bar{X}^T \tilde{D}_\lambda D^{-1}\bar{X},$$

where

$$\tilde{D}_\lambda = \text{diag}\left(-2e^{-\frac{\xi_1^T\lambda}{d_1}}, \cdots, -2e^{-\frac{\xi_n^T\lambda}{d_n}}\right),$$

which is a negative definite matrix. Then $H(\lambda)$ is nonsingular if $\bar{X}$ has full column rank. □

Here we show a semismooth Newton algorithm for solving the calibration estimation problem.

Algorithm 6 *(A Semismooth Newton Algorithm)*

Step 1 *Given the starting point $\lambda^0 \in \Re^p$. Choose σ, $\rho \in (0, \frac{1}{2})$. Set $K = 0$.*

Step 2 *If $\|\Psi(\lambda)\| = 0$, then stop. Otherwise go to Step 3.*

Step 3 *Compute $\triangle\mathbf{z}^K \in \Re^p$ by*

$$H_K \triangle\mathbf{z}^K = -\Psi(\lambda^K), \qquad (8)$$

where $H_K \in \partial\Psi(\lambda^K)$.

Step 4 *Set $\lambda^{K+1} := \lambda^K + \triangle\mathbf{z}^K$, and $K := K + 1$, Go to Step 2.*

Remark 7 *In Algorithm 6, we choose $H_K = H(\lambda^K)$, $K = 0, 1, 2, \ldots$. From Theorem 5, we get if $\bar{X}$ has full column rank, then $H(\lambda^K)$ is nonsingular. Therefore, Algorithm 6 is well defined.*

We test Algorithm 6 on an example included in the 'sampling' package in R. Nevertheless Algorithm 6 can also be applied to higher dimensional problems. For each sample size, we tested 10 times and calculated their average value. Each time, samples were randomly chosen from the population. Algorithm 6 was implemented in Matlab 2010b on an Inter(R) Core(TM)2 CPU T5500 @ 1.67GHz and 1GB RAM.

We consider the computing effort for the semismooth Newton method when solving problems with different sample size. Here we chose distance function Case 1. We used the optimal Lagrange multiplier of the optimization problem without box constraints as a starting point since we found out that it needed the least iteration number after several attempts. If we randomly chose a starting point, the algorithm might not converge. This is therefore the very disadvantage of the semismooth Newton method. The results of numerical experiments are shown in Table 2.

Table 2: Computing effort for different problem sizes

	$n = 185$	$\tilde{n} = 1,850$	n=18,500
Avg.It	1	1	1
Avg.t(s)	0.03	0.05	0.09
$\|\Psi(\lambda)\|$	1.32e-8	2.73e-8	7.62e-8
$\|\Psi(\lambda)\|_{\max}$	2.84e-8	5.37e-8	1.46e-7

In Table 2, **Avg. It** denotes the average number of iterations; **Avg.t(s)** denotes the average cpu time when

the algorithm terminates; $\|\bar{\Psi(\lambda)}\|$ denotes the average value of $\|\Psi(\lambda)\|$ when the algorithm terminates; $\|\Psi(\lambda)\|_{\max}$ denotes the maximal value of $\|\Psi(\lambda)\|$ when the algorithm terminates.

In Table 2, we find that the advantage of our improved semismooth Newton method is that it can converge after one iteration. With the increase of the sample size, the CPU time increases linearly, but is still very short. However, the semismooth Newton method cannot maintain its convergence property. Therefore, in the next section, we propose a smoothing Newton method to solve the same calibration estimation problem.

4   A smoothing Newton method

In this section, we consider using a smoothing Newton method to solve Formula (5). Let $\mathbf{z} := (\mu, \lambda) \in \Re_{++} \times \Re^p$, and $\Phi : \Re_{++} \times \Re^p \to \Re_{++} \times \Re^p$

$$\Phi(\mathbf{z}) = \begin{pmatrix} \mu \\ \Psi(\mu, \lambda) - \mu \end{pmatrix} \tag{9}$$

where

$$
\begin{aligned}
&\Psi(\mu, \lambda) \\
&= \tfrac{1}{2}\bar{X}^T \left(\mathbf{m} + \mathbf{M} - \sqrt{(\mathbf{M} - \mathbf{F}'^{-1}(\lambda))^2 + \mu^2 \mathbf{e}} \right. \\
&\quad \left. + \sqrt{(\mathbf{m} - \mathbf{F}'^{-1}(\lambda))^2 + \mu^2 \mathbf{e}} - \mathbf{t} \right)
\end{aligned}
\tag{10}
$$

and $\mathbf{e}$ is a column vector whose elements are all one,

$$\mathbf{F}'^{-1}(\lambda) = \left[F'^{-1}\left(-\frac{\xi_1^T \lambda}{d_1} \right), \cdots, F'^{-1}\left(-\frac{\xi_n^T \lambda}{d_n} \right) \right].$$

