THE PERFORMANCE OF EMPIRICAL LIKELIHOOD AND ITS GENERALIZATIONS

GUIDO W. IMBENS AND RICHARD H. SPADY

University of California, Berkeley and Nuffield College, Oxford

Abstract. We calculate higher-order asymptotic biases and mean squared errors (MSE) for a simple model with a sequence of moment conditions. In this setup, generalized empirical likelihood (GEL) and infeasible optimal GMM (OGMM) have the same higher-order biases, with GEL having an MSE that exceeds OGMM's by an additional term of order $(M-1)/N$, i.e. the degree of overidentification divided by sample size. In contrast, any 2-step GMM estimator has an additional bias relative to OGMM of order $(M-1)/N$ and an additional MSE of order $(M-1)^2/N$. Consequently GEL must be expected to dominate 2-step GMM. In our simple model all GEL's have equivalent next higher order behavior because generalized third moments of moment conditions are assumed to be zero; we explore, in further analysis and simulations, the implications of dropping this assumption.

1. Introduction

This paper has two parts. In the first part, we calculate higher-order asymptotic biases and mean squared errors (MSE) for a simple model with a sequence of moment conditions. In this setup, generalized empirical likelihood (GEL) and infeasible optimal GMM (OGMM) have the same higher-order biases, with GEL having an MSE that exceeds OGMM's by an additional term of order $(M-1)/N$, i.e. the degree of overidentification divided by sample size. In contrast, any 2-step GMM estimator has an additional bias relative to OGMM of order $(M-1)/N$ and an additional MSE of order $(M-1)^2/N$. Although these features do depend on the simple framework we have adopted, we cannot see how a more complicated framework will rescue 2-step GMM from these fundamental difficulties. Consequently, we conclude GEL must be expected to dominate 2-step GMM, and our interest shifts to distinguishing between variants of GEL and the closely related, (if not dual) empirical discrepancy (ED) estimators.

In our simple model all GEL's have equivalent next higher–order behavior because generalized third moments of moment conditions (i.e. products of the form $\psi_j\psi_k\psi_\ell$, where any or all of j, k, and ℓ may be equal) are assumed to be zero. We explore, in further analysis

Date: July 30, 2001.

Key words and phrases. Keywords.

and simulations, the implications of dropping this assumption. We find that when third moments are important, one variant of GEL/MD, which can be identified with 'continuously updated' GMM, is inferior to many other variants, including 'exponential tilting' (ET) and empirical likelihood (EL).

2. Framework

Consider a sequence of independent and identically distributed pairs of random vectors $\{(v_i, w_i)\}_{i=1}^N$. The dimension of v_i and w_i is $M \geq 1$. We are interested in a scalar parameter θ, satisfying

$$E[\psi(v_i, w_i, \theta)] = 0,$$

for $i = 1, \ldots, N$, where

$$\psi(v_i, w_i, \theta) = (v_i + e_1) \cdot \theta - w_i = \begin{pmatrix} (v_{i1} + 1) \cdot \theta - w_{i1} \\ v_{i2} \cdot \theta - w_{i2} \\ \vdots \\ v_{iM} \cdot \theta - w_{iM} \end{pmatrix},$$

and e_1 is an $M-$vector with the first element equal to one and the other elements equal to zero.

We are interested in the properties of various estimators for θ as the degree of overidentification $(M-1)$, increases. Following Donald and Newey (2000) who look at the behavior of various instrumental variables estimators as the number of instruments increases, and Newey and Smith (2001) who look at bias of GEL and GMM estimators, we look at the leading terms in the asymptotic expansion of the estimators and consider the rate at which the moments of these terms increase with M.

We make the following simplifying assumptions. The pairs (v_{im}, w_{im}) and (v_{jn}, w_{jn}) are independent if either $i \neq j$ or $n \neq m$ (or both), and have the same distribution. Let $\mu_{rp} = E[v_{im}^r \cdot w_{im}^p]$ denote the moments of this distribution. Moments up to order $p + r \leq 6$ are assumed to be finite. Without essential loss of generality, let $\mu_{10} = \mu_{01} = 0$, implying the true value of θ is $\theta^* = 0$, let $\mu_{20} = \mu_{02} = 1$, and let $\mu_{11} = \rho$ be the correlation coefficient of v_{im} and w_{im}.

Let

$$\overline{v} = \frac{1}{N} \sum_{i=1}^N v_i,$$

$$\overline{w} = \frac{1}{N} \sum_{i=1}^N w_i,$$

$$\overline{ww'} = \frac{1}{N}\sum_{i=1}^{N} w_i w_i',$$

denote sample averages, let $\overline{v}_j$ and $\overline{w}_j$ denote the jth element of $\overline{v}$ and $\overline{w}$ respectively, and let $\overline{ww'}_{ij}$ denote the (i,j)th element of $\overline{ww'}$.

Denote the optimal, infeasible, gmm estimator by

$$\hat{\theta}_{opt} = e_1'\overline{w} = \overline{w}_1.$$

Then,

$$\hat{\theta}_{opt} - \theta^* = O_p(N^{-1/2}),$$

$$E[\hat{\theta}_{opt}] - \theta^* = 0,$$

and

$$E[(\hat{\theta}_{opt} - \theta^*)^2] = 1/N.$$

The mean squared error for this estimator does not depend on the number of moments, as in fact increasing M does not affect the numerical value.

Lemma 1. (Expansion Optimal GMM Estimator)

$$\hat{\theta}_{opt} = \overline{w}_1 - \overline{w}_1\overline{v}_1 + o_p(1/N).$$

Proof: See Appendix.

Define
(i), $T_1 = \overline{w}_1$,
(ii) $R_1 = -\overline{w}_1\overline{v}_1$,
so that $\hat{\theta}_{opt} = T_1 + R_1 + o_p(1/N)$.

Lemma 2. (Bias of $\hat{\theta}_{opt}$)
The bias of the leading terms is

$$E[T_1 + R_1 - \theta^*] = -\rho/N.$$

Proof: See Appendix.

Lemma 3. (Mean-squared-error of $\hat{\theta}_{opt}$)
The mean-squared-error of the leading terms is

$$E[(T_1 + R_1 - \theta^*)^2] = 1/N - 2\mu_{12}/N^2 + (1 + 2\rho^2)/N^2 + o_p(1/N^2).$$

Proof: See Appendix.

3. Two-step GMM Estimator

The first estimator we consider is the standard two-step generalized method of moments estimator (GMM), due to Hansen (1984). Consider a generic gmm estimator, defined as the minimand of

$$\left(\frac{1}{N} \sum_{i=1}^{N} \psi(v_i, w_i, \theta) \right)' \cdot C \cdot \left(\frac{1}{N} \sum_{i=1}^{N} \psi(v_i, w_i, \theta) \right).$$

We focus here on the efficient gmm estimator, with the choice for the weight matrix C equal to

$$\left(\frac{1}{N} \sum_{i=1}^{N} \psi(v_i, w_i, \theta^*) \cdot \psi(v_i, w_i, \theta^*)' \right)^{-1} = (\overline{ww'})^{-1}.$$

Thus the gmm objective function is

$$\left(\frac{1}{N} \sum_{i=1}^{N} \psi(v_i, w_i, \theta) \right)' \cdot (\overline{ww'})^{-1} \cdot \left(\frac{1}{N} \sum_{i=1}^{N} \psi(v_i, w_i, \theta) \right)$$

$$= \Big((\overline{v} + e_1) \cdot \theta - \overline{w} \Big)' \cdot (\overline{ww'})^{-1} \cdot \Big((\overline{v} + e_1) \cdot \theta - \overline{w} \Big).$$

The first order condition for the gmm estimator is

$$0 = 2(\overline{v} + e_1)' \cdot (\overline{ww'})^{-1} \cdot ((\overline{v} + e_1) \cdot \theta - \overline{w}),$$

with the solution for the gmm estimator equal to

$$\hat{\theta}_{gmm} = \Big((\overline{v} + e_1)' \cdot (\overline{ww'})^{-1} \cdot (\overline{v} + e_1) \Big)^{-1} \cdot \Big((\overline{v} + e_1)' \cdot (\overline{ww'})^{-1} \cdot \overline{w} \Big).$$

The goal is to approximate this estimator up to terms of order $O_p(1/N)$ and evaluate the mean squared error of this approximation. In particular the terms whose moments depend on M are of interest, and specifically how fast the mean squared error increases with the number of excess moments.

Lemma 4. (Expansion of $\hat{\theta}_{gmm}$)

$$\hat{\theta}_{gmm} = \overline{w}_1 - 2\overline{v}_1 \overline{w}_1 + (\overline{ww'}_{11} - 1) \cdot \overline{w}_1 - e_1' \left(\overline{ww'} - \mathcal{I}_M \right) \overline{w} + \overline{v}' \overline{w} + o_p(1/N).$$

Proof: See Appendix.

Now define
(i), $T_1 = \overline{w}_1$,
(ii), $R_1 = -2\overline{v}_1\overline{w}_1$,
(iii), $R_2 = (\overline{ww'}_{11} - 1) \cdot \overline{w}_1$,
(iv), $R_3 = -e_1' \left(\overline{ww'} - \mathcal{I}_M\right) \overline{w}$,
(v), $R_4 = \overline{v}'\overline{w}$,
so that

$$\hat{\theta}_{gmm} = T_1 + R_1 + R_2 + R_3 + R_4 + o_p(1/N).$$

(Here, and throughout the paper we use T_j for terms of order $O_p(1/\sqrt{N})$ and R_j for terms of order $O_p(1/N)$.) For the bias and mean-squared-error of the GMM estimator we therefore investigate the moments of $S = T_1 + R_1 + R_2 + R_3 + R_4 - \theta^*$.

Lemma 5. (Bias of $\hat{\theta}_{gmm}$)
The expectation of the leading terms is

$$E[T_1 + R_1 + R_2 + R_3 + R_4 - \theta^*] = -\rho/N + \rho(M-1)/N.$$

Proof: See Appendix.

Lemma 6. (Mean-squared-error of $\hat{\theta}_{gmm}$)
The mean-squared-error of the leading terms is

$$E\left[(T_1 + R_1 + R_2 + R_3 + R_4 - \theta^*)^2\right] = 1/N - 2\mu_{12}/N^2 + (1+2\rho^2)/N^2$$

$$+\rho^2(M-1)^2/N^2 - \rho^2(M-1)/N^2 + 2(M-1)/N^2 + o_p(1/N^2).$$

Proof: See Appendix.

Note that the difference between the mse for $\hat{\theta}_{gmm}$ and $\hat{\theta}_{opt}$ is in the last three terms, $\rho^2(M-1)^2/N^2 - \rho^2(M-1)/N^2 + 2(M-1)/N^2$. All three of these are non-negative. If $M = 1$, they vanish as the optimal gmm estimator and feasible gmm estimator coincide.

4. Generalized Empirical Likelihood Estimators

In this section we consider alternatives to the standard two-step GMM estimators. The estimators consider include empirical likelihood (Qin and Lawless, 1984; Imbens, 1997), exponential tilting (Imbens, Spady and Johnson, 1997; Kitamura and Stutzer, 1997), and the continuously updating estimator (Hansen, Heaton and Yaron, 1996). The specific class of estimators we consider is related to that of the Cressie-Read family (e.g., Baggerly (), and

Corcoran ()), as well to that of the generalized empirical likelihood estimators, introduced by Smith (1997). For a given function $g(a)$, normalized to satisfy $g(0) = 1$, $g'(0) = 1$, and $g''(0) = \lambda$, the estimator for θ is defined through the system of equations

$$0 = \sum_{i=1}^{N} \psi(v_i, w_i, \theta) \cdot g(t'\psi(v_i, w_i, \theta)),$$

$$0 = \sum_{i=1}^{N} t' \frac{\partial \psi}{\partial \theta'}(v_i, w_i, \theta) \cdot g(t'\psi(v_i, w_i, \theta)),$$

solved as a function of θ and t. The leading choices for $g(a)$ are $g(a) = 1/(1-a)$ (empirical likelihood), $g(a) = \exp(a)$ (exponential tilting), and $g(a) = 1 + a$ (continuously updating).

Under standard conditions, the solution for t, denoted by $\hat{t}_\lambda$ converges to a vector of zeros, $\hat{\theta}_\lambda$ converges to θ^*, and

$$\hat{t}_\lambda = O_p(1/\sqrt{N}),$$

$$\hat{\theta}_\lambda = O_p(1/\sqrt{N}).$$

The choice of $g(a)$ does not matter for the standard large sample distribution, and

$$\hat{t}_{\lambda_1} - \hat{t}_{\lambda_2} = o_p(1/\sqrt{N}),$$

and

$$\hat{\theta}_\lambda - \hat{\theta}_{opt} = o_p(1/\sqrt{N}).$$

Lemma 7. (Expansion for $\hat{\theta}_\lambda$)

$$\hat{\theta}_\lambda = \overline{w}_1 + (\overline{ww'}_{11} - 1)\overline{w}_1 - e_1'(\overline{ww'} - \mathcal{I}_M)\overline{w}$$
$$+\overline{w}'\overline{v} - 2\overline{w}_1\overline{v}_1 - \rho\overline{w}'\overline{w} + \rho\overline{w}_1^2 + o_p(1/N).$$

Proof: See Appendix.

Note that the choice of λ in the family of generalized empirical likelihood estimators does not matter for the $O(1/N^2)$ term. This is special to our case. It relies on the fact that the first and other moments are independent. In general with a scalar parameter one can always renormalize the moments in such a way that only the derivative of the first moment is correlated with the parameters of interest, and that in addition the other moments are uncorrelated with the first one. This does not make the first and other moments independent, however, and the equivalence result here depends on the cross moments of the type $E[\psi_1\psi_2^2]$ being equal to zero.

Now define
(i), $T_1 = \overline{w}_1$,
(ii), $R_1 = -2\overline{v}_1\overline{w}_1$,
(iii), $R_2 = \overline{w}_1(\overline{ww'}_{11} - 1)$,
(iv), $R_3 = -\overline{w}'(\overline{ww'} - \mathcal{I}_M)e_1$,
(v), $R_4 = \overline{w}'\overline{v}$,
(vi), $R_5 = \rho\overline{w}_1^2$,
(vii), $R_6 = -\rho\overline{w}'\overline{w}$.

Lemma 8. (Bias of $\hat{\theta}_\lambda$)
The expectation of the leading terms is

$$E[T_1 + R_1 + R_2 + R_3 + R_4 + R_5 + R_6 - \theta^*] = -\rho/N + o_p(1/N).$$

Proof: See Appendix.

