

Dissertation on  
Generalized Empirical Likelihood Estimators

By  
Jin Xiang

A Dissertation Presented in Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy

Approved April 2013 by  
Graduate Supervisory Committee:

Seung Ahn, Co-chair  
Sunil Wahal, Co-chair  
Sreedhar Bharath  
Rajnish Mehra  
Yuri Tserlukevich

ARIZONA STATE UNIVERSITY

August 2013

## ABSTRACT

Schennach (2007) has shown that the Empirical Likelihood (EL) estimator may not be asymptotically normal when a misspecified model is estimated. This problem occurs because the empirical probabilities of individual observations are restricted to be positive. I find that even the EL estimator computed without the restriction can fail to be asymptotically normal for misspecified models if the sample moments weighted by unrestricted empirical probabilities do not have finite population moments. As a remedy for this problem, I propose a group of alternative estimators which I refer to as modified EL (MEL) estimators. For correctly specified models, these estimators have the same higher order asymptotic properties as the EL estimator. The MEL estimators are obtained by the Generalized Method of Moments (GMM) applied to an exactly identified model. The simulation results provide promising evidence for these estimators.

In the second chapter, I introduce an alternative group of estimators to the Generalized Empirical Likelihood (GEL) family. The new group is constructed by employing demeaned moment functions in the objective function while using the original moment functions in the constraints. This designation modifies the higher-order properties of estimators. I refer to these new estimators as Demeaned Generalized Empirical Likelihood (DGEL) estimators. Although Newey and Smith (2004) show that the EL estimator in the GEL family has fewer sources of bias and is higher-order efficient after bias-correction, the demeaned exponential tilting (DET) estimator in the DGEL group has those superior properties. In addition, if data are symmetrically distributed, every estimator in the DGEL family shares the same higher-order properties as the best member.

## ACKNOWLEDGMENTS

I would like to express my gratitude to my supervisors Dr. Seung Ahn and Dr. Sunil Wahal, whose expertise, understanding, and patience, added considerably to my graduate experience. I would like to thank the other members of my committee, Dr. Sreedhar Bharath, Dr. Rajnish Mehra and Dr. Yuri Tserlukevich, for their great guidance and assistance to my research.

I want to thank my family for their great support. A special thanks to my husband Huiyang Fei, whose understanding, encouragement and love contribute a lot to my completion of this dissertation.

# TABLE OF CONTENTS

	Page
LIST OF TABLES .....	v
CHAPTER	
1	MODIFIED EMPIRICAL LIKELIHOOD ESTIMATORS..... 1
	1.1 Introduction..... 1
	1.2 GEL and EL Estimators ..... 5
	1.3 Modified Empirical Likelihood Estimators ..... 11
	1.3.1 Definition..... 11
	1.3.2 Asymptotic Normality of MEL Estimator ..... 14
	1.3.3 Higher-Order Properties of MEL Estimator for Correctly Specified Models ..... 18
	1.4 Finite-Sample Properties of MEL Estimators..... 22
	1.5 Conclusion ..... 36
2	DEMEANED GENERALIZED EMPIRICAL LIKELIHOOD ESTIMATORS ..... 37
	2.1 Introduction..... 37
	2.2 Demeaned Generalized Empirical Likelihood Estimators .....39
	2.2.1 Review of Related Estimators .....39
	2.2.2 Definition of DGEL Estimators ..... 42
	2.2.3 Higher-Order Properties of DGEL Estimators ..... 43
	2.3 DGEL and GEL Estimators .....51
	2.3.1 Comparison of DGEL and GEL estimators .....51

CHAPTER	Page
2.3.2 Other Groups of Demeaned Estimators .....	52
2.4 Demeaned ET Estimator .....	54
2.5 Conclusion .....	57
REFERENCES .....	59
APPENDIX	
I PROOF FOR CHAPTER 1.....	62
II PROOF FOR CHAPTER 2.....	73
III COPYRIGHT.....	102

## LIST OF TABLES

Table	Page
1. The MEL Estimators and Their Higher-order Properties .....	13
2. Results from Misspecified Linear Moment Functions.....	23
3. Results from Misspecified Nonlinear Moment Functions .....	24
4. Results from Correctly Specified Nonlinear Moment Functions .....	25
5. Results from Misspecified Linear Regression Model.....	27
6. Results from Correctly Specified Linear Regression Model (Weak Identification with Different Sample Size) .....	29
7. Results from Correctly Specified Linear Regression Model (Weak and Strong Identification).....	30
8. Results from Correctly Specified Linear Regression Model (Different Numbers of Instruments) .....	32
9. Results from Correctly Specified Linear Regression Model (Different Sample Sizes) .....	33
10. Results from Correctly Specified Linear Regression Model (Different Error Distributions) .....	35
11. The Higher-order Properties of the DGEL and GEL Families .....	52

# CHAPTER 1<sup>1</sup>

## MODIFIED EMPIRICAL LIKELIHOOD ESTIMATORS

### 1.1 Introduction

Since the seminal paper of Hansen (1982), the Generalized Method of Moments (GMM) has been widely used in the economic and finance literature. An asymptotically efficient GMM estimator can be computed by two steps. At the first step, an initial consistent estimator is computed by minimizing a quadratic function of some sample moment functions weighted by an arbitrary weighting matrix. The initial estimator is then used to estimate the optimal weighting matrix. At the second step, an asymptotically more efficient estimator is computed by redoing GMM with the estimated optimal weighting matrix. The estimator from the second step is called two-step GMM estimator, which I simply refer to as the GMM estimator from now on. Despite its popularity, there is mounting evidence that the GMM estimator could have poor finite-sample properties. See for example, Hansen (1996), Christiano and Haan (1996), Altonji and Segal (1996), Clark (1996), and Andersen and Sorensen (1996), amongst others. The common conclusion presented by these papers is that the GMM estimator can have a large finite-sample bias.

Recently, Generalized Empirical Likelihood (GEL) estimators have been proposed as alternatives to the GMM estimator. The three well known members of the GEL family are the Empirical Likelihood (EL), Exponential Tilting (ET), and Continuous Updating (CU) estimators. Newey and Smith (2004) show that the GMM estimator

---

<sup>1</sup> This chapter is written with Seung Ahn.

suffers from biases due to four sources: the first caused by using sample moment functions in estimation instead of their population moments, the second caused by the correlation between sample moments and their Jacobians, the third caused by the use of an estimated optimal weighting matrix, and the last caused by using an arbitrary initial estimator to estimate the weighting matrix. They show that the GEL estimators remove the second and fourth bias terms while the EL estimator removes the third. Imbens (1997) and Newey and Smith (2004) also find that the bias in the EL estimator does not increase with the number of moment functions used, and that the estimator is invariant to linear transformations of the moment functions. The bias-corrected EL estimator is higher-order efficient among the GEL estimators in the sense that the  $O(n^{-2})$  component of its mean square error (MSE) is the smallest.

Despite these desirable properties, some studies have found that the finite-sample properties of the EL estimator are not always promising. Two problems related to the EL estimator have been documented. Simulations by Mittelhammer, Judge and Schoenberg (2003), Guggenberger and Hahn (2005) and Guggenberger (2008) show that the EL estimator can have large MSE (RMSE) depending on the model, and that its finite-sample distribution often has a heavy tail. The second problem of the EL estimator, which is identified by Schennach (2007), is that it may fail to be root- $n$  consistent when a misspecified model is estimated, where  $n$  denotes sample size. This problem can occur even if the model is only slightly misspecified. If so, the model specification tests based on the EL estimator may not have power.

This paper is motivated by my observations from simulations that the EL estimator's finite-sample distribution crucially depends on how it is computed. As Smith



(1997) and Newey and Smith (2004) show, the EL estimator can be computed by finding a saddle point solution of a minmax problem. This estimator is the one that is used by Guggenberger and Hahn (2005), Guggenberger and Smith (2007) and Guggenberger (2008). Alternatively, some studies (*e.g.*, Imbens, 2002; Newey and Smith, 2004) suggest computing the EL estimator by solving the first-order conditions for the minmax problem. Since this estimator is computed by the Method of Moments (MM), I refer to the estimator as the EL-MM estimator. The EL and EL-MM estimators are not necessarily asymptotically identical because the EL estimator is computed with the empirical probabilities of individual observations restricted to be non-negative, while the EL-MM estimator is not. The EL-MM estimator is easier to compute, and furthermore, the theoretical results of Newey and Smith (2004) are for the EL-MM estimator, not necessarily for the EL estimator.

However, the EL-MM estimator has problems as well. I find that the estimator is very sensitive to the initial parameter values used to compute the solution. I also find that the EL-MM estimator's finite-sample variance can increase with sample size,  $n$ , for some misspecified models. This result indicates that, similar to the EL estimator, the EL-MM estimator can fail to be root- $n$  consistent for some misspecified models. This problem occurs when the sample moment functions used for the EL-MM estimator do not have finite population moments. One example is the case in which sample moment functions are normally distributed.

Based upon these observations, I propose a group of estimators which I refer to as modified EL (MEL) estimators. These estimators are obtained by the Generalized Method of Moments (GMM) applied to exactly identifying first-order conditions. The

sample first-order conditions used are modified versions of the functions used for the EL-MM estimator.

It is generally assumed in the literature that for the cases with exact identification, the GMM estimator is numerically identical to the Method of Moments (MM) estimator. As I show later, however, this is not necessarily true for misspecified models. For the exactly identifying but misspecified moment functions, the population moments of the functions are not equal to zeros at any parameter value. For this case, the MM estimator has no point to converge in probability, and consequently, it cannot be root- $n$  consistent. In addition, the MM estimator may not exist in finite samples. In contrast, for the same case, the GMM estimator can be root- $n$  consistent. Thus, the GMM estimator is not necessarily numerically identical to the MM estimator for misspecified models.

Following Hall and Inoue (2003), I derive the asymptotic distributions of the MEL estimators for both correctly specified and misspecified models. While their approach is mainly for overidentified models, it can be easily extended to exactly identified cases. For correctly specified models, the MEL estimators have the same higher-order asymptotic properties as the EL-MM estimator. For misspecified models, the MEL estimators are root- $n$  consistent as long as data have finite (population) moments up to sufficiently high order. My simulation results show that the MEL estimators have much better finite sample properties than the EL-MM estimator.

The rest of this Chapter is organized as follows. Section 1.2 briefly reviews the GEL and EL estimators. Section 1.3 introduces the MEL estimators and their asymptotic properties. Section 1.4 presents my Monte Carlo simulation results. Some concluding remarks follow in Section 1.5.

## 1.2 GEL and EL Estimators

Let  $z_i$  ( $i = 1, \dots, n$ ) be the independently and identically distributed (*i.i.d.*) observations on a random vector  $z$ ; let  $\theta$  be a  $q \times 1$  parameter vector in a compact set  $\Theta \subset \mathbb{R}^q$ ; and  $g_i(\theta) = g(z_i, \theta)$  is a  $p \times 1$  vector of functions with  $p > q$ . I begin by assuming that  $E[g_i(\theta)] = 0$  if and only if  $\theta = \theta_o$ . From now on, whenever this condition holds, I will say the model is “correctly specified.” If the condition does not hold, that is, if  $E[g_i(\theta)] \neq 0$  for all  $\theta \in \Theta$ , I will say the model is “misspecified.” I also use the following notation:

$$G_i(\theta) = \frac{\partial g_i(\theta)}{\partial \theta'}; \hat{G}(\theta) = n^{-1} \sum_{i=1}^n G_i(\theta); G = E(G_i(\theta_o));$$

$$\Omega_i(\theta) = g_i(\theta)g_i(\theta)'; \hat{\Omega}(\theta) = n^{-1} \sum_{i=1}^n \Omega_i(\theta); \Omega = E(\Omega_i(\theta_o)).$$

The GEL estimators are defined as follows. Let  $\rho(v)$  be a strictly concave function on an open interval  $V$  containing zero. Let  $\lambda$  be a  $p \times 1$  vector of parameters depending on  $\theta$  such that for a given  $\theta$ ,  $\lambda \in \Lambda_n(\theta) = \{\lambda \mid \lambda' g_i(\theta) \in V, i = 1, \dots, n\}$ .<sup>2</sup> Then, a GEL estimator is the solution of the following saddle point problem:

$$\hat{\theta}_{GEL} = \arg \min_{\theta \in \Theta} \sup_{\lambda \in \Lambda_n(\theta)} n^{-1} \sum_{i=1}^n \rho(\lambda' g_i(\theta)). \quad (1.1)$$

The Empirical likelihood (EL), Exponential Tilting (ET), and Continuous Updating (CU) estimators all belong to the GEL family, with the objective functions  $\rho(v) = \ln(1-v)$ ,  $\rho(v) = -e^v$  and  $\rho(v) = -(1+v)^2/2$ , respectively. Let  $\rho_j(v)$  denote the  $j^{\text{th}}$  derivative of  $\rho(v)$ . For the GEL estimator, the function  $\rho$  is chosen such that  $\rho_1(0) = \rho_2(0) = -1$ .

---

<sup>2</sup> The parameter  $\lambda$  is the Lagrangean multiplier in the power-divergence dual of GEL.

The duality of the EL estimator developed by Qin and Lawless (1994) and Owen (2001) is the solution to the problem:

$$\hat{\theta}_{EL} = \arg \max_{\theta \in \Theta, \pi_1, \dots, \pi_n} \ln \pi_i, \quad (1.2)$$

subject to

$$\sum_{i=1}^n \pi_i g_i(\theta) = 0 \text{ and } \sum_{i=1}^n \pi_i = 1. \quad (1.3)$$

Newey and Smith (2004) have shown that the GEL estimator solving (1.1) with  $\rho(v) = \ln(1-v)$  is the same as the solution for (1.2) and (1.3). The empirical probability of the observation  $i$  is

$$\frac{1 / (1 - \hat{\lambda}'_{EL} g_i(\hat{\theta}_{EL}))}{\sum_{i=1}^n \left[ 1 / (1 - \hat{\lambda}'_{EL} g_i(\hat{\theta}_{EL})) \right]}.$$

Because the  $\rho$  function of the EL estimator is logarithmic, it should be that

$1 - \hat{\lambda}'_{EL} g_i(\hat{\theta}_{EL}) > 0$ , for any  $i$ . Thus, the empirical probabilities from the EL estimation should be all positive.

The GMM estimator is the solution to the following problem:

$$\hat{\theta}_{GMM} = \arg \min_{\theta \in \Theta} n^{-1} \hat{g}(\theta)' \Omega(\tilde{\theta})^{-1} \hat{g}(\theta), \quad (1.4)$$

where  $\tilde{\theta}$  is any initial consistent estimator. If an interior solution exists for (1.4), the GMM estimator satisfies:

$$\hat{G}(\hat{\theta}_{GMM})' [\hat{\Omega}(\tilde{\theta})]^{-1} \hat{g}(\hat{\theta}_{GMM}) = 0. \quad (1.5)$$

For correctly specified models, Newey and Smith (2004) show that the GMM estimator has four bias terms of  $O(n^{-1})$ :

$$Bias(\hat{\theta}_{GMM}) = B_I + B_G + B_\Omega + B_W,$$

where  $B_I$  is the bias by using  $\hat{g}(\theta)$  for (1.5) instead of  $E(g_i(\theta))$ ;  $B_G$ , by using  $\hat{G}(\theta)$  instead of  $G$ ;  $B_\Omega$ , by using  $\hat{\Omega}(\theta_o)$  instead of  $\Omega$ ; and  $B_W$  by using  $\hat{\Omega}(\tilde{\theta})$  instead of  $\hat{\Omega}(\theta_o)$ . In contrast, the GEL estimators have only two bias terms:

$$Bias(\hat{\theta}_{GEL}) = B_I + (1 + \rho_3(0)/2)B_\Omega.$$

Specifically,

$$B_I = HE(G_i(\theta_o)Hg_i(\theta_o))/n - H\Sigma_{j=1}^q E\left(\frac{\partial G_i(\theta_o)}{\partial \theta_j}\right)\Sigma e_j / (2n),$$

$$B_\Omega = E[g_i(\theta_o)g_i(\theta_o)'Pg_i(\theta_o)]/n,$$

where  $e_j$  is the  $j^{th}$  unit vector, and

$$\Sigma = (G'\Omega^{-1}G)^{-1}; H = \Sigma G'\Omega^{-1}; \text{ and } P = \Omega^{-1} - \Omega^{-1}G\Sigma G'\Omega^{-1}. \quad (1.6)$$

For the EL estimator,  $\rho_3(0) = -2$ . Thus, the EL estimator has only one bias term:

$$Bias(\hat{\theta}_{EL}) = B_I.$$

New and Smith (2004) and Schennach (2007) also find that

$$Var(\hat{\theta}_k) = \frac{\Sigma}{n} + \frac{\Xi_k}{n^2} + o(n^{-2}), \quad (1.7)$$

where  $\Sigma$  is the asymptotic variance matrix of  $Var[\sqrt{n}(\hat{\theta}_k - \theta_o)]$ , and  $k = \text{GEL or GMM or EL}$ . They also find that  $\Xi_k - \Xi_{EL}$  is positive semi-definite. In this sense, the EL estimator is said to be higher-order efficient.

However, it is important to note that the EL estimator may not be higher order-efficient in terms of mean square error (MSE). All of the GEL and GMM estimators have  $O(n^{-1})$  biases. If the biases are not corrected, the EL estimator can have greater

MSE than some other GEL estimators. In terms of MSE, only the bias-corrected EL estimator is higher-order efficient.

Despite its desirable higher-order asymptotic properties, the EL estimator is likely to suffer from some numerical problems. First, as Grendar and Judge (2009) show, even if the model is correctly specified (that is,  $E(g_i(\theta)) = 0$  if and only if  $\theta = \theta_o$ ), the EL estimator which solves the problem (1.1) may not exist, depending on the model and data. For small samples, the probability of no solution may not be negligible. If so, the estimator's finite-sample properties could be quite different from its asymptotic counterparts.

Second, as Schennach (2007) shows, when the model is misspecified and the moment function  $g_i(\theta)$  is unbounded for any  $\theta$ , the EL estimator could not be root- $n$  consistent. The reason is as follows. Let  $\phi^* = (\lambda^*, \theta^*)'$  be the solution for the population version of the problem (1.1):

$$\theta^* = \arg \min_{\theta \in \Theta} \sup_{\lambda \in \Lambda(\theta)} E[\rho(\lambda' g(z, \theta))], \Lambda(\theta) = \{\lambda \mid \lambda' g(z, \theta) < 1, z \in R_z\}, \quad (1.8)$$

where  $\rho(v) = \ln(1-v)$ , and  $R_z$  is the range of the random vector  $z$ . When the function  $g(\cdot, \theta)$  is not bounded for any  $\theta \in \Theta$ , the only possible value of  $\lambda$  is zero. That is,  $\Lambda(\theta) = \{0\}$ , and consequently,  $\lambda^* = 0$ . But, the first-order condition for  $\theta^*$  to minimize the objective function (1.8) is  $E(g(z, \theta^*)) = 0$ . However, such  $\theta^*$  does not exist for misspecified models. That is, for misspecified models there is no solution for the problem (1.8). Therefore, the EL estimator  $\hat{\theta}_{EL}$  has no point to converge in probability.

Based on this result, Schennach (2007) shows that the EL estimator is not root- $n$  consistent.

In response to this problem, Schennach (2007) proposes an alternative estimator named Exponentially Tilting Empirical Likelihood (ETEL), which has the same asymptotic properties as the EL estimator if the moment function  $g_i(\theta)$  is correctly specified, and which is still root- $n$  consistent for the cases with misspecified moment functions. The ETEL estimator is the solution to the following problem:

$$\hat{\theta}_{ETEL} = \arg \min_{\theta \in \Theta} \sum_{i=1}^n \exp(\hat{\lambda}(\theta)'(g_i(\theta) - \hat{g}(\theta))), \quad (1.9)$$

subject to

$$\sum_{i=1}^n \exp(\hat{\lambda}(\theta)' g_i(\theta)) g_i(\theta) = 0. \quad (1.10)$$

An important advantage of using this estimator is that the range of  $\hat{\lambda}(\theta)$  is not restricted by data. In addition, the empirical probabilities  $\exp(\hat{\lambda}(\theta)' g_i(\theta)) / \sum_{i=1}^n \exp(\hat{\lambda}(\theta)' g_i(\theta))$  are positive for any  $\theta$ .

While the ETEL estimator has promising asymptotic properties, its computation requires an optimization procedure with multiple constraints (1.10), which could be complicated when  $p$  or  $q$  are large.

I can expect that when the range of  $\lambda$  given  $\theta$ ,  $\Lambda_n(\theta)$ , is restricted by the data, any GEL estimator would have similar problems. One way to avoid this problem is to find the solution of the first-order conditions for the problem (1.1):

$$n^{-1} \sum_{i=1}^n f_i(\phi) = 0, \quad (1.11)$$

where  $\phi = (\lambda', \theta)'$  and

$$f_i(\phi) = \begin{pmatrix} \rho_1(\lambda' g_i(\theta)) g_i(\theta) \\ \rho_1(\lambda' g_i(\theta)) G_i(\theta)' \lambda \end{pmatrix}_{(p+q) \times 1}.$$

Imbens (2002) suggests computing the GEL or EL estimators by solving the equation (1.11). The resulting GEL estimators are MM estimators, as they are computed by the method of moments. In fact, Newey and Smith (2004) have derived the asymptotic distributions and higher-order biases of the GEL estimators computed by MM. I refer to the MM estimator computed with  $\rho(v) = \ln(1-v)$  as EL-MM estimator. Different from the EL estimator, the EL-MM estimator does not restrict the range of the parameter vector  $\phi$  depending on data. Thus, the EL-MM estimator of  $\lambda$  by solving (1.8) does not converge to zero in probability if a misspecified model is estimated.

While the EL-MM estimator can avoid the problem found by Schennach (2007), it may not be root- $n$  consistent for some misspecified models. There are two important conditions for the EL-MM estimator to be root- $n$  consistent for either correctly specified or misspecified models. The first is that  $E[f_i(\phi)]$  is finite for  $\phi = (\lambda', \theta)'$  in a compact set  $\Phi \times \Lambda$ . The second condition is that  $E[f_i(\phi)] = 0$  if and only if  $\phi = \phi^*$  and  $\phi^*$  is unique. The first condition can be violated if the functions in  $f_i(\phi)$  do not have finite moments. For the EL-MM estimator, the first condition implies that

$E[g_i(\theta) / (1 - \lambda' g_i(\theta))]$  should exist for some  $\lambda \neq 0$  for misspecified models. However, if  $g_i(\theta)$  is normal,  $E[g_i(\theta) / (1 - \lambda' g_i(\theta))]$  does not exist if  $\lambda \neq 0$ . This is because any ratio of normal random variables does not have finite moments (see Marsaglia, 1965; Marsaglia, 2006). For this case, there is no guarantee that the EL-MM estimator would be root- $n$  consistent. Even if  $E(f_i(\phi))$  exists, the second condition may not hold for



some misspecified models, since the function  $f_i(\phi)$  is nonlinear. Thus, although the number of functions in  $E(f_i(\phi))$  with  $\rho(v) = \ln(1-v)$  equals the number of parameters in  $\phi$  ( $p+q$ ), there is no guarantee that a solution exists for  $E[f_i(\phi)] = 0$ .

In summary, the EL and EL-MM estimators may not have well-defined asymptotic distributions for misspecified models. Two adjustments seem to be desirable. The first is to transform the function  $f_i(\phi)$  so that it does not have ratio forms, and the second is to use an alternative estimation procedure that produces an estimator with well-defined asymptotic distribution. I consider such adjustments in the next section.

### 1.3 Modified Empirical Likelihood Estimators

#### 1.3.1. Definition

One important reason for the problems of the EL and EL-MM estimators identified in the previous section is that the estimators use the logarithmic function  $\rho(v) = \ln(1-\rho)$ . A simple way to avoid the problem is to use the Taylor approximations of the logarithmic function. Some studies, *e.g.*, Owen (2001) and Schennach (2007), discussed this alternative method. However, in my knowledge, no study has explored the advantages of using a Taylor-approximated function. In this section, I examine the higher-order asymptotic properties of the estimators based on this approximation.

Define:

$$m_i(\phi) = \begin{pmatrix} m_{1i}(\phi) \\ m_{2i}(\phi) \end{pmatrix} = \begin{pmatrix} \rho_1^A(\lambda' g_i(\theta)) g_i(\theta) \\ \rho_1^B(\lambda' g_i(\theta)) G_i(\theta)' \lambda \end{pmatrix}, \quad (1.12)$$

where  $\rho^A(v)$  and  $\rho^B(v)$  are scalar functions and  $\rho_j^A$  and  $\rho_j^B$  denotes their  $j^{\text{th}}$  derivatives. The functions  $\rho^A$  and  $\rho^B$  need not be the same, but I impose the following restriction:

Restriction 1: (i) The  $\rho^A$  and  $\rho^B$  functions are four times and three times differentiable, respectively. (ii)  $\rho_1^A(0) = \rho_2^A(0) = -1$ ,  $\rho_3^A(0) = -2$ . (iii)  $\rho_1^B(0) = \rho_2^B(0) = -1$ .

This restriction is necessary to derive an estimator which preserves the higher-order asymptotic properties of the EL-MM estimator.

I refer to the solution of the following problem with Restriction 1 as modified EL (MEL) estimator:

$$\hat{\phi}_{MEL} = \arg \min_{\phi \in \Phi} \hat{m}(\phi)' \hat{m}(\phi), \quad (1.13)$$

where  $\hat{m}(\phi) = n^{-1} \sum_{i=1}^n m_i(\phi)$ , and  $\Phi \subset \mathbb{R}^{p+q}$  is a compact set. Because both  $m_i(\phi)$  and  $\phi$  are  $(p+q) \times 1$  vectors, the MEL estimator can be numerically identical to the MM estimator which solves  $\hat{m}(\phi) = 0$  if the solution exists. If I choose

$\rho^A(v) = \rho^B(v) = \ln(1-v)$ , the MM estimator equals the EL-MM estimator.

There are many possible choices for  $\rho^A$  and  $\rho^B$ . Among them, I consider five possible combinations that are listed in Table 1. The  $\rho^A$  functions are simply the third-order Taylor approximations. I have tried the fourth-order approximations in my

simulations, but the MEL estimators with the third-order approximations outperform them, thus I consider only the third-order approximations. The  $\rho^B$  functions are the functions used for some GEL estimators. For example, the  $\rho^B$  functions of the MEL1 and MEL2 estimators are the  $\rho$  functions of the ET and CU estimators. The  $\rho^B$  functions of the MEL4 and MEL5 estimators are the fourth- and third- order Taylor approximations of  $\ln(1-v)$  at  $v=0$ . The  $\rho^B$  function chosen for the MEL3 estimator is an ad hoc choice, but it satisfies Restriction 1.

Table 1

The MEL Estimators and Their Higher-order Properties

	$\rho^A(v)$	$\rho^B(v)$	$O(n^{-1})$ Bias	Higher-Order Efficiency after Bias-correction
MEL1	$-1-v-\frac{v^2}{2}-\frac{v^3}{3}$	$-e^v$	$B_I$	Yes
MEL2	$-1-v-\frac{v^2}{2}-\frac{v^3}{3}$	$-\frac{(v+1)^2}{2}$	$B_I$	Yes
MEL3	$-1-v-\frac{v^2}{2}-\frac{v^3}{3}$	$-v+e^v-ve^v$	$B_I$	Yes
MEL4	$-1-v-\frac{v^2}{2}-\frac{v^3}{3}$	$-1-v-\frac{v^2}{2}-\frac{v^3}{3}-\frac{v^4}{4}$	$B_I$	Yes
MEL5	$-1-v-\frac{v^2}{2}-\frac{v^3}{3}$	$-1-v-\frac{v^2}{2}-\frac{v^3}{3}$	$B_I$	Yes

When two different functions are used for  $\rho^A$  and  $\rho^B$ , the MEL estimator produces two different sets of empirical probabilities: one by  $\rho_1^A$  and the other by  $\rho_1^B$ . Thus, one may wish to use the same function for  $\rho^A$  and  $\rho^B$ . However, I find from

simulations that the MEL estimators with two different functions generally perform better than the MEL estimator with the same function. I do not explore which of the functions  $\rho^A$  and  $\rho^B$  should be used in practice in this paper. However, for either case, I can obtain non-negative empirical probabilities imposing the following optional restrictions.

Restriction 2: (i)  $\rho_1^A(v) < 0$  for all  $v \in \mathbb{R}$ . (ii)  $\rho_1^B(v) < 0$  for all  $v \in \mathbb{R}$ .

The MEL1 and MEL5 estimators satisfy both Restriction 2(i) and 2(ii), while the MEL2, MEL3 and MEL4 estimators satisfy Restriction 2(i) only.

### 1.3.2 Asymptotic Normality of MEL Estimators

In this section, I show that the MEL estimators are root- $n$  consistent whether the model is correctly specified or misspecified. Hall and Inoue (2003) show that the GMM estimators are asymptotically normal even for misspecified models. They consider general cases in which the model is over-identified and data are autocorrelated. While their focus was on over-identified models, their results can be used for exactly identified models. The MEL estimators can be viewed as GMM estimators applied to an exactly identified model. In this view, the MEL estimation is a special case in which data are *i.i.d.*

I begin by listing the assumptions which subsume misspecified models.

Assumption 1.1: (i) The function  $m_i(\phi)$  is continuous and twice differentiable for each  $\phi \in \Phi$ , where  $\Phi \subset \mathbb{R}^{p+q}$  is a compact set. (ii) The moment  $m(\phi) = E[m_i(\phi)]$  exists for every  $\phi \in \text{Int}(\Phi)$ . (iii) Let  $M_i(\phi) = \partial m_i(\phi) / \partial \phi'$  and  $D_i = \partial \text{vec}(M_i(\phi)) / \partial \phi'$ . Then,  $M(\theta) = E(M_i(\theta))$  and  $D(\phi) = E(D_i(\phi))$  exist for every  $\phi \in \text{Int}(\Phi)$ . (iv) There exists a function  $b_i = b(z_i)$  such that in a neighborhood of  $\phi^*$ ,  $N(\phi^*)$ ,  $m_i(\phi)$  and all of its partial derivatives (with respect to  $\phi$ ) up to order two are bounded by  $b_i$ , and  $E(b_i^2)$  is a finite constant. (v)  $E\left[\sup_{\phi \in N(\phi^*)} \|M_i(\phi)\|\right] < \infty$  and  $E\left(\sup_{\phi \in \Phi} \|m_i(\phi)\|^\alpha\right) < \infty$ , for some  $\alpha > 2$ .

Assumption 1.2: (i)  $\phi^* = (\lambda^*, \theta^*)'$  is the unique minimizer of  $m(\phi)'m(\phi)$  in  $\text{Int}(\Phi)$ . (ii) A matrix  $S_* = M_*'M_* + (\mu_*' \otimes I_q)D_*$  is nonsingular, where  $M_* = M(\theta^*)$ ,  $D_* = D(\phi^*)$ , and  $\mu_* = m(\phi^*)$ . (iii)  $\text{Var}(m_{1i}(\phi^*))$  is a positive definite matrix and  $\text{Var}[m_i(\phi^*)]$  is a positive semi-definite matrix. (iv) If  $\lambda^* = 0$ ,  $m(\phi^*) = 0$ .

Assumption 1.1 states the general regularity conditions for GMM or MM.

