# 行政院國家科學委員會補助專題研究計畫成果報告

※※※※※※※※※※※※※※※※※※※※※※※※※
※　　　　　　　　　　　　　　　　　　　　　　　※
※　　　　　　　　**修正估計函數**　　　　　　　　※
※　　　　　　　　　　　　　　　　　　　　　　　※
※※※※※※※※※※※※※※※※※※※※※※※※※

**計畫類別：**P**個別型計畫**　　□**整合型計畫**
**計畫編號：**NSC 89ー2118ーMー009ー014ー
**執行期間：**89 **年** 8 **月** 1 **日至** 90 **年** 7 **月** 31 **日**

**計畫主持人：陳志榮助理教授**
**共同主持人：**

**本成果報告包括以下應繳交之附件：**
□**赴國外出差或研習心得報告一份**
□**赴大陸地區出差或研習心得報告一份**
□**出席國際學術會議心得報告及發表之論文各一份**
□**國際合作研究計畫國外研究報告書一份**

**執行單位：**

**中　華　民　國**　90 **年** 10 **月** 31 **日**

# 行政院國家科學委員會專題研究計畫成果報告

## 修正估計函數

## Modified Estimating Functions

**計畫編號：NSC 89-2118-M-009-014**

**執行期限：89 年 8 月 1 日至 90 年 7 月 31 日**

**主持人：陳志榮助理教授　　國立交通大學統計學研究所**

**共同主持人：**

**計畫參與人員：**

### 一、中文摘要

**在這篇文章中，我們利用 Firth (1993) 的方法提出一個修正估計函數來降低估計量的漸近均方誤差。然後，我們給一個廣義最小平方法的例子來闡明此理論的有用性。**

**關鍵詞：有興趣的參數、有偏的估計函數、漸近均方誤差、修正估計函數**

### Abstract

In this paper, we utilize Firth's (1993) method to propose a modified estimating function to reduce the asymptotic mean square error of the original estimator. Then, we give an example from the generalized least squares method to illustrate the usefulness of this theory.

**Keywords**: Parameter of Interest, Biased Estimating Function, Asymptotic Mean Square Error, Modified Estimating Function

### 二、計畫緣由與目的

In this paper, consider a family $\{\boldsymbol{E}_n: n\in \boldsymbol{N}\}$ of experiments defined on a statistical space $(\Omega, \boldsymbol{F}, \{P_\theta: \theta\in\Theta\})$, where $\boldsymbol{N}=\{1,2,\ldots\}$, $\Omega$ is the sample space, $\boldsymbol{F}$ is the $\sigma$-field generated by some subsets of $\Omega$, and $P_\theta$ is a complete probability measure on $(\Omega, \boldsymbol{F})$ indexed by the parameter $\theta$ in the parameter space $\Theta$. Suppose that $\psi(\theta)$ is the $p$-dimensional parameter vector of interest and that $\lambda(\theta)$ is a $q$-dimensional nuisance parameter vector. Set $\Psi=\{\psi(\theta): \theta\in\Theta\}$ and $\Lambda=\{\lambda(\theta): \theta\in\Theta\}$. For simplicity of notation, set $\psi=\psi(\theta)$ and $\lambda=\lambda(\theta)$. Suppose that $\Psi$ and $\Lambda$ are open subsets of $\boldsymbol{R}^p$ and $\boldsymbol{R}^q$, respectively. Set $\boldsymbol{S}_n=Y_n(\Omega)$. For each $n\in\boldsymbol{N}$, let $\boldsymbol{G}_n$ be the family of $\boldsymbol{R}^p$-valued estimating functions $G_n$ for $\psi$ with respect to $\boldsymbol{E}_n$ such that for each $\theta\in\Theta$, the following conditions hold:

(i) $G_n'(\psi,Y_n)(\equiv\partial G_n(\psi,Y_n)/\partial\psi^T)$ exists $P_\theta$-a.s.;

(ii) both $E_\theta(G_n'(\psi,Y_n))$ and $Cov_\theta(G_n(\psi,Y_n))$ exist and are non-singular.

Suppose that for large $n$, $\hat{\psi}_n$ is a root of the estimating equation $G_n(\hat{\psi}_n,Y_n)=0$ for $\psi$, where $G_n\in\boldsymbol{G}_n$. The basis of the present work is the idea that the asymptotic mean square error of $\hat{\psi}_n$ can be reduced by adding some $A_n(\psi,Y_n)$ to $G_n(\psi,Y_n)$ such that the asymptotic mean square error of the root $\psi^*_n$ of the modified estimating equation $G_n^*(\psi^*_n,Y_n)=G_n(\psi^*_n,Y_n)+A_n(\psi^*_n,Y_n)=0$ is smaller than that of $\hat{\psi}_n$, where $A_n$ is some $\boldsymbol{R}^p$-valued function on $\Psi\times\boldsymbol{S}_n$.