Lemma 9. (Mean-squared-error of $\hat{\theta}_\lambda$)
The mean-squared-error of the leading terms is

$$E\left[(T_1 + R_1 + R_2 + R_3 + R_4 + r_5 + R_6 - \theta^*)^2\right] = 1/N - 2\mu_{12}/N^2 + (1 + 2\rho^2)/N^2$$
$$+\rho^2(M-1)/N^2 + 2(M-1)/N^2 + o_p(1/N^2).$$

Proof: See Appendix.

The difference in mse between this estimator and the feasible gmm estimator $\hat{\theta}_{gmm}$ is $\rho^2(M-1)(M-2)$. For $M = 1$ and $M = 2$ the two mse's agree up to order $O(1/N^2)$, but for higher degrees of over-identification the empirical likelihood estimator dominate in terms of mse. Note that although the bias of the empirical likelihood estimators does not increase with the degree of over-identification, the mse does.

3. Empirical Discrepancy Theory

Having established that generalized empirical likelihood offers 'asymptotic resistance' to the deterioration of estimation efficiency as moment conditions are added, we turn to analyzing *differences* between members of this class. To do this, it will be helpful to interpret these estimators from the point of view of empirical discrepancy (ED, also sometimes called minimum discrepancy) theory, as found in the statistics literature in Corcoran (1994,1998), and Baggerly (1998). We modify and extend some of the previous notation in order to deal with this more general context.

A random variable z is i.i.d. according to $F(\cdot)$ and we have a sample $z_1, z_2, ..., z_n$. In addition, for unknown θ of dimension k there is a (known) function $\psi(z,\theta)$ such that $\underset{F}{\mathrm{E}}\,\psi(z,\theta)$ =0. $\psi(z,\theta)$ is of dimension $m \geq k$. Empirical discrepancy theory considers choosing θ and probabilities $p_1, ..., p_n$ with support on the data such that

$$\sum_{i=1}^{n} h(p_i, \frac{1}{n}) \text{ is minimized subj. to } \underset{p}{\mathrm{E}}\,\psi(z,\theta) = 0 \text{ and } \sum p_i = 1$$

where $h(\cdot,\cdot)$ is a measure of the discrepancy between two discrete measures, with the property that $h(\frac{1}{n},\frac{1}{n}) = 0$; there are also some technical conditions on $h(\cdot,\cdot)$'s partial derivative with respect to its first argument.

Thus empirical discrepancy theory chooses θ and a reweighting of the data so that the moment conditions hold and a discrepancy measure is minimized.

$$Q(\theta,p) = \sum_{i=1}^{n} h(p_i, \frac{1}{n}) + \alpha(\sum p_i - 1) + t\sum_{i=1}^{n} p_i \psi_i(\theta) \tag{1}$$

Consider the determination of p first:

$$\frac{\partial Q(\theta,p)}{\partial p_i} = \frac{\partial h}{\partial p_i} + \alpha + t\psi_i(\theta) = 0 \tag{2}$$

$$\begin{aligned} \textstyle\sum_{i=1}^{n} \left\{ \frac{\partial Q(\theta,p)}{\partial p_i} p_i \right\} &= \sum \frac{\partial h}{\partial p_i} p_i \ + \ \alpha \textstyle\sum p_i \ + \ t \textstyle\sum_{i=1}^{n} p_i \psi_i(\theta) \\ &= \sum \frac{\partial h}{\partial p_i} p_i \ + \ \alpha \ + \ 0 \end{aligned}$$

So $\alpha = -\sum \frac{\partial h}{\partial p_i} p_i$; substituting into (2):

$$\frac{\partial Q(\theta,p)}{\partial p_i} = \frac{\partial h}{\partial p_i} - \sum \frac{\partial h}{\partial p_i} p_i + t\psi_i(\theta) = 0$$

Note:

$$\frac{\partial h}{\partial p_i} = -t\psi_i(\theta) \text{ is a solution}$$

Note that t is an m–dimensional Lagrange multiplier of the original problem.

Remaining with the problem of constructing p for given θ, there are three common choices for $h(\cdot,\cdot)$:

- $h(p_i, \frac{1}{n}) = p_i(p_i - \frac{1}{n})$, or effectively $\sum h(p_i, \frac{1}{n}) = \sum p_i^2$, in which case $p_i = k(1 + t\psi_i(\theta))$ and $t = -(\sum \psi_i \psi_i')^{-1}(\sum \psi_i)$; this is often called Euclidean likelihood.
- $h(p_i, \frac{1}{n}) = \frac{1}{n}\{\log(\frac{1}{n}) - \log p_i\}$, or effectively $-\sum h(p_i, \frac{1}{n}) = \sum \log p_i$; $p_i = k\frac{1}{1+t\psi_i(\theta)}$; this is Owen's (1988) empirical likelihood (EL).

- $h(p_i, \frac{1}{n}) = p_i\{\log(\frac{1}{n}) - \log p_i\}$, or effectively $-\sum h(p_i, \frac{1}{n}) = \sum p_i \log p_i$; $p_i = ke^{t \cdot \psi_i(\theta)}$; this is called 'exponential tilting' (ET).

Empirical likelihood and exponential tilting exchange the role of the empirical measure and the measure p that is under construction: ET finds the p to which the empirical measure is 'KLIC–closest', while EL finds the p that is KLIC-closest to the empirical measure. Thus ET 'imagines' that the data generating process is p, (which obeys $E(\psi) = 0$) while EL imagines the DGP as a repetition of the observed data, which does not obey the specificed moment conditions. To us, this suggests ET should be superior to EL; but EL enjoys an array of higher-order asymptotic properties, such as Bartlett correctability of its likelihood ratio test (but only when there are no nuisance parameters), which indicate that it may behave like parametric likelihood.

The preceding three cases are all members of the Cressie–Read family, with

$$h(p_i, \frac{1}{n}) = \left(\frac{p_i}{1/n}\right)^{-\lambda} - 1$$

$$p_i = k\left(\frac{1}{1 + t\psi_i(\theta)}\right)^{1/(\lambda+1)}$$

for $\lambda \in [-2, 1]$ so that $\lambda = -2$ is Euclidean likelihood, $\lambda = -1$ is ET and $\lambda = 0$ is EL.

Turning now to the problem of estimating θ, the minimum discrepancy estimate is obtained by differentiating (1) with respect to θ to obtain:

$$\frac{\partial Q(\theta, p)}{\partial \theta} = t\sum_{i=1}^{n} p_i \frac{\partial \psi_i(\theta)}{\partial \theta} = 0,$$

a system of equations in k elements of θ. Thus the entire system of $(m + k)$ equations can be written simply as:

$$\underset{p}{E}\psi(\theta) = 0 \qquad (m \text{ equations}) \tag{3a}$$

$$t \cdot \underset{p}{E}\frac{\partial \psi(\theta)}{\partial \theta} = 0 \qquad (k \text{ equations}) \tag{3b}$$

One way to think of these equations is that, having fixed θ and a formula for p (by choice of $h(\cdot, \cdot)$), the first m equations determine t. Similarly, for a fixed t and p, the remaining k equations determine θ.[1]

[1]This schema cannot be used to define a simple iterative procedure to compute θ, for in fact the saddlepoint nature of these equations makes the naive iterative procedure of (1) fix θ; (2) calculate t; (3) calculate new θ; unstable in a neighborhood of the solution θ_* of $\theta_* - \theta(t(\theta_*)) = 0$.

The duality of GEL and ED is examined in two papers of Newey and Smith (2000,2001), the latter of which will be presented in this conference. Writing the GEL estimator as:

$$\hat{\theta}_{GEL} = \arg\min_{\theta\in\Theta}\sup_{t\in T} n^{-1} g(t\cdot\psi_i(\theta))$$

the GEL estimator's estimating equations coincide with (3a) in cases where the derivative of $g(\cdot)$, denoted g', can be interpreted as being proportional to a probability. This can be done for the three cases under consideration here, as well as for all members of the Cressie-Read family. Newey and Smith (2000, 2001) show that for Euclidean likelihood, that $g(t\psi_i) = -t\psi_i - (t\psi_i)^2/2$ and the resulting GEL estimator coincides with the continuously updated GMM estimator of Hansen, Heaton and Yaron (1996). Consequently we will denote the three estimators as $\hat{\theta}_{CUE}, \hat{\theta}_{ET}$, and $\hat{\theta}_{EL}$.

4. A Further Characterization of ED/GEL Estimators

Rewriting the first equation of system (3a) as:

$$\sum_{i=1}^{n} p(t\cdot\psi_i(\theta))\psi_i(\theta) = 0 \tag{4}$$

we can express the probabilities associated with CUE, ET, and EL (after absorbing some sign changes into k) as:

$$\begin{aligned} p_i[CUE] &= k_{CUE}(1+t\psi_i(\theta)) \\ p_i[ET] &= k_{ET}(e^{t\psi_i(\theta)}) \\ p_i[EL] &= k_{EL}(\frac{1}{1-t\psi_i(\theta)}) \end{aligned}$$

Taking a Taylor series expansion of $p_i[ET]$ we can define a sequence of p functions:

$$\begin{aligned} p_i[ET,1] &= k_{ET,1}(1+t\psi_i(\theta)) \ = p_i[CUE] \\ p_i[ET,2] &= k_{ET,2}(1+t\psi_i(\theta)+\frac{(t\psi_i(\theta))^2}{2}) \\ p_i[ET,3] &= k_{ET,3}(1+t\psi_i(\theta)+\frac{(t\psi_i(\theta))^2}{2}+\frac{(t\psi_i(\theta))^3}{6}) \end{aligned}$$

$$\begin{aligned} p_i[ET,\infty] &= k_{ET,\infty}(1+t\psi_i(\theta)+\frac{(t\psi_i(\theta))^2}{2}+\frac{(t\psi_i(\theta))^3}{6}+...) \\ & k_\infty e^{t\cdot\psi_i(\theta)} \ = p_i[ET] \end{aligned}$$

And similarly for $p_i[EL]$ we have

$$
\begin{aligned}
p_i[EL,1] &= k_{EL,1}(1+t\psi_i(\theta)) \quad = p_i[CUE] \\
p_i[EL,2] &= k_{EL,2}(1+t\psi_i(\theta)+(t\psi_i(\theta))^2) \\
p_i[EL,3] &= k_{EL,3}(1+t\psi_i(\theta)+(t\psi_i(\theta))^2+(t\psi_i(\theta))^3) \\
. \quad &. \quad . \\
p_i[EL,\infty] &= k_{EL,\infty}(1+t\psi_i(\theta)+(t\psi_i(\theta))^2+(t\psi_i(\theta))^3+...) \\
&= k_{EL,\infty}(\frac{1}{1-t\psi_i(\theta)}) \quad = p_i[EL]
\end{aligned}
$$

Thus, all three p functions have the same first-order Taylor series expansion, coinciding exactly with $p[CUE]$. Then $p[ET]$ and $p[EL]$ include higher powers of $(t \cdot \psi_i)$, the former having factorially declining weights or coefficients and the latter the coefficients {1,1,...1}. Since t is an $O_p(n^{-1/2})$ object, the difference in the treatment of t^2 terms induces differences of $O_p(n^{-1})$ in $\hat{\theta}_{CUE}$, $\hat{\theta}_{ET}$, and $\hat{\theta}_{EL}$ and consequently their MSE behavior differs at $O_p(n^{-2})$. (This will be true for all members of the Cressie-Read family.[2])

To see the effect of these differences, let us consider the difference between the first two elements of the sequence of ET functions for the equation setting the expectation of the j^{th} component of ψ:

$$\sum_{i=1}^{n} p(t\cdot\psi_i(\theta))\psi_{ij}(\theta) = 0 \tag{5}$$

$$\sum_{i=1}^{n} k_{ET,1}(1+t\psi_i(\theta))\psi_{ij}(\theta) = 0 \tag{6}$$

$$\sum_{i=1}^{n} k_{ET,2}(1+t\psi_i(\theta)+\frac{(t\psi_i(\theta))^2}{2})\psi_{ij}(\theta) = 0 \tag{7}$$

Supressing the i subscript momentarily, the extra terms in (7) (relative to (6)) are of the form:

$$.5 * (t_1\psi_1 + t_2\psi_2 + ...t_m\psi_m)^2\psi_j$$

so that sums of these involve (generalized) third moments of ψ. Consequently, in problems where generalized third moments are zero, notably those in which ψ is symmetric, these terms will be converging rapidly to zero and thus have no effect even at $O_p(n^{-2})$.

[2]The Cressie-Read expansion is: $p[CR] = 1 + (t\cdot\psi_i)/(1+\lambda) + (2+\lambda)(t\cdot\psi_i)^2/2(1+\lambda)^2 + (2+\lambda)(3+2\lambda)(t\cdot\psi_i)^3/6(1+\lambda)^3 + ...$

5. A Detailed Analysis of Some Simple Examples

To examine further the relation between the choice of a GEL/ED and the higher order moments of the underlying data, consider the estimation of the scalar parameter θ from a scalar random variable x where it is known that x has mean θ and variance 2θ. Thus $\psi(x,\theta)$ is given by:

$$\begin{aligned} x-\theta &= 0 \\ x^2-\theta^2-2\theta &= 0 \end{aligned} \tag{8}$$

Writing the second moment condition in the way indicated (rather than $(x-\theta)^2-2\theta=0$) does not change the numerical values of the resulting estimates of θ, but it does simplify $\frac{\partial\psi}{\partial\theta}$ to:

$$\frac{\partial\psi}{\partial\theta}=\begin{cases} -1 \\ -2\theta-2 \end{cases}$$

Consequently, $\frac{\partial\psi}{\partial\theta}$ does not depend on the data so $\underset{p}{E}\frac{\partial\psi}{\partial\theta}$ does not depend on p. Using (3b) this means $\hat\theta$ can be determined from

$$\begin{aligned} t\cdot\underset{p}{E}\frac{\partial\psi}{\partial\theta} &= 0 \\ t_1(-1)+t_2(-2\theta-2) &= 0 \\ \theta &= \frac{-t_1}{2t_2}-1 \end{aligned}$$

It is apparent that our three estimators will differ, in this special case, only in their choice of t. For the CUE estimator, $t=(\psi'\psi)^{-1}\overline{\psi}$, i.e. the coefficients of the regression of a column of 1's on ψ, and so θ is determined by the fixed point of a function of five moment functions of ψ: the means of the two moment functions (expressed as functions of θ) and the corresponding three variances and covariances. CUE is thus committed to local (to θ) sufficiency of five statistics, and will ignore, for example, differences in skew between elements of the sample space. In cases where skew is zero, we can expect the difference between CUE and ET or EL to be negligible, whereas for non-zero skew, we might expect ET and/or EL to prove superior to CUE, but only at sample sizes at which $O_p(n^{-2})$ effects are operative.

