Nearly Singular Design in GMM and Generalized Empirical Likelihood Estimators

Mehmet Caner
University of Pittsburgh *

December 10, 2004

Abstract

This article analyzes Generalized Empirical Likelihood (GEL) estimators and GMM under nearly singular design. This design relaxes the nonsingularity assumption of the limit weight matrix in GMM, and nonsingularity of the limit variance matrix for the first order conditions in GEL. The sample versions of these matrices are nonsingular, but one or more of the eigenvalues are near zero. In large samples we assume these sample matrices converge to a singular matrix. Usage of the generalized inverses for the sample variance estimate does not solve this problem since the limit variance is singular not the sample version. This kind of problem can result in large size distortions for the overidentifying restrictions test and poor small sample performance of the estimators. This nearly singular design may occur because of the usage of similar instruments in these matrices. However, since the sample versions of these matrices are nonsingular practitioners may ignore deleting the problematic instruments. In this paper, we derive the large sample theory for GMM and GEL estimators under nearly singular design. We show that the rate of convergence of the estimators and the Lagrange Multiplier in GEL is slower than root n. This rate depends on the nature of the problem. However, the limits are the same as in the standard case. The test of overidentifying restrictions has the same limit. We also derive higher order expansions and show that bias converges to zero slowly compared with standard GEL and GMM. Empirical likelihood estimator has still the best bias property among other GEL estimators, this is much more clear in nearly singular design.

Keywords: Singular Matrix, Rate of Convergence, Small Sample Properties.

*Department of Economics, WW Posvar Hall 4509, Pittsburgh, PA 15260. email: caner@pitt.edu. We thank Keith Knight and Offer Lieberman for discussions related to the paper. We also thank the participants of NBER/NSF Time Series Workshop 2004 at SMU for their comments on the paper.

1 Introduction

Asymptotic normality provides a poor approximation to the sampling distribution of GMM estimators and test statistics in finite samples. Sampling distribution of estimates can be skewed. The tests of overidentifying restrictions exhibit substantial size distortions. These are described in various studies such as Tauchen (1986), Kocherlakota (1990), and Hansen, Heaton, and Yaron (1996). One possible source of these problems is weak identification as described in Stock and Wright (2000). This is low correlation between instruments and the relevant first order conditions. Some simulations suggest that the poor performance of the conventional normal approximation to the finite sample distributions can be explained by the differences between the GMM weight matrix and its population value as in Pagan and Robertson (1997).

We can relate this to the "nearly singular" design. This is a term used by Knight and Fu (2000) in least squares case. In least squares case, they relax the assumption of nonsingularity of the limit variance matrix for the estimators. Even though the sample variance matrix is invertible, in the limit it converges to a singular matrix. However, in a certain locality, a positive definite matrix exists on the nullspace of the singular variance matrix. This is obviously related to multicollinearity problem. This problem affects the rate of convergence of the estimators that they use.

We follow a similar route in GMM and GEL estimators. In GMM, even though the weight matrix is not singular the population value may be singular. However, there may be a matrix which is positive definite on a certain locality of the singular matrix. These problems arise when there are very similar instruments. The weight matrix may have one or more small eigenvalues. This can result in large values for tests of overidentifying restrictions as seen in finite sample studies. Empirical research is often conducted without checking this type of behavior or analyzing the consequences of this type of problem. Since the weight matrix is invertible, even though the eigenvalues may be small, the researchers do not delete some of the instruments or reparametrize the problem.

Generalized Empirical Likelihood (GEL) estimators are recently introduced by Newey and Smith (2004) as an alternative to GMM. GEL estimators have two advantages over GMM. First, the asymptotic bias does not grow with the number of moment restrictions. Then bias corrected empirical likelihood estimator is higher order efficient relative to the other GEL and GMM estimators. However we observe data related problems such as weak instruments affect the limit distributions of both GMM and GEL. This point can be clearly seen by Stock and Wright (2000) and Guggenberger and Smith (2003). GEL estimators depend crucially on orthogonality conditions and the nonsingularity of the limit variance of the first order conditions as can be seen in Assumption 1 of Newey and Smith (2004). As suggested above, usage of the similar instruments or similar first order conditions may render one or more of the eigenvalues to be small. Hence the sample variance matrix is invertible but may result in poor small sample performance. We can also model this as nearly singular design. Usage of the generalized inverses for the sample variance estimate does not solve this problem since the limit variance is singular not the sample version.

Nearly singular design can play an important role both in large sample and small sample properties of GEL estimators and GMM. In this paper we analyze nearly singular asymptotics and higher order expansions. We make several contributions to the literature. First, we show that the rate of convergence of GEL and GMM estimators are affected by nearly singular design. The worser the problem of similar instruments, the slower the rate of convergence gets. Second, the limits of GMM and GEL estimators do not change compared to standard GEL and GMM estimators on the nullspace of singular variance matrix. These two findings show that we need large samples in the case of nearly singular design to have meaningful results. Third, the limit of the overidentifying restrictions test stays the same as in the standard case. Fourth, we also develop higher order expansions for GMM and GEL estimators. These establish that in nearly singular design, the higher order terms converge to zero slower compared with the higher order terms in standard GMM and GEL of Newey and Smith (2004).

In nearly singular design we also observe that empirical likelihood estimator has better bias properties compared to other GEL estimators. But the effect here is more important compared with standard GEL since the bias terms converge to zero slower here, and these extra bias terms in other GEL estimators result in poor small sample performance. We also analyze a hybrid GMM estimator, which uses no weight matrix in the first step and uses a nearly singular one in the second step. We compare the higher order properties of this with regular GMM and GEL estimators.

Section 2 presents the model and the assumptions. Section 3 deals with asymptotics of GEL estimators under nearly singular design. Section 4 introduces the higher order expansion for GEL. Section 5 examines GMM under nearly singular design. The last section concludes. The appendix covers all the proofs except from section 5.1. Technical appendix covers all the proofs in section 5.1.

2 The Model and Assumptions

In this section we describe the Generalized Empirical Likelihood Estimators (from now on GEL) that are introduced by Newey and Smith (2004). Let the moment restrictions at the true parameter value β_0 be

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

for $i = 1, \cdots, n$, where β is a p-dimensional parameter vector with β_0 as the true value and this is assumed to be in the interior of the compact parameter space $\mathbf{B}$, $\mathbf{B} \subset R^p$, x_i is the R^l valued data.

In (1) we have m restrictions. We can rewrite (1) as

$$Eg_i(\beta_0) = 0, \tag{2}$$

where $g_i(\beta)$ is of m dimension. We assume $m \geq p$.

Now we define the GEL estimator of β_0 due to Smith (1997) and Newey and Smith (2004). Let ρ be a function of a scalar ν that is concave on its domain, an open interval O containing 0. Set $\hat{\Lambda}_n(\beta) = \{ \lambda \in R^m : \lambda' g_i(\beta) \in O\}$ for $i = 1, 2, \cdots, n$. ρ_0 represents $\rho(.)$ evaluated at 0. The GEL estimator is the solution to a saddle point problem

$$\hat{\beta}_{GEL} = \arg\min_{\beta \in \mathbf{B}} \sup_{\lambda \in \hat{\Lambda}_n(\beta)} \hat{P}(\beta, \lambda), \tag{3}$$

where

$$\hat{P}(\beta, \lambda) = [2 \sum_{i=1}^{n} \rho(\lambda' g_i(\beta))/n - 2\rho_0.]$$

This formulation is consistent with Newey and Smith (2004) and used by Guggenberger and Smith (2003). This provides some ease in our proofs, so we use the formulation in (3), compared to $P(\beta, \lambda) = \sum_{i=1}^{n} \rho(\lambda' g_i(\beta))/n$ which is used in Newey and Smith (2004).

GEL estimators are alternatives to two-step GMM. The GEL estimators that are used in the literature is Empirical Likelihood of Owen (2001), Qin and Lawless (1994) where $\rho(\nu) = ln(1 - \nu)$. The exponential tilting estimator sets $\rho(\nu) = -e^{\nu}$ and seen in Kitamura and Stutzer (1997), Smith (1997), Imbens, Spady and Johnson (1998). Continuous Updating Estimator is just another version of GMM and $\rho(\nu) = -(1 + \nu)^2/2$.

We define notation that is useful for the subsequent sections. Let

$$\hat{g}(\beta) = 1/n \sum_{i=1}^{n} g_i(\beta),$$

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

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

$$\Omega(\beta) = \lim_{n \to \infty} E n^{-1} \sum_{i=1}^{n} g_i(\beta) g_i(\beta)'.$$

The following Assumptions are needed.

Assumption 1.(Nearly Singular Design)

(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 $a_n = n^{\kappa}, 0 < \kappa < 1$, uniformly over β

$$a_n(\hat{\Omega}(\beta) - \Omega(\beta)) \xrightarrow{p} D(\beta), \tag{4}$$

where $D(\beta)$ is positive definite for all β on the null space of $\Omega(\beta)$. (i.e. for $u \neq 0$, $u'D(\beta)u > 0$, where $\Omega(\beta)u = 0$ for all β) $D(\beta)$ is continuous in β.

(iii).

$$\sup_{\beta \in \mathbf{B}} \|D(\beta)\| < \infty.$$

Assumption 2.

(i). $g_i(\beta)$ is independent.

(ii). $\sup_i E \sup_{\beta \in \mathbf{B}} \|g_i(\beta)\|^{\xi} < \infty$, where $\xi > 2/(1-\kappa)$.