In what follows, we present the classical theory with some appropriate technical conditions. For simplicity of notation, set $G_n(\psi)=G_n(\psi,Y_n)$, $A_n(\psi)=A_n(\psi,Y_n)$ and $G_n^*(\psi)=G_n^*(\psi,Y_n)$. Assume that the following conditions hold:

**Condition 1.1** For each $¯ \in n$, $m_n(¯)$ and all eigenvalues of $Cov_{¯}(G_n(\cdot))$ are of the same order as $n \grave{\imath} \infty$, where $m_n(¯) \to \infty$ as $n \to \infty$.

**Condition 1.2** For each $¯ \in n$, $E_{¯}(G_n(\cdot)) = O(m_n^{c}(¯))$ as $n \to \infty$ for some real number $c$.

**Condition 1.3** For each $¯ \in \Theta$, $m_n^{\max\{1,c\}}(¯)$ and all eigenvalues of $E_{¯}(G_n'(\cdot))$ are of the same order as $n \to \infty$.

**Condition 1.4** Under each $P_{¯}$, $G_n'(\cdot) = E_{¯}(G_n'(\cdot)) + o_p(m_n^{\max\{1,c\}}(¯))$ as $n \to \infty$.

**Condition 1.5** Under each $P_{¯}$, $G_n(\hat{\cdot}_n) = o_p(m_n^{\max\{1/2,c\}}(¯))$, $\hat{\cdot}_n = \cdot + o_p(1)$ and $G_n(\hat{\cdot}_n) = G_n(\cdot) + G_n'(\cdot)(\hat{\cdot}_n - \cdot) + o_p(m_n^{\max\{1/2,c\}}(¯))$ as $n \to \infty$.

**Theorem 1.1** Suppose that Conditions 1.1-1.5 hold. Then $c<1$ and the following hold:

(1) Under each $P_{¯}$, $\hat{\cdot}_n - \cdot = -[E_{¯}(G_n'(\cdot))]^{-1} G_n(\cdot) + R_n(¯) = O_p(m_n^{\max\{-1/2,c-1\}}(¯))$ and $(\hat{\cdot}_n - \cdot)(\hat{\cdot}_n - \cdot)^T = [E_{¯}(G_n'(\cdot))]^{-1} G_n(\cdot) G_n^T(\cdot) [E_{¯}(G_n'^T(\cdot))]^{-1} + S_n(¯) = O_p(m_n^{\max\{-1,2(c-1)\}}(¯))$ as $n \to \infty$, where $R_n(¯) = o_p(m_n^{\max\{-1/2,c-1\}}(¯))$ and $S_n(¯) = o_p(m_n^{\max\{-1,2(c-1)\}}(¯))$ as $n \to \infty$.

(2) For each $¯ \in n$, if $E_{¯}(R_n(¯)) = o(m_n^{\max\{-1/2,c-1\}}(¯))$ as $n \to \infty$, then $E_{¯}(\hat{\cdot}_n - \cdot) = -[E_{¯}(G_n'(\cdot))]^{-1} E_{¯}(G_n(\cdot)) + o(m_n^{\max\{-1/2,c-1\}}(¯)) = O(m_n^{\max\{-1/2,c-1\}}(¯))$ as $n \to \infty$.

(3) For each $¯ \in n$, if $E_{¯}(S_n(¯)) = o(m_n^{\max\{-1,2(c-1)\}}(¯))$ as $n \to \infty$, then $E_{¯}[(\hat{\cdot}_n - \cdot)(\hat{\cdot}_n - \cdot)^T] = [E_{¯}(G_n'(\cdot))]^{-1} [Cov_{¯}(G_n(\cdot)) + E_{¯}(G_n(\cdot)) E_{¯}(G_n^T(\cdot))] [E_{¯}(G_n'^T(\cdot))]^{-1} + o(m_n^{\max\{-1,2(c-1)\}}(¯)) = O(m_n^{\max\{-1,2(c-1)\}}(¯))$ as $n \to \infty$.

