CIRJE-F-184

# Improving Small Sample Properties of the Empirical Likelihood Estimation

Naoto Kunitomo
University of Tokyo

November 2002

# Improving Small Sample Properties of the Empirical Likelihood Estimation *

Naoto Kunitomo†

November 2002

**Abstract**

We propose to use a simple modification of the maximum empirical likelihood (MEL) method for estimating structural equations in econometrics. The modified estimator improves both the asymptotic bias and the mean squared error of the MEL estimator in the orders of $O(n^{-1})$ and $O(n^{-2})$, respectively, at the same time. It also improves the asymptotic bias of the generalized method of moments (GMM) estimation (or the estimating equation (EE) method) significantly when there are many instruments in the econometric literatures.

**Key Words**

Modified Empirical Likelihood Method, Estimating Structural Equation, Asymptotic Bias, Mean Squared Error, LIML Method, GMM Method

**JEL Code**: C-13, C-30

*The author thanks Mr. Yukitoshi Matsushita for his assistance on preparing numerical examples in Section 6 of the present work.

†Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, JAPAN.

## 1. Introduction

The study of estimating a single structural equation in econometric models has led to develop several estimation methods as the alternatives to the least squares estimation method. The classical examples in the econometric literatures are the limited information maximum likelihood (LIML) method and the instrumental variables (IV) method including the two-stage least squares (TSLS) method. See Anderson, Kunitomo, and Sawa (1982) and Anderson, Kunitomo, and Morimune (1986) for their finite sample properties, for instance. In addition to these classical methods the maximum empirical likelihood (MEL) method has been proposed and has gotten some attention recently in the statistical and econometric literatures. It is probably because the MEL method gives asymptotically efficient estimator in the semi-parametric sense and also improves the serious bias problem known in the estimating equation method or the generalized method of moments (GMM) method when the number of instruments is large in econometric models. See Owen (2001), Qin and Lawless (1994), and Kitamura, Tripathi, and Ahn (2001) on the details of the MEL method.

The main purpose of this study is to propose a modification of the MEL estimation method for estimating a single structural equation and show that it improves the small sample properties of the MEL estimator. Our modification method is simple and it has an intuitive interpretation. Thus it is quite appealing from the views of theory as well as practice. We shall show that the modified MEL estimator (which is abbreviated as the MMEL estimator) we are proposing in this paper has not only the smaller asymptotic bias in the order of $O(n^{-1})$ but also the smaller asymptotic mean squared errors in the order of $O(n^{-2})$ than the original MEL estimator at the same time where $n$ is the sample size. Thus the MMEL estimation method we are proposing dominates the MEL estimation method in the asymptotic higher order sense. Also by investigating a set of simulations systematically we have found that the modified MEL estimator has better small sample properties in the sense of the bias, the mean squared error, and the probability concentration than the MEL estimator in all cases.

In the econometric literatures the generalized method of moments (GMM) estimation method has been quite popular in the past decade. The GMM method was originally proposed by Hansen (1982) in the econometric literature and it is essentially the same as the estimating equation (EE) method proposed by Godambe (1960) which has been used in statistical applications. This approach has an attractive feature that it has rather broad applicability and it is easily implemented in statistical analyses. However, it has been known that there is a serious bias problem in the GMM estimation when there are many instruments in econometric models. In this respect we should notice that the MMEL estimator we are proposing is quite similar to the MEL estimator when there are many instruments. Hence the MMEL estimation method improves the MEL estimation while it retains the good small sample properties of the MEL estimation method. In our limited simulations the MMEL estimator has better small sample properties in the sense of the bias, the mean squared error, and the probability concentration than the GMM estimator in all cases when the number of instruments is large. Therefore our new method is quite attractive for the problem of estimating econometric models in the semi-parametric sense.

In Section 2 we state the estimation problem and the maximum empirical likelihood

(MEL) estimation method. In Section 3 we shall give a modified MEL estimation method for the problem of estimating a linear structural equation when there are instrumental variables and the disturbances are homoscedastic. Then in Sections 4 and 5 we shall discuss the modified MEL estimation method for the heteroscedastic disturbance case and the nonlinear structural equation case, respectively. In Section 6 we give some numerical examples and some conclusions are given in Section 7. The more detailed derivations of our results are given in the Appendix.

## 2. Estimating a Single Structural Equation by the Maximum Empirircal Likelihood Method

Let a single equation in the econometric model be given by

$$y_{1i} = h(\mathbf{y}_{2i}, \mathbf{z}_{1i}, \theta) + u_i \quad (i = 1, \cdots, n) , \tag{2.1}$$

where $h(\cdot,\cdot,\cdot)$ is a function, $y_{1i}$ and $\mathbf{y}_{2i}$ are $1 \times 1$ and $G_1 \times 1$ (vector of) endogenous variables, $\mathbf{z}_{1i}$ is a $K_1 \times 1$ vector of exogenous variables, $\theta$ is an $r \times 1$ vector of unknown parameters, and $\{u_i\}$ are mutually independent disturbance terms with $E(u_i) = 0$ $(i = 1, \cdots, n)$.

We assume that (2.1) is the first equation in a system of $(G_1 + 1)$ structural equations relating the vector of $G_1 + 1$ endogenous variables $\mathbf{y}_i^{'} = (\mathbf{y}_{1i}, \mathbf{y}_{2i}^{'})$ and the vector of $K$ $(= K_1 + K_2)$ exogenous variables $\{\mathbf{z}_i\}$ which includes $\{\mathbf{z}_{1i}\}$ . The set of exogenous variables $\{\mathbf{z}_i\}$ are often called the instrumental variables and we have the orthogonality condition

$$E(u_i \, \mathbf{z}_i) = \mathbf{0} \quad (i = 1, \cdots, n) . \tag{2.2}$$

Because we do not specify the equations except (2.1) and we only have the limited information on the set of instrumental variables or instruments, we only consider the limited information estimation methods. When the function $h(\cdot,\cdot,\cdot)$ is of the linear form, (2.1) can be written as

$$y_{1i} = (\mathbf{y}_{2i}^{'}, \mathbf{z}_{1i}^{'}) \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + u_i \quad (i = 1, \cdots, n), \tag{2.3}$$

where $\theta^{'} = (\beta^{'}, \gamma^{'})$ is a $1 \times p$ $(p = K_1 + G_1)$ vector of unknown coefficients.

Furthermore, when all structural equations in the econometric model are linear, the reduced form equations of $\mathbf{y}_i^{'} = (\mathbf{y}_{1i}, \mathbf{y}_{2i}^{'})$ can be defined by

$$\mathbf{y}_i = \mathbf{\Pi}^{'} \mathbf{z}_i + \mathbf{v}_i \quad (i = 1, \cdots, n) , \tag{2.4}$$

where $\mathbf{v}_i^{'} = (v_{1i}, \mathbf{v}_{2i}^{'})$ is a $1 \times (1 + G_1)$ disturbance terms with $E[\mathbf{v}_i] = \mathbf{0}^{'}$ and

$$\mathbf{\Pi}^{'} = \begin{pmatrix} \pi_1^{'} \\ \mathbf{\Pi}_2^{'} \end{pmatrix} \tag{2.5}$$

is a $(1 + G_1) \times K$ partitioned matrix of the linear reduced form coefficients. By multiplying $(1, -\beta^{'})$ from the left-hand side, we have the restriction

$$(1, -\beta^{'}) \mathbf{\Pi}^{'} = (\gamma^{'}, \mathbf{0}^{'}) . \tag{2.6}$$

The maximum empirical likelihood (MEL) estimator for the vector of unknown parameters $\theta$ in (2.1) is defined by maximizing the Langrangian form

$$L_n^*(\lambda,\theta) = \sum_{i=1}^{n} \log p_i - \mu(\sum_{i=1}^{n} p_i - 1) - n\lambda^{'} \sum_{i=1}^{n} p_i \ \mathbf{z}_i[y_{1i} - h(\mathbf{y}_{2i}, \mathbf{z}_{1i}, \theta)] \ , \tag{2.7}$$

where $\mu$ and $\lambda$ are a scalor and a $K \times 1$ vector of Lagrangian multipliers, and $p_i$ ($i = 1, \cdots, n$) are the weighted probability functions to be chosen. It has been known (see Qin and Lawles (1994) or Owen (2001)) that the above maximization problem is the same as to maximize

$$L_n(\lambda,\theta) = -\sum_{i=1}^{n} \log\{1 + \lambda^{'}\mathbf{z}_i \ [y_{1i} - h(\mathbf{y}_{2i}, \mathbf{z}_{1i}, \theta)]\} \ , \tag{2.8}$$

where we have the conditions $\hat{\mu} = n$ , and

$$[n\hat{p}_i]^{-1} = 1 + \lambda^{'}\mathbf{z}_i[y_{1i} - h(\mathbf{y}_{2i}, \mathbf{z}_{1i}, \theta)] \ . \tag{2.9}$$

By differentiating (2.7) with respect to $\lambda$ and combining the resulting equation with (2.9), we have the relation

$$\sum_{i=1}^{n} \hat{p}_i \mathbf{z}_i \ [y_{1i} - h(\mathbf{y}_{2i}, \mathbf{z}_{1i}, \theta)] = \mathbf{0} \tag{2.10}$$

and

$$\hat{\lambda} = [\sum_{i=1}^{n} \hat{p}_i u_i^2(\hat{\theta}) \mathbf{z}_i \mathbf{z}_i^{'}]^{-1} [\frac{1}{n} \sum_{i=1}^{n} u_i(\hat{\theta}) \mathbf{z}_i] \ , \tag{2.11}$$

where $u_i(\hat{\theta}) = y_{1i} - h(\mathbf{y}_{2i}, \mathbf{z}_{1i}, \hat{\theta})$ and $\hat{\theta}$ is the maximum empirical likelihood (MEL) estimator for the vector of unknown parameters $\theta$ . From (2.7) the MEL estimator of $\{\theta\}$ is the solution of the set of $p$ equations

$$\hat{\lambda}^{'} \sum_{i=1}^{n} \hat{p}_i \mathbf{z}_i [-\frac{\partial h(\mathbf{y}_{2i}, \mathbf{z}_{1i}, \hat{\theta})}{\partial \theta_j}] = 0 \ (j = 1, \cdots, p) \ . \tag{2.12}$$

When the structural equation of (2.1) is linear, the MEL estimator of the vector of coefficient parameters $\theta^{'} = (\beta^{'}, \gamma^{'})$ can be simplified as the solution of

$$\begin{aligned} & [\sum_{i=1}^{n} \hat{p}_i \begin{pmatrix} \mathbf{y}_{2i} \\ \mathbf{z}_{1i} \end{pmatrix} \mathbf{z}_i^{'}] [\sum_{i=1}^{n} \hat{p}_i u_i(\hat{\theta})^2 \mathbf{z}_i \mathbf{z}_i^{'}]^{-1} [\frac{1}{n} \sum_{i=1}^{n} \mathbf{z}_i \ y_{1i}] \\ = \ & [\sum_{i=1}^{n} \hat{p}_i \begin{pmatrix} \mathbf{y}_{2i} \\ \mathbf{z}_{1i} \end{pmatrix} \mathbf{z}_i^{'}] [\sum_{i=1}^{n} \hat{p}_i u_i(\hat{\theta})^2 \mathbf{z}_i \mathbf{z}_i^{'}]^{-1} [\frac{1}{n} \sum_{i=1}^{n} \mathbf{z}_i (\mathbf{y}_{2i}^{'}, \mathbf{z}_{1i}^{'})] \begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} \ . \end{aligned} \tag{2.13}$$

If we substitute $1/n$ for $\hat{p}_i$ ($i = 1, \cdots, n$) in (2.13), then we have the generalized method of moments (GMM) estimator for the vector of coefficient parameters $\theta^{'} = (\beta^{'}, \gamma^{'})$, which is the solution of

$$\begin{aligned} & [\frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} \mathbf{y}_{2i} \\ \mathbf{z}_{1i} \end{pmatrix} \mathbf{z}_i^{'}] [\frac{1}{n} \sum_{i=1}^{n} u_i(\hat{\theta})^2 \mathbf{z}_i \mathbf{z}_i^{'}]^{-1} [\frac{1}{n} \sum_{i=1}^{n} \mathbf{z}_i \ y_{1i}] \\ = \ & [\frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} \mathbf{y}_{2i} \\ \mathbf{z}_{1i} \end{pmatrix} \mathbf{z}_i^{'}] [\frac{1}{n} \sum_{i=1}^{n} u_i(\hat{\theta})^2 \mathbf{z}_i \mathbf{z}_i^{'}]^{-1} [\frac{1}{n} \sum_{i=1}^{n} \mathbf{z}_i (\mathbf{y}_{2i}^{'}, \mathbf{z}_{1i}^{'})] \begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} \ , \end{aligned} \tag{2.14}$$

where $\hat{\theta}$ is an initial (consistent) estimator of $\theta$ . ( See Hayashi (2000) on the details of the GMM method in econometrics, for instance. )

## 3. A Modified MEL Estimation in the Linear Homoscedastic Case

The most important feature of the MEL estimation is the use of the estimated weight functions $\hat{p}_i$ $(i = 1, \cdots, n)$ . We notice that if we substitute $(1/n)$ for $\hat{p}_i$ $(i = 1, \cdots, n)$, the resulting estimation method is identical to the estimating equations method or the generalized method of moments (GMM) in the econometric literatures. This is a simple fact which lead us to consider a simple modification of the MEL method we are proposing in this paper.

Let

(3.1) $$\hat{p}_i^* = \frac{1}{n[1 + \delta\, \lambda' \mathbf{z}_i u_i(\hat{\theta})]} ,$$

where $\delta$ is a positive constant $(0 \leq \delta \leq 1)$ and $\hat{\theta}$ is the MEL estimator of $\theta$ . Then we can define a modification of the MEL estimation by substituting $\hat{p}_i$ $(i = 1, \cdots, n)$ into (2.9)-(2.11). We shall denote the resulting Lagrangian multiplier and the modified estimator as $\hat{\lambda}^*$ and $\hat{\theta}^*$ .