Several things are noteworthy regarding Assumption 1.2. First, I do not assume that  $m(\phi^*) = 0$ . Instead, I assume that  $M(\phi^*)'m(\phi^*) = 0$  (Assumptions 1.2(i)). Thus, my result can apply even if there is no  $\phi$  that solves  $m(\phi) = 0$ . For misspecified models,  $m(\phi^*)$  could be equal to either a zero vector or non-zero vector. For the cases in which  $m(\phi^*) = 0$ , Assumption 1.2(ii) restricts  $M_*$  to be nonsingular. For the cases in which

$m(\phi^*) \neq 0$ ,  $M_* = M(\phi^*)$  cannot be a full-column matrix because  $M_*'m(\phi^*) = 0$  for nonzero  $m(\phi^*)$ . Nonetheless, as long as the matrix  $S_*$  is invertible, I can derive the asymptotic distribution of the MEL estimator. Third, I do not assume that  $\text{Var}[m_i(\phi^*)]$  is positive definite (Assumption 1.2 (iii)).<sup>3</sup> I do not do so because the variance matrix is not invertible for correctly specified models. If  $\phi^* = (0_{1 \times p}, \theta_o)'$ ,  $m_{2i}(\phi^*) = 0$  for any  $i$ . Thus, the rank of  $\text{Var}(m_i(\phi^*))$  becomes  $p$ . Fourth and finally, Assumptions 1.2(i) and 1.2(iv) imply that for a correctly specified model, a unique  $\theta_o$  exists such that  $E[g_i(\theta_o)] = 0$ .

In order to see why the possibility of  $m(\phi^*) \neq 0$  can be allowed, consider a simple case. Suppose that

$$g_i(\theta) = \begin{pmatrix} x_i - \theta \\ y_i - \theta \end{pmatrix}, \quad (1.14)$$

where

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right). \quad (1.15)$$

Because  $E(x_i) \neq E(y_i)$ , the model is misspecified with  $p = 2$  and  $q = 1$ . Consider the moment function  $m_i(\phi)$  of the MEL1 estimator ( $\rho_1^A(v) = -1 - v - v^2$  and  $\rho_1^B(v) = -e^v$ ).

Suppose that some  $\phi$  solves  $m(\phi) = 0$ . Because  $E(\rho^B(\lambda' g_i(\theta))) < 0$ ,

$E[m_{2i}(\phi)] = 0$  implies that  $\lambda_1 + \lambda_2 = 0$ . However, with such  $\lambda_1$  and  $\lambda_2$ , there is no  $\theta$  that

---

<sup>3</sup> Hall and Inoue (2003) assume that the moment function used for GMM has a positive-definite matrix (Assumption 4, p. 368.). However, the assumption is never used to derive the asymptotic distribution of the GMM estimator.

solves  $E[m_{1i}(\phi)] = 0$ . That is,  $m(\phi) \neq 0$  for any  $\phi$ . Proposition 1 of Hall and Inoue (2003) implies that the MM version of the MEL1 estimator may not exist even in finite samples. Thus, I can expect that computation of the MM version of the MEL1 estimator may suffer from some numerical problems. However, using software (Matlab and Gauss, Version 11), I found that the function  $m(\theta)'m(\theta)$  is minimized a point near  $\phi^* = (0.214, -0.214, 0.5)'$ . At this point,  $m(\phi^*) = (-0.339, 0.339, 0)'$ ,  $M_*'m(\phi^*) = 0$ ,  $\text{rank}(M_*) = 2$ , and  $\text{rank}(S_*) = 3$ . Thus, the moment function of the MEL1 estimator satisfies Assumption 1.2.

Theorem 1 of Hall and Inoue (2003) immediately follows under Assumption 1.1 and 1.2. I state the result without proof:

Theorem 1.1: Let

$$\text{Var} \left( \begin{pmatrix} m_i(\phi^*) - \mu_* \\ (M_i(\phi^*) - M_*)' \mu_* \end{pmatrix} \right) = \begin{pmatrix} \Upsilon_{11} & \Upsilon_{12} \\ \Upsilon_{21} & \Upsilon_{22} \end{pmatrix} \equiv \Upsilon.$$

Then, under Assumptions 1.1 – 1.2 and Restriction 1,

$$\sqrt{n}(\hat{\phi}_{MEL} - \phi^*) \rightarrow_d N(0, \Pi),$$

where,

$$\Pi = S_*^{-1} \begin{pmatrix} M_* \\ I_{(p+q)} \end{pmatrix}' \begin{pmatrix} \Upsilon_{11} & \Upsilon_{12} \\ \Upsilon_{21} & \Upsilon_{22} \end{pmatrix} \begin{pmatrix} M_* \\ I_{(p+q)} \end{pmatrix} (S_*')^{-1}. \quad (1.16)$$

Theorem 1 applies whether or not the model is correctly specified. The solution of  $E[m_i(\phi)] = 0$  does not have to exist. When  $E[m_i(\phi^*)] = \mu_* = 0$ , the variance matrix  $\Pi$  reduces

$$\Pi = M_*^{-1} \Upsilon_{11} M_*'^{-1}, \quad (1.17)$$

which is the asymptotic variance matrix of the MM-counterpart of the MEL estimator. A novelty of Theorem 1 is that for misspecified models, the MEL estimators can be root- $n$  consistent even if their MM-counterparts are not.

For correctly specified models,  $\phi^* = \phi_o = (0_{1 \times p}, \theta_o)'$  and  $\mu_* = 0$ . Thus, the asymptotic variance matrices of the MEL estimators have the form of (1.17). In addition,

$$M_* = M = \begin{pmatrix} \Omega & G \\ G' & 0_{p \times p} \end{pmatrix}; M^{-1} = \begin{bmatrix} P & H' \\ H & -\Sigma \end{bmatrix}; \Upsilon_{11} = \begin{pmatrix} \Omega & 0_{q \times p} \\ 0_{p \times q} & 0_{p \times p} \end{pmatrix}, \quad (1.18)$$

where  $P$ ,  $H$  and  $\Sigma$  are defined in (1.6). Using these results, I can show that for correctly specified models,

$$\sqrt{n}(\hat{\phi}_{MEL} - \phi_o) = \begin{pmatrix} \sqrt{n} \hat{\lambda}_{MEL} \\ \sqrt{n}(\hat{\theta}_{MEL} - \theta_o) \end{pmatrix} \rightarrow_d N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} P & 0 \\ 0 & \Sigma \end{pmatrix} \right).$$

Thus, the MEL estimators of  $\theta$  are asymptotically identical to the GMM estimator.

### 1.3.3 Higher-Order Properties of MEL Estimator for Correctly Specified Models

In this section, I show that the MEL estimators share the same higher-order asymptotic properties as the EL-MM estimator for correctly specified models. To do so, I need to impose stronger assumptions on the moment function  $g_i(\theta)$ . Assumptions 1.1 and 1.2 are made for  $m_i(\phi)$ , not directly for  $g_i(\theta)$ . For convenience, I list the



assumptions on the functions  $g_i(\theta)$ ,  $\rho^A$  and  $\rho^B$ , which are the same as those of Newey and Smith (2004) and Schennach (2007). Many of them are in fact implied by Assumptions 1.1 and 1.2.

Assumption 1.3: (i) For a compact set  $\Theta \subset \mathbb{R}^q$ ,  $\theta_o \in \text{Int}(\Theta)$  is the unique solution to  $E[g_i(\theta)] = 0$ . (ii) The moment function  $g_i(\theta)$  is four times continuously differentiable in a neighborhood of  $\theta_o$ ,  $N(\theta_o)$ , with probability 1. (iii) There exists a function  $b_i = b(z_i)$  such that in  $N(\theta_o)$ ,  $g_i(\theta)$  and all of its partial derivatives up to order four are bounded by  $b_i$  and  $E(b_i^2)$  is a finite constant. (iv) For some  $\alpha > 2$ ,  $E\left(\sup_{\theta \in \Theta} \|g_i(\theta)\|^\alpha\right) < \infty$ , and  $E\left(\sup_{\theta \in \Theta} \|G_i(\theta)\|\right) < \infty$ . (v)  $\Omega$  is nonsingular and finite, and  $\text{rank}(G) = p$ . (vi) The functions  $\rho^A$  and  $\rho^B$  are four times and three times continuously differentiable in a neighborhood of zero, respectively.

Newey and Smith (2004) derived the higher-order bias and variance of the EL-MM estimator under Assumption 1.3. An implicit assumption they made is that the EL-MM estimator exists for any given sample. This assumption can be relaxed.

Assumption 1.4: Let  $\hat{\phi}_{MM}(n)$  be the MM estimator using the moment function  $m_i(\phi)$  and a sample of  $n$ . There is a sequence  $\{\hat{\phi}_{MEL}(n)\}_{n=1}^\infty$  such that  $\hat{m}(\hat{\phi}_{MM}(n)) = 0$ , almost surely.

Assumption 1.4 states that the MM estimator using the moment function of the MEL estimator exists for sufficiently large samples. Whenever the MM estimator exists, it is equal to the MEL estimator. Thus, under Assumption 1.4, the higher-order bias and variance of the MEL estimator are equal to those of the MM estimator, almost surely. As I discussed in the previous sections, misspecified models would not have solution for  $E[m_i(\phi)] = 0$ . For such cases, Assumption 1.4 is violated (see Proposition 1 of Hall and Inoue (2003)). For correctly specified models, however,  $E[m_i(\phi_o)] = 0$ , where  $\phi_o = (0_{1 \times p}, \theta'_o)'$ . Thus, the probability of the MM estimator's existence increases with sample size  $n$ . Thus, Assumption 1.4 is likely to hold for correctly specified models.

Under Assumptions 1.3 and 1.4, the appendix shows that

$$\sqrt{n}(\hat{\theta}_{MEL} - \theta_0) = \tilde{\psi}_\theta + \tilde{Q}_{1,\theta} / \sqrt{n} + \tilde{Q}_{2,\theta} / n + O_p(n^{-2}), \quad (1.19)$$

where  $\tilde{\psi}_\theta$ ,  $\tilde{Q}_{1,\theta}$  and  $\tilde{Q}_{2,\theta}$  are  $O_p(1)$ , and  $E(\tilde{\psi}_\theta) = 0$  and  $Var(\tilde{\psi}_\theta) = \Sigma$ . What I mean by “higher-order bias” is the bias of  $O_p(n^{-1})$ . For any estimator of the form (1.13), its higher-order bias is  $E(\tilde{Q}_{1,\theta} / n)$  because  $E(\tilde{\psi}_\theta) = 0$ .

Theorem 1.2 (Higher-Order Bias): Under Assumptions 1.3 – 1.4 and with Restriction 1, any MEL estimator of  $\theta$  has

$$\text{Bias}(\hat{\theta}_{MEL}) = E(\tilde{Q}_{1,\theta} / n) = B_I.$$

Every MEL estimator has only one bias term  $B_7$  that results from the use of sample means of moment functions  $n^{-1}\sum_{i=1}^n g_i(\theta)$  instead of  $E[g_i(\theta)]$ . The higher-order bias of any MEL estimator is the same as that of the EL-MM estimator.

I now derive the higher-order variance of the MEL estimator. The equation (1.19) implies that

$$\text{Var}(\hat{\theta}_{MEL}) = \frac{\Sigma}{n} + \frac{\Xi_{MEL}}{n^2} + o(n^{-2}),$$

where  $\Xi_{MEL} = \text{Var}(\tilde{Q}_{1,\theta}) + E(\tilde{\psi}_\theta \tilde{Q}'_{2,\theta}) + E(\tilde{Q}_{2,\theta} \tilde{\psi}'_\theta)$ . The appendix shows that EL-MM and MEL estimators have the same  $\tilde{\psi}$  and  $\tilde{Q}_1$ , but different forms of  $\tilde{Q}_2$ . However, it turns out that the difference between the  $\tilde{Q}_2$  terms of the EL-MM estimator and a MEL estimator is uncorrelated with  $\tilde{\psi}$ . Accordingly, the difference does not influence the  $O_p(n^{-2})$  parts of the variance matrices. Thus, I obtain the following result:

**Theorem 1.3 (Higher-Order Efficiency):** Under Assumptions 1.3 – 1.4,  $\Xi_{MEL} = \Xi_{EL}$ .

As Newey and Smith (2004) note, the EL-MM estimator may not be higher-order efficient among GEL-MM estimators in terms of mean score error (MSE). Only the bias-corrected EL-MM estimator is efficient among bias-corrected GEL-MM estimators. Similarly, only the bias-corrected MEL estimator can be said to be more efficient than the biased-corrected GEL estimators.

## 1.4 Finite-Sample Properties of MEL Estimators

I conduct three different experiments for the Monte Carlo studies. For each case, I generate 10,000 different random samples. For each sample, I compute the GMM, EL-GMM, ETEL, and MEL estimators given in Table 1. Using the estimates from 10,000 simulated samples, I compute the median, mean, root mean-square-error (RMSE) and standard error of each estimator.

For my simulations, the EL-GMM estimator is computed by the MEL procedure. I do not compute the EL-MM estimator because of the following reasons. I often experience convergence problems while computing the EL-MM estimator. In addition, the EL-MM estimate is quite sensitive to the initial parameter values. These problems occur more often for misspecified models. The estimate from the MEL procedure is much less sensitive to the initial values. For the same reason, I also compute the ETEL estimator by the MEL procedure using the first-order conditions for the problem given in (1.9) and (1.10).

The first experiment is based on the misspecified model given in (1.14) and (1.15). The simulation results are reported in Table 2. The starting value of  $\theta$  is 1 and that of  $\lambda$  is a vector of zeros.<sup>4</sup> Interestingly, for each sample size ( $n = 100$  and  $500$ ) all of the MEL estimators have the same medians, means and standard errors. The EL-GMM estimator has the largest standard errors for either  $n = 100$  or  $n = 500$ . The estimator's standard error increases with sample size. This is an unexpected result if the EL-GMM estimator is root- $n$  consistent. The ETEL estimator has smaller standard error than the EL-GMM estimator. The standard error of the ETEL estimator decreases as  $n$  increases

---

<sup>4</sup> I tried many different starting values but the results were similar to those reported here.

from 100 to 500 but by a smaller factor than the expected factor of  $\sqrt{5}$  if the ETEL estimator is root- $n$  consistent. In contrast, the standard errors of the MEL estimators fall nearly by the factor of  $\sqrt{5}$ .

Table 2

Results from Misspecified Linear Moment Functions

The moment function used is  $g_i(\theta) = (x_i - \theta, y_i - \theta)'$ , where  $x_i$  and  $y_i$  are independent normal random variables from  $N(0,1)$  and  $N(1,1)$ , respectively.

	Median	Mean	Std	Median	Mean	Std
		$n = 100$			$n = 500$	
GMM	0.5007	0.5005	0.0577	0.5002	0.5002	0.0257
EL-GMM	0.4965	0.4408	1.2852	0.4899	-0.082	3.2061
EDEL	0.4977	0.5031	0.3831	0.5011	0.5042	0.2337
MEL1	0.5004	0.4993	0.0733	0.5006	0.5003	0.0333
MEL2	0.5004	0.4993	0.0733	0.5006	0.5003	0.0333
MEL3	0.5004	0.4993	0.0733	0.5006	0.5003	0.0333
MEL4	0.5004	0.4993	0.0733	0.5006	0.5003	0.0333
MEL5	0.5004	0.4993	0.0733	0.5006	0.5003	0.0333

For the second experiment, I use the following model:

$$g_i(\theta) = \begin{pmatrix} x_i - \theta \\ (x_i - \theta)^2 - 1 \end{pmatrix},$$

where  $x_i \sim N(0,c)$ . I try  $c = 0.8$  and 1. The model is misspecified if  $c = 0.8$  and correctly specified if  $c = 1$ . Table 3 reports the simulation results from the data generated with  $c = 0.8$ . The median and mean biases and RMSEs are computed using  $\theta_o = 0$ . The starting value of  $\theta$  is 1 and that of  $\lambda$  is a vector of zeros. The main results from Table 3

are as follows. First, the MEL estimators have smaller standard errors than the GMM, EL-GMM, and ETEL estimators. Second, the EL-GMM estimator has larger standard error than the GMM estimator for large sample size  $n = 500$ . Third, differently from Table 2, the standard error of the EL-GMM estimator now decreases with  $n$ . Comparing the results from Tables 2 and 3, I can see that for misspecified models, the finite-sample performance of the EL-GMM estimator crucially depends on estimated models.

Table 3

Results from Misspecified Nonlinear Moment Functions

The moment function used is  $g_i(\theta) = (x_i - \theta, (x_i - \theta)^2 - 1)$ , where the  $x_i$  follow  $N(0, c)$  and  $c = 0.8$ . The biases and RMSEs are computed using  $\theta_o = 0$ .

	Median Bias	Mean Bias	RMSE	Std
$n = 100$				
GMM	0.0026	0.0013	0.1195	0.1195
EL-GMM	0.0009	0.0010	0.1077	0.1077
ETEL	0.0008	-0.0086	0.1648	0.1646
MEL1	0.0019	0.0006	0.0948	0.0948
MEL2	0.0014	0.0005	0.1050	0.1050
MEL3	0.0020	0.0007	0.0892	0.0892
MEL4	0.0028	0.0013	0.0886	0.0886
MEL5	0.0021	0.0011	0.0955	0.0955
$n = 500$				
GMM	0.0004	0.0005	0.0549	0.0549
EL-GMM	0.0003	0.0013	0.0666	0.0666
ETEL	0.0006	-0.0008	0.0631	0.0631
MEL1	0.0003	0.0004	0.0431	0.0431
MEL2	0.0005	0.0004	0.0474	0.0474
MEL3	0.0008	0.0003	0.0396	0.0396
MEL4	0.0007	0.0004	0.0394	0.0394
MEL5	0.0005	0.0004	0.0432	0.0432

Table 4 reports the results from the data generated with  $c = 1$  (correctly specified model). All of the estimators show similar performances. They have very small biases even if sample size is small (e.g.,  $n = 100$ ).

Table 4

Results from Correctly Specified Nonlinear Moment Functions

The moment function used is  $g_i(\theta) = (x_i - \theta, (x_i - \theta)^2 - 1)$ , where the  $x_i$  follow  $N(0, c)$  and  $c = 1$ . The biases and RMSEs are computed using  $\theta_o = 0$ .

	Median Bias	Mean Bias	RMSE	Std
$n = 100$				
GMM	-0.0006	0.0003	0.1015	0.1015
EL-GMM	-0.0004	0.0003	0.1012	0.1012
ETEL	-0.0005	0.0001	0.1020	0.1020
MEL1	-0.0004	0.0001	0.1006	0.1006
MEL2	-0.0005	0.0001	0.1008	0.1008
MEL3	-0.0005	0.0001	0.1005	0.1005
MEL4	-0.0006	0.0001	0.1005	0.1005
MEL5	-0.0005	0.0001	0.1005	0.1005
$n = 500$				
GMM	-0.0011	-0.0006	0.0448	0.0448
EL-GMM	-0.0011	-0.0006	0.0448	0.0448
ETEL	-0.0011	-0.0006	0.0448	0.0448
MEL1	-0.0012	-0.0006	0.0448	0.0448
MEL2	-0.0011	-0.0006	0.0448	0.0448
MEL3	-0.0012	-0.0006	0.0448	0.0447
MEL4	-0.0012	-0.0006	0.0448	0.0448
MEL5	-0.0012	-0.0006	0.0448	0.0448

The foundation of my third and final experiment is the linear regression model that Donald and Newey (2001) used for their simulations studies:

$$y_i = x_i \theta_o + \varepsilon_i; \quad x_i = w_i' \delta + u_i, \quad (1.20)$$

where  $i = 1, 2, \dots, n$ ,  $\theta_o = 0$ , the  $x_i$  and  $\varepsilon_i$  are correlated through  $u_i$ , and

$\text{corr}(\varepsilon_i, u_i) = \rho_{\varepsilon u}$ . The  $w_i = (w'_{(K),i}, w_{K,i})'$  is a  $K \times 1$  vector of instrumental variables.

The data are drawn from a multivariate normal distribution:

$$\begin{pmatrix} w_{(K),i} \\ w_{K,i} \\ \varepsilon_i \\ u_i \end{pmatrix} \sim N \left( 0, \begin{pmatrix} I_{K-1} & 0_{(K-1) \times 1} & 0_{(K-1) \times 1} & 0_{(K-1) \times 1} \\ 0_{1 \times (K-1)} & 1 & \rho_{w\varepsilon} & 0 \\ 0_{1 \times (K-1)} & 0 & 1 & \rho_{\varepsilon u} \\ 0_{1 \times (K-1)} & 0 & \rho_{\varepsilon u} & 1 \end{pmatrix} \right), \quad (1.21)$$

where I fix  $\rho_{\varepsilon u}$  at 0.5. The coefficient vector  $\delta = (\delta_1, \dots, \delta_K)'$  is determined by

$$\delta_j^2 = c(K) \left( 1 - \frac{j}{K+1} \right)^4, \quad (1.22)$$

for  $j = 1, \dots, K$ . The value of  $c(K)$  is chosen given the value of the population  $R^2$  from the regression of  $x$  on  $w$  ( $R^2 = \delta' \delta / (1 + \delta' \delta)$ ), such that  $\delta' \delta = R^2 / (1 - R^2)$ .

If  $\rho_{w\varepsilon} \neq 0$ , the instrument  $w_{K,i}$  is an endogenous variable in the first equation of (1.20) and the model is misspecified. For my experiment I try  $\rho_{w\varepsilon} = 0.4$  and 0. For the case in which  $\rho_{w\varepsilon} = 0$ , the correlation  $\rho_{u\varepsilon}$  controls the degree of the endogeneity of the regressor  $x_i$  in the first equation in (1.20).

The results from the estimation of the misspecified model ( $\rho_{w\varepsilon} = 0.4$ ) are reported in Table 5. The starting value of  $\theta$  is 1 and that of  $\lambda$  is a vector of zeros. Since only one out of five instruments is endogenous, all of the estimators have only small biases. In terms of standard errors and RMSEs, the GMM estimator performs the best. Both the EL-GMM and ETEL estimators have larger standard errors and RMSEs than the GMM estimator, especially when  $n = 100$ . The MEL estimators have much smaller standard



errors and RMSEs than the EL-GMM and ETEL estimators. The GMM estimator outperforms the MEL estimators but only marginally so. The standard errors of all of the eight estimators fall as  $n$  increases. The GMM and MEL estimators have smaller RMSEs than the EL-GMM and ETEL estimators regardless of sample size.

Table 5

Results from Misspecified Linear Regression Model

The data are generated following model (1.20) and  $\theta_o = 0$ . The random vectors  $(w'_{(K),i}, w_{K,i}, \varepsilon_i, u_i)'$  are drawn from the multivariate normal distribution given in (1.21) with  $\rho_{w\varepsilon} = 0.4$ . The coefficient vector  $\delta = (\delta_1, \dots, \delta_K)'$  is determined by (1.22). The value of  $c(K)$  is chosen given the value of the population  $R$ -square from the regression of  $x$  on  $z$  ( $R^2 = \delta'\delta / (1 + \delta'\delta)$ ), such that  $\delta'\delta = R^2 / (1 - R^2)$ . The  $\rho_{\varepsilon u}$  is fixed at 0.5.

	Median Bias	Mean Bias	RMSE	Std
$n = 100, K = 5, R^2 = 0.5, \rho_{\varepsilon u} = 0.5$				
GMM	0.0208	0.0159	0.1190	0.1180
EL-GMM	0.0247	0.0245	0.2323	0.2310
ETEL	0.1397	0.1500	0.3386	0.3035
MEL1	0.0079	0.0029	0.1374	0.1373
MEL2	0.0080	0.0027	0.1379	0.1379
MEL3	0.0091	0.0029	0.1374	0.1374
MEL4	0.0078	0.0021	0.1378	0.1378
MEL5	0.0088	0.0029	0.1371	0.1371
$n = 500, K = 5, R^2 = 0.5, \rho_{\varepsilon u} = 0.5$				
GMM	0.0038	0.0025	0.0531	0.0530
EL-GMM	0.0181	0.0112	0.1205	0.1200
ETEL	0.0899	0.1160	0.2655	0.2389
MEL1	-0.0044	-0.0054	0.0665	0.0663
MEL2	-0.0043	-0.0057	0.0669	0.0667
MEL3	-0.0042	-0.0054	0.0661	0.0659
MEL4	-0.0040	-0.0055	0.0666	0.0663
MEL5	-0.0042	-0.0052	0.0654	0.0652

Tables 6 – 10 report the estimation results from correctly specified models ( $\rho_{we} = 0$ ). Table 6 reports the results from the cases in which the instrumental variables are weakly correlated with the regressor ( $R^2 = 0.1$ ). The results are from the estimation with two different sample sizes ( $n = 100$  and  $500$ ). For all of the two cases, the MEL estimators outperform the EL-GMM and ETEL estimators in terms of median bias, mean bias, standard deviation and RMSE. Among the MEL estimators, the MEL1 estimator is the best one that has the smallest median bias, RMSE and standard error. The MEL estimators have smaller median and mean biases than the GMM estimator, but greater RMSEs.

Somewhat unexpectedly, Table 6 does not provide clear evidence that any of the MEL and ETEL estimators outperform the GMM estimator. Guggenberger (2008) reports similar results. Using a slightly different simulation setup from my model, he compares the EL estimator and the Two-Stage Least Squares (2SLS) estimator. He computes the EL estimator using a grid search method and finds that the 2SLS estimator often outperforms the EL estimator. Although not reported here, I also examine the performances of the GMM, EL-GMM, ETEL and MEL estimators with data generated from his model. The performance of the GMM estimator is similar to the performance of the 2SLS estimator reported in Guggenberger (2008). In contrast, the performance of the EL-GMM estimator is much better than the performance of the EL estimator reported in Guggenberger (2008). I also find that the MEL estimators outperform the EL-GMM estimators. However, the RMSEs of the MEL estimators are greater than the RMSE of the GMM estimator, although they have smaller biases.

Table 6

Results from Correctly Specified Linear Regression Model (Weak Identification with Different Sample Size)

The data are generated following model (1.20) and  $\theta_o = 0$ . The random vectors  $(w'_{(K),i}, w_{K,i}, \varepsilon_i, u_i)'$  are drawn from the multivariate normal distribution given in (1.21) with  $\rho_{w\varepsilon} = 0$ . The coefficient vector  $\delta = (\delta_1, \dots, \delta_K)'$  is determined by (1.22). The value of  $c(K)$  is chosen given the value of the population  $R$ -square from the regression of  $x$  on  $z$  ( $R^2 = \delta'\delta / (1 + \delta'\delta)$ ), such that  $\delta'\delta = R^2 / (1 - R^2)$ . The  $\rho_{\varepsilon u}$  is fixed at 0.5.

	Median Bias	Mean Bias	RMSE	Std
$n = 100, K = 5, R^2 = 0.1$				
GMM	0.1422	0.1219	0.3045	0.2791
EL-GMM	0.1134	0.0566	0.5750	0.5722
ETEL	0.1468	0.0868	0.6064	0.6002
MEL1	0.0900	0.0469	0.4241	0.4216
MEL2	0.1003	0.0451	0.4589	0.4567
MEL3	0.1015	0.0474	0.4676	0.4652
MEL4	0.1006	0.0496	0.4600	0.4573
MEL5	0.1056	0.0462	0.4683	0.4661
$n = 500, K = 5, R^2 = 0.1$				
GMM	0.0334	0.0266	0.1337	0.1310
EL-GMM	0.0349	0.0050	0.2385	0.2384
ETEL	0.0439	0.0217	0.2655	0.2646
MEL1	0.0193	0.0030	0.1749	0.1749
MEL2	0.0322	0.0017	0.2184	0.2184
MEL3	0.0324	0.0008	0.2163	0.2163
MEL4	0.0301	-0.0007	0.2193	0.2193
MEL5	0.0307	0.0022	0.2151	0.2151

The performance of the estimators would depend on the degree of correlations between the instruments and regressor. Table 7 reports the estimation results from the cases of weak and strong instruments ( $R^2 = 0.1$  and 0.5). For both cases, the MEL estimators outperform the EL-GMM and ETEL estimators. Among the MEL estimators,

the MEL1 estimator is the best. The MEL estimators have smaller biases but larger RMSEs or standard errors than the GMM estimator.

Table 7

Results from Correctly Specified Linear Regression Model (Weak and Strong Identification)

The data are generated following model (1.20) and  $\theta_o = 0$ . The random vectors  $(w'_{(K),i}, w_{K,i}, \varepsilon_i, u_i)'$  are drawn from the multivariate normal distribution given in (1.21) with  $\rho_{w\varepsilon} = 0$ . The coefficient vector  $\delta = (\delta_1, \dots, \delta_K)'$  is determined by (1.22). The value of  $c(K)$  is chosen given the value of the population  $R$ -square from the regression of  $x$  on  $z$  ( $R^2 = \delta'\delta / (1 + \delta'\delta)$ ), such that  $\delta'\delta = R^2 / (1 - R^2)$ . The  $\rho_{eu}$  is fixed at 0.5.

	Median Bias	Mean Bias	RMSE	Std
$R^2 = 0.1$ (Weak Identification), $n = 500$ and $K = 5$				
GMM	0.0334	0.0266	0.1337	0.1310
EL-GMM	0.0349	0.0050	0.2385	0.2384
ETEL	0.0439	0.0217	0.2655	0.2646
MEL1	0.0193	0.0030	0.1749	0.1749
MEL2	0.0322	0.0017	0.2184	0.2184
MEL3	0.0324	0.0008	0.2163	0.2163
MEL4	0.0301	-0.0007	0.2193	0.2193
MEL5	0.0307	0.0022	0.2151	0.2151
$R^2 = 0.5$ (Strong Identification), $n = 500$ and $K = 5$				
GMM	0.0044	0.0036	0.0455	0.0453
EL-GMM	0.0058	0.0034	0.0504	0.0503
ETEL	0.0084	0.0063	0.0627	0.0623
MEL1	0.0038	0.0024	0.0470	0.0470
MEL2	0.0053	0.0033	0.0496	0.0495
MEL3	0.0055	0.0034	0.0495	0.0494
MEL4	0.0051	0.0033	0.0496	0.0495
MEL5	0.0054	0.0035	0.0494	0.0493

I now examine how the number of instruments can influence the performances of the estimators. The results are reported in Table 8. For both the cases with  $K = 5$  and 20, the MEL estimators again outperform the EL-GMM and ETEL estimators. The theoretical results of Newey and Smith (2004) predict that the EL-GMM estimator would have smaller mean bias than the GMM estimator, and my simulation results are consistent with this prediction. The MEL estimators also have smaller mean biases than the GMM estimators. Newey and Smith (2004) have shown that the bias-corrected EL-GMM estimator has smaller variance than the GMM estimator. My simulation results are not inconsistent with their finding because I do not make bias-corrections for the MEL estimators. Newey and Smith (2004) also predict that the bias in the EL-GMM estimator would not be sensitive to the number of instruments, but my results show that the bias in the EL-GMM estimator increases with  $K$ . The MEL estimators show the same pattern. However, as  $K$  increases, the RMSEs of the MEL estimators decrease faster than that of the GMM estimator. In terms of both bias and RMSE, the MEL estimators are better alternatives to the GMM estimator when a large number of instrumental variables are used.

Table 8

Results from Correctly Specified Linear Regression Model (Different Numbers of Instruments)

The data are generated following model (1.20) and  $\theta_o = 0$ . The random vectors  $(w'_{(K),i}, w_{K,i}, \varepsilon_i, u_i)'$  are drawn from the multivariate normal distribution given in (1.21) with  $\rho_{w\varepsilon} = 0$ . The coefficient vector  $\delta = (\delta_1, \dots, \delta_K)'$  is determined by (1.22). The value of  $c(K)$  is chosen given the value of the population  $R$ -square from the regression of  $x$  on  $z$  ( $R^2 = \delta'\delta / (1 + \delta'\delta)$ ), such that  $\delta'\delta = R^2 / (1 - R^2)$ . The value of  $R^2$  is fixed at 0.1. The  $\rho_{\varepsilon u}$  is fixed at 0.5.

