Jackknife CUE With Many Weak Instruments and Nearly Singular Design

Mehmet Caner
North Carolina State University
Nese Yildiz
University of Rochester

November 4, 2007

1 Nearly Singular Design and Many Weak Instruments

In this section we introduce the concept of nearly singular design in Knight and Fu (2000), Knight (2007) and Caner (2006). Then we analyze the case of weak instruments interacting with nearly singular design. Before these developments, the moment restrictions are as follows.

$$Eg(x_i, \beta_0) = 0, \tag{1}$$

for $i = 1, 2 \cdots, n$. In shortcut we write $g_i(\beta_0)$ instead of $g(x_i, \beta_0)$. As in Newey and Windmeijer (2005), we consider iid data x_i, with data generating process to depend on the sample size. Dependence on n is suppressed. β is $p \times 1$ parameter vector, β_0 is the true value. This is assumed to be in the interior of the compact parameter space $\mathbf{B}$, $\mathbf{B} \subset R^p$. The number of moment restrictions are "m" and the behavior of these will be discussed below $m \geq p$. Now we setup some notation for the subsequent sections.

$$\Psi_n(\beta) = n^{-1/2} \sum_{i=1}^{n} (g_i(\beta) - Eg_i(\beta)),$$

$$\hat{\Omega}(\beta) = n^{-1} \sum_{i=1}^{n} g_i(\beta) g_i(\beta)',$$

$$\Omega(\beta) = Eg_i(\beta) g_i(\beta)'.$$

We now provide nearly singular design assumption and a regularity condition that provide central theorem type of results in Caner (2006). These are extensions of the framework introduced by Knight and Fu (2000) from least squares based estimation to two-step GMM and Continuous Updating (CUE) estimation.

1.1 Assumptions

Assumption 1

(i). Uniformly over $\beta \in \mathbf{B}$

$$\hat{\Omega}(\beta) \xrightarrow{p} \Omega(\beta),$$

where $\Omega(\beta)$ *is singular for all* $\beta \in \mathbf{B}$.

(ii). However, for $a_n = n^s$, $0 < s < 1$, *uniformly over* $\mathbf{B}$

$$a_n(\hat{\Omega}(\beta) - \Omega(\beta)) \xrightarrow{p} D_1(\beta),$$

where $D_1(\beta)$ is assumed to be positive definite, for u which is on the nullspace of $\Omega(\beta)$ (i.e. for $u \neq 0$, $u'D_1(\beta)u > 0$, where $\Omega(\beta)u = 0$ for all β.) $D_1(\beta)$ is continuous in β.

(iii).

$$\sup_\beta \|D_1(\beta)\| < \infty.$$

This assumption is extended from nearly singular design assumption of Knight and Fu (2000) and Knight (2007). This basically assumes that there are similar moment conditions that give rise to nearly singular design. Empirical examples that show nearly singular design in asset pricing models are given in Caner (2006). The second assumption is more of a regularity assumption that is helpful in obtaining Central Limit Theorem type of results as in Caner (2006).

Assumption 2.

(i).

$$E \sup_\beta \|g_i(\beta)\|^\xi < \infty,$$

where $\xi > 2/(1-s)$.

(ii).

$$|g_i(\beta_2) - g_i(\beta_1)| \leq L_i|\beta_2 - \beta_1|,$$

where $EL_i^{2+\delta} < \infty$, for $\delta > 0$, and L_i does not depend on β.

Under Assumptions 1 and 2, we find that on the nullspace of $\Omega(\beta_0)$, setting $D_1(\beta_0) = D_1$

$$\frac{a_n^{1/2}}{n^{1/2}} \sum_{i=1}^n g_i(\beta_0) \xrightarrow{d} N(0, D_1). \tag{2}$$

Then also under the same assumptions

$$a_n^{1/2}\Psi_n(\beta) \Longrightarrow \Psi(\beta), \tag{3}$$

where $\Psi(\beta)$ is a zero-mean Gaussian process with $var(u'\Psi(\beta)) = u'D_1(\beta)u$ for all β, and u is in the nullspace of $\Omega(\beta)$. Also under Assumptions 1 and 2

$$\sup_\beta |\frac{a_n}{n} \sum_{i=1}^n (g_i(\beta) - Eg_i(\beta))| \xrightarrow{p} 0. \tag{4}$$

These are Theorem 1 and Lemma A.1 in Caner (2006). We see that nearly singular design slows down the rate of convergences in Central Limit Theorem and Law of Large Numbers.

Next we prove a Lemma that is useful in regulating the behavior of $\hat{\Omega}(\beta)^{-1}$ matrix when we have Assumption 1. This is a new result in the nearly-singular design literature. Note that det(M) represents the determinant of the matrix and adj (M) represents the adjoint of the matrix M.

Lemma 1. *Suppose* $\hat{\Omega}(\beta)$ *and* $\Omega(\beta)$ *are both* $m \times m$*, symmetric matrices that satisfy Assumption 1. Assume that* $\hat{\Omega}(\beta)^{-1} \xrightarrow{p} \infty$ *(*$adj(\Omega(\beta)) \neq 0$*) . Take an increasing sequence* $\{b_n\}_{n=1}^{\infty}$ *satisfying*

$$b_n - \frac{1}{det(\hat{\Omega}(\beta))} \xrightarrow{p} 0,$$

uniformly over β*. Then uniformly over* β *we obtain*

$$\hat{\Omega}^{-1}(\beta) - b_n adj(\Omega(\beta)) \xrightarrow{p} D_2(\beta),$$

where $D_2(\beta)$ *is positive definite on the nullspace of* $adj(\Omega(\beta))$*.*

We have the following Assumption that is needed for the asymptotic normality proof. This is a simple extension of Lemma 1, and controls the rate of convergence in that Lemma.

Assumption 3. *Uniformly over* β*,*

$$\hat{\Omega}(\beta)^{-1} - b_n adj((\Omega(\beta)) \xrightarrow{p} D_2(\beta) + o(\frac{1}{\sqrt{m a_n}}),$$

and the maximum eigenvalue of $D_2(\beta)$ *is finite for all* β*.*

The next two Assumptions are of more technical nature, and are used in Newey and Windmeijer (2007). These are modified to take into account the near singular design in our case. Set $\hat{g}(\beta) = \frac{1}{n}\sum_{i=1}^{n} g_i(\beta)$. Also we set $\bar{g}(\beta) = E g_i(\beta)$.

Assumption 4.

(i).There is a $p \times p$ *matrix* $S_n = \tilde{S}_n diag(\mu_{1n}, \cdots, \mu_{pn})$ *such that* $\tilde{S}_n$ *is bounded away from zero. For each j, either* $\mu_{jn} = \sqrt{n a_n}$ *or* $\mu_{jn}/\sqrt{n a_n} \to 0$*,* $\mu_n = \min_{1 \leq j \leq p} \mu_{jn} \to \infty$*. Also* $a_n^2 m/\mu_n^2$ *is bounded.*

(ii). Define $\delta(\beta) = S_n'(\beta - \beta_0)/\mu_n$*. For every* $\tilde{C} > 0$*, there is* C *and* $\hat{M} = O_p(1)$*, such that for all* $\beta_1, \beta_2 \in B$*,* $\|\delta(\beta_1)\| \leq \tilde{C}, \|\delta(\beta_2)\| \leq \tilde{C}$*,*

$$\frac{n a_n}{\mu_n^2}\|\bar{g}(\beta_2)' D_2(\beta_2)\bar{g}(\beta_2) - \bar{g}(\beta_1)' D_2(\beta_1)\bar{g}(\beta_1)\| \leq C\|\delta(\beta_2 - \beta_1)\|,$$

and

$$\frac{n a_n}{\mu_n^2}\|\hat{g}(\beta_2)' D_2(\beta_2)\hat{g}(\beta_2) - \hat{g}(\beta_1)' D_2(\beta_1)\hat{g}(\beta_1)\| \leq \hat{M}\|\delta(\beta_2 - \beta_1)\|.$$

(iii).

$$\sup_{\beta \in B}\|\hat{g}(\beta)\| = O_p(\mu_n/\sqrt{a_n n}).$$

Assumption 5.

(i).There exists $C > 0$, $\hat{M} = O_p(1)$ *such that wpa1,*

$$\|\delta(\beta)\| \leq C\sqrt{na_n}\|\hat{g}(\beta)\|/\mu_n + \hat{M},$$

for all $\beta \in B$.

(ii). There is $C > 0$, *with*

$$\|\delta(\beta)\| \leq C\sqrt{(n-1)a_n}\|\bar{g}(\beta)\|/\mu_n,$$

for all $\beta \in B$.

Assumption 4 is basically needed for stochastic equicontinuity proof for the uniform law of large numbers (Lemma A.2 in Appendix) result in the consistency argument. Assumption 5 is a global identification condition that is used in Newey and Windmeijer (2007). The main difference here is the convergence rates are affected by the near singular design (i.e. the rate a_n).

For example in Assumption 4i, in the strong identification case $\mu_n = \sqrt{na_n}$ because of the near singular design, rather than $\sqrt{n}$ in regular GEL case of Newey and Windmeijer (2007). The same is true for the less strong identification as denoted by $\mu_n/\sqrt{na_n} \to 0$.

Before the next assumption set $D_2 = D_2(\beta_0)$, and $g_i = g_i(\beta_0), G_i = \partial g_i(\beta_0)/\partial\beta$, $G = EG_i$.

Assumption 6.$g(.)$ *is twice continuously differentiable in a neighborhood* N *of* β_0

(i).

$$\frac{a_n^2 m}{n}[E\|g_i\|^4 + E\|G_i\|^4] \to 0,$$

(ii).

$$na_n S_n^{-1} G' D_2 D_1 D_2 G S_n^{-1'} \to H,$$
$$na_n S_n^{-1} G' D_2 G S_n^{-1'} \to \tilde{H}.$$

H *and* $\tilde{H}$ *is nonsingular*

This assumption is similar to Assumption 2 and 6 in Newey and Windmeijer (2007). The main difference is the rate corresponding to nearly singular design a_n. The main difference for example is the additional rate a_n in Assumption 6ii. Compared with many weak moments case of Newey and Windmeijer (2007), partial derivative converges to zero faster due to nearly-singular design in many weak moments. We also have two limit matrices, this is due to the nearly singular design as well, note that $D_2 D_1 \neq I$.

Before providing the assumption that is helpful in deriving the limit of the score and the Hessian matrix we set up some notation. Let $\hat{G}(\beta) = \sum_{i=1}^{n} G_i(\beta)/n$, and

$$\hat{\Omega}^k(\beta) = \frac{1}{n}\sum_{i=1}^{n} g_i(\beta)\frac{\partial g_i(\beta)'}{\partial \beta_k},\ \Omega^k(\beta) = E[\hat{\Omega}^k(\beta)],$$

$$\hat{\Omega}^{kl}(\beta) = \frac{1}{n}\sum_{i=1}^{n} g_i(\beta)\frac{\partial^2 g_i(\beta)'}{\partial \beta_k \beta_l},\ \Omega^{kl}(\beta) = E[\hat{\Omega}^{kl}(\beta)]$$

$$\hat{\Omega}^{k,l}(\beta) = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial g_i(\beta)'}{\partial \beta_k}\frac{\partial g_i(\beta)'}{\partial \beta_l},\ \Omega^{k,l}(\beta) = E[\hat{\Omega}^{k,l}(\beta)],$$

Assumption 7. *(i).*

$$\lambda_{max}(E[G_i G_i']) \le C,\ \lambda_{max}(E[\frac{\partial G_i(\beta_0)}{\partial \beta_j}\frac{\partial G_i(\beta_0)'}{\partial \beta_j}]) \le C,\ \sqrt{a_n n}\|E[\frac{\partial G_i(\beta_0)}{\partial \beta_j}]S_n^{-1'}\| \le C,$$

(ii). If $\bar{\beta} \xrightarrow{p} \beta_0$, *then*

$$\|\sqrt{a_n n}[\hat{G}(\bar{\beta}) - \hat{G}(\beta_0)]S_n^{-1'}\| \xrightarrow{p} 0,\ \|\sqrt{n a_n}[\partial \hat{G}(\bar{\beta})/\partial \beta_k - \partial \hat{G}(\beta_0)/\beta_k]S_n^{-1'}\| \xrightarrow{p} 0,\ k = 1, 2, \cdots, p$$

(iii). For all β *on a neighborhood* N *of* β_0 *and* A *equal to* $\Omega^k, \Omega^{kl}, \Omega^{k,l}$

$$\sup_{\beta \in N} \|\hat{A}(\beta) - A(\beta)\| \xrightarrow{p} 0,$$

$$|a'[A(\beta_2) - A(\beta_1)]b| \le C\|a\|\|b\|\|\beta_2 - \beta_1\|$$

.

1.2 Issues with GEL

The important issue is if we have weak instruments as defined in Stock and Wright (2000) how this may affect the results found by Caner (2006). In Stock and Wright (2000) weak identification is defined as

$$E g_i(\beta) = \frac{m_1(\beta)}{n^{1/2}},$$

where $m_1(\beta) = 0$, if and only of $\beta = \beta_0$. In standard identified models it is assumed , $E g_i(\beta) = m_1(\beta)$. So this weak identification assumption shows that in large samples the moment converges to zero regardless of parameter values and creating the identification problem. The weak identification assumption leads to inconsistent estimates of two-step GMM and Continuous Updating Estimators. Note that if we have nearly-singular design in two-step GMM and CUE the estimates are consistent. Nearly singular design only affects the rate of convergence of estimators, their rates are slower than root n.