In the rest of the present paper we shall consider the standardized error of estimators as

(3.2) $$\hat{\mathbf{e}} = \sqrt{n}\begin{pmatrix} \hat{\beta} - \beta \\ \hat{\gamma} - \gamma \end{pmatrix} ,$$

where $\hat{\theta}' = (\hat{\beta}', \hat{\gamma}')$ . We denote $\hat{\mathbf{e}}$ for the MEL estimator and its modification as $\hat{\mathbf{e}}_{EL}$ and $\hat{\mathbf{e}}^*$, respectively.

In this section we consider the situation that the disturbances are homoscedastic random variables. Under a set of regularity conditions, the asymptotic variance-covariance matrix of the asymptotically efficient estimators in the semi-parametric framework is given by

(3.3) $$\mathbf{Q}^{-1} = \mathbf{D}'\mathbf{M}\mathbf{C}^{-1}\mathbf{M}\mathbf{D} ,$$

where $\mathbf{C} = \sigma^2\mathbf{M}$ and

(3.4) $$\mathbf{D} = [\mathbf{\Pi}_2 , \begin{pmatrix} \mathbf{I}_{K_1} \\ \mathbf{O} \end{pmatrix}] ,$$

(3.5) $$\mathbf{M} = \operatorname{plim}_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i\mathbf{z}_i' .$$

Here we have implicitly assumed that $E(u_i^2) = \sigma^2$ $(> 0)$, the (constant) matrix $\mathbf{M}$ is positive definite, and the rank condition

(3.6) $$\mathbf{rank}(\mathbf{D}) = p \ (= G_1 + K_1) .$$

These conditions assure that the limiting variance-covariance matrix $\mathbf{Q}$ is non-degenerate. The rank condition implies the order condition

(3.7) $$L = K - p \geq 0 ,$$

which has been called the degree of overidentification in the econometric literatures.

In order to compare alternative efficient estimation methods in the asymptotic sense, we need to derive the asymptotic expansions of the density functions of the standardized estimators (3.2) in the form of

$$f(\xi) = \phi_{\mathbf{Q}}(\xi)[\,1 + \frac{1}{\sqrt{n}}H_1(\xi) + \frac{1}{n}H_2(\xi)\,] + o(\frac{1}{n})\ , \tag{3.8}$$

where $\xi = (\xi_1, \cdots, \xi_p)'$ , $\phi_{\mathbf{Q}}(\xi)$ is the multivariate normal density function with mean $\mathbf{0}$ and the variance-covariance matrix $\mathbf{Q}$ , and $H_i(\xi)$ $(i = 1, 2)$ are some polynomial functions of elements of $\xi$. For the rest of our arguments we need a set of regularity conditions.

**Assumption I** :
(i) The sequence of random vectors $\{u_i\ , \mathbf{v}_i'\}$ are independently and identically distributed, and their sixth order moments are bounded.
(ii) We have the rank condition given by (3.6).
(iii) The sequence of exogenous variables $\{\mathbf{z}_i\}$ are random vectors or non-random (i.e. deterministic) vectors, but they are i.i.d. random variables being independent of $\{u_i\}$, and their sixth order moments are bounded in the first case. They satisfy (iii-a)

$$\frac{1}{n}\max_{1\leq i\leq n} |\mathbf{z}_i|^2 \xrightarrow{p} 0 \tag{3.9}$$

as $n \longrightarrow \infty$ , (iii-b) there exists (constant) positive definite matrix $\mathbf{M}$ such that

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i\mathbf{z}_i'\ (= \mathbf{M}_n) = \mathbf{M} + O_p(\frac{1}{\sqrt{n}})\ , \tag{3.10}$$

and (iii-c) we have the condition

$$\frac{1}{n}\sum_{i=1}^{n} E[z_i^{(j)} z_i^{(k)} z_i^{(l)} u_i^3] = O(\frac{1}{\sqrt{n}})\ \ , \tag{3.11}$$

where we denote the $K \times 1$ vector $\mathbf{z}_i = (z_i^{(j)})$ .

When $\{\mathbf{z}_i\}$ are a sequence of i.i.d. random vectors, we have the notation $\mathbf{C} = E[u_i^2\mathbf{z}_i\mathbf{z}_i'] = \sigma^2 E(\mathbf{z}_i\mathbf{z}_i')$ $(i = 1, \cdots, n)$.
We notice that the conditions in **Assumption I** are rather strong, but the conditions lead to some simplifications in the derivations and the resulting expressions of the asymptotic bias (ABIAS) and the asymptotic meas suared errors (MSE). Nonetheless, these conditions can be relaxed considerably at the cost of complicated derivations and notations. Some possible directions will be given in the next two sections. We expect that the most of the results we are reporting in this paper are essentially true under a set of the weaker conditions.

We shall use the mean operator $AM_n(\hat{\mathbf{e}})$, which is defined as the mean of $\hat{\mathbf{e}}$ with respect to the asymptotic expansion of its density function of the standardized estimators up to $O(n^{-1})$ . Then we write the asymptotic bias and the asymptotic MSE of the standardized estimator by $ABIAS_n(\hat{\mathbf{e}}) = AM_n(\hat{\mathbf{e}})$ , and

$$AMSE_n(\hat{\mathbf{e}}) = AM_n(\hat{\mathbf{e}}\,\hat{\mathbf{e}}')\ .$$

Furthermore, as an important criterion we shall use the asymptotic probability of concentration (APC)

$$APC_n = P(\ \hat{\mathbf{e}} \in \mathbf{S}\ )\ ,$$

where the integrand is taken with respect to the asymptotic expansion of the density function of estimators up to $O(n^{-1})$ in the form of (3.8) and $\mathbf{S}$ is any star-shaped set. ( Here we define that for any real number $\alpha \in [0,1]$ $\alpha\mathbf{S} \in \mathbf{S}$ if $\mathbf{S}$ is a star-shaped. ) This criterion is important in the present case because there are important cases when the estimators do not posses moments. For instance, it has been known that the LIML estimator does not have any integer order moments. (see Anderson et. al. (1982)).
For the asymptotic bias of the modified MEL estimator, we have the following result and its derivation is given in the Appendix.

**Theorem 3.1** : *Under Assumption I, the asymptotic bias (ABISAS) of* $\hat{\mathbf{e}}^*$ *as* $n \to +\infty$ *is given by*

$$AM_n(\hat{\mathbf{e}}^*) = \frac{1}{\sqrt{n}}\mathbf{Q}\ \mathbf{q}[L - 1 - \delta L]\ , \tag{3.12}$$

*where* $\mathbf{q}$ *is the* $p \times 1$ *vector given by*

$$\mathbf{q} = \frac{1}{\sigma^2}\mathrm{Cov}((\ \begin{matrix}\mathbf{v}_{2i} \\ \mathbf{0}\end{matrix}\ )u_i)\ \ (i = 1, \cdots, n). \tag{3.13}$$

By using this result we immediately observe that the asymptotic bias of the MEL estimator and the GMM estimator are $(-1)\mathbf{Q}\ \mathbf{q}/\sqrt{n}$ and $(L-1)\mathbf{Q}\ \mathbf{q}/\sqrt{n}$, respectively. Then it is possible to remove the asymptotic bias of the MEL estimator by setting

$$\delta^* = \frac{L-1}{L}$$

provided that $L > 0$ . One interpretation of this modification is that $\delta^*$ is a kind of shrinkage factor to the estimated Lagrangian multiplier $\lambda$ and hence the estimators of probability parameters $\{p_i\}$. We now state the main result whose proof is given in the Appendix.

**Theorem 3.2** : *Suppose we choose* $\delta^* L = L - 1$ $(L \geq 1)$ *in the class of modified MEL estimators. Then under Assumption I,*

$$\lim_{n\to\infty} n[AMSE_n(\hat{\mathbf{e}}^*) - AMSE_n(\hat{\mathbf{e}}_{EL})] \leq 0 \tag{3.14}$$

*and*

$$\lim_{n\to\infty} n[APC_n(\hat{\mathbf{e}}^*) - APC_n(\hat{\mathbf{e}}_{EL})] \geq 0 \tag{3.15}$$

*as* $n \to \infty$ *. The strict inequalities in (3.14) and (3.15) hold when* $\mathbf{q} \neq \mathbf{0}$ *in the positive definite sense, where* $\mathbf{q}$ *is given by (3.13).*

## 4. The Linear Heteroscedastic Case

The results reported in Section 3 can be extended to the case when the disturbances are heteroscedastically distributed under a set of additional assumptions. The limiting variance-covariance matrix of the standardized errors for the estimator $\hat{\mathbf{e}}$ in this case shoud be modified as

$$\mathbf{Q}^{-1} = \mathbf{D}^{'}\mathbf{M}\mathbf{C}^{-1}\mathbf{M}\mathbf{D} \ , \tag{4.1}$$

where

$$\mathbf{C} = \lim \frac{1}{n}\sum_{i=1}^{n} E[u_i^2 \mathbf{z}_i \mathbf{z}_i^{'}] \ , \tag{4.2}$$

provided that the (constant) matrix $\mathbf{M}$ is positive definite, the probability limit (constant) matrix $\mathbf{C}$ is positive definite, and $\sup_{1 \le i \le n} E(u_i^2) < +\infty$ . For deriving the asymptotic bias (ABIAS) and the asymptotic mean squared errors (AMSE), we need stronger regularity conditions.

**Assumption II** :
(i) The sequence of random vectors $\{u_i, \mathbf{v}_i^{'}\}$ are a sequence of martingale differences with $E[(u_i, \mathbf{v}_i^{'})|\mathcal{F}_{i-1}] = \mathbf{0}^{'}$ $(i = 1, \cdots, n)$ and their sixth order moments are bounded, where $\mathcal{F}_{i-1}$ is the $\sigma-$field generalted by $(u_j, \mathbf{v}_j^{'})$ $(j \le i-1)$ and $\mathbf{z}_j$ $(j \le i)$ . ( We use the convention that $\mathcal{F}_0$ contains only the null set. ) Also there exists a $p \times 1$ constant vector $\mathbf{q}$ such that $\mathbf{w}_i^{'} = (\mathbf{v}_{2i}^{'}, \mathbf{0}^{'}) - \mathbf{q}^{'} u_i$ and

$$\mathbf{q} = \frac{1}{\sigma_i^2}\mathrm{Cov}(( \begin{array}{c} \mathbf{v}_{2i} \\ \mathbf{0} \end{array} ) \ u_i) + o(\frac{1}{\sqrt{n}}) \ , \tag{4.3}$$

where $E(u_i^2) = \sigma_i^2$ $(i = 1, \cdots, n)$.
(ii) We have the rank condition given by (3.6).
(iii) The sequence of exogenous variables $\{\mathbf{z}_i\}$ are random vectors or non-random (i.e. deterministic) vectors, but they are stationary, ergodic and their sixth order moments are bounded in the former case. Also they satisfy the conditions (3.9)-(3.11) in **Assumption I** and there exists a (constant) positive definite matrix $\mathbf{C}$ such that

$$\frac{1}{n}\sum_{i=1}^{n} u_i^2 \ \mathbf{z}_i \mathbf{z}_i^{'} = \mathbf{C} + O_p(\frac{1}{\sqrt{n}}) \ . \tag{4.4}$$

We notice that while some conditions are automatically satisfied for the heteroscedastic normal disturbances, they can be restrictive. In order to remove the asymptotic bias of the MEL estimator in the more general case, however, we need to have more complicated modifications. In this paper we restrict our discussions to the simple modification method which can be practical for real applications. If we have the situation

$$\mathrm{plim} \frac{1}{n}\sum_{i=1}^{n} E(u_i^2) = \sigma^2 \ (> 0) \tag{4.5}$$

and the variables $\{\mathbf{z}_i\}$ are a sequence of i.i.d. random vectors, and $\mathbf{z}_i \mathbf{z}_i^{'}$ are asymptotically uncorrelated with $u_i^2$ , then we have $\mathbf{C} = \sigma^2 \mathbf{M}$ and the asymptotic variance-covariance matrix reduces to (3.3). For the asymptotic bias of the modified MEL estimator, we have the following result.

**Theorem 4.1** : *Under Assumption II, the asymptotic mean function of* $\hat{\mathbf{e}}^*$ *as* $n \to +\infty$ *is given by*

$$AM_n(\hat{\mathbf{e}}^*) = \frac{1}{\sqrt{n}}\mathbf{Q} \ \mathbf{q}[L - 1 - \delta L] \ . \tag{4.6}$$

By using this result on the asymptotic bias of the general modified MEL estimator, we have the next result on the asymptotic MSE and the asymptotic PC as *Theorem 3.2.*

**Theorem 4.2** : *Suppose we choose* $\delta^* L = L - 1$ $(L \geq 1)$ *in the class of modified MEL estimators. Then under Assumption II,*

$$\lim_{n\to\infty} n[AMSE_n(\hat{e}^*) - AMSE_n(\hat{e}_{EL})] \leq 0 \tag{4.7}$$

*and*

$$\lim_{n\to\infty} n[APC_n(\hat{e}^*) - APC_n(\hat{e}_{EL})] \geq 0 \tag{4.8}$$

*as* $n \to \infty$ . *The strict inequalities in (4.7) and (4.8) hold if* $\mathbf{q} \neq \mathbf{0}$ *in the sense of positive definiteness and* $\mathbf{q}$ *is defined by (4.3).*

## 5. The Nonlinear Case

When the structural equation in (2.1) is nonlinear, we have similar arguments and our modification of the MEL estimation method can be still useful. However, we need complicated notations and a set of additional regularity conditions including the differentiabilities of the function $h(\cdot,\cdot,\cdot)$ . For the nonlinear function $h(\cdot,\cdot,\cdot)$, we use the notation

$$g_i(\theta) = \mathbf{z}_i \; [y_{1i} - h(\mathbf{y}_{2i}, \mathbf{z}_{1i}, \theta)] \;\; (i = 1, \cdots, n) \tag{5.1}$$

and let

$$\mathbf{C} \;=\; \text{plim}[\frac{1}{n}\sum_{i=1}^{n} g_i(\theta_0) g_i^{'}(\theta_0)] \; , \tag{5.2}$$

$$\mathbf{D}(M) \;=\; \text{plim}\frac{1}{n}[-\sum_{i=1}^{n} \partial g_i(\theta_0)] \; , \tag{5.3}$$

where $\theta_0 = (\theta_0^{(i)})$ is a $p \times 1$ vector of true value of the unknown parameters $\theta$ and $\partial g_i(\theta) = (\frac{\partial g_i(\theta)}{\partial \theta_i})$ .
We assume a $K \times K$ matrix $\mathbf{C}$ is positive definite and a $K \times p$ matrix $\mathbf{D}(M)$ is of full rank ( $p \leq K$ ). Then the asymptotic variance-covariance matrix of asymptotically efficient estimators in the semi-parametric sense can be given by

$$\mathbf{Q} = [\mathbf{D}(M)^{'} \mathbf{C}^{-1} \mathbf{D}(M)]^{-1} \; , \tag{5.4}$$

provided that it is non-singular.
In order to derive the asymptotic bias of the MEL estimator, we consider the situation when we have the nonlinear relations

$$\mathbf{y}_{2i} \;=\; \mathbf{\Pi}_2(\mathbf{z}_i, \mathbf{v}_{2i}, \pi_2) \; , \tag{5.5}$$

$$\frac{\partial h_i}{\partial \theta} \;=\; \mathbf{q}\; u_i + m(\mathbf{w}_i, \mathbf{z}_i, \theta) \; , \tag{5.6}$$

where $\pi_2 = (\pi_{2ij})$ is a vector of unknown parameters, $\mathbf{v}_{2i}$ is a vector of random terms,

$$\mathbf{w}_i = m(\mathbf{w}_i, \mathbf{z}_i, \theta) - E[m(\mathbf{w}_i, \mathbf{z}_i, \theta)] \; (i = 1, \cdots, n), \tag{5.7}$$

$p \times 1$ random vectors which are uncorrelated with $u_i$ and $E[\mathbf{w}_i] = \mathbf{0}$ .