To demonstrate these effects, we construct several data–generating processes which satisfy the moment conditions in (8) but have different properties for their higher–order moments. For each case we compute MSE and bias, and do this for CUE, ET, and EL. In addition, for ET and EL we compute p according to successive terms in the relevant Taylor series expansion, so that $p_i[EL,1]=k_{EL,1}(1+t\psi_i(\theta))\ =p_i[CUE]$, $p_i[EL,3]=k_{EL,3}(1+t\psi_i(\theta)+(t\psi_i(\theta))^2+(t\psi_i(\theta))^3)$, etc. In this way we can see whether the advantages, if any, of ET and EL over CUE set in after taking into account only a relatively small number of additional higher moments, and similarly if differences between EL and ET require the full limiting case of including some information about all higher–order moments.

Our first case takes x to be $\chi^2(1)$, so x has expectation 1 and variance 2. We consider MSE first. In each of the tables to follow, we report the value of the MSE in 6,000 replications for CUE, ET, and EL, together with the MSE's of the estimators based on the Taylor series expansion of degrees 3,5, and 7, and the simple mean. (CUE corresponds to a Taylor series expansion of degree 1, ET and EL to degree ∞). For each entry we report a jackknife estimate of the standard error.

In Table 1, case 1, based on $\chi^2(1)$ data and $n = 50$, we see for CUE the effect of adding an additional moment is to produce an estimator that is worse than the sample mean; the MSE's of the estimators at Taylor degrees 3,5, and 7 are about the same as the sample mean; and that ET and EL are better than their corresponding degree 7 estimators and also the sample mean; both of these effects are greater for EL than ET.

In Table 1, case 1, $n = 100$, we now see that CUE has smaller MSE than the sample average (.0185 vs. .020) while the third and successive powers showed marked improvemnt over the CUE level: (.0158 or less); full EL shows the best MSE at .01474, slightly better than ET's .01551; both have already reached the apparent asymptotic relative efficiency level of .75 of the sample mean's MSE.

At $n = 200$ the improvement in going from CUE to the degree 3 Taylor expansion is still discernible; ET, EL, and all the Taylor expansions have MSE's indistinguishable from the apparent asymptotic efficiency ratio, with ET showing a slight advantage. At $n = 400$, even CUE has reached the asymptotic relative efficiency limit and the performance of all the two-moment estimators is indistinguishable: second–order effects have been extinguished.

Table 2, case 2 reports results for x distributed $N(1, 2)$. These are simple to summarize: the asymptotic relative efficiency level is already reached at $n = 50$; there is no significant difference in the performance of any of the estimators. At each sample size, there is a very slight advantage in ET/EL over the previous estimators in the Taylor sequence: there seems to be no disadvantage at these sample sizes in taking higher–order moments into account.

So far, these results are in agreement with our expectations: in situations where skewness is present, estimators that can take this into account do better; where skewness is absent, the performance of the estimators is indistinguishable.

Case 3 moves from normality to symmetry, by taking a symmetric mixture of normals $(.5N(0, 1),\ .5N(2, 1))$ at sample sizes $n = 25, 50, 100$. The results here mirror those of case 2: the apparent asymptotic relative efficiency level of about .45 of the MSE of the sample mean is basically achieved by all the estimators at the smallest sample size.

A final example, case 4, takes an asymmetric mixture of normals, $(.25N(-1, 2/3), .75N(5/3, 2/3))$. Here the knowledge that the variance is twice the mean is very informatiove: MSE is reduced to 25% of that of the simple mean. This is achieved at $n = 50$; there is some indication that CUE is inferior to the other estimators; otherwise, they are indistinguishable.

In Table 2, we present biases for the same cases as considered in Table 1. Quite notably, in no case does the bias make a substantial difference to the MSE. This could perhaps be expected from the fact that the correlation between ψ and $\frac{\partial \psi}{\partial \theta}$ is zero in this example, because the latter does not depend on x. This suggests that all of the (sometimes erratic) effects seen in Table 2 are $O(n^{-2})$ or higher.

6. Summary

Higher–order asymptotic arguments suggest that GEL/ED/'one–step efficient' estimates of overidentified moment models will prove unambiguously superior to two–step GMM. Consequently interest shifts to distinguishing between elements of this family on the basis of estimation performance. With a simple argument and a simple example, it appears that the simplest GEL variant, the continuously updated or Euclidean likelihood estimator, is dominated by the more elaborate ET and EL estimators. The difference between these two variants can be seen to lie in their treatment of third and higher–order moments of moment conditions, with EL weighing these more heavily than ET.

Table 1: MSE for Estimators of θ when E(x)=θ =1 and V(x)=2θ=2; 6,000 replications

case 1: $\chi^2(1)$ n=50

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
MSE	0.0477737	0.0412037	0.0399501	0.0395447	0.0375955	0.0401685
s.e.	0.0010239	0.0009177	0.0008859	0.0008702	0.0007401	0.0007518
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
MSE	0.0477737	0.0417046	0.0408829	0.0406844	0.033915	0.0401685
s.e.	0.0010239	0.000948	0.0009407	0.0009408	0.0006789	0.0007518

case 1: $\chi^2(1)$ n=100

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
MSE	0.0185032	0.015828	0.0155639	0.0155207	0.0155082	0.0200417
s.e.	0.0003997	0.0003327	0.000321	0.0003175	0.0003151	0.0003652
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
MSE	0.0185032	0.0156564	0.0153857	0.0153269	0.0147410	0.0200417
s.e.	0.0003997	0.0003385	0.0003316	0.0003300	0.0002938	0.0003652

case 1: $\chi^2(1)$ n=200

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
MSE	0.0079449	0.0074472	0.0074264	0.0074256	0.0074245	0.0101593
s.e.	0.0001535	0.0001381	0.0001374	0.0001373	0.0001372	0.0001906
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
MSE	0.0079449	0.0074982	0.0075231	0.0075377	0.0075811	0.0101593
s.e.	0.0001535	0.0001392	0.0001396	0.0001400	0.0001405	0.0001906

case 1: $\chi^2(1)$ n=400

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
MSE	0.003708	0.0036713	0.0036719	0.0036721	0.0036725	0.0050607
s.e.	6.75e-05	6.72e-05	6.72e-05	6.72e-05	6.72e-05	9.09e-05
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
MSE	0.003708	0.0037128	0.0037281	0.003733	0.0037455	0.0050607
s.e.	6.75e-05	6.82e-05	6.85e-05	6.86e-05	6.88e-05	9.09e-05

case 2: N(1,2) n=50

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
MSE	0.0208211	0.0207016	0.0206913	0.0206884	0.0206869	0.0397528
s.e.	0.0003657	0.0003647	0.0003647	0.0003647	0.0003647	0.0007049
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
MSE	0.0208211	0.0206747	0.0206713	0.0206562	0.0205861	0.0397528
s.e.	0.0003657	0.000365	0.0003664	0.0003668	0.0003669	0.0007049

case 2: N(1,2) n=100

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
MSE	0.0103036	0.0102563	0.0102546	0.0102543	0.0102543	0.0199449
s.e.	0.0001887	0.0001881	0.0001881	0.0001881	0.0001881	0.0003529
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
MSE	0.0103036	0.0102351	0.0102265	0.0102245	0.0102176	0.0199449
s.e.	0.0001887	0.0001877	0.0001876	0.0001876	0.0001877	0.0003529

case 2: N(1,2) n=200

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
MSE	0.0049441	0.0049305	0.0049297	0.0049297	0.0049297	0.0098335
s.e.	8.74e-05	8.73e-05	8.72e-05	8.72e-05	8.72e-05	0.0001853
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
MSE	0.0049441	0.0049303	0.0049308	0.0049318	0.0049234	0.0098335
s.e.	8.74e-05	8.74e-05	8.75e-05	8.75e-05	8.73e-05	0.0001853

case 2: N(1,2) n=400

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
MSE	0.0025947	0.0025918	0.0025918	0.0025918	0.0025918	0.0051558
s.e.	4.75e-05	4.75e-05	4.75e-05	4.75e-05	4.75e-05	9.34e-05
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
MSE	0.0025947	0.0025917	0.0025920	0.0025923	0.0025916	0.0051558
s.e.	4.75e-05	4.76e-05	4.76e-05	4.76e-05	4.76e-05	9.34e-05

case 3: .5N(0,1)+.5N(2,1) n=25

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
MSE	0.0375929	0.0366819	0.0366552	0.0366509	0.0376198	0.0820290
s.e.	0.0007475	0.0006434	0.0006431	0.000643	0.0007462	0.0014929
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
MSE	0.0376024	0.0370686	0.0371873	0.0366086	0.0373127	0.0820290
s.e.	0.0007504	0.0006974	0.0007338	0.0006494	0.0007442	0.0014929

case 3: .5N(0,1)+.5N(2,1) n=50

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
MSE	0.0175622	0.0174594	0.0174505	0.0174491	0.0174489	0.0398252
s.e.	0.0003167	0.0003147	0.0003145	0.0003144	0.0003144	0.0007314
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
MSE	0.0175622	0.0174485	0.0174362	0.017434	0.017364	0.0398252
s.e.	0.0003167	0.0003148	0.0003146	0.0003146	0.0003128	0.0007314

case 3: .5N(0,1)+.5N(2,1) n=100

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
MSE	0.0089522	0.0089106	0.0089075	0.0089071	0.008907	0.0198064
s.e.	0.0001613	0.0001609	0.0001609	0.0001609	0.0001609	0.0003546
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
MSE	0.0089522	0.008902	0.0088993	0.0088995	0.0088751	0.0198064
s.e.	0.0001613	0.0001606	0.0001606	0.0001607	0.0001607	0.0003546

case 4 Asymmetric normal mixture: n=50

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
MSE	0.0102061	0.0100858	0.010078	0.0100775	0.0100774	0.0397681
s.e.	0.0001833	0.0001814	0.0001813	0.0001813	0.0001813	0.0007181
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
MSE	0.0102061	0.0100745	0.0100638	0.0100623	0.0100054	0.0397681
s.e.	0.0001833	0.0001811	0.000181	0.000181	0.0001805	0.0007181

case 4 Asymmetric normal mixture: n=100

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
MSE	0.0050106	0.0049707	0.0049698	0.0049698	0.0049698	0.0198901
s.e.	9.07e-05	9.02e-05	9.02e-05	9.02e-05	9.02e-05	0.0003614
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
MSE	0.0050106	0.0049567	0.0049506	0.004949	0.0049459	0.0198901
s.e.	9.07e-05	8.99e-05	8.99e-05	8.99e-05	9e-05	0.0003614

Table 2: Bias for Estimators of θ when E(x)=θ =1 and V(x)=2θ=2; 6,000 replications

case 1: $\chi^2(1)$ n=50

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
Bias	-0.0627946	-0.0382844	-0.0356822	-0.0352203	-0.0335646	-0.0020989
s.e.	0.0027030	0.0025737	0.0025391	0.0025269	0.0024656	0.0025875
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
Bias	-0.0627946	-0.0281142	-0.0202024	-0.0174114	-0.0016759	-0.0020989
s.e.	0.002703	0.0026115	0.0025975	0.0025945	0.0023776	0.0025875

case 1: $\chi^2(1)$ n=100

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
Bias	-0.0195812	-0.0033899	-0.0025044	-0.0025071	-0.0025731	0.0012786
s.e.	0.0017379	0.0016237	0.0016104	0.0016082	0.0016075	0.0018277
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
Bias	-0.0195812	0.0061319	0.0114108	0.0132869	0.0166805	0.0012786
s.e.	0.0017379	0.0016136	0.0015947	0.0015892	0.0015527	0.0018277

case 1: $\chi^2(1)$ n=200

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
Bias	-0.002773	0.0048958	0.0050594	0.0050137	0.0049701	0.0018738
s.e.	0.0011503	0.0011124	0.0011107	0.0011107	0.0011106	0.0013011
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
Bias	-0.002773	0.0101017	0.0118915	0.0124566	0.0125816	0.0018738
s.e.	0.0011503	0.0011104	0.0011093	0.0011093	0.0011124	0.0013011

case 1: $\chi^2(1)$ n=400

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
Bias	0.0016772	0.0044239	0.0043991	0.0043744	0.0043640	0.0011137
s.e.	0.0007859	0.0007802	0.0007803	0.0007803	0.0007804	0.0009184
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
Bias	0.0016772	0.0065287	0.0069163	0.0070145	0.0066197	0.0011137
s.e.	0.0007859	0.0007822	0.0007833	0.0007836	0.0007855	0.0009184

case 2: N(1,2) n=50

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
Bias	-0.015814	-0.0133767	-0.0131317	-0.0131018	-0.0130948	0.0005364
s.e.	0.0018518	0.0018496	0.0018494	0.0018493	0.0018493	0.0025742
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
Bias	-0.015814	-0.0131027	-0.0126967	-0.0124400	-0.0106091	0.0005364
s.e.	0.0018518	0.0018487	0.001849	0.0018486	0.0018474	0.0025742

case 2: N(1,2) n=100

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
Bias	-0.0064192	-0.0046947	-0.0045749	-0.0045645	-0.0045634	-0.0005053
s.e.	0.0013079	0.0013061	0.0013061	0.0013061	0.0013061	0.0018234
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
Bias	-0.0064192	-0.0042371	-0.0038316	-0.003693	-0.0027664	-0.0005053
s.e.	0.0013079	0.001305	0.0013047	0.0013046	0.0013046	0.0018234

case 2: N(1,2) n=200

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
Bias	-0.0035925	-0.0025934	-0.0025558	-0.0025544	-0.0025544	7.35e-05
s.e.	0.0009066	0.000906	0.0009059	0.0009059	0.0009059	0.0012803
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
Bias	-0.0035925	-0.0021008	-0.0018504	-0.0017735	-0.0014979	7.35e-05
s.e.	0.0009066	0.0009062	0.0009063	0.0009064	0.0009057	0.0012803

case 2: N(1,2) n=400

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
Bias	-0.0014226	-0.0008502	-0.0008379	-0.0008377	-0.0008377	0.0003372
s.e.	0.0006574	0.0006572	0.0006572	0.0006572	0.0006572	0.0009271
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
Bias	-0.0014226	-0.0004787	-0.0003439	-0.0003064	-0.0002362	0.0003372
s.e.	0.0006574	0.0006573	0.0006573	0.0006573	0.0006573	0.0009271