Definition 8 (*Consistently smoothing approximation function*) *Given $G : \Re^n \to \Re^n$, $G(\mu, \cdot) : \Re^n \to \Re^n$ is called the smoothing approximation function of G, if for every $x \in \Re^n$, there exists $\kappa > 0$, such that*

$$\|G(\mu, x) - G(\mu, x)\| \le \kappa \mu, \quad \forall \mu > 0.$$

Furthermore, if κ is independent of x, then $G(\mu, \cdot)$ is called the consistently smoothing approximation function of G.

Now, we claim that $\Psi(\mu, \lambda)$ is a consistently smoothing approximation function of $\Psi(\lambda)$.

Theorem 9 *$\Psi(\mu, \lambda)$ is a consistently smoothing approximation function of $\Psi(\lambda)$.*

Proof: For $i \in \{1, \ldots, p\}$, we have

$$
\begin{aligned}
&|\Psi_i(\mu, \lambda) - \Psi_i(\lambda)| \\
&= \tfrac{1}{2} \sum_{k=1}^n \xi_{ki} \left| \sqrt{\left(M_k - F'^{-1}\left(-\frac{\xi_k^T \lambda}{d_k} \right) \right)^2} - \right. \\
&\quad \sqrt{\left(M_k - F'^{-1}\left(-\frac{\xi_k^T \lambda}{d_k} \right) \right)^2 + \mu^2} \\
&\quad + \sqrt{\left(m_k - F'^{-1}\left(-\frac{\xi_k^T \lambda}{d_k} \right) \right)^2 + \mu^2} \\
&\quad \left. - \sqrt{\left(m_k - F'^{-1}\left(-\frac{\xi_k^T \lambda}{d_k} \right) \right)^2} \right| \\
&= \tfrac{1}{2} \sum_{k=1}^n \xi_{ki} \\
&\quad \left| \frac{-\mu^2}{\sqrt{\left(M_k - F'^{-1}\left(-\frac{\xi_k^T \lambda}{d_k} \right) \right)^2} + \sqrt{\left(M_k - F'^{-1}\left(-\frac{\xi_k^T \lambda}{d_k} \right) \right)^2 + \mu^2}} \right. \\
&\quad \left. + \frac{\mu^2}{\sqrt{\left(m_k - F'^{-1}\left(-\frac{\xi_k^T \lambda}{d_k} \right) \right)^2} + \sqrt{\left(m_k - F'^{-1}\left(-\frac{\xi_k^T \lambda}{d_k} \right) \right)^2 + \mu^2}} \right| \\
&\le \tfrac{1}{2} \sum_{k=1}^n \xi_{ki} \mu^2 \left| \frac{2}{\sqrt{\mu^2}} \right| \\
&= \sum_{k=1}^n \xi_{ki} \mu.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\|\Psi(\mu, \lambda) - \Psi(\lambda)\| \\
&\le \sqrt{\sum_{i=1}^p \left(\Psi_i(\mu, \lambda) - \Psi_i(\lambda) \right)^2} \\
&\le \sqrt{p} \max_{1 \le i \le p} |\Psi_i(\mu, \lambda) - \Psi_i(\lambda)| \\
&\le \sqrt{p} \max_{1 \le i \le p} \left(\sum_{k=1}^n \xi_{ki} \mu \right) \\
&= \sqrt{p} \max_{1 \le i \le p} \left(\sum_{k=1}^n \xi_{ki} \right) \mu \\
&= \kappa \mu,
\end{aligned}
$$

where

$$\kappa = \sqrt{p} \max_{1 \le i \le p} \left(\sum_{k=1}^n \xi_{ki} \right).$$

As κ is independent of λ, $\Psi(\mu, \lambda)$ is a consistently smoothing approximation of $\Psi(\lambda)$. $\square$