**Assumption III** :
(i) The same conditions as $(i)$ in *Assumption II.*
(ii) The $K \times p$ matrix $\mathbf{D}(M)$ is of full rank and it is p.
(iii) We assume the same conditions on $\{\mathbf{z}_i\}$ as $(iii)$ of *Assumption II.* Also the functions $g_i(\theta)$ are twice continuously differentiable and their sixth order moments are bounded. There exist constant matricess $\mathbf{C}$ and $\mathbf{D}(M)$ such that

$$\frac{1}{n}\sum_{i=1}^{n} g_i(\theta)g_i^{'}(\theta) = \mathbf{C} + O_p(\frac{1}{\sqrt{n}}) , \tag{5.8}$$

$$(-1)\frac{1}{n}\sum_{i=1}^{n} \partial g_i(\theta) = \mathbf{D}(M) + O_p(\frac{1}{\sqrt{n}}) . \tag{5.9}$$

(iv) The true value $\theta_0$ of unknown parameters is an interior point of the compact parameter space $\mathbf{\Theta}$ .

We note that the conditions in *Assumption III* are pararell to those in *Assumption I* and *Assumption II* in the linear case. In the simplest linear homoscedastic case when $\{\mathbf{z}_i\}$ are a sequence of i.i.d. random vectors, then $\mathbf{D}(M) = \mathbf{MD}$ and $\mathbf{C} = \sigma^2\mathbf{M}$ .
For the asymptotic bias of the modified MEL estimator, we have the following result and its proof is given in the Appendix.

**Theorem 5.1** : *Under Assumption III, the asymptotic bias of the standardized estimator* $\hat{\mathbf{e}}^*$ *as* $n \to +\infty$ *is given by*

$$AM_n(\hat{\mathbf{e}}^*) = \frac{1}{\sqrt{n}}\mathbf{Q}\ \mathbf{q}[L - 1 - \delta L] - \frac{1}{\sqrt{n}}[(\frac{1}{2})\mathbf{Q}\ \mathbf{D}(M)^{'}\mathbf{C}^{-1}tr(\mathbf{FQ})] , \tag{5.10}$$

*where* $\mathbf{F} = (f_{jk})$ *and*

$$f_{jk} = \text{plim}\frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i \frac{\partial^2 h_i}{\partial\theta_j\partial\theta_k} .$$

By using this result we immediately observe that the asymptotic bias of the MEL estimator and the GMM estimator have the corresponding terms as in the linear case and there is a common extra term due to the nonlinear relation in (2.1). Then it is possible to reduce the asymptotic bias of the MEL estimator by using the same modification of the MEL estimation method given that the second bias term in (5.10) is not very large. This is the case when the degrees of overidentification L is large.

The asymptotic expansions of the density functions of the modified MEL estimator and its asymptotic mean squared error have many terms in the general nonlinear case. However, we expect that the similar results as *Theorem 3.2* would hold in many situations under a set of additional regularity conditions.

## 6. Some Simulations

In order to investigate the finite sample properties of the MEL estimation method and our modified MEL estimation method, we have done a set of numerical simulations. For this purpose we set a simple linear structural model

$$y_{1i} = \beta_1 + \beta_2 y_{2i} + \gamma_1 z_{1i} + u_i , \tag{6.1}$$

where $y_{ki}$ $(k = 1, 2)$ are the endogenous variables, $z_{1i}$ is an exogenous variable, and $\beta_i$ $(i = 1, 2)$ and $\gamma_1$ are constant coefficients. We have investigated the situation when the endogenous variable $y_{2i}$ can be solved as

$$y_{2i} = \pi_{20} + \pi_{21} z_{1i} + \pi_{22}^{'} \mathbf{z}_{2i} + v_{2i} \ , \tag{6.2}$$

where $\pi_{2j}$ $(j = 0, 1)$ are scalor coefficients, $\pi_{22}$ is a $K_2 \times 1$ vector of coefficients, and $\mathbf{z}_{2i}$ is a vector of instrumental variables. Because $G_1 = 1$, the degrees of overidentification $L = K_2 - 1$ in the present case. We first set $L = 3, 5, 10, 20$; $n = 50, 100$, and then we simulated the normal random numbers for the exogenous variables $\mathbf{z}_i^{'} = (z_{1i}, \mathbf{z}_{2i}^{'})$ , the disturbance terms $u_i$ , and the endogenous variables $(y_{1i}, y_{2i})$ $(i = 1, \cdots, n)$ by using the Gaussian number generators. The number of replications in each case is 5,000.

We have summarized our simulation results in Table 6. For the sake of comparison we have given the mean squared error (MSE), the mean absolute error (MAE), and the probability of concentration (PC) for the GMM estimator, the MEL estimator, and the Modified MEL (MMEL) estimator. The PC has been calculated as

$$PC = P( \ |\sqrt{n} Q_{11}^{-1/2} (\hat{\beta} - \beta)| \leq c \ ) \ , \tag{6.3}$$

where $\hat{\beta}$ is the estimator of $\beta$ and $Q_{11}$ is the $(1, 1)$ element of $\mathbf{Q}$ which is the asymptotic variance-covariance matrix of the standardized estimator $\hat{\mathbf{e}}$ . We have used the normalization or standardization because it is often easy to make comparisons of alternative estimation methods and we set $c = 1$ for our numerical analysis.

Table 6.1 :Finite Sample Properties of Estimators
(L=5, n=50)

| | Bias | MSE | MAE | PC |
|---|---|---|---|---|
| GMM | -0.0567 | 0.8367 | 0.7479 | 0.6018 |
| MEL | 0.0180 | 1.0930 | 0.7700 | 0.6418 |
| MMEL | 0.0005 | 0.9107 | 0.7345 | 0.6422 |

Table 6.2 :Finite Sample Properties of Estimators
(L=10, n=50)

| | Bias | MSE | MAE | PC |
|---|---|---|---|---|
| GMM | -0.0671 | 0.6146 | 0.6526 | 0.5100 |
| MEL | 0.0093 | 0.6911 | 0.6373 | 0.5672 |
| MMEL | 0.0003 | 0.6361 | 0.6189 | 0.5734 |

Table 6.3 :Finite Sample Properties of Estimators
(L=3, n=100)

| | Bias | MSE | MAE | PC |
|---|---|---|---|---|
| GMM | -0.0207 | 0.5475 | 0.5905 | 0.6552 |
| MEL | 0.0115 | 0.6164 | 0.6068 | 0.6662 |
| MMEL | -0.0002 | 0.5684 | 0.5898 | 0.6724 |

Table 6.4 :Finite Sample Properties of Estimators
(L=10, n=100)

| | Bias | MSE | MAE | PC |
|---|---|---|---|---|
| GMM | -0.0337 | 0.4996 | 0.5769 | 0.5766 |
| MEL | 0.0044 | 0.4804 | 0.5451 | 0.6278 |
| MMEL | -0.0003 | 0.4614 | 0.5377 | 0.6310 |

Table 6.5 :Finite Sample Properties of Estimators
(L=15, n=100)

| | Bias | MSE | MAE | PC |
|---|---|---|---|---|
| GMM | -0.0358 | 0.3369 | 0.4759 | 0.5118 |
| MEL | 0.0037 | 0.3148 | 0.4392 | 0.5810 |
| MMEL | 0.0013 | 0.3046 | 0.4337 | 0.5892 |

There are several interesting findings on the small sample properties of the alternative estimation methods from the set of our experiments.

First, in terms of the MSE and MAE criteria the GMM estimator often performs well when L is small. In such cases the MEL estimator perform well in term of the probability of concentration. Thus we should be careful on the choice of loss functions when we want to compare alternative estimation methods in order to make a fair comparison of alternative estimation methods. ( See Anderson et. al. (1982) on the related issues. ) Second, the good performance of the GMM estimator becomes deteriolated quickly as the number of instruments L becomes large. This is the case regardless of the choice of criteria for comparison. On the other hand, the MEL method outperforms the GMM estimator in this situation by using any criteria.

Most importantly, in all cases the modifoed MEL (MMEL) estimator outperforms the MEL estimator in the sense of the Bias, the MSE, the MAE, and the PC. When L is large, the MMEL estimator is better than the MEL estimator, but the differences are not large.

On the whole these findings agree with our investigations on the finite sample properties based on the asymptotic expansions of the density functions of estimators up to $O(n^{-2})$ in the previous sections. It is consistent with the study on the small sample properties of the MEL estimator and the GMM estimator by Kunitomo and Matsushita (2002). They have given extensive tables of the distribution functions of two estimators in the linear homoscedastic case.

## 7. Conclusions

We have proposed a new estimation method of a structural equation by modifying the maximum empirical likelihood (MEL) method. Our estimator (MMEL) has not only the smaller asymptotic bias than the MEL estimator but also has the smaller asymptotic mean squared error when the sample size is large. Also we have given the numerical examples which suggest that the new estimation method has better small sample properties than the original MEL estimation method. Therefore we should use the modified MEL estimation method whenever we want to use the MEL estimation method for practical purposes. Also our modified estimator has good small sample properties when the degree of overidentification is large. It has been known that the GMM estimator has large bias in such situations and our method gives an alternative estimation method in this respect.

We should mention that the empirical likelihood (EL) method has been originally developped as the non-parametric testing and constructing confidence interval problems. (See Owen (2001) in the details.) In this respect there would be some useful ways to incorporate the modified MEL estimation method proposed in this paper for the related problems. It is currently under investigation.

## References

[1] Anderson, T.W., N. Kunitomo, and T. Sawa (1982), "Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator," *Econometrica*, Vol. 50-4, 1009-1027.

[2] Anderson, T.W., N. Kunitomo, and K. Morimune (1986), "Comparing Single Equation Estimators in a Simultaneous Equation System," *Econometric Theory*, Vol. 2, 1-32.

[3] Fujikoshi, Y., K. Morimune, N. Kunitomo, and M. Taniguchi (1982), "Asymptotic Expansions of the Distributions of the Estimates of Coefficients in a Simultaneous Equation System," *Journal of Econometrics*, Vol. 18, 2, 191-205.

[4] Godambe, V.P. (1960), "An Optimum Property of Regular Maximum Likelihood Equation," *Annals of Mathematical Statistics*, Vol. 31, 1208-1211.

[5] Hansen, L. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, Vol. 50, 1029-1054.

[6] Hayashi, F. (2000), *Econometrics*, Princeton University Press.

[7] Kitamura, Y. G. Tripathi, and H. Ahn (2001), "Empirical Likelihood-Based Inference in Conditional Moment Restriction Models," Unpublished Manuscript.

[8] Kunitomo, N. and Y. Matsushita (2002), "On Finite Sample Distributions of the Empirical Likelihood Estimator and the GMM Estimator," Unpublished Manuscript (in preparation).

[9] Morimune, K. (1985), *Estimation and Testing of Econometric Models*, Kyoritsu (in Japanese).

[10] Newey, W. K. and R. Smith (2001), "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimator," Unpublished Manuscript.

[11] Owen, A. B. (1990), "Empirical Likelihood Ratio Confidence Regions," *The Annals of Statistics*, Vol. 22, 300-325.

[12] Owen, A. B. (2001), *Empirical Likelihood*, Chapman and Hall.

[13] Qin, J. and Lawless, J. (1994), "Empirical Likelihood and General Estimating Equations," *The Annals of Statistics*, Vol. 22, 300-325.

## Mathematical Appendix

In this appendix, we give the mathematical details of derivations omitted in the previous sections. There are two cases depending on whether the sequence of exogenous variables $\{\mathbf{z}_i\}$ are random vectors or non-random (i.e. deterministic) vectors. In order to simplify our proofs and to avoid tedious derivations, we shall discuss the non-random case under the additional conditon $\mathbf{M}_n = \mathbf{M} + O(\frac{1}{n})$ . Nonetheless, other cases can be handled in the similar ways as we shall discuss in this Appendix.

### Appendix A : Derivations of Asymptotic Expansions

**[A-1]** : First we apply the similar arguments used in Owen (1990) and Qin and Lawless (1994) on the probability limits and the consistency of the MEL estimator. Then we have $n\hat{p}_i \xrightarrow{p} 1$, $\hat{\theta}_{EL} \xrightarrow{p} \theta_0$, ($\theta_0$ is the true value of $\theta$) and $\sqrt{n}\hat{\lambda}$ converges to a random vector as $n \to \infty$ .