(iii).

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

where

$$\lim_{n \to \infty} \frac{a_n}{n} \sum_{i=1}^{n} EL_i^{2+\delta} < \infty$$

for some $\delta > 0$, and L_i does not depend on β.

(iv). $\mathbf{B}$ is a compact subset of R^p, β_0 is in the interior of $\mathbf{B}$.

Assumption 3. *$\rho(\nu)$ is twice continuously differentiable in a neighborhood of zero, $\rho(.)$ is concave on the domain of $\nu \in O$, ρ_1 represents the first-order partial derivative and ρ_2 represents the second-order partial derivative and $\rho_1(0) = \rho_2(0) = -1$.*

Note that given Assumption 2, $m_1(\beta)$ is defined as uniformly over $\beta \in \mathbf{B}$

$$\lim_{n \to \infty} E a_n n^{-1} \sum_{i=1}^{n} g_i(\beta) = m_1(\beta) < \infty.$$

Assumption 4. *$m_1(\beta)$ is continuous in β and $m_1(\beta) = 0$ iff $\beta = \beta_0$.*

The first assumption is about the singularity of $\Omega(\beta)$. This is the point of departure from the GEL literature and the standard econometrics literature as in Newey and Smith (2004). Even though $\Omega(\beta)$ is singular a suitable centering and scaling may give a nonsingular limit matrix. The same idea is used for the least squares case in model selection literature in Knight and Fu (2000). With this assumption we can learn what may happen when similar instruments are used. Assumption 1 also plays a crucial role in determining the rates of convergence in the functional central limit theorem and uniform law of large numbers. An interesting point is what if $a_n = n$? The problem in our case with $a_n = n$, is the functional central limit theorem result is impossible to derive because of the problems in Assumption 2ii, when κ takes the value of 1. So it is not clear how to proceed to get limits. In order to understand the nature of this assumption better we adapt an example of Knight and Fu (2000). They analyze a case in LS and with iid errors. Set $\hat{\Omega}_{xx} = \frac{1}{n}\sum_{i=1}^n X_i X_i'$ where X_i represent regressors. They normalize $\hat{\Omega}_{xx}$ in the following way

$$\hat{\Omega}_{xx} = \begin{pmatrix} 1 & \rho_n & \cdots & \rho_n \\ \rho_n & 1 & \cdots & \rho_n \\ \vdots & \vdots & \ddots & \vdots \\ \rho_n & \cdots & \rho_n & 1 \end{pmatrix},$$

where $\rho_n \xrightarrow{p} 1$ and $a_n(1-\rho_n) \xrightarrow{p} \tau > 0$. We can see that $\hat{\Omega}_{xx}$ converges in probability to a matrix Ω_{xx} of all 1's. Then clearly $a_n(\hat{\Omega}_{xx} - \Omega_{xx}) \xrightarrow{p} D_{xx}$ where

$$D_{xx} = \begin{pmatrix} 0 & -\tau & \cdots & -\tau \\ -\tau & 0 & \cdots & -\tau \\ \vdots & \vdots & \ddots & \vdots \\ -\tau & \cdots & -\tau & 0 \end{pmatrix}.$$

Assumption 2 is standard in the GEL literature as can be seen in articles of Newey and Smith (2004), Guggenberger and Smith (2003). However, Assumption 2ii ties the moment conditions to "κ" the coefficient of near singularity. Assumption 2iii is a Lipschitz condition used for obtaining Donsker type of result in the literature as in Andrews (1994). We strengthen the moment existence and finiteness in order to take into account the singularity problem. Because of Assumption 1 we take $f_{ni}(.)$ in Theorem 10.6 of Pollard (1990) to be $\frac{a_n^{1/2}}{n^{1/2}} g_i(\beta)$ in our case. This results in the specific rate of Assumption 2iii.

Assumption 3 describes $\rho(.)$ and standard in the literature. Assumption 4 is a standard identification condition which takes into account the nearly singular design in Assumption 1. In Assumption 4 the convergence rate of the first moments a_n/n provides us with the consistency. Without this assumption consistency is not possible. In standard GEL of Newey and Smith (2004), this rate is $1/n$ and in weakly identified GEL of Guggenberger and Smith (2003) this rate is $1/n^{1/2}$. This rate in Assumption 4 is not arbitrary, this benefits from Assumption 1. This rate will be used in getting a Uniform Law of Large Numbers in Lemma A.1.

3 Asymptotics

In this section we first provide two results that benefit the other theorems in the paper. The first one is modified Lindeberg Central Limit Theorem. The second one is a empirical central limit theorem which takes into account nearly singular design of Assumption 1.

Lemma 1. *Under Assumptions 1, 2i, ii, on the nullspace of $\Omega(\beta_0)$, the following result holds jointly with (4)*

$$[a_n^{1/2} n^{-1/2} \sum_{i=1}^n g_i(\beta_0)] \xrightarrow{d} N(0, D).$$

Remark. This is a trivial extension of Lindeberg Central Limit Theorem given Assumption 1. This Lemma also helps us understand where the nearly singular design will affect our problem. We see that in contrast to usual $n^{1/2}$ rate for $\hat{g}(\beta_0)$ we have $a_n^{1/2} n^{1/2}$ because of Assumption 1 and singularity of $\Omega(\beta_0)$.

Now we provide one of the important results of the study. This is a Donsker type result and it also generalizes Lemma 1. The proof of this Theorem is given in the Appendix.

Theorem 1. *Under Assumptions 1-2, and on the nullspace of $\Omega(\beta)$, the following result holds jointly with (4)*

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

where $\Psi(\beta)$ is a zero-mean Gaussian process with covariance $var(\Psi(\beta)) = D(\beta)$, for all $\beta \in \mathbf{B}$.

Remark. This also shows that the empirical process $\Psi_n(\beta)$ converges at rate $a_n^{1/2}$. This is mainly because of the finite dimensional convergence rate and the nearly singular design of $\hat{\Omega}(\beta)$. Theorem 1 also extends the Donsker result in Pollard (1990) from standard case to nearly singular case. This is a new result and helps us understand the behavior of estimators in nearly singular context.

Next we provide one of the key results in the paper. This provides the consistency of GEL under nearly singular design and provides rate of convergence for $\hat{\lambda}$, the estimate of the Lagrange Multiplier.

Theorem 2. *Under Assumptions 1-4, on the nullspace of $\Omega(\beta)$, the following results hold jointly with (4)*

(i).

$$\hat{\beta} \xrightarrow{p} \beta_0,$$

(ii).

$$\hat{g}(\hat{\beta}) = O_p(n^{-1/2} a_n^{-1/2}),$$

(iii).

$$\hat{\lambda} = \arg \max_{\lambda \in \hat{\Lambda}_n(\hat{\beta})} \hat{P}(\hat{\beta}, \lambda)$$

exists wpa1.

(iv).

$$\hat{\lambda} = O_p(a_n^{1/2} n^{-1/2}).$$

Remark. This theorem provides two new results compared to Theorem 3.1 of Newey and Smith (2004). First of all consistency is preserved even though we may have a nearly singular design as in Assumption 1. Secondly, we observe changes in the convergence rates both in the sample moment and the estimator of Lagrange Multiplier. The nearly singular design particularly slows down the rate of convergence of $\hat{\lambda}$, this is unlike the standard GEL case of Newey and Smith (2004) and the weakly identified GEL of (Guggenberger and Smith (2003)). In both standard and weakly identified GEL, $\hat{\lambda}$ converges at rate $n^{1/2}$. This slowing down of the rate of convergence happens because of the nearly singular behavior of $\hat{\Omega}(\beta)$.

Now set $G_i(\beta) = \frac{\partial g_i(\beta)}{\partial \beta'}$ which is the partial derivative matrix and dimension of $m \times p$. We have the following assumption that is helpful in deriving the limit theory.

Assumption 5. *Uniformly over $\beta \in \mathbf{B}$*

$$\frac{a_n}{n} \sum_{i=1}^{n} G_i(\beta) \xrightarrow{p} G(\beta),$$

where $G(\beta) = \lim_{n\to\infty} a_n/n \sum_{i=1}^{n} EG_i(\beta)$, $G(\beta)$ is continuous in β and $G(\beta_0)$ has full column rank p.

This Assumption is somewhat similar to Assumption 4. Note that in Assumption 4 with the rate of a_n/n we are able to derive the limit theory for $\hat{\beta}, \hat{\lambda}$ and this can be seen from the proof of Theorem 3 . Without this rate, the derivation is not possible. This Uniform Law of Large Numbers Assumption can be proven under primitive assumptions such as Assumption 2iii applied to $G_i(\beta)$ (i.e. Assuming Lipschitz continuity and bounded moments). Following the proof of Lemma A1 provides us then the Uniform Law of Large Numbers mentioned in Assumption 5.

The following Theorem is one of the main results of the paper. We derive the limit law under the nearly singular design for GEL estimators. Set $G(\beta_0) = G$ for notational convenience.

Theorem 3. *Under Assumptions 1-5, on the nullspace of $\Omega(\beta)$, the following result holds jointly with (4)*

$$\frac{n^{1/2}}{a_n^{1/2}}[(\hat{\beta} - \beta_0)', \hat{\lambda}']' \xrightarrow{d} N(0, diag(\Sigma, P)),$$

where $D = D(\beta_0), \Sigma = (G'D^{-1}G)^{-1}, H = \Sigma G' D^{-1}, P = D^{-1} - D^{-1}G\Sigma G' D^{-1}$.