## 三、結果與討論

In this section, consider a fairly general modification of the estimating function $G_n$ for $\cdot$ with respect to $\boldsymbol{E}_n$, of the form $G_n^* = G_n + A_n$, where $A_n$ is some $\boldsymbol{R}^p$-valued function on $\Psi \times \boldsymbol{S}_n$ such that $A_n'(\cdot, Y_n)$ $(\equiv \hat{o} A_n(\cdot, Y_n)/\hat{o}\cdot^T)$ exists $P_{¯}$-a.s. For each $¯ \in n$ and $n \in \boldsymbol{N}$, set $A_n(\cdot) = A_n(\cdot, Y_n)$, $A_n'(\cdot) = A_n'(\cdot, Y_n)$, $G_n^*(\cdot) = G_n^*(\cdot, Y_n)$ and $G_n^{*\prime}(\cdot) = G_n'(\cdot) + A_n'(\cdot)$. First of all, we assume that Conditions 1.1-1.4 and the following hold:

**Condition 2.1** For each $¯ \in n$, $A_n(\cdot) = -E_{¯}(G_n(\cdot)) + O_p(m_n^{c-1/2}(¯))$ and $A_n'(\cdot) = -\hat{o} E_{¯}(G_n(\cdot))/\hat{o}\cdot^T + O_p(m_n^{c-1/2}(¯)) = O_p(m_n^{c}(¯))$ as $n \to \infty$ such that $E_{¯}(A_n(\cdot)) = -E_{¯}(G_n(\cdot)) + O(m_n^{c-1/2}(¯))$, $Cov_{¯}(A_n(\cdot)) = O(m_n^{2c-1}(¯))$ and $E_{¯}(A_n'(\cdot)) = -\hat{o} E_{¯}(G_n(\cdot))/\hat{o}\cdot^T + O(m_n^{c-1/2}(¯)) = O(m_n^{c}(¯))$ as $n \to \infty$.

**Condition 2.2** For each $¯ \in n$, $m_n^{\max\{1,c\}}(¯)$ and all eigenvalues of $E_{¯}(G_n^{*\prime}(\cdot))$ are of the same order as $n \to \infty$.

**Condition 2.3** Under each $P_{¯}$, $G_n^*(\cdot^*_n) = o_p(m_n^{\max\{1/2,c-1/2\}}(¯))$, $\cdot^*_n = \cdot + o_p(1)$ and $G_n^*(\cdot^*_n) = G_n^*(\cdot) + G_n^{*\prime}(\cdot)(\cdot^*_n - \cdot) + o_p(m_n^{\max\{1/2,c-1/2\}}(¯))$ as $n \to \infty$.

**Theorem 2.1** Suppose that Conditions 1.1-1.4 and 2.1-2.3 hold. Then the following hold:

(1) Under each $P_{¯}$, $\cdot^*_n - \cdot = -[E_{¯}(G_n^{*\prime}(\cdot))]^{-1} G_n^*(\cdot) + R_n^*(¯) = O_p(m_n^{-1/2}(¯))$ and $(\cdot^*_n - \cdot)(\cdot^*_n - \cdot)^T = [E_{¯}(G_n^{*\prime}(\cdot))]^{-1} G_n^*(\cdot) G_n^{*T}(\cdot) [E_{¯}(G_n^{*\prime T}(\cdot))]^{-1} + S_n^*(¯) = O_p(m_n^{-1}(¯))$ as $n \to \infty$, where $R_n^*(¯)$

$=o_p(m_n^{-1/2}(¯))$ and $S_n^*(¯)=o_p(m_n^{-1}(¯))$ as $n\to\infty$.

(2) For each ¯∈ *n*, if $E¯(R_n^*(¯)) =o(m_n^{-1/2}(¯))$ as $n\to\infty$, then $E¯(·^*_n-·) =-[E¯(G_n^{*\prime}(·))]^{-1}E¯(G_n^*(·))+o(m_n^{-1/2}(¯)) =O(m_n^{\min\{-1/2,c-3/2\}}(¯))+o(m_n^{-1/2}(¯))$ as $n\to\infty$.

(3) For each ¯∈ *n*, if $E¯(S_n^*(¯))=o(m_n^{-1}(¯))$ as $n\to\infty$, then $E¯[(·^*_n-·)(·^*_n-·)^T] =[E¯(G_n^{*\prime}(·))]^{-1} [Cov¯(G_n^*(·))+E¯(G_n^*(·))E¯(G_n^{*T}(·))] [E¯(G_n^{*\prime T}(·))]^{-1}+o(m_n^{-1}(¯))=O(m_n^{-1}(¯))$ as $n\to\infty$.