	Median Bias	Mean Bias	RMSE	Std
$K = 5, n = 500$ and $R^2 = 0.1$				
GMM	0.0334	0.0266	0.1337	0.1310
EL-GMM	0.0349	0.0050	0.2385	0.2384
ETEL	0.0439	0.0217	0.2655	0.2646
MEL1	0.0193	0.0030	0.1749	0.1749
MEL2	0.0322	0.0017	0.2184	0.2184
MEL3	0.0324	0.0008	0.2163	0.2163
MEL4	0.0301	-0.0007	0.2193	0.2193
MEL5	0.0307	0.0022	0.2151	0.2151
$K = 20, n = 500$ and $R^2 = 0.1$				
GMM	0.1309	0.1262	0.1707	0.1149
EL-GMM	0.1159	0.1070	0.1753	0.1388
ETEL	0.1212	0.1110	0.1802	0.1420
MEL1	0.1102	0.1039	0.1629	0.1255
MEL2	0.1134	0.1034	0.1702	0.1352
MEL3	0.1131	0.1042	0.1701	0.1344
MEL4	0.1134	0.1033	0.1701	0.1351
MEL5	0.1125	0.1035	0.1702	0.1351

Table 9 reports the estimation results from the data with two different sample sizes and with strong instruments. For the case with  $n = 1000$ , all estimators show similar performances.

Table 9

Results from Correctly Specified Linear Regression Model (Different Sample Sizes)

The data are generated following model (1.20) and  $\theta_o = 0$ . The random vectors  $(w'_{(K),i}, w_{K,i}, \varepsilon_i, u_i)'$  are drawn from the multivariate normal distribution given in (1.21) with  $\rho_{w\varepsilon} = 0$ . The coefficient vector  $\delta = (\delta_1, \dots, \delta_K)'$  is determined by (1.22). The value of  $c(K)$  is chosen given the value of the population  $R$ -square from the regression of  $x$  on  $z$  ( $R^2 = \delta'\delta / (1 + \delta'\delta)$ ), such that  $\delta'\delta = R^2 / (1 - R^2)$ . The  $\rho_{\varepsilon u}$  is fixed at 0.5.

	Median Bias	Mean Bias	RMSE	Std
$n = 500, K = 5, R^2 = 0.5$				
GMM	0.0334	0.0266	0.1337	0.1310
EL-GMM	0.0349	0.0050	0.2385	0.2384
ETEL	0.0439	0.0217	0.2655	0.2646
MEL1	0.0193	0.0030	0.1749	0.1749
MEL2	0.0322	0.0017	0.2184	0.2184
MEL3	0.0324	0.0008	0.2163	0.2163
MEL4	0.0301	-0.0007	0.2193	0.2193
MEL5	0.0307	0.0022	0.2151	0.2151
$n = 1000, K = 5, R^2 = 0.5$				
GMM	0.0020	0.0018	0.0314	0.0313
EL-GMM	0.0036	0.0027	0.0344	0.0343
ETEL	0.0056	0.0048	0.0419	0.0417
MEL1	0.0018	0.0015	0.0325	0.0325
MEL2	0.0038	0.0026	0.0342	0.0341
MEL3	0.0037	0.0027	0.0341	0.0340
MEL4	0.0038	0.0027	0.0340	0.0339
MEL5	0.0038	0.0027	0.0342	0.0341

Finally, I consider the cases in which the error terms  $\varepsilon_i$  and  $u_i$  are conditionally heteroskedastic or non-normal. The estimation results reported in Table 10 indicate that my simulation results from the cases with normal errors are robust to error distributions. The MEL estimators outperform the EL-GMM and ETEL estimators. The MEL estimators have smaller biases and higher RMSEs than the GMM estimator. Among the MEL estimators, the MEL1 estimator is the best.

My simulation results can be summarized as follows. First, the MEL estimators are better alternatives to the EL-GMM and ETEL estimators. For misspecified models, the finite-sample performances of the MEL estimators are consistent with the prediction that they are root- $n$  consistent, and for both misspecified and correctly specified models, they outperform the EL-GMM and ETEL estimators in terms of both bias and RMSE. Among the MEL estimators, the MEL1 estimator is the best. Second, I do not find decisive evidence that the MEL estimators, as well as the EL-GMM and ETEL estimators, outperform the GMM estimator. Consistent with the theoretical result of Newey and Smith (2004), the MEL estimators have smaller biases than the GMM estimators. However, they often have larger standard errors than the GMM estimator, and their biases may need to be corrected to have smaller standard errors. Third, there is some evidence that the MEL estimators would be good alternatives to the GMM estimators for nonlinear models or linear regression models with a large number of instrumental variables.



Table 10

## Results from Correctly Specified Linear Regression Model (Different Error Distributions)

The data are generated with  $y_i = x_i\theta_o + \varepsilon_i$  and  $x_i = w_i'\delta + u_i$ , where  $i = 1, 2, \dots, n$ , and  $\theta_o = 0$ . The instruments  $w_i$  are drawn from the a multivariate normal distribution  $N(0, I_K)$  and independent of the error vectors  $(\varepsilon_i, u_i)$ . Heteroskedastic errors are drawn in the following way: generate random errors from a bivariate normal distribution with  $\rho_{\varepsilon u} = 0.5$ , multiply the  $i^{th}$  error vector by  $\sqrt{|w_i'\delta| / \text{var}(w_i'\delta)}$ . For the case with  $\chi^2(3)$  errors, the error vectors  $(\varepsilon_i, u_i)'$  are standardized with  $\text{corr}(\varepsilon_i, u_i) = 0.5$ . The coefficient vector  $\delta = (\delta_1, \dots, \delta_K)'$  is determined by  $\delta_j^2 = c(K)(1 - j / (K + 1))^4$ , for  $j = 1, \dots, K$ . The value of  $c(K)$  is chosen given the value of the population  $R$ -square from the regression of  $x$  on  $z$  ( $R^2 = \delta'\delta / (1 + \delta'\delta)$ ), such that  $\delta'\delta = R^2 / (1 - R^2)$ .

	Median Bias	Mean Bias	RMSE	Std
Normal but Heteroskedastic Errors, $n = 500$ and $K = 5$				
2GMM	0.1333	0.1141	0.2893	0.2658
EL-GMM	0.1816	0.1355	0.7373	0.7247
ETEL	0.1945	0.1449	0.7874	0.7740
MEL1	0.1116	0.0549	0.5135	0.5106
MEL2	0.1466	0.0689	0.6456	0.6420
MEL3	0.1473	0.0660	0.6605	0.6573
MEL4	0.1480	0.0764	0.6354	0.6308
MEL5	0.1454	0.0619	0.6508	0.6479
$\chi^2(3)$ Errors, $n = 500$ and $K = 5$				
2GMM	0.0342	0.0273	0.1329	0.1301
EL-GMM	0.0341	0.0031	0.2411	0.2411
ETEL	0.0420	0.0191	0.2775	0.2769
MEL1	0.0216	0.0073	0.1769	0.1768
MEL2	0.0303	0.0032	0.2076	0.2075
MEL3	0.0310	0.0044	0.2084	0.2083
MEL4	0.0299	0.0030	0.2083	0.2083
MEL5	0.0312	0.0064	0.2058	0.2057

## 1.5 Conclusion

This paper proposed a group of alternative estimators to the Empirical Likelihood (EL) estimator, which I refer to as Modified Empirical Likelihood (MEL) estimators. The estimators are obtained by GMM applied to exactly identifying moment functions. For correctly specified models, the estimators share the same higher-order asymptotic properties as the EL estimator. My simulation results also show that the MEL estimators outperform the EL estimator.

They are infinitely many possible MEL estimators, and although I tried many in my simulations, I only report the results from five. I found from the reported and unreported simulations that use of higher than the third order Taylor approximation of the EL estimator does not improve the finite sample properties of the MEL estimators. The MEL estimators are obtained by using two sets of nonlinear moment functions using two different weights. My simulation results indicate that the MEL estimators using different weights for the two sets generally perform better than those which use the same weights.

## CHAPTER 2

### DEMEANED GENERALIZED EMPIRICAL LIKELIHOOD ESTIMATORS

#### 2.1 Introduction

In this chapter, I introduce an alternative group of estimators to the family of Generalized Empirical Likelihood (GEL) estimators. The new estimators are obtained from a constrained maximization problem. Different from the GEL estimators, the new group employs demeaned moment functions in the objective function. Demeaned moment functions are those with means subtracted from the original moment functions. However, in the constraints, the original moment functions rather than the demeaned moment functions are used. I refer to these new estimators as Demeaned Generalized Empirical Likelihood (DGEL) estimators. With this designation, there are more terms in the first-order conditions of the maximization problem with respect to the parameters of interest, relative to the GEL estimators. These additional terms modify the higher-order properties of the estimated parameters.

Although use of demeaned moment functions for the objective function plays an important role in determining the estimators' higher-order properties, it is important to keep the original moment functions in the constraints. If demeaned moment functions were used in the constraints, the optimal problem would have every moment function demeaned. Such a "double-demeaned" estimator would not converge to the true values of parameters because expected values of demeaned moment functions are always zeros

for any value of the model's parameters.<sup>5</sup> Thus, parameters are not identifiable, and there are infinitely many parameter values that satisfy the double-demeaned maximization problem. As a result, demeaned moment functions should not be used in the constraints.

The DGEL estimators can be considered as the GEL versions of the Exponential Tilting Empirical Likelihood (ETEL) estimator. The ETEL estimator, which is proposed by Schennach (2007), is a combination of Empirical Likelihood (EL) estimator and Exponential Tilting (ET) estimator. The ETEL estimator can be written as a constrained maximization problem with demeaned moment functions in the objective function, which coincides with the Demeaned Exponential Tilting (DET) estimator in the DGEL group. In order to document these DGEL estimators, I investigate their asymptotic normality, higher-order biases and efficiency.

If data are symmetrically distributed (i.e., drawn from normal distribution), all estimators in the DGEL family have only one source of bias, and they are higher-order efficient after bias-correction. However, in a more general setting, the DET estimator is the unique member that has these superior higher-order properties. In addition, the DET estimator does not suffer from the same problem that Schennach (2007) has pointed out regarding the EL estimator in a misspecified model with unbounded moment functions.

The DET estimator coincides with the empirical exponential likelihood (EEL) estimator, which belongs to the empirical exponential family (Corcoran, 1998), as well as the ETEL estimator proposed by Schennach (2007), because they share the same first-order-conditions. I refer to the empirical exponential family in Corcoran (1998) as the Generalized EEL family, or GEEL family. The DET estimator is a transformation of

---

<sup>5</sup> The true values of parameter are those satisfying the correctly-specified non-demeaned moment functions.

Corcoran's EEL estimator from the power-divergence version to the GEL version, which involves fewer parameters. Lee and Young (1999) find the EEL estimator has the same higher-order properties as the EL estimator; Schennach (2007) also finds the ETEL estimator has the same higher-order properties as the EL estimator. In addition, I find that the EEL estimator is the best among the Generalized EEL family. Other members in this family either suffer from the misspecification problem pointed out by Schennach (2007), or they are computationally complicated. Thus, not only is the DET estimator the best estimator in the DGEL group, but also its power-divergence counterpart - the EEL estimator - is the best estimator in the GEEL family.

Overall, this paper shows that demeaning the moment functions plays an important role in determining the higher-order properties of estimators. Demeaning reduces the bias and increases the efficiency of the empirical tilting (ET) estimator, while it increases the bias and reduces the efficiency of the EL estimator.

The rest of the chapter is organized as follows. Section 2.2 discusses the DGEL estimators and their higher-order properties; section 2.3 compares the DGEL and GEL estimators; section 2.4 focuses on the DET estimator; and section 2.5 concludes.

## 2.2 Demeaned Generalized Empirical Likelihood Estimators

### 2.2.1 Review of Related Estimators

Let  $z_i$  ( $i = 1, \dots, n$ ) be the independently and identically distributed (*i.i.d.*) observations on a random vector  $z$ ; let  $\theta$  be a  $q \times 1$  parameter vector in a set  $\Theta \subset \mathbb{R}^q$ ;

and  $g_i(\theta) = g(z_i, \theta)$  is a  $p \times 1$  vector of functions with  $p > q$ . I begin by assuming that  $E[g_i(\theta)] = 0$  if and only if  $\theta = \theta_o$ . I also use the following notation:

$$G_i(\theta) = \frac{\partial g_i(\theta)}{\partial \theta'}; \hat{G}(\theta) = n^{-1} \sum_{i=1}^n G_i(\theta); G = E(G_i(\theta_o));$$

$$\Omega_i(\theta) = g_i(\theta)g_i(\theta)'; \hat{\Omega}(\theta) = n^{-1} \sum_{i=1}^n \Omega_i(\theta); \Omega = E(\Omega_i(\theta_o)).$$

Let  $\rho(v)$  be a strictly concave function on an open interval  $V$  containing zero, and  $\lambda$  be a  $p \times 1$  vector of parameters.

The Generalized Empirical Likelihood (GEL) estimator proposed by Newey and Smith (2004) and the Exponential Tilting Empirical Likelihood estimators (ETEL) proposed by Schennach (2007) are reviewed. A GEL estimator is the solution to the following saddle point problem:

$$\hat{\theta}_{GEL} = \arg \min_{\theta \in \Theta} \sup_{\lambda \in \Lambda_n(\theta)} n^{-1} \sum_{i=1}^n \rho(\lambda' g_i(\theta)),$$

where  $\rho(v) = -(1 + \gamma v)^{(\gamma+1)/\gamma} / (\gamma + 1)$  with  $\gamma \in (-\infty, \infty)$ , and

$\Lambda_n(\theta) = \{\lambda \mid \lambda' g_i(\theta) \in V, i = 1, \dots, n\}$ .<sup>6</sup> The Empirical likelihood (EL), Exponential Tilting (ET), and Continuous Updating (CU) have the objective functions  $\rho(v) = \ln(1 - v)$ ,  $\rho(v) = -e^v$  and  $\rho(v) = -(1 + v)^2 / 2$ , respectively. The GEL estimators can also be written as a constrained minimization problem (Imbens, Spady and Johnson, 1998; Schennach, 2007),

$$\min_{\theta} n^{-1} \sum_{i=1}^n \rho(\lambda' g_i(\theta)),$$

with  $\lambda$  a function of  $\theta$  satisfying

---

<sup>6</sup> The parameter  $\lambda$  is the Lagrangean multiplier in the power-divergence dual of the GEL estimators.

$$\sum_{i=1}^n \rho_1(\lambda' g_i(\theta)) g_i(\theta) = 0. \quad (2.1)$$

The ETEL estimator (Schennach, 2007) or the EEL estimator (Corcoran, 1998) is a combination of the EL and ET estimators in the Cressie-Read power-divergence form,

$$\hat{\theta}_{ETEL} = \arg \max_{\theta} \left\{ n^{-1} \sum_{i=1}^n \ln(n \hat{\pi}_i(\theta)) \right\},$$

where  $\hat{\pi}(\theta)$  is the solution to

$$\begin{aligned} & \min_{\{\pi_i\}_{i=1}^n} n^{-1} \sum_{i=1}^n n \pi_i \ln(n \pi_i), \\ \text{s.t. } & \sum_{i=1}^n \pi_i g_i(\theta) = 0 \quad \text{and} \quad \sum_{i=1}^n \pi_i = 0. \end{aligned}$$

The ETEL estimator can also be written as a constrained maximization problem in the GEL version (Schennach, 2007),

$$\hat{\theta}_{ETEL} = \arg \max_{\theta \in \Theta} - \sum_{i=1}^n \exp\left(\hat{\lambda}(\theta)'(g_i(\theta) - \hat{g}(\theta))\right),$$

with  $\lambda$  a function of  $\theta$  satisfying

$$\sum_{i=1}^n \exp(\lambda' g_i(\theta)) g_i(\theta) = 0.$$

Comparing the ETEL estimator and the ET estimator in the GEL family, the ETEL estimator is obtained from a maximization rather than a minimization problem. In addition, the ETEL estimator has moment functions demeaned in the objective function. These differences improve the higher-order properties of the ET estimator. Thus, in the next section, I propose a group of alternative estimators that have demeaned moment functions in the objective function.

### 2.2.2 Definition of DGEL Estimators

The proposed DGEL estimators, which are obtained by maximizing the objective functions with demeaned moment functions over a set of constraints with un-demeaned original moment functions, can be thought of as the GEL versions of the Exponential Tilting Empirical Likelihood estimators (ETEL). A DGEL estimator is a solution to the following problem:

$$\max_{\theta} n^{-1} \sum_{i=1}^n \rho(\lambda'(g_i(\theta) - \hat{g}(\theta))), \quad (2.2)$$

where  $\hat{g}(\theta) = n^{-1} \sum_{i=1}^n g(z_i, \theta)$ , and  $\lambda$  is a function of  $\theta$  such that

$$\sum_{i=1}^n \rho_1(\lambda' g_i(\theta)) g_i(\theta) = 0. \quad (2.3)$$

The DGEL estimators can take any form of  $\rho(v) = -(1 + \gamma v)^{(\gamma+1)/\gamma} / (\gamma + 1)$  with  $\gamma \in (-\infty, +\infty)$ . The demeaned empirical likelihood estimator (DEL), the demeaned exponential tilting estimator (DET), and the demeaned continuous updating estimator (DCU) have  $\rho(v) = \ln(1 - v)$ ,  $\rho(v) = -e^v$  and  $\rho(v) = -(1 + v)^2 / 2$ , respectively.

Using demeaned moment functions in the objective function and transferring the minimization problem into a maximization problem make the DGEL estimators different from the GEL estimators. The GEL optimization problem (2.1) is a minimization problem, because after transfer it into an unconstrained problem by replacing  $\lambda(\theta)$  that is solved from the constraint into (2.1), the objective function in (2.1) is convex in  $\theta$ . Similar, after replacing  $\lambda(\theta)$  which is solved from the constraint (2.3) into the minimization problem (2.2), the objective function in (2.2) is concave in  $\theta$ . Therefore, the DGEL estimators are the maximum point of the maximization problems. In the



following section I will show that demeaning plays a very important role in determining estimators' higher-order properties.

### 2.2.3 Higher-Order Properties of DGEL Estimators

In this section, I show the asymptotic normality of the DGEL estimators as well as their higher-order bias and efficiency. Assumption 2.1a is listed for the existence of a unique solution for the maximization problem (2.2) and (2.3).

Assumption 2.1a: (i)  $\theta \in \Theta$  is the unique solution to  $E(g(z, \theta)) = 0$ , and  $\Theta$  is compact. (ii)  $g(z, \theta)$  is continuous at each  $\theta \in \Theta$  with probability 1. (iii)  $g(z, \theta)$  is continuous differentiable in  $\theta$  in a neighborhood  $N$  of  $\theta_0$ . (iv)  $\rho(\cdot)$  is twice continuous differentiable in a neighborhood of zero.

Assumption 2.1a (i) is a necessary condition for the optimization problem (2.2) and (2.3) to have a unique solution. The continuous and differentiation properties of  $g(z, \theta)$  and  $\rho(\cdot)$  functions are required for deriving the first order conditions.

For simplicity, let  $g_i = g_i(\theta)$ , and  $G_i = G_i(\theta)$ . The FOCs of the DGEL estimator are

$$\begin{cases} n^{-1} \sum_{i=1}^n \rho_1(\lambda'(g_i - \hat{g})) \left( \frac{\partial \lambda'}{\partial \theta} (g_i - \hat{g}) + \lambda'(G_i - \hat{G}) \right) = 0, \\ n^{-1} \sum_{i=1}^n \rho_1(\lambda' g_i) g_i = 0 \end{cases}, \quad (2.4)$$

where  $\frac{\partial \lambda'}{\partial \theta} = - \frac{\sum_{i=1}^n \rho_2(\lambda' g_i) G_i \lambda' g_i + \sum_{i=1}^n \rho_1(\lambda' g_i) G_i}{\sum_{i=1}^n \rho_2(\lambda' g_i) g_i g_i'}$  and it is derived from (2.3).

Since  $\rho(\cdot)$  is twice continuous and differentiable and  $g(z, \theta)$  is continuous differentiable in  $\theta$ ,  $\hat{\lambda}(\theta)$  derived from (2.3) is also continuous and differentiable in  $\theta$ . Replacing  $\hat{\lambda}(\theta)$  into the objective function (2.2) transforms the constrained maximization problem into an unconstrained one. It is easy to show that the objective function of the unconstrained problem is continuous in  $\theta$ . Moreover, Assumption 2.1a (i) states that  $\theta \in \Theta$  and  $\Theta$  is compact. Therefore, by the Weierstrass Theorem, there exists a unique global solution to this unconstrained maximization problem.

Next, I would like to show that the solution to the FOCs is the maximum point. The second-order conditions (SOCs) of the unconstrained maximization problem with respect to  $\theta$  are checked. Replacing  $\hat{\lambda}(\theta)$  into the objective function of (2.2), the unconstrained optimization problem over  $\theta$  is:

$$\max_{\theta} n^{-1} \sum_{i=1}^n \rho \left( \hat{\lambda}(\theta)' (g_i(\theta) - \hat{g}(\theta)) \right).$$

In the Appendix, I show that at  $\lambda_0 = \mathbf{0}$ , the SOC of the above unconstrained maximization problem with respect to  $\theta$  is negative definite for any  $\theta \in \Theta$ . Therefore, at  $\lambda_0 = 0$ ,  $\beta_0$  is the unique globe maximizer for the optimization problem (2.2) and (2.3). Thus, I have the following theorem.

**Theorem 2.1 (Unique Solution):** If Assumption 2.1a is satisfied, there exists a unique DGEL estimator that maximizes the objective function.

Theorem 2.1 shows that the maximizing problem is well-defined and the DGEL estimator is well-designed. It is now safe to discuss the higher-order properties of the DGEL estimators based on the Taylor expansion of the first-order conditions.

In order to facilitate the description and comparison of the higher-order properties of the DGEL estimators, I follow Schennach (2007) to introduce two groups of auxiliary parameters to eliminate the nonlinear elements in the FOCs.

For simplicity, let  $\dot{\tau} = \rho_1(\lambda'(g_i - \hat{g}))$ . The two auxiliary parameters are  $\tau$  and  $\kappa$ , where  $\tau = \sum_{i=1}^n \dot{\tau} / n$  and  $\kappa = \left[ \sum_{i=1}^n \rho_2(\lambda' g_i) g_i g_i' \right]^{-1} \sum_{i=1}^n (\dot{\tau} - \tau) g_i$ .

Lemma 2.1: The two auxiliary conditions for parameters  $\tau$  and  $\kappa$  are

$$\tau - \sum_{i=1}^n \dot{\tau} / n = 0,$$

$$\sum_{i=1}^n \left( (\tau - \dot{\tau}) g_i + \rho_2(\lambda' g_i) g_i g_i' \kappa \right) = 0.$$

Thus, the augmented FOCs of the DGEL estimator are  $n^{-1} \sum_{i=1}^n m_i(\phi) = 0$ , where

$\phi = (\tau, \kappa, \lambda, \theta)$  and  $m_i(\phi)$  is

$$m_i(\phi) = \begin{bmatrix} \dot{\tau} - \tau \\ \rho_1(\lambda' g_i) g_i \\ (\tau - \dot{\tau}) g_i + \rho_2(\lambda' g_i) g_i g_i' \kappa \\ \left[ \rho_2(\lambda' g_i) G_i' \lambda g_i' + \rho_1(\lambda' g_i) G_i' \right] \kappa + (\tau - \dot{\tau}) G_i' \lambda \end{bmatrix}. \quad (2.5)$$

Some regular assumptions are needed to drive the asymptotic normality of the DGEL estimators. First, the solution to the augmented FOCs should be unique. Second,  $G(z_i, \theta)$  and the higher-order moment of  $g(z_i, \theta)$  should be bounded. Otherwise, the estimators would not converge to the true value. Third,  $\Omega$  should be invertible.

Assumption 2.1b: (i)  $E\left(\sup_{\theta \in \Theta} \|g(z_i, \theta)\|^\alpha\right) < \infty$  for some  $\alpha > 2$ , and

$E\left(\sup_{\theta \in \Theta} \|G(z_i, \theta)\|\right) < \infty$ . (ii)  $\Omega$  is nonsingular and finite, and  $\text{rank}(G)=p$ . (iii)

$\phi^* = (\tau^*, \kappa^*, \lambda^*, \theta^*)'$  is the unique solution to  $m(\phi) = 0$  in  $\text{Int}(\Phi)$ .

With these assumptions, the asymptotic normality can be derived from a mean-value expansion of the FOCs. I have proved the following theorem in the Appendix.

Theorem 2.2 (Asymptotic Normality): Under Assumptions 2.1a and 2.1b, the DGEL estimators have the limiting distribution,

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d N(0, \Sigma),$$

where  $\Sigma = (G' \Omega G)^{-1}$  and  $\Omega = E(g_i(\theta)g_i(\theta)')$ .

Theorem 2.2 implies that every DGEL member has the same asymptotic distribution as those of the GEL and GMM estimators. Thus, in a correctly specified model, when sample size is very large, all estimators, including the GMM, GEL and DGEL estimators, should exhibit similar performance. The following theorem presents the higher-order biases of the DGEL estimators when sample size is finite. First, I list the assumptions needed for sketching the higher-order properties.

Assumption 2.2: (i)  $g(z, \theta)$  is three times continuously differentiable in  $\theta$  in a neighborhood  $N$  of  $\theta_0$ . (ii) There is  $b(z)$  with  $E[b(z_i)^6] < \infty$  such that in a

neighborhood of  $N$  of  $\theta_0$ , all partial derivatives of  $g(z, \theta)$  with respect to  $\theta$  up to order four exist, are bounded by  $b(z_i)$  for all  $z$ , and are Lipschitz in  $\theta$  with prefactor  $b(z_i)$ . (iii)  $\rho(\cdot)$  is four times continuously differentiable in a neighborhood of zero.

Since the higher-order bias and variance terms involve the third derivative of  $g(z, \theta)$  and up to the fourth derivative of  $\rho(\cdot)$ , the differentiable properties (i) and (iii) are required. The bounded assumption (ii) is for terms with up to six moment of  $g(z, \theta)$  and its derivatives.

If Assumptions 2.1a, 2.1b and 2.2 are satisfied, the augmented FOCs can be used to derive the higher-order expansion for the DGEL estimators as the following,

$$\sqrt{n}(\hat{\theta}_{DGEL} - \theta_0) = \tilde{\psi} + \tilde{Q}_1 / \sqrt{n} + \tilde{Q}_2 / n + O_p(n^{-2}). \quad (2.6)$$

Here  $\tilde{\psi}$ ,  $\tilde{Q}_1$  and  $\tilde{Q}_2$  are  $O_p(1)$ . The higher-order bias is the bias of order  $O_p(n^{-1})$ . For any estimator  $\hat{\theta}$  that can be expanded as (2.6), the higher order bias is  $E(\hat{\theta}_{DGEL} - \theta_0) = E(\tilde{Q}_{1,\theta} / n)$ , as the expectation of  $\tilde{\psi}$  is zero.

**Theorem 2.3 (Higher-order Bias):** Under Assumptions 2.1a, 2.1b and 2.2, the DGEL estimator of  $\theta$  has

$$Bias(\hat{\theta}_{DGEL}) = E(\tilde{Q}_{1,\theta} / n) = B_I + (1 + \rho_3(0))B_\Omega \quad (2.7)$$

where  $B_I = HE[G_i(\theta)Hg_i(\theta)]/n - H\Sigma_{i=1}^n E[\partial G/\partial\theta_j]\Sigma e_j/2n$ ,  $B_\Omega = HE[g_i g_i' P g_i]/n$ , and  $H = \Sigma G' \Omega^{-1}$ ,  $P = \Omega^{-1} - \Omega^{-1} G H$ . Any estimator with  $\rho_3(0) = -1$  in the DGEL group has one source of bias  $B_I$ .

Similar to the GEL estimators, the DGEL estimators have the bias terms  $B_I$  and  $B_\Omega$  only. The DGEL estimators do not have the term  $B_G$  because they use estimated efficient weights for the Jacobian matrix of the moment conditions. Therefore, they remove the bias arising from the correlation between the moment conditions and their Jacobian matrix.

The precisions of different estimators can be compared based on their higher-order mean square error (MSE). Since I focus on the higher-order efficiencies of the bias-corrected estimators, I say one estimator is relative efficient to the other if its variance is smaller than that of the other. I now derive the higher-order variance of the DGEL estimators. The equation (2.6) implies that

$$Var(\hat{\theta}_{DGEL}) = \frac{\Sigma}{n} + \frac{\Xi_{DGEL}}{n^2} + o(n^{-1/2}),$$

where  $\Xi_{DGEL} = Var(\tilde{Q}_{1,\theta}) + E(\tilde{\psi}_\theta \tilde{Q}'_{2,\theta}) + E(\tilde{Q}_{2,\theta} \tilde{\psi}'_\theta)$ .

The appendix shows that the DET estimator or any estimator with  $\rho_3(0) = -1$  has a  $\tilde{\psi}$  term within an  $O_p(n^{-1})$  difference from that of EL. This  $O_p(n^{-1})$  difference can only contribute to the  $o_p(n^{-1})$  difference in bias and the  $o_p(n^{-2})$  difference in variance. Thus, it will not influence the higher-order bias and efficiency of the estimator. The  $\tilde{Q}_1$  and  $\tilde{Q}_2$

terms for the DET estimator or any estimator with  $\rho_3(0) = -1$  are also different from those for the EL estimator. However, these differences in  $\tilde{Q}_1$  and  $\tilde{Q}_2$  are uncorrelated to  $\tilde{\psi}$ . Therefore, the DET estimator or any estimator with  $\rho_3(0) = -1$  shares the same  $O_p(n^{-1})$  bias and  $O_p(n^{-2})$  variance with the EL estimators. I obtain the following result:

**Theorem 2.4 (Higher-order Efficiency):** Any estimator with  $\rho_3(0) = -1$  in the DGEL group is higher-order efficient after bias-correction.

As Newey and Smith (2004) note, the EL estimator may not be higher-order efficient among the GEL estimators in terms of mean square error (MSE). Only the bias-corrected EL estimator is efficient among the bias-corrected GEL estimators. Similarly, only the bias-corrected DET estimator is more efficient than other biased-corrected DGEL estimators.

Demeaning plays a key role in changing the higher-order bias and efficiency. In the GEL group, any estimator with  $\rho_3(0) = -2$  has one single bias term  $B_1$ , while in the DGEL group, any estimator with  $\rho_3(0) = -1$  does. Without demeaning in the objective function, the maximization problem of DGEL reduced to the GEL problem (2.1). Therefore, demeaning is the major difference between the DGEL and GEL estimators. Comparing the results in Theorem 2.3 and 2.4 and that in Newey and Smith (2004), I can conclude that demeaning modifies the higher-order properties of the GEL estimators.

Corollary 2.1: If data are symmetrically distributed, every DGEL estimator has one source of bias  $B_I$ , and every bias-corrected DGEL estimator is higher-order efficient. If data are not symmetric, only the DET estimator has one source of bias and the bias-corrected DET estimator is higher-order efficient.

If data are symmetric,  $E[g_j g g'] = 0$ . The appendix shows that the differences in  $\tilde{Q}_1$  and  $\tilde{Q}_2$  between the DGEL and EL estimators are in terms of  $E[g_j g g']$  or uncorrelated to  $\tilde{\phi}_\theta$ . Therefore, all of the DGEL estimators share the same higher-order properties as the EL estimator when data are symmetric.

Corollary 2.2: Using different  $\rho(v)$  functions in the objective function and constraint:  $\rho^B(v)$  in the objective function and  $\rho^A(v)$  in the constraint. Any DGEL estimator using the  $\rho(v)$  functions satisfying  $2 - \rho_3^B(0) + 3\rho_3^A(0) = 0$  has only one source of bias  $B_I$ .

Similar to the MEL estimators, having different  $\rho(v)$  function in the objective function and constraints creates new estimators in the DGEL group. With the choice of  $\rho^A$  and  $\rho^B$  satisfies  $2 - \rho_3^B(0) + 3\rho_3^A(0) = 0$ , the higher-order bias in  $\theta$  is  $B_I$ .



## 2.3 DGEL and GEL Estimators

### 2.3.1 Comparison of DGEL and GEL estimators

Table 11 compares the higher-order properties of the DGEL and GEL estimators. Although they have the same bias terms  $B_I$  and  $B_\Omega$ , the coefficients for the bias term  $B_\Omega$  are different. In the DGEL group, the estimator with the fewest sources of bias is the DET estimator, while in the GEL group it is the EL estimator. For the CU estimator, since it has  $\rho_3(0) = 0$ , the CU and DCU estimators share the same higher-order bias.