We can define the weak identification in a nearly-singular design as follows:

$$Eg_i(\beta) = \frac{m_1(\beta)}{n^{1/2} a_n^{1/2}}, \tag{5}$$

where $m_1(\beta) \neq 0$, for $\beta \neq \beta_0$. We can easily show that this results in inconsistent estimates in two-step GMM and CUE. We only analyze the case of CUE, the two-step GMM is very similar hence omitted.

$$\begin{aligned} n a_n \hat{Q}_C(\beta) &= \left[\frac{a_n^{1/2}}{n^{1/2}} \sum_{i=1}^n g_i(\beta) \right]' \left[\hat{\Omega}(\beta) \right]^{-1} \left[\frac{a_n^{1/2}}{n^{1/2}} \sum_{i=1}^n g_i(\beta) \right] \\ &= \left[\frac{a_n^{1/2}}{n^{1/2}} \sum_{i=1}^n (g_i(\beta) - Eg_i(\beta)) + a_n^{1/2} n^{1/2} (Eg_i(\beta)) \right]' \left[\hat{\Omega}(\beta) \right]^{-1} \\ &\times \left[\frac{a_n^{1/2}}{n^{1/2}} \sum_{i=1}^n (g_i(\beta) - Eg_i(\beta)) + a_n^{1/2} n^{1/2} (Eg_i(\beta)) \right] \end{aligned}$$

So that by (3) and Lemma 1, adding and subtracting $adj(\Omega(\beta))$ to weight $\hat{\Omega}(\beta)^{-1}$,

$$n a_n \hat{Q}_C(\beta) \Longrightarrow [\Psi(\beta) + m_1(\beta)]' [D_2(\beta)] [\Psi(\beta) + m_1(\beta)],$$

on the nullspace of $adj(\Omega(\beta))$. If we multiply $\hat{Q}_C(\beta)$ by slower rate than a_n, $r_n = o(a_n)$, we can easily show that by (2)-(4) $r_n \hat{Q}_C(\beta) \xrightarrow{p} 0$. So $\hat{\beta}$ is inconsistent when there is weak identification and nearly singular design. Weak identification manifests itself in inconsistent estimates. Equation (5) provides this in nearly singular design. That is the reason we did not have any other rate in (5). So compared with Stock and Wright (2000), information decays to zero much faster because of nearly singular design

The motivation for having a nearly singular design is the case of many instruments. It will be easy to have near linear dependencies among many of them. This point is well made in Knight (2007) in the case of many similar regressors. Now we need to incorporate many weak instruments into nearly singular design. In that case we try to use the framework of Newey and Windmeijer (2007). They assume, given $G = E(\frac{\partial g_i(\beta_0)}{\partial \beta'})$ and $\Omega(\beta_0) = \Omega$,

$$n S_n^{-1} G' \Omega^{-1} G S_n^{-1'} \to \bar{H},$$

$\bar{H}$ is nonsingular matrix, $S_n = \tilde{S}_n diag(\mu_{1n}, \cdots, \mu_{pn})$ such that $\tilde{S}_n$ is bounded, the smallest eigenvalue of $\tilde{S}_n \tilde{S}_n'$ is bounded away from zero, S_n is a $p \times p$ matrix, for each j, either

$\mu_{jn} = \sqrt{n}$, or $\mu_{jn}/\sqrt{n} \to 0$, $\mu_n = \min_{1 \le j \le p} \mu_{jn} \to \infty$, $m/\mu_n^2 \to \iota$, $0 \le \iota \le 1$, and $\mu_n^2/n \to 0$. Where μ_n^2 controls the number of instruments, and "n" controls the "weakness". We can rewrite the many weak instruments assumption in Newey and Windmeijer (2005)

$$\frac{m}{\mu_n^2} \frac{(n^{1/2}G')\Omega^{-1}(n^{1/2}G)}{m} \to \bar{H}.$$

In our case in (5) the partial derivative can also be handled in the same way so we assume by setting $D_2(\beta_0) = D_2$

$$n a_n S_n^{-1} G' D_2 G S_n^{-1'} \to \tilde{H}. \tag{6}$$

Now we analyze some estimators. It is well known that two-step GMM is not consistent when there are many weak instruments. However, Generalized Empirical Likelihood Estimators (GEL) are consistent as well as jackknife GMM as shown in Newey and Windmeijer (2005). Caner (2006) shows that with only nearly singular design, GEL are consistent as well as two-step GMM. Nearly singular design only affects the rate of convergence of estimators. Their rates are slower than root n. Here we show an interesting result that GEL estimators with many weak instruments facing nearly singular design will not be consistent. So even though many weak moments (instruments) do not reverse consistency and nearly singular design does not by itself affect consistency, the combination of these problems cause inconsistent GEL estimators. However, we then show in detail that jackknife GMM (which is not GEL type of estimator) estimators are consistent with many weak instruments and nearly singular design.

The main issue with GEL is the expectation of the limit objective function is not minimized at β_0. There is a bias term. Unlike many weak moments in GEL of Newey and Windmeijer (2007), nearly singular design introduces this bias which depends on β.

We briefly show the nature of the problem in GEL. A detailed proof is provided in the Appendix. An example illustrating CUE will be given. For consistency we need the limit of the objective function to be minimized at the true values of the parameter. The CUE objective function is

$$\hat{Q}_C(\beta) = \frac{1}{2} \left(\frac{\sum_{i=1}^n g_i(\beta)}{n} \right)' [\hat{\Omega}(\beta)]^{-1} \left(\frac{\sum_{i=1}^n g_i(\beta)}{n} \right).$$

By taking into account the nearly singular design (equations (2-5)), and normalizing for many weak moments, replacing $\hat{\Omega}(\beta)^{-1}$ with $D_2(\beta)$ and denoting this function by $\tilde{Q}_C(\beta)$

$$\begin{aligned} \frac{na_n}{\mu_n^2}\tilde{Q}_C(\beta) &= \frac{na_n}{2\mu_n^2}\left(\frac{1}{n}\sum_{i=1}^n g_i(\beta)\right)'(D_2(\beta))\left(\frac{1}{n}\sum_{i=1}^n g_i(\beta)\right) \\ &= \frac{na_n}{2\mu_n^2}\hat{g}(\beta)'[D_2(\beta)]\hat{g}(\beta). \end{aligned}$$

Note that $\hat{g}(\beta) = \frac{1}{n}\sum_{i=1}^n g_i(\beta)$.

Taking expected value as in Newey and Windmeijer (2007)

$$\begin{aligned} \frac{na_n}{2\mu_n^2}E\left[\hat{g}(\beta)'D_2(\beta)\hat{g}(\beta)\right] &= \frac{na_n}{\mu_n^2}\{\frac{1}{2n^2}E[\sum_{i\neq j} g_i(\beta)'D_2(\beta)g_j(\beta)] + \frac{1}{2n^2}E[\sum_{i=1}^n g_i(\beta)'D_2(\beta)g_i(\beta)]\} \\ &= \frac{na_n}{\mu_n^2}\frac{1-n^{-1}}{2}\bar{g}(\beta)'D_2(\beta)\bar{g}(\beta) + \frac{a_n}{2\mu_n^2}E[g_i(\beta)'D_2(\beta)g_i(\beta)] \\ &= (\frac{na_n}{\mu_n^2})(\frac{1-n^{-1}}{2})\bar{g}(\beta)'D_2(\beta)\bar{g}(\beta) + a_n tr(D_2(\beta)\Omega(\beta))/2\mu_n^2. \end{aligned} \quad (7)$$

Note that the first term is minimized at β_0, but there is an addition to that as the second term. This term depends on β, and bounded given Assumption 4i. Clearly total of these two terms are not minimized at β_0. We prove in the technical appendix that $D_2(\beta)\Omega(\beta) \neq 0$, and $D_2(\beta)\Omega(\beta) \neq I$.

Now we compare our finding above in (7) with the result in p.9 of Newey and Windmeijer (2007). Without nearly singular design, Newey and Windmeijer (2007) have the following expression

$$\frac{n}{\mu_n^2}E\left[\left(\sum_{i=1}^n g_i(\beta)/n\right)\Omega(\beta)^{-1}\left(\sum_{i=1}^n g_i(\beta)/n\right)\right]/2 = (\frac{n}{\mu_n^2})(\frac{1-n^{-1}}{2})\bar{g}(\beta)'\Omega(\beta)^{-1}\bar{g}(\beta) + tr(\Omega(\beta)^{-1}\Omega(\beta))/2\mu_n^2. \quad (8)$$

It is clear that with nearly singular design in (7) $D_2(\beta)\Omega(\beta) \neq 0$ or it is also not equal to identity matrix as in (8), where $\Omega(\beta)^{-1}\Omega(\beta) = I_m$. So $\beta = \beta_0$ does not minimize right hand side of (7). But in regular CUE with many instruments (i.e (8) the second term on the right hand side is equivalent to

$$tr(\Omega(\beta)^{-1}\Omega(\beta))/2\mu_n^2 = m/2\mu_n^2,$$

which is bounded by Assumption in Newey and Windmeijer (2007), and the right hand side of (8) is minimized at $\beta = \beta_0$. The main difference between the two cases is the "noise" term. Because of the nearly singular design, the limit weight is $D_2(\beta)$ not $\Omega(\beta)^{-1}$, hence the source of the inconsistency. Formal proof for inconsistency of CUE estimator can be seen in Lemma A.6, Lemma A.7.

However, jackknife GMM and jackknife CUE estimators do not contain the "noise" term in its expectation. Hence we have consistent estimates. Specifically, we provide jackknife CUE,

$$\sum_{i\neq j}^{n} g_i(\beta)'[\hat{\Omega}(\beta)]^{-1}g_j(\beta)/2n^2.$$

Taking the expectations and appropriately normalizing, replacing $\hat{\Omega}(\beta)$ with $D_2(\beta)$

$$\frac{na_n}{\mu_n^2}E[\frac{1}{2n^2}\sum_{i\neq j}^{n} g_i(\beta)'D_2(\beta)g_j(\beta)] = \frac{na_n}{\mu_n^2}(1-n^{-1})\bar{g}(\beta)'D_2(\beta)\bar{g}(\beta)/2.$$

This is minimized at $\beta = \beta_0$. The details are in the subsequent sections. Note that a jackknife GMM solves the same problem as well. This is defined as

$$\sum_{i\neq j} g_i(\beta)'\hat{\Omega}(\bar{\beta}_F)^{-1}g_j(\beta)/n^2.$$

where $\hat{\Omega}(\bar{\beta}_F)$ is an estimate of the variance covariance matrix . To get this variance estimator, first-step jackknife GMM estimator $\bar{\beta}_F$ is used. Following the same steps for jackknife CUE, this will not have the bias term either.

2 Asymptotics

This section considers consistency, large sample limit and inference with jackknife estimators with near singular design. Defining the jackknife CUE estimator as

$$\hat{\beta} = \arg\min_{\beta\in B}\hat{Q}(\beta),$$

where

$$\hat{Q}(\beta) = \sum_{i\neq j} g_i(\beta)'\hat{\Omega}(\beta)^{-1}g_j(\beta)/2n^2.$$

Jackknife GMM estimator is defined in the same way but $\hat{\Omega}(\beta)^{-1}$ is replaced by $\hat{\Omega}(\hat{\beta}_F)^{-1}$, which is a weight matrix based on first step GMM estimates, $\hat{\beta}_F$. The consistency applies to jackknife GMM estimator, too, but the proof is given for jackknife CUE in the appendix.

Theorem 1. *Under Assumptions 1-5, and* $\hat{\Omega}(\beta)^{-1} \xrightarrow{p} \infty$, $adj(\Omega(\beta)) \neq 0$,

$$S_n'(\hat{\beta}-\beta_0)/\mu_n \xrightarrow{p} 0.$$

The result shows that jackknife CUE is consistent, this is also true for jackknife GMM. As mentioned above GEL estimators are not consistent due to bias term in the near singular design.

We now introduce some notation that is helpful in understanding the limit. Note that $U_i^j = G_i e_j - E[G_i e_j] - B^{j\prime} g_i$, $U_i = [U_i^1, \cdots, U_i^p]$, e_j is the j th unit vector, $G_i = \partial g_i(0)/\partial \delta$, $G = EG_i$. Then we set $D_2 = D_2(\beta_0)$. The following is the limit for jackknife CUE estimator with nearly-singular design and many weak moments. The rate of convergence is the same as in GEL estimators with many weak moments case in Newey and Windmeijer (2007).

Theorem 2. *Under Assumptions 1-7, $\hat{\Omega}(\beta)^{-1} \xrightarrow{p} \infty$, $adj(\Omega(\beta)) \neq 0$ with*

$$\frac{a_n^2}{n^2} \sum_{i=2}^{n} \sum_{j=1}^{i-1} [S_n^{-1} U_i' D_2 g_j g_j' D_2 U_i S_n^{-1} + S_n^{-1} U_j' D_2 g_i g_i' D_2 U_j S_n^{-1}] \xrightarrow{p} \Lambda,$$

Then

$$S_n'(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V),$$

where

$$V = \tilde{H}^{-1} H \tilde{H}^{-1} + \tilde{H}^{-1} \Lambda \tilde{H}^{-1},$$

on the nullspace of $adj(\Omega(\beta)), \Omega$.

Remark. The limit term is similar to Newey and Windmeijer (2007) limit, although here the first term is not $\tilde{H}^{-1}$ alone. It is not necessarily true that $H = \tilde{H}$ in this analysis. Also, the assumption about double summed term is the sample counterpart of the assumption used in Newey and Windmeijer (2007). Since we have nearly singular design we have sample version as well as rate a_n^2/n^2 because of the nearly singular design. This rate is compatible with the rate in Assumption 1.