Remark. This result shows that in a nearly singular design (Assumption 1) the rate of convergence slows down. If the problem is acute; for example if the instruments are very similar to each other (i.e. large a_n, large κ, κ is near 1) then the rate of convergence is $\frac{n^{1/2}}{a_n^{1/2}} = n^{1/2-\kappa/2}$ and this shows that the rate of convergence is very slow. We need very large samples to achieve meaningful results in such a case. This theorem extends Theorem 3.2 of Newey and Smith (2004) from standard GEL to nearly singular GEL.

Corollary 1. *Under Assumptions 1-5, on the nullspace of $\Omega(\beta)$, the following result holds jointly with (4)*

$$n\hat{P}(\hat{\beta}, \hat{\lambda}) \xrightarrow{d} \chi^2_{m-p}.$$

Remark. This result shows that in large samples the nearly singular design does not play an important role in terms of the overidentifying restrictions test since the limit is the same as in the standard GEL. Corollary 1 extends the standard GEL overidentifying restrictions test in Newey and Smith (2004) to nearly singular case. Even though the J test has the same limit as in the standard GEL case, as we shall see in higher order expansions the bias term converges to zero slower than the case in standard GEL. This should affect adversely the small sample properties of this test.

In the subsequent section we consider higher-order expansions of the estimators which provides us some hints about the small sample behavior of estimators.

4 Higher Order Expansion

In this part we analyze the higher order properties of GEL estimators with nearly singular design. The following assumption is needed for stochastic expansion for our estimators. Let ∇^j denote the vector of all distinct partial derivatives with respect to β of order j.

Assumption 6.*$\nabla^j g(x_i,\beta)$ exists on a neighborhood $\mathcal{N}$ of β_0,*

$$\max_i \sup_{\beta\in\mathcal{N}} \|\nabla^j g(x_i,\beta)\| \le l(x),$$

for $0 \le j \le 4$ and for each $\beta \in \mathcal{N}$

$$\|\nabla^4 g(x_i,\beta) - \nabla^4 g(x_i,\beta_0)\| \le l(x_i)\|\beta-\beta_0\|,$$

where

$$\lim_{n\to\infty} \frac{a_n}{n}\sum_{i=1}^{n} El(x_i)^{10+5\delta} < \infty,$$

for $\delta > 0$. $\rho(\nu)$ is four times continuously differentiable with Lipschitz fourth derivative in a neighborhood of zero.

Using the first-order condition for GEL in (21)-(22) and set $g(x_i,\beta) = g_i(\beta)$ we define

$$m(x_i,\theta) = \rho_1(\lambda' g_i(\beta)) \begin{pmatrix} G_i(\beta)'\lambda \\ g_i(\beta) \end{pmatrix}. \tag{5}$$

Then suppose

$$\frac{a_n^{1/2}}{n^{1/2}}\sum_{i=1}^{n} vec\left[\frac{\partial m(x_i,\theta_0)}{\partial\theta'} - E\frac{\partial m(x_i,\theta_0)}{\partial\theta'}\right] = O_p(1), \tag{6}$$

$$\frac{a_n^{1/2}}{n^{1/2}}\sum_{i=1}^{n} vec\left[\frac{\partial^2 m(x_i,\theta_0)}{\partial\theta_j\partial\theta'} - E\frac{\partial^2 m(x_i,\theta_0)}{\partial\theta_j\partial\theta'}\right] = O_p(1), \tag{7}$$

for $j = 1, \cdots, p, p+1, \cdots, m+p$. $\theta = (\beta', \lambda')'$. $\theta_j = \beta_j$, for $j = 1, 2, \cdots, p$ and $\theta_j = \lambda_{j-p}$, for $j = p+1, \cdots, p+m$.

$$\frac{a_n^{1/2}}{n^{1/2}} \sum_{i=1}^{n} vec\left[\frac{\partial^3 m(x_i, \theta_0)}{\partial\theta_k \partial\theta_j \partial\theta'} - E\frac{\partial^3 m(x_i, \theta_0)}{\partial\theta_k \partial\theta_j \partial\theta'}\right] = O_p(1), \tag{8}$$

for $j, k = 1, 2 \cdots, p, p+1, \cdots, p+m$.

Expressions for (6)-(8) in our case is shown in the proof of Theorem 4. These correspond to Assumption d in iid nonsingular GEL in Lemma A.4 of Newey and Smith (2004). In iid nonsingular case, the primitive assumptions on (6)-(8) amount to uniform bound on partial derivatives of $m(.)$. We discuss the primitive assumptions for (6)-(8) in our case in the proof of Theorem 4. Even though (6)-(8) are high level, this saves a lot of notation and needed for stochastic expansion in nearly singular design. (6)-(8) are nothing more than Central Limit Theorems for partial derivatives of the first order conditions for GEL (first, second, and third order partial derivatives).

As in Newey and Smith (2004) let F denote the distribution of the data x, $\psi(x, F)$ is a function of x and F with $E\psi(x, F_0) = 0$, and $\tilde{\psi} = a_n^{1/2}/n^{1/2}\sum_{i=1}^{n}\psi(x_i)$. $\psi(.)$ function will represent the first order conditions in GEL as can be seen from the proof of Theorem 4.

Theorem 4. *Under Assumptions 1-6, and with (6)-(8) holding we have the following stochastic expansion for $\hat{\theta}$ on the nullspace of $\Omega(\beta)$*

$$r_n(\hat{\theta} - \theta_0) = \tilde{\psi} + Q_1(\tilde{\psi}, \tilde{a}, F_0)/r_n + Q_2(\tilde{\psi}, \tilde{a}, \tilde{b}, F_0)/r_n^2 + O_p(r_n^{-3}),$$

where $r_n = n^{1/2}/a_n^{1/2} = n^{(1-\kappa)/2}, 0 < \kappa < 1$. $\tilde{\psi}, Q_1(.), Q_2(.)$ are $O_p(1)$ terms, and with them $\tilde{a}, \tilde{b}$ are all explained in Lemma A.5 and the proof of Theorem 4.

Remark. Theorem 4 is the nearly singular version of the stochastic expansion in standard GEL of Theorem 3.4 of Newey and Smith (2004).

Now we consider the asymptotic bias. This is analyzed carefully in standard GEL case by Newey and Smith (2004). In our case

$$Bias(\hat{\beta}) = E(Q_1(\tilde{\psi}, \tilde{a}, F_0))/r_n^2.$$

This is a little bit different than the one in Newey and Smith (2004). Bias declines at rate $r_n^2 = n^{1-\kappa}, 0 < \kappa < 1$, which is much slower than the rate "n" in Newey and Smith (2004). This is clear from Theorem 4. The key question will be whether the nearly singular design causes changes in the bias of standard GEL. In standard GEL of Newey and Smith (2004), the bias term consists of optimal linear combination of $G'\Omega(\beta_0)^{-1}g(x, \beta)$ and the term arising from the estimation of second moment matrix $\Omega(\beta_0)$. In our case $\Omega(\beta_0)$ is singular, instead we use Assumption 1. Let c be an $m \times 1$ vector and

$$c_j \equiv tr[\Sigma(\lim_{n\to\infty} \frac{a_n}{n} \sum_{i=1}^{n} E\frac{\partial^2 g_{ij}(\beta_0)}{\partial\beta\partial\beta'})]/2,$$

where $g_{ij}(\beta)$ denotes the jth element of $g_i(\beta)$, $j = 1, 2 \cdots, m$.

Theorem 5. *Under Assumptions 1-6, and with (6)-(8) holding we have the following on the nullspace of* $\Omega(\beta)$

$$Bias(\hat{\beta}) = B_I + (1 + \rho_3/2) B_\Omega + o(r_n^{-2}),$$

where

$$B_I = H(-c + \lim_{n\to\infty} \frac{a_n}{n} \sum_{i=1}^{n} E G_i H g_i)/r_n^2,$$

$$B_\Omega = H(\lim_{n\to\infty} \frac{a_n}{n} \sum_{i=1}^{n} E g_i g_i' P g_i)/r_n^2.$$

Remark. This extends Theorem 4.2 of Newey and Smith (2004) to nearly singular design in GEL. There are some differences of B_I, B_Ω terms here compared with standard GEL case. The first and the most important one is : the bias terms converge to zero at a slower rate. This is because of the nearly singular design. The second difference concerns the rate of convergence of the terms in the parenthesis for the bias terms. These converge to limit slowly compared with standard case in Newey and Smith (2004). This is due to the behavior of partial derivatives of the first order conditions. The third difference is that the bias terms involve not $\Omega(\beta_0)$ but D via matrix H. The fourth difference stems from the independent rather than iid data, this is the $o(r_n^{-2})$ remainder term. However, with the nearly singular design we do not see any additional bias terms for GEL estimators. As explained in Newey and Smith (2004), B_I represents the asymptotic bias for a GMM estimator with the optimal linear combination. B_Ω arises from estimation of the second moment matrix D.

We have the special result for empirical likelihood estimator.

Corollary 2. *Under Assumptions 1-6, and (6)-(8) holding on the nullspace of* $\Omega(\beta)$

$$Bias(\hat{\beta}_{EL}) = B_I + o(r_n^{-2}).$$

This extends Corollary 4.3 in Newey and Smith (2004) to nearly singular GEL. This result is important since the bias terms converge to zero slowly in nearly singular GEL setting . EL estimator will have an advantage in this setting compared with other GEL estimators. This advantage is accelerated in nearly singular GEL setting compared to standard GEL. But any GEL estimator with $\rho_3 = -2$ or with zero third moments will share this property.