In the linear case we substitute (2.1) into (2.13) and we have the corresponding representation of the standardized estimator $\hat{\mathbf{e}}$ as

$$\text{(A.1)} \qquad [\sum_{i=1}^{n} \hat{p}_i \begin{pmatrix} \mathbf{y}_{2i} \\ \mathbf{z}_{1i} \end{pmatrix} \mathbf{z}_i^{'}][\sum_{i=1}^{n} \hat{p}_i u_i(\hat{\theta})^2 \mathbf{z}_i \mathbf{z}_i^{'}]^{-1}[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{z}_i u_i]$$

$$= [\sum_{i=1}^{n} \hat{p}_i \begin{pmatrix} \mathbf{y}_{2i} \\ \mathbf{z}_{1i} \end{pmatrix} \mathbf{z}_i^{'}][\sum_{i=1}^{n} \hat{p}_i u_i(\hat{\theta})^2 \mathbf{z}_i \mathbf{z}_i^{'}]^{-1}[\frac{1}{n} \sum_{i=1}^{n} \mathbf{z}_i (\mathbf{y}_{2i}^{'}, \mathbf{z}_{1i}^{'})]\hat{\mathbf{e}} ,$$

where we use the notation $\hat{\theta}$ for $\hat{\theta}_{EL}$ without any subscript whenever we do not have any confusion. As $n \to \infty$, we write the first order term of $\hat{\mathbf{e}}$ as $\tilde{\mathbf{e}}_0$, which is given by

$$\text{(A.2)} \quad \tilde{\mathbf{e}}_0 = [\mathbf{D}^{'}(\text{plim}\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^{'})(\text{plim}\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^{'}u_i^2(\hat{\theta}))^{-1}(\text{plim}\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^{'})\mathbf{D}]^{-1}$$

$$\times[\mathbf{D}^{'}(\text{plim}\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^{'})(\text{plim}\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^{'}u_i^2(\hat{\theta}))^{-1}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_iu_i) .$$

The probability limits and the random variable on the right hand side of (A.1) have been defined properly because the matrices $\mathbf{M}$ and $\mathbf{C}$ are non-singular and $\mathbf{D}$ is of full rank by our assumptions. By using the central limit theorem (CLT) to the last term, we have the weak convergence

$$\text{(A.3)} \qquad \tilde{\mathbf{e}}_0 \xrightarrow{d} N_p(\mathbf{0}, \mathbf{Q}) ,$$

where a $p \times p$ matrix $\mathbf{Q}$ has been defined by (3.3) and $\xrightarrow{d}$ means the convergence of distribution as $n \to \infty$ . Also as $n \to \infty$ we notice that

$$\text{(A.4)} \qquad \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_iu_i(\hat{\theta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_iu_i + \frac{1}{n}[-\sum_{i=1}^{n}\mathbf{z}_i(\mathbf{y}_{2i}^{'}, \mathbf{z}_{1i}^{'})\hat{\mathbf{e}}]$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_iu_i - (\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i)\mathbf{D}\hat{\mathbf{e}} + O_p(\frac{1}{\sqrt{n}}) .$$

Then by utilizing the representation of (2.11) for $\hat{\lambda}$ , we have

$$\sqrt{n}\lambda - \lambda_0 \xrightarrow{p} 0 \ , \tag{A.5}$$

where

$$\lambda_0 = \mathbf{C}_n^{-1/2}[\mathbf{I}_K - \mathbf{C}_n^{-1/2}\mathbf{M}_n\mathbf{D}(\mathbf{D}^{'}\mathbf{M}_n\mathbf{C}_n^{-1}\mathbf{M}_n\mathbf{D})^{-1}\mathbf{D}^{'}\mathbf{M}_n\mathbf{C}_n^{-1/2}][\mathbf{C}_n^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i u_i] \ ,$$

and we have denoted $K \times K$ matrices $\mathbf{M}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^{'}$ and $\mathbf{C}_n = \frac{1}{n}\sum_{i=1}^{n}u_i^2\mathbf{z}_i\mathbf{z}_i^{'}$ . The random variable

$$\mathbf{B}_n = \mathbf{C}^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i u_i \tag{A.6}$$

converges to $N_K(\mathbf{0}, \mathbf{I}_K)$ by CLT, and $\mathbf{M}_n \xrightarrow{p} \mathbf{M}$, $\mathbf{C}_n \xrightarrow{p} \mathbf{C}$ as $n \longrightarrow +\infty$ under *Assumption I*, where $\mathbf{C}$ is defined by (4.2). Hence we also have the convergence

$$\mathbf{C}_n^{1/2}\sqrt{n}\hat{\lambda} \xrightarrow{d} N_K(\mathbf{0}, \bar{\mathbf{P}}_E) \ , \tag{A.7}$$

and the projection matrix on $\mathbf{E}$ is defined by

$$\bar{\mathbf{P}}_E = \mathbf{I}_K - \mathbf{E}(\mathbf{E}^{'}\mathbf{E})^{-1}\mathbf{E}^{'} \ , \tag{A.8}$$

which is constructed by a $K \times p$ matrix $\mathbf{E} = \mathbf{C}^{-1/2}\mathbf{M}\mathbf{D}$ and $\mathbf{E}_n = \mathbf{C}_n^{-1/2}\mathbf{M}_n\mathbf{D} \xrightarrow{p} \mathbf{E}$ as $n \longrightarrow +\infty$ .

**[A-2]**: The method we shall use to derive the asymptotic expansion of the density function of the standardized estimator $\hat{\mathbf{e}}$ is similar to the one used in Fujikoshi et. al. (1982) and Anderson et. al. (1986). The validity of the asymptotic expansions can be given by lengthy arguments which are similar to Appendix C of Fujikoshi et. at. (1982). We first derive the asymptotic expansion of the density function of the standardized estimator when the disturbance terms are normally distributed. Then we shall consider the same problem for more general disturbances.

By expanding (3.2) and (2.13) with respect to $\tilde{\mathbf{e}}_0$, formally we can write

$$\hat{\mathbf{e}} = \tilde{\mathbf{e}}_0 + [\mathbf{e}_0 - \tilde{\mathbf{e}}_0] + \frac{1}{\sqrt{n}}\mathbf{e}_1 + \frac{1}{n}\mathbf{e}_2 + o_p(\frac{1}{n}) \ , \tag{A.9}$$

and

$$\sqrt{n}\hat{\lambda} = \lambda_0 + \frac{1}{\sqrt{n}}\lambda_1 + \frac{1}{n}\lambda_2 + o_p(\frac{1}{n}) \ , \tag{A.10}$$

where we denote

$$\mathbf{e}_0 = [\mathbf{D}^{'}\mathbf{M}_n\mathbf{C}_n^{-1}\mathbf{M}_n\mathbf{D}]^{-1}[\mathbf{D}^{'}\mathbf{M}_n\mathbf{C}_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i u_i] \ . \tag{A.11}$$

By substituting these expansions into (2.9), we can also expand the estimated probability function as

$$n\ \hat{p}_i = 1 + \frac{1}{\sqrt{n}}p_i^{(1)} + \frac{1}{n}p_i^{(2)} + o_p(\frac{1}{n}) \ , \tag{A.12}$$

where

$$p_i^{(1)} = -\lambda_0^{'}\mathbf{z}_i[u_i - \frac{1}{\sqrt{n}}\mathbf{z}_i^{'}\mathbf{D}\mathbf{e}_0]$$

$$p_i^{(2)} = -\lambda_1^{'}\mathbf{z}_i[u_i - \frac{1}{\sqrt{n}}(\mathbf{y}_{2i}^{'},\mathbf{z}_{1i}^{'})\mathbf{e}_0]$$

$$+\lambda_0^{'}\mathbf{z}_i[\frac{1}{\sqrt{n}}(\mathbf{y}_{2i}^{'},\mathbf{z}_{1i}^{'})\mathbf{e}_1 + (\mathbf{v}_{2i}^{'},\mathbf{0}^{'})\mathbf{e}_0] + (\lambda_0^{'}\mathbf{z}_i)^2[u_i - \frac{1}{\sqrt{n}}(\mathbf{y}_{2i}^{'},\mathbf{z}_{1i}^{'})]^2 .$$

By using the relation

(A.13) $$(\mathbf{y}_{2i}^{'},\mathbf{z}_{1i}^{'}) = \mathbf{z}_i^{'}\mathbf{D} + \mathbf{w}_i^{'} + \mathbf{q}^{'}u_i ,$$

we shall expand

(A.14)
$$\hat{\mathbf{C}}_n = \sum_{i=1}^{n}\hat{p}_i u_i^2(\hat{\theta})\mathbf{z}_i\mathbf{z}_i^{'}$$
$$= \mathbf{C}_n + \frac{1}{\sqrt{n}}\hat{\mathbf{C}}_n^{(1)} + \frac{1}{n}\hat{\mathbf{C}}_n^{(2)} + o_p(\frac{1}{n})$$
$$= \mathbf{C}_n + \frac{1}{\sqrt{n}}\mathbf{C}_n^{(1)} + \frac{1}{n}\mathbf{C}_n^{(2)} + o_p(\frac{1}{n}) ,$$

and

(A.15) $$\sum_{i=1}^{n}\hat{p}_i\begin{pmatrix}\mathbf{y}_{2i}\\ \mathbf{z}_{1i}\end{pmatrix}\mathbf{z}_i^{'} = [\mathbf{D}^{'}\mathbf{M}_n] + \frac{1}{\sqrt{n}}\mathbf{E}_n^{(1)} + \frac{1}{n}\mathbf{E}_n^{(2)} + o_p(\frac{1}{n}) ,$$

where we denote the random matrices as

(A.16) $$\hat{\mathbf{C}}_n^{(j)} = \frac{1}{n}\sum_{i=1}^{n}p_i^{(j)}\mathbf{z}_i\mathbf{z}_i^{'}u_i^2(\hat{\theta}) \quad (j=1,2) ,$$

(A.17) $$\mathbf{E}_n^{(1)} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix}\mathbf{v}_{2i}\\ \mathbf{0}\end{pmatrix}\mathbf{z}_i^{'} + \mathbf{D}^{'}\frac{1}{n}\sum_{i=1}^{n}p_i^{(1)}\mathbf{z}_i\mathbf{z}_i^{'} + \frac{1}{n}\sum_{i=1}^{n}p_i^{(1)}\begin{pmatrix}\mathbf{v}_{2i}\\ \mathbf{0}\end{pmatrix}\mathbf{z}_i^{'} ,$$

(A.18) $$\mathbf{E}_n^{(2)} = \mathbf{D}^{'}\frac{1}{n}\sum_{i=1}^{n}p_i^{(2)}\mathbf{z}_i\mathbf{z}_i^{'} + \frac{1}{n}\sum_{i=1}^{n}p_i^{(2)}\begin{pmatrix}\mathbf{v}_{2i}\\ \mathbf{0}\end{pmatrix}\mathbf{z}_i^{'} .$$

By using (A.12)-(A.14), we notice that $\mathbf{C}_n^{(1)}$ can be rewritten as

$$\mathbf{C}_n^{(1)}$$
$$= \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^{'}[p_i^{(1)}u_i^2 + 2u_i(-1)(\mathbf{y}_{2i}^{'},\mathbf{z}_{1i}^{'})\mathbf{e}_0]$$
$$= \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^{'}[-2(\mathbf{q}^{'}\mathbf{e}_0)u_i^2 - 2u_i(\mathbf{w}_i^{'} + \mathbf{z}_i^{'}\mathbf{D})\mathbf{e}_0 - \lambda^{'}\mathbf{z}_i u_i^3] + O_p(\frac{1}{\sqrt{n}})$$
$$= (-2)(\mathbf{q}^{'}\mathbf{e}_0)\mathbf{C}_n + [-2\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^{'}u_i(\mathbf{w}_i^{'} + \mathbf{z}_i^{'}\mathbf{D})\mathbf{e}_0 - \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^{'}u_i^3\lambda_0] + O_p(\frac{1}{\sqrt{n}}) .$$

By substituting the above expressions into (A.1) for $\hat{\mathbf{e}}$ , $\hat{\lambda}$, and $\hat{p}_i$ $(i = 1, \cdots, n)$, we can determine each terms of the stochastic expansions in the recursive way as

$$\text{(A.19)} \qquad \mathbf{e}_1 = -\mathbf{Q}_n \mathbf{D}^{'} \mathbf{M}_n \mathbf{C}_n^{-1} [\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{z}_i (\mathbf{v}_{2i}^{'}, \mathbf{0})] \mathbf{e}_0$$

$$+\mathbf{Q}_n [\mathbf{A}_1] [\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{z}_i u_i - \mathbf{M}_n \mathbf{D}^{'} \mathbf{e}_0] ,$$

$$\text{(A.20)} \qquad \mathbf{e}_2 = \mathbf{Q}_n [\mathbf{A}_2] [\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{z}_i u_i - \mathbf{M}_n \mathbf{D}^{'} \mathbf{e}_0]$$

$$-\mathbf{Q}_n [\mathbf{A}_1] [\mathbf{M}_n \mathbf{D} \mathbf{e}_1 + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{z}_i (\mathbf{v}_{2i}^{'}, \mathbf{0}^{'}) \mathbf{e}_0]$$

$$-\mathbf{Q}_n \mathbf{D}^{'} \mathbf{M}_n \mathbf{C}_n^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathbf{v}_{2i}^{'}, \mathbf{0}^{'}) \mathbf{e}_1 ,$$

where we have used the corresponding notations

$$\mathbf{Q}_n^{-1} = \mathbf{D}^{'} \mathbf{M}_n \mathbf{C}_n^{-1} \mathbf{M}_n \mathbf{D} ,$$

$$\mathbf{A}_1 = -\mathbf{D}^{'} \mathbf{M}_n \mathbf{C}_n^{-1} \mathbf{C}_n^{(1)} \mathbf{C}_n^{-1} + \mathbf{E}_n^{(1)} \mathbf{C}_n^{-1} ,$$

$$\mathbf{A}_2 = \mathbf{D}^{'} \mathbf{M}_n [-\mathbf{C}_n^{-1} \mathbf{C}_n^{(2)} \mathbf{C}_n^{-1} + \mathbf{C}_n^{-1} \mathbf{C}_n^{(1)} \mathbf{C}_n^{-1} \mathbf{C}_n^{(1)} \mathbf{C}_n^{-1}] - \mathbf{E}_n^{(1)} \mathbf{C}_n^{-1} \mathbf{C}_n^{(1)} \mathbf{C}_n^{-1}$$
$$+ \mathbf{E}_n^{(2)} \mathbf{C}_n^{-1} .$$