case 3: .5N(0,1)+.5N(2,1) n=25

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
Bias	-0.027177	-0.023619	-0.0231924	-0.0231363	-0.0223505	0.0016949
s.e.	0.0024786	0.0024539	0.0024537	0.0024536	0.0024875	0.0036977
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
Bias	-0.027181	-0.0245965	-0.0241462	-0.0232051	-0.0193877	0.0016949
s.e.	0.0024789	0.0024654	0.0024702	0.0024521	0.0024814	0.0036977

case 3: .5N(0,1)+.5N(2,1) n=50

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
Bias	-0.0122021	-0.0097789	-0.0095632	-0.0095442	-0.0095427	0.0023067
s.e.	0.0017037	0.0017013	0.0017011	0.001701	0.001701	0.0025764
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
Bias	-0.0122021	-0.0093893	-0.0088237	-0.0086242	-0.006993	0.0023067
s.e.	0.0017037	0.0017011	0.001701	0.0017011	0.0016989	0.0025764

case 3: .5N(0,1)+.5N(2,1) n=100

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
Bias	-0.0072486	-0.0057475	-0.0056751	-0.0056713	-0.005671	-0.0006089
s.e.	0.001218	0.0012165	0.0012163	0.0012163	0.0012163	0.001817
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
Bias	-0.0072486	-0.0051148	-0.004709	-0.0045766	-0.0040612	-0.0006089
s.e.	0.001218	0.0012164	0.0012165	0.0012166	0.0012152	0.001817

case 4 Asymmetric normal mixture: n=50

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
Bias	-0.0106045	-0.007696	-0.0074922	-0.0074792	-0.0074785	0.003173
s.e.	0.0012971	0.0012928	0.0012925	0.0012925	0.0012925	0.0025744
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
Bias	-0.0106045	-0.0069114	-0.006191	-0.0059425	-0.0044577	0.003173
s.e.	0.0012971	0.0012928	0.0012927	0.0012928	0.0012902	0.0025744

case 4 Asymmetric normal mixture: n=100

ET sequence	1=CUE	3	5	7	ET(Infinity)	Sample Mean
Bias	-0.0049305	-0.0032471	-0.0031752	-0.003172	-0.0031718	3.25e-05
s.e.	0.0009117	0.0009093	0.0009093	0.0009093	0.0009093	0.0018209
EL sequence	1=CUE	3	5	7	EL(Infinity)	Sample Mean
Bias	-0.0049305	-0.0024855	-0.0020272	-0.001878	-0.001382	3.25e-05
s.e.	0.0009117	0.0009084	0.0009081	0.000908	0.0009078	0.0018209

Appendix

Lemma 10. (Expansion of Matrix Inversion)
Let A and B be $M \times M$ matrices, with A invertible and both A and B of order $O_p(1)$. Then
(i): $(A + B/\sqrt{N})^{-1} = A^{-1} + o_p(1)$,
(ii): $(A + B/\sqrt{N})^{-1} = A^{-1} - A^{-1}BA^{-1}/\sqrt{N} + o_p(1/\sqrt{N})$,
(iii): $(A + B/\sqrt{N})^{-1} = A^{-1} - A^{-1}BA^{-1}/\sqrt{N} + A^{-1}BA^{-1}BA^{-1}/N + o_p(1/N)$,

Proof of Lemma 10

$$(A + B/\sqrt{N})^{-1} = \Big(A^{-1} - A^{-1}BA^{-1}/\sqrt{N}\Big) \cdot \Big(A^{-1} - A^{-1}BA^{-1}/\sqrt{N}\Big)^{-1} \cdot \Big(A + B/\sqrt{N}\Big)^{-1}$$

$$= (A^{-1} - A^{-1}BA^{-1}/\sqrt{N}) \cdot \Big((A + B/\sqrt{N}) \cdot (A^{-1} - A^{-1}BA^{-1}/\sqrt{N})\Big)^{-1}$$

$$= (A^{-1} - A^{-1}BA^{-1}/\sqrt{N}) \cdot \Big(\mathcal{I} + BA^{-1}/\sqrt{N} - AA^{-1}BA^{-1}/\sqrt{N} - BA^{-1}BA^{-1}/N)\Big)^{-1}$$

$$= (A^{-1} - A^{-1}BA^{-1}/\sqrt{N}) \cdot (\mathcal{I} - BA^{-1}BA^{-1}/N))^{-1}$$

$$= A^{-1} - A^{-1}BA^{-1}/\sqrt{N} + o_p(1/\sqrt{N}).$$

For the next step consider the second factor in the last expression:

$$(\mathcal{I} - BA^{-1}BA^{-1}/N))^{-1}$$

$$= (\mathcal{I} + BA^{-1}BA^{-1}/N) \cdot (\mathcal{I} + BA^{-1}BA^{-1}/N)^{-1}(\mathcal{I} - BA^{-1}BA^{-1}/N)^{-1}$$

$$= (\mathcal{I} + BA^{-1}BA^{-1}/N) \cdot \Big((\mathcal{I} - BA^{-1}BA^{-1}/N) \cdot (\mathcal{I} + BA^{-1}BA^{-1}/N)\Big)^{-1}$$

$$= (\mathcal{I} + BA^{-1}BA^{-1}/N) \cdot (\mathcal{I} - BA^{-1}BA^{-1}BA^{-1}BA^{-1}/N^2)^{-1}$$

$$= (\mathcal{I} + BA^{-1}BA^{-1}/N) + o_p(1/N).$$

Hence:

$$(A + B/\sqrt{N})^{-1} = (A^{-1} - A^{-1}BA^{-1}/\sqrt{N}) \cdot (\mathcal{I} - BA^{-1}BA^{-1}/N))^{-1}$$

$$= (A^{-1} - A^{-1}BA^{-1}/\sqrt{N}) \cdot \Big(\mathcal{I} + BA^{-1}BA^{-1}/N) + o_p(1/N)\Big)$$

$$= A^{-1} - A^{-1}BA^{-1}/\sqrt{N} + A^{-1}BA^{-1}BA^{-1}/N + o_p(1/N)).$$

□

Proof of Lemma 1

Because by Lemma 10, $1/(1+\overline{v}_1) = 1 - \overline{v}_1 + o_p(1/\sqrt{N})$, we have

$$\hat{\theta}_{opt} = \overline{w}_1/(1+\overline{v}_1) = \overline{w}_1(1-\overline{v}_1) + o_p(1/N) = \overline{w}_1 - \overline{w}_1\overline{v}_1 + o_p(1/N).$$

□

Proof of Lemma 2

We show the following two results, which then imply the main result:
(i), $E[T_1] = \theta^*$,
(ii), $E[R_1] = -\rho/N$.
(i) This is immediate.
(ii):

$$E[R_1] = -E[\overline{v}_1\overline{w}_1] = -E\left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} v_{i1}w_{i1}\right]$$

$$= -E\left[\frac{1}{N^2}\sum_{i=1}^{N} v_{i1}w_{i1}\right] = -\rho/N.$$

□

Proof of Lemma 3

We first show the following results.
(i): $E[T_1^2] = 1/N$,
(ii): $E[T_1 \cdot R_1] = -\mu_{12}/N^2$,
(iii): $E[R_1^2] = (1+2\rho^2)/N^2 + o(1/N^2)$.
In the following, let $\delta_{mn} = 1$ if $m = n$ and zero otherwise.
(i):

$$E[T_1^2] = E\left[\sum_{i=1}^{N}\sum_{j=1}^{N} w_{i1}w_{j1}\right] = E\left[\sum_{i=1}^{N} w_{i1}^2\right] = 1/N.$$

(ii):

$$E[T_1 \cdot R_1] = -E[\overline{w}_1\overline{w}_1\overline{v}_1] = -E\left[\frac{1}{N^3}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} w_{i1}w_{j1}v_{k1}\right]$$

$$= -E\left[\frac{1}{N^3}\sum_{i=1}^{N} w_{i1}^2 v_{i1}\right] = -\mu_{12}/N^2.$$

(iii):

$$E[R_1^2] = E\left[(\overline{v}_1\overline{w}_1)^2\right] = E\left[\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} v_{i1}w_{j1}\right)^2\right]$$

$$= E\left[\frac{1}{N^4}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N} v_{i1}w_{j1}v_{k1}w_{l1}\right].$$

Because the (v_{im}, w_{im}) is independent of (v_{jn}, w_{jn}) if either $i \neq j$ or $m \neq n$, we can ignore all terms where one of the four indices i, j, k, and l, does not match up with at least one of the others. Ignoring also the N terms with all four indices matching up because they are of lower order, we only consider terms with $(i = j, k = l, i \neq k)$, $(i = l, j = k, i \neq j)$ or $(i = k, j = l, i \neq j)$, leading to

$$E[R_1^2] = \frac{1}{N^4}E\left[\sum_{i=1}^{N}\sum_{k\neq i} v_{i1}w_{i1}v_{k1}w_{k1} + \sum_{i=1}^{N}\sum_{j\neq i} v_{i1}w_{j1}v_{j1}w_{i1} + \sum_{i=1}^{N}\sum_{j\neq i} v_{i1}w_{j1}v_{i1}w_{j1}\right] + o_p(1/N^2)$$

$$= \frac{1}{N^2}\left(\rho^2 + \rho^2 + 1\right) = (2\rho^2 + 1)/N^2.$$

Then, adding up the three components:

$$E[(T_1 + R_1 - \theta^*)^2)] = E[T_1^2 + 2T_1R_1 + R_1^2] = 1/N - 2\mu_{12}/N^2 + (1 + 2\rho^2)/N^2 + o_p(1/N^2).$$

□

Proof of Lemma 4:

First we proves the following set of preliminary results:
(i), $\overline{w} = O_p(1/\sqrt{N})$,
(ii), $\overline{v} = O_p(1/\sqrt{N})$,
(iii), $\overline{v}'\overline{w} = O_p(1/N)$,
(iv), $(\overline{v} + e_1)'\overline{w} = O_p(1/\sqrt{N})$,
(v), $\overline{ww'} = \mathcal{I}_M + O_p(1/\sqrt{N})$,
(iv), $(\overline{ww'})^{-1} = \mathcal{I}_M - (\overline{ww'} - \mathcal{I}_M) + o_p(1/\sqrt{N})$,
(vii), $(\overline{v} + e_1)'(\overline{ww'})^{-1}\overline{w} = \overline{w}_1 + o_p(1/\sqrt{N})$,
$(viii)$, $\left((\overline{v} + e_1)'(\overline{ww'})^{-1}(\overline{v} + e_1)\right)^{-1} = 1 - 2\overline{v}_1 + (\overline{ww'}_{11} - 1) + o_p(1/\sqrt{N})$.
(i) and (ii) are obvious by a central limit theorem.
(iii): As the product of two factors that are both $O_p(1/\sqrt{N})$ this is $O_p(1/N)$.
(iv): $(\overline{v} + e_1)'\overline{w} = \overline{v}'\overline{w} + \overline{w}_1$. The first term is $O_p(1/N)$ and the second term is $O_p(1/\sqrt{N})$ so that the sum is $O_p(1/\sqrt{N})$.
(v). The expectation opf $\overline{ww'}$ is $\mathcal{I}_M$. Combined with a central limit theorem this proves the result.
(vi): this is an application of Lemma 10
(vii). The first and second factor are $O_p(1)$, and the third one is $O_p(1/\sqrt{N})$. Hence the leading term is $e_1\mathcal{I}_M\overline{w} = \overline{w}_1$, and the remainder is $O_p/(1\sqrt{N})$.

$(viii)$: by (vi), $\overline{ww'}^{-1} = \mathcal{I}_M - (\overline{ww'} - \mathcal{I}) + o_p(1/\sqrt{N})$. Hence

$$(\overline{v}+e_1)'(\overline{ww'})^{-1}(\overline{v}+e_1) = (\overline{v}+e_1)'(\mathcal{I} - (\overline{ww'} - \mathcal{I}))(\overline{ww'})^{-1}(\overline{v}+e_1) + o_p(1/\sqrt{N})$$

$$= e1'e1 + \overline{v}'e_1 + e_1'\overline{v} - e_1'(\overline{ww'} - \mathcal{I}_M)e_1 + o_p(1/\sqrt{N})$$

$$= 1 + 2\overline{v}_1 - (\overline{ww'}_{11} - 1) + o_p(1/\sqrt{N}).$$

Nest we prove the main result. We write

$$\hat{\theta}_{gmm} = \left((\overline{v}+e_1)' \cdot (\overline{ww'})^{-1} \cdot (\overline{v}+e_1)\right)^{-1} \cdot \left((\overline{v}+e_1)' \cdot (\overline{ww'})^{-1} \cdot \overline{w}\right)$$

$$= \left(\left((\overline{v}+e_1)' \cdot (\overline{ww'})^{-1} \cdot (\overline{v}+e_1)\right)^{-1} - 1\right) \cdot \left((\overline{v}+e_1)' \cdot (\overline{ww'})^{-1} \cdot \overline{w}\right) \tag{9}$$

$$+ \left((\overline{v}+e_1)' \cdot (\overline{ww'})^{-1} \cdot \overline{w}\right) - \left(\overline{v}' \cdot (\overline{ww'})^{-1} \cdot \overline{w}\right) - \left(e_1' \cdot (\overline{ww'})^{-1} \cdot \overline{w}\right) \tag{10}$$

$$+ \left(e_1' \cdot (\overline{ww'})^{-1} \cdot \overline{w}\right) - \overline{w}_1 \tag{11}$$

$$+ \left(\overline{v}' \cdot (\overline{ww'})^{-1} \cdot \overline{w}\right) - \overline{v}'\overline{w} \tag{12}$$

$$+ \overline{v}'\overline{w} \tag{13}$$

$$+ \overline{w}_1. \tag{14}$$

Next we consider each of these six terms in order. The first term is

$$\left(\left((\overline{v}+e_1)' \cdot (\overline{ww'})^{-1} \cdot (\overline{v}+e_1)\right)^{-1} - 1\right) \cdot \left((\overline{v}+e_1)' \cdot (\overline{ww'})^{-1} \cdot \overline{w}\right).$$

By Lemma 2(vi) the second factor is $O_p(1/\sqrt{N})$, and by Lemma 2(vii) the first factor is $O_p(1/\sqrt{N})$. Therefore the leading term of the product is $O_p(1/N)$. Since we are only interested in the $O_p(1/N)$ terms, we can ignore all terms of order $o_p(1/\sqrt{N})$ in the two factors. For the first factor the leading $O_p(1/\sqrt{N})$ term is

$$-2v_1 + (\overline{ww'}_{11} - 1),$$

and for the second term the leading $O_p(1/\sqrt{N})$ term is $\overline{w}_1$. Hence the leading $O_p(1/N)$ term of the product is

$$-2\overline{v}_1\overline{w}_1 + (\overline{ww'}_{11} - 1)\overline{w}_1. \tag{15}$$

The second term is identically equal to zero. The third term is

$$\left(e_1' \cdot (\overline{ww'})^{-1} \cdot \overline{w}\right) - \overline{w}_1 = e_1' \cdot \left((\overline{ww'})^{-1} - \mathcal{I}_M\right) \overline{w}.$$

Using result (vi) to approximate $(\overline{ww'})^{-1}$, this can be written as

$$e_1' \cdot \left(\mathcal{I}_M - (\overline{ww'} - \mathcal{I}_M) - \mathcal{I}_M\right) \overline{w} + o_p(1/N)$$

(16) $$= -e_1'(\overline{ww'} - \mathcal{I}_M)\overline{w} + o_p(1/N).$$

The fourth term is

(17) $$\left(\overline{v}' \cdot (\overline{ww'})^{-1} \cdot \overline{w}\right) - \overline{v}'\overline{w} = \overline{v}'((\overline{ww'})^{-1} - \mathcal{I}_M)\overline{w} = o_p(1/N),$$

because all three factors are $O_p(1/\sqrt{N})$. The fifth and sixth terms are $O_p(1/N)$, so adding them to the sum of (15)-(17) gives the result. □

Proof of Lemma 5:

We show the following results:

(i), $E[T_1] = \theta^*$,

(ii), $E[R_1] = -2\rho/N$,

(iii), $E[R_2] = \mu_{03}/N$,

(iv), $E[R_3] = -\mu_{03}/N$,

(v), $E[R_4] = M\rho/N$,

which then by adding up imply the result in Lemma 5. In this we use the notation δ_{mn} for the indicator function, $\delta_{mn} = 1$ if $m = n$ and zero otherwise.