Following Theorem 5 and Corollary 2 and Theorem 8 below we see that as in Theorem 4.5 of Newey and Smith (2004) asymptotic bias of EL does not grow with the number of moment restrictions unlike GMM in some applications. Bias-corrected GEL estimators can be obtained as in Newey and Smith (2004). Since this is the same as in their case we do not include here. In nearly singular GEL we were not able to establish higher-order asymptotic efficiency of bias corrected EL over other bias corrected GEL estimators as shown in standard GEL of Newey and Smith (2004). Even if we had used iid setup and change our assumption accordingly, it is not clear how Pfanzagl and Wefelmeyer (1978) article, which establishes the third order efficiency of MLE, will work under Assumption 1.

5 Asymptotics and Higher Order Expansion for GMM

In this section we first consider the GMM estimators with nearly singular design (i.e. under Assumption 1). GMM estimators $\hat{\beta}_{GMM}$ minimize the following objective function over $B \subset R^p$.

$$S_n(\beta) = [n^{-1/2} \sum_{i=1}^{n} g_i(\beta)]' (\hat{V}(\hat{\beta}_1))^{-1} [n^{-1/2} \sum_{i=1}^{n} g_i(\beta)],$$

where

$$\hat{V}(\hat{\beta}_1) = \frac{1}{n} \sum_{i=1}^{n} [g_i(\hat{\beta}_1) - \bar{g}_1][g_i(\hat{\beta}_1) - \bar{g}_1]',$$

and $\bar{g}_1 = 1/n \sum_{i=1}^{n} g_i(\hat{\beta}_1)$, $\hat{\beta}_1$ represents the first step GMM estimator which is assumed to be consistent and faces also nearly singular design. We provide the limit theory for GMM estimators with nearly singular design.

Theorem 6. *Under Assumptions 1, 2, 4, 5 on the nullspace of* $\Omega(\beta_0)$*, the following result holds jointly with (4)*

$$\frac{n^{1/2}}{a_n^{1/2}} (\hat{\beta}_{GMM} - \beta_0) \xrightarrow{d} N(0, \Sigma).$$

Remark. Theorem 6 extends the standard GMM limit theory to the nearly singular designs for the first time in the literature. Theorem 6 implicitly assumes the first step GMM estimator faces nearly singular design.

Corollary 3. *Under Assumptions 1, 2, 4, 5 on the nullspace of* $\Omega(\beta_0)$*, the following result holds jointly with (4)*

$$n S_n(\hat{\beta}) \xrightarrow{d} \chi^2_{m-p}.$$

Remark. Corollary 3 follows from the proof of Corollary 1. In GMM test of overidentifying restrictions has the same limit even though we have a nearly singular design. Even though the J test has the same limit as in the standard GMM case, as we shall see in higher order expansions the bias term converges to zero slower than the case in standard GMM. This should affect adversely the small sample properties of this test.

We now introduce an Assumption which is useful for developing higher-order expansion.

Assumption 7. *Let initial weighting matrix in GMM be $\hat{W}$. There exists W and $\tilde{\xi}$ such that*

$$a_n(\hat{W}-W)=K_W+\frac{\tilde{\xi}}{n^{1/2}/a_n^{1/2}}+O_p(a_n n^{-1}),$$

where $\tilde{\xi}=O_p(n^{-1/2}a_n^{1/2})$, W is singular, and K_W is positive definite on the nullspace of W, $\tilde{\xi}=\frac{a_n}{n}\sum_{i=1}^n \xi(x_i), E\xi(x_i)=0$.

This Assumption is application of Assumption 1 to initial weighting matrix in GMM (near-singularity in the first step of GMM). This extends Assumption 4 in Newey and Smith (2004) to nearly singular settings in GMM.

The following Theorem is stochastic expansion for GMM under nearly singular design. This extends the standard GMM expansion in Newey and Smith (2004).

Theorem 7. *Under Assumptions 1-7 and (6)-(8) holding for the m(.) in GMM case (equation (77)), with addition to Assumption 1ii of $a_n(\hat{\Omega}(\beta_0)-\Omega(\beta_0))-D=O_p(r_n^{-1})$, we have the following expansion for two-step GMM (on the nullspace of $\Omega(\beta_0)$ and W)*

$$r_n(\hat{\beta}_{GMM}-\beta_0)=\tilde{\psi}+\frac{Q_1(\tilde{\psi},\tilde{a},F_0)}{r_n}+\frac{Q_2(\tilde{\psi},\tilde{a},\tilde{b},F_0)}{r_n^2}+O_p(r_n^{-3}),$$

where the expansion terms are defined in the proof, $r_n=n^{1/2}/a_n^{1/2}$

Remark. The assumptions are similar to Newey and Smith (2004). The additional assumption to Assumption 1ii and (6)-(8) are extra compared with Newey and Smith (2004). These are needed to get a stochastic expansion in nearly singular design case. They amount to assuming central limit theorems for the derivatives of the first-order conditions in nearly singular designs.

The following theorem provides the bias term for GMM estimators with nearly singular design.

Theorem 8. *If Assumptions 1-7 are satisfied and (6)-(8) are holding and the addition to Assumption 1ii*

$$a_n\hat{\Omega}(\beta_0)-a_n\Omega-D=O_p(r_n^{-1}),$$

is satisfied then on the nullspace of $\Omega(\beta_0)$ and W

$$Bias(\hat{\beta}_{GMM})=B_I+B_G+B_\Omega+B_W+o(r_n^{-2}),$$

where

$$B_I=H(-c+\lim_{n\to\infty}\frac{a_n}{n}\sum_{i=1}^n EG_iHg_i)/r_n^2,$$

$$B_G=-\Sigma\lim_{n\to\infty}\sum_{i=1}^n EG_i'Pg_i/r_n^2,$$

$$B_\Omega=H(\lim_{n\to\infty}\frac{a_n}{n}\sum_{i=1}^n Eg_ig_i'Pg_i)/r_n^2,$$

$$B_W = -H \sum_{j=1}^{p} \bar{\Omega}_{\beta_j} (H_W - H)' e_j / r_n^2,$$

where H_W is defined in Lemma A.6.

Remark. We need $m(.)$ term in (77) , so that we can use it in (6)-(8). This result extends Theorem 4.1 of Newey and Smith (2004) to nearly singular design for standard GMM. We see that by comparing two theorems, the bias of GMM goes to zero much slower in the nearly singular case. So this shows that in order to get meaningful results in GMM with nearly singular design we need large samples. Also we see that comparing Corollary 1 with Theorem 8 , empirical likelihood estimator has only one bias term compared to four terms in GMM (including B_I). It is possible that empirical likelihood estimator does better in terms of bias compared to GMM even in small samples.

5.1 A Peculiar Case

We now consider some interesting issue. This involves a standard first-step GMM estimator, which is not subject to nearly singular design. This can happen if we set the weight matrix to be identity in the first step.

For the second step GMM estimator, we face again nearly singular design and our model is subject to Assumption 1. So for the first-step GMM estimator,$\hat{\beta}_1$ we have

$$n^{1/2}(\hat{\beta}_1 - \beta_0) \xrightarrow{d} N[0, (G'W^{-1}G)^{-1}], \tag{9}$$

where W may be nonsingular limit weight or $W = I$. For the second-step in GMM on the nullspace of $\Omega(\beta_0)$

$$\frac{n^{1/2}}{a_n^{1/2}}(\hat{\beta}_{GMM} - \beta_0) \xrightarrow{d} N(0, \Sigma), \tag{10}$$

where $\Sigma = (G'D^{-1}G)^{-1}$. These limit results are obtained via standard GMM and nearly singular GMM decribed in Theorem 6.

Here we try to find the higher order expansion for GMM estimators in such a setup as well as the bias term. To that effect for the first step weight matrix we use the following Assumption 4 from Newey and Smith (2004). This is independent version of Assumption 4 of Newey and Smith (2004).

Assumption 8. There exists first step weight matrix W which is nonsingular and and estimator $\hat{W}$, and $\xi(x_i)$ such that $\sum_{i=1}^n \xi(x_i)/n = O_p(n^{-1/2})$,$E\xi(x_i) = 0$,

$$\hat{W} = W + \frac{\sum_{i=1}^n \xi(x_i)}{n} + O_p(n^{-1}).$$

The next theorem provides the higher-order expansion for GMM subject to nearly singular design only in the second step of estimation. Proofs for this section are in technical appendix.

Theorem 9. *Under Assumptions 1-6, and 8 , (6)-(8) are holding, with this addition to Assumption 1ii*

$$a_n(\hat{\Omega}(\beta_0) - \Omega) - D = O_p(a_n^{1/2} n^{-1/2}),$$

then we have the following on the nullspace of $\Omega(\beta_0)$

$$\begin{aligned}\frac{n^{1/2}}{a_n^{1/2}}(\hat{\beta}_{GMM} - \beta_0) &= \tilde{\psi} + \frac{Q_{11}(\tilde{\psi}, \tilde{a})}{n^{1/2} a_n^{-1/2}} + \frac{Q_{12}(\tilde{\psi}, \tilde{a}^{\omega})}{n^{1/2}} + \frac{Q_{21}(\tilde{\psi}, \tilde{a}, b^1)}{n a_n^{-1}} \\ &+ M^{-1} \frac{diag[0, \tilde{Q}_{11}^{\Omega}] \tilde{\psi}}{n a_n^{-1/2}} + M^{-1} \frac{diag[0, \tilde{Q}_{12}^{\Omega}] \tilde{\psi}}{n} + O_p(n^{-3/2} a_n^{1/2}).\end{aligned}$$

The terms of the expansion is described in the proof in detail.