Choose $\bar{\mu} \in \Re_{++}$ and $\gamma \in (0, 1)$, such that $\gamma \bar{\mu} < 1$. Let

$$\bar{\mathbf{z}} = (\bar{\mu}, \mathbf{0}) \in \Re_{++} \times \Re^p$$

and $\Theta : \Re^{p+1} \to \Re_+$,

$$\Theta(\mathbf{z}) = \|\Phi(\mathbf{z})\|^2,$$

and let $\beta : \Re^{p+1} \rightarrow \Re_+$,

$$\beta(\mathbf{z}) = \gamma \min\{1, \Theta(\mathbf{z})\}.$$

Then define a neighborhood

$$\Omega := \{\mathbf{z} = (\mu, \lambda) \in \Re \times \Re^p | \mu \geq \beta(\mathbf{z})\bar{\mu}\}.$$

Because $\beta(\mathbf{z}) \leq \gamma < 1$, we have

$$(\mu, \lambda) \in \Omega, \forall \lambda \in \Re^p.$$

According to the definition above, we have the following relationship between those functions.

Lemma 10 $\Phi(\mathbf{z}) = 0 \iff \beta(\mathbf{z}) = 0 \iff \Phi(\mathbf{z}) = \beta(\mathbf{z})\bar{\mathbf{z}}.$

Proof: By the definitions of $\Phi(\mathbf{z})$ and $\beta(\mathbf{z})$, we have

$$\Phi(\mathbf{z}) = 0 \iff \beta(\mathbf{z}) = 0$$

and

$$\beta(\mathbf{z}) = 0 \implies \Phi(\mathbf{z}) = \beta(\mathbf{z})\bar{\mathbf{z}}.$$

Therefore, it just need to prove:

$$\Phi(\mathbf{z}) = \beta(\mathbf{z})\bar{\mathbf{z}} \implies \beta(\mathbf{z}) = 0.$$

If $\Phi(\mathbf{z}) = \beta(\mathbf{z})\bar{\mathbf{z}}$, then

$$\mu = \beta(\mathbf{z})\bar{\mu} \text{ and } \Psi(\mu, \lambda) - \mu\lambda = 0.$$

Hence, by the definitions of $\Theta(\mathbf{z})$ and $\beta(\mathbf{z})$ and together with $\gamma\bar{\mu} < 1$, we have

$$\begin{aligned} \Theta(\mathbf{z}) &= \mu^2 + \|\Psi(\mu, \lambda) - \mu\lambda\|^2 \\ &= \mu^2 = (\beta(\mathbf{z})\bar{\mu})^2 \\ &\leq \gamma^2\bar{\mu}^2 < 1. \end{aligned}$$

i.e.,

$$\beta(\mathbf{z}) = \gamma\Theta(\mathbf{z}) = \gamma(\beta(\mathbf{z})\bar{\mu})^2.$$

If $\beta(\mathbf{z}) \neq 0$, from the equation above and $\beta(\mathbf{z}) \leq \gamma$, we have

$$1 = \gamma\beta(\mathbf{z})\bar{\mu}^2 \leq \gamma^2\bar{\mu}^2.$$

This contradicts to $\gamma\bar{\mu} < 1$, so $\beta(\mathbf{z}) = 0$. This completes the proof. $\square$

We can easily see that $\Phi(\mathbf{z})$ is differentiable on $\Re_{++} \times \Re^p$. And the Jacobian of $\Phi(\mathbf{z})$ is

$$\Phi'(\mathbf{z}) = \begin{pmatrix} 1 & \mathbf{0} \\ \Psi_\mu - \lambda & \Psi_\lambda - \mu\mathbf{E} \end{pmatrix} \qquad (11)$$

where $\mathbf{E}$ is an identity matrix of dimension p and

$$\Psi_\mu = \left(\frac{\partial\Psi_1}{\partial\mu}, \cdots, \frac{\partial\Psi_p}{\partial\mu}\right)^T,$$