The bias-corrected DET estimator is the most efficient one among the DGEL group, while the bias-corrected DEL estimator is not. However, in the GEL group, the bias-corrected EL estimator is higher-order efficient but not the ET estimator. Even though the best one in the DGEL group and the best one in the GEL family have the same higher-order properties, the DET estimator is better than the EL estimator when model is misspecified. Schennach (2007) shows the ETEL estimator is root- $n$  consistent when model is misspecified. Since the DET estimator shares the same FOCs with the ETEL estimator, the DET estimator is also better than the EL estimator in the misspecified models.

Table 11

The Higher-order Properties of the DGEL and GEL Families

$\rho(v)$	DGEL		GEL	
	Bias	Higher-order Efficiency*	Bias	Higher-order Efficiency*
EL $Ln(1-v)$	$B_I - B_\Omega$		$B_I$	Yes
ET $-\exp(v)$	$B_I$	Yes	$B_I + B_\Omega / 2$	
CUE $-(1+v)^2 / 2$	$B_I + B_\Omega$		$B_I + B_\Omega$	
$\rho(v)$ $-(1+\gamma v)^{(\gamma+1)/\gamma} / (\gamma+1)$	$B_I + (1 + \rho_3(0))B_\Omega$		$B_I + \left(1 + \frac{\rho_3(0)}{2}\right)B_\Omega$	

\* Higher-order efficient is the higher-order efficiency after bias-correction.

### 2.3.2 Other Groups of Demeaned Estimators

The group of the Demeaned Generalized Empirical Likelihood estimators involves demeaning in the objective function but not in the constraints. However, there are other groups of estimators that apply demeaned moment functions. The second group can be defined by the following optimization problem:

$$\max_{\theta} n^{-1} \sum_{i=1}^n \rho(\lambda'(g_i(\theta) - \hat{g}(\theta))),$$

$$\text{s.t. } \lambda \text{ satisfies } \sum_{i=1}^n \rho_1(\lambda'(g_i(\theta) - \hat{g}(\theta))) g_i(\theta) = 0.$$

Since  $\hat{g}(\theta)$  and  $\hat{G}(\theta)$  only contribute to the  $O_p(n^{-3/2})$  terms in the bias and  $o_p(n^{-2})$  terms in the variance, the additional demeaning term in the constraints does not change the higher-order properties of the estimators. In other words, the estimators in the second group share the same higher-order properties as the DGEL estimators.

Two other groups of demeaned estimators are defined as the following. The third group is defined as:

$$\max_{\theta} n^{-1} \sum_{i=1}^n \rho(\lambda'(g_i(\theta) - \hat{g}(\theta))),$$

$$\text{s.t. } \lambda \text{ satisfies } \sum_{i=1}^n \rho_1(\lambda' g_i(\theta))(g_i(\theta) - \hat{g}(\theta)) = 0.$$

The fourth group has an additional part of demeaning in the constraints. Thus, I refer to them as double-demeaned estimators,

$$\max_{\theta} n^{-1} \sum_{i=1}^n \rho(\lambda'(g_i(\theta) - \hat{g}(\theta))),$$

$$\text{s.t. } \lambda \text{ satisfies } \sum_{i=1}^n \rho_1(\lambda'(g_i(\theta) - \hat{g}(\theta)))(g_i(\theta) - \hat{g}(\theta)) = 0.$$

Let  $h_i(\theta) = g_i(\theta) - \hat{g}(\theta)$ . The FOCs of the fourth group are

$$\begin{cases} \sum_{i=1}^n \frac{1}{1 - \lambda' h_i(\theta)} H_i(\theta)' \lambda = 0 \\ \sum_{i=1}^n \frac{1}{1 - \lambda' h_i(\theta)} h_i(\theta) = 0 \end{cases},$$

where  $H_i(\theta) = \partial h_i(\theta) / \partial \theta$ . In a correctly specified model, any value of  $\theta$  satisfies

$$E(h_i(\theta)) = 0, \text{ while only } \theta_0 \text{ satisfies } E(g_i(\theta_0)) = 0. \text{ Therefore, there are infinitely many}$$

parameter values at which the demeaned moment functions are satisfied. As a result, this optimization problem cannot identify the parameters of interest. Moreover, in a

misspecified model,  $\hat{\lambda} = 0$  can be a solution even if the moment functions are bounded.

With  $\hat{\lambda} = 0$ , any value of  $\hat{\theta}$  satisfies the optimization problem. Thus, the double-

demeaned estimator  $\hat{\theta}$  does not converge to its pseudo-true value  $\theta^*$  that satisfies

$$E(g_i(\theta^*)) = 0. \text{ Furthermore, there is no solution for } \hat{\theta} \text{ in the minimum distance model,}$$

because  $\theta$  disappears from the optimization problem. Thus, the double-demeaned estimators are not desirable.

The same intuition of comparing the DGEL and the second demeaned groups applies to the comparison of the third and fourth groups. The demeaning part  $\rho_1(\lambda'(g_i(\theta) - \hat{g}(\theta)))$  in the constraints does not influence the higher-order properties. Thus, the third and fourth groups share the same higher-order properties, and these two groups are not desirable because the interested parameters are not identifiable.

## 2.4 Demeaned ET Estimator

The DET estimator coincides with the Empirical Exponential Likelihood (EEL) estimator (Corcoran (2000)) and the Exponential Tilting Empirical Likelihood (ETEL) estimator (Schennach (2007)), in the sense that they share the same FOCs. The EEL estimator is a member in the Generalized EEL (GEEL) family, and a GEEL estimator is a combination of the Empirical Likelihood estimator and any other estimator in the Cressie-Read power-divergence family (Cressie and Read (1984)). The GEEL estimator is a solution to the following optimization problem:

$$\max_{\theta} n^{-1} \sum_{i=1}^n \log(n\pi_i(\theta)), \quad (2.8)$$

where  $\hat{\pi}(\theta)$  is the solution to

$$\begin{aligned} & \min_{\{\pi_i\}_{i=1}^n} n^{-1} \sum_{i=1}^n h(\pi_i), \\ & s.t. \sum_{i=1}^n \pi_i g_i(\theta) = 0, \sum_{i=1}^n \pi_i = 0, \end{aligned} \quad (2.9)$$

and  $h(\pi) = \left( (n\pi)^{\gamma+1} - 1 \right) / \left( \gamma(\gamma + 1) \right)$  with  $\gamma \in (-\infty, +\infty)$ .<sup>7</sup> The EL representation is at  $\gamma = -1$  and  $h(\pi) = -\ln(n\pi)$ , and the ET representation is at  $\gamma = 0$  and  $h(\pi) = n\pi \ln(n\pi)$ . Any member in the power-divergence family can be applied to solve for  $\hat{\pi}(\theta)$ . The EEL estimator is with ET representation in the first step to solve for  $\hat{\pi}(\theta)$ . While the EL estimator minimizes the forward Kullback-Leibler discrepancy between distributions  $\pi_i$  and  $n^{-1}$ , the EEL estimator minimizes the backward Kullback-Leibler discrepancy between them. In addition, the objective functions of the GEEL estimators are corresponding to different likelihood ratio statistics (Corcoran, 1998).

The FOCs of the GEEL estimators are provided in Appendix 2. Here are some remarks on the FOCs. First, all of the GEEL members have the same bias as the EL estimator (see Schennach and Spady (2003))<sup>8</sup>. Second, only if the representation of  $\hat{\pi}(\theta)$  takes the form of those which are for the ET, EL, and CU estimators, the FOCs of the corresponding GEEL estimators can be presented in a simple form similar to those of the DET, DEL and DCU estimators in the DGEL family. Other members in the GEEL family have very complicated FOCs. Thus, the computation of other estimators using FOCs is very difficult. Moreover, solving the original constrained optimization problem rather than solving the FOCs is also difficult. Thus, members with  $\hat{\pi}(\theta)$  solved from ET, EL and CU representations in the first step are computationally much simpler than others.

---

<sup>7</sup>The Cressie-Read power-divergence estimators are obtained from the optimization problem (2.9), which is a minimize problem. However, since the EL estimator has  $h(\pi) = -\ln(n\pi)$ , the optimization problem for the EL estimator is  $\max_{\theta} n^{-1} \sum_{i=1}^n \log(n\pi_i(\theta))$ , *s.t.*  $\sum_{i=1}^n \pi_i g_i(\theta) = 0$ ,

$\sum_{i=1}^n \pi_i = 0$ .

<sup>8</sup> A more detailed proof of bias for CU-GEEL estimator is provided in the appendix.

Third, the GEEL estimators with the EL and CU representations suffer from the same inconsistency problem as the EL estimator in the misspecified model. The GEEL estimator with  $\hat{\pi}(\theta)$  solved from EL is identical to the EL estimator in the GEL family.

The GEEL estimator with CU representation is a solution to:

$$\max_{\theta} n^{-1} \sum_{i=1}^n \log(n\pi_i(\theta))$$

where  $\hat{\pi}(\theta)$  is the solution to

$$\begin{aligned} & \min_{\{\pi_i\}_{i=1}^n} n^{-1} \sum_{i=1}^n (n\pi_i)^2 \\ & \text{s.t. } \sum_{i=1}^n \pi_i g_i(\theta) = 0, \sum_{i=1}^n \pi_i = 0. \end{aligned}$$

Since the solution of  $\hat{\pi}_i$  is  $\hat{\pi}_i(\theta) = (1 + \lambda' g_i(\theta)) / n$ , the optimization problem becomes

$$\max_{\theta} \sum_{i=1}^n \log(1 + \lambda' g_i(\theta))$$

$$\text{with } \hat{\lambda}(\theta) \text{ from } \sum_{i=1}^n (1 + \lambda' g_i(\theta)) g_i(\theta) = 0.$$

It suffers from the same inconsistency problem as the EL estimator when model is misspecified. With unbounded moment functions, the estimated  $\hat{\lambda}$  converges to zero.

Accordingly, there is no  $\hat{\theta}$  that would satisfy  $E[g_i(\theta)] = 0$ . Therefore, the best one in this GEEL group has the ET representation of  $\hat{\pi}(\theta)$ , and it shares the same FOCs of the DET estimator. Not only the DET estimator is the best in the DGEL group, but also its power-divergence counterpart - the EEL estimator - is the best among the GEEL family, in the sense that they have the same superior higher-order properties as the EL estimator in a correctly specified model and are root- $n$  consistent in a misspecified model.

## 2.5 Conclusion

This chapter provides a new group of estimators named Demeaned Generalized Empirical Likelihood (DGEL) estimators. A DGEL estimator is obtained by solving the constrained maximization problem with demeaned moment conditions in the objective function.

I investigate the higher-order bias and efficiency of the DGEL estimators. All members have two sources of bias: the bias from using sample moment functions in the estimation instead of their population moments, and the bias by using an estimated optimal weighting matrix. If data are symmetrically distributed, for example if the sample follows a normal distribution, all members in the DGEL family have only one source of bias, and they are also higher-order efficient after bias-correction. However, in a more general setting, only the Demeaned ET estimator has the same superior higher-order properties as the EL estimator.

Demeaning plays a very important role in determining the higher-order properties of estimators. The best estimator in the GEL group is the EL estimator, which is better than the ET estimator in terms of higher-order bias and efficiency. After demeaning, the higher-order bias and efficiency of the ET estimator improve. The Demeaned ET estimator becomes the best estimator in the DGEL family, and shares the same higher-order bias and efficiency as the EL estimator in the GEL family. Moreover, the DET estimator performs better than the EL estimator when the model is misspecified.

The DET estimator coincides with the Empirical Exponential Likelihood estimator which belongs to the Empirical Exponential family (Corcoran (1998)). My version of it reduces the dimension of the parameters and makes computation easier. In

addition, I find that the EEL estimator is the best among the Generalized EEL family-- other members in this family either suffer from the misspecification problem pointed out by Schennach (2007), or they are computationally difficult. Thus, the DET estimator is the best in the DGEL group, and its power-divergence counterpart - the EEL estimator - is also the best in the GEEL group.

I also discuss other groups of demeaned estimators and their high-order properties. Those discussions illustrate how the demeaning in the objective function and the constraint changes the higher-order properties of the estimators. Results in this chapter support wider application of demeaned estimators and offer econometric guidance to the empiricists who use this method.



## REFERENCES

- Altonji, J.G., and L.M. Segal, 1996, Small-sample bias in GMM estimation of covariance structures, *Journal of Business & Economic Statistics*, 14, 353 – 366.
- Andersen, T.G., and B.E. Sorensen, 1996, GMM estimation of a stochastic volatility model: a Monte Carlo study, *Journal of Business & Economic Statistics*, 14, 328 – 352.
- Baggerly, K.A., 1998, Empirical likelihood as a goodness-of-fit measure, *Biometrika*, 85, 535 – 547.
- Chrisiano, L.J., and W.J. Haan, 1996, Small-sample properties of GMM for business-cycle analysis, *Journal of Business & Economic Statistics*, 14, 309 – 327.
- Clark, T.E., 1996, Small-sample properties of estimators of nonlinear models of covariance structure, *Journal of Business & Economic Statistics*, 14, 367 – 373.
- Corcoran, S.A., 1998, Bartlett adjustment of empirical discrepancy statistics, *Biometrika*, 85, 967-972.
- Corcoran, S.A., 2000, Empirical exponential family likelihood using several moment conditions, *Statistica Sinica*, 10, 545-557.
- Cressie, N., and T. Read, 1984, Multinomial goodness-of-fit tests, *Journal of the Royal Statistical Society Series, B* 46, 440-464.
- Grendar, M., 2009, Empty set problem of maximum empirical likelihood methods, *Electronic Journal of Statistics*, 3, 1542 – 1555.
- Guggenberger, P., and J. Hahn, 2005, Finite sample properties of the two-step empirical likelihood estimator, *Econometric Review*, 24, 247 – 263.
- Guggenberger, P., 2008, Finite sample evidence suggesting a heavy tail problem of the generalized empirical likelihood estimator, *Econometric Review*, 26, 526 – 541.
- Hahn, J., and J. Hausman, 2002, A new specification test for the validity of instrumental variables, *Econometrica*, 70, 163 – 189.
- Hall, A., and A. Inoue, 2003, The large sample behaviour of the generalized method of moments estimator in misspecified models, *Journal of Econometrics*, 114, 361 – 394.
- Hansen, L.P., 1982, Large sample properties of generalized method of moments estimators, *Econometrica*, 50, 1029 – 1054.

- Hansen, L. P., J. Heaton, and A. Yaron, 1996, Finite-sample properties of some alternative GMM estimators, *Journal of Business & Economic Statistics*, 14, 262 – 280.
- Imbens, G.W., 1997, One-step estimators for overidentified generalized method of moments models, *Review of Economic Studies*, 64, 359 – 383.
- Imbens, G.W., 2002, Generalized method of moments and empirical likelihood, *Journal of Business & Economic Statistics*, 20, 493 – 506.
- Imbens, G.W., R.H. Spady and P. Johnson, 1998, Information theoretic approaches to inference in moment condition models." *Econometrica*, 66, 333–357.
- Jagannathan, R., and Z.Wang, 1998, An asymptotic theory for estimating beta-pricing models using cross-sectional regression, *Journal of Finance*, 53, 1285 – 1309.
- Lee, M.S. and G.A. Young, 1999, Nonparametric likelihood ratio confidence intervals, *Biometrika*, 86, 1, 107-118.
- Marsaglia, G., 1965, Ratios of normal variables and ratios of sums of uniform variables, *Journal of the American Statistical Association*, 60, 193 – 204.
- Marsaglia, G., 2006, Ratios of normal variables, *Journal of Statistical Software*, 16, 1 – 10.
- Mittelhammer, R, G. Judge, and R. Schoenberg, 2003, Empirical evidence concerning the finite sample performance of EL-type structural equation estimation and inference methods, Working paper.
- Newey, W.K., and R.J. Smith, 2004, Higher order properties of GMM and generalized empirical likelihood estimators, *Econometrica*, 72, 219 – 255.
- Owen, A., 2001, *Empirical likelihood*, Chapman and Hall/CRC, New York.
- Qin, J., and J. Lawless, 1994, Empirical likelihood and general estimating equations, *Annals of Statistics*, 22, 300 – 325.
- Schennach, S.M., 2007, Point estimation with exponentially tilted empirical likelihood, *The Annals of Statistics*, 35, 634 – 672.
- Schennach, S.M., 2005, Accompanying document to “Point estimation with exponentially tilted empirical likelihood.”
- Schennach, S.M. and R.H. Spady, 2003, Higher-order properties of GEL/EL estimators, working paper.

Smith, R.J., 1997, Alternative semi-parametric likelihood approaches to generalized method of moments estimation, *Economic Journal*, 107, 503 – 519.

APPENDIX I  
PROOF FOR CHAPTER 1

I use the following notations to advance the proofs. For  $\phi_o = (0_{1 \times p}, \theta_o)'$ , let  $g_i = g_i(\theta_o)$ ,  $G_i = G_i(\theta_o)$  and

$$M = E \left[ \frac{\partial m_i(\phi_o)}{\partial \phi'} \right]; M_s = E \left[ \frac{\partial^2 m_i(\phi_o)}{\partial \phi_s \partial \phi'} \right]; M_{st} = E \left[ \frac{\partial^3 m_i(\phi_o)}{\partial \phi_s \partial \phi_t \partial \phi'} \right];$$

$$A_i = \frac{\partial m_i(\phi_o)}{\partial \phi'} - M; B_{i,s} = \frac{\partial^2 m_i(\phi_o)}{\partial \phi_s \partial \phi'} - M_s; \psi_i = -M^{-1} m_i(\phi_o);$$

$$K_i = g_i g_i' - \Omega; R_i = G_i - G;$$

$$\tilde{\psi} = n^{-1/2} \sum_{i=1}^n \psi_i; \tilde{A} = n^{-1/2} \sum_{i=1}^n A_i; \tilde{B}_s = n^{-1/2} \sum_{i=1}^n B_{i,s}; \tilde{g} = n^{-1/2} \sum_{i=1}^n g_i;$$

$$\tilde{K} = n^{-1/2} \sum_{i=1}^n K_i; \tilde{R} = n^{-1/2} \sum_{i=1}^n R_i.$$

The following facts are also useful:

$$PG = 0; P' = P; P\Omega P = P; HG = I; P\Omega H' = 0; H\Omega H' = \Sigma; P\Omega - GH = I.$$

Proof of Theorem 1.2

Here I derive the higher-order bias on a MEL estimator without the condition  $\rho_3^B(0) = -2$ . Let  $\rho_j^l = -\rho_j^l(0)$ , where  $l = A, B$  and  $j = 1, 2, 3, 4$ . Using the Taylor expansion of  $\hat{m}(\hat{\phi}) = 0$  at  $\phi_o$ , I can show

$$\hat{\phi} - \phi_o = \tilde{\psi} / \sqrt{n} + \tilde{Q}_1 / n + \tilde{Q}_2 / n^{3/2} + O_p(n^{-2}), \quad (\text{A1.1})$$

where

$$\tilde{Q}_1 = -M^{-1} \left[ \tilde{A}\tilde{\psi} + \sum_{s=1}^{p+q} \tilde{\psi}_s M_s \tilde{\psi} / 2 \right]; \quad (\text{A1.2})$$

$$\tilde{Q}_2 = -M^{-1} \left[ \tilde{A}\tilde{Q}_1 + \sum_{s=1}^{p+q} (\tilde{\psi}_s M_s \tilde{Q}_1 + \tilde{Q}_{1s} M_s \tilde{\psi} + \tilde{\psi}_s \tilde{B}_s \tilde{\psi}) / 2 + \sum_{s=1}^{p+q} \sum_{k=1}^{p+q} \tilde{\psi}_s \tilde{\psi}_k M_{sk} \tilde{\psi} / 6 \right], \quad (\text{A1.3})$$

and  $M$  is defined in (1.18).

Under Assumption 1.3, Newey and Smith (2004) show that  $\tilde{\psi}$ ,  $\tilde{Q}_1$  and  $\tilde{Q}_2$  have finite first- and second-order moments. My goal here is to find  $E(\tilde{Q}_1)$ . By the form of  $M$  and because  $\lambda_o = 0$ ,

$$\psi_i = - \begin{bmatrix} P & H' \\ H & -\Sigma \end{bmatrix}^{-1} \begin{bmatrix} \rho_1^A g_i \\ \rho_1^B G_i' 0 \end{bmatrix} = - \begin{bmatrix} P g_i \\ H g_i \end{bmatrix}.$$

Thus, I have

$$E(\psi_i \psi_i') = E \left[ \begin{pmatrix} P g_i \\ H g_i \end{pmatrix} (g_i' P' \quad g_i' H') \right] = \begin{bmatrix} P\Omega P' & P\Omega H' \\ H\Omega P' & H\Omega H' \end{bmatrix} = \begin{bmatrix} P & 0 \\ 0 & \Sigma \end{bmatrix}.$$

Using these results I can show:

$$\begin{aligned}
E[A_i \psi_i] &= -E \left[ \begin{bmatrix} g_i g_i' & G_i \\ G_i' & 0 \end{bmatrix} - \begin{bmatrix} \Omega & G \\ G' & 0 \end{bmatrix} \right] \begin{bmatrix} P g_i \\ H g_i \end{bmatrix} \\
&= -E \begin{bmatrix} g_i g_i' P g_i + G_i H g_i \\ G_i' P g_i \end{bmatrix} - E \begin{bmatrix} \Omega P g_i + G H g_i \\ G' P g_i \end{bmatrix} \\
&= -E \begin{bmatrix} g_i g_i' P g_i + G_i H g_i \\ G_i' P g_i \end{bmatrix},
\end{aligned}$$

because  $E(\Omega P g_i + G H g_i) = 0$  and  $G' P = 0$ . Also,

$$\begin{aligned}
&E \left( \sum_{s=1}^{p+q} \tilde{\psi}_s M_s \tilde{\psi}' / 2 \right) \\
&= \sum_{s=1}^{p+q} M_s E(\psi_i \psi_i') e_s / 2 \\
&= \sum_{j=1}^p \begin{bmatrix} \rho_3^A E[g_{i,j} g_i g_i'] & E \left[ \frac{\partial(g_{i,j} g_i)}{\partial \theta'} \right] \\ E \left[ \frac{\partial(g_{i,j} g_i')}{\partial \theta} \right] & E \left[ \frac{\partial G_{i,j,\cdot}}{\partial \theta'} \right] \end{bmatrix} \begin{bmatrix} P \\ 0 \end{bmatrix} e_j / 2 \\
&\quad + \sum_{j=1}^q \begin{bmatrix} E \left[ \frac{\partial(g_i g_i')}{\partial \theta_j} \right] & E \left[ \frac{\partial G_i}{\partial \theta_j} \right] \\ E \left[ \frac{\partial G_i'}{\partial \theta_j} \right] & 0 \end{bmatrix} \begin{bmatrix} 0 \\ \Sigma \end{bmatrix} e_j / 2 \\
&= \begin{bmatrix} \sum_{j=1}^p \rho_3^A E[g_{i,j} g_i g_i'] P e_j / 2 + \sum_{j=1}^q E \left[ \frac{\partial G_i}{\partial \theta_j} \right] \Sigma e_j / 2 \\ \sum_{j=1}^p E \left[ \frac{\partial(g_{i,j} g_i')}{\partial \theta} \right] P e_j / 2 \end{bmatrix},
\end{aligned}$$

where  $g_{i,j}$  is the  $j^{\text{th}}$  moment function in  $g_i$  ( $j = 1, \dots, p$ ),  $\partial G_i / \partial \theta_j = \partial G_i(\theta_0) / \partial \theta_j$  ( $j = 1, \dots, q$ ),  $G_{i,j,\cdot}$  is the  $j^{\text{th}}$  row of  $G_i$  ( $j = 1, \dots, p$ ), and

$$\frac{\partial(g_{i,j} g_i)}{\partial \theta'} = \frac{\partial(g_{i,j}(\theta) g_i(\theta))}{\partial \theta'} \Big|_{\theta=\theta_0}; \quad \frac{\partial(g_i g_i')}{\partial \theta_j} = \frac{\partial(g_i(\theta) g_i'(\theta))}{\partial \theta_j} \Big|_{\theta=\theta_0}.$$

Since  $E(\tilde{Q}_1) = -M^{-1} \left[ E[A_i \psi_i] + \sum_{s=1}^{p+q} \tilde{\psi}_s M_s \tilde{\psi}' / 2 \right]$  and partition  $\tilde{Q}_1 = (\tilde{Q}'_{1,\lambda}, \tilde{Q}'_{1,\theta})'$  corresponding to the dimensions of  $\lambda$  and  $\theta$ ,

$$\begin{aligned}
E[\tilde{Q}_{1,\theta}] &= -\begin{bmatrix} 0 \\ I \end{bmatrix}' M^{-1} \left[ -E \begin{bmatrix} g_i g_i' P g_i + G_i H g_i \\ G_i' P g_i \end{bmatrix} + E \begin{bmatrix} \left( \sum_{j=1}^p \rho_3^A E[g_{i,j} g_i g_i'] P e_j / 2 \right) \\ + \sum_{j=1}^q E \left[ \frac{\partial G_i}{\partial \theta_j} \right] \Sigma e_j / 2 \\ \sum_{j=1}^p E \left[ \frac{\partial (g_{i,j} g_i')}{\partial \theta} \right] P e_j / 2 \end{bmatrix} \right] \\
&= -H \left[ -E[G_i H g_i] + \sum_{j=1}^q E \left[ \frac{\partial G_i}{\partial \theta_j} \right] \Sigma e_j / 2 + (1 - \rho_3^A / 2) E[g_i g_i' P g_i] \right] \\
&\quad + \Sigma \left[ -E[G_i' P g_i] + E \left[ \sum_{j=1}^p (E[G_{i,j} g_i'] + E[G_i' g_j]) P e_j / 2 \right] \right].
\end{aligned}$$

As  $\sum_{j=1}^p E(g_{i,j} g_i g_i') P e_j = E(g_i g_i' P g_i)$  and  $\sum_{j=1}^p (E[G_{i,j} g_i'] + E[G_i' g_j]) P e_j / 2 = E(G_i' P g_i)$ ,

I have

$$\begin{aligned}
E[\tilde{Q}_{1,\theta} / n] &= -H \begin{bmatrix} -E[G_i H g_i] + \sum_{j=1}^q E \left[ \frac{\partial G_i}{\partial \theta_j} \right] \Sigma e_j / 2 \\ -\sum_{j=1}^q (1 - \rho_3^A / 2) E[g_i g_i' P g_i] \end{bmatrix} / n \\
&= B_I + (1 - \rho_3^A / 2) B_\Omega.
\end{aligned}$$

All MEL members have  $\rho_3^A = 2$ , thus  $\text{Bias}(\hat{\theta}_{MEL}) = E(\tilde{Q}_{1,\theta} / n) = B_I$ .

### Proof of Theorem 1.3

By (A1.1), the variance matrix of a MEL estimator ( $\hat{\phi}$ ) is given:

$$\text{Var}(\hat{\phi}) = \frac{\text{Var}(\tilde{\psi})}{n} + \frac{\text{Var}(\tilde{Q}_1)}{n^2} + \frac{E(\tilde{\psi} \tilde{Q}_2') + E(\tilde{Q}_2 \tilde{\psi}')}{n^2} + o(1).$$

Partition  $\tilde{Q}_2 = (\tilde{Q}_{2,\lambda}', \tilde{Q}_{2,\theta}')'$  and  $\tilde{\psi} = (\tilde{\psi}_\lambda', \tilde{\psi}_\theta')'$  similarly to  $\tilde{Q}_1$ . Then,

$$\text{Var}(\hat{\theta}) = \frac{\text{Var}(\tilde{\psi}_\theta)}{n} + \frac{\text{Var}(\tilde{Q}_{1,\theta})}{n^2} + \frac{E(\tilde{\psi}_\theta \tilde{Q}_{2,\theta}') + E(\tilde{Q}_{2,\theta} \tilde{\psi}_\theta')}{n^2} + o(1),$$

where  $\text{Var}(\tilde{\psi}_\theta) = \Sigma$ . Under Assumption 1.4, the EL-MM estimator is a MEL estimator using  $\rho^A(v) = \rho^B(v) = \ln(1-v)$ . Thus, I will complete the proof by showing that all of the MEL estimators have the same  $\tilde{Q}_{1,\theta}$  and  $\tilde{Q}_{2,\theta}$  functions.

Define  $\bar{P} = P \tilde{g} \tilde{g}' P'$ ;  $\bar{0} = P \tilde{g} \tilde{g}' H'$ ;  $\bar{\Sigma} = H \tilde{g} \tilde{g}' H'$ . By definition,

$$\tilde{A} = n^{-1/2} \sum_{i=1}^n A_i = \begin{bmatrix} \tilde{K} & \tilde{R} \\ \tilde{R}' & 0 \end{bmatrix}.$$

Also,

$$\begin{aligned}
M_h &= \begin{bmatrix} \rho_3^A E[g_{i,h} g_i g'_i] & E\left[\frac{\partial(g_{i,h} g_i)}{\partial \theta'}\right] \\ E\left[\frac{\partial(g_{i,h} g'_i)}{\partial \theta}\right] & E\left[\frac{\partial G'_{i,h,\cdot}}{\partial \theta}\right] \end{bmatrix}; M_{p+l} = \begin{bmatrix} E\left[\frac{\partial(g_i g'_i)}{\partial \theta_l}\right] & E\left[\frac{\partial G_i}{\partial \theta_l}\right] \\ E\left[\frac{\partial G'_i}{\partial \theta_l}\right] & 0 \end{bmatrix}; \\
\tilde{B}_h &= n^{-1/2} \sum_{i=1}^n \begin{bmatrix} \rho_3^A (g_{i,h} g_i g'_i - E[g_{i,h} g_i g'_i]) & \frac{\partial(g_{i,h} g_i)}{\partial \theta'} - E\left[\frac{\partial(g_{i,h} g_i)}{\partial \theta'}\right] \\ \frac{\partial(g_{i,h} g'_i)}{\partial \theta} - E\left[\frac{\partial(g_{i,h} g'_i)}{\partial \theta}\right] & \frac{\partial G'_{i,h,\cdot}}{\partial \theta'} - E\left[\frac{\partial G'_{i,h,\cdot}}{\partial \theta'}\right] \end{bmatrix}; \\
\tilde{B}_{p+l} &= n^{-1/2} \sum_{i=1}^n \begin{bmatrix} \frac{\partial(g_i g'_i)}{\partial \theta_l} - E\left[\frac{\partial(g_i g'_i)}{\partial \theta_l}\right] & \frac{\partial G_i}{\partial \theta_l} - E\left[\frac{\partial G_i}{\partial \theta_l}\right] \\ \frac{\partial G'_i}{\partial \theta_l} - E\left[\frac{\partial G'_i}{\partial \theta_l}\right] & 0 \end{bmatrix},
\end{aligned}$$

where  $h = 1, \dots, p, l = 1, \dots, q$ .