Remark. Note that the rates of decay to zero is different here compared to Theorem 7. Even though the first two terms on the right hand side has the same order, the other terms mainly converge to zero at different rates. Instead of Q_1 term in Theorem 7 we have Q_{11}, Q_{12} terms and the term involving Q_{12} converge to zero rapidly than $(n^{1/2} a_n^{-1/2})$ rate. This will have a strong implication when we characterize the bias term next.

The bias is defined in the following way for this case:

$$Bias = \frac{EQ_{11}(\tilde{\psi}, \tilde{a})}{r_n^2}.$$

It can be seen that bias is defined in terms of function $Q(.)$ in the other sections. However, from the proof of Theorem 9 we see that $Q = Q_{11} + Q_{12}$ and Q_{12} converges to zero in probability whereas $Q_{11} = O_p(1)$.

Theorem 10. *If Assumptions 1-6 and Assumption 8 are satisfied and (6)-(8) are holding and the addition to Assumption 1ii*

$$a_n \hat{\Omega}(\beta_0) - a_n \Omega - D = O_p(r_n^{-1}),$$

is satisfied then on the nullspace of $\Omega(\beta_0)$

$$Bias(\hat{\beta}_{GMM}) = B_I + B_G + B_{\Omega} + o(r_n^{-2}),$$

where

$$B_I = H(-c + \lim_{n\to\infty} \frac{a_n}{n} \sum_{i=1}^{n} EG_i H g_i)/r_n^2,$$

$$B_G = -\Sigma \lim_{n\to\infty} \sum_{i=1}^{n} EG_i' P g_i / r_n^2,$$

$$B_{\Omega} = H(\lim_{n\to\infty} \frac{a_n}{n} \sum_{i=1}^{n} E g_i g_i' P g_i)/r_n^2,$$

$r_n = n^{1/2}/a_n^{1/2}$

Remark. This result is intuitive in the sense that, bias arising from the estimation of weight matrix in the first step does not play any role here, since GMM estimator in this case is not subject to nearly singular design in the first step.

6 Conclusion

This paper covers nearly singular design for GMM and GEL estimators. We derive the limiting distribution of the estimators and provide higher order expansions. The natural extension to this paper involves many instruments with nearly singular design. Another extension we are considering is the joint display of weak instruments and nearly singular design to better explain small sample properties of GMM and GEL estimators.

APPENDIX

First we provide a lemma that is useful for the proof of Theorem 1 and the consistency result.

Lemma A.1. *Under Assumptions A.2i,iii*

$$\sup_{\beta \in \mathbf{B}} |a_n n^{-1} \sum_{i=1}^{n} g_i(\beta) - E g_i(\beta)| \xrightarrow{p} 0.$$

Proof of Lemma A.1. Theorem 21.11i in Davidson (1994) shows that Assumption 2i,iii gives stochastic equicontinuity condition for Uniform Law of Large Numbers. Then again using Assumption 2iii with the stochastic equicontinuity (Theorem 21.11i of Davidson (1994)) Theorem 21.9 of Davidson (1994) provides the Uniform Law of Large Numbers for triangular arrays. **Q.E.D.**

Proof of Theorem 1.

First rewrite $\hat{\Omega}(\beta)$ as

$$\begin{aligned} a_n \hat{\Omega}(\beta) &= \frac{a_n}{n} \sum_{i=1}^{n} g_i(\beta) g_i(\beta)' \\ &= \frac{a_n}{n} \sum_{i=1}^{n} [g_i(\beta) - \bar{g} + \bar{g}][g_i(\beta) - \bar{g} + \bar{g}]' \\ &= a_n \hat{V}(\beta) + a_n \bar{g}\bar{g}', \end{aligned}$$

where $\bar{g} = 1/\sum_{i=1}^{n} g_i(\beta)$ and

$$\hat{V}(\beta) = \frac{1}{n} \sum_{i=1}^{n} [g_i(\beta) - \bar{g}][g_i(\beta) - \bar{g}]'.$$

But then via Lemma A.1

$$a_n \bar{g}\bar{g}' = o_p(1).$$

This means

$$a_n \hat{\Omega}(\beta) = a_n \hat{V}(\beta) + o_p(1). \tag{11}$$

So Assumption 1 with (11), Assumption 2i, ii satisfy finite dimensional convergence and hence satisfy conditions ii and iv of Pollard (Theorem 10.6, 1990). Assumption 2 i, iii satisfy the stochastic equicontinuity condition in Theorem 10.6 of Pollard (1990). This can be seen easily since our

Assumption 2 implies conditions i, iii,v of Theorem 10.6 of Pollard (1990) with the choice of a simple modification of pseudometric in Andrews (1994). This can also be seen by the proofs of Theorems 1 and 2 of Andrews (1994) given our Assumption 2. **Q.E.D**

To prove the consistency of the GEL estimators we follow closely the proof of consistency in Newey and Smith (2004). The notation is similar to the one used in Guggenberger and Smith (2003) since their proofs cover more general case than the iid case covered in Newey and Smith (2004).

For $n \in N$, let $B_n \subset \mathbf{B}$. Let $c_n = n^{-1/2} a_n^{1/2} \max_i \sup_{\beta \in B_n} \|g_i(\beta)\|$ and $\wedge_n = \{\lambda \in R^m : \|\lambda\| \leq n^{-1/2} a_n^{1/2} c_n^{-1/2}\}$ if $c_n \neq 0$. The case for $c_n = 0$ is trivial and without losing any generality, we will not be dealing with it as in Guggenberger and Smith (2003). Note that in our case c_n converges slowly to zero compared with Guggenberger and Smith (2003). This is because of the singularity of the variance covariance matrix of the sample moments. "uwpa1" denotes uniformly in β_n with probability approaching one. "wpa1" denotes with probability approaching one.

We need the following three lemmata to have the consistency result for $\hat{\beta}$ and the rate of convergence for $\hat{\lambda}$.

Lemma A.2. *Assume* $\max_i \sup_{\beta_n} \|g_i(\beta)\| = o_p(n^{1/2} a_n^{-1/2})$, *then*

$$\sup_{\beta \in B_n, \lambda \in \wedge_n, 1 \leq i \leq n} |\lambda' g_i(\beta)| \xrightarrow{p} 0,$$

and $\wedge_n \subset \hat{\wedge}_n(\beta)$ *uwpa1.*

Proof of Lemma A.2. First

$$c_n = n^{-1/2} a_n^{1/2} \max_i \sup_\beta \|g_i(\beta)\| = n^{(\kappa-1)/2} o_p(n^{(1-\kappa)/2}) = o_p(1),$$

for $0 < \kappa < 1$ and $a_n = n^\kappa$. Then

$$\begin{aligned} \sup |\lambda' g_i(\beta)| &\leq n^{-1/2} a_n^{1/2} c_n^{-1/2} \max_i \sup_\beta \|g_i(\beta)\| \\ &= n^{-1/2} a_n^{1/2} c_n^{-1/2} c_n n^{1/2} a_n^{-1/2} \\ &= c_n^{1/2} = o_p(1). \end{aligned}$$

This result also implies the second result of Lemma A.2.**Q.E.D**

Remark. Also $\sup_i E \sup_\beta \|g_i(\beta)\|^\xi < \infty$, for some $\xi > 2/(1-\kappa)$ implies the assumption of Lemma A.2. The proof of this assertion can be seen in equation (2.4) of Guggenberger and Smith (2003).

See that compared to Guggenberger and Smith (2003) and Newey and Smith (2004) nearly singular design warrants the existence of higher moment conditions. The existence of moment conditions are tied to the "near singularity coefficient" κ. Lemma A.2 is similar to Lemma A.2 of Newey and Smith (2004) and Lemma 7 of Guggenberger and Smith (2003).

Lemma A.3 *Assume* $\max_i \sup_\beta \|g_i(\beta)\| = o_p(n^{1/2}a_n^{-1/2})$, *for any* $\bar{\beta} \xrightarrow{p} \beta_0$ *where* $\bar{\beta} \in B_n$, $\hat{g}(\bar{\beta}) = O_p(n^{-1/2}a_n^{-1/2})$, *and Assumptions 1-3 hold then on the nullspace of* $\Omega(\beta_0)$,

$$\bar{\lambda} = \arg \max_{\lambda \in \hat{\Lambda}_n(\bar{\beta})} \hat{P}(\bar{\beta}, \lambda),$$

exists wpa1,

$$\bar{\lambda} = O_p(a_n^{1/2} n^{-1/2}),$$

and

$$\sup_{\lambda \in \hat{\Lambda}_n(\bar{\beta})} a_n \hat{P}(\bar{\beta}, \lambda) = O_p(n^{-1}a_n).$$

Remark. Lemma A.3 is similar to Lemma A.2 of Newey and Smith (2004). When we compare Lemma A.3 with the standard GEL of Newey and Smith (2004) and weakly identified GEL of Guggenberger and Smith (2003), we see that $\bar{\lambda}$ converges to zero at a slower rate than $n^{1/2}$. This happens because $\bar{\beta}$ depends on the behavior of $\hat{\Omega}(\bar{\beta})$ which has nearly singular design.