We also define two random vectors $\tilde{\mathbf{e}}_1$ and $\tilde{\mathbf{e}}_2$ by substituting $\mathbf{C}$ for $\mathbf{C}_n$ in (A.19) and (A.20), respectively. Our next strategy is to derive the asymptotic expansion of the density function of the random vector

$$\text{(A.21)} \qquad \tilde{\mathbf{e}} = \tilde{\mathbf{e}}_0 + \frac{1}{\sqrt{n}} \tilde{\mathbf{e}}_1 + \frac{1}{n} \tilde{\mathbf{e}}_2 + o_p(\frac{1}{n})$$

and then we shall evaluate the effects of the differences in the form

$$\text{(A.22)} \qquad [\hat{\mathbf{e}} - \tilde{\mathbf{e}}] = [\mathbf{e}_0 - \tilde{\mathbf{e}}_0] + \frac{1}{\sqrt{n}} [\mathbf{e}_1 - \tilde{\mathbf{e}}_1] + \frac{1}{n} [\mathbf{e}_2 - \tilde{\mathbf{e}}_2] + o_p(\frac{1}{n}) .$$

Although there are many terms in the above stochastic expansion of $\tilde{\mathbf{e}}$ , many of them can be ignored for order calculations. This consideration leads to some simplifications of its stochastic expansions.
Because

$$(\frac{1}{n}) \sum_{i=1}^{n} \hat{p}_i^{(1)} \mathbf{z}_i \mathbf{z}_i^{'} = (-1) \frac{1}{n} \sum_{i=1}^{n} \mathbf{z}_i \mathbf{z}_i^{'} \lambda^{'} \mathbf{z}_i u_i = O_p(\frac{1}{\sqrt{n}}) ,$$

and we have the relation that

$$\text{(A.23)} \qquad \mathbf{E}_n^{(1)} \cong \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{w}_i \mathbf{z}_i^{'} + \mathbf{q} [\frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i \mathbf{z}_i^{'} - \delta \lambda_0^{'} \mathbf{C}_n] + O_p(\frac{1}{\sqrt{n}}) ,$$

we have the effect of our modification of the MEL estmation only in the last term for the asymptotic bias. By using the relation

$$-\mathbf{Q}\mathbf{D}^{'}\mathbf{M}\mathbf{C}^{-1}[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i u_i](\mathbf{q}^{'}\mathbf{e}_0) \cong -\tilde{\mathbf{e}}_0(\mathbf{q}^{'}\tilde{\mathbf{e}}_0) + O_p(\frac{1}{\sqrt{n}}) ,$$

we can calculate the conditional expectation of each terms in $\mathbf{e}_1$ given $\mathbf{x}$, which is the limiting random vector of $\mathbf{e}_0$, under the assumption of normal disturbances. Since the random vectors $\hat{\mathbf{e}}$ and $\sqrt{n}\lambda$ converge to the normal random variables, the effects of nonnormality occur in higher orders which should be evaluated later. The conditional asymptotic bias is given as

$$\text{(A.24)} \qquad E[\tilde{\mathbf{e}}_1|\mathbf{x}] = -(\mathbf{q}^{'}\mathbf{x})\mathbf{x} + \mathbf{Q}\,\mathbf{q}[L-\delta L] + O_p(\frac{1}{\sqrt{n}}) ,$$

where we have used the relation on the projection operators such as $\mathbf{P}_E\bar{\mathbf{P}}_E = \mathbf{O}$ ( $\mathbf{P}_E = \mathbf{E}(\mathbf{E}^{'}\mathbf{E})^{-1}\mathbf{E}^{'}$ is a $K \times K$ projection matrix ) and

$$\text{(A.25)} \qquad \lambda_0^{'}\mathbf{C}_n^{1/2}\bar{\mathbf{P}}_E\mathbf{B}_n = \mathbf{B}_n^{'}\bar{\mathbf{P}}_E\mathbf{B}_n \stackrel{d}{\longrightarrow} \chi^2(L)$$

as $n \to \infty$ . We notice that $\mathbf{e}_0$ and $\mathbf{B}_n^{'}\bar{\mathbf{P}}_E\mathbf{B}_n$ are independent under the assumption of normal disturbances. ( In fact they are asymptotically independent in the general case. ) We also have

$$\mathbf{e}_0\,\mathbf{e}_0^{'} = \mathbf{Q}_n\mathbf{D}^{'}\mathbf{M}_n\mathbf{C}_n^{-1}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}u_i\mathbf{z}_i\frac{1}{\sqrt{n}}\sum_{i=1}^{n}u_i\mathbf{z}_i^{'})\mathbf{C}_n^{-1}\mathbf{M}_n\mathbf{D}\mathbf{Q}_n \stackrel{p}{\longrightarrow} \mathbf{Q}$$

as $n \to \infty$ and $\mathbf{Q}$ is non-singular. Then we have the results on the asymptotic bias for the class of modified MEL estimators in *Theorem 3.1* and *Theorem 4.1.* It is easily seen that the formula for the asymptotic bias does not depend on the distribution of disturbances.

**[A-3]**: Since there are many terms in the expression of $\mathbf{e}_2$ and $\tilde{\mathbf{e}}_2$, at first it looks formidable to evaluate the stochastic orders of these terms. Fortunately, we can show that we can ignore many terms because their stochastic orders do not affect the asymptotic bias and the asymptotic mean squared error.
We use the notation $\mathbf{x}$ be the random limit vector of $\tilde{\mathbf{e}}_0$ as $n \to \infty$ . After straightforward but quite tedious calculations as illustrated in *Appendix B*, we have the conditional expectation of $\tilde{\mathbf{e}}_2$ given $\mathbf{x}$ under the normal disturbances as

$$\text{(A.26)} \qquad \begin{aligned} E[\tilde{\mathbf{e}}_2|\mathbf{x}] &= \mathbf{x}^{'}\mathbf{C}_1^{*}\mathbf{x}\cdot\mathbf{x} + \mathbf{Q}\mathbf{Q}^{*}\mathbf{Q}\mathbf{C}_2^{*}\mathbf{x} - (1-\delta)\mathbf{Q}\mathbf{C}_2^{*}\mathbf{x}\,tr[\bar{\mathbf{P}}_E\mathbf{M}^{*}] \\ &\quad -(1-\delta)[2L\mathbf{Q}\mathbf{C}_1^{*}\mathbf{x} + \mathbf{x}\,tr\{\mathbf{C}_1^{*}\mathbf{Q}]\} + O_p(\frac{1}{\sqrt{n}}) , \end{aligned}$$

where we have used the matrix notations $\mathbf{C}_1^{*} = \mathbf{q}\mathbf{q}^{'}$, $\mathbf{C}_2^{*} = E[\mathbf{w}_i\mathbf{w}_i^{'}]/\sigma^2$, $\mathbf{M}^{*} = \mathbf{C}^{-1/2}\mathbf{M}\mathbf{C}^{-1/2}$, and $\mathbf{Q}^{*} = \mathbf{D}^{'}\mathbf{M}\mathbf{C}^{-1}\mathbf{M}\mathbf{C}^{-1}\mathbf{M}\mathbf{D}$ . The trace notation has been used as $tr(\mathbf{A}) = \sum_i a_{ii}$ for any conformable matrix $\mathbf{A} = (a_{ij})$ .
Also by utilizing the expression of $\tilde{\mathbf{e}}_1$ and by using lengthy calculations, we can derive the conditional expectation of $\tilde{\mathbf{e}}_1\tilde{\mathbf{e}}_1^{'}$ as

$$\text{(A.27)} \qquad E[\tilde{\mathbf{e}}_1\,\tilde{\mathbf{e}}_1^{'}|\mathbf{x}]$$

$$= \mathbf{x}'\mathbf{C}_1^*\mathbf{x}\,\mathbf{x}\mathbf{x}' + \mathbf{Q}\mathbf{Q}^*\mathbf{Q}\mathbf{x}'\mathbf{C}_2^*\mathbf{x} + \mathbf{Q}\mathbf{C}_2^*\mathbf{Q}\; tr[\bar{\mathbf{P}}_E\mathbf{M}^*]$$
$$+(1-\delta)^2L(L+2)\mathbf{Q}\mathbf{C}_1^*\mathbf{Q} - (1-\delta)L[\mathbf{Q}\mathbf{C}_1^*\mathbf{x}\mathbf{x} + \mathbf{x}\mathbf{x}'\mathbf{C}_1^*\mathbf{Q}] + O_p(\frac{1}{\sqrt{n}}) \ .$$

Next we consider the characteristic function of the standardized estimator $\tilde{\mathbf{e}}$ which can be written as

$$\text{(A.28)} \qquad C(t) = E[\exp(i\mathbf{t}'\mathbf{x})]$$
$$+\frac{1}{\sqrt{n}}E[i\mathbf{t}'E(\tilde{\mathbf{e}}_1|\mathbf{x})\exp(i\mathbf{t}'\mathbf{x})]$$
$$+\frac{1}{2n}E\{2i\mathbf{t}'E(\tilde{\mathbf{e}}_2|\mathbf{x})\exp(i\mathbf{t}'\mathbf{x}) + i^2\mathbf{t}'E(\tilde{\mathbf{e}}_1\tilde{\mathbf{e}}_1'|\mathbf{x})\mathbf{t}\exp(i\mathbf{t}'\mathbf{x})\} + O(\frac{1}{n\sqrt{n}}) \ ,$$

where $\mathbf{t} = (t_i)$ is a $p \times 1$ vector of real variables and $i^2 = -1$ . By using the Fourier Inversion Formulae developed by *Appendix B* of Fujikoshi et. al. (1982), we can invert the characteristic function in (A.28). The intermediate computations are quite tedious but straightforward and they are similar to those reported in Appendix of Anderson et. al. (1986). By arranging each terms in the Fourier Inversions we finally have the next result.

**Theorem A.1** : Suppose that the conditions in *Assumption II* hold and the disturbances are normally distributed. Then the asymptotic expansion of the joint density function of $\tilde{\mathbf{e}}^*$ for the class of modified (MEL) estimators as $n \to \infty$ is given by

$$f(\xi) = \phi_Q(\xi)\Big(1 + \frac{1}{\sqrt{n}}(\mathbf{q}'\xi)[p+1+(1-\delta)L - \xi'\mathbf{Q}^{-1}\xi]$$
$$+\frac{1}{2n}\Big(\xi'\mathbf{C}_1^*\xi\{[p+1+(1-\delta)L-\xi'\mathbf{Q}^{-1}\xi]^2 + p + 1 - 3\xi'\mathbf{Q}^{-1}\xi + 2(1-\delta)^2L\}$$
$$+tr(\mathbf{C}_1^*\mathbf{Q})[(1-\delta)L][2-(1-\delta)(L+2)]$$
$$+\xi'\mathbf{C}_2^*\xi\{tr[\bar{\mathbf{P}}_E\mathbf{M}^*][1-2(1-\delta)] - p - 4 + \xi'\mathbf{Q}^{-1}\xi\}$$
$$+tr(\mathbf{C}_2^*\mathbf{Q})\{tr[\bar{\mathbf{P}}_E\mathbf{M}^*][2(1-\delta)-1]\} + 2\xi'\mathbf{Q}^*\mathbf{Q}\mathbf{C}_2^*\xi\Big)\Big)$$
$$+o(\frac{1}{n}) \ ,$$

where $\xi$ is a $p \times 1$ ($p = G_1 + K_1$) vector and $\phi_Q(\xi)$ is the multivariate normal density function with mean $\mathbf{0}$ and the variance-covariance matrix $\mathbf{Q}$ .

By using the asymptotic expansion of the density function, we can evaluate the asymptotic mean squared errors of the modified MEL estimator. We summarize the resulting formulae.

**Theorem A.2** : Suppose that the conditions in *Assumption II* hold and the disturbances are normally distributed. Then the asymptotic mean squared errors of $\tilde{\mathbf{e}}^*$ for the modified (MEL) estimators based on the asymptotic expansion of the density function as $n \to \infty$ up to $O(n^{-1})$ is given by

$$AM_n(\tilde{\mathbf{e}}\,\tilde{\mathbf{e}}')$$

$$= \mathbf{Q} + \frac{1}{n}\Big\{\mathbf{Q}\mathbf{C}_1^*\mathbf{Q}[6-6(1-\delta)L+(1-\delta)^2L(L+2)]$$
$$+\mathbf{Q}tr(\mathbf{C}_1^*\mathbf{Q})[3-2(1-\delta)L]+\mathbf{Q}\mathbf{Q}^*\mathbf{Q}tr(\mathbf{C}_2^*\mathbf{Q})+\mathbf{Q}\mathbf{C}_2^*\mathbf{Q}\ tr[\bar{\mathbf{P}}_E\mathbf{M}^*][1-2(1-\delta)]$$
$$+2\mathbf{Q}\mathbf{Q}^*\mathbf{Q}\mathbf{C}_2^*\mathbf{Q}\Big\}\ .$$

**Remark A.1**: When the disturbance terms are homoscedastic as in *Assumption I* and normaly distributed random vectors, we can show the relations

$$\text{(A.29)} \qquad \mathbf{Q}^* = \sigma^{-2}\mathbf{Q}^{-1}$$

$$\text{(A.30)} \qquad tr[\bar{\mathbf{P}}_E\mathbf{M}^*] = \sigma^{-2}L\ ,$$

because $tr[\mathbf{I}_K-\mathbf{E}(\mathbf{E}^{'}\mathbf{E})^{-1}\mathbf{E}^{'}]=L\ (=K-p)$ and $\sigma^2=E(u_i^2)\ (i=1,\cdots,n)$ . Then the formulae given in *Theorem A.1* and *Theorem A.2* are identical to the corresponding formulae for the limited information maximum likelihood (LIML) estimator when $\delta=1$ except some differences in the parametrizations. Also the the formulae given in *Theorem A.1* and *Theorem A.2* are identical to the corresponding formulae for the two satge least squares (TSLS) estimator when $\delta=0$ (the GMM case). They have been already derived by Fujikoshi et. al. (1982) and extensively used by Anderson et. al. (1986) for the comparison of alternative single equation parametric estimation methods. Thus we have extended their results to the non-parametric or the semi-parametric single equation estimation methods in this Appendix.