As an example, suppose that under each *P*¯ and for each $n\in \boldsymbol{N}$, $Y_n$ has mean $«_n(·)$ and covariance $ªρ_n(·)$, where $Y_n$ is the response up to time *n*, $«_n$ is a known $\boldsymbol{R}^n$-valued function on Ψ, and $ρ_n$ is a known $n\times n$ positive definite covariance matrix function on Ψ. In this case, ¯ is the parameter, ·(¯)(≡·) is the *p*-dimensional parameter vector of interest and ª(¯) (≡ª) is a positive nuisance parameter. Set $·=(·_1,\ldots,·_p)^T$ and $ª=(ª_1,\ldots,ª_q)^T$.

For each $n\in \boldsymbol{N}$, set $G_n^{\sim}(·) =G_n^{\sim}(·,Y_n)=«_n^T(·)ρ_n^{-1}(·)[Y_n-«_n(·)]$. Then the classical iteratively reweighted generalized least-squares estimator $·^{\sim}_n$ of · is the root of the asymptotically optimal linear unbiased estimating equation $G_n^{\sim}(·^{\sim}_n)=0$ for ·. Set $i(·) =«_n^T(·)ρ_n^{-1}(·)«_n(·)$. Then, under regularity conditions we have $E¯[(·^{\sim}_n-·)(·^{\sim}_n-·)^T]= ªi^{-1}(·)+o(n^{-1})=O(n^{-1})$ as $n\to\infty$.

However, the classical non-iteratively reweighted least-squares estimator $·^{\wedge}_n$ of · is in general inconsistent, where $·^*_n$ minimizes the generalized sum of squares $GSS(·,Y_n) =[Y_n-«_n(·)]^Tρ_n^{-1}(·)[Y_n-«_n(·)]$. Set $G_n(·) =G_n(·,Y_n)=\hat{o}GSS(·,Y_n)/\hat{o}·$. Then $·^{\wedge}_n$ is the root of the estimating equation $G_n(·^{\wedge}_n)=0$ for ·. Note that $G_n(·)=-2G_n^{\sim}(·) +([Y_n-«_n(·)]^T\hat{o}ρ_n^{-1}(·)/\hat{o}·_1[Y_n-«_n(·)],\ldots, [Y_n-«_n(·)]^T\hat{o}ρ_n^{-1}(·)/\hat{o}·_p[Y_n-«_n(·)])^T$. Then $E¯(G_n(·))=ª(\mathrm{tr}[ρ_n(·)\hat{o}ρ_n^{-1}(·)/\hat{o}·_1], \ldots,\mathrm{tr}[ρ_n(·)\hat{o}ρ_n^{-1}(·)/\hat{o}·_p])^T$, $G_n'(·)=2i(·) +([Y_n-«_n(·)]^T\hat{o}^2ρ_n^{-1}(·)/\hat{o}·_i\hat{o}·_j[Y_n-«_n(·)]) +W_n(·)$ and $E¯(G_n'(·))=2i(·) +ª(\mathrm{tr}[ρ_n(·)\hat{o}^2ρ_n^{-1}(·)/\hat{o}·_i\hat{o}·_j])$, where $W_n(·)=O_p(n^{-1/2})$ is the remainder term. In general, $·^{\sim}_n$ is more efficient than $·^{\wedge}_n$.

First of all, assume that the nuisance parameter ª is known. For example, when each component of $Y_n$ is binomial or Poisson distributed, where the relationship between mean and variance is known.

*Case 1*: For each $n\in \boldsymbol{N}$, set $A_n(·)=-E¯(G_n(·))$, $G_n^*=G_n+A_n$ and $G_n^*(·^*_n) =0$. Then $-ªG_n^*(·)/2$ is the score estimating function for · when $Y_n$ is jointly normally distributed. Thus $·^*_n$ is the maximum likelihood estimator of · when $Y_n$ is jointly normally distributed. Therefore, $·^*_n$ is more efficient than $·^{\sim}_n$ when $Y_n$ is approximately jointly normally distributed.

*Case 2*: For each $n\in \boldsymbol{N}$, set $A_n(·) =-([Y_n-«_n(·)]^T\hat{o}ρ_n^{-1}(·)/\hat{o}·_1[Y_n-«_n(·)],\ldots, [Y_n-«_n(·)]^T\hat{o}ρ_n^{-1}(·)/\hat{o}·_p[Y_n-«_n(·)])^T$, $G_n^* =G_n+A_n$ and $G_n^*(·^*_n)=0$. Then $G_n^*=-2G_n^{\sim}$. Thus $·^*_n=·^{\sim}_n$ is the classical iteratively reweighted generalized least-squares estimator $·^{\sim}_n$ of ·.