Proof of Lemma A.3. Assume $\wedge_n$ is compact since without losing any generality $c_n \neq 0$. Define $\tilde{\lambda} \in \wedge_n$ such that wpa1

$$\hat{P}(\bar{\beta}, \tilde{\lambda}) = \max_{\lambda \in \wedge_n} \hat{P}(\bar{\beta}, \lambda).$$

This holds by Assumption 3, Lemma A.2 and $\wedge_n$ being compact. Next we show that wpa1

$$\hat{P}(\bar{\beta}, \tilde{\lambda}) = \sup_{\lambda \in \hat{\Lambda}_n(\bar{\beta})} \hat{P}(\bar{\beta}, \lambda).$$

To show this last equality first we have to prove $\tilde{\lambda} \in int(\wedge_n)$ wpa1. To show this use a Taylor series expansion around $\lambda = 0$ then multiply the term by a_n and $\lambda^* \in (0, \tilde{\lambda})$

$$0 = a_n \hat{P}(\bar{\beta}, 0) \leq a_n \hat{P}(\bar{\beta}, \tilde{\lambda}) = -2a_n \tilde{\lambda}' \hat{g}(\bar{\beta}) + \tilde{\lambda}'[a_n \frac{\sum_{i=1}^n \rho_2(\lambda^{*'} g_i(\bar{\beta})) g_i(\bar{\beta}) g_i(\bar{\beta})'}{n}]\tilde{\lambda}. \tag{12}$$

Note that as in Newey and Smith (2004) $\max_i \rho_2(\lambda^{*'} g_i(\bar{\beta})) < -C_1$ wpa1, where $C_1 > 0$. Then wpa1

$$0 = a_n \hat{P}(\bar{\beta}, 0) \leq a_n \hat{P}(\bar{\beta}, \tilde{\lambda}) \leq -2a_n \tilde{\lambda}' \hat{g}(\bar{\beta}) - C_1 \tilde{\lambda}'(a_n \hat{\Omega}(\bar{\beta}))\tilde{\lambda}. \tag{13}$$

Then use Assumption 1 to have wpa1 (on the nullspace of $\Omega(\beta_0)$)

$$0 \leq a_n \hat{P}(\bar{\beta}, \tilde{\lambda}) \leq 2a_n \|\tilde{\lambda}\| \|\hat{g}(\bar{\beta})\| - C_2 \|\tilde{\lambda}\|^2, \tag{14}$$

for $C_2 > 0$. So

$$C_2/2 \|\tilde{\lambda}\| \leq a_n \|\hat{g}(\bar{\beta})\|. \tag{15}$$

Then by (15) and $\|\hat{g}(\bar{\beta})\| = O_p(n^{-1/2}a_n^{-1/2})$ we have

$$\|\tilde{\lambda}\| = O_p(n^{-1/2}a_n^{1/2}) = o_p(n^{-1/2}a_n^{1/2}c_n^{-1/2}). \tag{16}$$

So $\tilde{\lambda} \in int(\wedge_n)$ wpa1 because of (16). The first order conditions for an interior maximum hold $\partial \hat{P}(\bar{\beta}, \tilde{\lambda})/\partial \tilde{\lambda} = 0$. . Next by Lemma A.2, $\tilde{\lambda} \in \hat{\wedge}_n(\bar{\beta})$ wpa1. By concavity of $\hat{P}(\bar{\beta}, \tilde{\lambda})$ over the convex $\hat{\wedge}_n(\bar{\beta})$

$$\hat{P}(\bar{\beta}, \tilde{\lambda}) = \sup_{\lambda \in \Lambda_n(\bar{\beta})} \hat{P}(\bar{\beta}, \lambda).$$

So $\tilde{\lambda} = \bar{\lambda}$ and the first conclusion is derived. The second conclusion of the Lemma A.3 is reached by (16) and the above result. Then the last conclusion follows by applying the first equality of (16), $\hat{g}(\bar{\beta}) = O_p(n^{-1/2} a_n^{-1/2})$ to (14). **Q.E.D.**

Lemma A.4. *Under Assumptions 1 and 3 and* $\max_i \sup_\beta \|g_i(\beta)\| = o_p(n^{1/2} a_n^{-1/2})$ *then on the nullspace of* $\Omega(\beta)$

$$\|\hat{g}(\hat{\beta})\| = O_p(n^{-1/2} a_n^{-1/2}).$$

Proof of Lemma A.4. Let $\tilde{\lambda} = -n^{-1/2} a_n^{1/2} \hat{g}(\hat{\beta})/\|\hat{g}(\hat{\beta})\|$. Note that $\tilde{\lambda} \in \wedge_n$ and by Lemma A.2, $\max_i |\tilde{\lambda}' g_i(\hat{\beta})| \xrightarrow{p} 0$, so $\tilde{\lambda} \in \wedge_n(\hat{\beta})$ wpa1. Then have a second degree Taylor series expansion around $\lambda = 0$ and for $\lambda^{\cdot} \in (0, \tilde{\lambda})$

$$\hat{P}(\hat{\beta}, \tilde{\lambda}) = -2\tilde{\lambda}' \hat{g}(\hat{\beta}) + \tilde{\lambda}' [\frac{\sum_{i=1}^n \rho_2(\lambda^{\cdot\prime} g_i(\hat{\beta})) g_i(\hat{\beta}) g_i(\hat{\beta})'}{n}] \tilde{\lambda}. \tag{17}$$

First, as in Newey and Smith (2004), $\rho_2(\lambda^{\cdot\prime} g_i(\hat{\beta})) \geq -C_3$ wpa1 with $C_3 > 0$. Then multiply the term by a_n and use $\tilde{\lambda}$ definition to have wpa1

$$a_n \hat{P}(\hat{\beta}, \tilde{\lambda}) \geq a_n (2n^{-1/2} a_n^{1/2} \|\hat{g}(\hat{\beta})\|) - C_3 \tilde{\lambda}' [a_n \hat{\Omega}(\hat{\beta})] \tilde{\lambda}. \tag{18}$$

Use Assumption 1ii-iii (on the nullspace of $\Omega(\beta)$) to have the following wpa1

$$\begin{aligned} a_n \hat{P}(\hat{\beta}, \tilde{\lambda}) &\geq 2n^{-1/2} a_n^{3/2} \|\hat{g}(\hat{\beta})\| - C_4 (n^{-1/2} a_n^{1/2})^2 \\ &= 2n^{-1/2} a_n^{3/2} \|\hat{g}(\hat{\beta})\| - C_4 (a_n/n), \end{aligned} \tag{19}$$

where $C_4 > 0$. Clearly by Lemma 1, the hypotheses of Lemma A.3 are satisfied at $\bar{\beta} = \beta_0$, $\hat{g}(\beta_0) = O_p(n^{-1/2} a_n^{-1/2})$. Since $\hat{\beta}$ and $\hat{\lambda}$ are saddlepoints, and by Lemma A.3 at $\bar{\beta} = \beta_0$

$$a_n \hat{P}(\hat{\beta}, \tilde{\lambda}) \leq a_n \hat{P}(\hat{\beta}, \hat{\lambda}) \leq a_n \sup_{\lambda \in \Lambda_n(\beta_0)} \hat{P}(\beta_0, \lambda) = O_p(a_n n^{-1}). \tag{20}$$

Combine (19) and (20)

$$2n^{-1/2} a_n^{3/2} \|\hat{g}(\hat{\beta})\| \leq C_4 a_n n^{-1} + O_p(a_n n^{-1}),$$

which provides us the desired result.**Q.E.D.**

Proof of Theorem 2. By Lemma A.4

$$a_n \hat{g}(\hat{\beta}) \xrightarrow{p} 0,$$

since $a_n\hat{g}(\hat{\beta}) = O_p(a_n^{1/2}n^{-1/2}) = o_p(1)$, and by $a_n = n^\kappa, 0 < \kappa < 1$. Then use Lemma A1 (or Theorem 1 with Slutsky's Lemma) to have

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

Since $g_i(\beta)$ is continuous $\frac{a_n}{n}\sum_{i=1}^n Eg_i(\hat{\beta}) \xrightarrow{p} 0$ by triangle inequality. Note that by Assumption 4 we know that $\lim_{n\to\infty} Ea_n n^{-1}\sum_{i=1}^n g_i(\beta) = m_1(\beta)$ is zero iff $\beta = \beta_0$. So outside the neighborhood of β_0, $m_1(\beta)$ must be bounded away from zero, so $\hat{\beta} \xrightarrow{p} \beta_0$.

Next the second conclusion follows by Lemma A.4, then using the consistency and Lemma A.3 we obtain the existence of $\hat{\lambda}$ and $\hat{\lambda} = O_p(n^{-1/2}a_n^{1/2})$. **Q.E.D.**

Proof of Theorem 3. This proof is similar to the derivation of asymptotic theory of estimators in Newey and Smith (2004). The main difference is Assumption 1 here. We assume a nearly singular design in this article. This is not a trivial change since this affects central limit theorem and rate of convergence rates. The proof consists of two major steps. First step involves the existence of first order conditions wpa1. The second step is to use the first order conditions and a Taylor series expansion to derive the limit.