(i): This is immediate.

(ii): This follows directly from Lemma 2, part (ii).

(iii):

$$E[R_2] = -E[e_1'(\overline{ww'} - \mathcal{I}_M)\overline{w} = -E\left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M} e_{1m}(w_{im}w_{in} - \delta_{mn})w_{jn}\right]$$

$$= -E\left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{n=1}^{M}(w_{i1}w_{in} - \delta_{1n})w_{jn}\right]$$

$$= -E\left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(w_{i1}w_{i1} - 1)w_{j1}\right]$$

$$= -E\left[\frac{1}{N^2}\sum_{i=1}^{N}(w_{i1}w_{i1} - 1)w_{i1}\right] = -\mu_{03}/N.$$

(iv):

$$E[R_3] = -E[e_1'(\overline{ww'} - \mathcal{I}_M)\overline{w}] = -E\left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M} e_{1m}(w_{im}w_{in} - \delta_{mn})w_{jn}\right]$$

$$= -E\left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{n=1}^{M}(w_{i1}w_{in} - \delta_{1n})w_{jn}\right]$$

$$= -E\left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(w_{i1}w_{i1}-\delta_{11})w_{j1}\right]$$

$$= -E\left[\frac{1}{N^2}\sum_{i=1}^{N}(w_{i1}w_{i1}-1)w_{i1}\right] = -\mu_{03}/N.$$

(v):

$$E[R_4] = E[\overline{v}'\overline{w}] = E\left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}v_{im}w_{jm}\right]$$

$$= E\left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{m=1}^{M}v_{im}w_{im}\right] = M\rho/N.$$

$\square$.

Proof of Lemma 6

We first show the following results.

(i): $E[T_1^2] = 1/N$,
(ii): $E[T_1 \cdot R_1] = -2\mu_{12}/N^2$,
(iii): $E[T_1 \cdot R_2] = (\mu_{04}-1)/N^2$,
(iv): $E[T_1 \cdot R_3] = -(\mu_{04}-1)/N^2$,
(v): $E[T_1 \cdot R_4] = \mu_{12}/N^2$,
(vi): $E[R_1^2] = 4\cdot(2\rho^2+1)/N^2 + o(1/N^2)$,
(vii): $E[R_2^2] = (2\mu_{03}^2+\mu_{04}-1)/N^2 + o(1/N^2)$,
$(viii)$: $E[R_1 \cdot R_2] = -2\cdot(2\rho\mu_{03}+\mu_{12})/N^2 + o(1/N^2)$,
(ix): $E[R_3^2] = (2\mu_{03}^2+\mu_{04}+M-2)/N^2 + o(1/N^2)$,
(x): $E[R_1 \cdot R_3] = 2\cdot(2\rho\mu_{03}+\mu_{12})/N^2 + o(1/N^2)$,
(xi): $E[R_2 \cdot R_3] = -(2\mu_{03}^2+\mu_{04}-1)/N^2 + o(1/N^2)$,
(xii): $E[R_4^2] = (M^2\rho^2+M\rho^2+M)/N^2 + o(1/N^2)$,
$(xiii)$: $E[R_1 \cdot R_4] = -2\cdot(M\rho^2+\rho^2+1)/N^2 + o(1/N^2)$,
(xiv): $E[R_2 \cdot R_4] = (M\rho\mu_{03}+\rho\mu_{03}+\mu_{12})/N^2 + o(1/N^2)$,
(xv): $E[R_3 \cdot R_4] = -\left(M\rho\mu_{03}+\rho\mu_{03}+\mu_{12}\right)/N^2 + o(1/N^2)$.
(i): See Lemma 3, part (i).
(ii): This follows directly from Lemma 3, part (ii).
(iii):

$$E[T_1 \cdot R_2] = E\left[\overline{w}_1\overline{w}_1(\overline{ww'}_{11}-1)\right] = E\left[\frac{1}{N^3}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}w_{i1}w_{j1}(w_{k1}^2-1)\right]$$

$$= E\left[\frac{1}{N^3}\sum_{i=1}^{N} w_{i1}w_{i1}(w_{i1}^2-1)\right] = (\mu_{04}-1)/N^2.$$

(iv):

$$E[T_1\cdot R_3] = -E\left[\overline{w}_1\overline{w}'(\overline{ww'}-\mathcal{I}_M)e_1\right] = -E\left[\frac{1}{N^3}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} w_{i1}w_j'(w_k w_k'-\mathcal{I}_M)e_1\right]$$

$$= -E\left[\frac{1}{N^3}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M} w_{i1}w_{jm}(w_{km}w_{kn}-\delta_{mn})e_{1n}\right]$$

$$= -E\left[\frac{1}{N^3}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{m=1}^{M} w_{i1}w_{jm}(w_{km}w_{k1}-\delta_{m1})\right]$$

$$= -E\left[\frac{1}{N^3}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} w_{i1}w_{j1}(w_{k1}^2-1)\right]$$

$$= -E\left[\frac{1}{N^3}\sum_{i=1}^{N} w_{i1}w_{i1}(w_{i1}^2-1)\right] = -(\mu_{04}-1)/N^2.$$

(v):

$$E[T_1\cdot R_4] = E[\overline{w}_1\overline{v}'\overline{w}] = E\left[\frac{1}{N^3}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{m=1}^{M} w_{i1}v_{jm}w_{km}\right]$$

$$= E\left[\frac{1}{N^3}\sum_{i=1}^{N} w_{i1}v_{i1}w_{i1}\right] = \mu_{12}/N^2.$$

(vi): This follows directly from Lemma 3, part (iii).

(vii):

$$E[R_2^2] = E\left[\left(\overline{ww'}_1-1)\overline{w}_1\right)^2\right] = E\left[\left(\frac{1}{N^2}\sum_{j=1}^{N}\sum_{i=1}^{N}(w_{i1}^2-1)\cdot w_{j1}\right)^2\right]$$

$$= \frac{1}{N^4}E\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}(w_{i1}^2-1)\cdot w_{j1}\cdot(w_{k1}^2-1)\cdot w_{l1}\right]$$

$$= \frac{1}{N^4}E\left[\sum_{i=1}^{N}\sum_{k\neq i}(w_{i1}^2-\text{‘})\cdot w_{i1}\cdot(w_{k1}^2-1)\cdot w_{k1}\right.$$

$$+\sum_{i=1}^{N}\sum_{j\neq i}(w_{i1}^2-1)\cdot w_{j1}\cdot(w_{j1}^2-1)\cdot w_{i1}$$

$$+\sum_{i=1}^{N}\sum_{j\neq i}(w_{i1}^2-1)\cdot w_{j1}\cdot(w_{i1}^2-1)\cdot w_{j1}\Bigg]+o_p(1/N^2)$$

$$=\frac{1}{N^2}\left(\mu_{03}^2+\mu_{03}^2+\mu_{04}-1)\right)/N^2+o_p(1/N^2)$$

$$=\frac{1}{N^2}\left(2\mu_{03}^2+\mu_{04}-1\right)/N^2+o_p(1/N^2).$$

$(viii)$,

$$E[R_1\cdot R_2]=E\left[-2\overline{v}_1\overline{w}_1(\overline{ww'}_{11}-1)\overline{w}_1\right]=-\frac{2}{N^4}\cdot E\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}v_{i1}w_{j1}(w_{k1}^2-1)w_{l1}\right]$$

$$=-\frac{1}{N^4}E\Bigg[\sum_{i=1}^{N}\sum_{k\neq i}v_{i1}w_{i1}(w_{k1}^2-1)w_{k1}$$

$$+\sum_{i=1}^{N}\sum_{j\neq i}v_{i1}w_{j1}(w_{j1}^2-1)w_{i1}$$

$$+\sum_{i=1}^{N}\sum_{j\neq i}v_{i1}w_{j1}(w_{i1}^2-1)w_{j1}\Bigg]+o_p(1/N^2)$$

$$=-\frac{2}{N^2}\left(\rho\mu_{03}+\rho\mu_{03}+\mu_{12}\right)+o_p(1/N^2)$$

$$=-2(2\rho\mu_{03}+\mu_{12})/N^2+o_p(1/N^2).$$

(ix),

$$E[R_3^2]=E\left[\left(e_1'(\overline{ww'}-\mathcal{I}_M)\overline{w}\right)^2\right]$$

$$=\frac{1}{N^4}E\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M}(w_{i1}w_{im}-\delta_{1m})w_{jm}(w_{k1}w_{kn}-\delta_{1n})w_{ln}\right]$$

$$=\frac{1}{N^4}E\Bigg[\sum_{i=1}^{N}\sum_{k\neq i}\sum_{m=1}^{M}\sum_{n=1}^{M}(w_{i1}w_{im}-\delta_{1m})w_{im}(w_{k1}w_{kn}-\delta_{1n})w_{kn}$$

$$+\sum_{i=1}^{N}\sum_{j\neq i}\sum_{m=1}^{M}\sum_{n=1}^{M}(w_{i1}w_{im}-\delta_{1m})w_{jm}(w_{j1}w_{jn}-\delta_{1n})w_{in}$$

$$+\sum_{i=1}^{N}\sum_{k\neq i}\sum_{m=1}^{M}\sum_{n=1}^{M}(w_{i1}w_{im}-\delta_{1m})w_{jm}(w_{i1}w_{in}-\delta_{1n})w_{jn}\Bigg]+o_p(1/N^2)$$

$$=\frac{1}{N^4}E\Bigg[\sum_{i=1}^{N}\sum_{k\neq i}(w_{i1}w_{i1}-)w_{i1}(w_{k1}w_{k1}-1)w_{k1}$$

$$+\sum_{i=1}^{N}\sum_{j\neq i}(w_{i1}w_{i1}-1)w_{j1}(w_{j1}w_{j1}-1)w_{i1}$$

$$+\sum_{i=1}^{N}\sum_{k\neq i}(w_{i1}w_{i1}-1)w_{j1}(w_{i1}w_{i1}-1)w_{j1}$$

$$+\sum_{i=1}^{N}\sum_{k\neq i}\sum_{m\neq 1}w_{i1}w_{im}w_{jm}w_{i1}w_{im}w_{jm}\Bigg]+o_p(1/N^2)$$

$$=\frac{1}{N^2}\left(\mu_{03}^2+\mu_{03}^2+\mu_{04}-1+M-1\right)+o_p(1/N^2).$$

$$=\frac{1}{N^2}\left(2\mu_{03}^2+\mu_{04}+M-2\right)+o_p(1/N^2).$$

(x),

$$E[R_1\cdot R_3]=E[(-2\overline{v}_1\overline{w}_2)(-e_1(\overline{ww'}-\mathcal{I}_M)\overline{w})]$$

$$=2\cdot E\left[\overline{v}_1\overline{w}_1e_1'(\overline{ww'}-\mathcal{I}_M)\overline{w}\right]$$

$$=\frac{2}{N^4}\cdot E\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{M}v_{i1}w_{j1}(w_{k1}w_{km}-\delta_{m1})w_{lm}\right]$$

$$=\frac{2}{N^4}E\Bigg[\sum_{i=1}^{N}\sum_{k\neq i}\sum_{m=1}^{M}v_{i1}w_{i1}(w_{k1}w_{km}-\delta_{m1})w_{km}$$

$$+\sum_{i=1}^{N}\sum_{j\neq i}\sum_{m=1}^{M}v_{i1}w_{j1}(w_{j1}w_{jm}-\delta_{m1})w_{im}$$

$$+\sum_{i=1}^{N}\sum_{j\neq i}\sum_{m=1}^{M} v_{i1}w_{j1}(w_{i1}w_{im}-\delta_{m1})w_{jm}\Bigg] + o_p(1/N^2)$$

$$= \frac{2}{N^4}E\left[\sum_{i=1}^{N}\sum_{k\neq i} v_{i1}w_{i1}(w_{k1}w_{k1}-1)w_{k1}\right.$$

$$+\sum_{i=1}^{N}\sum_{j\neq i} v_{i1}w_{j1}(w_{j1}w_{j1}-1)w_{i1}$$

$$\left.+\sum_{i=1}^{N}\sum_{j\neq i} v_{i1}w_{j1}(w_{i1}w_{i1}-1)w_{j1}\right] + o_p(1/N^2)$$

$$= \frac{2}{N^2}(\rho\mu_{03}+\rho\mu_{03}+\mu_{12}) + o_p(1/N^2).$$

$$= \frac{2}{N^2}(2\rho\mu_{03}+\mu_{12}) + o_p(1/N^2).$$

(xi),

$$E[R_2\cdot R_3] = -E\left[(\overline{ww'}_{11}-1)\overline{w}_1 e_1'(\overline{ww'}-\mathcal{I}_M)\overline{w}\right]$$