APPENDIX

Proof of Lemma 1. Note that

$$\begin{aligned}
\hat{\Omega}(\beta)^{-1} - b_n adj(\Omega(\beta)) &= \frac{1}{det(\hat{\Omega}(\beta))}[adj(\hat{\Omega}(\beta)) - adj(\Omega(\beta))] + adj(\Omega(\beta))[\frac{1}{det(\hat{\Omega}(\beta))} - b_n] \\
&= \frac{1}{det(\hat{\Omega}(\beta))}[adj(\hat{\Omega}(\beta)) - adj(\Omega(\beta))] + o_p(1) \\
&= \frac{a_n[adj(\hat{\Omega}(\beta)) - adj(\Omega(\beta))]}{a_n[det(\hat{\Omega}(\beta)) - det(\Omega(\beta))]} + o_p(1),
\end{aligned}$$

where we use $det(\Omega(\beta)) = 0$. Define any matrix A of $m \times m$ dimension, where a_{ij} represents the i th row, jth column cell in that matrix. Then

$$det(A) = \sum_{\sigma \in S_m} sgn(\sigma) \Pi_{i=1}^m a_{i,\sigma(i)},$$

where S_m represents all the permutations σ of the numbers $1, 2, \cdots, m$, and $\sigma(i)$ represents the specific permutation, sgn σ denotes the sign of the permutation of σ. This is 1 if σ is an even permutation, and it is -1, if this is an odd one.

Now, use Assumption 1 to have

$$a_n[det(\hat{\Omega}(\beta)) - det(\Omega(\beta))] \xrightarrow{p} \kappa(\beta),$$

$$a_n[adj(\hat{\Omega}(\beta) - adj(\Omega(\beta)))] \xrightarrow{p} \Delta(\beta).$$

We show the definitions of $\kappa(.), \Delta(.)$ at the end of the proof for interested readers. To prove Lemma 1 , we first need to show $\kappa(\beta) \neq 0$, for all $\beta \in B$. Note that, for any square matrix A, $Aadj(A) = adj(A)A = detAI$. Since $det(\Omega(\beta)) = 0$, in our case we have

$$a_n\hat{\Omega}(\beta)adj(\hat{\Omega}(\beta)) = a_n[det(\hat{\Omega}(\beta) - det(\Omega(\beta))]I,$$

Then add and subtract to have

$$\hat{\Omega}(\beta)a_n[adj(\hat{\Omega}(\beta)) - adj(\Omega(\beta))] + a_n[\hat{\Omega}(\beta) - \Omega(\beta)]adj(\Omega(\beta)) = a_n[det(\hat{\Omega}(\beta)) - det(\Omega(\beta))]I. \tag{9}$$

Then by Assumption 1, the left hand-side of the equation above converges in probability to (uniformly over β)

$$\hat{\Omega}(\beta)a_n[adj(\hat{\Omega}(\beta)) - adj(\Omega(\beta))] + a_n[\hat{\Omega}(\beta) - \Omega(\beta)]adj(\Omega(\beta)) \xrightarrow{p} \Omega(\beta)\Delta(\beta) + D_1(\beta)adj(\Omega(\beta)).$$

The right-hand side of (9) converges in probability to (uniformly over β)

$$a_n[det(\hat{\Omega}(\beta)) - det(\Omega(\beta))]I \xrightarrow{p} \kappa(\beta)I.$$

By (9) and the result above we have

$$\Omega(\beta)\Delta(\beta) + D_1(\beta)adj(\Omega(\beta)) = \kappa(\beta)I. \tag{10}$$

Since $\Omega(\beta)adj(\Omega(\beta)) = det(\Omega(\beta)) = 0$, and $adj(\Omega(\beta))$ is symmetric, every column and every row of $adj(\Omega(\beta))$ is in the nullspace of $\Omega(\beta)$ and vice versa. Take an $m \times 1$ vector, which has 1 in first cell and zeroes in the other cells, denote this by e_1. Then premultiply (10) by $e_1'adj(\Omega(\beta))$ and postmultiply by e_1 to have

$$e_1'adj(\Omega(\beta))\Omega(\beta)\Delta(\beta)e_1 + e_1'adj(\Omega(\beta))D_1(\beta)adj(\Omega(\beta))e_1 = \kappa(\beta)e_1'adj(\Omega(\beta))e_1. \tag{11}$$

In (11) note that $adj(\Omega(\beta))\Omega(\beta) = 0$ so

$$e_1' adj(\Omega(\beta)) D_1(\beta) adj(\Omega(\beta)) e_1 = \kappa(\beta) e_1' adj(\Omega(\beta)) e_1. \quad (12)$$

At (12), $adj(\Omega(\beta))e_1$ equals the first column of $adj(\Omega(\beta))$ which is in the nullspace of $\Omega(\beta)$. This means that the left hand side of (12) is positive. Then we have $e_1' adj(\Omega(\beta))e_1$ is nonzero since $adj(\Omega(\beta)) \neq 0$ and so $\kappa(\beta) \neq 0$, for all $\beta \in B$.

Next part of the proof shows $D_2(\beta) = \Delta(\beta)/\kappa(\beta)$ is positive definite on the nullspace of $adj(\Omega(\beta))$. Toward this goal, we first note that for all $\nu \in R^m$ be such that for $\Omega(\beta) \neq 0$ premultiply each side of (10) with ν' and postmultiply by $\Omega(\beta)\nu$ to get

$$\nu' \Omega(\beta) \Delta(\beta) \Omega(\beta) \nu = \kappa(\beta) \nu' \Omega(\beta) \nu,$$

by $adj(\Omega(\beta))\Omega(\beta) = det\Omega(\beta) = 0$. Then $\kappa(\beta) \neq 0$, so divide each side by $\kappa(\beta)$ to have

$$\nu' \Omega(\beta) D_2(\beta) \Omega(\beta) \nu = \nu' \Omega(\beta) \nu. \quad (13)$$

Then since $\Omega(\beta)$ is positive semidefinite and symmetric, there exists a matrix $\Omega(\beta)^{1/2}$ such that $\Omega(\beta)^{1/2}\Omega(\beta)^{1/2} = \Omega(\beta)$. Moreover $\Omega(\beta)\nu \neq 0$ implies $\Omega(\beta)^{1/2}\nu \neq 0$ which implies $\nu'\Omega(\beta)\nu \neq 0$. This shows that $D_2(\beta)$ is positive definite on the set of $u \neq 0$, such that $u = \Omega(\beta)\nu$ for some $\nu \in R^m$. Let S denote this set. If we show that S coincides with the nullspace of $adj(\Omega(\beta))$, the proof will be complete. Since $adj(\Omega(\beta))\Omega(\beta)\nu = 0$, we know that image of $\Omega(\beta)$ (column space, Im (.)) is contained in the nullspace of $adj(\Omega(\beta))$. But by p.60 of Abadir and Magnus (2005)

$$\begin{aligned} dim(S) &= dim(Im(\Omega(\beta))) \\ &= rank(\Omega(\beta)) \\ &= m - 1 \\ &= dim(nullspaceofadj(\Omega(\beta))). \end{aligned}$$

For the last two equalities we benefit from Sinkhorn (1993). dim (M) represents the dimension of matrix M. **Q.E.D**.

Now we provide the definition of $\kappa(\beta), \Delta(\beta)$. Denoting the cells of $\hat{\Omega}(\beta)$ as $\hat{\omega}_{ij}$, and $\Omega(\beta)$ as ω_{ij}. Also using Assumption 1 we have $a_n[\hat{\omega}_{ij} - \omega_{ij}] \xrightarrow{p} d_{i,j}$ where $d_{i,j}$ is the ith row, jth column element of $D_1(\beta)$ matrix. Then

$$\kappa(\beta) = \sum_{\sigma \in S_m} sgn(\sigma) \sum_{\{i=1,\cdots,m\}} d_{i,\sigma(i)} \Pi_{j \in \{1,\cdots,m\}-\{i\}} \omega_{j,\sigma(j)}.$$

$$\Delta_{i,j}(\beta) = (-1)^{i+j} \sum_{\sigma \in S_{m-1}} sgn(\sigma) \sum_{l \in \{1,\cdots,m\}-\{i\}} d_{l,c^j(\sigma(r^i(l)))} \Pi_{k \in \{1,\cdots,m\}-\{l,k\}} \omega_{k,c^j(\sigma(r^i(k)))},$$

where $r^i(l) = l-1$ if $i=1$, or $i > 1 \wedge l > i, r^i(l) = l$, if $i > 1 \wedge l < i$, and $c^j(k) = k+1$ if $j = 1$ or $j > 1 \wedge k > j, c^j(k) = k$ if $j > 1 \wedge k > j$.

Let Y_i, Z_i be iid $m \times 1$ vector, that can depend on n, but this is suppressed. Define

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i,\ \bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i,$$

$$\Sigma_{yy} = EY_iY_i',\ \Sigma_{zz} = EZ_iZ_i',$$

Given (4), we have the following as well

$$EY_i = \mu_y,\ EZ_i = \mu_z.$$

The following is crucial to the consistency proof. This is similar to the analysis of Newey and Windmeijer (2007). We analyze a jackknifed term with near singular design in many weak moments. This result is different from Newey and Windmeijer (2007). This does not involve any "'bias"' term as in Newey and Windmeijer (2007), and involves near singular design rate "a_n". The result in Newey and Windmeijer (2007) is for GEL and this is for jackknife estimators with near singular design. Our result is different and only contains the "signal" term.

Lemma A.1. *If (Y_i, Z_i) are iid, and $\lambda_{max}(AA') \leq C, \lambda_{max}(A'A) \leq C$, $\lambda_{max}(\Sigma_{yy}) \leq C, \lambda_{max}(\Sigma_{zz}) \leq C$, $a_n^2 m/d_n^2 \to 0$, $(n-1)a_n^2\mu_y'\mu_y/d_n^2 \to 0, (n-1)a_n^2\mu_z'\mu_z/d_n^2 \to 0$, then*

$$a_n\left[\frac{\sum_{i\neq j} Y_i'AZ_j}{n}\right]/d_n = ((n-1)a_n)\frac{\mu_y' A\mu_z}{d_n} + o_p(1).$$

Remark. Note that we have also fewer conditions compared with Newey and Windmeijer (2007). This results from the jackknife form of our term.

Proof of Lemma A.1. First rewrite the function as

$$a_n\left(\frac{\sum_{i\neq j} Y_i'AZ_j}{n}\right)/d_n = \frac{na_n}{d_n}(\bar{Y}'A\bar{Z}) - a_n\left(\frac{\sum_{i=1}^{n} Y_i'AZ_i}{nd_n}\right), \tag{14}$$

where $\bar{Y} = \sum_{i=1}^{n}\frac{1}{n}Y_i, \bar{Z} = \frac{1}{n}Z_i$ as in their definitions. Then without losing any generality we can ignore the matrix A as in the proof of Lemma A.1 in Newey and Windmeijer (2007). Suppose $\mu_y = \mu_z = 0$, and setting $W_n = \frac{a_n}{d_n}(\frac{\sum_{i\neq j} Y_i'Z_j}{n})$,

$$EW_n = \frac{a_n}{d_n}(\frac{\sum_{i\neq j} Y_i'Z_j}{n}) = 0,$$

by iid property of Z_i, Y_i. Then

$$EW_n^2 = \frac{n-1}{n}\frac{a_n^2}{d_n^2}(E[Y_i'Z_jY_j'Z_i] + E[Y_i'Z_jZ_j'Y_i]). \tag{15}$$

Then in (15), by $\lambda_{max}(\Sigma_{zz}) \leq C$

$$\begin{aligned} E[Y_i'Z_jZ_j'Y_i] &= E[Y_i'\Sigma_{zz}Y_i] \\ &\leq CE[Y_i'Y_i] = Ctr(\Sigma_{yy}) \\ &\leq Cm, \end{aligned} \tag{16}$$

$$\begin{aligned} E[Y_i'Z_jY_j'Z_i] &\leq C(E[Y_i'Z_jZ_j'Y_i] + E[Y_j'Z_iZ_i'Y_j]) \\ &\leq Cm, \end{aligned} \tag{17}$$

The last two equations are taken from the proof of Lemma A.1 on p.36 of Newey and Windmeijer (2007). Using $a_n^2 m/d_n^2 \to 0$ we have $EW_n^2 \to 0$, so by Markov's inequality

$$W_n = o_p(1). \tag{18}$$

Now we analyze the case of when $\mu_y, \mu_z \neq 0$. Note that $EY_i = \mu_y, EZ_i = \mu_z$. So use this in (14)

$$\begin{aligned} W_n &= \frac{na_n}{d_n}(\bar{Y} - \mu_y)'(\bar{Z} - \mu_z) \\ &+ na_n\mu_y'(\bar{Z} - \mu_z)/d_n + na_n(\bar{Y} - \mu_y)'\mu_z/d_n \\ &+ na_n\mu_y'\mu_z/d_n - na_n\left(\frac{\sum_{i=1}^n (Y_i - \mu_y)'(Z_i - \mu_z)}{n^2}\right)/d_n \\ &- na_n\left(\frac{\sum_{i=1}^n (Y_i - \mu_y)'\mu_z}{n^2}\right)/d_n - na_n\left(\frac{\sum_{i=1}^n \mu_y'(Z_i - \mu_z)}{n^2}\right)/d_n \\ &- na_n\left(\frac{\sum_{i=1}^n \mu_y'\mu_z}{n^2}\right)/d_n. \end{aligned} \tag{19}$$

Since

$$na_n\left(\sum_{i=1}^n \frac{(Y_i - \mu_y)'\mu_z}{n^2}\right)/d_n = a_n(\bar{Y} - \mu_y)'\mu_z/d_n,$$