$$\begin{aligned} \Psi_\lambda &= \begin{pmatrix} \frac{\partial\Psi_1}{\partial\lambda_1}, & \cdots, & \frac{\partial\Psi_1}{\partial\lambda_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial\Psi_p}{\partial\lambda_1}, & \cdots, & \frac{\partial\Psi_p}{\partial\lambda_p} \end{pmatrix} \\ &= \frac{1}{2}\bar{X}^T D_{\lambda\mu} D^{-1}\bar{X}, \end{aligned}$$

with

$$\frac{\partial\Psi_i}{\partial\mu} = \frac{\mu}{2}\sum_{k=1}^n \xi_{ki}\left(\left(\sqrt{\left(m_k - F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right)\right)^2 + \mu^2}\right)^{-1} - \left(\sqrt{\left(M_k - F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right)\right)^2 + \mu^2}\right)^{-1}\right),$$

and

$$D_{\lambda\mu} = \text{diag}\left(\bar{d}_1, \cdots, \bar{d}_n\right),$$

with

$$\bar{d}_k = (F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right))'\left(\frac{m_k - F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right)}{\sqrt{\left(m_k - F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right)\right)^2 + \mu^2}} - \frac{M_k - F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right)}{\sqrt{\left(M_k - F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right)\right)^2 + \mu^2}}\right)$$

for $k = 1, \ldots, n$.

Next, we show a smoothing Newton algorithm to solving the calibration estimation problem and prove that it is well defined.

Algorithm 11 *A Smoothing Newton Algorithm*

Step 0 *Choose $\delta \in (0, 1)$, $\sigma \in (0, \frac{1}{2})$, and let $\mu_0 = \bar{\mu}$, $\lambda^0 \in \Re^p$. Set $K = 0$.*

Step 1 *If $\Phi(\mathbf{z}^K) = 0$, then stop. Otherwise, let $\beta_K = \beta(\mathbf{z}^K)$, and go to Step 2.*

Step 2 *Solve $\triangle\mathbf{z}^K = (\triangle\mu_K, \triangle\lambda^K) \in \Re \times \Re^p$ by*

$$\Phi(\mathbf{z}^K) + \Phi'(\mathbf{z}^K)\triangle\mathbf{z}^K = \beta_K\bar{\mathbf{z}}. \qquad (12)$$

Step 3 *Let l_K be the smallest nonnegative integer l satisfying*

$$\Theta(\mathbf{z}^K + \delta^l \triangle \mathbf{z}^K) \leq \left(1 - 2\sigma(1 - \gamma\bar{\mu})\delta^l\right) \Theta(\mathbf{z}^K).$$
(13)

Step 4 *Set $\mathbf{z}^{K+1} := \mathbf{z}^K + \delta^{l_K}\triangle \mathbf{z}^K$, $K := K+1$. Go to Step 1.*

Lemma 12 *For any $(\mu, \lambda) \in \Re_{++} \times \Re^p$, $\Phi'(\mathbf{z})$ is nonsingular.*

Proof: From (11), we get that to prove $\Phi'(\mathbf{z})$ is nonsingular is equivalent to prove

$$\Psi_\lambda - \mu\mathbf{E} = \frac{1}{2}\bar{X}^T D_{\lambda\mu} D^{-1} \bar{X} - \mu\mathbf{E}$$

is nonsingular.

Firstly, we show that Ψ_λ negative semidefinite.

Since $D_{\lambda\mu} = \text{diag}\left(\bar{d}_1, \cdots, \bar{d}_n\right)$, with

$$\bar{d}_k =$$

$$(F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right))' \left(\frac{m_k - F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right)}{\sqrt{\left(m_k - F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right)\right)^2 + \mu^2}} \right.$$

$$\left. -\frac{M_k - F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right)}{\sqrt{\left(M_k - F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right)\right)^2 + \mu^2}} \right)$$

for $k = 1, \ldots, n$.

From the next two part, we show that $D_{\lambda\mu}$ is negative semidefinite.

(i) For any distance function F in Table 1, its derivative F' is continuous, strictly increasing. Therefore, its inverse function F'^{-1} is continuous, strictly increasing. Then, the value of $(F'^{-1})'$ is always nonnegative.