$$= -\frac{1}{N^4}E\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{M}(w_{i1}^2-1)w_{j1}(w_{k1}w_{km}-\delta_{1m})w_{lm}\right]$$

$$= -\frac{1}{N^4}E\left[\sum_{i=1}^{N}\sum_{k\neq i}\sum_{m=1}^{M}(w_{i1}^2-1)w_{i1}(w_{k1}w_{km}-\delta_{1m})w_{km}\right.$$

$$+\sum_{i=1}^{N}\sum_{j\neq i}\sum_{m=1}^{M}(w_{i1}^2-1)w_{j1}(w_{j1}w_{jm}-\delta_{1m})w_{im}$$

$$\left.+\sum_{i=1}^{N}\sum_{j\neq i}\sum_{m=1}^{M}(w_{i1}^2-1)w_{j1}(w_{i1}w_{im}-\delta_{1m})w_{jm}\right] + o_p(1/N^2)$$

$$= -\frac{1}{N^4}E\left[\sum_{i=1}^{N}\sum_{k\neq i}(w_{i1}^2-1)w_{i1}(w_{k1}^2-1)w_{k1}\right.$$

$$+\sum_{i=1}^{N}\sum_{j\neq i}(w_{i1}^2-1)w_{j1}(w_{j1}^2-\delta_{1m})w_{i1}$$

$$+\sum_{i=1}^{N}\sum_{j\neq i}(w_{i1}^2-1)w_{j1}(w_{i1}^2-1)w_{j1}\Bigg] + o_p(1/N^2)$$

$$= -\frac{1}{N^2}(\mu_{03}^2+\mu_{03}^2+\mu_{04}-1) + o_p(1/N^2)$$

$$= -\frac{1}{N^2}(2\mu_{03}^2+\mu_{04}-1) + o_p(1/N^2)$$

(xii),

$$E[R_4^2] = E[(\overline{v}'\overline{w})^2] = \frac{1}{N^4}E\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M} v_{im}w_{jm}v_{kn}w_{ln}\right]$$

$$= \frac{1}{N^4}E\Bigg[\sum_{i=1}^{N}\sum_{k\neq i}\sum_{m=1}^{M}\sum_{n=1}^{M} v_{im}w_{im}v_{kn}w_{kn}$$

$$+\sum_{i=1}^{N}\sum_{j\neq i}\sum_{m=1}^{M}\sum_{n=1}^{M} v_{im}w_{jm}v_{jn}w_{in}$$

$$+\sum_{i=1}^{N}\sum_{j\neq i}\sum_{m=1}^{M}\sum_{n=1}^{M} v_{im}w_{jm}v_{in}w_{jn}\Bigg] + o_p(1/N^2)$$

$$= \frac{1}{N^4}E\Bigg[\sum_{i=1}^{N}\sum_{k\neq i}\sum_{m=1}^{M}\sum_{n=1}^{M} v_{im}w_{im}v_{kn}w_{kn}$$

$$+\sum_{i=1}^{N}\sum_{j\neq i}\sum_{m=1}^{M} v_{im}w_{jm}v_{jm}w_{im}$$

$$+\sum_{i=1}^{N}\sum_{j\neq i}\sum_{m=1}^{M} v_{im}w_{jm}v_{im}w_{jm}\Bigg] + o_p(1/N^2)$$

$$= \frac{1}{N^2}(M^2\rho^2+M\rho^2+M) + o_p(1/N^2).$$

$(xiii)$,

$$E[R_1\cdot R_4] = -2\cdot E[\overline{v}_1\overline{w}_1\overline{v}'\overline{w}] = -\frac{2}{N^4}\cdot E\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{M} v_{i1}w_{j1}v_{km}w_{lm}\right]$$

$$= -\frac{2}{N^4}E\Bigg[\sum_{i=1}^{N}\sum_{k\neq i}\sum_{m=1}^{M} v_{i1}w_{i1}v_{km}w_{km}$$

$$\sum_{i=1}^{N}\sum_{j\neq i}\sum_{m=1}^{M} v_{i1}w_{j1}v_{jm}w_{im}$$

$$\left.\sum_{i=1}^{N}\sum_{j\neq i}\sum_{m=1}^{M} v_{i1}w_{j1}v_{im}w_{jm}\right] + o_p(1/N^2)$$

$$= -\frac{2}{N^4}E\left[\sum_{i=1}^{N}\sum_{k\neq i}\sum_{m=1}^{M} v_{i1}w_{i1}v_{km}w_{km}\right.$$

$$\sum_{i=1}^{N}\sum_{j\neq i} v_{i1}w_{j1}v_{j1}w_{i1}$$

$$\left.\sum_{i=1}^{N}\sum_{j\neq i} v_{i1}w_{j1}v_{i1}w_{j1}\right] + o_p(1/N^2)$$

$$= -\frac{2}{N^2}(M\rho^2 + \rho^2 + 1) + o_p(1/N^2).$$

(xiv),

$$E[R_2 \cdot R_4] = E[(\overline{ww'}_{11} - 1)\overline{w}_1\overline{v}'\overline{w}]$$

$$\frac{1}{N^4}E\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{M}(w_{i1}^2 - 1)w_{j1}v_{km}w_{lm}\right]$$

$$= \frac{1}{N^4}E\left[\sum_{i=1}^{N}\sum_{k\neq i}\sum_{m=1}^{M}(w_{i1}^2 - 1)w_{i1}v_{km}w_{km}\right.$$

$$+\sum_{i=1}^{N}\sum_{j\neq i}\sum_{m=1}^{M}(w_{i1}^2 - 1)w_{j1}v_{jm}w_{im}$$

$$\left.+\sum_{i=1}^{N}\sum_{j\neq i}\sum_{m=1}^{M}(w_{i1}^2 - 1)w_{j1}v_{im}w_{jm}\right] + o_p(1/N^2)$$

$$= \frac{1}{N^4}E\left[\sum_{i=1}^{N}\sum_{k\neq i}\sum_{m=1}^{M}(w_{i1}^2 - 1)w_{i1}v_{km}w_{km}\right.$$

$$+\sum_{i=1}^{N}\sum_{j\neq i}(w_{i1}^2 - 1)w_{j1}v_{j1}w_{i1}$$

$$+\sum_{i=1}^{N}\sum_{j\neq i}(w_{i1}^2-1)w_{j1}v_{i1}w_{j1}\Bigg]+o_p(1/N^2)$$

$$=\frac{1}{N^2}(M\mu_{03}\rho+\mu_{03}\rho+\mu_{12})+o_p(1/N^2).$$

(xv),

$$E[R_3\cdot R_4]=-E[e_1'(\overline{ww'}-\mathcal{I}_M)\overline{wv}'\overline{w}]$$

$$=-\frac{1}{N^4}E\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M}(w_{i1}w_{im}-\delta_{1m})w_{jm}v_{kn}w_{ln}\right]$$

$$=-\frac{1}{N^4}E\Bigg[\sum_{i=1}^{N}\sum_{k\neq i}\sum_{m=1}^{M}\sum_{n=1}^{M}(w_{i1}w_{im}-\delta_{1m})w_{im}v_{kn}w_{kn}$$

$$+\sum_{i=1}^{N}\sum_{j\neq i}\sum_{m=1}^{M}\sum_{n=1}^{M}(w_{i1}w_{im}-\delta_{1m})w_{jm}v_{jn}w_{in}$$

$$+\sum_{i=1}^{N}\sum_{k\neq i}\sum_{m=1}^{M}\sum_{n=1}^{M}(w_{i1}w_{im}-\delta_{1m})w_{jm}v_{in}w_{jn}\Bigg]+o_p(1/N^2)$$

$$=-\frac{1}{N^4}E\Bigg[\sum_{i=1}^{N}\sum_{k\neq i}\sum_{n=1}^{M}(w_{i1}w_{i1}-1)w_{i1}v_{kn}w_{kn}$$

$$+\sum_{i=1}^{N}\sum_{j\neq i}(w_{i1}w_{i1}-1)w_{j1}v_{j1}w_{i1}$$

$$+\sum_{i=1}^{N}\sum_{k\neq i}(w_{i1}w_{i1}-1)w_{j1}v_{i1}w_{j1}\Bigg]+o_p(1/N^2)$$

$$=-\frac{1}{N^2}(M\mu_{03}\rho+\mu_{03}\rho+\mu_{12}).$$

Adding up all the terms leads to

$$E\left[(T_1+R_1+R_2+R_2+R_4-\theta^*)^2\right]$$

$$=1/N-2\mu_{12}/N^2+\rho^2(M^2-3M+4)+2M-1+o_p(1/N^2)$$

$$=1/N-2\mu_{12}/N^2+(1+2\rho^2)/N^2$$

$$+(M-1)^2\rho^2/N^2 + (M-1)(2-\rho^2)/N^2 + o_p(1/N^2).$$

□

Before proving Lemma 7, it is useful to consider the solution for t given θ. Define $\hat{t}(\theta)$ implicitly through the first equation:

$$0 = \sum_{i=1}^{N} \psi(v_i, w_i, \theta) \cdot g(t(\theta)'\psi(v_i, w_i, \theta)).$$

Lemma 11. (Expansion for $\hat{t}(\theta)$)
If $\theta = \hat{\theta}_{\text{opt}} + o_p(1/\sqrt{N})$, *then*

$$\hat{t}(\theta) = -e_1\theta + \overline{w} - \overline{vw}_1 - \overline{ww'(\overline{w} - e_1\overline{w}_1)w'}(\overline{w} - e_1\overline{w}_1)\lambda/2 + (\overline{ww'} - \mathcal{I}_M)e_1\overline{w}_1$$

$$-(\overline{ww'} - \mathcal{I}_M)\overline{w} - 2\rho\overline{w}_1^2 e_1 + 2\rho\overline{w}_1\overline{w} + o_p(1/N).$$

Proof of Lemma 11:
Use a Taylor series expansion around zero for $g(a)$, $g(a) = g(0) + g'(0)a + g''(\tilde{a})a^2/2 = 1 + a + g''(\tilde{a})a^2/2$, to write the equation characterizing $\hat{t}(\theta)$ as

$$0 = \sum_{i=1}^{N} \psi(v_i, w_i, \theta) \cdot \Big(1 + t(\theta)'\psi(v_i, w_i, \theta) + g''(a)\left(t(\theta)'\psi(v_i, w_i, \theta)\right)/2\Big),$$

for some a between zero and $t(\theta)'\psi(v_i, w_i, \theta)$. Hence

$$\hat{t}(\theta) = -\left(\frac{1}{N}\sum_{i=1}^{N} \psi(v_i, w_i, \theta)\psi(v_i, w_i, \theta)'\right)^{-1}$$

$$\times \left(\sum_{i=1}^{N} \psi(v_i, w_i, \theta) + \psi(v_i, w_i, \theta)g''(a)\left(t(\theta)'\psi(v_i, w_i, \theta)\right)^2/2\right).$$

The second step is to show that

(18)
$$\frac{1}{N}\sum_{i=1}^{N} \psi(v_i, w_i, \theta)g''(a)\left(t(\theta)'\psi(v_i, w_i, \theta)\right)^2/2 = \overline{ww'(\overline{w} - e_1\overline{w}_1)w'}(\overline{w} - e_1\overline{w}_1)\lambda/2 + o_p(1/N),$$

To see this, first note that because $\theta = \overline{w}_1 + o_p(1/\sqrt{N})$, we have $\hat{t}(\theta) = \overline{w} - e_1\overline{w}_1 + o_p(1/\sqrt{N})$. Hence

$$\frac{1}{N}\sum_{i=1}^{N} \psi(v_i, w_i, \theta)\left(t(\theta)'\psi(v_i, w_i, \theta)\right)^2$$

$$= \frac{1}{N}\sum_{i=1}^{N}\psi(v_i, w_i, \theta^*)\left(t(\theta)'\psi(v_i, w_i, \theta^*)\right)^2 + o_p(1/N)$$

$$= \frac{1}{N}\sum_{i=1}^{N} w_i\left(t(\theta)'w_i\right)^2 + o_p(1/N)$$

$$= \frac{1}{N}\sum_{i=1}^{N} w_i\left((\overline{w} - e_1\overline{w}_1)'w_i\right)^2 + o_p(1/N)$$

$$= \overline{ww'(\overline{w} - e_1\overline{w}_1)w'}(\overline{w} - e_1\overline{w}_1) + o_p(1/N).$$

Since $t = o_p(1)$, $a = o_p(1)$, and $g''(a) = \lambda + o_p(1)$, so that the result in equation (18) follows.
The third step is to show that, with $\theta = \overline{w}_1 + o_p(1/\sqrt{N})$, we have

$$\left[\frac{1}{N}\sum_{i=1}^{N}\psi(v_i, w_i, \theta)\psi(v_i, w_i, \theta)'\right]^{-1} = \mathcal{I}_M + (\overline{ww'} - \mathcal{I}_M) + 2\rho\overline{w}_1\mathcal{I}_M + o_p(1/\sqrt{N}). \tag{19}$$

To see this, first write out

$$\frac{1}{N}\sum_{i=1}^{N}\psi(v_i, w_i, \theta)\psi(v_i, w_i, \theta)' = \overline{ww'} - 2\overline{vw'}\theta - 2e_1\overline{w}'\theta + \theta^2\overline{(v+e_1)(v+e_1)'}$$

$$= \mathcal{I}_M + (\overline{ww'} - \mathcal{I}_M) - 2\rho\mathcal{I}_M\theta + o_p(1/\sqrt{N})$$

$$= \mathcal{I}_M + (\overline{ww'} - \mathcal{I}_M) - 2\rho\mathcal{I}_M\overline{w}_1 + o_p(1/\sqrt{N}).$$