We can simplify (19) as

$$\begin{aligned} W_n &= [na_n(\bar{Y} - \mu_y)'(\bar{Z} - \mu_z)/d_n - na_n\frac{\sum_{i=1}^n (Y_i - \mu_y)'(Z_i - \mu_z)}{n^2}/d_n] \\ &+ [((n-1)a_n)\,\mu_y'(\bar{Z} - \mu_z)/d_n] + [((n-1)a_n)\,(\bar{Y} - \mu_y)'\mu_z/d_n] \\ &+ [((n-1)a_n)\,\mu_y'\mu_z/d_n]. \end{aligned} \tag{20}$$

The first term with the square brackets on the right-hand side of (20) converges in probability to zero by the analysis in (18). Then for the second term

$$
\begin{aligned}
E\left[(n-1)a_n\mu_y'(\bar{Z}-\mu_z)/d_n\right]^2 &= \frac{(n-1)^2a_n^2}{d_n^2}\mu_y'\{E[(\bar{Z}-\mu_z)(\bar{Z}-\mu_z)']\}\mu_y \\
&\leq \frac{(n-1)^2a_n^2}{nd_n^2}\mu_y'\Sigma_{zz}\mu_y \\
&\leq C\frac{(n-1)a_n^2}{d_n^2}\mu_y'\mu_y \to 0, \qquad (21)
\end{aligned}
$$

by $\lambda_{max}(\Sigma_{zz}) \leq C$, and $\bar{Z} = \sum_{i=1}^n Z_i/n$, and the assumption of $\frac{(n-1)a_n^2}{d_n^2}\mu_y'\mu_y \to 0$. The same analysis applies to the third term on the right hand side of (20), which provides

$$
E\left[(n-1)a_n(\bar{Y}-\mu_y)'\mu_z/d_n\right]^2 \to 0. \qquad (22)
$$

So the second and third terms on the right hand side of (21)(22) converge in probability to zero. So we have

$$
W_n = ((n-1)a_n)\,\frac{\mu_y'\mu_z}{d_n} + o_p(1).
$$

Q.E.D

The next lemma is extension of Lemma A.2 in Newey and Windmeijer (2007) from GEL estimators to jackknife CUE estimator with nearly singular design. This is the uniform law of large numbers result that is needed for the consistency. Set $\bar{g}(\beta) = Eg_i(\beta)$. Note that before Lemma A.2, we define the limit of the sample jackknife CUE objective function

$$
Q(\beta) = \frac{n-1}{n}\frac{\bar{g}(\beta)'D_2(\beta)\bar{g}(\beta)}{2}.
$$

Lemma A.2. *Under Assumptions 1-4, with* $\hat{\Omega}(\beta)^{-1} \xrightarrow{p} \infty$, *and* $adj(\Omega(\beta)) \neq 0$

$$
\sup_{\beta\in B,} na_n\frac{1}{\mu_n^2}|\hat{Q}(\beta) - Q(\beta)| \xrightarrow{p} 0.
$$

Proof of Lemma A.2. The sample objective function, appropriately normalized, is

$$
\hat{Q}(\beta) = \frac{1}{n^2}\frac{\sum_{i\neq j} g_i(\beta)'\hat{\Omega}(\beta)^{-1}g_j(\beta)}{2}. \qquad (23)
$$

This can be rewritten as

$$
\hat{Q}(\beta) = \frac{\hat{g}(\beta)'\hat{\Omega}(\beta)^{-1}\hat{g}(\beta)}{2} - \frac{1}{n^2}\frac{\sum_{i=1}^n g_i(\beta)'\hat{\Omega}(\beta)^{-1}g_i(\beta)}{2}.
$$

We analyze the behavior of

$$na_n \frac{1}{\mu_n^2}\hat{Q}(\beta) = na_n \frac{\hat{g}(\beta)'\hat{\Omega}(\beta)^{-1}\hat{g}(\beta)}{2\mu_n^2} - \frac{a_n}{n}\frac{\sum_{i=1}^n g_i(\beta)'\hat{\Omega}(\beta)^{-1}g_i(\beta)}{2\mu_n^2}. \tag{24}$$

Denote

$$\tilde{Q}(\beta) = \frac{\hat{g}(\beta)'D_2(\beta)\hat{g}(\beta)}{2} - \frac{1}{n^2}\frac{\sum_{i=1}^n g_i(\beta)'D_2(\beta)g_i(\beta)}{2}.$$

or

$$\tilde{Q}(\beta) = \frac{1}{n^2}\frac{\sum_{i\neq j} g_i(\beta)'D_2(\beta)g_j(\beta)}{2}. \tag{25}$$

We first show that

$$\sup_{\beta} |na_n \frac{1}{\mu_n^2}(\hat{Q}(\beta) - \tilde{Q}(\beta))| \xrightarrow{p} 0. \tag{26}$$

As in Newey and Windmeijer (2007), we reparametrize $\delta = S_n'(\beta - \beta_0)/\mu_n$. Then $\hat{Q}(\delta)$ will denote $\hat{Q}(\beta_0 + \mu_n S_n^{-1\prime}\delta)$. Let $\tilde{Q}(\delta)$ denote $\tilde{Q}(\beta_0 + \mu_n S_n^{-1\prime}\delta)$. Then on the nullspace of $adj(\Omega(\delta))$

$$\begin{aligned} na_n \frac{1}{\mu_n^2}|\hat{g}(\delta)'[\hat{\Omega}(\delta)^{-1} - D_2(\delta)]\hat{g}(\delta)| &= \frac{na_n}{\mu_n^2}|\hat{g}(\delta)[(\hat{\Omega}(\delta)^{-1} - b_n adj(\Omega(\delta))) + b_n adj(\Omega(\delta)) - D_2(\delta)]\hat{g}(\delta)| \\ &= \frac{na_n}{\mu_n^2}|\hat{g}(\delta)[(\hat{\Omega}(\delta)^{-1} - b_n adj(\Omega(\delta))) - D_2(\delta)]\hat{g}(\delta)| \\ &\leq na_n \frac{1}{\mu_n^2}\|\hat{g}(\delta)\|^2|(\hat{\Omega}(\delta)^{-1} - b_n adj(\Omega(\delta))) - D_2(\delta)| \\ &\xrightarrow{p} 0, \end{aligned} \tag{27}$$

uniformly over δ by Assumption 4iii (i.e.$\sup_{\{\delta\leq C\}} \|\hat{g}(\delta)\| = O_p(\mu_n \frac{1}{\sqrt{a_n}\sqrt{n}})$) and by Lemma 1. Then in the same way as in (27) on the nullspace of $adj(\Omega(\delta))$ with Lemma 1

$$\begin{aligned} \frac{a_n}{n}\left(\frac{\sum_{i=1}^n g_i(\delta)'\hat{\Omega}(\delta)^{-1}g_i(\delta)}{\mu_n^2}\right) \quad - \quad \frac{a_n}{n}\left(\frac{\sum_{i=1}^n g_i(\delta)'D_2(\delta)g_i(\delta)}{\mu_n^2}\right) \\ \xrightarrow{p} 0. \end{aligned} \tag{28}$$

So we prove (26) by (27)(28). Next we need to prove stochastic equicontinuity of $na_n\mu_n^{-2}\tilde{Q}(\delta)$.

So, with Assumption 4ii, clearly

$$na_n \frac{1}{\mu_n^2}|\tilde{Q}(\delta_2) - \tilde{Q}(\delta_1)\| \leq \hat{M}\|\delta_2 - \delta_1\|,$$

on $\|\delta_2\| \leq C, \|\delta_1\| \leq C$ with $\hat{M} = O_p(1)$. Hence, $\mu_n^{-2}na_n\tilde{Q}(\delta)$ is stochastically equicontinuous. Next, we consider equicontinuity of $na_n\frac{1}{\mu_n^2}Q(\delta)$ on $\|\delta_2\| \leq C, \|\delta_1\| \leq C$, where

$$Q(\delta) = \frac{n-1}{n}\frac{\bar{g}(\delta)'D_2(\delta)\bar{g}(\delta)}{2}, \tag{29}$$

where $\bar{g}(\delta) = Eg_i(\delta)$. This property is obtained using Assumption 4ii, iv exactly like the stochastic equicontinuity of $\tilde{Q}(\delta)$, hence skipped. We now apply Lemma A.1 to show point-wise convergence results in δ.

$$\mu_n^{-2} n a_n \tilde{Q}(\delta) = \mu_n^{-2} n a_n Q(\delta) + o_p(1).$$

Set $Y_i = g_i(\delta) = Z_i$, $A = D_2(\delta)$, $d_n = \mu_n^2$. By Assumption 3, $\lambda_{max}(AA') \leq C, \lambda_{max}(A'A) \leq C$, and also $\lambda_{\max}(\Sigma_{yy}) \leq C$, by Assumption 4i we have $a_n^2 m/\mu_n^4 \to 0$, and

$$\begin{aligned} \frac{(n-1)a_n^2}{\mu_n^4}\bar{g}(\delta)'\bar{g}(\delta) &\leq \frac{C}{\mu_n^4}(n-1)a_n^2 \bar{g}(\delta)' D_2(\delta)\bar{g}(\delta) \\ &\leq \frac{Ca_n}{\mu_n^2}\left[na_n \frac{Q(\delta)}{\mu_n^2}\right] \\ &\to 0, \end{aligned}$$

where $\lambda_{max}(D_2(\delta)) \leq C$ by Assumption 3, $a_n/\mu_n^2 \to 0$ by Assumption 4i, and $na_n\mu_n^{-2}Q(\delta)$ is equicontinuous is used. Since Assumptions of Lemma A.1 are satisfied

$$\begin{aligned} \mu_n^{-2} n a_n \tilde{Q}(\delta) &= \mu_n^{-2} n a_n Q(\delta) + o_p(1) \\ &= (n-1)a_n \frac{\bar{g}(\delta)' D_2(\delta)\bar{g}(\delta)}{2\mu_n^2} + o_p(1). \end{aligned}$$

Clearly combining these results,

$$\sup_{\|\delta\| \leq C} \mu_n^{-2} n a_n |\hat{Q}(\delta) - Q(\delta)| \xrightarrow{p} 0.$$

Q.E.D

Proof of Theorem 1.

The proof is similar to the proof of Theorem 1 in Newey and Windmeijer (2007) given Lemma A.1 and A.2 here. The main differences are the issue of near singular design, and $Q(\delta)$ is defined differently with Lemma A.1 result in jackknife CUE form rather than CUE in Newey and Windmeijer (2007).

First note that,

$$\|\hat{\delta}\| = O_p(1), \tag{30}$$

by Assumptions 4iii, and 5i. Since $\delta_0 = 0$, by $\hat{\delta}$ definition

$$\mu_n^{-2} n a_n \hat{Q}(\hat{\delta}) \leq \mu_n^{-2} n a_n \hat{Q}(0). \tag{31}$$

By (30),

$$P(\|\hat{\delta}\| \leq C) \geq 1 - \epsilon/3, \tag{32}$$

for $\epsilon > 0$. Then for $\gamma > 0$, via Lemma A.2

$$P(\sup_{\|\delta\| \leq C} \mu_n^{-2} n a_n |\hat{Q}(\delta) - Q(\delta)| < \gamma/3) \geq 1 - \epsilon/3. \tag{33}$$

By (30)-(33), the joint probability of these events are greater than or equal to $1 - \epsilon$. Then on the joint event set

$$\begin{aligned} \mu_n^{-2} n a_n Q(\hat{\delta}) &\leq \mu_n^{-2} n a_n \hat{Q}(\hat{\delta}) + \gamma/3 \\ &\leq \mu_n^{-2} n a_n \hat{Q}(0) + 2\gamma/3 \\ &\leq \mu_n^{-2} n a_n Q(0) + \gamma, \end{aligned} \tag{34}$$

which implies by (29)

$$\frac{(n-1)a_n}{2\mu_n^2} \bar{g}(\hat{\delta})' D_2(\hat{\delta}) \bar{g}(\hat{\delta}) \leq \gamma. \tag{35}$$

Since ϵ, γ can be small positive constants

$$\mu_n^{-2} \frac{(n-1)a_n}{2} \bar{g}(\hat{\delta})' D_2(\hat{\delta}) \bar{g}(\hat{\delta}) \xrightarrow{p} 0. \tag{36}$$

By Assumptions 3 and 5ii

$$\begin{aligned} (n-1)a_n \frac{1}{2\mu_n^2} \bar{g}(\hat{\delta})' D_2(\hat{\delta}) \bar{g}(\hat{\delta}) &\geq C\mu_n^{-2}(n-1)a_n \bar{g}(\hat{\beta})' \bar{g}(\hat{\beta}) \\ &\geq C\|\hat{\delta}\|^2. \end{aligned} \tag{37}$$

So

$$\|\hat{\delta}\|^2 \xrightarrow{p} 0.$$

Q.E.D.

To derive the asymptotic normality result, we need the following notation. Set up as before $\delta = S_n^{-1}(\beta - \beta_0)$. Denote $\tilde{g}_{\delta_k} = \partial \hat{g}(0)/\partial \delta_k = \sum G_i S_n^{-1'} e_k \mu_n$, and $\tilde{g} = \hat{g}(0)$, for $\delta = 0$. Also we have $\tilde{\Omega}^k = \sum_{i=1}^n g_i g_{i\delta_k}'/n, g_{i\delta_k} = \partial g_i(0)/\partial \delta_k = G_i' S_n^{-1} e_k' \mu_n$, $g_i = g_i(0)$, $\bar{g}_{\delta_k} = E[\partial g_i(0)/\partial \delta_k] = G S_n^{-1'} e_k \mu_n$. $\tilde{\Omega}^{-1} = \hat{\Omega}^{-1}(0)$, $\tilde{B}^k = \tilde{\Omega}^{-1}\tilde{\Omega}^k$, $B^k = D_2(0)\Omega^k, \Omega_k = E g_i g_{i\delta_k}', D_2(0)$ is $D_2(\delta)$ that is evaluated at $\delta = 0$. Note that $U_i^j = G_i e_j - E[G_i e_j] - B^{j'} g_i$, $U_i = [U_i^1, \cdots, U_i^p]$, e_j is the j th unit vector, $G_i = \partial g_i(0)/\partial \delta$, $G = E G_i$. Then we set $D_1 = D_1(0), D_2 = D_2(0)$.