I now derive the detailed form of  $\tilde{Q}_1$  given in (A1.2). Notice that,

$$\tilde{A}\tilde{\psi} = - \begin{bmatrix} \tilde{K} & \tilde{R} \\ \tilde{R}' & 0 \end{bmatrix} \begin{bmatrix} P\tilde{g} \\ H\tilde{g} \end{bmatrix} = - \begin{bmatrix} \tilde{K}P\tilde{g} + \tilde{R}H\tilde{g} \\ \tilde{R}'H\tilde{g} \end{bmatrix},$$

and

$$\begin{aligned}
& \sum_{s=1}^{p+q} \tilde{\psi}'_s M_s \tilde{\psi} / 2 \\
&= \sum_{s=1}^{p+q} M_s (\tilde{\psi} \tilde{\psi}') e_s / 2 \\
&= \begin{bmatrix} \sum_{j=1}^p \rho_3^A E[g_{i,j} g_i g'_i] \bar{P} e_j / 2 + \sum_{j=1}^p E\left[\frac{\partial g_{i,j} g_i}{\partial \theta'}\right] \bar{0} e_j / 2 \\ \sum_{j=1}^p E\left[\frac{\partial g_{i,j} g'_i}{\partial \theta}\right] \bar{P} e_j / 2 + \sum_{j=1}^p E\left(\frac{\partial G'_{i,j,\cdot}}{\partial \theta'}\right) \bar{0} e_j / 2 \\ + \begin{bmatrix} \sum_{j=1}^q E\left(\frac{\partial g_i g'_i}{\partial \theta_j}\right) \bar{0} e_j / 2 + \sum_{j=1}^q E\left(\frac{\partial G_i}{\partial \theta_j}\right) \bar{\Sigma} e_j / 2 \\ \sum_{j=1}^q E\left(\frac{\partial G'_i}{\partial \theta_j}\right) \bar{0} e_j / 2 \end{bmatrix} \end{bmatrix}.
\end{aligned}$$

Thus, I have

$$\tilde{A}\tilde{\psi} + \sum_{j=1}^{p+q} \tilde{\psi}'_j M_j \tilde{\psi} / 2$$



$$= \begin{pmatrix} \left( \begin{array}{l} \Sigma_{j=1}^p \rho_3^A E [g_{i,j} g_i g_i'] \bar{P} e_j / 2 + \Sigma_{j=1}^p E \left[ \frac{\partial (g_{i,j} g_i)}{\partial \theta'} \right] \bar{0} e_j / 2 \\ + \Sigma_{j=1}^q E \left[ \frac{\partial (g_i g_i')}{\partial \theta_j} \right] \bar{0} e_j / 2 + \Sigma_{j=1}^q E \left[ \frac{\partial G_i}{\partial \theta_j} \right] \bar{\Sigma} e_j / 2 - \tilde{K} P \tilde{g} - \tilde{R} H \tilde{g} \end{array} \right) \\ \left( \begin{array}{l} \Sigma_{j=1}^p E \left[ \frac{\partial (g_{i,j} g_i)}{\partial \theta'} \right] \bar{P} e_j / 2 + \Sigma_{j=1}^p E \left[ \frac{\partial G'_{i,j \cdot}}{\partial \theta'} \right] \bar{0} e_j / 2 \\ + \Sigma_{j=1}^q E \left[ \frac{\partial G'_i}{\partial \theta_j} \right] \bar{0} e_j / 2 - \tilde{R} H \tilde{g} \end{array} \right) \end{pmatrix},$$

which is the same for all of the MEL estimators including the EL-MM estimator. Also,

$$\begin{aligned} \tilde{Q}_{1,\theta} &= - \begin{bmatrix} 0 \\ I \end{bmatrix}' M^{-1} \left[ \tilde{A} \tilde{\psi} + \Sigma_{s=1}^{p+q} \tilde{\psi}_s M_s \tilde{\psi} / 2 \right] \\ &= - \begin{bmatrix} H & -\Sigma \end{bmatrix} \left[ \tilde{A} \tilde{\psi} + \Sigma_{s=1}^{p+q} \tilde{\psi}_s M_s \tilde{\psi} / 2 \right] \\ &= - \begin{pmatrix} H \left( \begin{array}{l} \Sigma_{j=1}^p \rho_3^A E [g_{i,j} g_i g_i'] \bar{P} e_j / 2 + \Sigma_{j=1}^p E \left[ \frac{\partial (g_{i,j} g_i)}{\partial \theta'} \right] \bar{0} e_j / 2 \\ + \Sigma_{j=1}^q E \left[ \frac{\partial (g_i g_i')}{\partial \theta_j} \right] \bar{0} e_j / 2 + \Sigma_{j=1}^q E \left[ \frac{\partial G_i}{\partial \theta_j} \right] \bar{\Sigma} e_j / 2 - \tilde{K} P \tilde{g} - \tilde{R} H \tilde{g} \end{array} \right) \\ -\Sigma \left( \begin{array}{l} \Sigma_{j=1}^p E \left[ \frac{\partial (g_{i,j} g_i)}{\partial \theta'} \right] \bar{P} e_j / 2 + \Sigma_{j=1}^p E \left[ \frac{\partial G'_{i,j \cdot}}{\partial \theta'} \right] \bar{0} e_j / 2 \\ + \Sigma_{j=1}^q E \left[ \frac{\partial G'_i}{\partial \theta_j} \right] \bar{0} e_j / 2 - \tilde{R} H \tilde{g} \end{array} \right) \end{pmatrix}. \end{aligned}$$

Therefore, the  $\tilde{Q}_{1,\theta}$  is the same for all of the MEL estimators and EL estimator.

I now consider the form of  $\tilde{Q}_2$ :

$$\begin{aligned} \tilde{Q}_2 &= -M^{-1} \begin{bmatrix} \tilde{A} \tilde{Q}_1 + \Sigma_{s=1}^{p+q} \left( \tilde{\psi}_s M_s \tilde{Q}_1 + \tilde{Q}_{1,s} M_s \tilde{\psi} + \tilde{\psi}_s \tilde{B}_s \tilde{\psi} \right) / 2 \\ + \Sigma_{s=1}^{p+q} \Sigma_{l=1}^{p+q} \tilde{\psi}_s \tilde{\psi}_l M_{sk} \tilde{\psi} / 6 \end{bmatrix} \\ &= -M^{-1} (I + II + III + IV + V). \end{aligned} \tag{A1.4}$$

First,

$$I = \tilde{A} \tilde{Q}_1 = - \begin{bmatrix} \tilde{K} & \tilde{R} \\ \tilde{R}' & 0 \end{bmatrix} \begin{bmatrix} \tilde{Q}_{1,\lambda} \\ \tilde{Q}_{1,\theta} \end{bmatrix} = - \begin{bmatrix} \tilde{K} \tilde{Q}_{1,\lambda} + \tilde{R} \tilde{Q}_{1,\theta} \\ \tilde{R}' \tilde{Q}_{1,\lambda} \end{bmatrix}.$$

Second, because

$$\tilde{Q}_1 \tilde{\psi}' = - \begin{bmatrix} \tilde{Q}_{1,\lambda} \tilde{g}' P & \tilde{Q}_{1,\lambda} \tilde{g}' H \\ \tilde{Q}_{1,\theta} \tilde{g}' P & \tilde{Q}_{1,\theta} \tilde{g}' H \end{bmatrix},$$

I have

$$\begin{aligned} II &= \sum_{s=1}^{p+q} \tilde{\psi}_s M_s \tilde{Q}_1 / 2 = \sum_{s=1}^{p+q} M_s \tilde{Q}_1 \tilde{\psi}' e_s / 2 \\ &= -\sum_{j=1}^p \begin{bmatrix} \rho_3^A E[g_{i,j} g_i g_i'] & E\left[\frac{\partial(g_{i,j} g_i)}{\partial \theta'}\right] \\ E\left[\frac{\partial(g_{i,j} g_i')}{\partial \theta}\right] & E\left[\frac{\partial G'_{i,j,\cdot}}{\partial \theta'}\right] \end{bmatrix} \begin{bmatrix} \tilde{Q}_{1,\lambda} \tilde{g}' P \\ \tilde{Q}_{1,\theta} \tilde{g}' P \end{bmatrix} e_j / 2 \\ &\quad -\sum_{j=1}^q \begin{bmatrix} E\left[\frac{\partial(g_i g_i')}{\partial \theta_j}\right] & E\left[\frac{\partial G_i}{\partial \theta_j}\right] \\ E\left[\frac{\partial G'_i}{\partial \theta_j}\right] & 0 \end{bmatrix} \begin{bmatrix} \tilde{Q}_{1,\lambda} \tilde{g}' H \\ \tilde{Q}_{1,\theta} \tilde{g}' H \end{bmatrix} e_j / 2 \\ &= - \left( \begin{array}{l} \left( \sum_{j=1}^p \rho_3 E[g_{i,j} g_i g_i'] \tilde{Q}_{1,\lambda} \tilde{g}' P e_j / 2 + \sum_{j=1}^p E\left[\frac{\partial(g_{i,j} g_i)}{\partial \theta'}\right] \tilde{Q}_{1,\theta} \tilde{g}' P e_j / 2 \right) \\ + \sum_{j=1}^q E\left[\frac{\partial(g_i g_i')}{\partial \theta_j}\right] \tilde{Q}_{1,\lambda} \tilde{g}' H e_j / 2 + \sum_{j=1}^q E\left[\frac{\partial G_i}{\partial \theta_j}\right] \tilde{Q}_{1,\theta} \tilde{g}' H e_j / 2 \\ \left( \sum_{j=1}^p E\left[\frac{\partial(g_{i,j} g_i')}{\partial \theta}\right] \tilde{Q}_{1,\lambda} \tilde{g}' P e_j / 2 + \sum_{j=1}^p E\left[\frac{\partial G'_{i,j,\cdot}}{\partial \theta'}\right] \tilde{Q}_{1,\theta} \tilde{g}' P e_j / 2 \right) \\ + \sum_{j=1}^q E\left[\frac{\partial G'_i}{\partial \theta_j}\right] \tilde{Q}_{1,\lambda} \tilde{g}' H e_j / 2 \end{array} \right) \end{aligned}$$

Third, because

$$\tilde{\psi} \tilde{Q}'_1 = -n^{-1/2} \sum_{i=1}^n \begin{bmatrix} P g_i \\ H g_i \end{bmatrix} \begin{bmatrix} \tilde{Q}'_{1,\lambda} & \tilde{Q}'_{1,\theta} \end{bmatrix} = - \begin{bmatrix} P \tilde{g} \tilde{Q}'_{1,\lambda} & P \tilde{g} \tilde{Q}'_{1,\theta} \\ H \tilde{g} \tilde{Q}'_{1,\lambda} & H \tilde{g} \tilde{Q}'_{1,\theta} \end{bmatrix},$$

I have

$$\begin{aligned} III &= \sum_{s=1}^{p+q} \tilde{Q}_{1,s} M_s \tilde{\psi} / 2 = \sum_{s=1}^{p+q} M_s \tilde{\psi} \tilde{Q}'_1 e_s / 2 \\ &= -\sum_{j=1}^p \begin{bmatrix} \rho_3^A E[g_{i,j} g_i g_i'] & E\left[\frac{\partial(g_{i,j} g_i)}{\partial \theta'}\right] \\ E\left[\frac{\partial(g_{i,j} g_i')}{\partial \theta}\right] & E\left[\frac{\partial G'_{i,j,\cdot}}{\partial \theta'}\right] \end{bmatrix} \begin{bmatrix} P \tilde{g} \tilde{Q}'_{1,\lambda} \\ H \tilde{g} \tilde{Q}'_{1,\lambda} \end{bmatrix} e_j / 2 \end{aligned}$$

$$\begin{aligned}
& -\Sigma_{j=1}^q \begin{bmatrix} E \left[ \frac{\partial(g_i g'_i)}{\partial \theta_j} \right] & E \left[ \frac{\partial G_i}{\partial \theta_j} \right] \\ E \left[ \frac{\partial G'_i}{\partial \theta_j} \right] & 0 \end{bmatrix} \begin{bmatrix} P\tilde{g}\tilde{Q}'_{1,\theta} \\ H\tilde{g}\tilde{Q}'_{1,\theta} \end{bmatrix} e_j / 2 \\
= & - \left[ \begin{array}{l} \left( \Sigma_{j=1}^p \rho_3^A E[g_{i,j} g_i g'_i] P\tilde{g}Q'_{1,\lambda} e_j / 2 + \Sigma_{j=1}^p E \left[ \frac{\partial(g_{i,j} g_i)}{\partial \theta'} \right] H\tilde{g}Q'_{1,\lambda} e_j / 2 \right) \\ + \Sigma_{j=1}^q E \left[ \frac{\partial(g_i g'_i)}{\partial \theta_j} \right] P\tilde{g}Q'_{1,\theta} e_j / 2 + \Sigma_{j=1}^q E \left[ \frac{\partial G_i}{\partial \theta_j} \right] H\tilde{g}Q'_{1,\theta} e_j / 2 \\ \left( \Sigma_{j=1}^p E \left[ \frac{\partial(g_{i,j} g'_i)}{\partial \theta} \right] P\tilde{g}Q'_{1,\lambda} e_j / 2 + \Sigma_{j=1}^p E \left[ \frac{\partial G'_{i,j,\cdot}}{\partial \theta'} \right] H\tilde{g}Q'_{1,\lambda} e_j / 2 \right) \\ + \Sigma_{j=1}^q E \left[ \frac{\partial G'_i}{\partial \theta_j} \right] P\tilde{g}Q'_{1,\theta} e_j / 2 \end{array} \right].
\end{aligned}$$

Fourth,

$$\begin{aligned}
& \Sigma_{s=1}^{p+q} \tilde{\psi}'_s \tilde{B}_s \tilde{\psi}_s / 2 = \Sigma_{s=1}^{p+q} \tilde{B}_s \tilde{\psi} \tilde{\psi}'_s e_s / 2 \\
= & \Sigma_{j=1}^p n^{-1/2} \Sigma_{i=1}^n \begin{bmatrix} \rho_3^A (g_{i,j} g_i g'_i - E[g_{i,j} g_i g'_i]) & \frac{\partial(g_{i,j} g_i)}{\partial \theta'} - E \left[ \frac{\partial(g_{i,j} g_i)}{\partial \theta'} \right] \\ \frac{\partial(g_{i,j} g'_i)}{\partial \theta} - E \left[ \frac{\partial(g_{i,j} g'_i)}{\partial \theta} \right] & \frac{\partial G'_{i,j,\cdot}}{\partial \theta'} - E \left[ \frac{\partial G'_{i,j,\cdot}}{\partial \theta'} \right] \end{bmatrix} \begin{bmatrix} \bar{P} \\ \bar{0} \end{bmatrix} e_j / 2 \\
+ & \Sigma_{j=1}^p n^{-1/2} \Sigma_{i=1}^n \begin{bmatrix} \frac{\partial(g_i g'_i)}{\partial \theta_j} - E \left[ \frac{\partial(g_i g'_i)}{\partial \theta_j} \right] & \frac{\partial G_i}{\partial \theta_j} - E \left[ \frac{\partial G_i}{\partial \theta_j} \right] \\ \frac{\partial G'_i}{\partial \theta_j} - E \left[ \frac{\partial G'_i}{\partial \theta_j} \right] & 0 \end{bmatrix} \begin{bmatrix} \bar{0} \\ \bar{\Sigma} \end{bmatrix} / 2
\end{aligned}$$

$$\begin{aligned}
& \left[ \begin{aligned}
& \sum_{j=1}^p \rho_3^A (g_{i,j} g_i g_i' - E[g_{i,j} g_i g_i']) \bar{P} e_j / 2 \\
& + \sum_{j=1}^p \left( \frac{\partial (g_{i,j} g_i)}{\partial \theta'} - E \left[ \frac{\partial (g_{i,j} g_i)}{\partial \theta'} \right] \right) \bar{0} e_j / 2 \\
& + \sum_{j=1}^q \left( \frac{\partial (g_i g_i')}{\partial \theta_j} - E \left[ \frac{\partial (g_i g_i')}{\partial \theta_j} \right] \right) \bar{0} e_j / 2 \\
& + \sum_{j=1}^q \left( \frac{\partial G_i}{\partial \theta_j} - E \left[ \frac{\partial G_i}{\partial \theta_j} \right] \right) \bar{\Sigma} e_j / 2
\end{aligned} \right] \\
= n^{-1/2} \sum_{i=1}^n & \left[ \begin{aligned}
& \sum_{j=1}^p \left( \frac{\partial (g_{i,j} g_i')}{\partial \theta} - E \left[ \frac{\partial (g_{i,j} g_i')}{\partial \theta} \right] \right) \bar{P} e_j / 2 \\
& + \sum_{j=1}^p \left( \frac{\partial G'_{i,j \cdot}}{\partial \theta'} - E \left[ \frac{\partial G'_{i,j \cdot}}{\partial \theta'} \right] \right) \bar{0} e_j / 2 \\
& + \sum_{j=1}^q \left( \frac{\partial G'_i}{\partial \theta_j} - E \left[ \frac{\partial G'_i}{\partial \theta_j} \right] \right) \bar{0} e_j / 2
\end{aligned} \right].
\end{aligned}$$

Thus, the terms *I*, *II*, *III* and *IV* are the same for all of the MEL estimators and EL estimator.

Finally, I derive the detailed form of *V* in (A1.4). For  $M_{sk}$  ( $s, k = 1, \dots, p+q$ ),

$$\begin{aligned}
M_{h,l} &= \begin{bmatrix} \rho_4^A E(g_{i,h} g_{i,l} g_i g_i') & \rho_3^A E \left( \frac{\partial (g_{i,h} g_{i,l} g_i)}{\partial \theta'} \right) \\ \rho_3^B E \left( \frac{\partial (g_{i,h} g_{i,l} g_i')}{\partial \theta} \right) & E \left( \frac{\partial (g_{i,h} g_{i,l})}{\partial \theta \partial \theta'} \right) \end{bmatrix}, h, l = 1, \dots, p; \\
M_{h,p+l} &= \begin{bmatrix} \rho_3^A E \left( \frac{\partial (g_{i,h} g_i g_i')}{\partial \theta_l} \right) & E \left( \frac{\partial^2 (g_{i,h} g_i)}{\partial \theta' \partial \theta_l} \right) \\ E \left( \frac{\partial^2 (g_{i,h} g_i')}{\partial \theta \partial \theta_l} \right) & E \left( \frac{\partial^2 G'_{i,h \cdot}}{\partial \theta' \partial \theta_l} \right) \end{bmatrix}, h = 1, \dots, p, l = 1, \dots, q; \\
M_{p+h,p+l} &= \begin{bmatrix} E \left( \frac{\partial^2 (g_i g_i')}{\partial \theta_h \partial \theta_l} \right) & E \left( \frac{\partial^2 G_i}{\partial \theta_h \partial \theta_l} \right) \\ E \left( \frac{\partial^2 G'_i}{\partial \theta_h \partial \theta_l} \right) & 0 \end{bmatrix}, h, l = 1, \dots, q.
\end{aligned}$$

Thus,  $\sum_{s=1}^{p+q} \sum_{k=1}^{p+q} \tilde{\psi}_s \tilde{\psi}_k M_{sk} \tilde{\psi}$

$$\begin{aligned}
&= -\sum_{h=1}^p \sum_{l=1}^p \tilde{\psi}_h \tilde{\psi}_l \begin{bmatrix} \rho_4^A E(g_{i,h} g_{i,l} g_i g_i') & \rho_3^A E\left(\frac{\partial(g_{i,h} g_{i,l} g_i)}{\partial \theta'}\right) \\ \rho_3^B E\left(\frac{\partial(g_{i,h} g_{i,l} g_i')}{\partial \theta}\right) & E\left(\frac{\partial(g_{i,h} g_{i,l})}{\partial \theta \theta'}\right) \end{bmatrix} \begin{bmatrix} P\tilde{g} \\ H\tilde{g} \end{bmatrix} \\
&\quad -2\sum_{h=1}^p \sum_{l=1}^q \tilde{\psi}_h \tilde{\psi}_{p+l} \begin{bmatrix} \rho_3^A E\left(\frac{\partial(g_{i,h} g_i g_i')}{\partial \theta_l}\right) & E\left(\frac{\partial^2(g_{i,h} g_i)}{\partial \theta' \partial \theta_l}\right) \\ E\left(\frac{\partial^2(g_{i,h} g_i')}{\partial \theta \theta_l}\right) & E\left(\frac{\partial^2 G'_{i,h \cdot}}{\partial \theta' \partial \theta_l}\right) \end{bmatrix} \begin{bmatrix} P\tilde{g} \\ H\tilde{g} \end{bmatrix} \\
&\quad -\sum_{h=1}^q \sum_{l=1}^q \tilde{\psi}_{p+h} \tilde{\psi}_{p+l} \begin{bmatrix} E\left(\frac{\partial^2(g_i g_i')}{\partial \theta_h \partial \theta_l}\right) & E\left(\frac{\partial^2 G_i}{\partial \theta_h \partial \theta_l}\right) \\ E\left(\frac{\partial^2 G'_i}{\partial \theta_h \partial \theta_l}\right) & 0 \end{bmatrix} \begin{bmatrix} P\tilde{g} \\ H\tilde{g} \end{bmatrix}.
\end{aligned}$$

Because  $\tilde{\psi}_\lambda = P\tilde{g}$  and  $\tilde{\psi}_\theta = H\tilde{g}$ , I have

$$\begin{aligned}
V &= \sum_{s=1}^{p+q} \sum_{k=1}^{p+q} \tilde{\psi}_s \tilde{\psi}_k M_{sk} \tilde{\psi} / 6 \\
&= -\sum_{h=1}^p \sum_{l=1}^p \tilde{\psi}_h \tilde{\psi}_l \begin{bmatrix} \rho_4^A E(g_{i,l} g_{i,h} g_i g_i') & \rho_3^A E\left(\frac{\partial(g_{i,h} g_{i,l} g_i)}{\partial \theta'}\right) \\ \rho_3^B E\left(\frac{\partial(g_{i,l} g_{i,h} g_i')}{\partial \theta}\right) & E\left(\frac{\partial(g_{i,h} g_{i,l})}{\partial \theta \theta'}\right) \end{bmatrix} / 6 \\
&\quad -2\sum_{h=1}^p \sum_{l=1}^q \tilde{\psi}_h \tilde{\psi}_{p+l} \begin{bmatrix} \rho_3^A E\left(\frac{\partial(g_{i,h} g_i g_i')}{\partial \theta_l}\right) & E\left(\frac{\partial^2(g_{i,h} g_i)}{\partial \theta' \partial \theta_l}\right) \\ E\left(\frac{\partial^2(g_{i,h} g_i')}{\partial \theta \theta_l}\right) & E\left(\frac{\partial^2 G'_{i,h \cdot}}{\partial \theta' \partial \theta_l}\right) \end{bmatrix} / 6 \\
&\quad -\sum_{h=1}^q \sum_{l=1}^q \tilde{\psi}_{p+h} \tilde{\psi}_{p+l} \begin{bmatrix} E\left(\frac{\partial^2(g_i g_i')}{\partial \theta_h \partial \theta_l}\right) & E\left(\frac{\partial^2 G_i}{\partial \theta_h \partial \theta_l}\right) \\ E\left(\frac{\partial^2 G'_i}{\partial \theta_h \partial \theta_l}\right) & 0 \end{bmatrix} / 6.
\end{aligned}$$

The first term of  $V$  depends on  $\rho_4^A$  and  $\rho_3^B$ , which are not restricted for the MEL estimators. Thus, the first term can differ for different MEL estimators. However, observe that the first term of  $V$  is a function of  $\tilde{\psi}_\lambda = P\tilde{g}$  which is uncorrelated with  $\tilde{\psi}_\theta$ . Thus, the first term does not influence  $E(\tilde{\psi}_\theta \tilde{Q}'_2)$ . The rest of the terms in  $V$  are the same for all of the MEL estimators. Consequently, all of the MEL estimators have the same  $E(\tilde{\psi}_\theta \tilde{Q}'_{2,\theta})$ .

APPENDIX II  
PROOF FOR CHAPTER 2

Proof of Theorem 2.1:

The FOCs of the DGEL estimators are

$$\begin{cases} n^{-1}\sum_{i=1}^n \rho_1(\lambda'(g_i - \hat{g})) \left( \frac{\partial \lambda'}{\partial \theta}(g_i - \hat{g}) + \lambda'(G_i - \hat{G}) \right) = 0 \\ n^{-1}\sum_{i=1}^n \rho_1(\lambda' g_i) g_i = 0 \end{cases}.$$

By taking the derivative with respect to  $\theta$  in the second FOC,

$$\sum_{i=1}^n \rho_2(\lambda' g_i) \left( \frac{\partial \lambda'}{\partial \theta} g_i + \lambda' G_i \right) g_i' + \sum_{i=1}^n \rho_1(\lambda' g_i) G_i = 0.$$

I have

$$\frac{\partial \lambda'}{\partial \theta} = - \left( \sum_{i=1}^n \rho_2(\lambda' g_i) g_i g_i' \right)^{-1} \left( \sum_{i=1}^n \rho_2(\lambda' g_i) G_i \lambda' g_i + \sum_{i=1}^n \rho_1(\lambda' g_i) G_i \right). \quad (\text{A2.1})$$

From the second FOC, I can solve  $\lambda$  as a function of  $\theta$ . Replacing  $\lambda(\theta)$  into the objective function of the maximization problem (2.2) transforms the constrained optimization problem into an unconstrained optimization problem, which is maximizing over  $\theta$ :

$$\max_{\theta} n^{-1} \sum_{i=1}^n \rho(\lambda(\theta)'(g_i(\theta) - \hat{g}(\theta))).$$

The SOC of the objective function with respect to  $\theta$  is

$$\begin{aligned} & n^{-1} \sum_{i=1}^n \rho_2(\lambda'(g_i - \hat{g})) \left( \frac{\partial \lambda'}{\partial \theta}(g_i - \hat{g}) + \lambda'(G_i - \hat{G}) \right) \left( \frac{\partial \lambda'}{\partial \theta}(g_i - \hat{g}) + \lambda'(G_i - \hat{G}) \right)' \\ & + n^{-1} \sum_{i=1}^n \rho_1(\lambda'(g_i - \hat{g})) \left( \frac{\partial^2 \lambda'}{\partial \theta^2}(g_i - \hat{g}) + 2 \frac{\partial \lambda'}{\partial \theta}(G_i - \hat{G}) + \lambda' \frac{\partial(G_i - \hat{G})}{\partial \theta} \right). \end{aligned}$$

At  $\lambda_0 = \mathbf{0}$ , the SOC is

$$SOC = -n^{-1} \sum_{i=1}^n \left( \frac{\partial \lambda'}{\partial \theta}(g_i - \hat{g}) \right) \left( \frac{\partial \lambda'}{\partial \theta}(g_i - \hat{g}) \right)'.$$

Since not all  $g_i$  with  $i=1, \dots, n$ , equal  $\hat{g}$ , the SOC for the maximization problem with respect to  $\theta$  is negative definite. Thus, at  $\lambda_0 = \mathbf{0}$ , this SOC is globally negative definite for any  $\theta \in \Theta$ .

Proof of Lemma 2.1:

The FOCs of the DGEL estimators are

$$\begin{cases} n^{-1}\sum_{i=1}^n \rho_1(\lambda'(g_i - \hat{g})) \left( \frac{\partial \lambda'}{\partial \theta}(g_i - \hat{g}) + \lambda'(G_i - \hat{G}) \right) = 0 \\ n^{-1}\sum_{i=1}^n \rho_1(\lambda' g_i) g_i = 0 \end{cases}.$$



Let  $\dot{\tau} = \rho_1(\lambda'(g_i - \hat{g}))$ . Introduce two auxiliary parameters  $\tau$  and  $\kappa$ , where  $\tau = \sum_{i=1}^n \dot{\tau} / n$  and  $\kappa = \left[ \sum_{i=1}^n \rho_2(\lambda' g_i) g_i g_i' \right]^{-1} \sum_{i=1}^n (\dot{\tau} - \tau) g_i$ . Use (A2.1) and the definition of  $\tau$  and  $\dot{\tau}$ ,  $\partial \lambda' / \partial \theta$  can be rewritten as

$$\frac{\partial \lambda'}{\partial \theta} = - \frac{\left( \sum_{i=1}^n \rho_2(\lambda' g_i) G_i \lambda' g_i + \sum_{i=1}^n \rho_1(\lambda' g_i) G_i \right) \kappa}{\sum_{i=1}^n (\dot{\tau} - \tau) g_i}. \quad (\text{A2.2})$$

Replace (A2.2) into the first FOC, after several steps of algebra, the first FOC becomes

$$\left[ \rho_2(\lambda' g_i) G_i' \lambda g_i' + \rho_1(\lambda' g_i) G_i' \right] \kappa + (\tau - \dot{\tau}) G_i' \lambda = 0.$$

Next, I introduce two additional FOCs to make the system exactly identified,

$$\begin{aligned} \tau - \sum_{i=1}^n \dot{\tau} / n &= 0, \\ \sum_{i=1}^n \left( (\tau - \dot{\tau}) g_i + \rho_2(\lambda' g_i) g_i g_i' \kappa \right) &= 0. \end{aligned}$$

Therefore, the augmented FOCs are

$$m_i(\phi) = \begin{bmatrix} \dot{\tau} - \tau \\ \rho_1(\lambda' g_i) g_i \\ (\tau - \dot{\tau}) g_i + \rho_2(\lambda' g_i) g_i g_i' \kappa \\ \left[ \rho_2(\lambda' g_i) G_i' \lambda g_i' + \rho_1(\lambda' g_i) G_i' \right] \kappa + (\tau - \dot{\tau}) G_i' \lambda \end{bmatrix},$$

and  $\phi = (\tau, \kappa', \lambda', \theta')'$ .

Proof of Theorem 2.2:

Using the mean value expansion of the FOCs at  $\phi_0 = (1, 0, 0, \theta_0)$ ,

$$0 = \begin{pmatrix} 0 \\ \hat{g}(\theta_0) \\ 0 \\ 0 \end{pmatrix} + \bar{M}(\hat{\phi} - \phi_0).$$

Let  $\hat{g}_i = g_i(\hat{\theta})$ ,  $\hat{G}_i = G_i(\hat{\theta})$ ,  $\bar{g}_i = g_i(\bar{\theta})$  and  $\bar{G}_i = G_i(\bar{\theta})$  then

$$\bar{M} = \begin{bmatrix} -1 & 0 & \sum_{i=1}^n \rho_2(\bar{\lambda}'(\hat{g}_i - \hat{g})) \bar{g}_i' / n & 0 \\ 0 & 0 & \sum_{i=1}^n \rho_2(\bar{\lambda}' \hat{g}_i) \bar{g}_i \hat{g}_i' / n & \sum_{i=1}^n \rho_1(\bar{\lambda}' \hat{g}_i) \bar{G}_i' / n \\ \hat{g}_i & \sum_{i=1}^n \rho_2(\hat{\lambda}' \hat{g}_i) \hat{g}_i \hat{g}_i' / n & -\sum_{i=1}^n \rho_2(\hat{\lambda}'(\hat{g}_i - \hat{g})) \hat{g}_i \hat{g}_i' / n & 0 \\ 0 & \sum_{i=1}^n \rho_1(\hat{\lambda}' \hat{g}_i) \hat{G}_i' / n & 0 & 0 \end{bmatrix},$$

where  $\bar{\theta}$  and  $\bar{\lambda}$  are mean values.

I will ignore the  $O_p(n^{-1})$  terms, since it only contributes to  $O_p(n^{-3/2})$  terms to the bias and  $o_p(n^{-2})$  terms to the variance of parameters. Ignoring the  $O_p(n^{-1})$  terms in  $M$ ,

$$M = E\left(\frac{\partial m_i}{\partial \theta}\right) = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 0 & \Omega & G \\ 0 & \Omega & -\Omega & 0 \\ 0 & G' & 0 & 0 \end{bmatrix}, \quad M^{-1} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & P & P & H' \\ 0 & P & P - \Omega^{-1} & H' \\ 0 & H & H & -\Sigma \end{bmatrix}.$$

Since  $\bar{M} \rightarrow M$ , I have

$$\sqrt{n}(\hat{\phi} - \phi_o) = -\bar{M}^{-1} \begin{pmatrix} 0 \\ \hat{g}(\theta_o) \\ 0 \\ 0 \end{pmatrix} = -M^{-1} \begin{pmatrix} 0 \\ \hat{g}(\theta_o) \\ 0 \\ 0 \end{pmatrix} + o_p(1) = - \begin{pmatrix} 0 \\ \sqrt{n}P\hat{g}(\theta_o) \\ \sqrt{n}P\hat{g}(\theta_o) \\ \sqrt{n}H\hat{g}(\theta_o) \end{pmatrix} + o_p(1)$$

where  $\Sigma = (G'\Omega^{-1}G)^{-1}$ ,  $P = \Omega^{-1} - \Omega^{-1}G\Sigma G'\Omega^{-1}$  and  $H = \Sigma G'\Omega^{-1}$ . Using these results, I can show that for correctly specified models,

$$\sqrt{n}(\hat{\phi}_{DGEL} - \phi_o) = \begin{pmatrix} \hat{\tau} - 1 \\ \sqrt{n}\hat{\kappa}_{DGEL} \\ \sqrt{n}\hat{\lambda}_{DGEL} \\ \sqrt{n}(\hat{\theta}_{DGEL} - \theta_o) \end{pmatrix} \rightarrow_d N \left( \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & P & P & 0 \\ 0 & P & P & 0 \\ 0 & 0 & 0 & \Sigma \end{pmatrix} \right).$$

Thus,  $\sqrt{n}(\theta_{DGEL} - \theta_o) \rightarrow_d N(0, \Sigma)$ .