Benefiting from Theorem 2, Assumption 1 (nonsingular $D(\beta)$), Assumption 3, Lemma A.2 and Lemma A.3 shows that $\hat{\lambda}$ exists wpa1 and the first order conditions are satisfied wpa1.

$$\sum_{i=1}^{n} \rho_1(\hat{\lambda}'\hat{g}_i(\hat{\beta}))\hat{g}_i(\hat{\beta}) = 0, \tag{21}$$

$$\sum_{i=1}^{n} \rho_1(\hat{\lambda}'\hat{g}_i(\hat{\beta}))G_i(\hat{\beta})'\hat{\lambda} = 0, \tag{22}$$

where $G_i(\beta) = \partial g_i(\beta)/\partial\beta'$ and $g_i(\beta)$ is defined before equation (2).

Now we have a Taylor series expansion for (21) and substitute (22) together with $\hat{\theta} = (\hat{\beta}', \hat{\lambda}')'$ and $\theta_0 = (\beta_0', 0')'$, then multiply each side of the first order conditions by $a_n^{1/2}/n^{1/2}$ to have

$$0 = \begin{pmatrix} 0 \\ -n^{1/2}a_n^{1/2}\hat{g}(\beta_0) \end{pmatrix} + (a_n\bar{M})\left(\frac{n^{1/2}}{a_n^{1/2}}(\hat{\theta} - \theta_0)\right). \tag{23}$$

where

$$\bar{M} = \begin{pmatrix} 0 & \frac{1}{n}\sum_{i=1}^n \rho_1(\hat{\lambda}'g_i(\hat{\beta}))G_i(\hat{\beta})' \\ \frac{1}{n}\sum_{i=1}^n \rho_1(\bar{\lambda}'g_i(\bar{\beta}))G_i(\bar{\beta}) & \frac{1}{n}\sum_{i=1}^n \rho_2(\bar{\lambda}'g_i(\bar{\beta}))g_i(\bar{\beta})g_i(\hat{\beta})' \end{pmatrix},$$

and $\bar{\lambda} \in (0, \hat{\lambda})$ and $\bar{\beta} \in (\beta_0, \hat{\beta})$.

Then by Assumption 3, Lemma A.1, and Theorem 2

$$\rho_1(\bar{\lambda}'g_i(\bar{\beta})) \xrightarrow{p} -1, \tag{24}$$

$$\rho_2(\bar{\lambda}' g_i(\bar{\beta})) \xrightarrow{p} -1. \tag{25}$$

Use Assumption 5 in combination with (24) and Theorem 2 to have

$$\frac{a_n}{n} \sum_{i=1}^{n} \rho_1(\bar{\lambda}' g_i(\bar{\beta})) G_i(\bar{\beta}) \xrightarrow{p} -G. \tag{26}$$

Using Assumption 1, on the nullspace of $\Omega(\beta_0)$, rewrite (23)

$$0 = \begin{pmatrix} 0 \\ -n^{1/2} a_n^{1/2} \hat{g}(\beta_0) \end{pmatrix} + M \left(\frac{n^{1/2}}{a_n^{1/2}} (\hat{\theta} - \theta_0) \right),$$

where

$$M = -\begin{pmatrix} 0 & G' \\ G & D \end{pmatrix}.$$

Then we use Assumption 1 and Theorem 2 in combination with (25) and (24) to have (on the nullspace of $\Omega(\beta_0)$)

$$\begin{aligned} \frac{n^{1/2}}{a_n^{1/2}} (\hat{\theta} - \theta_0) &= M^{-1} \begin{pmatrix} 0 \\ n^{1/2} a_n^{1/2} \hat{g}(\beta_0) \end{pmatrix} + o_p(1) \\ &= -(H', P)' [n^{1/2} a_n^{1/2} \hat{g}(\beta_0)] + o_p(1), \end{aligned} \tag{27}$$

where we use

$$M^{-1} = \begin{pmatrix} -\Sigma & H \\ H' & P \end{pmatrix}.$$

Then use Lemma 1 to have the desired result.**Q.E.D**

We should note that here the limit results for estimators (the proof of Theorem 3) are holding on the nullspace of $\Omega(\beta_0)$, however in consistency proof we need the results to hold on the nullspace of $\Omega(\beta)$. Because all results of this paper depend on consistency we denote that all the statements of Theorems include "results hold on the nullpsace of $\Omega(\beta)$".

Proof of Corollary 1. First we have the Taylor series expansion, on the nullspace of $\Omega(\beta)$

$$\begin{aligned} a_n \hat{g}(\hat{\beta}) &= \frac{a_n}{n} \sum_{i=1}^{n} g_i(\hat{\beta}) = \frac{a_n}{n} \sum_{i=1}^{n} g_i(\beta_0) - \left(\frac{a_n}{n} \sum_{i=1}^{n} G_i(\beta_0) \right) (\hat{\beta} - \beta_0) + o_p(n^{-1/2} a_n^{1/2}) \\ &= a_n \hat{g}(\beta_0) - GH(a_n \hat{g}(\beta_0)) + o_p(n^{-1/2} a_n^{1/2}), \end{aligned}$$

by Assumption 5 and (27), $\hat{g}(\beta_0) = 1/n \sum_{i=1}^{n} g_i(\beta_0)$. Next using $D^{-1}(I - GH) = P$, and by (27)

$$\begin{aligned} a_n \hat{g}(\hat{\beta}) &= D P a_n \hat{g}(\beta_0) + o_p(n^{-1/2} a_n^{1/2}) \\ &= -D \hat{\lambda} + o_p(n^{-1/2} a_n^{1/2}). \end{aligned} \tag{28}$$

Furthermore a Taylor series expansion of $\hat{\lambda}$ around 0 provides (on the nullspace of $\Omega(\beta)$)

$$\begin{aligned} a_n \hat{P}(\hat{\beta}, \hat{\lambda}) &= -2\hat{\lambda}'(a_n \hat{g}(\hat{\beta})) + \hat{\lambda}'[\frac{a_n}{n}\sum_{i=1}^{n} \rho_2(\bar{\lambda}' g_i(\hat{\beta})) g_i(\hat{\beta}) g_i(\hat{\beta})']\hat{\lambda} \\ &= -2\hat{\lambda}'(a_n \hat{g}(\hat{\beta})) + \hat{\lambda}' D \hat{\lambda} + o_p(n^{-1} a_n), \end{aligned} \tag{29}$$

where we use (25) and (27) and Assumption 1.

Next use (28) in (29) to have

$$a_n \hat{P}(\hat{\beta}, \hat{\lambda}) = [a_n \hat{g}(\hat{\beta})]' D^{-1} [a_n \hat{g}(\hat{\beta})] + o_p(n^{-1} a_n).$$

Using the above equation and expressing the objective function in the following way

$$n \hat{P}(\hat{\beta}, \hat{\lambda}) = [n^{1/2} a_n^{1/2} \hat{g}(\hat{\beta})]' D^{-1} [n^{1/2} a_n^{1/2} \hat{g}(\hat{\beta})] + o_p(1). \tag{30}$$

Note that by using Taylor series expansion

$$\begin{aligned} n^{1/2} a_n^{1/2} \hat{g}(\hat{\beta}) = \frac{a_n^{1/2}}{n^{1/2}} \sum_{i=1}^{n} g_i(\hat{\beta}) &= \frac{a_n^{1/2}}{n^{1/2}} \sum_{i=1}^{n} g_i(\beta_0) - \left(\frac{a_n}{n} \sum_{i=1}^{n} G_i(\beta_0) \right) \frac{n^{1/2}}{a_n^{1/2}} (\hat{\beta} - \beta_0) + o_p(n^{-1/2} a_n^{1/2}) \\ &\xrightarrow{d} N(0, D - G\Sigma G'), \end{aligned} \tag{31}$$

where we use Lemma 1, (27) and Assumption 5 in the last step. Then combine (30)-(31) to have

$$n \hat{P}(\hat{\beta}, \hat{\lambda}) \xrightarrow{d} \chi^2_{m-p}.$$

Q.E.D

The following Lemma extends Lemma A.4 of Newey and Smith (2004) to nearly singular designs. The proof is similar but takes into account nearly singular design for the variance-covariance function for the vector function for first order conditions and its partial derivatives. We need to introduce some notation that will be necessary to understand the results and the proofs. First let F denote the distribution of the data x, $\psi(x, F)$ a function of x and F with $E\psi(x, F_0) = 0$, and $\tilde{\psi} = \frac{a_n^{1/2}}{n^{1/2}} \sum_{i=1}^{n} \psi(x_i)$. We suppress the F argument. The assumptions c and d can be obtained under more primitive conditions and these are discussed immediately below the proof. This is a general result than for GEL estimator, then we supply the specific result in Theorem 4.