Lemma A.3. *Under Assumptions 1, 3 and 6, on the nullspace of* $adj(\Omega(\delta=0))$

$$a_n\sqrt{m}\|\tilde{\Omega}^k-\Omega^k\| \xrightarrow{p} 0,$$
$$a_n\sqrt{m}\|\tilde{B}^k-B^k\| \xrightarrow{p} 0.$$

Proof of Lemma A.3. Note that first result is due to Lemma A.11 of Newey and Windmeijer (2007) given Assumption 6 and the nearly singular design. We proceed through the last result by adding and subtracting $b_n adj(\Omega(0))$ from the second term on the right hand side of the inequality below

$$a_n\sqrt{m}\|\tilde{B}^k-B^k\| \le a_n\sqrt{m}\|(\tilde{\Omega}^{k'}-\Omega^{k'})\tilde{\Omega}^{-1}\| + a_n\|\sqrt{m}\|\Omega^{k'}(D_2-\tilde{\Omega}^{-1})\| \xrightarrow{p} 0,$$

by the first result Lemma 1, and Assumption 3 on the nullspace of $adj(\Omega(\delta=0))$.**Q.E.D.**

Lemma A.4.*Under Assumptions 1-6, with*

$$\frac{a_n^2}{n^2}\sum_{i=2}^{n}\sum_{j=1}^{i-1}[S_n^{-1}U_i'D_2g_jg_j'D_2U_iS_n^{-1}+S_n^{-1}U_j'D_2g_ig_i'D_2U_jS_n^{-1}] \xrightarrow{p} \Lambda.$$

we obtain

$$\frac{na_n}{\mu_n}\frac{\partial\hat{Q}(0)}{\partial\delta} \xrightarrow{d} N(0,H+\Lambda),$$

on the nullspace of $adj(\Omega(\delta=0))$*, and* Ω*.*

Proof of Lemma A.4. The objective function is

$$\sum_{i\neq j} g_i(\beta)'\hat{\Omega}(\beta)^{-1}g_j(\beta)/2n^2 = \frac{\hat{g}(\beta)'\hat{\Omega}(\beta)^{-1}\hat{g}(\beta)}{2} - \left[\frac{1}{n}\sum_{i=1}^{n} g_i(\beta)'\hat{\Omega}(\beta)^{-1}g_i(\beta)\right]/2n. \tag{38}$$

Using (38) to get the derivative of the objective function at $\delta_k=0$

$$\begin{aligned}\frac{\partial\hat{Q}(0)}{\partial\delta_k} &= \tilde{g}_{\delta_k}'\tilde{\Omega}^{-1}\tilde{g}-\tilde{g}'\tilde{B}^k\tilde{\Omega}^{-1}\tilde{g}\\ &- \{[\frac{1}{n}\sum_{i=1}^{n} g_{i\delta_k}'\tilde{\Omega}^{-1}g_i]-[\frac{1}{n}\sum_{i=1}^{n} g_i'\tilde{B}^k\tilde{\Omega}^{-1}g_i]\}/n.\end{aligned}$$

Add and subtract $\bar{g}_{\delta_k}'\tilde{\Omega}^{-1}\tilde{g}$, and $\{\frac{1}{n}\sum_{i=1}^{n}\bar{g}_{\delta_k}'\tilde{\Omega}^{-1}g_i\}/n$ to the right-hand side of the above equation to have

$$\begin{aligned}\frac{\partial\hat{Q}(0)}{\partial\delta_k} &= \bar{g}_{\delta_k}'\tilde{\Omega}^{-1}\tilde{g}-\{\frac{1}{n}\sum_{i=1}^{n}\bar{g}_{\delta_k}'\tilde{\Omega}^{-1}g_i\}/n\\ &+ [\tilde{g}_{\delta_k}'\tilde{\Omega}^{-1}\tilde{g}-\bar{g}_{\delta_k}'\tilde{\Omega}^{-1}\tilde{g}-\tilde{g}'\tilde{B}^k\tilde{\Omega}^{-1}\tilde{g}]\\ &- \{[\frac{1}{n}\sum_{i=1}^{n} g_{i\delta_k}'\tilde{\Omega}^{-1}g_i-\frac{1}{n}\sum_{i=1}^{n}\bar{g}_{\delta_k}'\tilde{\Omega}^{-1}g_i-\frac{1}{n}\sum_{i=1}^{n} g_i'\tilde{B}^k\tilde{\Omega}^{-1}g_i]\}/n.\end{aligned}$$

For $k = 1, \cdots, p$, let

$$\hat{U}^k = \tilde{g}_{\delta_k} - \bar{g}_{\delta_k} - \tilde{B}^{k'} \tilde{g}, \tag{39}$$

$$\hat{U}_i^k = g_{i\delta_k} - \bar{g}_{\delta_k} - \tilde{B}^{k'} g_i,$$

$$\hat{U}^k = \frac{1}{n} \sum_{i=1}^n \hat{U}_i^k.$$

So

$$\frac{\partial \hat{Q}(0)}{\partial \delta_k} = \bar{g}'_{\delta_k} \tilde{\Omega}^{-1} \tilde{g}(1 - 1/n) + \hat{U}^{k'} \tilde{\Omega}^{-1} \tilde{g} - \{\frac{1}{n} \sum_{i=1}^n \hat{U}_i^{k'} \tilde{\Omega}^{-1} g_i\}/n. \tag{40}$$

The following results are on the nullspace of $adj(\Omega(\delta = 0))$. We consider $\bar{g}'_{\delta_k} \tilde{\Omega}^{-1} \tilde{g}$ in (40) above. We show that

$$\frac{na_n}{\mu_n} |\bar{g}'_{\delta_k} \tilde{\Omega}^{-1} \tilde{g} - \bar{g}'_{\delta_k} D_2 \tilde{g}| \xrightarrow{p} 0, \tag{41}$$

where $D_2 = D_2(0)$ when $\delta = 0$. First note that by Assumption 6, $\sqrt{na_n} \| EG_i(\beta_0) \| S_n^{-1'} \leq C$. Then we have

$$\|\bar{g}_{\delta_k}\| = O_p(\mu_n / \sqrt{a_n n}).$$

Next, by Assumption 4, $\|\tilde{g}\| = O_p(\mu_n / \sqrt{a_n n})$.

$$\frac{na_n}{\mu_n} |\bar{g}'_{\delta_k} \tilde{\Omega}^{-1} \tilde{g} - \bar{g}'_{\delta_k} D_2 \tilde{g}\| = \frac{na_n}{\mu_n} |(\bar{g}'_{\delta_k} \tilde{\Omega}^{-1} \tilde{g} - \bar{g}'_{\delta_k} b_n adj(\Omega(0)) \tilde{g}) + \bar{g}'_{\delta_k} b_n adj(\Omega(0)) \tilde{g} - \bar{g}'_{\delta_k} D_2 \tilde{g}|. \tag{42}$$

Since $\bar{g}'_{\delta_k} b_n adj(\Omega(0)) \tilde{g} = 0$ (nullspace of $adj(\Omega(0))$), (42) can be rewritten as

$$\begin{aligned} \frac{na_n}{\mu_n} |(\bar{g}'_{\delta_k} \tilde{\Omega}^{-1} \tilde{g} - \bar{g}'_{\delta_k} b_n adj(\Omega(0)) \tilde{g}) - \bar{g}'_{\delta_k} D_2 \tilde{g}| &\leq \frac{na_n}{\mu_n} \|\bar{g}_{\delta_k}\| \|\tilde{g}\| \|\tilde{\Omega}^{-1} - b_n adj(\Omega(0)) - D_2\| \\ &= \frac{na_n}{\mu_n} O_p(\frac{\mu_n}{\sqrt{na_n}}) O_p(\frac{\mu_n}{\sqrt{a_n n}}) o_p(\frac{1}{\sqrt{m} a_n}) \\ &\xrightarrow{p} 0, \end{aligned} \tag{43}$$

by Lemma 1. Since (41) is proved, we consider the following term in (40), $\hat{U}^{k'} \tilde{\Omega}^{-1} \tilde{g}$. We now show

$$\frac{na_n}{\mu_n} |\hat{U}^{k'} \tilde{\Omega}^{-1} \tilde{g} - \tilde{U}^{k'} \tilde{\Omega}^{-1} \tilde{g}| \xrightarrow{p} 0, \tag{44}$$

where $\tilde{U}^k = \tilde{g}_{\delta_k} - \bar{g}_{\delta_k} - B^{k'}\tilde{g}$. Note that $\hat{U}^k$ is defined in (39). First by Lemma 1

$$\tilde{\Omega}^{-1} - b_n adj(\Omega(0)) + b_n adj(\Omega(0)) - D_2\| \xrightarrow{p} 0.$$

So

$$\begin{aligned}\frac{na_n}{\mu_n}|\hat{U}^{k'}\tilde{\Omega}^{-1}\tilde{g} - \tilde{U}^{k'}\tilde{\Omega}^{-1}\tilde{g}| &\leq \frac{na_n}{\mu_n}|\tilde{g}'(\tilde{B}^k - B^k)\tilde{\Omega}^{-1}\tilde{g}| \\ &\leq \frac{na_n}{\mu_n}\|\tilde{g}\|^2\|\tilde{B}^k - B^k\|\|\tilde{\Omega}^{-1}\| \\ &\xrightarrow{p} 0,\end{aligned}$$

by $\|\tilde{g}\| = O_p(\mu_n/\sqrt{a_n n})$, by Lemma A.3, and Assumption 3, on the nullspace of $adj(\Omega(\delta = 0))$.

Then we show

$$\frac{na_n}{\mu_n}|\tilde{U}^{k'}\tilde{\Omega}^{-1}\tilde{g} - \tilde{U}^{k'}D_2\tilde{g}| \xrightarrow{p} 0, \tag{45}$$

on the nullspace of $adj(\Omega(\delta = 0))$. To prove the above equation, note that

$$na_n E\|\tilde{U}^k\|^2 \leq Ca_n E\|g_{i\delta_k}\|^2 \leq Cm,$$

by Assumption 6. Clearly by the same analysis in (42) (43)

$$\begin{aligned}\frac{na_n}{\mu_n}|\tilde{U}^{k'}\tilde{\Omega}^{-1}\tilde{g} - \tilde{U}^{k'}D_2\tilde{g}| &\leq \frac{na_n}{\mu_n}O_p(m/\sqrt{a_n n})O_p(\mu_n/\sqrt{na_n})o_p(\frac{1}{\sqrt{m}a_n}) \\ &\xrightarrow{p} 0,\end{aligned}$$

where the last step is on the nullspace of $adj(\Omega(\delta = 0))$, by Assumption 3, Lemma 1.

We need to simplify the following term in the partial derivative

$$\frac{1}{n}\sum_{i=1}^n \hat{U}_i^{k'}\tilde{\Omega}^{-1}g_i = tr[\frac{1}{n}\sum_{i=1}^n g_i\hat{U}_i^{k'}]\tilde{\Omega}^{-1}.$$

In the same manner as in (44)-(45), on the nullspace of $adj(\Omega(\delta = 0))$ we have , by using the above representation

$$\frac{na_n}{\mu_n}|(\frac{1}{n}\sum_{i=1}^n \hat{U}_i^{k'}\tilde{\Omega}^{-1}g_i)/n - (\frac{1}{n}\sum_{i=1}^n \tilde{U}_i^{k'}D_2 g_i)/n| \xrightarrow{p} 0. \tag{46}$$

Collecting (41)-(46) in (40)

$$\frac{na_n}{\mu_n}\frac{\partial\hat{Q}(0)}{\partial\delta_k} = \frac{na_n}{\mu_n}[\bar{g}'_{\delta_k}D_2\tilde{g}(1 - 1/n) + \tilde{U}^{k'}D_2\tilde{g} - \{\frac{1}{n}\sum_{i=1}^n \tilde{U}_i^{k'}D_2 g_i\}/n] + o_p(1). \tag{47}$$

We now try to derive the limit for the above expression. Note the following notation

$$\tilde{U}^k = n^{-1}\sum_{i=1}^{n} U_i S_n^{-1'} e_k \mu_n,$$

$$\bar{g}_{\delta_k} = G S_n^{-1'} e_k \mu_n.$$

Stack over k, and by simple algebra

$$\frac{na_n}{\mu_n}\frac{\partial \hat{Q}(0)}{\partial \delta} = na_n S_n^{-1}[G' D_2 \tilde{g}(1-1/n) + (n^{-1}\sum_{i\neq j} U_i' D_2 g_j)/n] + o_p(1). \tag{48}$$

Next for any vector $\|\lambda\| = 1$, we benefit from Technical Lemma 2. So write as

$$na_n \mu_n^{-1}\lambda' \frac{\partial \hat{Q}(0)}{\partial \delta} = (1-1/n)\sum_{i=1}^{n} X_i + \sum_{i\neq j} Z_i' Y_j + o_p(1).$$

$$X_i = a_n \lambda' S_n^{-1} G' D_2 g_i, \quad Z_i = U_i S_n^{-1'}\lambda/n, \quad Y_j = a_n D_2 g_j.$$