Proof of Theorem 2.3:

Using the Taylor expansion of  $\hat{m}(\hat{\phi}) = 0$  at  $\phi_o$ , I can show

$$\hat{\phi} - \phi_o = \tilde{\psi} / \sqrt{n} + \tilde{Q}_1 / n + \tilde{Q}_2 / n^{3/2} + O_p(n^{-2}), \quad (\text{A2.3})$$

where

$$\tilde{Q}_1 = -M^{-1} \left[ \tilde{A}\tilde{\psi} + \sum_{s=1}^{1+2p+q} \tilde{\psi}_s M_s \tilde{\psi} / 2 \right]; \quad (\text{A2.4})$$

$$\tilde{Q}_2 = -M^{-1} \left[ \tilde{A}\tilde{Q}_1 + \sum_{s=1}^{1+2p+q} (\tilde{\psi}_s M_s \tilde{Q}_1 + \tilde{Q}_{1s} M_s \tilde{\psi} + \tilde{\psi}_s \tilde{B}_s \tilde{\psi}) / 2 \right. \\ \left. + \sum_{s=1}^{1+2p+q} \sum_{k=1}^{1+2p+q} \tilde{\psi}_s \tilde{\psi}_k M_{sk} \tilde{\psi} / 6 \right]. \quad (\text{A2.5})$$

By definition of  $M$  and  $E(g_i \hat{g}') = E(n^{-1} g_i g_i') = O_p(n^{-1})$ ,

$$M = E\left(\frac{\partial m_i(\phi)}{\partial \phi'}\right) = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 0 & \Omega & G \\ 0 & \Omega & -\Omega + O_p(n^{-1}) & 0 \\ 0 & G' & 0 & 0 \end{bmatrix}.$$

I will ignore the  $O_p(n^{-1})$  terms, since it only contributes to  $O_p(n^{-3/2})$  terms to the bias and  $o_p(n^{-2})$  terms to the variance of parameters. Therefore,

$$M^{-1} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & P & P & H' \\ 0 & P & P - \Omega^{-1} & H' \\ 0 & H & H & -\Sigma \end{bmatrix}$$

and

$$\psi_i = - \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & P & P & H' \\ 0 & P & P - \Omega^{-1} & H' \\ 0 & H & H & -\Sigma \end{bmatrix} \begin{bmatrix} \hat{\tau} - \tau \\ \rho_1(\lambda' g_i) g_i \\ (\tau - \hat{\tau}) g_i + \rho_2(\lambda' g_i) g_i g_i' \kappa \\ \left( \begin{array}{l} \rho_2(\lambda'(g_i - \hat{g})) G_i' \lambda g_i' \kappa \\ + \rho_1(\lambda'(g_i - \hat{g})) G_i' \kappa + (\tau - \hat{\tau}) G_i' \lambda \end{array} \right) \end{bmatrix} \Big|_{\phi^*=(1,0,0,\theta^*)} = - \begin{bmatrix} 0 \\ P g_i \\ P g_i \\ H g_i \end{bmatrix}.$$

Let  $\rho_j = -\rho_j(0)$ . I have

$$E(\psi_i \psi_i') = \begin{bmatrix} 0 \\ P g_i \\ P g_i \\ H g_i \end{bmatrix} \begin{bmatrix} 0 & g_i' P' & g_i' P' & g_i' H' \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & P & P & 0 \\ 0 & P & P & 0 \\ 0 & 0 & 0 & \Sigma \end{bmatrix}.$$

Using these results I can show:

$$\begin{aligned} E[A_i \psi_i] &= -E \left[ \begin{bmatrix} -1 & 0 & (g_i - \hat{g})' & 0 \\ 0 & 0 & g_i g_i' & G_i \\ g_i & g_i g_i' & -g_i g_i' + g_i \hat{g}' & 0 \\ 0 & G_i' & 0 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 0 & \Omega & G \\ 0 & \Omega & -\Omega & 0 \\ 0 & G' & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ P g_i \\ P g_i \\ H g_i \end{bmatrix} \right] \\ &= -E \begin{bmatrix} (g_i - \hat{g})' P g_i \\ g_i g_i' P g_i + G_i H g_i \\ 0 \\ G_i' P g_i \end{bmatrix} - E \begin{bmatrix} 0 \\ \Omega P g_i + G H g_i \\ 0 \\ G' P g_i \end{bmatrix} = -E \begin{bmatrix} g_i' P g_i \\ g_i g_i' P g_i + G_i H g_i \\ 0 \\ G_i' P g_i \end{bmatrix}, \end{aligned}$$

because  $E(\Omega P g_i + G H g_i) = 0$  and  $G' P = 0$ .

By definition,

$$M_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & E[G] \\ 0 & 0 & E[G] & 0 \end{bmatrix};$$

$$\begin{aligned}
M_{1+h} &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \rho_3 E[g_{i,j} g_i g'_i] & E\left[\frac{\partial(g_{i,j} g_i)}{\partial \theta'}\right] \\ 0 & 0 & E\left[\frac{\partial(g_{i,j} g'_i)}{\partial \theta}\right] & E\left[\frac{\partial G'_{i,j,\cdot}}{\partial \theta'}\right] \end{bmatrix}; \\
M_{1+p+h} &= \begin{bmatrix} 0 & 0 & \rho_3 E[g_{i,j} g'_i] & 0 \\ 0 & 0 & \rho_3 E[g_{i,j} g_i g'_i] & E\left[\frac{\partial(g_{i,j} g_i)}{\partial \theta'}\right] \\ 0 & \rho_3 E[g_{i,j} g_i g'_i] & -\rho_3 E[g_{i,j} g_i g'_i] & -E\left[\frac{\partial(g_{i,j} g_i)}{\partial \theta'}\right] \\ E[G'_{i,j,\cdot}] & E\left[\frac{\partial(g_{i,j} g'_i)}{\partial \theta}\right] & -E\left[\frac{\partial(g_{i,j} g'_i)}{\partial \theta}\right] & 0 \end{bmatrix}; \\
M_{1+2p+l} &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & E\left[\frac{\partial(g_i g'_i)}{\partial \theta_j}\right] & E\left[\frac{\partial G_i}{\partial \theta_j}\right] \\ E[G_{i,\cdot,j}] & E\left[\frac{\partial(g_i g'_i)}{\partial \theta_j}\right] & -E\left[\frac{\partial(g_i g'_i)}{\partial \theta_j}\right] & 0 \\ 0 & E\left[\frac{\partial G'_i}{\partial \theta_j}\right] & 0 & 0 \end{bmatrix};
\end{aligned}$$

where  $h = 1, \dots, p, l = 1, \dots, q$ . Thus,

$$\begin{aligned}
E\left(\sum_{s=1}^{1+2p+q} \tilde{\psi}_s M_s \tilde{\psi}' / 2\right) &= \sum_{s=1}^{1+2p+q} M_s E(\psi_s \psi'_s) e_s / 2 \\
&= \begin{bmatrix} 0 \\ 0 \\ \sum_{j=1}^p \rho_3 E[g_{i,j} g_i g'_i] P \\ \sum_{j=1}^p E\left[\frac{\partial(g_{i,j} g'_i)}{\partial \theta}\right] P \end{bmatrix} \frac{e_j}{2} + \begin{bmatrix} \sum_{j=1}^p \rho_3 E[g_{i,j} g'_i] P \\ \sum_{j=1}^p \rho_3 E[g_{i,j} g_i g'_i] P \\ 0 \\ 0 \end{bmatrix} \frac{e_j}{2} + \begin{bmatrix} 0 \\ \sum_{j=1}^q E\left[\frac{\partial G_i}{\partial \theta_j}\right] \Sigma \\ 0 \\ 0 \end{bmatrix} \frac{e_j}{2} \\
&= \begin{bmatrix} \sum_{j=1}^p \rho_3 E[g_{i,j} g'_i] P e_j / 2 \\ \sum_{j=1}^p \rho_3 E[g_{i,j} g_i g'_i] P e_j / 2 + \sum_{j=1}^q E[\partial G_i / \partial \theta_j] \Sigma e_j / 2 \\ \sum_{j=1}^p \rho_3 E[g_{i,j} g_i g'_i] P e_j / 2 \\ \sum_{j=1}^p E[\partial(g_{i,j} g'_i) / \partial \theta] P e_j / 2 \end{bmatrix},
\end{aligned}$$

where  $g_{i,j}$  is the  $j^{\text{th}}$  moment function in  $g_i$  ( $j = 1, \dots, p$ ),  $\partial G_i / \partial \theta_j = \partial G_i(\theta_o) / \partial \theta_j$  ( $j = 1, \dots, q$ ),  $G_{i,j}$  is the  $j^{\text{th}}$  row of  $G_i$  ( $j = 1, \dots, p$ ), and

$$\frac{\partial(g_{i,j}g_i)}{\partial \theta'} = \frac{\partial(g_{i,j}(\theta)g_i(\theta))}{\partial \theta'} \Big|_{\theta=\theta_o}; \quad \frac{\partial(g_i g_i')}{\partial \theta_j} = \frac{\partial(g_i(\theta)g_i'(\theta))}{\partial \theta_j} \Big|_{\theta=\theta_o}.$$

Then, I have

$$E(Q_1) = -M^{-1} \left[ E[A_i \psi_i] + \sum_{s=1}^{1+2p+q} E[\psi_s M_s \psi] / 2 \right].$$

Partition  $\tilde{Q}_1 = (\tilde{Q}'_{1,\tau}, \tilde{Q}'_{1,\kappa}, \tilde{Q}'_{1,\lambda}, \tilde{Q}'_{1,\theta})'$  corresponding to the dimensions of  $\tau, \kappa, \lambda$  and  $\theta$ .

Then,

$$\begin{aligned} E[Q_{1,\theta}] &= - \begin{bmatrix} 0 \\ 0 \\ 0 \\ I \end{bmatrix}' M^{-1} \left[ \begin{array}{c} \begin{bmatrix} g_i' P g_i \\ g_i g_i' P g_i + G_i H g_i \\ 0 \\ G_i' P g \end{bmatrix} \\ \begin{bmatrix} \sum_{j=1}^p \rho_3 E[g_{i,j} g_i'] P e_j / 2 \\ \sum_{j=1}^p \rho_3 E[g_{i,j} g_i g_i'] P e_j / 2 + \sum_{j=1}^q E \left[ \frac{\partial G_i}{\partial \theta_j} \right] \Sigma e_j / 2 \\ \sum_{j=1}^p \rho_3 E[g_{i,j} g_i g_i'] P e_j / 2 \\ \sum_{j=1}^p E \left[ \frac{\partial(g_{i,j} g_i')}{\partial \theta} \right] P e_j / 2 \end{bmatrix} \right] \\ &= - [0 \quad H \quad H \quad -\Sigma] \begin{bmatrix} \sum_{j=1}^p \rho_3 E[g_{i,j} g_i'] P e_j / 2 - E[g_i' P g_i] \\ \left( -E[g_i g_i' P g_i + G_i H g_i] + \sum_{j=1}^p \rho_3 E[g_{i,j} g_i g_i'] P e_j / 2 + \right. \\ \left. \sum_{j=1}^q E \left[ \frac{\partial G_i}{\partial \theta_j} \right] \Sigma e_j / 2 \right) \\ \sum_{j=1}^p \rho_3 E[g_{i,j} g_i g_i'] P e_j / 2 \\ -E[G_i' P g_i] + E \left[ \sum_{j=1}^p (E(G'_{i,j} g_i') + E(G'_i g_{i,j})) P e_j / 2 \right] \end{bmatrix} \\ &= -H \begin{bmatrix} -E[G_i H g_i] + \sum_{j=1}^q E \left[ \frac{\partial G_i}{\partial \theta_j} \right] \Sigma e_j / 2 - (1 - \rho_3) E[g_i g_i' P g_i] \\ -E[G_i' P g_i] + E \left[ \sum_{j=1}^p (E(G'_{i,j} g_i') + E(G'_i g_{i,j})) P e_j / 2 \right] \end{bmatrix}. \end{aligned}$$

As  $\sum_{j=1}^p E(g_{i,j} g_i g_i') P e_j = E(g_i g_i' P g_i)$  and  $\sum_{j=1}^p (E[G'_{i,j} g_i'] + E[G'_i g_{i,j}]) P e_j / 2 = E(G'_i P g_i)$ ,

I have

$$\begin{aligned}
E[Q_{1,\theta}] &= H \left[ E[G_i H g_i] - \sum_{j=1}^q E \left[ \frac{\partial G_i}{\partial \theta_j} \right] \Sigma e_j / 2 \right] + (1 - \rho_3) H E[g_i g_i' P g_i] \\
&= B_I + (1 - \rho_3) B_\Omega
\end{aligned}$$

which completes the proof.

Proof of Theorem 2.4:

By (A2.3), the variance matrix of a DGEL estimator ( $\hat{\phi}$ ) is given:

$$\text{Var}(\hat{\phi}) = \frac{\text{Var}(\tilde{\psi})}{n} + \frac{\text{Var}(\tilde{Q}_1)}{n^2} + \frac{E(\tilde{\psi} \tilde{Q}_2') + E(\tilde{Q}_2 \tilde{\psi}')}{n^2} + o(1).$$

Similarly to  $\tilde{Q}_1$ , I have partitions  $\tilde{Q}_2 = (\tilde{Q}_{2,\tau}', \tilde{Q}_{2,\kappa}', \tilde{Q}_{2,\lambda}', \tilde{Q}_{2,\theta}')'$  and  $\tilde{\psi} = (\tilde{\psi}'_\tau, \tilde{\psi}'_\kappa, \tilde{\psi}'_\lambda, \tilde{\psi}'_\theta) = (0, \tilde{g}'P', \tilde{g}'H', \tilde{g}'H')$ . Then,

$$\text{Var}(\hat{\theta}) = \frac{\text{Var}(\tilde{\psi}_\theta)}{n} + \frac{\text{Var}(\tilde{Q}_{1,\theta})}{n^2} + \frac{E(\tilde{\psi}_\theta \tilde{Q}_{2,\theta}') + E(\tilde{Q}_{2,\theta} \tilde{\psi}'_\theta)}{n^2} + o(1).$$

Define  $\bar{P} = P \tilde{g} \tilde{g}' P'$ ;  $\bar{0} = P \tilde{g} \tilde{g}' H'$ ;  $\bar{\Sigma} = H \tilde{g} \tilde{g}' H'$ , where  $E(\bar{P}) = P$ ,  $E(\bar{0}) = 0$ ,  $E(\bar{\Sigma}) = \Sigma$ ,  $E(\bar{g}) = 0$  and  $E(\tilde{g} \tilde{g}') = \Omega$ . By definition,

$$\tilde{\psi} = - \begin{bmatrix} 0 \\ P \tilde{g} \\ P \tilde{g} \\ H \tilde{g} \end{bmatrix}; \quad \tilde{A} = n^{-1/2} \sum_{i=1}^n A_i = \begin{bmatrix} 0 & 0 & \tilde{g}' & 0 \\ 0 & 0 & \tilde{K} & \tilde{R} \\ \tilde{g} & \tilde{K} & -\tilde{K} & 0 \\ 0 & \tilde{R}' & 0 & 0 \end{bmatrix}.$$

I now derive the detailed form of  $\tilde{Q}_1$  given in (A2.4). Notice that

$$\tilde{A} \tilde{\psi} = - \begin{bmatrix} 0 & 0 & \tilde{g}' & 0 \\ 0 & 0 & \tilde{K} & \tilde{R} \\ \tilde{g} & \tilde{K} & -\tilde{K} & 0 \\ 0 & \tilde{R}' & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ P \tilde{g} \\ P \tilde{g} \\ H \tilde{g} \end{bmatrix} = - \begin{bmatrix} \tilde{g}' P \tilde{g} \\ \tilde{K} P \tilde{g} + \tilde{R} H \tilde{g} \\ 0 \\ \tilde{R}' H \tilde{g} \end{bmatrix}; \quad \tilde{\psi} \tilde{\psi}' = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \bar{P} & \bar{P} & \bar{0} \\ 0 & \bar{P} & \bar{P} & \bar{0} \\ 0 & \bar{0} & \bar{0} & \bar{\Sigma} \end{bmatrix}.$$

Therefore,

$$\begin{aligned}
&\sum_{s=1}^{1+2p+q} \tilde{\psi}'_s M_s \tilde{\psi} / 2 = \sum_{s=1}^{1+2p+q} M_s (\tilde{\psi} \tilde{\psi}') e_s / 2 \\
&= \sum_{j=1}^p \left[ \begin{array}{c} 0 \\ 0 \\ \rho_3 E[g_{i,j} g_i g_i'] \bar{P} + E \left[ \frac{\partial (g_{i,j} g_i)}{\partial \theta'} \right] \bar{0} \\ E \left[ \frac{\partial (g_{i,j} g_i')}{\partial \theta} \right] \bar{P} + E \left[ \frac{\partial G'_{i,j}}{\partial \theta} \right] \bar{0} \end{array} \right] \frac{e_j}{2}
\end{aligned}$$

$$\begin{aligned}
& + \Sigma_{j=1}^p \begin{bmatrix} \rho_3 E[g_{i,j} g'_i] \bar{P} \\ \rho_3 E[g_{i,j} g'_i] \bar{P} + E\left[\frac{\partial(g_{i,j} g'_i)}{\partial \theta'}\right] \bar{0} \\ -E\left[\frac{\partial(g_{i,j} g'_i)}{\partial \theta'}\right] \bar{0} \\ 0 \end{bmatrix} \frac{e_j}{2} + \Sigma_{j=1}^q \begin{bmatrix} 0 \\ E\left[\frac{\partial(g_i g'_i)}{\partial \theta_j}\right] \bar{0} + E\left[\frac{\partial G_i}{\partial \theta_j}\right] \bar{\Sigma} \\ 0 \\ E\left[\frac{\partial G'_i}{\partial \theta_j}\right] \bar{0} \end{bmatrix} \frac{e_j}{2} \\
& = \begin{bmatrix} \Sigma_{j=1}^p \rho_3 E[g_{i,j} g'_i] \bar{P} e_j / 2 \\ \left( \Sigma_{j=1}^p \rho_3 E[g_{i,j} g'_i] \bar{P} e_j / 2 + \Sigma_{j=1}^p E\left[\frac{\partial(g_{i,j} g'_i)}{\partial \theta'}\right] \bar{0} e_j / 2 \right) \\ + \Sigma_{j=1}^p E\left[\frac{\partial(g_i g'_i)}{\partial \theta_j}\right] \bar{0} e_j / 2 + \Sigma_{j=1}^p E\left[\frac{\partial G_i}{\partial \theta_j}\right] \bar{\Sigma} e_j / 2 \\ \Sigma_{j=1}^p \rho_3 E[g_{i,j} g'_i] \bar{P} e_j / 2 \\ \Sigma_{j=1}^p E\left[\frac{\partial(g_{i,j} g'_i)}{\partial \theta}\right] \bar{P} e_j / 2 + \Sigma_{j=1}^p E\left[\frac{\partial G'_{i,j \cdot}}{\partial \theta}\right] \bar{0} e_j / 2 + \Sigma_{j=1}^q E\left[\frac{\partial G'_i}{\partial \theta_j}\right] \bar{0} e_j / 2 \end{bmatrix} \\
& \tilde{A} \tilde{\psi} + \Sigma_{j=1}^{1+2p+q} \tilde{\psi}_j M_j \tilde{\psi} / 2 \\
& = \begin{bmatrix} \Sigma_{j=1}^p \rho_3 E[g_{i,j} g'_i] \bar{P} e_j / 2 - \tilde{g}' P \tilde{g} \\ \left( \Sigma_{j=1}^p \rho_3 E[g_{i,j} g'_i] \bar{P} e_j / 2 + \Sigma_{j=1}^p E\left[\frac{\partial(g_{i,j} g'_i)}{\partial \theta'}\right] \bar{0} e_j / 2 \right) \\ + \Sigma_{j=1}^p E\left[\frac{\partial(g_i g'_i)}{\partial \theta_j}\right] \bar{0} e_j / 2 + \Sigma_{j=1}^p E\left[\frac{\partial G_i}{\partial \theta_j}\right] \bar{\Sigma} e_j / 2 - \tilde{K} P \tilde{g} - \tilde{R} H \tilde{g} \\ \Sigma_{j=1}^p \rho_3 E[g_{i,j} g'_i] \bar{P} e_j / 2 \\ \left( \Sigma_{j=1}^p E\left[\frac{\partial(g_{i,j} g'_i)}{\partial \theta}\right] \bar{P} e_j / 2 + \Sigma_{j=1}^p E\left[\frac{\partial G'_{i,j \cdot}}{\partial \theta}\right] \bar{0} e_j / 2 \right) \\ + \Sigma_{j=1}^q E\left[\frac{\partial G'_i}{\partial \theta_j}\right] \bar{0} e_j / 2 - \tilde{R}' H \tilde{g} \end{bmatrix}
\end{aligned}$$

Since  $\tilde{Q}_1 = -M^{-1} [\tilde{A}_i \tilde{\psi}_i + \Sigma_{s=1}^{1+2p+q} \tilde{\psi}_s M_s \tilde{\psi} / 2]$ , then

$$\begin{aligned}
\tilde{Q}_{1,\tau} &= -[-1 \ 0 \ 0 \ 0] [\tilde{A}_i \tilde{\psi}_i + \Sigma_{s=1}^{1+2p+q} \tilde{\psi}_s M_s \tilde{\psi} / 2] \\
&= \Sigma_{j=1}^p \rho_3 E[g_{i,j} g'_i] \bar{P} e_j / 2 - \tilde{g}' P \tilde{g} \\
&= \tilde{Q}_{1,\tau}^{EL};
\end{aligned}$$

$$\begin{aligned}
\tilde{Q}_{1,\kappa} &= -[0 \quad P \quad P \quad H'] [\tilde{A}_i \tilde{\psi}_i + \sum_{s=1}^{1+2p+q} \tilde{\psi}_s M_s \tilde{\psi} / 2] \\
&= - \left[ \begin{array}{l} P \left( \begin{array}{l} \sum_{j=1}^p \rho_3 E [g_{i,j} g_i g'_i] \bar{P} e_j + \sum_{j=1}^p E \left[ \frac{\partial (g_{i,j} g_i)}{\partial \theta'} \right] \bar{0} e_j / 2 \\ + \sum_{j=1}^p E \left[ \frac{\partial (g_i g'_i)}{\partial \theta_j} \right] \bar{0} e_j / 2 + \sum_{j=1}^p E \left[ \frac{\partial G_i}{\partial \theta_j} \right] \bar{\Sigma} e_j / 2 - \tilde{K} P \tilde{g} - \tilde{R} H \tilde{g} \end{array} \right) \\ + H' \left( \begin{array}{l} \sum_{j=1}^p E \left[ \frac{\partial (g_{i,j} g'_i)}{\partial \theta} \right] \bar{P} e_j / 2 + \sum_{j=1}^p E \left[ \frac{\partial G'_{i,j,\cdot}}{\partial \theta} \right] \bar{0} e_j / 2 \\ + \sum_{j=1}^q E \left[ \frac{\partial G'_i}{\partial \theta_j} \right] \bar{0} e_j / 2 - \tilde{R}' H \tilde{g} \end{array} \right) \end{array} \right] \\
&= \tilde{Q}_{1,\kappa}^{EL} + P(1 - \rho_3) \sum_{j=1}^p E [g_{i,j} g_i g'_i] \bar{P} e_j; \\
\tilde{Q}_{1,\lambda} &= -[0 \quad P \quad P - \Omega^{-1} \quad H'] [\tilde{A}_i \tilde{\psi}_i + \sum_{s=1}^{1+2p+q} \tilde{\psi}_s M_s \tilde{\psi} / 2] \\
&= - \left[ \begin{array}{l} P \left( \begin{array}{l} \sum_{j=1}^p \rho_3 E [g_{i,j} g_i g'_i] \bar{P} e_j + \sum_{j=1}^p E \left[ \frac{\partial (g_{i,j} g_i)}{\partial \theta'} \right] \bar{0} e_j / 2 \\ + \sum_{j=1}^p E \left[ \frac{\partial (g_i g'_i)}{\partial \theta_j} \right] \bar{0} e_j / 2 + \sum_{j=1}^p E \left[ \frac{\partial G_i}{\partial \theta_j} \right] \bar{\Sigma} e_j / 2 - \tilde{K} P \tilde{g} - \tilde{R} H \tilde{g} \end{array} \right) \\ + H' \left( \begin{array}{l} -\Omega^{-1} \sum_{j=1}^p \rho_3 E [g_{i,j} g_i g'_i] \bar{P} e_j \\ \sum_{j=1}^p E \left[ \frac{\partial (g_{i,j} g'_i)}{\partial \theta} \right] \bar{P} e_j / 2 + \sum_{j=1}^p E \left[ \frac{\partial G'_{i,j,\cdot}}{\partial \theta} \right] \bar{0} e_j / 2 \\ + \sum_{j=1}^q E \left[ \frac{\partial G'_i}{\partial \theta_j} \right] \bar{0} e_j / 2 - \tilde{R}' H \tilde{g} \end{array} \right) \end{array} \right] \\
&= \tilde{Q}_{1,\lambda}^{EL} + (P - \Omega^{-1})(1 - \rho_3) \sum_{j=1}^p E [g_{i,j} g_i g'_i] \bar{P} e_j; \\
\tilde{Q}_{1,\theta} &= -[0 \quad H \quad H \quad -\Sigma] [\tilde{A}_i \tilde{\psi}_i + \sum_{s=1}^{1+2p+q} \tilde{\psi}_s M_s \tilde{\psi} / 2]
\end{aligned}$$



$$\begin{aligned}
&= - \left[ \begin{array}{c} H \left( \begin{array}{c} \Sigma_{j=1}^p \rho_3 E [g_{i,j} g_i g'_i] \bar{P} e_j + \Sigma_{j=1}^p E \left[ \frac{\partial (g_{i,j} g_i)}{\partial \theta'} \right] \bar{0} e_j / 2 \\ + \Sigma_{j=1}^p E \left[ \frac{\partial (g_i g'_i)}{\partial \theta_j} \right] \bar{0} e_j / 2 + \Sigma_{j=1}^p E \left[ \frac{\partial G_i}{\partial \theta_j} \right] \bar{\Sigma} e_j / 2 - \tilde{K} P \tilde{g} - \tilde{R} H \tilde{g} \end{array} \right) \\ -\Sigma \left( \begin{array}{c} \Sigma_{j=1}^p E \left[ \frac{\partial (g_{i,j} g'_i)}{\partial \theta} \right] \bar{P} e_j / 2 + \Sigma_{j=1}^p E \left[ \frac{\partial G'_{i,j}}{\partial \theta} \right] \bar{0} e_j / 2 \\ + \Sigma_{j=1}^q E \left[ \frac{\partial G'_i}{\partial \theta_j} \right] \bar{0} e_j / 2 - \tilde{R}' H \tilde{g} \end{array} \right) \end{array} \right] \\
&= \tilde{Q}_{1,\theta}^{EL} + H(1 - \rho_3) \Sigma_{j=1}^p E [g_{i,j} g_i g'_i] \bar{P} e_j.
\end{aligned}$$

Let  $\bar{Q} = (1 - \rho_3) \Sigma_{j=1}^p E [g_{i,j} g_i g'_i] \bar{P} e_j$ ,  $\tilde{Q}_1$  of DGEL can be written as functions of  $\tilde{Q}_1^{EL}$  as

$$\tilde{Q}_1 = \begin{bmatrix} \sim Q_{1,\tau}^{EL} \\ \sim Q_{1,\kappa}^{EL} + P\bar{Q} \\ \sim Q_{1,\lambda}^{EL} + (P - \Omega^{-1})\bar{Q} \\ \sim Q_{1,\theta}^{EL} + H\bar{Q} \end{bmatrix}.$$

For any DGEL estimators with  $\rho_3 = 1$  or  $E [g_{i,j} g_i g'_i] = 0$ , its  $\tilde{Q}_1 = \tilde{Q}_1^{EL}$ .

I now consider the form of  $\tilde{Q}_2$ :

$$\begin{aligned}
\tilde{Q}_2 &= -M^{-1} \left[ \begin{array}{c} \tilde{A}\tilde{Q}_1 + \Sigma_{s=1}^{1+2p+q} (\tilde{\psi}_s M_s \tilde{Q}_1 + \tilde{Q}_{1s} M_s \tilde{\psi} + \tilde{\psi}_s \tilde{B}_s \tilde{\psi}) / 2 \\ + \Sigma_{s=1}^{1+2p+q} \Sigma_{l=1}^{1+2p+q} \tilde{\psi}_s \tilde{\psi}_k M_{sk} \tilde{\psi} / 6 \end{array} \right] \\
&\equiv I + II + III + IV + V.
\end{aligned}$$

First, calculate  $\tilde{A}\tilde{Q}_1$ :

$$\tilde{A}\tilde{Q}_1 = \begin{bmatrix} 0 & 0 & \tilde{g}' & 0 \\ 0 & 0 & \tilde{K} & \tilde{R} \\ \tilde{g} & \tilde{K} & -\tilde{K} & 0 \\ 0 & \tilde{R}' & 0 & 0 \end{bmatrix} \begin{bmatrix} \tilde{Q}_{1,\tau} \\ \tilde{Q}_{1,\kappa} \\ \tilde{Q}_{1,\lambda} \\ \tilde{Q}_{1,\theta} \end{bmatrix} = \begin{bmatrix} \tilde{g}' \tilde{Q}_{1,\lambda} \\ \tilde{K} \tilde{Q}_{1,\lambda} + \tilde{R} \tilde{Q}_{1,\theta} \\ \tilde{K} (\tilde{Q}_{1,\kappa} - \tilde{Q}_{1,\lambda}) + \tilde{g} \tilde{Q}_{1,\tau} \\ \tilde{R}' \tilde{Q}_{1,\kappa} \end{bmatrix}.$$

Then the difference of  $\tilde{A}\tilde{Q}_1$  between DGEL and EL is

$$\left[ (\tilde{A}\tilde{Q}_1)^{DGEL} - (\tilde{A}\tilde{Q}_1)^{EL} \right] = \begin{bmatrix} \tilde{g}' (P - \Omega^{-1}) \bar{Q} \\ \tilde{K} (P - \Omega^{-1}) \bar{Q} + \tilde{R} H \bar{Q} \\ \tilde{g} Q_{1,\tau}^{EL} + \tilde{K} \Omega^{-1} \bar{Q} \\ \tilde{R}' P \bar{Q} \end{bmatrix}.$$

$$\begin{aligned}
I^{DGEL} - I^{EL} &= - \begin{bmatrix} 0 \\ H' \\ H' \\ -\Sigma' \end{bmatrix}' \left( (\tilde{A}\tilde{Q}_1)^{DGEL} - (\tilde{A}\tilde{Q}_1)^{EL} \right) = - \begin{bmatrix} 0 \\ H' \\ H' \\ -\Sigma' \end{bmatrix}' \begin{bmatrix} \tilde{g}'(P - \Omega^{-1})\tilde{Q} \\ K(P - \Omega^{-1})\tilde{Q} + RH\tilde{Q} \\ \tilde{g}Q_{1,\tau}^{EL} + K\Omega^{-1}\tilde{Q} \\ R'P\tilde{Q} \end{bmatrix} \\
&= -H(KP\tilde{Q} + RH\tilde{Q} + \tilde{g}Q_{1,\tau}^{EL}) + \Sigma R'P\tilde{Q} \\
&= -H(KP\tilde{Q} + RH\tilde{Q} + \tilde{g}\sum_{j=1}^p \rho_3 E[g_{i,j}g_i'] \bar{P}e_j / 2 - \tilde{g}\tilde{g}'P\tilde{g}) + \Sigma R'P\tilde{Q}.
\end{aligned}$$

If  $\rho_3 = 1$ ,  $I^{DGEL} - I^{EL} = -H\tilde{g}(\sum_{j=1}^p E[g_{i,j}g_i'] \bar{P}e_j / 2 - \tilde{g}\tilde{g}'P\tilde{g})$ . Since  $E(\bar{P}) = P$ ,  $E(\tilde{\psi}'_\theta(I^{DGEL} - I^{EL})) = -E(\tilde{g}'H'H\tilde{g})E(g_i'Pg_i) + E(\tilde{g}'H'H\tilde{g}) * E(\tilde{g}'P\tilde{g}) = 0$ .

$$\text{If } E[g_{i,j}g_i'] = 0, \quad I^{DGEL} - I^{EL} = -H\tilde{g}Q_{1,\tau}^{EL} = -H\tilde{g}(\sum_{j=1}^p \rho_3 E[g_{i,j}g_i'] \bar{P}e_j / 2 - \tilde{g}\tilde{g}'P\tilde{g}).$$

This difference will be canceled out with the first term in II and the remaining term in III.