**[A-4] Proof of Theorem 3.2 and Theorem 4.2** :
Our method of proof consists of two steps and also we use one Lemma.
[*i*] We take two standardized estimators $\hat{\mathbf{e}}_{EL}$ and $\tilde{\mathbf{e}}^*$ with an arbitrary $\delta$ $(0\le\delta\le1)$ and apply *Theorem A.2*. We first compare the MSE of $\tilde{\mathbf{e}}_{EL}$ and $\tilde{\mathbf{e}}^*$, which are the corresponding main parts of $\hat{\mathbf{e}}_{EL}$ and $\tilde{\mathbf{e}}^*$ . Then we have

$$\text{(A.31)} \qquad n[AM_n(\tilde{\mathbf{e}}^*\,\tilde{\mathbf{e}}^{*'})-AM_n(\tilde{\mathbf{e}}_{EL}\,\tilde{\mathbf{e}}_{EL}^{'})]$$
$$= \mathbf{Q}\mathbf{C}_1^*\mathbf{Q}[(1-\delta)^2L(L+2)-6(1-\delta)L]+\mathbf{Q}tr(\mathbf{C}_1^*\mathbf{Q})[-2(1-\delta)L]$$
$$+\mathbf{Q}\mathbf{C}_2^*\mathbf{Q}\ tr[\bar{\mathbf{P}}_E\mathbf{M}^*][-2(1-\delta)]\ .$$

We notice that the fourth order term $\mathbf{x}^{'}\mathbf{C}_1\mathbf{x}\cdot\mathbf{x}\mathbf{x}^{'}$ in the conditional stochastic expansion of $\tilde{\mathbf{e}}_{EL}$ and $\tilde{\mathbf{e}}^*$ has disappeared in the above expression. Hence the differences of the asymptotic MSE (AMSE) of two main parts of estimators do not depend on the fourth order moments of $\{u_i\}$ . That is, it does not depend on the normality of disturbance terms.
[*ii*] Next we consider the effects of the differences between $\hat{\mathbf{e}}$ and $\tilde{\mathbf{e}}$ on the AMSE's for the modified MEL estimators. By ignoring some higher order terms $o(n^{-1})$, we can write

$$\text{(A.32)} \quad AM_n(\hat{\mathbf{e}}\,\hat{\mathbf{e}}^{'})-AM_n(\tilde{\mathbf{e}}\,\tilde{\mathbf{e}}^{'})$$
$$= E\Big[(\tilde{\mathbf{e}}_0+\frac{1}{\sqrt{n}}\tilde{\mathbf{e}}_1)^{'})(\mathbf{e}_0-\tilde{\mathbf{e}}_0)^{'}+(\mathbf{e}_0-\tilde{\mathbf{e}}_0)^{'}(\tilde{\mathbf{e}}_0+\frac{1}{\sqrt{n}}\tilde{\mathbf{e}}_1)^{'}+(\mathbf{e}_0-\tilde{\mathbf{e}}_0)(\mathbf{e}_0-\tilde{\mathbf{e}}_0)^{'}$$

$$+\frac{1}{\sqrt{n}}\tilde{\mathbf{e}}_0(\mathbf{e}_1-\tilde{\mathbf{e}}_1)+\frac{1}{\sqrt{n}}(\mathbf{e}_1-\tilde{\mathbf{e}}_1)\tilde{\mathbf{e}}_0)^{'}\Big] \ .$$

Then we need to evaluate each terms up to $O_p(\frac{1}{n})$ . Because of *Assumption I* or *Assumption II* on the third order moments, it is straightforward to show that

$$E[(\mathbf{e}_1-\tilde{\mathbf{e}}_1)\tilde{\mathbf{e}}_0^{'}]=o(\frac{1}{\sqrt{n}}) \ . \tag{A.33}$$

Also

$$\begin{aligned}(A.34)\quad & E[(\mathbf{e}_0-\tilde{\mathbf{e}}_0)\tilde{\mathbf{e}}_0^{'}]\\ = \ & E[(\mathbf{Q}_n\mathbf{D}^{'}\mathbf{M}_n\mathbf{C}_n^{-1}-\mathbf{Q}\mathbf{D}^{'}\mathbf{M}\mathbf{C}^{-1})(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_iu_i)(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_iu_i)^{'}\mathbf{C}_n^{-1}\mathbf{M}\mathbf{D}\mathbf{Q}] \ .\end{aligned}$$

We notice that under our assumptions we can use the order relations $\mathbf{C}_n-\mathbf{C}=O_p(1/\sqrt{n})$ and also

$$[(\frac{1}{\sqrt{n}})\sum_{i=1}^{n}\mathbf{z}_iu_i][(\frac{1}{\sqrt{n}})\sum_{i=1}^{n}\mathbf{z}_iu_i]^{'}-\mathbf{C}=O_p(\frac{1}{\sqrt{n}}) \ ,$$

this term is asymptotically the same as

$$\begin{aligned}(A.35)\quad & -\mathbf{Q}\mathbf{D}^{'}\mathbf{M}\mathbf{C}^{-1}E[\mathbf{C}_n\mathbf{A}(\mathbf{C}_n-\mathbf{C})\mathbf{C}_n^{-1}\mathbf{M}_n\mathbf{D}\mathbf{Q}_n]\\ = \ & (-\frac{\kappa_4}{n})\mathbf{Q}\mathbf{D}^{'}\mathbf{M}\mathbf{C}^{-1}[\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^{'}\ tr(\mathbf{A}\mathbf{z}_i\mathbf{z}_i^{'})]\mathbf{C}^{-1}\mathbf{M}\mathbf{D}\mathbf{Q}+O(\frac{1}{n\sqrt{n}}) \ ,\end{aligned}$$

where $\kappa_4$ is the fourth cumulant given by $\kappa_4=E(u_i^4)-3(E(u_i^2))^2$ and $\mathbf{A}=\mathbf{C}^{-1/2}\bar{\mathbf{P}}_E\mathbf{C}^{-1/2}$ . Similarly we can evaluate

$$\begin{aligned}(A.36)\quad & E[(\mathbf{e}_0-\tilde{\mathbf{e}}_0)\ (\mathbf{e}_0-\tilde{\mathbf{e}}_0)^{'}]\\ = \ & \mathbf{Q}\mathbf{D}^{'}\mathbf{M}\mathbf{C}^{-1}E[(\mathbf{C}_n-\mathbf{C})\mathbf{A}\mathbf{C}_n\mathbf{A}(\mathbf{C}_n-\mathbf{C})]\mathbf{C}^{-1}\mathbf{M}\mathbf{D}\mathbf{Q}\\ = \ & \frac{\kappa_4}{n}\mathbf{Q}\mathbf{D}^{'}\mathbf{M}\mathbf{C}^{-1}[\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^{'}\ tr(\mathbf{A}\mathbf{z}_i\mathbf{z}_i^{'})]\mathbf{C}^{-1}\mathbf{M}\mathbf{D}\mathbf{Q}+O(\frac{1}{n\sqrt{n}}) \ .\end{aligned}$$

By using the calculations illustrated in *Appendix B*, it is possible to show that other terms are asymptotically negligible. Hence we have shown that the above terms depend on the nonnormality, but they are common in the class of estimators we are comparing. Then by using the following lemma, which is **Lemma 1** of Anderson et. al. (1986), we can immediately obtain the second parts on the APC's of estimators in *Theorem 3.2* and *Theorem 4.2*. (*Q.E.D.*)

**Lemma A.3** : Let $\mathbf{S}$ be a star-shaped set in $\mathbf{R}^p$ . Then for any arbitrary $p\times p$ positive semidefinite matrix $\mathbf{C}^*$,

$$\int_{\mathbf{S}}[\xi^{'}\mathbf{C}^*\xi-\ tr(\mathbf{C}^*\mathbf{Q})]\phi_{\mathbf{Q}}(\xi)d\xi\leq 0 \ . \tag{A.37}$$

Furthermore, for any arbitrary vector $\mathbf{b}$,

$$\int_{\mathbf{S}}(\mathbf{b}^{'}\xi)^2[p+2-\xi^{'}\mathbf{Q}\xi]\phi_{\mathbf{Q}}(\xi)d\xi\geq 0 \ , \tag{A.38}$$

where $\xi$ is a $p \times 1$ vector and $\phi_Q(\xi)$ is the multivariate normal density function with mean $\mathbf{0}$ and the variance-covariance matrix $\mathbf{Q}$ .

## Appendix B : Derivations of (A.26) and (A.27)

In this subsection we give some details of our calculations on $E(\mathbf{e}_2|\mathbf{x})$ we have omitted in *Appendix A*, where $\mathbf{x}$ is the limiting random vector of $\mathbf{e}_0$ . Since there are three terms in (A.20), we write $\mathbf{e}_2 = \mathbf{e}_2^{(1)} + \mathbf{e}_2^{(2)} + \mathbf{e}_2^{(3)}$ as their corresponding orders. All terms involving the differences between $\tilde{\mathbf{e}}_2$ and $\mathbf{e}_2$ are of order $O_p(n^{-1/2})$ , which can be ignored for the present purpose. First by using the stochastic expansion of $\mathbf{e}_2$, we notice that

$$
\begin{aligned}
\mathbf{e}_2^{(3)} &= -\mathbf{Q}_n\mathbf{D}^{'}\mathbf{M}_n\mathbf{C}_n^{-1}[(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbf{z}_i\mathbf{w}_i^{'}) + (\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbf{z}_iu_i)\mathbf{q}^{'}] \\
&\quad \times\{\mathbf{Q}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbf{w}_i\mathbf{z}_i^{'})\mathbf{C}^{-1}\bar{\mathbf{P}}_E\mathbf{B}_n - \mathbf{Q}\mathbf{D}^{'}\mathbf{M}\mathbf{C}^{-1}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{w}_i^{'})\mathbf{e}_0 - \mathbf{e}_0(\mathbf{q}^{'}\mathbf{e}_0) \\
&\qquad +(1-\delta)\mathbf{Q}\mathbf{q}\mathbf{B}_n^{'}\bar{\mathbf{P}}_E\mathbf{B}_n\} \\
&= -\mathbf{Q}_n\mathbf{D}^{'}\mathbf{M}_n\mathbf{C}_n^{-1}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{w}_i^{'})[-\mathbf{Q}_n\mathbf{D}^{'}\mathbf{M}_n\mathbf{C}_n^{-1}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{w}_i^{'}]\mathbf{e}_0] \\
&\quad -\mathbf{Q}_n\mathbf{D}^{'}\mathbf{M}_n\mathbf{C}_n^{-1}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_iu_i)\mathbf{q}^{'}[-\mathbf{e}_0(\mathbf{q}^{'}\mathbf{e}_0) + (1-\delta)\mathbf{Q}\mathbf{q}\mathbf{B}_n^{'}\bar{\mathbf{P}}_E\mathbf{B}_n + \mathbf{e}_2^{(3*)} \ ,
\end{aligned}
$$

where the last term $\mathbf{e}_2^{(3*)}$ denotes other terms in $\mathbf{e}_1^{(3)}$ which cab be ignored. Then by using the relations that $E(\mathbf{e}_2^{(3*)}|\mathbf{x}) = O(n^{-1/2})$ and $\mathbf{e}_0 = \mathbf{Q}_n\mathbf{D}^{'}\mathbf{M}_n\mathbf{C}_n^{-1}[(1/\sqrt{n})\sum_{i=1}^{n}\mathbf{z}_iu_i]$ , we have the expression

$$
\begin{aligned}
\text{(A.39)} \quad E[\mathbf{e}_2^{(3)}|\mathbf{x}] &= \mathbf{Q}\mathbf{D}^{'}\mathbf{M}\mathbf{C}^{-1}\mathbf{M}\mathbf{D}\mathbf{Q}\mathbf{C}_2^{*}\mathbf{x} + \mathbf{x}(\mathbf{q}^{'}\mathbf{x})^2 - (1-\delta)L\mathbf{q}^{'}\mathbf{Q}\mathbf{q}\mathbf{x} \\
&= \mathbf{Q}\mathbf{Q}^{*}\mathbf{Q}\mathbf{C}_2^{*}\mathbf{x} + \mathbf{x}^{'}\mathbf{C}_1^{*}\mathbf{x} - (1-\delta)L\ tr(\mathbf{C}_1^{*}\mathbf{Q}) + O_p(\frac{1}{\sqrt{n}}) \ .
\end{aligned}
$$

For the second term we rewrite

$$
\begin{aligned}
\mathbf{e}_2^{(2)} &= -\mathbf{Q}_n[\mathbf{A}_1]\{\mathbf{M}\mathbf{D}[-\mathbf{Q}_n\mathbf{D}^{'}\mathbf{M}_n\mathbf{C}_n^{-1}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i(\mathbf{v}_{2i}^{'},\mathbf{0}))\mathbf{e}_0 + \mathbf{Q}_n[\mathbf{A}_1]\mathbf{C}_n^{1/2}\bar{\mathbf{P}}_E\mathbf{B}_n] \\
&\quad +(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i(\mathbf{v}_{2i}^{'},\mathbf{0}))\mathbf{e}_0\} \\
&= -\mathbf{Q}_n[\mathbf{A}_1]\mathbf{M}\mathbf{D}\mathbf{Q}_n[\mathbf{A}_1]\mathbf{C}_n^{1/2}\bar{\mathbf{P}}_E\mathbf{B}_n - \mathbf{Q}_n[\mathbf{A}_1]\mathbf{C}_n^{1/2}\bar{\mathbf{P}}_E\mathbf{C}_n^{-1/2}[(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i(\mathbf{v}_{2i}^{'},\mathbf{0}))\mathbf{e}_0] \\
&\quad +O_p(\frac{1}{\sqrt{n}}) \ .
\end{aligned}
$$