Now consider the general case where the nuisance parameter ª is unknown. First choose any estimator $ª^*_{·,n}$ of ª such that $ª^*_{·,n}=ª+O_p(n^{-1/2})$ as $n\to\infty$. For example, one such $ª^*_{·,n}$ can be chosen as the root of the generalized estimating equation (GEE) for ª:
$\hat{o}h_n(·,ª)/\hat{o}ª \quad W_n(·,ª)[H_n(·,Y_n)-h_n(·,ª)]=0$, where $H_n(·,Y_n)$ is the vector of different components of $[Y_n-«_n(·)][Y_n-«_n(·)]^T$, $h_n(·,ª)=E¯(H_n(·,Y_n))$ and $W_n(·,ª)$ is some working covariance matrix. In the following, we choose $ª^*_{·,n}$ as the moment estimator $[Y_n-«_n(·)]^Tρ_n^{-1}(·)[Y_n-«_n(·)]/n$ of ª.

*Case 1*: For each $n \in N$, set $B_n(\cdot, ^a) = -E^-(G_n(\cdot))$, $A_n(\cdot) = B_n(\cdot, ^{a^*}{}_{\cdot,n})$, $G_n^* = G_n + A_n$ and $G_n^*(\cdot^*_n) = 0$. Then $-{}^{a^*}{}_{\cdot,n} G_n^*(\cdot)/2$ is the profile score estimating function for · when $Y_n$ is jointly normally distributed. Thus $\cdot^*_n$ is the maximum likelihood estimator of · when $Y_n$ is jointly normally distributed. Therefore, $\cdot^*_n$ is more efficient than $\cdot^{\sim}_n$ when $Y_n$ is approximately jointly normally distributed.

*Case 2*: For each $n \in N$, set $A_n(\cdot) = -([Y_n - «_n(\cdot)]^T \hat{o}p_n^{-1}(\cdot)/\hat{o}\cdot_1[Y_n - «_n(\cdot)], \ldots, [Y_n - «_n(\cdot)]^T \hat{o}p_n^{-1}(\cdot)/\hat{o}\cdot_p[Y_n - «_n(\cdot)])^T$, $G_n^* = G_n + A_n$ and $G_n^*(\cdot^*_n) = 0$. Then $G_n^* = -2G_n^{\sim}$. Thus $\cdot^*_n = \cdot^{\sim}_n$ is the classical iteratively reweighted generalized least-squares estimator $\cdot^{\sim}_n$ of ·.

For the general finite-dimensional nuisance parameter case, it is similar to the above one-dimensional case.

## 四、計畫成果自評

In this paper, we propose a modified estimating function for the parameter vector · of interest to reduce the asymptotic mean square error of the original estimator of ·. It is found from the example that the classical iteratively reweighted generalized least-squares estimator of · can be derived from the heuristic inconsistent classical non-iteratively reweighted generalized least-squares estimator of ·. Moreover, when the data is jointly normally distributed, the maximum likelihood estimator of · can also be derived from the heuristic inconsistent classical non-iteratively reweighted generalized least-squares estimator of ·. Thus, more efficient estimators of · than the original estimator of · may be obtained by utilizing our proposed method to modify the original estimating function for ·. Hence, this paper can be treated as a supplement to the theory of estimating functions. Consequently, it deserves reading when studying and/or doing the research in estimation of the parameter vector of interest in the presence of nuisance parameter vector.

## 五、參考文獻


[1] Crowder, M. (1987). On linear and quadratic estimating functions. *Biometrika* **74**, 591-597.

[2] Firth, D. (1987). On efficiency of quasi-likelihood estimation. *Biometrika* **74**, 233-246.

[3] Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 27-38.

[4] Godambe, V. P. (Ed.) (1991). *Estimating functions.* Oxford Science Publications, Oxford.

[5] Godambe, V. P. and Heyde, C. C. (1987). Quasi-likelihood and optimal estimation. *Int. Statist. Rev.* **55**, 231-244.

[6] Godambe, V. P. and Thompson, M. E. (1974). Estimating equations in the presence of a nuisance parameter. *Ann. Statist.* **2**, 568-571.

[7] Heyde, C. C. (1988). Quasi-likelihood and optimality for estimating functions: some current unifying themes. *Bull. Internat. Statist. Inst.* **53**, 19-29.

[8] Heyde, C. C. (1997). *Quasi-Likelihood and its Application: A General Approach to Optimal Parameter Estimation.* Springer, New York.

[9] McCullagh, P. (1991). Quasi-likelihood and estimating functions. In D. V. Hinkley, N. Reid and E. J. Snell, Eds., *Statistical Theory and Modeling: In Honour of Sir David Cox, FRS.* Chapman and Hall, London, 256-286.

[10] Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* **61**, 439-447.