Second, because

$$\tilde{Q}_1 \tilde{\psi}' = - \begin{bmatrix} \tilde{Q}_{1,\tau} \\ \tilde{Q}_{1,\kappa} \\ \tilde{Q}_{1,\lambda} \\ \tilde{Q}_{1,\theta} \end{bmatrix}' \begin{bmatrix} 0 \\ P\tilde{g} \\ P\tilde{g} \\ H\tilde{g} \end{bmatrix} = - \begin{bmatrix} 0 & \tilde{Q}_{1,\tau}\tilde{g}'P & \tilde{Q}_{1,\tau}\tilde{g}'P & \tilde{Q}_{1,\tau}\tilde{g}'H' \\ 0 & \tilde{Q}_{1,\kappa}\tilde{g}'P & \tilde{Q}_{1,\kappa}\tilde{g}'P & \tilde{Q}_{1,\kappa}\tilde{g}'H' \\ 0 & \tilde{Q}_{1,\lambda}\tilde{g}'P & \tilde{Q}_{1,\lambda}\tilde{g}'P & \tilde{Q}_{1,\lambda}\tilde{g}'H' \\ 0 & \tilde{Q}_{1,\theta}\tilde{g}'P & \tilde{Q}_{1,\theta}\tilde{g}'P & \tilde{Q}_{1,\theta}\tilde{g}'H' \end{bmatrix},$$

I have

$$\begin{aligned}
&\sum_{s=1}^{1+2p+q} \tilde{\psi}'_s M_s \tilde{Q}_1 / 2 = \sum_{s=1}^{1+2p+q} M_s \tilde{Q}_1 \tilde{\psi}'_s e_s / 2 \\
&= -\sum_{j=1}^p \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \rho_3 E[g_{i,j}g_i g_i'] & E\left[\frac{\partial(g_{i,j}g_i)}{\partial\theta'}\right] \\ 0 & 0 & E\left[\frac{\partial(g_{i,j}g_i')}{\partial\theta}\right] & E\left[\frac{\partial G'_{i,j,\cdot}}{\partial\theta'}\right] \end{bmatrix} \begin{bmatrix} \tilde{Q}_{1,\tau}\tilde{g}'P \\ \tilde{Q}_{1,\kappa}\tilde{g}'P \\ \tilde{Q}_{1,\lambda}\tilde{g}'P \\ \tilde{Q}_{1,\theta}\tilde{g}'P \end{bmatrix} e_j / 2 \\
&-\sum_{j=1}^p \begin{bmatrix} 0 & 0 & \rho_3 E[g_{i,j}g_i'] & 0 \\ 0 & 0 & \rho_3 E[g_{i,j}g_i g_i'] & E\left[\frac{\partial(g_{i,j}g_i)}{\partial\theta'}\right] \\ 0 & \rho_3 E[g_{i,j}g_i g_i'] & -\rho_3 E[g_{i,j}g_i g_i'] & -E\left[\frac{\partial(g_{i,j}g_i)}{\partial\theta'}\right] \\ E[G_{i,j,\cdot}] & E\left[\frac{\partial(g_{i,j}g_i')}{\partial\theta}\right] & -E\left[\frac{\partial(g_{i,j}g_i')}{\partial\theta}\right] & 0 \end{bmatrix} \begin{bmatrix} \tilde{Q}_{1,\tau}\tilde{g}'P \\ \tilde{Q}_{1,\kappa}\tilde{g}'P \\ \tilde{Q}_{1,\lambda}\tilde{g}'P \\ \tilde{Q}_{1,\theta}\tilde{g}'P \end{bmatrix} \frac{e_j}{2}
\end{aligned}$$

$$\begin{aligned}
& -\sum_{j=1}^q \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & E\left[\frac{\partial(g_i g_i')}{\partial \theta_j}\right] & E\left[\frac{\partial G_i}{\partial \theta_j}\right] \\ E[G_{i,\cdot,j}] & E\left[\frac{\partial(g_i g_i')}{\partial \theta_j}\right] & -E\left[\frac{\partial(g_i g_i')_i}{\partial \theta_j}\right] & 0 \\ 0 & E[\partial G_i' / \partial \theta_j] & 0 & 0 \end{bmatrix} \begin{bmatrix} \tilde{Q}_{1,\tau} \tilde{g}' H' \\ \tilde{Q}_{1,\kappa} \tilde{g}' H' \\ \tilde{Q}_{1,\lambda} \tilde{g}' H' \\ \tilde{Q}_{1,\theta} \tilde{g}' H' \end{bmatrix} \frac{e_j}{2} \\
& = - \begin{bmatrix} \sum_{j=1}^p \rho_3 E[g_{i,j} g_i'] \tilde{Q}_{1,\lambda} \tilde{g}' P e_j / 2 \\ \left( \sum_{j=1}^p \left( \rho_3 E[g_{i,j} g_i g_i'] \tilde{Q}_{1,\lambda} \tilde{g}' P + E\left[\frac{\partial g_{i,j} g_i}{\partial \theta'}\right] \tilde{Q}_{1,\theta} \tilde{g}' P \right) e_j / 2 \right. \\ \left. + \sum_{j=1}^q E\left[\frac{\partial g_i g_i'}{\partial \theta_j}\right] \tilde{Q}_{1,\lambda} \tilde{g}' H' e_j / 2 + \sum_{j=1}^q E\left[\frac{\partial G_i}{\partial \theta_j}\right] \tilde{Q}_{1,\theta} \tilde{g}' H' e_j / 2 \right) \\ \left( \sum_{j=1}^p \rho_3 E[g_{i,j} g_i g_i'] \tilde{Q}_{1,\kappa} \tilde{g}' P e_j / 2 + \sum_{j=1}^q E[G_{i,\cdot,j}] \tilde{Q}_{1,\tau} \tilde{g}' H' e_j / 2 \right. \\ \left. + \sum_{j=1}^q E\left[\frac{\partial g_i g_i'}{\partial \theta_j}\right] (\tilde{Q}_{1,\kappa} - \tilde{Q}_{1,\lambda}) \tilde{g}' H' e_j / 2 \right) \\ \left( \sum_{j=1}^p E\left[\frac{\partial G_{i,j,\cdot}'}{\partial \theta}\right] \tilde{Q}_{1,\theta} \tilde{g}' P e_j / 2 + \sum_{j=1}^p (E[G_{i,j,\cdot}]) \tilde{Q}_{1,\tau} \tilde{g}' P e_j / 2 \right. \\ \left. + \sum_{j=1}^p E\left[\frac{\partial(g_{i,j} g_i')}{\partial \theta}\right] \tilde{Q}_{1,\kappa} \tilde{g}' P e_j / 2 + \sum_{j=1}^q E\left[\frac{\partial G_i'}{\partial \theta_j}\right] \tilde{Q}_{1,\kappa} \tilde{g}' H' e_j / 2 \right) \end{bmatrix}.
\end{aligned}$$

Therefore, the difference of  $\sum_{s=1}^{1+2p+q} \tilde{\psi}_s M_s \tilde{Q}_1 / 2$  between DGEL and EL is

$$\left( \sum_{s=1}^{1+2p+q} \tilde{\psi}_s M_s \tilde{Q}_1 / 2 \right)^{DGEL} - \left( \sum_{s=1}^{1+2p+q} \tilde{\psi}_s M_s \tilde{Q}_1 / 2 \right)^{EL} =$$

$$\left[ \begin{array}{c} \Sigma_{j=1}^p \rho_3 E[g_{i,j} g_i'] (P - \Omega^{-1}) \bar{Q} \tilde{g}' P e_j / 2 \\ \left( \Sigma_{j=1}^p \left( \rho_3 E[g_{i,j} g_i g_i'] (P - \Omega^{-1}) \bar{Q} \tilde{g}' P + E \left[ \frac{\partial (g_{i,j} g_i)}{\partial \theta'} \right] H \bar{Q} \tilde{g}' P \right) e_j / 2 \right. \\ \left. + \Sigma_{j=1}^q \left( E \left[ \frac{\partial (g_i g_i')}{\partial \theta_j} \right] (P - \Omega^{-1}) \bar{Q} \tilde{g}' H' + E \left[ \frac{\partial G_i}{\partial \theta_j} \right] H \bar{Q} \tilde{g}' H \right) e_j / 2 \right) \\ \left( \Sigma_{j=1}^p \rho_3 E[g_{i,j} g_i g_i'] Q_{1,\lambda}^{EL} \tilde{g}' P e_j / 2 - \Sigma_{j=1}^p (2 - \rho_3) E[g_j g g'] Q_{1,\kappa}^{EL} \tilde{g}' P e_j / 2 \right. \\ \left. + \Sigma_{j=1}^q E[G_{i,\cdot,j}] Q_{1,\tau}^{EL} \tilde{g}' H / 2 + \Sigma_{j=1}^p \rho_3 E[g_{i,j} g_i g_i'] P \bar{Q} \tilde{g}' P e_j / 2 \right. \\ \left. + \Sigma_{j=1}^q E \left[ \frac{\partial (g_i g_i')}{\partial \theta_j} \right] \Omega^{-1} \bar{Q} \tilde{g}' H / 2 \right) \\ \left( \Sigma_{j=1}^p E \left[ \frac{\partial G'_{i,j,\cdot}}{\partial \theta} \right] H \bar{Q} \tilde{g}' P e_j / 2 + \Sigma_{j=1}^p E \left[ \frac{\partial (g_{i,j} g_i')}{\partial \theta} \right] P \bar{Q} \tilde{g}' P e_j / 2 \right) \\ \left. + \Sigma_{j=1}^q E \left[ \frac{\partial G'_i}{\partial \theta_j} \right] P \bar{Q} \tilde{g}' H' e_j / 2 + \Sigma_{j=1}^p \left( E[G_{i,j,\cdot}] \tilde{Q}_{1,\tau} \tilde{g}' P \right) e_j / 2 \right) \end{array} \right]$$

If  $\rho_3 = 1$  or  $E[g_j g g'] = 0$ , then  $\bar{Q} = 0$ . Thus,

$$\begin{aligned} \Pi^{DGEL} - \Pi^{EL} &= - \begin{bmatrix} 0 \\ H' \\ H' \\ -\Sigma' \end{bmatrix}' \left( \left( \Sigma_{s=1}^{1+2p+q} \tilde{\psi}_s M_s \tilde{Q}_1 / 2 \right)^{DGEL} - \left( \Sigma_{s=1}^{1+2p+q} \tilde{\psi}_s M_s \tilde{Q}_1 / 2 \right)^{EL} \right) \\ &= \begin{bmatrix} 0 \\ H' \\ H' \\ -\Sigma' \end{bmatrix}' \begin{bmatrix} 0 \\ 0 \\ \left( \Sigma_{j=1}^p \rho_3 E[g_{i,j} g_i g_i'] Q_{1,\lambda}^{EL} \tilde{g}' P e_j / 2 \right. \\ \left. - \Sigma_{j=1}^p (2 - \rho_3) E[g_{i,j} g_i g_i'] Q_{1,\kappa}^{EL} \tilde{g}' P e_j / 2 \right. \\ \left. + \Sigma_{j=1}^q E[G_{i,\cdot,j}] Q_{1,\tau}^{EL} \tilde{g}' H / 2 \right. \\ \left. \Sigma_{j=1}^p \left( E[G_{i,j,\cdot}] \tilde{Q}_{1,\tau} \tilde{g}' P \right) e_j / 2 \right) \end{bmatrix}. \end{aligned}$$

Furthermore, if  $\rho_3 = 1$ ,

$$\Pi^{DGEL} - \Pi^{EL}$$

$$\begin{aligned}
&= \begin{bmatrix} 0 \\ H' \\ H' \\ -\Sigma' \end{bmatrix}' \begin{bmatrix} 0 \\ 0 \\ \left( \Sigma_{s=1}^p E[g_{i,j} g_i g_i'] \left[ \Omega \Sigma_{s=1}^p E[g_{i,j} g_i g_i'] \bar{P} e_j / 2 \right] \tilde{g}' P e_j / 2 \right) \\ + \Sigma_{j=1}^q E[G_{i,\cdot,j}] Q_{1,\tau}^{EL} \tilde{g}' H / 2 \\ \Sigma_{j=1}^p E[G_{i,j,\cdot}] \tilde{Q}_{1,\tau}^{EL} \tilde{g}' P e_j / 2 \end{bmatrix} \\
&= H \begin{pmatrix} \Sigma_{s=1}^p E[g_{i,j} g_i g_i'] \left[ \Omega \Sigma_{s=1}^p E[g_{i,j} g_i g_i'] \bar{P} e_j / 2 \right] \tilde{g}' P e_j / 2 \\ + \Sigma_{j=1}^q E[G_{i,\cdot,j}] Q_{1,\tau}^{EL} \tilde{g}' H / 2 \\ - \Sigma_{j=1}^p E[G_{i,j,\cdot}] Q_{1,\tau}^{EL} \tilde{g}' P e_j / 2 \end{pmatrix} \\
&= H \begin{pmatrix} \Sigma_{s=1}^p E[g_{i,j} g_i g_i'] \left[ \Omega \Sigma_{s=1}^p E[g_{i,j} g_i g_i'] \bar{P} e_j / 2 \right] \tilde{g}' P e_j / 2 \\ + \Sigma_{j=1}^q E[G_{i,\cdot,j}] Q_{1,\tau}^{EL} \tilde{g}' H / 2 \end{pmatrix}.
\end{aligned}$$

The last step is derived from  $\Sigma \Sigma_{j=1}^p (E[G_{i,j,\cdot}] Q_{1,\tau}^{EL} \tilde{g}' P) e_j / 2 = 0$ . Since  $\tilde{\psi}_\theta = H \tilde{g}$ ,  $H \tilde{g}$  is uncorrelated with  $P \tilde{g}$ , I get

$$\begin{aligned}
&E(\tilde{\psi}'_\theta (II^{DGEL} - II^{EL})) \\
&= E \left[ \tilde{g}' H' H \Sigma_{j=1}^q E[G_{i,\cdot,j}] \left( \Sigma_{j=1}^p E[g_{i,j} g_i g_i'] \bar{P} e_j / 2 - \tilde{g}' P \tilde{g} \right) \tilde{g}' H / 2 \right] \\
&= E \left[ \tilde{g}' H' \left( \Sigma_{j=1}^p E[g_{i,j} g_i g_i'] \bar{P} e_j / 2 - \tilde{g}' P \tilde{g} \right) \tilde{g}' H / 2 \right] = 0.
\end{aligned}$$

The last step is derived upon the facts  $HG=I$ ,  $H \tilde{g}$  is uncorrelated with  $P \tilde{g}$ , and  $E(Q_{1,\tau}^{EL}) = 0$  at  $\rho_3 = 1$ .

Furthermore, if data are symmetric with  $E[g_{i,j} g_i g_i'] = 0$ ,

$$II^{DGEL} - II^{EL} = H \Sigma_{j=1}^q E[G_{i,\cdot,j}] Q_{1,\tau}^{EL} \tilde{g}' H e_j / 2 - \Sigma_{j=1}^p E[G_{i,j,\cdot}] \tilde{Q}_{1,\tau}^{EL} \tilde{g}' P e_j / 2.$$

Since  $HG=I$ , the first term here together with the remaining term in III cancels out with term in  $I^{DGEL} - I^{EL}$ . Because  $H \tilde{g}$  is uncorrelated with  $P \tilde{g}$  and  $\tilde{Q}_{1,\tau}^{EL} = (\Sigma_{j=1}^p \rho_3 E[g_{i,j} g_i g_i'] \bar{P} e_j / 2 - \tilde{g}' P \tilde{g})$ , the second term is a function of  $P \tilde{g}$ , which are uncorrelated with  $H \tilde{g}$ .

Third, since

$$\begin{aligned} \tilde{\psi} \tilde{Q}'_1 &= - \begin{bmatrix} 0 \\ P\tilde{g} \\ P\tilde{g} \\ H\tilde{g} \end{bmatrix} \begin{bmatrix} \tilde{Q}_{1,\tau}^{EL} \\ \tilde{Q}_{1,\kappa}^{EL} + P\bar{Q} \\ \tilde{Q}_{1,\lambda}^{EL} + (P - \Omega^{-1})\bar{Q} \\ \tilde{Q}_{1,\theta}^{EL} + H\bar{Q} \end{bmatrix}' \\ &= - \begin{bmatrix} 0 & 0 & 0 & 0 \\ P\tilde{g}Q_{1,\tau}^{EL'} & P\tilde{g}Q_{1,\kappa}^{EL'} & P\tilde{g}Q_{1,\lambda}^{EL'} & P\tilde{g}Q_{1,\theta}^{EL'} \\ P\tilde{g}Q_{1,\tau}^{EL'} & P\tilde{g}Q_{1,\kappa}^{EL'} & P\tilde{g}Q_{1,\lambda}^{EL'} & P\tilde{g}Q_{1,\theta}^{EL'} \\ H\tilde{g}Q_{1,\tau}^{EL'} & H\tilde{g}Q_{1,\kappa}^{EL'} & H\tilde{g}Q_{1,\lambda}^{EL'} & H\tilde{g}Q_{1,\theta}^{EL'} \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & P\tilde{g}\bar{Q}'P & P\tilde{g}\bar{Q}'(P - \Omega^{-1}) & P\tilde{g}\bar{Q}'H' \\ 0 & P\tilde{g}\bar{Q}'P & P\tilde{g}\bar{Q}'(P - \Omega^{-1}) & P\tilde{g}\bar{Q}'H' \\ 0 & H\tilde{g}\bar{Q}'P & H\tilde{g}\bar{Q}'(P - \Omega^{-1}) & H\tilde{g}\bar{Q}'H' \end{bmatrix}, \end{aligned}$$

then

$$\begin{aligned} \sum_{s=1}^{1+2p+q} \tilde{Q}_{1s} M_s \tilde{\psi} / 2 &= \sum_{s=1}^{1+2p+q} M_s \tilde{\psi} \tilde{Q}'_1 e_s / 2 \\ &= - \begin{bmatrix} 0 \\ 0 \\ GH\tilde{g}Q_{1,\tau}^{EL'} / 2 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \sum_{j=1}^p \left( \rho_3 E[g_{i,j} g_i g_i'] P\tilde{g}Q_{1,\kappa}^{EL'} + E \left[ \frac{\partial(g_{i,j} g_i)}{\partial \theta'} \right] H\tilde{g}Q_{1,\kappa}^{EL'} \right) e_j / 2 \\ \sum_{j=1}^p \left( E \left[ \frac{\partial(g_{i,j} g_i')}{\partial \theta} \right] P\tilde{g}Q_{1,\kappa}^{EL'} + E \left[ \frac{\partial G'_{i,j \cdot}}{\partial \theta} \right] H\tilde{g}Q_{1,\kappa}^{EL'} \right) e_j / 2 \end{bmatrix} \\ &= - \begin{bmatrix} \sum_{j=1}^p \rho_3 E[g_{i,j} g_i'] P\tilde{g}Q_{1,\lambda}^{EL'} e_j / 2 \\ \left( \sum_{j=1}^p \rho_3 E[g_{i,j} g_i g_i'] P\tilde{g}Q_{1,\lambda}^{EL'} \right) e_j / 2 \\ + \sum_{j=1}^p E \left[ \frac{\partial(g_{i,j} g_i)}{\partial \theta'} \right] H\tilde{g}Q_{1,\lambda}^{EL'} e_j / 2 \\ - \sum_{j=1}^p E \left[ \frac{\partial(g_{i,j} g_i)}{\partial \theta'} \right] H\tilde{g}Q_{1,\lambda}^{EL'} e_j / 2 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ \left( \sum_{j=1}^q E \left[ \frac{\partial(g_i g_i')}{\partial \theta_j} \right] P\tilde{g}Q_{1,\theta}^{EL'} e_j / 2 \right) \\ + \sum_{j=1}^q E \left[ \frac{\partial G_i}{\partial \theta_j} \right] H\tilde{g}Q_{1,\theta}^{EL'} e_j / 2 \\ 0 \\ \sum_{j=1}^q E \left[ \frac{\partial G'_i}{\partial \theta_j} \right] P\tilde{g}Q_{1,\theta}^{EL'} e_j / 2 \end{bmatrix} \end{aligned}$$

$$\begin{bmatrix}
0 \\
0 \\
\Sigma_{j=1}^p \left( \rho_3 E [g_{i,j} g_i'] P \tilde{g} \bar{Q}' P + E \left[ \frac{\partial (g_{i,j} g_i)}{\partial \theta'} \right] H \tilde{g} \bar{Q}' P \right) e_j / 2 \\
\Sigma_{j=1}^p \left( E \left[ \frac{\partial (g_{i,j} g_i')}{\partial \theta} \right] P \tilde{g} \bar{Q}' P + E \left[ \frac{\partial G'_{i,j \cdot}}{\partial \theta} \right] H \tilde{g} \bar{Q}' P \right) e_j / 2
\end{bmatrix}$$

$$\begin{aligned}
& \begin{bmatrix} \Sigma_{j=1}^p \rho_3 E [g_{i,j} g'_i] P \tilde{g} \bar{Q}' (P - \Omega^{-1}) e_j / 2 \\ \left( \Sigma_{j=1}^p \rho_3 E [g_{i,j} g_i g'_i] P \tilde{g} \bar{Q}' (P - \Omega^{-1}) e_j / 2 \right) \\ + \Sigma_{j=1}^p E \left[ \frac{\partial (g_{i,j} g_i)}{\partial \theta'} \right] H \tilde{g} \bar{Q}' (P - \Omega^{-1}) e_j / 2 \\ - \Sigma_{j=1}^p E \left[ \frac{\partial (g_{i,j} g_i)}{\partial \theta'} \right] H \tilde{g} \bar{Q}' (P - \Omega^{-1}) e_j / 2 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ \left( \Sigma_{j=1}^q E \left[ \frac{\partial (g_i g'_i)}{\partial \theta_j} \right] P \tilde{g} \bar{Q}' H' e_j / 2 \right) \\ + \Sigma_{j=1}^q E \left[ \frac{\partial G_i}{\partial \theta_j} \right] H \tilde{g} \bar{Q}' H' e_j / 2 \\ 0 \\ \Sigma_{j=1}^q E \left[ \frac{\partial G'_i}{\partial \theta_j} \right] P \tilde{g} \bar{Q}' H' / 2 \end{bmatrix} \\
& = \begin{bmatrix} \left( \Sigma_{j=1}^p \rho_3 E [g_{i,j} g'_i] P \tilde{g} Q_{\lambda}^{EL'} e_j / 2 + \Sigma_{j=1}^p \rho_3 E [g_{i,j} g'_i] P \tilde{g} \bar{Q}' (P - \Omega^{-1}) e_j / 2 \right) \\ \left( \Sigma_{j=1}^p \left( \rho_3 E [g_{i,j} g_i g'_i] P \tilde{g} \tilde{Q}_{\lambda}^{EL'} + E \left[ \frac{\partial (g_{i,j} g_i)}{\partial \theta'} \right] H \tilde{g} \tilde{Q}_{\lambda}^{EL'} \right) e_j / 2 \right) \\ + \Sigma_{j=1}^q E \left[ \frac{\partial (g_i g'_i)}{\partial \theta_j} \right] P \tilde{g} \tilde{Q}_{\theta}^{EL'} e_j / 2 + \Sigma_{j=1}^q E \left[ \frac{\partial G_i}{\partial \theta_j} \right] H \tilde{g} \tilde{Q}_{\theta}^{EL'} e_j / 2 \\ + \Sigma_{j=1}^p \rho_3 E [g_{i,j} g_i g'_i] P \tilde{g} \bar{Q}' (P - \Omega^{-1}) e_j / 2 \\ + \Sigma_{j=1}^p E \left[ \frac{\partial (g_{i,j} g_i)}{\partial \theta'} \right] H \tilde{g} \bar{Q}' (P - \Omega^{-1}) e_j / 2 \\ + \Sigma_{j=1}^q E \left[ \frac{\partial (g_i g'_i)}{\partial \theta_j} \right] P \tilde{g} \bar{Q}' H' e_j / 2 + \Sigma_{j=1}^q E \left[ \frac{\partial G_i}{\partial \theta_j} \right] H \tilde{g} \bar{Q}' H' e_j / 2 \end{bmatrix} \\
& = \begin{bmatrix} G H \tilde{g} Q_{\tau}^{EL'} / 2 + \Sigma_{j=1}^p \rho_3 E [g_{i,j} g_i g'_i] P \tilde{g} Q_{\kappa}^{EL'} e_j / 2 \\ + \Sigma_{j=1}^p E \left[ \frac{\partial (g_{i,j} g_i)}{\partial \theta'} \right] H \tilde{g} Q_{\kappa}^{EL'} e_j / 2 - \Sigma_{j=1}^p E \left[ \frac{\partial (g_{i,j} g_i)}{\partial \theta'} \right] H \tilde{g} Q_{\lambda}^{EL'} e_j / 2 \\ + \Sigma_{j=1}^p \left( \rho_3 E [g_{i,j} g_i g'_i] P \tilde{g} \bar{Q}' P + E \left[ \frac{\partial (g_{i,j} g_i)}{\partial \theta'} \right] H \tilde{g} \bar{Q}' P \right) e_j / 2 \\ - \Sigma_{j=1}^p E \left[ \frac{\partial (g_{i,j} g_i)}{\partial \theta'} \right] H \tilde{g} \bar{Q}' (P - \Omega^{-1}) e_j / 2 \\ \left( \Sigma_{j=1}^p \left( E \left[ \frac{\partial (g_{i,j} g'_i)}{\partial \theta} \right] P \tilde{g} Q_{\kappa}^{EL'} + E \left[ \frac{\partial G'_{i,j \cdot}}{\partial \theta} \right] H \tilde{g} Q_{\kappa}^{EL'} \right) e_j / 2 \right) \\ + \Sigma_{j=1}^q E \left[ \frac{\partial G'_i}{\partial \theta_j} \right] P \tilde{g} Q_{\theta}^{EL'} e_j / 2 + \Sigma_{j=1}^q E \left[ \frac{\partial G'_i}{\partial \theta_j} \right] P \tilde{g} \bar{Q}' H' / 2 \end{bmatrix} .
\end{aligned}$$



$$\begin{aligned}
& \left( \sum_{s=1}^{1+2p+q} \tilde{Q}_{1s} M_s \tilde{\psi} / 2 \right)^{DGEL} - \left( \sum_{s=1}^{1+2p+q} \tilde{Q}_{1s} M_s \tilde{\psi} / 2 \right)^{EL} = \\
& \left[ \begin{array}{c}
\Sigma_{s=1}^p \rho_3 E \left[ g_{i,j} g_i' \right] P \tilde{g} \bar{Q}' (P - \Omega^{-1}) e_j / 2 \\
\left( \begin{array}{c}
\Sigma_{s=1}^p \rho_3 E \left[ g_{i,j} g_i g_i' \right] P \tilde{g} \bar{Q}' (P - \Omega^{-1}) e_j / 2 \\
+ \Sigma_{s=1}^p E \left[ \frac{\partial (g_{i,j} g_i)}{\partial \theta'} \right] H \tilde{g} \bar{Q}' (P - \Omega^{-1}) e_j / 2 \\
+ \Sigma_{s=1}^p E \left[ \frac{\partial (g_i g_i')}{\partial \theta_j} \right] P \tilde{g} \bar{Q}' H' e_j / 2 + \Sigma_{s=1}^p E \left[ \frac{\partial G_i}{\partial \theta_j} \right] H \tilde{g} \bar{Q}' H' e_j / 2
\end{array} \right) \\
\left( \begin{array}{c}
GH \tilde{g} Q_{1,\tau}^{EL'} / 2 + \Sigma_{s=1}^p (\rho_3 - 2) E \left[ g_{i,j} g_i g_i' \right] P \tilde{g} Q_{1,\kappa}^{EL'} e_j / 2 \\
+ \Sigma_{s=1}^p \rho_3 E \left[ g_{i,j} g_i g_i' \right] P \tilde{g} Q_{1,\lambda}^{EL'} e_j / 2 \\
+ \Sigma_{s=1}^p \rho_3 E \left[ g_{i,j} g_i g_i' \right] P \tilde{g} \bar{Q}' P e_j / 2 + \Sigma_{s=1}^p E \left[ \frac{\partial (g_{i,j} g_i)}{\partial \theta'} \right] H \tilde{g} \bar{Q}' \Omega^{-1} e_j / 2
\end{array} \right) \\
\Sigma_{s=1}^p \left( \left[ \frac{\partial (g_{i,j} g_i')}{\partial \theta} \right] P \tilde{g} \bar{Q}' P + E \left[ \frac{\partial G_{i,j}}{\partial \theta'} \right] H \tilde{g} \bar{Q}' P \right) e_j / 2 + \Sigma_{s=1}^p E \left[ \frac{\partial G_i'}{\partial \theta_j} \right] P \tilde{g} \bar{Q}' H' / 2
\end{array} \right].
\end{aligned}$$

Since,

$$III^{DGEL} - III^{EL} = - \begin{bmatrix} 0 \\ H' \\ H' \\ -\Sigma' \end{bmatrix}' \left( \left( \sum_{s=1}^{1+2p+q} \tilde{Q}_{1s} M_s \tilde{\psi} / 2 \right)^{DGEL} - \left( \sum_{s=1}^{1+2p+q} \tilde{Q}_{1s} M_s \tilde{\psi} / 2 \right)^{EL} \right),$$

if  $\rho_3 = 1$ , then

$$\begin{aligned}
III^{DGEL} - III^{EL} &= \\
&= \begin{bmatrix} 0 \\ H' \\ H' \\ -\Sigma' \end{bmatrix}' \left[ \begin{array}{c}
0 \\
0 \\
\left( -GH \tilde{g} \left( \Sigma_{s=1}^p E \left[ g_{i,j} g_i' \right] \bar{P} e_j / 2 - \tilde{g}' P \tilde{g} \right) / 2 \right. \\
\left. - \Sigma_{s=1}^p E \left[ g_{i,j} g_i g_i' \right] P \tilde{g} \left[ \Sigma_{s=1}^p \Omega^{-1} E \left[ g_{i,j} g_i g_i' \right] \bar{P} e_j / 2 \right] e_j / 2 \right) \\
0
\end{array} \right] \\
&= H \left( \begin{array}{c}
GH \tilde{g} \left( \Sigma_{s=1}^p E \left[ g_{i,j} g_i' \right] \bar{P} e_j / 2 - \tilde{g}' P \tilde{g} \right) / 2 \\
+ \Sigma_{s=1}^p E \left[ g_{i,j} g_i g_i' \right] P \tilde{g} \left[ \Sigma_{s=1}^p \Omega^{-1} E \left[ g_{i,j} g_i g_i' \right] \bar{P} e_j / 2 \right]' e_j / 2
\end{array} \right).
\end{aligned}$$

As  $HG = I$ ,  $E(\bar{P}) = P$ ,  $E(Q_{1,\tau}^{EL}) = 0$  at  $\rho_3 = 1$ , and  $\tilde{\psi}_\lambda = P \tilde{g}$  is uncorrelated with  $\tilde{\psi}_\theta = H \tilde{g}$ , I have  $III^{DGEL} - III^{EL} = 0$ .