Hence, using Lemma 10,

$$\left[\frac{1}{N}\sum_{i=1}^{N}\psi(v_i, w_i, \theta)\psi(v_i, w_i, \theta)'\right]^{-1} = \mathcal{I}_M - (\overline{ww'} - \mathcal{I}_M) + 2\rho\mathcal{I}_M\overline{w}_1 + o_p(1/\sqrt{N}),$$

which proves the equality in equation (19).
Then, using the fact that $\overline{\psi(v, w, \theta)} = (\overline{v} + e_1)\theta - \overline{w}$, we can approximate the expression for $\hat{t}(\theta)$ as

$$\hat{t}(\theta) = -\left(\mathcal{I}_M - (\overline{ww'} - \mathcal{I}_M) + 2\rho\mathcal{I}_M\overline{w}_1\right)\left((\overline{v} + e_1)\theta - \overline{w} + \overline{ww'(\overline{w} - e_1\overline{w}_1)w'}(\overline{w} - e_1\overline{w}_1)\lambda/2\right)$$

$$= -(\overline{v} + e_1)\theta + \overline{w} - \overline{ww'(\overline{w} - e_1\overline{w}_1)w'}(\overline{w} - e_1\overline{w}_1)\lambda/2 + (\overline{ww'} - \mathcal{I}_M)e_1\overline{w}_1$$

$$-(\overline{ww'} - \mathcal{I}_M)\overline{w} - 2\rho\overline{w}_1^2 e_1 + 2\rho\overline{w}_1\overline{w} + o_p(1/N).$$

$$= -e_1\theta + \overline{w} - \overline{w}_1\overline{v} - \overline{ww'(\overline{w} - e_1\overline{w}_1)w'}(\overline{w} - e_1\overline{w}_1)\lambda/2 + (\overline{ww'} - \mathcal{I}_M)e_1\overline{w}_1$$

$$-(\overline{ww'} - \mathcal{I}_M)\overline{w} - 2\rho\overline{w}_1^2 e_1 + 2\rho\overline{w}_1\overline{w} + o_p(1/N).$$

□

Proof of Lemma 7

The solution for $\hat{\theta}_\lambda$ is characterized by the equation

$$0 = \hat{t}(\theta)'\frac{1}{N}\sum_{i=1}^{N}\frac{\partial\psi}{\partial\theta'}(v_i, w_i, \theta)\cdot g(\hat{t}(\theta)'\psi(v_i, w_i, \theta)).$$

We can write this as

$$0 = \Big(\Big[-e_1\theta + \overline{w} - \overline{w}_1\overline{v} - \overline{ww'(\overline{w} - e_1\overline{w}_1)w'}(\overline{w} - e_1\overline{w}_1)\lambda/2 + (\overline{ww'} - \mathcal{I}_M)e_1\overline{w}_1 - (\overline{ww'} - \mathcal{I}_M)\overline{w}$$

$$-2\rho\overline{w}_1^2 e_1 + 2\rho\overline{w}_1\overline{w}\Big] + \hat{t}(\theta) - \Big[e_1\theta + \overline{w} - \overline{w}_1\overline{v} - \overline{ww'(\overline{w} - e_1\overline{w}_1)w'}(\overline{w} - e_1\overline{w}_1)\lambda/2$$

$$+(\overline{ww'} - \mathcal{I}_M)e_1\overline{w}_1 - (\overline{ww'} - \mathcal{I}_M)\overline{w} - 2\rho\overline{w}_1^2 e_1 + 2\rho\overline{w}_1\overline{w}\Big]\Big)'$$

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\partial\psi}{\partial\theta'}(v_i, w_i, \theta)\cdot g(\hat{t}(\theta)'\psi(v_i, w_i, \theta)).$$

Hence,

$$\hat{\theta}_\lambda = \left(e_1'\frac{1}{N}\sum_{i=1}^{N}\frac{\partial\psi}{\partial\theta'}(v_i, w_i, \theta)\cdot g(\hat{t}(\theta)'\psi(v_i, w_i, \theta))\right)^{-1}$$

$$\times\Big(\Big[\overline{w} - \overline{w}_1\overline{v} - \overline{ww'(\overline{w} - e_1\overline{w}_1)w'}(\overline{w} - e_1\overline{w}_1)\lambda/2 + (\overline{ww'} - \mathcal{I}_M)e_1\overline{w}_1 - (\overline{ww'} - \mathcal{I}_M)\overline{w}$$

$$-2\rho\overline{w}_1^2 e_1 + 2\rho\overline{w}_1\overline{w}\Big] + \hat{t}(\theta) - \Big[e_1\theta + \overline{w} - \overline{w}_1\overline{v} - \overline{ww'(\overline{w} - e_1\overline{w}_1)w'}(\overline{w} - e_1\overline{w}_1)\lambda/2$$

$$+(\overline{ww'} - \mathcal{I}_M)e_1\overline{w}_1 - (\overline{ww'} - \mathcal{I}_M)\overline{w} - 2\rho\overline{w}_1^2 e_1 + 2\rho\overline{w}_1\overline{w}\Big]\Big)'$$

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\partial\psi}{\partial\theta'}(v_i, w_i, \theta)\cdot g(\hat{t}(\theta)'\psi(v_i, w_i, \theta)).$$

We break this up in a couple of parts. First we show that

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\partial\psi}{\partial\theta'}(v_i, w_i, \theta)\cdot g(\hat{t}(\theta)'\psi(v_i, w_i, \theta)) = e_1 + \overline{v} - \rho\overline{w} + \rho e_1\overline{w}_1 + o_p(1/\sqrt{N}). \tag{20}$$

To see this, write out $\psi(v_i, w_i, \theta) = (v_i + e_1)\theta - w_i$ to get

$$\frac{1}{N}\sum_{i=1}^{N}(v_i + e_1) \cdot g(\hat{t}(\theta)'(v_i\theta + e_1\theta - w_i))$$

$$= \frac{1}{N}\sum_{i=1}^{N}(v_i + e_1) \cdot (1 + t(\theta)'(v_i\theta + e_1\theta - w_i)) + o_p(1/\sqrt{N})$$

$$= e_1 + \overline{v} - \frac{1}{N}\sum_{i=1}^{N} v_i\overline{w}'w_i + o_p(1/\sqrt{N}) = e_1 + \overline{v} - \rho\overline{w} + \rho e_1\overline{w}_1 + o_p(1/\sqrt{N}),$$

which proves the equality in equation (20). A direct implication is that

$$\left(e_1'\frac{1}{N}\sum_{i=1}^{N}\frac{\partial\psi}{\partial\theta'}(v_i, w_i, \theta) \cdot g(\hat{t}(\theta)'\psi(v_i, w_i, \theta))\right)^{-1} = 1 - \overline{v}_1 + o_p(1/\sqrt{N}). \tag{21}$$

Second, we show that

$$\Big(\hat{t}(\theta) - \Big[e_1\theta + \overline{w} - \overline{w}_1\overline{v} - \overline{ww'(\overline{w} - e_1\overline{w}_1)w'}(\overline{w} - e_1\overline{w}_1)\lambda/2$$

$$+(\overline{ww'} - \mathcal{I}_M)e_1\overline{w}_1 - (\overline{ww'} - \mathcal{I}_M)\overline{w} - 2\rho\overline{w}_1^2 e_1 + 2\rho\overline{w}_1\overline{w}\Big]\Big)'$$

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\partial\psi}{\partial\theta'}(v_i, w_i, \theta) \cdot g(\hat{t}(\theta)'\psi(v_i, w_i, \theta)) = o_p(1/N).$$

This follows from Lemma 7, which implies that the first factor is $o_p(1/N)$, combined with the fact that the left hand side of equation (20) is $O_p(1)$. Third, we show that

$$\Big(\overline{w} - \overline{w}_1\overline{v} - \overline{ww'(\overline{w} - e_1\overline{w}_1)w'}(\overline{w} - e_1\overline{w}_1)\lambda/2 + (\overline{ww'} - \mathcal{I}_M)e_1\overline{w}_1 - (\overline{ww'} - \mathcal{I}_M)\overline{w} - 2\rho\overline{w}_1^2 e_1 + 2\rho\overline{w}_1\overline{w}\Big)'$$

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\partial\psi}{\partial\theta'}(v_i, w_i, \theta) \cdot g(\hat{t}(\theta)'\psi(v_i, w_i, \theta))$$

$$= \overline{w}_1 - e_1'\overline{w}_1\overline{v} - e_1'\overline{ww'(\overline{w} - e_1\overline{w}_1)w'}(\overline{w} - e_1\overline{w}_1)\lambda/2 + e_1'(\overline{ww'} - \mathcal{I}_M)e_1\overline{w}_1 - e_1'(\overline{ww'} - \mathcal{I}_M)\overline{w} - 2\rho\overline{\iota}$$

$$+2\rho\overline{w}_1^2 + \overline{w}'\overline{v} - \rho\overline{w}'\overline{w} + \rho\overline{w}_1^2 + o_p(1/N).$$

$$= \overline{w}_1 - \overline{w}_1\overline{v}_1 - e_1'\overline{ww'(\overline{w} - e_1\overline{w}_1)w'}(\overline{w} - e_1\overline{w}_1)\lambda/2 + e_1'(\overline{ww'} - \mathcal{I}_M)e_1\overline{w}_1 - e_1'(\overline{ww'} - \mathcal{I}_M)\overline{w}$$

$$+\overline{w}'\overline{v} - \rho\overline{w}'\overline{w} + \rho\overline{w}_1^2 + o_p(1/N).$$

Now note that although $\overline{ww'(\overline{w}-e_1\overline{w}_1)w'}(\overline{w}-e_1\overline{w}_1) = O_p(1/N)$, $e_1'\overline{ww'(\overline{w}-e_1\overline{w}_1)w'}(\overline{w}-e_1\overline{w}_1) = o_p(1/N)$, because the subtraction of $e_1\overline{w}_1$ from $\overline{w}$ makes $(\overline{w}-e_1\overline{w}_1)$ independent of $e_1 w$. This relies on the full independence assumption we are using in the sequence of the moments. Because of this the term $e_1'\overline{ww'(\overline{w}-e_1\overline{w}_1)w'}(\overline{w}-e_1\overline{w}_1)$ is of lower order, and the above expression reduces to

$$\overline{w}_1 - \overline{w}_1\overline{v}_1 + e_1'(\overline{ww'} - \mathcal{I}_M)e_1\overline{w}_1 - e_1'(\overline{ww'} - \mathcal{I}_M)\overline{w}$$

$$+\overline{w}'\overline{v} - \rho\overline{w}'\overline{w} + \rho\overline{w}_1^2 + o_p(1/N).$$

Finally bringing all the terms together, we get

$$\hat{\theta}_\lambda = \overline{w}_1 + e_1'(\overline{ww'} - \mathcal{I}_M)e_1\overline{w}_1 - e_1'(\overline{ww'} - \mathcal{I}_M)\overline{w}$$

$$+\overline{w}'\overline{v} - \rho\overline{w}'\overline{w} - 2\overline{w}_1\overline{v}_1 + \rho\overline{w}_1^2 + o_p(1/N).$$

□

Proof of Lemma 8:

We first prove
(i), $E[T_1] = \theta^*$,
(ii), $E[R_3] = -\mu_{03}/N$,
(iii) $E[R_5] = \rho/N$,
(iv) $E[R_4] = \rho M/N$,
(v), $E[R_6] = -\rho M/N$,
(vi), $E[R_1] = -\rho/N$,
(vii), $E[R_2] = \mu_{03}/N$,
$(viii)$, $E[R_7] = o_p(1/N)$.
The result then follows from adding up the expectations.
(i): This is immediate.
(ii): See proof of Lemma 5, part (iv).
(iii):

$$E[R_5] = \rho E[\overline{w}_1^2] = \rho E\left[\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N w_{i1}w_{j1}\right] = \rho\frac{1}{N^2}E\left[\sum_{i=1}^N w_{i1}^2\right] = \rho/N.$$

(iv): See proof of Lemma 5, part (v).
(v):

$$E[R_6] = -\rho E[\overline{w}'\overline{w}] = -\rho E\left[\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N\sum_{m=1}^M w_{im}w_{jm}\right] = -\rho\frac{1}{N^2}E\left[\sum_{i=1}^N\sum_{m=1}^M w_{im}^2\right] = -\rho M/N.$$

(vi): See proof of Lemma 5, part (ii).
(vii): See proof of Lemma 5, part (iii).

$(viii)$: First consider

$$E\left[e_1'\overline{ww'\overline{w}w'\overline{w}}\right]\lambda/2 = \frac{\lambda}{2}\frac{1}{N^3}E\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}e_1'w_iw_i'w_jw_i'w_k\right]$$

$$= \frac{\lambda}{2}\frac{1}{N^3}E\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M}w_{i1}w_{im}w_{jm}w_{in}w_{kn}\right]$$

$$= \frac{\lambda}{2}\frac{1}{N^3}E\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M}w_{i1}w_{im}w_{jm}w_{in}w_{jn}\right] + o_p(1/N)$$

$$= \frac{\lambda}{2}\frac{1}{N^3}E\left[\sum_{i=1}^{N}\sum_{j=1}^{N}w_{i1}w_{i1}w_{j1}w_{i1}w_{j1}\right] + o_p(1/N) = \mu_{03}\lambda/(2N) + o_p(1/N).$$

The same argument shows that

$$E\left[e_1'\overline{ww'e_1\overline{w}_1w'\overline{w}}\right]\lambda/2 = \mu_{03}\lambda/(2N) + o_p(1/N),$$

and

$$E\left[e_1'\overline{ww'e_1\overline{w}_1w'}e_1\overline{w}_1\right]\lambda/2 = \mu_{03}\lambda/(2N) + o_p(1/N),$$

which after adding the terms up gives the desired result that

$$E[R_7] = o_p(1/N).$$

□

Proof of Lemma 9:

We first prove
(i), $E[T_1^2] = 1/N$,
(ii), $E[T_1 \cdot R_3] = -(\mu_{04} - 1)/N^2 + o_p(1/N^2)$,
(iii), $E[T_1 \cdot R_5] = \rho\mu_{03}/N^2 + o_p(1/N^2)$,
(iv), $E[T_1 \cdot R_4] = \mu_{12}/N^2 + o_p(1/N^2)$,
(v), $E[T_1 \cdot R_6] = -\rho\mu_{03}/N^2 + o_p(1/N^2)$,
(vi), $E[T_1 \cdot R_1] = -2\mu_{12}/N^2 + o_p(1/N^2)$,
(vii), $E[T_1 \cdot R_2] = (\mu_{04} - 1)/N^2 + o_p(1/N^2)$,
$(viii)$, $E[R_3^2] = (2\mu_{03}^2 + \mu_{04} + M - 2)/N^2 + o_p(1/N^2)$,
(ix), $E[R_3 \cdot R_5] = -3\rho\mu_{03}/N^2 + o_p(1/N^2)$,
(x), $E[R_3 \cdot R_4] = -(M\rho\mu_{03} + \rho\mu_{03} + \mu_{12})/N^2 + o_p(1/N^2)$,
(xi), $E[R_3 \cdot R_6] = \rho\mu_{03}(M + 2)/N^2 + o_p(1/N^2)$,
(xii), $E[R_3 \cdot R_1] = 2(2\rho\mu_{03} + \mu_{12}/N^2 + o_p(1/N^2)$,
$(xiii)$ $E[R_3 \cdot R_2] = -(2\mu_{03}^2 + \mu_{04} - 1)/N^2 + o_p(1/N^2)$,
(xiv), $E[R_5^2] = 3\rho^2/N^2 + o_p(1/N^2)$,

(xv), $E[R_5 \cdot R_4] = \rho(M+2)/N^2 + o_p(1/N^2)$,
(xvi), $E[R_5 \cdot R_6] = -\rho(M+2)/N^2 + o_p(1/N^2)$,
$(xvii)$, $E[R_5 \cdot R_1] = -6\rho^2/N^2 + o_p(1/N^2)$,
$(xviii)$, $E[R_5 \cdot R_2] = 3\rho\mu_{03}/N^2 + o_p(1/N^2)$,
(xix), $E[R_4^2] = (M^2\rho^2 + M\rho^2 + M)/N^2 + o_p(1/N^2)$,
(xx), $E[R_4 \cdot R_6] = -\rho^2(M^2+2)/N^2 + o_p(1/N^2)$,
(xxi), $E[R_4 \cdot R_1] = -2(M\rho^2 + \rho^2 + 1)/N^2 + o_p(1/N^2)$,
$(xxii)$, $E[R_4 \cdot R_2] = (M\rho\mu_{03} + \rho\mu_{03} + \mu_{12})/N^2 + o_p(1/N^2)$,
$(xxiii)$, $E[R_6^2] = \rho^2(M^2+2)/N^2 + o_p(1/N^2)$,
$(xxiv)$, $E[R_6 \cdot R_1] = \rho^2(2M+4)/N^2 + o_p(1/N^2)$,
(xxv), $E[R_6 \cdot R_2] = -\rho\mu_{03}(M+2)/N^2 + o_p(1/N^2)$,
$(xxvi)$, $E[R_1^2] = 4(2\rho^2+1)/N^2 + o_p(1/N^2)$,
$(xxvii)$, $E[R_1 \cdot R_2] = -2(2\rho\mu_{03} + \mu_{12})/N^2 + o_p(1/N^2)$,
$(xxviii)$, $E[R_2^2] = (2\mu_{03}^2 + \mu_{04} - 1)/N^2 + o_p(1/N^2)$.