(ii) For any $(\mu, \lambda) \in \Re_{++} \times \Re^p$, we consider the function

$$f(x) = \frac{x - F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right)}{\sqrt{\left(x - F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right)\right)^2 + \mu^2}}.$$

Its derivative of the first order is

$$f'(x) = \frac{\mu^2}{\left[\left(x - F'^{-1}\left(-\frac{\xi_k^T\lambda}{d_k}\right)\right)^2 + \mu^2\right]^{\frac{3}{2}}} > 0.$$

Thus, $f(x)$ is a decreasing function. Then for $m_k < M_k$, $f(m_k) \leq f(M_k)$, i.e.

$$f(m_k) - f(M_k) \leq 0.$$

Therefore, from the two parts (i) and (ii), we obtain that the diagonal elements of $D_{\lambda\mu}$ are all nonpositive, which means that $D_{\lambda\mu}$ is negative semidefinite.

Denote $\sqrt{D^{-1}}$ be a diagonal matrix whose elements are the square root of the diagonal elements of D^{-1}. Then $D^{-1} = \sqrt{D^{-1}}\sqrt{D^{-1}}$. So we can rewrite Ψ_λ as a symmetric form

$$\Psi_\lambda = \frac{1}{2}\left(\sqrt{D^{-1}}\bar{X}\right)^T D_{\lambda\mu} \left(\sqrt{D^{-1}}\bar{X}\right).$$

For $\forall q \in \Re^p$,

$$q^T \Psi_\lambda q = \frac{1}{2}\left(\sqrt{D^{-1}}\bar{X}q\right)^T D_{\lambda\mu}\left(\sqrt{D^{-1}}\bar{X}q\right)$$
$$\leq 0.$$

Thus, Ψ_λ is negative semidefinite.

As $\mu \in \Re_{++}$, then we have $\Psi_\lambda - \mu\mathbf{E}$ is negative definite.

Above all, $\Phi'(\mathbf{z})$ is nonsingular. This completes the proof. □

Lemma 13 *For any $\tilde{\mathbf{z}} = (\tilde{\mu}, \tilde{\lambda}) \in \Re_{++} \times \Re^p$, there exists a closed neighborhood $N(\tilde{\mathbf{z}})$ of $\tilde{\mathbf{z}}$ and a positive number $\tilde{\alpha} \in (0, 1]$, such that for any $\mathbf{z} = (\mu, \lambda) \in N(\tilde{\mathbf{z}})$ and all $\alpha \in [0, \tilde{\alpha}]$, we have for any $\mu \in \Re_{++}$,*

$$\Theta(\mathbf{z} + \alpha\triangle\mathbf{z}) \leq (1 - 2\sigma(1 - \gamma\bar{\mu})\alpha) \Theta(\mathbf{z}) \quad (14)$$

holds, where $\triangle\mathbf{z} = (\triangle\mu, \triangle\lambda) \in \Re \times \Re^p$ is the unique solution to the following equation:

$$\Phi(\mathbf{z}) + \Phi'(\mathbf{z})\triangle\mathbf{z} = \beta\bar{\mathbf{z}}.$$

Proof: The proof is similar to that of Lemma 5 in [15]. The detail is omitted. □

Remark 14 *Algorithm 11 is well defined. As*

(i) *the Newton equation (12) is well defined, since $\Phi'(\mathbf{z}^K)$ is nonsingular from Lemma 12.*

(ii) *the linesearch step (13) is well defined from Lemma 13*

By a similar discussion with [15, Proposition 8], we can get the following lemma.

Lemma 15 *Assume that $\bar{X}$ has full column rank, then Algorithm 11 generates an infinite sequence $\mathbf{z}^K$ and $\{\mu_K\} \subseteq \Re_{++}, \{\mathbf{z}^K\} \subseteq \Omega$.*

Lemma 16 *The sequence $\{\mathbf{z}^K\}$ generated by Algorithm 11 is bounded.*

Proof: From the linesearch (13), $\Theta(\mathbf{z}^K)$ is decreasing. It is sufficient to show that the level set

$$L(\mathbf{z}^0) := \{\mathbf{z} | \Theta(\mathbf{z}) \leq \Theta(\mathbf{z}^0)\}$$

is bounded.