First, see that

$$\sum_{i=1}^{n} X_i^2 = na_n^2 \lambda' S_n^{-1} G' D_2 (\sum_{i=1}^{n} g_i g_i') D_2 G S_n^{-1}\lambda.$$

Assuming that $\lambda' S_n^{-1} G' D_2$ is on the nullspace of Ω, by adding and subtracting

$$\begin{aligned}
\sum X_i^2 &= [na_n \lambda' S_n^{-1} G' D_2 (\frac{a_n \sum g_i g_i'}{n}) D_2 G S_n^{-1}\lambda - na_n \lambda' S_n^{-1} G' D_2 (a_n \Omega) D_2 G S_n^{-1}\lambda] \\
&+ na_n \lambda' S_n^{-1} G' D_2 (a_n\Omega) D_2 G S_n^{-1}\lambda \\
&= [na_n \lambda' S_n^{-1} G' D_2 (\frac{a_n \sum g_i g_i'}{n} - \frac{a_n}{n}\Omega) D_2 G S_n^{-1}\lambda]
\end{aligned}$$

Then by Assumptions 1 and 6

$$[na_n \lambda' S_n^{-1} G' D_2 (\frac{a_n \sum g_i g_i'}{n} - \frac{a_n}{n}\Omega) D_2 G S_n^{-1}\lambda] - na_n \lambda' S_n^{-1} G' D_2 D_1 D_2 G S_n^{-1}\lambda \xrightarrow{p} 0,$$

and

$$na_n \lambda' S_n^{-1} G' D_2 D_1 D_2 G S_n^{-1}\lambda \to \lambda' H \lambda.$$

Next we benefit from the assumption that we make in the statement of Lemma A.4

$$\frac{a_n^2}{n^2}\sum_{i=2}^{n}\sum_{j=1}^{i-1}[\lambda' S_n^{-1} U_i' D_2 g_j g_j' D_2 U_i S_n^{-1}\lambda + \lambda' S_n^{-1} U_j' D_2 g_i g_i' D_2 U_j S_n^{-1}\lambda] \xrightarrow{p} \lambda' \Lambda \lambda.$$

This is the nearly-singular design counterpart of the assumption that is made in Newey and Windmeijer (2007). In Newey and Windmeijer (2007) they make the assumption that

$$\lambda' S_n^{-1} E[U_i' \Omega^{-1} U_i] S_n^{-1}\lambda \to \lambda' \Lambda \lambda.$$

We consider then the following condition

$$nE|X_i|^4 \leq na_n^2 E\|\lambda' \sqrt{na_n} S_n^{-1} G' D_2 g_i\|^4 / n^2.$$

Note that $\|\sqrt{na_n} S_n^{-1} G' D_2\| \leq C$, by Assumption 6. So

$$nE|X_i|^4 \leq C a_n^2 E\|g_i\|^4 / n \to 0,$$

by Assumption 6.

Note by Assumption 7i, the definition of U_i we have $\bar{\lambda}_z \leq C/\mu_n^2 n^2$,

$$mn^4 \bar{\lambda}_y^2 \bar{\lambda}_z^2 \leq Cmn^4 a_n^4 / (\mu_n^2 n^2)^2 \leq Cma_n^4/\mu_n^4 \to 0,$$

by Assumption 4i. Then

$$\begin{aligned} n^2[\bar{\lambda}_z^2 E\|Y_i\|^4 + \bar{\lambda}_y^2 E\|Z_i\|^4 &\leq n^2[\left(\frac{C}{\mu_n^4 n^4}\right) a_n^4 E\|g_i\|^4 + \frac{E\|G_i\|^4 a_n^4}{(n^4 \mu_n^4)}] \\ &\leq \frac{Cn^2 a_n^4}{\mu_n^4 n^4}[E\|g_i\|^4 + E\|G_i\|^4] \\ &= \frac{Ca_n^4}{n^2 \mu_n^4}[E\|g_i\|^4 + E\|G_i\|^4] \to 0, \end{aligned}$$

by Assumption 6 and Assumption 4i. Then the last condition is

$$\begin{aligned} n^2 E\|Y_i\|^4 E\|Z_i\|^4 &\leq Cn^2 a_n^4 E\|g_i\|^4 [E\|g_i\|^4 + E\|G_i\|^4]/\mu_n^4 n^4 \\ &= \frac{Ca_n^4}{n^2 \mu_n^4} E\|g_i\|^4 (E\|g_i\|^4 + E\|G_i\|^4) \to 0, \end{aligned}$$

by Assumption 6 and Assumption 4i. The conclusion follows from Technical Lemma 2 and Cramer Wold device. So

$$\frac{na_n}{\mu_n} \frac{\partial \hat{Q}(0)}{\partial \delta} \xrightarrow{d} N(0, \Lambda + H).$$

Q.E.D.

Lemma A.5. *Under Assumptions 1-7, there is an open convex set N_n such that $0 \in N_n$, and wpa1 $\hat{\delta} \in N_n$, $\hat{Q}(\delta)$ is twice continuously differentiable on N_n and for any $\bar{\delta}$ that is an element of N_n wpa1*

$$\frac{na_n}{\mu_n^2} \partial^2 \hat{Q}(\bar{\delta}) / \partial \delta \partial \delta' \xrightarrow{p} \tilde{H},$$

on the nullspace of $adj(\Omega(\bar{\delta}))$.

Proof of Lemma A.5. By the definition of our objective function

$$\hat{Q}(\delta) = \frac{\hat{g}(\bar{\delta})' \hat{\Omega}(\bar{\delta})^{-1} \hat{g}(\bar{\delta})}{2} - [\frac{1}{n} \sum_{i=1}^{n} g_i(\bar{\delta})' \hat{\Omega}(\bar{\delta})^{-1} g_i(\bar{\delta})]/2n.$$

Before taking second order partial derivative, we need to define notation to simplify the expression. Let $\hat{g}_i = g_i(\bar{\delta}), \hat{g} = \hat{g}(\bar{\delta}), \hat{\Omega} = \sum_{i=1}^n \hat{g}_i\hat{g}_i'/n$, $\hat{g}_{i\delta_k} = \partial g_i(\bar{\delta})/\partial\delta_k$, $\hat{g}_{\delta_k} = \partial\hat{g}(\bar{\delta})/\partial\delta_k$, $\hat{\Omega}^k = \sum_{i=1}^n \hat{g}_i\hat{g}_{i\delta_k}'/n, \hat{g}_{i\delta_k\delta_l} = \partial^2 g_i(\bar{\delta})/\partial\delta_k\partial\delta_l$, $\hat{g}_{\delta_k\delta_l} = \partial^2\hat{g}(\bar{\delta})/\partial\delta_k\partial\delta_l$, $\hat{\Omega}^{k,l} = \sum_{i=1}^n \hat{g}_{i\delta_k}\hat{g}_{i\delta_l}'/n, \hat{\Omega}^{kl} = \sum_{i=1}^n \hat{g}_i\hat{g}_{i\delta_k\delta_l}'/n$. Now we can take the second order partial derivative

$$\begin{aligned}
\frac{\partial^2\hat{Q}(\bar{\delta})}{\partial\delta_k\partial\delta_l} &= \hat{g}_{\delta_k}'\hat{\Omega}^{-1}\hat{g}_{\delta_l} + \hat{g}'\hat{\Omega}^{-1}\hat{g}_{\delta_k\delta_l} - \hat{g}'\hat{\Omega}^{-1}(\hat{\Omega}^k + \hat{\Omega}^{k'})\hat{\Omega}^{-1}\hat{g}_{\delta_l} \\
&- \hat{g}'\hat{\Omega}^{-1}(\hat{\Omega}^l + \hat{\Omega}^{l'})\hat{\Omega}^{-1}\hat{g}_{\delta_l} + \hat{g}'\hat{\Omega}^{-1}(\hat{\Omega}^l + \hat{\Omega}^{l'})\hat{\Omega}^{-1}(\hat{\Omega}^{k'} + \hat{\Omega}^k)\hat{\Omega}^{-1}\hat{g} - \hat{g}'\hat{\Omega}^{-1}(\hat{\Omega}^{kl} + \hat{\Omega}^{k,l})\hat{\Omega}^{-1}\hat{g} \\
&- \frac{1}{n}\sum_{i=1}^n[\hat{g}_{i\delta_k}'\hat{\Omega}^{-1}\hat{g}_{i\delta_l} + \hat{g}_i'\hat{\Omega}^{-1}\hat{g}_{i\delta_k\delta_l} - \hat{g}_i'\hat{\Omega}^{-1}(\hat{\Omega}^k + \hat{\Omega}^{k'})\hat{\Omega}^{-1}\hat{g}_{i\delta_l} - \hat{g}_i'\hat{\Omega}^{-1}(\hat{\Omega}^l + \hat{\Omega}^{l'})\hat{\Omega}^{-1}\hat{g}_{i\delta_k} \\
&+ \hat{g}_i'\hat{\Omega}^{-1}(\hat{\Omega}^l + \hat{\Omega}^{l'})\hat{\Omega}^{-1}(\hat{\Omega}^{k'} + \hat{\Omega}^k)\hat{\Omega}^{-1}\hat{g}_i - \hat{g}_i'\hat{\Omega}^{-1}(\hat{\Omega}^{kl} + \hat{\Omega}^{k,l})\hat{\Omega}^{-1}\hat{g}_i]/n. \qquad (49)
\end{aligned}$$

The next task is to replace the terms $\hat{\Omega}^{-1}, \hat{\Omega}^l, \hat{\Omega}^k, \hat{\Omega}^{kl}, \hat{\Omega}^{k,l}$ with the respective limits at $\bar{\delta}$, taking into account the nearly singular design for $\hat{\Omega}^{-1}$. For $\hat{\Omega}^{-1}$ that will be $D_2(\bar{\delta})$, the limit for $\hat{\Omega}^l$ is $\Omega^l(\bar{\delta}) = Eg_i(\bar{\delta})g_{i\delta_l}(\bar{\delta})'$. For $\hat{\Omega}^{kl}$ the limit is $\Omega^{kl}(\bar{\delta}) = Eg_i(\bar{\delta})g_{i\delta_k\delta_l}(\bar{\delta})'$ and for $\hat{\Omega}^{k,l}$, the limit is $\Omega^{k,l}(\bar{\delta}) = Eg_{i\delta_k}(\bar{\delta})g_{i\delta_l}(\bar{\delta})'$. In (49) see that

$$\|\hat{\Omega}^k - \Omega^k(\bar{\delta})\| \xrightarrow{p} 0,$$

by Assumption 7iii and $S_n^{-1}\mu_n$ bounded. Then in the same way

$$\|\hat{\Omega}^{kl} - \Omega^{kl}(\bar{\delta})\| \xrightarrow{p} 0,$$

$$\|\hat{\Omega}^{k,l} - \Omega^{k,l}(\bar{\delta})\| \xrightarrow{p} 0.$$

Next, see that

$$\mu_n^{-1}\sqrt{na_n}\|\frac{\partial\hat{g}(\bar{\delta})}{\partial\delta}\| = \sqrt{na_n}\|\hat{G}(\beta)S_n^{-1'}\| = \sqrt{na_n}\|\hat{G}(\beta_0)S_n^{-1'}\| + o_p(1),$$

by Assumption 7ii. Then

$$\|\sqrt{na_n}[\hat{G}(\beta_0) - G]S_n^{-1'}\|^2 = O_p(a_n E\|G_i\|^2)/\mu_n^2 = o_p(1), \qquad (50)$$

by Assumption 4i and 7ii, since $E\|G_i\|^2 \leq Cm$.

By Assumption 6ii and (50)

$$\sqrt{na_n}\|\hat{G}(\beta_0)S_n^{-1}\| \leq \sqrt{na_n}\|(\hat{G}(\beta_0) - G)S_n^{-1}\| + \sqrt{na_n}\|GS_n^{-1}\| = O_p(1).$$

So

$$\sqrt{na_n}\mu_n^{-1}\|\partial\hat{g}(\bar{\delta})/\partial\delta\| = O_p(1). \qquad (51)$$

By Assumption 7ii, similar to the previous result

$$\sqrt{na_n}\mu_n^{-1}\|\partial^2\hat{g}(\bar{\delta})/\partial\delta\partial\delta'\| = O_p(1). \qquad (52)$$

Next see that on the nullspace of $adj(\Omega(\bar{\delta}))$

$$\frac{na_n}{\mu_n^2}|\hat{g}'\hat{\Omega}^{-1}\hat{\Omega}^k\hat{\Omega}^{-1}\hat{g}_{\delta_l} - \hat{g}'D_2(\bar{\delta})\Omega^k(\bar{\delta})D_2(\bar{\delta})\hat{g}_{\delta_l}| \xrightarrow{p} 0,$$

by Lemma 1, 4iii, and (51)(52). (Nullspace of $adj(\Omega(\bar{\delta}))$ is the limit values of $\hat{g}_i$ and $\hat{g}_{i\delta_k}$, for each $i = 1, \cdots, n$) To see this result, note that by Lemma 1, adding and subtracting $adj(\Omega(\bar{\delta}))$, follow the analysis in (42)(43) on the nullspace of $adj(\Omega(\bar{\delta}))$,

$$\begin{aligned} \|\hat{\Omega}^{-1}\hat{\Omega}^k\hat{\Omega}^{-1} - D_2(\bar{\delta})\Omega^k(\bar{\delta})D_2(\bar{\delta})\| &\leq \|\hat{\Omega}^{-1}\hat{\Omega}^k(\hat{\Omega}^{-1} - D_2(\bar{\delta})\| + \|\hat{\Omega}^{-1}(\hat{\Omega}^k - \Omega^k(\bar{\delta}))D_2(\bar{\delta}))\| \\ &+ \|(\hat{\Omega}^{-1} - D_2(\bar{\delta}))\Omega^k(\bar{\delta})D_2(\bar{\delta})\| \xrightarrow{p} 0. \end{aligned}$$