(i): See Lemma 3, part (i).
(ii): This follows directly from Lemma 6, part (iv).
(iii):

$$E[T_1 \cdot R_5] = E[\rho\overline{w}_1^3] = \rho\frac{1}{N^3}E\left[\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^N w_{i1}w_{j1}w_{k1}\right] = \rho\mu_{03}/N^2.$$

(iv): This follows directly from Lemma 6, part (v).
(v):

$$E[T_1 \cdot R_6] = E[-\rho\overline{w}_1\overline{w}'\overline{w}] = -\rho\frac{1}{N^3}\left[\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^N\sum_{m=1}^M w_{1i}w_{jm}w_{km}\right]$$

$$= -\rho\frac{1}{N^3}\left[\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^N w_{1i}w_{j1}w_{k1}\right] = -\rho\mu_{03}/N^2.$$

(vi): This follows directly from Lemma 6, part (ii).
(vii): This follows directly from Lemma 6, part (iii).
$(viii)$: This follows directly from Lemma 6, part (ix).
(ix):

$$E[R_3 \cdot R_5] = E\left[-\rho\overline{w}_1^2 e_1'(\overline{ww'} - \mathcal{I}_M)\overline{w}\right]$$

$$= -\rho\frac{1}{N^4}E\left[\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^N\sum_{l=1}^N\sum_{m=1}^M w_{i1}w_{j1}(w_{k1}w_{km} - \delta_{m1})w_{lm}\right]$$

$$-\rho\frac{1}{N^3}E\ \square\sum_{i=1}^{N}\sum_{k=1}^{N}\sum_{m=1}^{M}w_{i1}^2(w_{k1}w_{km}-\delta_{m1})w_{km}$$

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}w_{i1}w_{j1}(w_{j1}w_{jm}-\delta_{m1})w_{im}$$

$$\left.\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}w_{i1}w_{j1}(w_{i1}w_{im}-\delta_{m1})w_{jm}\right]+o_p(1/N^2)$$

$$=-\rho\frac{1}{N^4}(\mu_{03}+\mu_{03}+\mu_{03})=-3\rho\mu_{03}/N^2+o_p(1/N^2).$$

(x): This follows directly from Lemma 6, part (xv).
(xi):

$$E[R_3\cdot R_6]=E[\rho\overline{w}'\overline{w}e_1'(\overline{ww'}-\mathcal{I}_M)\overline{w}]$$

$$=\rho\frac{1}{N^4}E\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M}w_{im}w_{jm}(w_{k1}w_{kn}-\delta_{n1})w_{ln}\right]$$

$$=\rho\frac{1}{N^4}E\left[\sum_{i=1}^{N}\sum_{k=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M}w_{im}w_{im}(w_{k1}w_{kn}-\delta_{n1})w_{kn}\right.$$

$$+\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M}w_{im}w_{jm}(w_{j1}w_{jn}-\delta_{n1})w_{in}$$

$$\left.+\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M}w_{im}w_{jm}(w_{i1}w_{in}-\delta_{n1})w_{jn}\right]+o_p(1/N^2)$$

$$=\rho(M\mu_{03}+\mu_{03}+\mu_{03})/N^2+o(1/N^2)=\rho\mu_{03}(M+2)/N^2+o_p(1/N^2).$$

(xii): This follows directly from Lemma 6, part (x).
$(xiii)$: This follows directly from Lemma 6, part (xi).
(xiv):

$$E[R_5^2]=E[\rho^2\overline{w}_1^4]=\rho^2\frac{1}{N^3}E\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}w_{i1}w_{j1}w_{k1}w_{l1}\right]$$

$$=\rho^2\frac{1}{N^3}E\left[\sum_{i=1}^{N}\sum_{k=1}^{N}w_{i1}w_{i1}w_{k1}w_{k1}\right.$$

$$+\sum_{i=1}^{N}\sum_{j=1}^{N} w_{i1}w_{j1}w_{j1}w_{i1}$$

$$\left.+\sum_{i=1}^{N}\sum_{j=1}^{N} w_{i1}w_{j1}w_{i1}w_{j1}\right] + o(1/N^2)$$

$$= 3\rho^2/N^2 + o(1/N^2).$$

(xv):

$$E[R_4 \cdot R_5] = E[\rho\overline{w}_1^2\overline{w}'\overline{v}] = \rho\frac{1}{N^4}E\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{M} w_{i1}w_{j1}w_{km}v_{lm}\right]$$

$$\rho\frac{1}{N^4}E[\left[+\sum_{i=1}^{N}\sum_{k=1}^{N} w_{i1}w_{i1}w_{km}v_{km}\right.$$

$$+\sum_{i=1}^{N}\sum_{j=1}^{N} w_{i1}w_{j1}w_{jm}v_{im}$$

$$+\sum_{i=1}^{N}\sum_{j=1}^{N} w_{i1}w_{j1}w_{im}v_{jm} + o(1/N^2)$$

$$= \rho(M\rho + \rho + \rho)/N^2 + o(1/N^2) = \rho^2(M+2)/N^2 + o(1/N^2).$$

(xvi):

$$E[R_5 \cdot R_6] = E[-\rho\overline{w}_1^2\overline{w}'\overline{v}] = -\rho\frac{1}{N^4}E\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{M} w_{i1}w_{j1}w_{km}v_{lm}\right]$$

$$= -\rho\frac{1}{N^4}E\left[\sum_{i=1}^{N}\sum_{k=1}^{N}\sum_{m=1}^{M} w_{i1}w_{i1}w_{km}v_{km}\right.$$

$$+\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M} w_{i1}w_{j1}w_{jm}v_{im}$$

$$+\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M} w_{i1}w_{j1}w_{im}v_{jm} + o(1/N^2)$$

$$= -\rho((M\rho + \rho + \rho)/N^2 + o(1/N^2) = -\rho^2(M+2)/N^2 + o(1/N^2).$$

$(xvii)$:

$$E[R_1 \cdot R_5] = E\left[-2\rho\overline{w}_1^3\overline{v}_1\right] = -2\rho\frac{1}{N^4}E\left[\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^N\sum_{l=1}^N w_{i1}w_{j1}w_{k1}v_{l1}\right]$$

$$= -2\rho\frac{1}{N^4}E\Bigg[\sum_{i=1}^N\sum_{k=1}^N w_{i1}w_{i1}w_{k1}v_{k1} + \sum_{i=1}^N\sum_{j=1}^N w_{i1}w_{j1}w_{j1}v_{i1}$$

$$+ \sum_{i=1}^N\sum_{j=1}^N w_{i1}w_{j1}w_{i1}v_{j1}\Bigg] + o(1/N^2)$$

$$= -2\rho(\rho+\rho+\rho)/N^2 + o(1/N^2) = -6\rho^2/N^2 + o(1/N^2).$$

$(xviii)$:

$$E[R_1 \cdot R_5] = E\left[\rho\overline{w}_1^2\overline{w}_1(\overline{ww'}_{11} - 1)\right] = \rho\frac{1}{N^4}E\left[\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^N\sum_{l=1}^N w_{i1}w_{j1}w_{k1}(w_{l1}^2 - 1)\right]$$

$$= \rho\frac{1}{N^4}E\Bigg[\sum_{i=1}^N\sum_{k=1}^N w_{i1}w_{i1}w_{k1}(w_{k1}^2 - 1) + \sum_{i=1}^N\sum_{j=1}^N w_{i1}w_{j1}w_{j1}(w_{i1}^2 - 1)$$

$$+ \sum_{i=1}^N\sum_{j=1}^N w_{i1}w_{j1}w_{i1}(w_{j1}^2 - 1)\Bigg] + o(1/N^2)$$

$$= \rho(\mu_{03} + \mu_{03} + \mu_{03})/N^2 + o(1/N^2) = 3\rho\mu_{03}/N^2 + o(1/N^2).$$

(xix): This follows directly from Lemma 6, part (xii).

(xx):

$$E[R_4 \cdot R_6] = E\left[-\rho\overline{w}'\overline{ww}'\overline{v}\right]$$

$$= -\rho\frac{1}{N^4}E\left[\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^N\sum_{l=1}^N\sum_{m=1}^M\sum_{n=1}^M w_{im}w_{jm}w_{kn}v_{ln}\right]$$

$$= -\rho\frac{1}{N^4}E\Bigg[\sum_{i=1}^N\sum_{k=1}^N\sum_{m=1}^M\sum_{n=1}^M w_{im}w_{im}w_{kn}v_{kn}$$

$$+ \sum_{i=1}^N\sum_{j=1}^N\sum_{m=1}^M\sum_{n=1}^M w_{im}w_{jm}w_{jn}v_{in}$$

$$+\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M} w_{im}w_{jm}w_{in}v_{jn} + o(1/N^2)$$

$$\left. = -\rho(M^2\rho+\rho+\rho)/N^2 + o(1/N^2) = -\rho^2(M^2+2)/N^2 \right] + o(1/N^2).$$

(xxi): This follows directly from Lemma 6, part $(xiii)$.
$(xxii)$: This follows directly from Lemma 6, part (xiv).
$(xxiii)$:

$$E[R_6^2] = E[\rho^2\overline{w}'\overline{w}\overline{w}'\overline{w}]$$

$$= \rho^2\frac{1}{N^4}E\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M} w_{im}w_{jm}w_{kn}w_{ln}\right]$$

$$= \rho^2\frac{1}{N^4}E\Bigg[\sum_{i=1}^{N}\sum_{k=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M} w_{im}w_{im}w_{kn}w_{kn}$$

$$+\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M} w_{im}w_{jm}w_{jn}w_{in}$$

$$+\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M}\sum_{n=1}^{M} w_{im}w_{jm}w_{in}w_{jn} + o(1/N^2)$$

$$\left. = \rho^2(M^2+1+1)/N^2 + o(1/N^2) = \rho^2(M^2+2)/N^2 \right] + o(1/N^2).$$

$(xxxiv)$:

$$E[R_1R_6] = E[2\rho\overline{w}'\overline{w}\overline{w}_1\overline{v}_1]$$

$$= 2\rho\frac{1}{N^4}\cdot E\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{M} w_{i1}w_{j1}v_{km}w_{lm}\right]$$

$$= 2\rho\frac{1}{N^4}E\Bigg[\sum_{i=1}^{N}\sum_{k\neq i}\sum_{m=1}^{M} w_{i1}w_{i1}v_{km}w_{km}$$

$$+\sum_{i=1}^{N}\sum_{j\neq i}\sum_{m=1}^{M} w_{i1}w_{j1}v_{jm}w_{im}$$

$$+\sum_{i=1}^{N}\sum_{j\neq i}\sum_{m=1}^{M} w_{i1}w_{j1}v_{im}w_{jm}\Bigg] + o_p(1/N^2)$$

$$= \rho^2(2M+4)/N^2 + o_p(1/N^2).$$

(xv):

$$E[R_2\cdot R_6] = E\left[-\rho\overline{w}'\overline{ww}_1(\overline{ww'}_{11}-1)\right] = -\rho\frac{1}{N^4}E\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{M} w_{im}w_{jm}w_{k1}(w_{l1}^2-1)\right]$$

$$-\rho\frac{1}{N^4}E\Bigg[\sum_{i=1}^{N}\sum_{k=1}^{N}\sum_{m=1}^{M} w_{im}w_{im}w_{k1}(w_{k1}^2-1)$$

$$+\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M} w_{im}w_{jm}w_{j1}(w_{i1}^2-1)$$

$$+\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{M} w_{im}w_{jm}w_{i1}(w_{j1}^2-1) + o_p(1/N^2)$$

$$= -\rho(M\mu_{03}+\mu_{03}+\mu_{03})/N^2 + o(1/N^2) = -\rho\mu_{03}(M+2)/N^2 + o(1/N^2).$$

$(xxvi)$: This follows directly from Lemma 6, part (vi).
$(xxvii)$: This follows directly from Lemma 6, part $(viii)$.
$(xxviii)$: This follows directly from Lemma 6, part (vii).

References

(preliminary and VERY incomplete).

Corcoran, S.A., (1998): "Bartlett adjustment of empirical discrepancy statistics," *Biometrika,* 85, 967–972.

Newey, W.K., and Smith, R.J., (2000): "Asymptotic bias and equivalence of GMM and GEL," mimeo, MIT and Bristol.

Newey, W.K., and Smith, R.J., (2001): "Higher order properties of GMM and generalized empirical likelihood estimators" mimeo, MIT and Bristol.

Rothenberg, T. (1999): "Higher order properties of empirical likelihood for simulataneous equations," seminar notes, Harvard–MIT econometrics seminar.

7.