Suppose that $L(\mathbf{z}^0)$ is unbounded, we will induce a contradiction. Then, for the sequence $\{\mathbf{z}^K\}$, there exists a subsequence, without generation, we still denote it as $\{\mathbf{z}^K\}$ such that $\{\mathbf{z}^K\} \subset L(\mathbf{z}^0)$ and

$$\|\mathbf{z}^K\| \to \infty.$$

As $\{\mu_K\}$ is nonnegative and decreasing, we have $\{\mu_K\} < \bar{\mu}$. Therefore, there must exist at least one sequence $\{\lambda_i^K\}, i \in \{1, \ldots, p\}$ of the components of $\{\lambda^K\}$, such that $|\lambda_i^K| \to \infty$.

We also have

$$\Theta(\mathbf{z}) = \|\Phi(\mathbf{z})\|^2 = \mu^2 + \|\Psi(\mu, \lambda)\|^2$$
$$= \mu^2 +$$
$$\sum_{i=1}^{p} \left(\frac{1}{2} \sum_{k=1}^{n} \xi_{ki} \left(m_k + M_k - \right.\right.$$
$$\sqrt{\left(M_k - F'^{-1}\left(-\frac{\xi_k^T \lambda}{d_k}\right)\right)^2} +$$
$$\left.\left. \sqrt{\left(m_k - F'^{-1}\left(-\frac{\xi_k^T \lambda}{d_k}\right)\right)^2}\right) - t_i - \mu\lambda_i \right)^2$$
$$= \mu^2 +$$
$$\sum_{i=1}^{p} \left(\frac{1}{2} \sum_{k=1}^{n} \xi_{ki} \left(m_k + M_k - \right.\right.$$
$$\frac{\left(m_k - F'^{-1}\left(-\frac{\xi_k^T \lambda}{d_k}\right)\right)^2 - \left(M_k - F'^{-1}\left(-\frac{\xi_k^T \lambda}{d_k}\right)\right)^2}{\left|m_k - F'^{-1}\left(-\frac{\xi_k^T \lambda}{d_k}\right)\right| + \left|M_k - F'^{-1}\left(-\frac{\xi_k^T \lambda}{d_k}\right)\right|} \right)$$
$$- t_i - \mu\lambda_i)^2$$
$$= \mu^2 +$$
$$\sum_{i=1}^{p} \left(\frac{1}{2} \sum_{k=1}^{n} \xi_{ki} \left(m_k + M_k - \right.\right.$$
$$\frac{m_k^2 - M_k^2 - 2F'^{-1}\left(-\frac{\xi_k^T \lambda}{d_k}\right)(m_k - M_k)}{\left|m_k - F'^{-1}\left(-\frac{\xi_k^T \lambda}{d_k}\right)\right| + \left|M_k - F'^{-1}\left(-\frac{\xi_k^T \lambda}{d_k}\right)\right|} \right)$$
$$- t_i - \mu\lambda_i)^2$$

When $\{\lambda_i^K\}$ approaches infinity, by using the fact that F'^{-1} is a monotonic function, we have

$$\lim_{\|\mathbf{z}^K\| \to \infty} \Theta(\mathbf{z}^K) = \infty,$$

which contradicts that

$$0 \leq \Theta(\mathbf{z}^K) \leq \Theta(\mathbf{z}^0), \ \forall k = 1, 2, \ldots.$$

Therefore, the sequence $\{\mathbf{z}^K\}$ is bounded. □

The following theorem shows that Algorithm 11 is global convergent.

Theorem 17 *Assume that $\{\mathbf{z}^K\}$ is generated by Algorithm 11, then any accumulation point of $\{\mathbf{z}^K\}$ is a solution of $\Phi(\mathbf{z}) = 0$.*

Proof: The proof is similar to that of Theorem 4 in [15]. □

The following theorem shows the superlinear and quadratic convergence of Algorithm 11.

Theorem 18 *Assume that $\{\mathbf{z}^K\}$ is generated by Algorithm 11, then the sequence $\{\mathbf{z}^K\}$ converges to z*, and*

$$\|\mathbf{z}^{K+1} - \mathbf{z}^*\| = o(\|\mathbf{z}^K - \mathbf{z}^*\|)$$

and

$$\mu_{K+1} = o(\mu_K).$$

Meanwhile, if $\Phi(\mathbf{z})$ is strongly semismooth at the point $\mathbf{z}^*$, then

$$\|\mathbf{z}^{K+1} - \mathbf{z}^*\| = O(\|\mathbf{z}^K - \mathbf{z}^*\|)$$

and

$$\mu_{K+1} = O(\mu_K).$$

Proof: The proof is similar to that of [15], Theorem 8. □

From Theorem 4 and Theorem 5, we can see that the smoothing Newton method is global convergent without any assumptions and locally superlinearly convergent with some conditions.