Then in (49) note that we can write

$$\frac{1}{n}\sum_{i=1}^n \hat{g}'_{i\delta_k}\hat{\Omega}^{-1}\hat{g}_{i\delta_l} = tr[(\hat{\Omega}^{k,l})'\hat{\Omega}^{-1}].$$

We can use

$$\frac{na_n}{\mu_n^2}[\frac{1}{n}\sum_{i=1}^n \hat{g}'_{i\delta_k}\hat{\Omega}^{-1}\hat{g}_{i\delta_l}]/n = \frac{a_n}{\mu_n^2}tr[(\hat{\Omega}^{k,l})'\hat{\Omega}^{-1}].$$

The issue is whether, on the nullspace of $adj(\Omega(\bar{\delta}))$

$$\frac{a_n}{\mu_n^2}[tr(\hat{\Omega}^{kl'}\hat{\Omega}^{-1}) - tr(\hat{\Omega}^{kl'}D_2(\bar{\delta}))] \xrightarrow{p} 0.$$

This is true by Assumption 7i, $a_n m/\mu_n^2$ being bounded and Lemma 1. So apply these ideas to all the terms in (49) given Assumptions 4, 7 and Lemma 1, we can replace the estimator of variance terms with their respective limits on the nullspace of $adj(\Omega(\bar{\delta}))$:

$$\frac{na_n}{\mu_n^2}|\hat{Q}_{\delta_k\delta_l}(\bar{\delta}) - \tilde{Q}_{\delta_k\delta_l}(\bar{\delta})| \xrightarrow{p} 0, \tag{53}$$

where in $\tilde{Q}_{\delta_k\delta_l}$ we replace $\hat{\Omega}^{-1}, \hat{\Omega}^k, \hat{\Omega}^{kl}, \hat{\Omega}^{k,l}$ with $D_2(\bar{\delta}), \Omega^k(\bar{\delta}), \Omega^{kl}(\bar{\delta}), \Omega^{k,l}(\bar{\delta})$. Next we would like to show for $\tilde{Q}_{\delta_k\delta_l}$ that is evaluated at $\delta = 0$

$$\frac{na_n}{\mu_n^2}|\tilde{Q}_{\delta_k\delta_l}(\bar{\delta}) - \tilde{Q}_{\delta_k\delta_l}(0)| \xrightarrow{p} 0.$$

Same arguments that lead to (53) with (51)(52), and Assumption 7iii provides the result above. To simplify the notation we substitute $\Omega^k = \Omega^k(0)$, $\Omega^{kl} = \Omega^{kl}(0)$, $\Omega^{k,l} = \Omega^{k,l}(0), \tilde{g} = \hat{g}(0), \tilde{g}_{\delta_k} = \partial\hat{g}(0)/\partial\delta_k$, $\tilde{g}_{\delta_k\delta_l} = \partial^2\hat{g}(0)/\partial\delta_k\partial\delta_l$, $\tilde{g}_{i\delta_k} = \hat{g}_{i\delta_k}(0), \tilde{g}_{i\delta_k} = \hat{g}_{i\delta_k\delta_l}(0), \tilde{g}_i = \hat{g}_i(0), D_2(\delta = 0) = D_2$.

Expressing the second order partial derivative of the objective function with variance terms (estimators) replaced by their limits and evaluated at $\delta = 0$

$$\begin{aligned} \tilde{Q}_{\delta_k\delta_l}(0) &= \tilde{g}'_{\delta_k}D_2\tilde{g}_{\delta_l} + \tilde{g}'D_2\tilde{g}_{\delta_k\delta_l} - \tilde{g}'D_2(\Omega^k + \Omega^{k'})D_2\tilde{g}_{\delta_l} \\ &- \tilde{g}'D_2(\Omega^l + \Omega^{l'})D_2\tilde{g}_{\delta_k} + \tilde{g}'D_2(\Omega^l + \Omega^{l'})D_2(\Omega^{k'} + \Omega^k)D_2\tilde{g} - \tilde{g}'D_2(\Omega^{kl} + \Omega^{k,l})D_2\tilde{g} \\ &- \frac{1}{n}\sum_{i=1}^n[\tilde{g}'_{i\delta_k}D_2\tilde{g}_{i\delta_l} + \tilde{g}'_iD_2\tilde{g}_{i\delta_k\delta_l} - \tilde{g}'_iD_2(\Omega^k + \Omega^{k'})D_2\tilde{g}_{i\delta_l} \\ &- \tilde{g}'_iD_2(\Omega^l + \Omega^{l'})D_2\tilde{g}_{i\delta_l} + \tilde{g}'_iD_2(\Omega^l + \Omega^{l'})D_2(\Omega^{k'} + \Omega^k)D_2\tilde{g}_i - \tilde{g}'_iD_2(\Omega^{kl} + \Omega^{k,l})D_2\tilde{g}_i]/n. \end{aligned}$$

We can rewrite that by subtracting the terms in the large square bracket from the others

$$\begin{aligned}\frac{na_n}{\mu_n^2}\tilde{Q}_{\delta_k\delta_l}(0) &= \frac{a_n}{\mu_n^2}\sum_{i\neq j}(\tilde{g}'_{i\delta_k}D_2\tilde{g}_{j\delta_l}+\tilde{g}'_iD_2\tilde{g}_{j\delta_k\delta_l}-\tilde{g}'_iD_2(\Omega^k+\Omega^{k'})D_2\tilde{g}_{j\delta_l}\\ &- \tilde{g}'_iD_2(\Omega^l+\Omega^{l'})D_2\tilde{g}_{j\delta_k}+\tilde{g}'_iD_2(\Omega^l+\Omega^{l'})D_2(\Omega^k+\Omega^{k'})D_2\tilde{g}_j-\tilde{g}'_iD_2(\Omega^{kl}+\Omega^{k,l})D_2\tilde{g}_j)/n.\end{aligned}$$

We consider now applying Lemma A.1 to each term above. Analyze the first term

$$\frac{a_n}{\mu_n^2}(\frac{\sum_{i\neq j}\tilde{g}'_{i\delta_k}D_2\tilde{g}_{j\delta_l}}{n}).$$

Now set $d_n=\mu_n^2, Y_i=G_iS_n^{-1'}\mu_ne_k,\ Z_j=G'_jS_n^{-1'}\mu_ne_l,\ A=D_2$, to have

$$\frac{a_n}{\mu_n^2}(\frac{\sum_{i\neq j}\tilde{g}'_{i\delta_k}D_2\tilde{g}_{j\delta_l}}{n})=(n-1)a_nS_n^{-1}G'_kD_2G_lS_n^{-1'}+o_p(1),$$

where G_k,G_l represent G's k the and l th columns respectively. We obtain the result since our Assumptions 3, 4i, 6 and 7 satisfy Lemma A.1 conditions. Then analyzing the other terms in the $\tilde{Q}_{\delta_k\delta_l}(0)$ expression, using Lemma A.1, and $E\tilde{g}_i=Eg_i=0$

$$\frac{a_n}{\mu_n^2}(\frac{\sum_{i\neq j}\tilde{g}'_iD_2\tilde{g}_{j\delta_k\delta_l}}{n})=o_p(1).$$

This is true for all the other terms in the second order partial derivative of our objective function that is evaluated at $\delta=0$.

Then clearly

$$\frac{na_n}{\mu_n^2}\tilde{Q}_{\delta_k\delta_l}\xrightarrow{p}\tilde{H}_{k,l},$$

where

$$\tilde{H}=\lim_{n\to\infty}na_nS_n^{-1}G'D_2GS_n^{-1'},$$

since by Assumption 6, it is clear that $\lim_{n\to\infty}a_nS_n^{-1}G'D_2GS_n^{-1'}\to 0$. So having $(n-1)$ or n in front of the sequence above does not matter asymptotically. **Q.E.D.**

As a brief note, the main difference between the proof of Lemma A.13 in Newey and Windmeijer (2007) is the usage of Lemma A.1 here (jackknife term) instead of the quadratic term. In Newey and Windmeijer (2007) proof, usage of their Lemma A.1 results in additional "bias" that cancel each other. Here, these terms do not exist. The main result is the same but we have D_2 in the limit instead of Ω^{-1} because of the nearly singular design.

Proof of Theorem 2. First have a Taylor series expansion for the first order condition

$$\frac{\partial\hat{Q}(\hat{\beta})}{\partial\beta}=0,$$

as

$$S'_n(\hat{\beta}-\beta_0)=-[na_nS_n^{-1}\frac{\partial^2\hat{Q}(\bar{\beta})}{\partial\beta\partial\beta'}S_n^{-1'}]^{-1}[na_nS_n^{-1}\frac{\partial\hat{Q}(\beta_0)}{\partial\beta}],$$

where $\bar{\beta} \in (\beta_0, \hat{\beta})$. Then apply Lemma A.4 and Lemma A.5 to have the desired result.**Q.E.D**

The next two Lemmata shows clearly why CUE estimator is inconsistent given the nearly-singular design. The following Lemma is an extension of Lemma A.1 in Newey and Windmeijer (2007) to the nearly-singular design case. The main difference is that the terms in their Lemma is multiplied by the near-singularity rate "a_n" here. Before Lemma A.6, notation is the same in Lemma A.1, and $\Sigma_{YZ} = EY_iZ_i'$.

Lemma A.6.*If Y_i, Z_i are iid and $\lambda_{max}(AA') \leq C$, $\lambda_{max}(A'A) \leq C$, $\lambda_{max}(\Sigma_{yy}) \leq C, \lambda_{max}(\Sigma_{zz}) \leq C, a_n^2 m/d_n^2 \to 0$, $na_n^2\mu_y'\mu_y/d_n^2 \to 0, na_n^2\mu_z'\mu_z/d_n^2 \to 0$, $a_n^2 E[(Y_i'Y_i)^2]/nd_n^2 \to 0, a_n^2E[(Z_i'Z_i)^2]/nd_n^2 \to 0$, then*

$$a_n n\frac{\bar{Y}'A\bar{Z}}{d_n} = a_n tr(A\Sigma_{YZ}')/d_n + na_n\frac{\mu_y'A\mu_z}{d_n} + o_p(1).$$

Remark. Compared with Lemma A.1 for jackknife CUE, we have an additional "bias" term here. This is the term with trace. In Lemma A.1, that is zero.

Proof. This is similar to the proof for Lemma A.1 here. We try to skip similar steps, and instead show the steps where there are differences with the proof of Lemma A.1.

Without losing any generality, we can ignore the matrix "A" as done in the proof of Lemma A.1. Suppose first $\mu_y = \mu_z = 0$. Then we have

$$W_n = \frac{na_n}{d_n}(\bar{Y}'\bar{Z}).$$

Then

$$EW_n = E[a_nY_i'Z_i]/d_n = \frac{a_n}{d_n}tr(\Sigma_{YZ}').$$

Then follow the steps in the proof of Lemma A.1 in Newey and Windmeijer (2007)

$$E[(Y_i'Z_i)^2] \leq E[Y_i'Y_i]^2 + E[Z_i'Z_i]^2. \tag{54}$$

In addition to that equation we have (16)(17)

$$E[Y_i'Z_jZ_j'Y_i] \leq Cm,$$

$$E[Y_i'Z_jY_j'Z_i] \leq Cm.$$

Then

$$\begin{aligned} E[W_n^2]/n &\leq a_n^2E[Y_i'Z_i]^2/nd_n^2 \\ &\leq \frac{a_n^2}{d_n^2}\frac{E[Y_i'Y_i]^2 + E[Z_i'Z_i]^2}{n} \\ &\to 0. \end{aligned} \tag{55}$$

given the conditions in the statement of Lemma A.6. Given (54)-(55) exactly as in the proof of Lemma A.1 in Newey and Windmeijer (2007), $var(W_n) = o(1)$, and we get

$$W_n = a_ntr(\Sigma_{YZ}')/d_n + o_p(1).$$

The case for $\mu_y \neq 0, \mu_z \neq 0$ simply follows from Newey and Windmeijer (2007) after their equation (9.1), or from the proof of Lemma A.1 here by the analysis of the first four terms on the right-hand side of (19) as in (20)-(22) (Instead of $n-1$ use n in those terms). **Q.E.D**

The next Lemma extends Lemma A.2 of Newey and Windmeijer (2007) to the case of the nearly-singular design, and shows us why the CUE is inconsistent. The limit in the result of our Lemma A.7 below is different than the limit of CUE in Newey and Windmeijer (2007). This is due to the nearly singular design, and the intuition behind this result is described in the issues with GEL section. We briefly repeat some of those issues.

Note that the limit of the CUE objective function when there is nearly-singular design can be written as

$$Q_c(\beta) = \frac{\bar{g}(\beta)' D_2(\beta) \bar{g}(\beta)}{2} + a_n \frac{tr(D_2(\beta)\Omega(\beta))}{2n},$$

where $\bar{g}(\beta) = Eg_i(\beta)$. The following Lemma clearly shows that the "bias" term : $tr(D_2(\beta)\Omega(\beta))$ exists and depends on β. Furthermore in technical appendix we show that $D_2(\beta)\Omega(\beta) \neq 0$, $D_2(\beta)\Omega(\beta) \neq I$. The CUE objective function is

$$\hat{Q}_c(\beta) = \frac{\hat{g}(\beta)' \hat{\Omega}(\beta)^{-1} \hat{g}(\beta)}{2}.$$

Lemma A.7. *Under Assumptions 1-4, with* $\hat{\Omega}(\beta)^{-1} \xrightarrow{p} \infty$, *and* $adj(\Omega(\beta)) \neq 0$, *and* $E(g_i(\beta)' g_i(\beta))^2 / n a_n^2 \to 0$, *we have*

$$\sup_\beta \frac{n a_n}{\mu_n^2} |\hat{Q}_c(\beta) - Q_c(\beta)| \xrightarrow{p} 0.$$

Remark. It is clear that β_0 cannot be minimized at $Q_c(\beta)$ from Lemma A.7. In Newey and Windmeijer (2007), we see that the bias term is $tr(I_m)$, rather than $tr(D_2(\beta)\Omega(\beta))$ here.