As the convenient notations we denote the leading two terms as $\mathbf{e}_2^{(2.1)}$ and $\mathbf{e}_2^{(2.2)}$, respectively. By using (A.17) for $\mathbf{E}_n^{(1)}$ and noting that $\bar{\mathbf{P}}_E$ is a projection operator, we have

$$\begin{aligned}
E[\mathbf{e}_2^{(2.1)}|\mathbf{x}] &= -\mathbf{Q}_n E[(\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{w}_i\mathbf{z}_i^{'})\mathbf{C}_n^{-1}\mathbf{M}_n\mathbf{D}\mathbf{Q}_n(\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{w}_i\mathbf{z}_i^{'})\mathbf{C}_n^{-1/2}\bar{\mathbf{P}}_E\mathbf{B}_n|\mathbf{x}] \\
&\quad -\mathbf{Q}\mathbf{q}E\left\{[(\frac{1}{\sqrt{n}}\sum_{i=1}^n u_i\mathbf{z}_i^{'}) - \delta\mathbf{B}_n^{'}\bar{\mathbf{P}}_E\mathbf{C}_n^{1/2}]\mathbf{C}_n^{-1}\mathbf{M}_n\mathbf{D}\mathbf{Q}_n\mathbf{q}\right. \\
&\quad \left.\times[(\frac{1}{n}\sum_{i=1}^n u_i\mathbf{z}_i^{'}) - \delta\mathbf{B}_n^{'}\bar{\mathbf{P}}_E\mathbf{C}_n^{1/2}]\mathbf{C}_n^{-1/2}\bar{\mathbf{P}}_E\mathbf{B}_n|\mathbf{x}\right\} \\
&= -\mathbf{Q}\mathbf{q}(\mathbf{e}_0^{'}\mathbf{q})(1-\delta)L + O_p(\frac{1}{\sqrt{n}}) \ ,
\end{aligned}$$

where we have used the relations $\bar{\mathbf{P}}_{E_n}\mathbf{C}_n^{-1/2}\mathbf{M}\mathbf{D} = \bar{\mathbf{P}}_{E_n}\mathbf{E}_n = \mathbf{O}$ , and $E[\mathbf{B}^{'}\bar{\mathbf{P}}_E\mathbf{B}] = L$ for the random vector $\mathbf{B}$, which is the limiting normal random vector of $\mathbf{B}_n$, and we already know that

$$\bar{\mathbf{P}}_{E_n} \xrightarrow{p} \bar{\mathbf{P}}_E$$

as $n \to \infty$ . Also we can calculate

$$\begin{aligned}
\text{(A.40)} \qquad & E[\mathbf{e}_2^{(2.2)}|\mathbf{x}] \\
&= -\mathbf{Q}_n E\left\{[(\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{w}_i\mathbf{z}_i^{'}) + \mathbf{q}(\frac{1}{\sqrt{n}}\sum_{i=1}^n u_i\mathbf{z}_i^{'} - \delta\mathbf{B}_n^{'}\bar{\mathbf{P}}_E\mathbf{C}_n^{1/2})]\right. \\
&\quad \left.\times\mathbf{C}_n^{-1/2}\bar{\mathbf{P}}_E\mathbf{C}_n^{-1/2}[(\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{z}_i\mathbf{w}_i^{'}) + (\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{z}_iu_i\mathbf{q}^{'})]\mathbf{e}_0|\mathbf{x}\right\} \\
&= -\mathbf{Q}E\{(\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{w}_i\mathbf{z}_i^{'})\mathbf{C}_n^{-1/2}\bar{\mathbf{P}}_E\mathbf{C}_n^{-1/2}(\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{w}_i\mathbf{z}_i^{'})\mathbf{e}_0|\mathbf{x}\} \\
&\quad -\mathbf{Q}\mathbf{q}E\{\mathbf{B}_n^{'}\bar{\mathbf{P}}_E\mathbf{B}_n\mathbf{q}^{'} - (1-\delta)\mathbf{B}_n^{'}\bar{\mathbf{P}}_E\bar{\mathbf{P}}_E\mathbf{B}_n\mathbf{q}^{'}\mathbf{e}_0|\mathbf{x}\} \\
&\cong -\mathbf{Q}\mathbf{C}_2^{*}\mathbf{e}_0 \ tr[\mathbf{C}^{-1/2}\bar{\mathbf{P}}_E\mathbf{C}^{-1/2}\mathbf{M}] - (1-\delta)L\mathbf{Q}\mathbf{C}_1^{*}\mathbf{e}_0 \ .
\end{aligned}$$

We now turn to the term of $\mathbf{e}_2^{(1)}$ which has many terms involved. Because we can show $\mathbf{C}_n^{(1)} = O_p(1/\sqrt{n})$ , we rewrite

$$\begin{aligned}
\text{(A.41)} \qquad \mathbf{e}_2^{(1)} &= \mathbf{Q}_n[-\mathbf{D}^{'}\mathbf{M}_n\mathbf{C}_n^{-1}\mathbf{C}_n^{(2)}\mathbf{C}_n^{-1} + \mathbf{E}_n^{(2)}\mathbf{C}_n^{-1}]\mathbf{C}_n^{1/2}\bar{\mathbf{P}}_E\mathbf{B}_n + O_p(\frac{1}{\sqrt{n}}) \\
&= \left\{-\mathbf{Q}\mathbf{D}^{'}\mathbf{M}\mathbf{C}^{-1}(\frac{1}{n}\sum_{i=1}^n p_i^{(2)}\hat{u}_i^2\mathbf{z}_i\mathbf{z}_i^{'}) + \mathbf{Q}\mathbf{D}^{'}(\frac{1}{n}\sum_{i=1}^n p_i^{(2)}\hat{u}_i^2\mathbf{z}_i\mathbf{z}_i^{'})\right. \\
&\quad \left.+\mathbf{Q}(\frac{1}{n}\sum_{i=1}^n p_i^{(2)}(\mathbf{v}_{2i}^{'},\mathbf{0}^{'})\mathbf{z}_i^{'})\mathbf{C}_n^{-1/2}\bar{\mathbf{P}}_E\mathbf{B}_n\right\} + O_p(\frac{1}{\sqrt{n}}) \ .
\end{aligned}$$

As the convenient notations we denote the leading three terms as $\mathbf{e}_2^{(1.1)}$, $\mathbf{e}_2^{(1.2)}$ and $\mathbf{e}_2^{(1.3)}$ , respectively. For the second term, we need to use the explicit expression for $\lambda_1$ , which is the solution of the equation

$$\lambda_0 + \frac{1}{\sqrt{n}}\lambda_1 + O_p(\frac{1}{n}) = \{\mathbf{C}_n^{-1} + \frac{1}{\sqrt{n}}[-\mathbf{C}_n^{-1}\mathbf{C}_n^{(1)}\mathbf{C}_n^{-1}]\}$$

$$\times\{[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i u_i - \mathbf{M}_n\mathbf{D}\mathbf{e}_0] + \frac{1}{\sqrt{n}}[-\mathbf{M}_n\mathbf{D}\mathbf{e}_1 - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i(\mathbf{v}_{2i}^{'}, \mathbf{0}^{'})\mathbf{e}_0]\} \ .$$

Then we have the representation as

$$\begin{aligned}
& \lambda_1 \\
= \ & -\mathbf{C}_n^{-1}\mathbf{M}_n\mathbf{D}\mathbf{e}_1 - \mathbf{C}_n^{-1}(\frac{1}{\sqrt{n}})\sum_{i=1}^{n}\mathbf{z}_i(\mathbf{v}_{2i}^{'}, \mathbf{0}^{'})\mathbf{e}_0 - \mathbf{C}_n^{-1}\mathbf{C}_n^{(1)}\mathbf{C}_n^{-1}\bar{\mathbf{P}}_{E_n}\mathbf{B}_n \\
= \ & -\mathbf{C}_n^{-1}\mathbf{M}_n\mathbf{D}[-\mathbf{Q}_n\mathbf{D}^{'}\mathbf{M}_n\mathbf{C}_n^{-1}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{w}_i^{'})\mathbf{e}_0 - (\mathbf{q}^{'}\mathbf{e}_0)\mathbf{e}_0 + \mathbf{Q}_n(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{w}_i^{'})\mathbf{C}_n^{-1/2}\bar{\mathbf{P}}_E\mathbf{B}_n] \\
& \quad +\sum_{i=1}^{n}\mathbf{z}_i(\mathbf{v}_{2i}^{'}, \mathbf{0}^{'})\mathbf{e}_0 + O_p(\frac{1}{\sqrt{n}}) \\
= \ & -\mathbf{C}_n^{-1}[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i u_i(\mathbf{q}^{'}\mathbf{e}_0) + (\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{w}_i^{'})\mathbf{e}_0 \\
& -\mathbf{C}_n^{-1}[-2\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^{'}u_i^2 + (\mathbf{q}^{'}\mathbf{e}_0)\mathbf{C}_n^{-1/2}\bar{\mathbf{P}}_E\mathbf{B}_n] + O_p(\frac{1}{\sqrt{n}}) \\
= \ & -\mathbf{C}_n^{-1/2}\bar{\mathbf{P}}_E\mathbf{C}_n^{-1/2}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{w}_i^{'})\mathbf{e}_0 + \mathbf{C}_n^{-1}\mathbf{M}_n\mathbf{D}^{'}\mathbf{e}_0(\mathbf{q}^{'}\mathbf{e}_0) \\
& -\mathbf{C}_n^{-1}\mathbf{M}_n\mathbf{D}^{'}\mathbf{Q}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{w}_i\mathbf{z}_i^{'})\mathbf{C}_n^{-1/2}\bar{\mathbf{P}}_E\mathbf{B}_n - \mathbf{C}_n^{-1}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i u_i)(\mathbf{q}^{'}\mathbf{e}_0) \\
& \quad +2(\mathbf{q}^{'}\mathbf{e}_0)\mathbf{C}_n^{-1/2}\bar{\mathbf{P}}_E\mathbf{B}_n + O_p(\frac{1}{\sqrt{n}}) \ .
\end{aligned}$$

Then we summarize the second term of $\mathbf{e}_2^{(1)}$ as

$$\mathbf{e}_2^{(1.2)} = \mathbf{Q}\mathbf{D}^{'}\mathbf{M}_n[\lambda_0^{'}\mathbf{z}_i(\mathbf{v}_{2i}^{'}, \mathbf{O})\mathbf{e}_0 - \lambda_1^{'}\mathbf{z}_i u_i + (\lambda_0^{'}\mathbf{z}_i u_i)^2]\mathbf{C}_n^{-1/2}\bar{\mathbf{P}}_E\mathbf{B}_n + O_p(\frac{1}{\sqrt{n}}) \ .$$

Hence we have

(A.42) $$E[\mathbf{e}_2^{(1.2)}|\mathbf{x}] = O_p(n^{-1/2}) \ .$$

In order to deal with the terms in $\mathbf{e}_2^{(1.1)}$, we notice that

$$\begin{aligned}
& \mathbf{C}_n^{(2)}\mathbf{C}_n^{-1/2}\bar{\mathbf{P}}_E\mathbf{B}_n \\
= \ & \mathbf{M}_n\{(\mathbf{z}_i^{'}\mathbf{D}\mathbf{e}_0)^2 + [\mathbf{v}_{2i}^{'}, \mathbf{O})\mathbf{e}_0 - 2p_i^{(1)}(\mathbf{z}_i^{'}\mathbf{D}\mathbf{e}_0 + (\mathbf{v}_{2i}^{'}, \mathbf{O})\mathbf{e}_0 - 2u_i\mathbf{z}_i^{'}\mathbf{D}\mathbf{e}_1] + p_i^{(2)}u_i^2\} \\
& \quad +\mathbf{C}_n^{-1/2}\bar{\mathbf{P}}_E\mathbf{B}_n + O_p(\frac{1}{\sqrt{n}}) \ .
\end{aligned}$$

Then we can evaluate the conditional expectations of each term and it is possible to show that

(A.43) $$E[\mathbf{e}_2^{(1.1)}|\mathbf{x}] = O_p(\frac{1}{\sqrt{n}}) \; .$$

Hence the only remaining contribution on the order of $E[\mathbf{e}_2^{(1)}|\mathbf{x}]$ comes from the terms in $E[\mathbf{e}_2^{(1.3)}|\mathbf{x}]$ . Then we write

(A.44)
$$\begin{aligned}
&\mathbf{e}_2^{(1.3)} \\
&= \delta\mathbf{Q}_n[(\frac{1}{n}\sum_{i=1}^{n}\mathbf{w}_i\;\mathbf{z}_i^{'}) + \mathbf{q}(\frac{1}{n}\sum_{i=1}^{n}u_i\;\mathbf{z}_i^{'})] \\
&\quad\times[\lambda_0^{'}\mathbf{z}_i(\mathbf{w}_i^{'}+\mathbf{q}^{'}u_i)\mathbf{e}_0 - \lambda_1^{'}\mathbf{z}_i u_i + (\lambda_0^{'}\mathbf{z}_i u_i)^2]\mathbf{C}_n^{-1/2}\bar{\mathbf{P}}_E\mathbf{B}_n + O_p(\frac{1}{\sqrt{n}}) \\
&= \delta\mathbf{Q}_n[\frac{1}{n}\sum_{i=1}^{n}(\mathbf{w}_i+\mathbf{q}u_i)(\mathbf{w}_i^{'}+\mathbf{q}^{'}u_i)\mathbf{e}_0\lambda_0^{'}\mathbf{z}_i\mathbf{z}_i^{'}]\mathbf{C}_n^{-1/2}\bar{\mathbf{P}}_E\mathbf{B}_n \\
&\quad-\delta\mathbf{Q}_n[\frac{1}{n}\sum_{i=1}^{n}(\mathbf{w}_i+\mathbf{q}u_i)\mathbf{z}_i^{'}\lambda_1^{'}\mathbf{z}_i u_i]\mathbf{C}_n^{-1/2}\bar{\mathbf{P}}_E\mathbf{B}_n \\
&\quad+\delta\mathbf{Q}_n[\frac{1}{n}\sum_{i=1}^{n}(\mathbf{w}_i+\mathbf{q}u_i)\mathbf{z}_i^{'}(\lambda_0^{'}\mathbf{z}_i)^2u_i^2]\mathbf{C}_n^{-1/2}\bar{\mathbf{P}}_E\mathbf{B}_n + O(\frac{1}{\sqrt{n}}) \; .
\end{aligned}$$