If data are symmetric with  $E[g_{i,j}g_i g'_i] = 0$ ,  $III^{DGEL} - III^{EL} = H\tilde{g}Q_{1,\tau}^{EL'} / 2$ . This term together with the first term in II cancels out the remaining term in I.

Thus, if  $\rho_3 = 1$  or when data are symmetric, the sum of difference in I, II, III between the DGEL estimator and the EL estimator does not contribute to the difference in variance.

Fourth, calculate  $\sum_{s=1}^{1+2p+q} \tilde{\psi}_s \tilde{B}_s \tilde{\psi}' / 2$ .

$$\begin{aligned} \sum_{s=1}^{1+2p+q} \tilde{\psi}_s \tilde{B}_s \tilde{\psi}' / 2 &= \sum_{s=1}^{1+2p+q} \tilde{B}_s \tilde{\psi} \tilde{\psi}' e_s / 2 \\ &= \sum_{s=1}^p n^{-1/2} \sum_{i=1}^n \begin{bmatrix} 0 \\ 0 \\ \rho_3 (g_{i,j} g_i g'_i - E[g_{i,j} g_i g'_i]) \bar{P} e_j / 2 + \left( \frac{\partial(g_{i,j} g_i)}{\partial \theta'} - E \left[ \frac{\partial(g_{i,j} g_i)}{\partial \theta'} \right] \right) \bar{0} e_j / 2 \\ \left( \frac{\partial(g_{i,j} g'_i)}{\partial \theta} - E \left[ \frac{\partial(g_{i,j} g'_i)}{\partial \theta} \right] \right) \bar{P} e_j / 2 + \left( \frac{\partial \tilde{G}_j}{\partial \theta} - E \left[ \frac{\partial \tilde{G}_j}{\partial \theta} \right] \right) \bar{0} e_j / 2 \end{bmatrix} \\ &+ \sum_{s=1}^p n^{-1/2} \sum_{i=1}^n \begin{bmatrix} (\rho_3 g_{i,j} g'_i - \rho_3 E[g_{i,j} g'_i]) \bar{P} e_j / 2 \\ \rho_3 (g_{i,j} g_i g'_i - E[g_{i,j} g_i g'_i]) \bar{P} e_j / 2 + \left( \frac{\partial(g_{i,j} g_i)}{\partial \theta'} - E \left[ \frac{\partial(g_{i,j} g_i)}{\partial \theta'} \right] \right) \bar{0} e_j / 2 \\ \left( -\frac{\partial(g_{i,j} g_i)}{\partial \theta'} + E \left[ \frac{\partial(g_{i,j} g_i)}{\partial \theta'} \right] \right) \bar{0} e_j / 2 \\ 0 \end{bmatrix} \\ &+ \sum_{s=1}^q n^{-1/2} \sum_{i=1}^n \begin{bmatrix} 0 \\ \left( \left( \frac{\partial(g_i g'_i)}{\partial \theta_j} - E \left[ \frac{\partial(g_i g'_i)}{\partial \theta_j} \right] \right) \bar{0} + \left( \frac{\partial G_i}{\partial \theta_j} - E \left[ \frac{\partial G_i}{\partial \theta_j} \right] \right) \bar{\Sigma} \right) e_j / 2 \\ 0 \\ \left( \frac{\partial G'_i}{\partial \theta_j} - E \left[ \frac{\partial G'_i}{\partial \theta_j} \right] \right) \bar{0} e_j / 2 \end{bmatrix}. \end{aligned}$$

Thus,

$$IV^{DGEL} - IV^{EL} = - \begin{bmatrix} 0 \\ H' \\ H' \\ -\Sigma' \end{bmatrix}' \left( \left( \sum_{s=1}^{1+2p+q} \tilde{\psi}_s \tilde{B}_s \tilde{\psi}' / 2 \right)^{DGEL} - \left( \sum_{s=1}^{1+2p+q} \tilde{\psi}_s \tilde{B}_s \tilde{\psi}' / 2 \right)^{EL} \right)$$

$$\begin{aligned}
&= - \begin{bmatrix} 0 \\ H' \\ H' \\ -\Sigma' \end{bmatrix}' \begin{bmatrix} 0 \\ 0 \\ \sum_{s=1}^p n^{-1/2} \sum_{i=1}^n (\rho_3 - 1) (g_{i,j} g_i g_i' - E[g_{i,j} g_i g_i']) \bar{P} e_j \\ 0 \end{bmatrix} \\
&= -H \sum_{s=1}^p n^{-1/2} \sum_{i=1}^n (\rho_3 - 1) (g_{i,j} g_i g_i' - E[g_{i,j} g_i g_i']) \bar{P} e_j.
\end{aligned}$$

If  $\rho_3 = 1$ , then  $IV^{DGEL} - IV^{EL} = 0$ .

If data are symmetric, I get  $E(\tilde{\psi}'_{\theta} (IV^{DGEL} - IV^{EL})) = 0$ .

Fifth, calculate  $(\sum_{s=1}^{1+2p+q} \sum_{k=1}^{1+2p+q} \tilde{\psi}_s \tilde{\psi}_k M_{sk} \tilde{\psi} / 6)^{DGEL} - (\sum_{s=1}^{1+2p+q} \sum_{k=1}^{1+2p+q} \tilde{\psi}_s \tilde{\psi}_k M_{sk} \tilde{\psi} / 6)^{EL}$ . Since the EL and DGEL estimators have the same  $\tilde{\psi}$ , the difference is

$$\sum_{s=1}^{1+2p+q} \sum_{k=1}^{1+2p+q} \tilde{\psi}_s \tilde{\psi}_k (M_{sk}^{DGEL} - M_{sk}^{EL}) \tilde{\psi} / 6.$$

$$M_{h,l}^{DGEL} - M_{h,l}^{EL} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, h=1, l=1, \dots, 1+p;$$

$$M_{h,1+p+l}^{DGEL} - M_{h,1+p+l}^{EL} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & E\left(\frac{\partial G'_{i,l,\cdot}}{\partial \theta}\right) \end{bmatrix}, h=1, l=1, \dots, p;$$

$$M_{h,1+2p+l}^{DGEL} - M_{h,1+2p+l}^{EL} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & E\left(\frac{\partial G_i}{\partial \theta_l}\right) \\ 0 & 0 & E\left(\frac{\partial G'_i}{\partial \theta_l}\right) & 0 \end{bmatrix}, h=1, l=1, \dots, q;$$

$$M_{1+h,1+l}^{DGEL} - M_{1+h,1+l}^{EL} = - \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 6E(g_{i,h}g_{i,l}g_i g'_i) & 0 & 2E\left(\frac{\partial(g_{i,l}g_{i,h}g_i)}{\partial\theta'}\right) \\ 0 & \begin{pmatrix} 2E(g_{i,h}G'_{i,l},g'_i) \\ +2E(g_{i,l}G'_{i,h},g'_i) \\ +2E(g_{i,l}g_{i,h}G'_i) \end{pmatrix} & 0 & E\left(\frac{\partial(g_{i,h}G'_{i,l},\cdot)}{\partial\theta}\right) + E\left(\frac{\partial(g_{i,l}G'_{i,h},\cdot)}{\partial\theta}\right) \end{bmatrix},$$

$h, l = 1, \dots, p;$

$$M_{1+h,1+p+l}^{DGEL} - M_{1+h,1+p+l}^{EL} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \rho_4 E(g_{i,l}g_{i,h}g_i g'_i) & \rho_3 E\left(\frac{\partial(g_{i,l}g_{i,h}g_i)}{\partial\theta'}\right) \\ 0 & 0 & \begin{pmatrix} \rho_3 E(g_{i,h}G'_{i,l},g'_i) \\ +\rho_3 E(g_{i,l}G'_{i,h},g'_i) \\ +\rho_3 E(g_{i,l}g_{i,h}G'_i) \end{pmatrix} & \begin{pmatrix} E\left(\frac{\partial(g_{i,h}G'_{i,l},\cdot)}{\partial\theta'}\right) \\ +E\left(\frac{\partial(g_{i,l}G'_{i,h},\cdot)}{\partial\theta'}\right) \end{pmatrix} \end{bmatrix},$$

$h, l = 1, \dots, p;$

$$M_{1+h,1+2p+l}^{DGEL} - M_{1+h,1+2p+l}^{EL} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -2E\left(\frac{\partial(g_{i,h}g_i g'_i)}{\partial\theta_l}\right) & \rho_3 E\left(\frac{\partial(g_{i,h}g_i g'_i)}{\partial\theta_l}\right) & 0 \\ 0 & -E\left(\frac{\partial(G'_i g_{i,h} + G'_{i,h}, g'_i)}{\partial\theta_l}\right) & E\left(\frac{\partial(G'_i g_{i,h} + G'_{i,h}, g'_i)}{\partial\theta_l}\right) & 0 \end{bmatrix},$$

$h = 1, \dots, p, l = 1, \dots, q;$

$$M_{1+p+h,1+p+l}^{DGEL} - M_{1+p+h,1+p+l}^{EL} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \rho_4 E(g_{i,l} g_{i,h} g_i g'_i) & 0 & 0 \\ 0 & \begin{pmatrix} \rho_3 E(g_{i,h} G'_{i,l}, g'_i) \\ + \rho_3 E(g_{i,l} G'_{i,h}, g'_i) \\ + \rho_3 E(g_{i,l} g_{i,h} G'_i) \end{pmatrix} & \begin{pmatrix} \rho_3 E(g_{i,h} G'_{i,l}, g'_i) \\ + \rho_3 E(g_{i,l} G'_{i,h}, g'_i) \\ + \rho_3 E(g_{i,l} g_{i,h} G'_i) \end{pmatrix} & \begin{pmatrix} E\left(\frac{\partial(g_{i,h} G'_{i,l}, \cdot)}{\partial \theta'}\right) \\ + E\left(\frac{\partial(g_{i,l} G'_{i,h}, \cdot)}{\partial \theta'}\right) \end{pmatrix} \end{bmatrix},$$

$$h, l = 1, \dots, p;$$

$$M_{1+p+h,1+2p+l}^{DGEL} - M_{1+p+h,1+2p+l}^{EL} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \rho_3 E\left(\frac{\partial(g_{i,h} g_i g'_i)}{\partial \theta_l}\right) & 0 & 0 \\ E\left(\frac{\partial G'_{i,h}, \cdot}{\partial \theta_l}\right) & E\left(\frac{\partial(g_{i,h} G'_i + G'_{i,h}, g'_i)}{\partial \theta_l}\right) & -E\left(\frac{\partial(g_{i,h} G'_i + G'_{i,h}, g'_i)}{\partial \theta_l}\right) & 0 \end{bmatrix},$$

$$h = 1, \dots, p, l = 1, \dots, q;$$

$$M_{1+2p+h,1+2p+l}^{DGEL} - M_{1+2p+h,1+2p+l}^{EL} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ E\left(\frac{\partial^2 g_i}{\partial \theta_h \partial \theta_l}\right) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, h, l = 1, \dots, q.$$

Therefore,

$$\begin{aligned} & \sum_{s=1}^{1+2p+q} \sum_{k=1}^{1+2p+q} \tilde{\psi}_s \tilde{\psi}_k (M_{sk}^{DGEL} - M_{sk}^{EL}) \tilde{\psi} / 6 \\ & = -2 \sum_{h=1}^p \tilde{\psi}_\tau \tilde{\psi}_{\lambda+h} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & E\left(\frac{\partial G'_{i,h}, \cdot}{\partial \theta'}\right) \end{bmatrix} \begin{bmatrix} 0 \\ P\tilde{g} \\ P\tilde{g} \\ H\tilde{g} \end{bmatrix} \end{aligned}$$

$$\begin{aligned}
& -2\sum_{h=1}^q \tilde{\psi}_\tau \tilde{\psi}_{\theta+h} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & E\left(\frac{\partial G_i}{\partial \theta_h}\right) \\ 0 & 0 & E\left(\frac{\partial G'_i}{\partial \theta_h}\right) & 0 \end{bmatrix} \begin{bmatrix} 0 \\ P\tilde{g} \\ P\tilde{g} \\ H\tilde{g} \end{bmatrix} \\
& + \sum_{h=1}^p \sum_{l=1}^p \tilde{\psi}_{\kappa+l} \tilde{\psi}_{\kappa+h} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 6E(g_{i,h}g_{i,l}g_i g'_i) & 0 & 2E\left(\frac{\partial(g_{i,l}g_{i,h}g_i)}{\partial \theta'}\right) \\ 0 & \begin{pmatrix} 2E(g_{i,h}G'_{i,l},g'_i) \\ +2E(g_{i,l}G'_{i,h},g'_i) \\ +2E(g_{i,l}g_{i,h}G'_i) \end{pmatrix} & 0 & \begin{pmatrix} E\left(\frac{\partial(g_{i,h}G'_{i,l},\cdot)}{\partial \theta'}\right) \\ +E\left(\frac{\partial(g_{i,l}G'_{i,h},\cdot)}{\partial \theta'}\right) \end{pmatrix} \end{bmatrix} \begin{bmatrix} 0 \\ P\tilde{g} \\ P\tilde{g} \\ H\tilde{g} \end{bmatrix} \\
& -2\sum_{h=1}^p \sum_{l=1}^p \tilde{\psi}_{\kappa+l} \tilde{\psi}_{\lambda+h} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \rho_4 E(g_{i,l}g_{i,h}g_i g'_i) & \rho_3 E\left(\frac{\partial(g_{i,l}g_{i,h}g_i)}{\partial \theta'}\right) \\ 0 & 0 & \begin{pmatrix} \rho_3 E(g_{i,h}G'_{i,l},g'_i) \\ +\rho_3 E(g_{i,l}G'_{i,h},g'_i) \\ +\rho_3 E(g_{i,l}g_{i,h}G'_i) \end{pmatrix} & \begin{pmatrix} E\left(\frac{\partial(g_{i,h}G'_{i,l},\cdot)}{\partial \theta'}\right) \\ +E\left(\frac{\partial(g_{i,l}G'_{i,h},\cdot)}{\partial \theta'}\right) \end{pmatrix} \end{bmatrix} \begin{bmatrix} 0 \\ P\tilde{g} \\ P\tilde{g} \\ H\tilde{g} \end{bmatrix} \\
& -2\sum_{h=1}^p \sum_{l=1}^q \tilde{\psi}_{\kappa+h} \tilde{\psi}_{\theta+l} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -2E\left(\frac{\partial(g_{i,h}g_i g'_i)}{\partial \theta_l}\right) & \rho_3 E\left(\frac{\partial(g_{i,h}g_i g'_i)}{\partial \theta_l}\right) & 0 \\ 0 & -E\left(\frac{\partial(G'_i g_{i,h} + G'_{i,h}, g'_i)}{\partial \theta_l}\right) & E\left(\frac{\partial(G'_i g_{i,h} + G'_{i,h}, g'_i)}{\partial \theta_l}\right) & 0 \end{bmatrix} \begin{bmatrix} 0 \\ P\tilde{g} \\ P\tilde{g} \\ H\tilde{g} \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
& -\sum_{h=1}^p \sum_{l=1}^p \tilde{\psi}_{\lambda+h} \tilde{\psi}_{\lambda+l} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \rho_4 E(g_{i,l} g_{i,h} g_i g'_i) & 0 & 0 \\ \left( \begin{array}{c} \rho_3 E(g_{i,h} G'_{i,l}, g'_i) \\ + \rho_3 E(g_{i,l} G'_{i,h}, g'_i) \\ + \rho_3 E(g_{i,l} g_{i,h} G'_i) \end{array} \right) & - \left( \begin{array}{c} \rho_3 E(g_{i,h} G'_{i,l}, g'_i) \\ + \rho_3 E(g_{i,l} G'_{i,h}, g'_i) \\ + \rho_3 E(g_{i,l} g_{i,h} G'_i) \end{array} \right) & - \left( \begin{array}{c} E \left( \frac{\partial g_{i,h} G'_{i,l}}{\partial \theta'} \right) \\ + E \left( \frac{\partial g_{i,l} G'_{i,h}}{\partial \theta'} \right) \end{array} \right) \end{bmatrix} \begin{bmatrix} 0 \\ P\tilde{g} \\ P\tilde{g} \\ H\tilde{g} \end{bmatrix} \\
& -2\sum_{h=1}^p \sum_{l=1}^q \tilde{\psi}_{\lambda+h} \tilde{\psi}_{\theta+l} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \rho_3 E \left( \frac{\partial (g_{i,h} g_i g'_i)}{\partial \theta_l} \right) & 0 & 0 \\ E \left( \frac{\partial G'_{i,h}}{\partial \theta_l} \right) & E \left( \frac{\partial \left( \begin{array}{c} g_{i,h} G'_i \\ + G'_{i,h}, g'_i \end{array} \right)}{\partial \theta_l} \right) & -E \left( \frac{\partial \left( \begin{array}{c} g_{i,h} G'_i \\ + G'_{i,h}, g'_i \end{array} \right)}{\partial \theta_l} \right) & 0 \end{bmatrix} \begin{bmatrix} 0 \\ P\tilde{g} \\ P\tilde{g} \\ H\tilde{g} \end{bmatrix} \\
& -\sum_{h=1}^q \sum_{l=1}^q \tilde{\psi}_{\theta+h} \tilde{\psi}_{\theta+l} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ E \left( \frac{\partial^2 g_i}{\partial \theta_h \partial \theta_l} \right) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ P\tilde{g} \\ P\tilde{g} \\ H\tilde{g} \end{bmatrix}.
\end{aligned}$$

Since  $\tilde{\psi} = \begin{bmatrix} 0 \\ P\tilde{g} \\ P\tilde{g} \\ H\tilde{g} \end{bmatrix}$ ,  $\tilde{\psi}_\tau = 0$ ,  $\tilde{\psi}_{\kappa+h} = \tilde{\psi}_{\lambda+h} = (P\tilde{g})_h$ , and  $(\tilde{\psi}_\theta)_{\theta+h} = H\tilde{g}$ ,

$$V^{DGEL} - V^{EL} = \begin{bmatrix} 0 \\ H' \\ H' \\ -\Sigma' \end{bmatrix}' \left( \sum_{s=1}^{1+2p+q} \sum_{k=1}^{1+2p+q} \tilde{\psi}_s \tilde{\psi}_k (M_{sk}^{DGEL} - M_{sk}^{EL}) \tilde{\psi} / 6 \right)$$

$$= \left( \begin{array}{c} \left( \begin{array}{c} (\rho_4 / 2 - 1) HE(g_{i,h} g_{i,l} g_i g'_i) P\tilde{g} \\ + (\rho_3 / 3 - 1 / 3) HE \left( \frac{\partial(g_{i,l} g_{i,h} g_i)}{\partial \theta'} \right) H\tilde{g} \\ + (1 / 3 - \rho_3 / 3) \Sigma \left( \begin{array}{c} E(g_{i,h} G'_{i,l} \cdot g'_i) \\ + E(g_{i,l} G'_{i,h} \cdot g'_i) + E(g_{i,l} g_{i,h} G'_i) \end{array} \right) P\tilde{g} \end{array} \right) \\ + (2\rho_3 / 3 - 2 / 3) \Sigma_{h=1}^p \Sigma_{l=1}^q (P\tilde{g})_h (H\tilde{g})_l HE \left( \frac{\partial(g_{i,h} g_i g'_i)}{\partial \theta_l} \right) P\tilde{g} \end{array} \right).$$

If  $\rho_3 = 1$ , then

$$V^{DGEL} - V^{EL} = H \Sigma_{h=1}^p \Sigma_{l=1}^q \tilde{\psi}_{\kappa+l} \tilde{\psi}_{\kappa+h} (\rho_4 / 2 - 1) E(g_{i,l} g_{i,h} g_i g'_i) P\tilde{g}.$$

Observe that  $V$  is a function of  $\tilde{\psi}_\lambda = P\tilde{g}$  which is uncorrelated with  $\tilde{\psi}_\theta$ . Thus, the difference in  $V$  does not influence  $E(\tilde{\psi}_\theta \tilde{Q}'_2)$ .

If data are symmetric, because  $P\tilde{g}$  is uncorrelated with  $\tilde{\psi}_\theta$  and  $PG = 0$ ,

$$E\left(\tilde{g}'H(V^{DGEL} - V^{EL})\right) = E\left(\tilde{g}'H \left( \begin{array}{c} (2\rho_3 - 2) \Sigma_{h=1}^p \Sigma_{l=1}^q (P\tilde{g})_h (P\tilde{g})_l HE \left( \frac{\partial(g_{i,l} g_{i,h} g_i)}{\partial \theta'} \right) H\tilde{g} \\ + (4\rho_3 - 4) \Sigma_{h=1}^p \Sigma_{l=1}^q (P\tilde{g})_h (H\tilde{g})_l HE \left( \frac{\partial(g_{i,h} g_i g'_i)}{\partial \theta_l} \right) P\tilde{g} \end{array} \right) \right) = 0.$$

To sum up, any estimator with  $\rho_3 = 1$  has a  $\tilde{\psi}$  term with an  $O_p(n^{-1})$  difference from that of the EL estimator. This  $O_p(n^{-1})$  difference only contributes to the  $o_p(n^{-1})$  difference in bias and  $o_p(n^{-2})$  difference in variance. The  $\tilde{Q}_1$  and  $\tilde{Q}_2$  terms for any estimator with  $\rho_3 = 1$  are also different from that of the EL estimator. However, these differences in  $\tilde{Q}_1$  and  $\tilde{Q}_2$  are uncorrelated to  $\tilde{\psi}$ . Therefore, the DET estimator or any estimator with  $\rho_3 = 1$  shares the same  $O_p(n^{-2})$  variance with the EL estimators. Similarly, if data are symmetrically distributed, all DGEL estimators share the same  $O_p(n^{-2})$  variance with the EL estimators.



Proof of Corollary 2.1

Higher-order Bias: from Theorem 2.2, the bias of the DGEL estimators are

$$\begin{aligned} E[Q_{1\theta}^{DGEL}] &= H \left[ E[G_i H g_i] - \sum_{j=1}^q E \left[ \frac{\partial G_i}{\partial \theta_j} \right] \Sigma e_j / 2 \right] + (1 - \rho_3) H E[g_i g_i' P g_i] \\ &= B_I + (1 - \rho_3) B_\Omega. \end{aligned}$$

If data are symmetric ( $E[g_j g g'] = 0$ ), then  $E[Q_{1\theta}^{DGEL}] = B_I$ .

Higher-order Efficiency: from Theorem 2.3 and its proof, the difference between  $Q_{1,\theta}^{DGEL}$  and  $Q_{1,\theta}^{EL}$  is

$$Q_{1,\theta}^{DGEL} - Q_{1,\theta}^{EL} = H \Sigma_{i=1}^n (1 - \rho_3) E[g_j g g'] \bar{P} e_j + o_p(n^{-1/2}),$$

where  $\bar{P} = P \tilde{g} \tilde{g}' P$ . This difference is  $o_p(n^{-1/2})$  if  $E[g_{i,j} g_i g_i'] = 0$ . Similarly, the difference between  $Q_{2,\theta}^{DGEL}$  and  $Q_{2,\theta}^{EL}$  are in terms of  $E[g_{i,j} g_i g_i']$ . When  $E[g_{i,j} g_i g_i'] = 0$ , this difference becomes  $o_p(n^{-1/2})$ . Thus, the difference in variance between  $\hat{\theta}^{DGEL}$  and  $\hat{\theta}^{EL}$  is  $o_p(n^{-2})$  when data are symmetrically distributed.

Proof of Corollary 2.2

With  $\rho^B(v)$  in the objective function and  $\rho^A(v)$  in the constraint, the augmented FOCs of the DGEL estimators are  $n^{-1} \Sigma_{i=1}^n m_i(\phi) = 0$ , where  $\phi = (\tau, \kappa', \lambda', \theta')'$ ,  $m_i(\phi)$  is

$$m_i(\phi) = \begin{bmatrix} \dot{\tau} - \tau \\ \rho_1^A(\lambda' g_i) g_i \\ (\tau - \dot{\tau}) g_i + \rho_2^A(\lambda' g_i) g_i g_i' \kappa \\ \left[ \rho_2^A G_i' \lambda g_i' + \rho_1^A G_i \right] \kappa + (\tau - \dot{\tau}) G_i' \lambda \end{bmatrix},$$

and  $\dot{\tau} = \rho_1^B(\lambda'(g_i - \hat{g}))$ ,  $\tau = \Sigma_{i=1}^n \dot{\tau} / n$ , and  $\kappa = \left[ \Sigma_{i=1}^n \rho_2^A g_i g_i' \right]^{-1} \left[ \Sigma_{i=1}^n (\dot{\tau} - \tau) g_i \right]$ .

Following the same procedure as in proving Theorem 2.2,

$$\begin{aligned} E[Q_{1,\theta}] &= H \left[ E[G_i H g_i] - \sum_{j=1}^q E \left[ \frac{\partial G_i}{\partial \theta_j} \right] \Sigma e_j / 2 \right] + (1 + \rho_3 / 2 - 3\rho_3^* / 2) H E[g_i g_i' P g_i] \\ &\quad + \Sigma \left[ -E[G_i P g_i] + E \left[ \sum \left( E[G_j g_j'] + E[G_j' g_j] \right) P e_j / 2 \right] \right]. \end{aligned}$$

Since  $\Sigma_{i=1}^n \left( E[G_j' g_j'] + E[G_j' g_j] \right) P e_j / 2 = E[G_i P g_i]$ ,

$$E[Q_{1,\theta}] = B_I + \frac{2 + \rho_3^B - 3\rho_3^A}{2} B_\Omega.$$

## FOCs of the GEEL estimators

The first step of estimating the GEEL estimators is to solve  $\pi_i(\theta)$ . Transform the power-divergence form into GEL form, and obtain  $\pi_i(\theta) = \frac{\rho_1(\lambda' g_i(\theta))}{\sum_{i=1}^n \rho_1(\lambda' g_i(\theta))}$ , where the  $\rho(\bullet)$  is in the GEL form, i.e., EL with  $\rho(v) = \ln(1-v)$ , ET with  $\rho(v) = -e^v$  and CUE with  $\rho(v) = -(1+v)^2/2$ . Replace  $\pi_i(\theta)$  in the second step, and the FOCs w.r.t  $\theta$  becomes

$$\sum_{i=1}^n \left( \frac{n\rho_2(\lambda' g_i(\theta))}{\sum_{i=1}^n \rho_1(\lambda' g_i(\theta))} - \frac{\rho_2(\lambda' g_i(\theta))}{\rho_1(\lambda' g_i(\theta))} \right) \left( \frac{\partial \lambda'}{\partial \theta} g_i(\theta) + \lambda' G_i(\theta) \right) = 0,$$

where  $\frac{\partial \lambda'}{\partial \theta}$  is derived from  $\sum_{i=1}^n \rho_1(\lambda' g_i(\theta)) g_i(\theta) = 0$ . Therefore, the FOCs of the GEEL estimators are

$$\begin{cases} \sum_{i=1}^n \left( \frac{n\rho_2(\lambda' g_i(\theta))}{\sum_{i=1}^n \rho_1(\lambda' g_i(\theta))} - \frac{\rho_2(\lambda' g_i(\theta))}{\rho_1(\lambda' g_i(\theta))} \right) \left( \frac{\partial \lambda'}{\partial \theta} g_i(\theta) + \lambda' G_i(\theta) \right) = 0 \\ \sum_{i=1}^n \rho_1(\lambda' g_i(\theta)) g_i(\theta) = 0 \end{cases}.$$

Specifically, with  $\pi_i(\theta)$  derived from the EL problem, the FOCs of the corresponding GEEL estimators are

$$\begin{cases} \sum_{i=1}^n \frac{1}{1 - \lambda' g_i(\theta)} G_i(\theta)' \lambda = 0 \\ \sum_{i=1}^n \frac{1}{1 - \lambda' g_i(\theta)} g_i(\theta) = 0 \end{cases},$$

which is the same as the EL estimator in the GEL group.

With  $\pi_i(\theta)$  derived from the CUE problem, the FOCs of the corresponding GEEL estimators are

$$\begin{cases} \sum_{i=1}^n \left( 1 - \frac{1}{1 + \lambda' g_i(\theta)} \right) \left( \frac{\partial \lambda'}{\partial \theta} g_i(\theta) + G_i(\theta)' \lambda \right) = 0 \\ \sum_{i=1}^n (1 + \lambda' g_i(\theta)) g_i(\theta) = 0 \end{cases}.$$

With  $\pi_i(\theta)$  derived from the ET problem, the FOCs of the corresponding GEEL estimators are

$$\begin{cases} \sum_{i=1}^n (1 - n\pi_i^{ET}) \left( \frac{\partial \lambda'}{\partial \theta} g_i(\theta) + \lambda' G_i(\theta) \right) = 0 \\ \sum_{i=1}^n e^{(\lambda' g_i(\theta))} g_i(\theta) = 0 \end{cases}$$

where  $\pi_i^{ET} = \frac{e^{(\lambda' g_i(\theta))}}{\sum_{i=1}^n e^{(\lambda' g_i(\theta))}}$ .

## Bias of the GEEL estimators-example of CU-GEEL estimator

The augmented FOCs of CU-GEEL estimators are  $n^{-1}\sum_{i=1}^n m_i(\phi) = 0$ , where  $\phi = (\tau, \kappa', \lambda', \theta')'$ ,  $m_i(\phi)$  is

$$m_i(\phi) = \begin{bmatrix} \frac{1}{1 + \lambda' g_i} - \tau \\ (1 + \lambda' g_i) g_i \\ \frac{1}{1 + \lambda' g_i} g_i + g_i g_i' \kappa \\ \left[ G_i' \lambda g_i' + (1 + \lambda' g_i) G_i' \right] \kappa + \frac{1}{1 + \lambda' g_i} G_i' \lambda \end{bmatrix},$$

where  $\tau = \sum_{i=1}^n \left( \frac{1}{1 + \lambda' g_i(\theta)} \right) / n$ , and  $\kappa = - \left[ \sum_{i=1}^n g_i g_i' \right]^{-1} \left[ \sum_{i=1}^n \frac{1}{1 + \lambda' g_i} g_i \right]$ .

Following the same procedure as in proving Theorem 2.2,

$$E \left[ Q_{1,\theta}^{CU-GEEL} \right] = H \left[ E \left[ G_i H g_i \right] - \sum_{j=1}^q E \left[ \frac{\partial G_i}{\partial \theta_j} \right] \Sigma e_j / 2 \right] \\ + \Sigma \left[ -E \left[ G_i P g_i \right] + E \left[ \sum \left( E \left[ G_j g_j' \right] + E \left[ G_j' g_j \right] \right) P e_j / 2 \right] \right],$$

since  $\sum_{i=1}^n \left( E \left[ G_j' g_j' \right] + E \left[ G_j' g_j \right] \right) P e_j / 2 = E \left[ G_i P g_i \right]$ , then  $E \left[ Q_{1,\theta}^{CU-GEEL} \right] = B_I$ .

APPENDIX III  
COPYRIGHT

This dissertation includes some of my own work and the co-authored work with Dr. Seung Ahn. The first chapter is reproduced with permission from Dr. Seung Ahn.