We chose distance functions Case 1 and Case 2, and applied our smoothing Newton method to solving the corresponding calibration estimation problem with the same sample size. Then we compared their computing efforts in Table 3 and Table 4. After several attempts, we used the optimal Lagrange multiplier of the optimization problem without box constraints as a starting point for the same reason as we mentioned in the semismooth Newton method.

Table 3: Computing effort for distance functions Case 1

n	It	t(s)	$\|\mathbf{\Psi}(\lambda)\|$	$\|\mathbf{\Psi}(\lambda)\|_{\max}$
185	2	0.09	5.02e-7	8.69e-7
1,850	1	0.10	7.33e-7	8.35e-7
18,500	2	0.62	7.68e-7	9.84e-7

Both the semismooth Newton method and the smoothing Newton method can be used to solve the

Table 4: Computing effort for distance functions Case 2

n	It	t(s)	$\|\Psi(\lambda)\|$	$\|\Psi(\lambda)\|_{\max}$
185	6	0.13	1.59e-7	9.81e-7
1,850	4	0.20	4.54e-7	8.26e-7
18,500	4	1.34	5.03e-7	9.94e-7

Table 5: Comparison among three algorithms

Semismooth Newton method	$n = 185$	$n = 1,850$	n=18,500
Avg.It	1	1	1
Avg.t(s)	0.03	0.05	0.09
$\|\overline{\Psi}(\lambda)\|$	1.32e-8	2.73e-8	7.62e-8
$\|\Psi(\lambda)\|_{\max}$	2.84e-8	5.37e-8	1.46e-7
Smoothing Newton method	$n = 185$	$n = 1,850$	n=18,500
Avg.It	2	1	2
Avg.t(s)	0.09	0.10	0.62
$\|\overline{\Psi}(\lambda)\|$	5.02e-7	7.33e-7	7.68e-7
$\|\Psi(\lambda)\|_{\max}$	8.69e-7	8.35e-7	9.84e-7
Calib	$n = 185$	$n = 1,850$	n=18,500
Avg.It	1	1	1
Avg.t(s)	0.06	0.07	0.09
$\|\overline{\Psi}(\lambda)\|$	9.86e-13	5.30e-12	1.48e-10
$\|\Psi(\lambda)\|_{\max}$	1.42e-12	8.35e-12	3.46e-10

calibration problem with different distance functions. Meanwhile, they both show good efficiency and results according to the numerical experiments. For distance function Case 1, the semismooth Newton method only runs one iteration, while the smoothing Newton method is as good as it. Due to the complexity of distance function Case 2, the number of iteration is larger than and the time is longer than those of Case 1. However, the results of the problem under distance function Case 2 may be much accurate than that of Case 1. Therefore, it has both advantages and disadvantages when choosing different distance functions. In the application, we should carefully choose a proper distance function according to different demand.

We also made a comparison among the semismooth Newton method, the smoothing Newton method and the 'Calib'. Here we chose the same distance function Case 1 to solve problems with different sample size. The results are shown in Table 5.

In Table 3-5, **Avg.It** denotes the average number of iterations; **Avg.t(s)** denotes the average cpu time when the algorithm terminates; $\|\overline{\Psi}(\lambda)\|$ denotes the average value of $\|\Psi(\lambda)\|$ when the algorithm terminates; $\|\Psi(\lambda)\|_{\max}$ denotes the maximal value of $\|\Psi(\lambda)\|$ when the algorithm terminates.

In Table 5, we can conclude that 'Calib' is still the quickest algorithm to solve the calibration estimation problem. The Semismooth Newton method is as good as 'Calib' since it only runs one iteration. However, the very disadvantage of those two algorithms is that they cannot maintain convergent property. On the other hand, the smoothing Newton method has been proved to have the global convergence without any assumptions and the local superlinear convergence under certain assumptions, and its numerical results are equally perfect compared to the other two algorithms.