Hence the conditional expectation is given by

(A.45)
$$\begin{aligned}
&E[\mathbf{e}_2^{(1.3)}|\mathbf{x}] \\
&= \delta\mathbf{Q}\mathbf{C}_2\mathbf{e}_0\; tr(\mathbf{C}^{-1/2}\mathbf{M}\mathbf{C}^{-1/2}\bar{\mathbf{P}}_E) + \delta\; L\;\mathbf{Q}\mathbf{C}_1^{*}\mathbf{e}_0 \\
&\quad-\delta\mathbf{Q}E[(\frac{1}{n}\sum_{i=1}^{n}(\mathbf{w}_i+\mathbf{q}u_i)\mathbf{z}_i^{'}\lambda_1^{'}\mathbf{z}_i u_i\mathbf{C}_n^{-1/2}\bar{\mathbf{P}}_E\mathbf{B}_n|\mathbf{x}] + O_p(\frac{1}{\sqrt{n}}) \; .
\end{aligned}$$

By substituting $\lambda_1$ into the last term in the above expression, the thrid part of (A.45) can be rewritten as

$$\begin{aligned}
&-\delta\mathbf{Q}E\Big\{(\frac{1}{n}\sum_{i=1}^{n}\mathbf{w}_i\mathbf{z}_i^{'})[-\mathbf{C}_n^{-1/2}\bar{\mathbf{P}}_E\mathbf{C}_n^{-1/2} - (\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{w}_i^{'})\mathbf{e}_0 \\
&-\delta\mathbf{Q}\mathbf{q}(\frac{1}{n}\sum_{i=1}^{n}u_i\mathbf{z}_i^{'})[\mathbf{C}_n^{-1}\mathbf{M}\mathbf{D}\mathbf{e}_0(\mathbf{q}^{'}\mathbf{e}_0 - \mathbf{C}_n(\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i u_i(\mathbf{q}^{'}\mathbf{e}_0)]^{'}u_i\mathbf{z}_i)\mathbf{C}_n^{-1/2}\bar{\mathbf{P}}_E\mathbf{B}_n|\mathbf{x}\Big\} \\
&-2\mathbf{Q}\mathbf{q}(\mathbf{q}^{'}\mathbf{e}_0)E[\mathbf{B}_n^{'}\bar{\mathbf{P}}_E\mathbf{B}_n] + O_p(\frac{1}{\sqrt{n}}) \\
= \;&-\delta\; L\;\mathbf{Q}\mathbf{q}\mathbf{q}^{'}\mathbf{e}_0 + O_p(\frac{1}{\sqrt{n}})
\end{aligned}$$

Then we evaluate each term and we can find that the second term and the third term of (A.45) are cancelled out. Hence we have

(A.46) $$E[\mathbf{e}_2^{(1.3)}|\mathbf{x}] = \delta\;\mathbf{Q}\mathbf{C}_2^{*}\mathbf{e}_0\; tr(\mathbf{M}^{*}\bar{\mathbf{P}}_E) + O(\frac{1}{\sqrt{n}}) \; .$$

## Appendix C : Nonlinear Case

In the nonlinear case we use the notation

$$g(\mathbf{y}_i, \mathbf{z}_i, \theta) = \mathbf{z}_i \; [y_{1i} - h(\mathbf{y}_{2i}, \mathbf{z}_{1i}, \theta)] \tag{A.47}$$

and $\hat{\theta}_{EL}$ be the MEL estimator for the structural parameters satisfying (2.10)-(2.12). Then we have the representation

$$\begin{aligned}
(\text{A.48}) \qquad & [-\sum_{i=1}^n \hat{p}_i \; (\partial g_i(\hat{\theta})]' [\sum_{i=1}^n \hat{p}_i g_i(\hat{\theta}) g_i(\hat{\theta})']^{-1} [\frac{1}{\sqrt{n}} \sum_{i=1}^n g_i(\theta_0)] \\
= \; & [-\sum_{i=1}^n \hat{p}_i \; (\partial g_i(\hat{\theta})]' [\sum_{i=1}^n \hat{p}_i g_i(\hat{\theta}) g_i(\hat{\theta})']^{-1} \\
& \times \Big\{ [-\frac{1}{n} \sum_{i=1}^n g_i(\theta_0)] \hat{\mathbf{e}} + \frac{1}{\sqrt{n}} [-\frac{1}{n} \sum_{i=1}^n (\frac{1}{2}) \partial^{(j,k)} g_i(\theta_0) \hat{\mathbf{e}}^{(j)} \hat{\mathbf{e}}^{(k)}] \\
& \quad + \frac{1}{n} [-\frac{1}{n} \sum_{i=1}^n (\frac{1}{6}) \partial^{(j,k,l)} g_i(\theta_0) \hat{\mathbf{e}}^{(j)} \hat{\mathbf{e}}^{(k)}] \hat{\mathbf{e}}^{(l)} + O_p(\frac{1}{\sqrt{n}})] \Big\}
\end{aligned}$$

where we have used the notations as $\partial^{(j,k)} g_i = \sum_{j,k} \frac{\partial^2}{\partial \theta_j \partial \theta_k} g_i$ and $\hat{e}^{(i)}$ is the i-th component of $\hat{\mathbf{e}}$ . for instance. We use the convergence in probability as $n \to \infty$ that

$$-\sum_{i=1}^n \hat{p}_i \; (\partial g_i(\hat{\theta}) \cong (-\frac{1}{n}) \sum_{i=1}^n \partial g_i(\theta_0) \xrightarrow{p} \mathbf{D}(M) \; , \tag{A.49}$$

$$\sum_{i=1}^n \hat{p}_i g_i(\hat{\theta}) g_i(\hat{\theta})' \xrightarrow{p} \mathbf{C} = \text{plim} \frac{1}{n} \sum_{i=1}^n g_i(\theta_0) g_i(\theta_0)' \; . \tag{A.50}$$

Then we have

$$[\mathbf{D}(M)' \; \mathbf{C}^{-1} \; \mathbf{D}(M)] \; \mathbf{e}_0 = \mathbf{D}(M)' \; \mathbf{C}^{-1} \; [\frac{1}{\sqrt{n}} \sum_{i=1}^n g_i(\theta_0)] \; , \tag{A.51}$$

where we note that $g_i(\theta_0) = \mathbf{z}_i u_i(\theta_0)$ and $\mathbf{e}_0$ is the first order term of $\hat{\mathbf{e}}$ . Then by using the CLT and using the notation $\mathbf{Q} = \mathbf{D}(M)' \; \mathbf{C}^{-1} \; \mathbf{D}(M)$ , we have

$$\mathbf{e}_0 = \mathbf{Q} \; \mathbf{D}(M)' \; \mathbf{C}^{-1} \; [\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{z}_i u_i] \xrightarrow{w} N[\mathbf{0}, \mathbf{Q}] \; . \tag{A.52}$$

In the nonlinear case we also denote $\mathbf{x}$ as the limiting random vector of $\mathbf{e}_0$ . We expand the random variable $\sqrt{n}\hat{\lambda}$ at the true value $\hat{\theta} = \theta_0$ which can be formally written as

$$\begin{aligned}
(\text{A.53}) \qquad & \sqrt{n}\hat{\lambda} \\
= \; & [\mathbf{C}_n + \frac{1}{\sqrt{n}} \mathbf{C}_n^{(1)} + \frac{1}{n} \mathbf{C}_n^{(2)}]^{-1} [\frac{1}{\sqrt{n}} \sum_{i=1}^n g_i(\hat{\theta})]
\end{aligned}$$

$$= [\mathbf{C}_n^{-1}\frac{1}{n}\sum_{i=1}^{n} g_i(\theta_0) + \mathbf{C}_n^{-1}\frac{1}{n}\sum_{i=1}^{n}\partial^j g_i(\theta_0)\mathbf{e}_0^{(j)}]$$

$$\times\frac{1}{\sqrt{n}}\{[\mathbf{C}_n^{-1}\frac{1}{n}\sum_{i=1}^{n}\partial^j g_i(\theta_0)\mathbf{e}_1^{(j)}] + [\mathbf{C}_n^{-1}\frac{1}{n}\sum_{i=1}^{n}(\frac{1}{2})\partial^{j,k} g_i(\theta_0)\mathbf{e}_0^{(j)}\mathbf{e}_0^{(k)}]$$

$$-\mathbf{C}_n^{-1}\mathbf{C}_n^{(1)}\mathbf{C}_n^{-1}[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_i(\theta_0) + \frac{1}{n}\sum_{i=1}^{n}\partial^j g_i(\theta_0)]\} + O_p(\frac{1}{n}) \ .$$

Hence we have the first order term as

$$\lambda_0 = \mathbf{C}_n^{-1}[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_i(\theta_0) + \frac{1}{n}\sum_{i=1}^{n}\partial^j g_i(\theta_0)\mathbf{e}_0^{(j)}]$$

$$\cong \mathbf{C}_n^{-1}[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_i(\theta_0) - \mathbf{D}(M)\mathbf{e}_0]$$

$$\cong \mathbf{C}_n^{-1/2}\bar{\mathbf{P}}_E\mathbf{B}_n \ . \tag{A.54}$$

As for the linear case we write

$$-\sum_{i=1}^{n}\hat{p}_i\ (\partial g_i(\hat{\theta}) = \mathbf{D}(M) + \frac{1}{\sqrt{n}}\mathbf{E}_n^{(1)} + O_p(\frac{1}{n}) \ , \tag{A.55}$$

where we denote

$$\mathbf{E}_n^{(1)} = \sqrt{n}[-\frac{1}{n}\sum_{i=1}^{n}\partial g_i(\theta_0) - \mathbf{D}(M)] + (-1)\frac{1}{n}\sum_{i=1}^{n} p_i^{(1)}\partial g_i(\theta_0) + (-1)\frac{1}{n}\sum_{i=1}^{n} p_i^{(1)}\partial^{j,k} g_i(\theta_0)\mathbf{e}_0^{(k)} \ ,$$

Then we can rewrite

$$[-\sum_{i=1}^{n}\hat{p}_i\partial g_i(\hat{\theta}_0)][\mathbf{e}_0 + \frac{1}{\sqrt{n}}\mathbf{e}_1 + \frac{1}{n}\mathbf{e}_2]$$

$$= \mathbf{D}(M)\mathbf{e}_0 + \frac{1}{\sqrt{n}}\left\{\mathbf{D}(M)\mathbf{e}_1 + \sqrt{n}[-\frac{1}{n}\sum_{i=1}^{n}\partial g_i(\theta_0) - \mathbf{D}(M)]\mathbf{e}_0 + \frac{1}{n}(-1)\sum_{i=1}^{n}\frac{1}{2}\partial^{j,k} g_i(\theta_0)\mathbf{e}_0^{(j)}\mathbf{e}_0^{(k)}\right\}$$

$$+O_p(\frac{1}{n}) \ .$$

Hence we have

$$(\mathbf{D}(M)'\ \mathbf{C}_n^{-1}\ \mathbf{D}(M))\mathbf{e}_1$$

$$= [\mathbf{A}_1][-\frac{1}{n}\sum_{i=1}^{n}\partial g_i(\theta_0) - \mathbf{D}(M)] - \mathbf{D}(M)'\ \mathbf{C}_n^{-1}\ \left\{\sqrt{n}[-\frac{1}{n}\sum_{i=1}^{n}\partial g_i(\theta_0) - \mathbf{D}(M)]\mathbf{e}_0\right.$$

$$\left. + \frac{1}{n}(-1)\sum_{i=1}^{n}\frac{1}{2}\partial^{j,k} g_i(\theta_0)\mathbf{e}_0^{(j)}\mathbf{e}_0^{(k)}\right\} \ .$$

We notice that in the nonlinear case there is an extra term due to the nonlinearity in the above expression. By using the similar arguments as in the linear case, we have

$$E[\mathbf{C}_n^{(1)}\ \mathbf{C}_n^{-1}\mathbf{C}_n^{1/2}\bar{\mathbf{P}}_E\mathbf{B}_n|\mathbf{x}] = O_p(\frac{1}{\sqrt{n}}) \ , \tag{A.56}$$

and

$$
\begin{aligned}
[\mathbf{A}_1] &= \mathbf{E}_n^{(1)'}\,\mathbf{C}_n^{-1} \\
&= \Big\{\sqrt{n}[-\frac{1}{n}\sum_{i=1}^{n}\partial g_i(\theta_0) - E(-\frac{1}{n}\sum_{i=1}^{n}\partial g_i(\theta_0))] - \lambda_0^{'}\frac{1}{n}\sum_{i=1}^{n}u_i\mathbf{z}_i\mathbf{z}_i^{'}\partial h_i(\theta_0)) \\
&\qquad -\frac{1}{n}\sum_{i=1}^{n}\partial^{j,k}g_i(\theta_0)\mathbf{e}_0^{(j)}\mathbf{e}_0^{(k)}\Big\}\,\mathbf{C}_n^{-1}\ ,
\end{aligned}
$$

where $E(\cdot)$ on the righ thand side is the expectation operator and we have used the notation that $\mathbf{x}$ is the limiting random vector of $\mathbf{e}_0$ as $n \longrightarrow +\infty$ . Then we take the conditional expectation of $\mathbf{e}_1$ given $\mathbf{x}$ . It is easily seen that the first two terms are of the same form which are given by

$$
-\mathbf{q}^{'}\mathbf{x}\mathbf{x}^{'} + (1-\delta)L\ \mathbf{Q}\ \mathbf{q}
$$

as in the linear case. The expectation of the extra last term becomes

$$
E\{\frac{1}{2}\mathbf{Q}\mathbf{D}(M)^{'}\,\mathbf{C}_n^{-1}\,\frac{1}{n}(-1)\sum_{i=1}^{n}(\partial^{(j,k)}h_i(\theta_0))(-\mathbf{z}_i)\mathbf{e}_0^{(j)}\mathbf{e}_0^{(k)}\} \tag{A.57}
$$

$$
= (-\frac{1}{2})\mathbf{Q}\mathbf{D}(M)^{'}\,\mathbf{C}_n^{-1}\ tr[\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i(\partial^{(j,k)}h_i(\theta_0))\mathbf{e}_0^{(j)}\mathbf{e}_0^{(k)}]\ .
$$

Hence by summarizing each terms we have the formula in *Theorem 5.1.*