Proof of Lemma A.7. This is derived easily by the proof of Lemma A.2 here. Note that by (24) in the proof of Lemma we see that CUE is the first term on the right-hand side of that equation. Following the proof of Lemma A.2, (ignoring the extra term (28)), and replacing $Q(\delta)$ there with $Q_c(\delta)$ here and applying Lemma A.6 here provides the result ($\delta = S_n(\beta - \beta_0)/\mu_n$).**Q.E.D.**

TECHNICAL APPENDIX:

Technical Lemma 1. *Under Assumption 1 with* $\hat{\Omega}(\beta)^{-1} \xrightarrow{p} \infty$, *(*$adj\Omega(\beta) \neq 0$*).*

(i). $D_2(\beta)\Omega(\beta) \neq I$.

(ii). $D_2(\beta)\Omega(\beta) \neq 0$.

Proofs.

(i). We prove this by contradiction. By (10) in Lemma 1, and since $\kappa(\beta) \neq 0$ from Lemma 1 we have

$$\Omega(\beta)[\frac{\Delta(\beta)}{\kappa(\beta)}] + D_1(\beta)[\frac{adj(\Omega(\beta))}{\kappa(\beta)}] = I.$$

Since $\Delta(\beta)/\kappa(\beta) = D_2(\beta)$ in Lemma 1, the equation above can be rewritten as

$$\Omega(\beta)D_2(\beta) + D_1(\beta)[\frac{adj(\Omega(\beta))}{\kappa(\beta)}] = I. \quad (56)$$

If it were to be that $\Omega(\beta)D_2(\beta) = I$, then in (56) we should have obtained

$$D_1(\beta)\frac{adj(\Omega(\beta))}{\kappa(\beta)} = 0. \quad (57)$$

Since $\kappa(\beta) \neq 0$, from Lemma 1 we can rewrite that as

$$D_1(\beta)adj(\Omega(\beta)) = 0. \quad (58)$$

We, now show that (58) cannot be true. Note that by Assumption 1, $D_1(\beta)$ is positive definite on the nullspace of $\Omega(\beta)$. Define $u = adj(\Omega(\beta))v, \quad v \neq 0$. Note that by Sinkhorn (1993) u is in the nullspace of $\Omega(\beta)$. Premultiply left hand side of (58) by $v'adj(\Omega(\beta))$ and postmultiply by v to have

$$v'adj(\Omega(\beta))D_1(\beta)adj(\Omega(\beta))v = u'D_1(\beta)u.$$

But since u is in the nullspace of $\Omega(\beta)$ and $D_1(\beta)$ is positive definite in that nullspace we have

$$u'D_1(\beta)u > 0. \quad (59)$$

But (59) contradicts the right hand side of (58) which is zero. Hence we have the desired result by contradiction $\Omega(\beta)D_2(\beta) \neq I$.

(ii). This is obtained by using equation (13) and the discussion below that in the proof of Lemma 1. **Q.E.D.**

Before Technical Lemma 2, we need the following notation. $EY_iY_i' = \Sigma_{yy}, EZ_iZ_i' = \Sigma_{zz}$. This extends Lemma A.10 of Newey and Windmeijer (2007) to our case of the nearly-singular design.

Technical Lemma 2. *If* (X_i, Y_i, Z_i) *are iid,* $EX_i = 0$, $EY_i = EZ_i = 0$, $EZ_iY_i' = 0$, $Z_i'\Sigma_{yy}Z_i = 0$

$$\sum_{i=1}^{n} X_i^2 \xrightarrow{p} A,$$

$$\sum_{i=2}^{n}\sum_{j=1}^{i-1}(Z_i'Y_j)^2+(Z_j'Y_i)^2\xrightarrow{p}\Lambda.$$

For $\bar{\lambda}_z=\lambda_{max}(\Sigma_{zz}),\bar{\lambda}_y=\lambda_{max}(\Sigma_{yy})$*,*$mn^4\bar{\lambda}_z^2\bar{\lambda}_y^2\to 0$*,*

$$n^2[\bar{\lambda}_zE\|Y_i\|^4+\bar{\lambda}_yE\|Z_i\|^4]\to 0.$$

We also assume $nEX_i^4\to 0$*,* $n^2E\|Z_i\|^4E\|Y_i\|^4\to 0$*, then*

$$\sum_{i=}^{n}X_i+\sum_{i\neq j}Z_i'Y_j\xrightarrow{d}N(0,A+\Lambda).$$

Remark. This is similar to Lemma A.10 of Newey and Windmeijer (2007). The limit result is the same, however since there is nearly-singular design we cannot use that Lemma. Newey and Windmeijer (2007) benefit from the variance of data in their Lemma. This is due to Corollary 3.1 in Hall and Heyde (1980) in proving the central limit theorem. Instead we use Theorem 3.2 in Hall and Heyde (1980) so that we can use sample versions of the second moments of data. This is the main difference between the Lemma here and Lemma A.10 of Newey and Windmeijer (2007).

Proof of Technical Lemma 2. Let $w_i=(X_i,Y_i,Z_i)$ and for any $j<i$,

$$\psi_{ij}=Z_i'Y_j+Z_j'Y_i.$$

Set

$$\sum_{i=1}^{n}X_i+\sum_{i\neq j}Z_i'Y_j=\sum_{i=2}^{n}(X_i+B_{in})+X_1,$$

where X_1 will be analyzed later, we consider the summation in the above equation first, and

$$B_{in}=\sum_{j<i}\psi_{ij}=(\sum_{j<i}Z_j)'Y_i+(\sum_{j<i}Y_j)'Z_i,$$

and $EX_iB_{in}=0$. Note that by Assumption

$$\sum_{i=2}^{n}X_i^2\xrightarrow{p}A. \tag{60}$$

Then consider

$$\begin{aligned}B_{in}^2 &= \sum_{j=1}^{i-1}\sum_{k=1}^{i-1}\psi_{ij}\psi_{ik}\\ &= \sum_{j=1}^{i-1}\psi_{ij}^2+\sum_{j\neq k,j,k<i}\psi_{ij}\psi_{ik}\end{aligned}$$

Set

$$C_{in}^2=\sum_{j=1}^{i-1}\psi_{ij}^2=\sum_{j=1}^{i-1}C_{ijn1}^2+\sum_{j=1}^{i-1}C_{ijn2},$$

where $C_{ijn1}^2 = (Z_i'Y_j)^2 + (Z_j'Y_i)^2$, and $C_{ijn2} = 2(Z_i'Y_jZ_j'Y_i)$. Note that $E(C_{ijn2}|w_{i-1}, \cdots, w_1) = 0$, since $EZ_iY_i' = 0$. So that term will play no role in our analysis. Set also

$$D_{in} = \sum_{j \neq k, j,k<i} \psi_{ij}\psi_{ik}.$$

We assume

$$\sum_{i=2}^{n} C_{ijn1}^2 \xrightarrow{p} \Lambda. \tag{61}$$

Conditions (60)(61) are the same as Condition (3.19) in Theorem 3.2 of Hall and Heyde (1980) for martingale central limit theorem. From section 3.2iii of Hall and Heyde (1980) it is clear that conditions

$$\sum_{i=2}^{n} E[D_{in}|w_{i-1}, \cdots, w_1] \xrightarrow{p} 0. \tag{62}$$

$$\sum_{i=2}^{n} E[X_iB_{in}|w_{i-1}, \cdots, w_1] \xrightarrow{p} 0. \tag{63}$$

$$\sum_{i=2}^{n} E[(X_i + B_{in})^4] \to 0. \tag{64}$$

are sufficient conditions for (3.18)(3.20) in Theorem 3.2 in Hall and Heyde (1980). Note that conditions (63)(64) are used Lemma A.10 of Newey and Windmeijer (2007).

Now we prove sufficient conditions (62)-(64) for our Lemma. We start with condition (62).

$$E[D_{in}|w_{i-1}, \cdots, w_1] = (\sum_{j<i} Y_j)'\Sigma_{zz}(\sum_{k<i} Y_j),$$

by $EZ_iY_i' = 0$, and $Z_i'\Sigma_{yy}Z_i = 0$. This last equation is equivalent to

$$E[D_{in}|w_{i-1}, \cdots, w_1] = R_{1i} + R_{2i},$$

where

$$R_{1i} = \sum_{k<i} S_k = \sum_{k<i}(\sum_{j<k} Y_j)'\Sigma_{zz}Y_k,$$

where

$$S_k = (\sum_{j<k} Y_j)'\Sigma_{zz}Y_k,$$

$$R_{2i} = \sum_{j<k<i} Y_k'\Sigma_{zz}Y_j.$$

Consider

$$\begin{aligned} E[Y_i'\Sigma_{zz}\Sigma_{yy}\Sigma_{zz}Y_i] &\leq \bar{\lambda}_y E[Y_i'\Sigma_{zz}\Sigma_{zz}Y_i] \\ &= \bar{\lambda}_y tr(\Sigma_{zz}\Sigma_{yy}\Sigma_{zz}) \\ &\leq m\bar{\lambda}_y^2\bar{\lambda}_z^2. \end{aligned}$$

So

$$ES_i^2 \leq (i-1)m\bar{\lambda}_y^2\bar{\lambda}_z^2,$$

Then we know that

$$E[S_i|w_{i-1},\cdots,w_1] = 0,$$

$$\begin{aligned} E[(\sum_{i=2}^n R_{1i})^2] &= E[(\sum_{i=2}^n (n-i+1)S_i)^2] \\ &= \sum_{i=2}^n (n-i+1)^2 ES_i^2 \\ &\leq \sum_{i=2}^n (n-i+1)^2(i-1)m\bar{\lambda}_y^2\bar{\lambda}_z^2 \\ &\leq mn^4\bar{\lambda}_z^2\bar{\lambda}_y^2 \\ &\rightarrow 0. \end{aligned}$$

by Assumption in the statement of Technical Lemma 2. So $\sum_{i=2}^n R_{1i} \xrightarrow{p} 0$. The same analysis applies

$$\sum_{i=2}^n R_{2i} \xrightarrow{p} 0.$$

So we obtain

$$\sum_{i=2}^n E(D_{in}|w_{i-1},\cdots,w_1) \xrightarrow{p} 0.$$

Now we show (63). Following exactly p.48 of Newey and Windmeijer (2007)

$$\sum_{i=2}^n E[X_iB_{in}|w_{i-1},\cdots,w_1] = E[X_iZ_i']\sum_{i=1}^n (n-i)Y_i + [EX_iY_i']\sum_{i=1}^{n-1}(n-i)Z_i.$$

Then we have

$$\begin{aligned} E[(\sum_{i=2}^n E[X_iB_{in}|w_{i-1},\cdots,w_1])^2 &\leq C(E[X_iY_i']\Sigma_{zz}E[Y_iX_i] + E[X_iZ_i']\Sigma_{yy}E[Z_iX_i])\sum_{i=1}^{n-1}(n-i)^2 \\ &\leq Cn^3E\|X_i\|^4[\bar{\lambda}_zE\|Y_i\|^4 + \bar{\lambda}_yE\|Z_i\|^4] \\ &\rightarrow 0, \end{aligned}$$

by the Assumptions $nEX_i^4 \to 0, n^2[\bar{\lambda}_z E\|Y_i\|^4 + \bar{\lambda}_y E\|Z_i\|^4] \to 0$. So clearly we satisfy

$$\sum_{i=2}^{n} E[X_i B_{in} | w_{i-1}, \cdots, w_1] \xrightarrow{p} 0.$$

Next we try to prove (64)

$$\sum_{i=2}^{n} E[(X_i + B_{in})^4] \leq CnEX_i^4 + C\sum_{i=2}^{n} EB_{in}^4.$$

Now

$$\sum_{i=2}^{n} EB_{in}^4 \leq \sum_{i=2}^{n} \{CE[\sum_{j<i}(Y_j' Z_i)^4] + CE(\sum_{j<i} Z_j' Y_i)^4\}.$$

Analyze each term on the right-hand side of the above equation. Proceeding exactly in the proof of Lemma A.10 in Newey and Windmeijer (2007), by using $Z_i'\Sigma_{yy}Z_i = 0$ in our case we have

$$\begin{aligned}
\sum_{i=2}^{n} E[(\sum_{j<i} Y_j' Z_i)^4] &= E(Z_i'\Sigma_{yy}Z_i)^2 \sum_{i=2}^{n} 3(i-1)(i-2) + E(Z_1' Y_2)^4 \sum_{i=2}^{n} (i-1) \\
&\leq n^2 E\|Z_i\|^4 E\|Y_i\|^4 \\
&\to 0,
\end{aligned}$$

by assumption. Same analysis applies to the term $\sum_{i=2}^{n} E[(\sum_{j<i} Z_j' Y_i)^4]$. So we show that $\sum_{i=2}^{n} EB_{in}^4 \to 0$. This proves (64). So we show that conditions (62)-(64) are holding. Note that by Assumption $nEX_i^4 \to 0$, we obtain $X_1 \xrightarrow{p} 0$. Then with (60)-(64) provides us the desired result. **Q.E.D**