5 Conclusion

This paper focuses on solving the calibration estimation problem with box constraints. On the basis of the optimal model with box constraints, improvement has been made to refine the existing semismooth Newton

method. Corresponding numerical experiments have been done to test it. Furthermore, a smoothing Newton method has been designed to solve the same problem. A comparison has been made through numerical experiments among semismooth Newton method, smoothing Newton method and 'Calib'. From the results of numerical experiments, 'Calib' algorithm is still the quickest and most efficient algorithm for solving the problem. After the improvement of semismooth Newton method, it also comes out of one step of iteration. Its efficiency has been improved significantly. The smoothing Newton method can have the global convergence without any assumptions and the local superlinear convergence under certain assumptions. Since the core content of calibration estimation method is an optimization process, it is possible to introduce some other optimize methods to solve this problem, for example, alternating direction method, proximal-like algorithm et al. These deserve to be further investigated.

References:

[1] J. Neyman, On two different aspects of the representative method: The method of stratified sampling and the method of purposive selection, *J. Royal Stat. Soc. B.*, 97, 1934, pp. 558-606.

[2] M. H. Hansen and W. H. Hurwitz, On the theory of sampling from finite populations, *Ann. Math. Statist.*, 14, 1943, pp. 333-362.

[3] D. G. Horvitz and D. J. Thompson, A generalization of sampling without replacement from a finite universe, *J. Amer. Statist. Assoc.*, 47, 1952, pp. 663-685.

[4] J. C. Deville, Estimation Lineaire et Redressement sur Informations Auxihaires d'Enquetes par Sondage, *Essais en l'Honneur d'Edmond Mahnvaud*, 1988, pp. 915-927.

[5] J. C. Deville and C. E. Särnda, Calibration estimators in survey sampling, *Journal of the American Statistical Association*, 87, 1992, pp. 376-382.

[6] J. C. Deville, C. E. Särnda and C. E. Sautory, Generalized raking procedures in survey sampling, *Journal of the American Statistical Association*, 88, 1993, pp. 1013-1020.

[7] C. E. Särnda, The calibration approach in survey theory and pratice, *Survey Methodology*, 33, 2007, pp. 99-119.

[8] J. LeGuennec and C. E. Sautory, CALMAR2: une nouvelle version de la macro CALMAR de redressement d'échantillon parcalage, Actes des Journées de Méthodologie, INSEE: Pairs, 2002.

[9] C. Vanderhoeft, E. Waeytens and J. M. Museux, Generalised calibration with SPSS 9.0 for Windows baser, In Enqutes, Modles et Applications (Eds. J. J. Droesbeke and L. Lebart): Paris, Dunod, 2001.

[10] C. Vanderhoeft, Generalised calibration at Statistics Belgium, http://www. statbel. fgov.be/studies/paper03en. asp, 2012.

[11] N. J. Nieuwenbroek and H. J. Boonstra, Bascula 4.0 for weighting sample survey data with estimation of variances, The Survey Statistician, 2002.

[12] Y. Tillé and A. Matei, Sampling: Survey Sampling, http://cran.rproject.org/ package=sampling, 2012.

[13] R. T. Münnich, E. W. Sachs and M. Wagner, Calibration of estimators-weights via semismooth Newton method, *Journal of Global Optimization*, 52, 2011, pp. 471-485..

[14] L. Q. Qi and J. Sun, A nonsmooth version of Newton's method, *Math. Program*, 58, 1993, pp. 353-367.

[15] L. Q. Qi, D. F. Sun and G. L. Zhou, A New Look at Smoothing Newton Methods for Nonlinear Complementarity Problems and Box Constrained Variational Inequalities, *Math. Program*, 87, 2000, pp. 1-35.

[16] Z. H. Huang and W. Z. Gu, A smooth-type algorithm for solving linear complementarity problems with strong convergence properties. *Applied Mathematics and Optimization*, 57, 2008, pp. 17-29.

[17] L. Y. Lu and W. Z. Gu, A non-interior continuation algorithm for the CP based on a generalized smoothing function. *Journal of Computational and Applied Mathematics*, 235, 2011, pp. 2300-2313.

[18] Z. H. Huang, S. L. Hu and J. Y. Han, Convergence of a smoothing algorithm for symmetric cone complementarity problems with a nonmonotone line search, *Science in China (Series A)*, 52, 2009, pp.833-848.