Lemma A.5. *Suppose that the estimator $\hat{\theta}$ is a $q \times 1$ vector and vector of functions $m(x_i, \theta)$ satisfy*

(a).

$$\hat{\theta} = \theta_0 + O_p(n^{-1/2} a_n^{1/2});$$

(b).

$$\hat{m}(\hat{\theta}) = \frac{a_n}{n} \sum_{i=1}^{n} m(x_i, \hat{\theta}) = 0, wpa1;$$

(c). $m(x_i,\theta)$ is three times continuously differentiable on T_n where

$$T_n = \{\theta : \|\theta-\theta_0\| \le n^{-1/2}a_n^{1/2}c_n^{-1/2}\}$$

wpa1. For $\theta \in T_n$,

$$\|\frac{\partial^3 m(x_i,\theta)}{\partial\theta_j\partial\theta_k\partial\theta_l} - \frac{\partial^3 m(x_i,\theta_0)}{\partial\theta_j\partial\theta_k\partial\theta_l}\| \le d(x_i)\|\theta-\theta_0\|$$

where

$$\lim_{n\to\infty}\frac{a_n}{n}\sum_{i=1}^n Ed(x_i)^{2+\delta} < \infty,$$

for some $\delta > 0$;

(d). First we need the following notation, for $j,k = 1,2,\cdots q$,$i = 1,\cdots,n$,

$$\hat{M}(\theta) = \frac{a_n}{n}\sum_{i=1}^n \frac{\partial m(x_i,\theta)}{\partial\theta'},$$

$$M_j = \lim_{n\to\infty} E\frac{a_n}{n}\sum_{i=1}^n \frac{\partial^2 m(x_i,\theta_0)}{\partial\theta_j\partial\theta'}$$

$$M_{jk} = \lim_{n\to\infty} E\frac{a_n}{n}\sum_{i=1}^n \frac{\partial^3 m(x_i,\theta_0)}{\partial\theta_k\partial\theta_j\partial\theta'}$$

$$A(x_i) = \frac{\partial m(x_i,\theta_0)}{\partial\theta'} - E\frac{\partial m(x_i,\theta_0)}{\partial\theta'},$$

$$B_j(x_i) = \frac{\partial^2 m(x_i,\theta_0)}{\partial\theta_j\partial\theta'} - E\frac{\partial^2 m(x_i,\theta_0)}{\partial\theta_j\partial\theta'},$$

and $a(x_i) = vecA(x_i), b(x_i) = vec[B_1(x_i),\cdots,B_q(x_i)]$, $\tilde{A} = \frac{a_n^{1/2}}{n^{1/2}}\sum_{i=1}^n A(x_i)$, $\tilde{B}_j = \frac{a_n^{1/2}}{n^{1/2}}\sum_{i=1}^n B_j(x_i)$, $\tilde{a} = vec\tilde{A}, \tilde{b}_j = vec\tilde{B}_j$.

We state Assumption d as follows

(i).

$$\|\frac{\partial^2\hat{M}(\theta_0)}{\partial\theta_j\partial\theta_k} - M_{jk}\| = O_p(n^{-1/2}a_n^{1/2}), forj,k = 1,2,\cdots q$$

(ii). $\tilde{a} = O_p(1), \tilde{b}_j = O_p(1)$, $j = 1,\cdots q$,

(iii). $\frac{a_n^{1/2}}{n^{1/2}}\sum_{i=1}^n m(x_i,\theta_0) = O_p(1)$;

(e). $Em(x_i,\theta_0) = 0$, $\hat{M}(\theta_0) \xrightarrow{p} M$, and M exists and nonsingular.

Then under Assumptions a-e, we have the following expansion on the nullspace of $\Omega(\beta)$

$$\frac{n^{1/2}}{a_n^{1/2}}(\hat{\theta}-\theta_0) = \tilde{\psi} + Q_1(\tilde{\psi},\tilde{a})/(n^{1/2}/a_n^{1/2}) + Q_2(\tilde{\psi},\tilde{a},\tilde{b})/(n/a_n) + O_p\left(\frac{1}{(n/a_n)^{3/2}}\right),$$

where $\psi(x_i) = -M^{-1}m(x_i,\theta_0)$ and $\tilde{\psi} = \frac{a_n^{1/2}}{n^{1/2}}\sum_{i=1}^n \psi(x_i)$,

$$Q_1(\tilde{\psi},\tilde{a}) = -M^{-1}[\tilde{A}\tilde{\psi} + \sum_{j=1}^q \tilde{\psi}_j M_j\tilde{\psi}/2,]$$

$$
\begin{aligned}
Q_2(\tilde{\psi},\tilde{a},\tilde{b}) &= -M^{-1}[\tilde{A}Q_1(\tilde{\psi},\tilde{a}) + \sum_{j=1}^{q}\{\tilde{\psi}_j M_j Q_1(\tilde{\psi},\tilde{a}) + Q_{1j}(\tilde{\psi},\tilde{a})M_j\tilde{\psi} + \tilde{\psi}_j\tilde{B}_j\tilde{\psi}\}/2 \\
&+ \sum_{j,k=1}^{q}\tilde{\psi}_j\tilde{\psi}_k M_{jk}\tilde{\psi}/6].
\end{aligned}
$$

Remark. The difference in the Lemma compared to Lemma A.4 of Newey and Smith (2004) is the rate of convergence. There in their article the rate of convergence was $n^{1/2}$ and in this article due to the the near singularity this is $n^{1/2}/a_n^{1/2} = n^{(1-\kappa)/2}, 0 < \kappa < 1$. So we need large samples in this setup. Assumptions a and b are similar to the Assumptions a-b of Lemma A.4 in Newey and Smith (2004). The difference is the rate of convergence in a and multiplication of the first order conditions by a_n/n in this article compared to $1/n$ in Newey and Smith (2004). Assumption d provides high-level conditions but this saves a lot of notation. For example instead of providing $\tilde{a} = O_p(1)$ in Assumption d, we could have written instead that

$$
a_n\left[\frac{\sum_{i=1}^n a(x_i)a(x_i)'}{n} - \Omega_A\right] \xrightarrow{p} D_A,
$$

where $\Omega_A = \lim_{n\to\infty} E\frac{\sum_{i=1}^n a(x_i)a(x_i)'}{n}$ is singular and D_A is positive definite on the nullspace of Ω_A. Further assume that $\sup_i E\|a(x_i)\|^{\xi} < \infty$, and $a(x_i)$ being independent. These three conditions provide primitive assumptions for $\tilde{a} = O_p(1)$. This can be seen by applying Lemma 1 to the primitives.

Similar primitive conditions may be provided for Assumptions d-i, d-iii and $\tilde{B}_j = O_p(1)$.

Proof of Lemma A.5. First take a Taylor series expansion around θ_0 for $\hat{m}(\hat{\theta})$

$$
\begin{aligned}
\hat{m}(\hat{\theta}) &= \hat{m}(\theta_0) + \hat{M}(\theta_0)(\hat{\theta}-\theta_0) + \sum_{j=1}^{q}(\hat{\theta}_j - \theta_{j0})\left[\frac{\partial \hat{M}(\theta_0)}{\partial\theta_j}\right](\hat{\theta}-\theta_0)/2 \\
&+ \sum_{j,k=1}^{q}(\hat{\theta}_j-\theta_{j0})(\hat{\theta}_k-\theta_{k0})\left[\frac{\partial^2\hat{M}(\bar{\theta})}{\partial\theta_k\partial\theta_j}\right](\hat{\theta}-\theta_0)/6, \qquad (32)
\end{aligned}
$$

where $\bar{\theta} \in (\theta_0, \hat{\theta})$. Then we consider the fourth term on the right-hand side of (32). By triangle inequality

$$
\|\frac{\partial^2\hat{M}(\bar{\theta})}{\partial\theta_k\partial\theta_j} - M_{jk}\| \le \|\frac{\partial^2\hat{M}(\bar{\theta})}{\partial\theta_k\partial\theta_j} - \frac{\partial^2\hat{M}(\theta_0)}{\partial\theta_k\partial\theta_j}\| + \|\frac{\partial^2\hat{M}(\theta_0)}{\partial\theta_k\partial\theta_j} - M_{jk}\|. \qquad (33)
$$

In (33) apply Assumption c to the first right hand side term and apply Assumption d to the second right-hand side term to have

$$
\|\frac{\partial^2\hat{M}(\bar{\theta})}{\partial\theta_k\partial\theta_j} - M_{jk}\| \le \left[\frac{a_n}{n}\sum_{i=1}^{n} d(x_i)\right]\|\bar{\theta}-\theta_0\| + O_p(n^{-1/2}a_n^{1/2}). \qquad (34)
$$

Then apply Assumption c and use Lemma A.1 proof to have $\left[\frac{a_n}{n}\sum_{i=1}^{n} d(x_i)\right] \xrightarrow{p} D_x$ where $D_x = \lim_{n\to\infty} E\frac{a_n}{n}\sum_{i=1}^{n} d(x_i)$ which is finite. Then benefit from Assumption a to have

$$\|\frac{\partial^2 \hat{M}(\bar{\theta})}{\partial\theta_k\partial\theta_j} - M_{jk}\| = O_p(n^{-1/2}a_n^{1/2}). \tag{35}$$

Use Assumption b with (35) and Assumption a to rewrite (32)

$$\begin{aligned} 0 &= \hat{m}(\theta_0) + [\hat{M}(\theta_0) - M + M](\hat{\theta} - \theta_0) \\ &+ \sum_{j=1}^{q}(\hat{\theta}_j - \theta_{j0})\left[\frac{\partial \hat{M}(\theta_0)}{\partial\theta_j} - M_j + M_j\right](\hat{\theta} - \theta_0)/2 \\ &+ \sum_{j,k=1}^{q}(\hat{\theta}_k - \theta_{k0})(\hat{\theta}_j - \theta_{j0})M_{jk}(\hat{\theta} - \theta_0)/6 + O_p(n^{-2}a_n^2). \end{aligned} \tag{36}$$

Now rewrite (36) using the definitions of $\tilde{\psi}, \tilde{A}, \tilde{B}_j$, Assumption e and benefiting from the first and second right hand side terms in (36)

$$\begin{aligned} (\hat{\theta} - \theta_0) &= \frac{a_n^{1/2}\tilde{\psi}}{n^{1/2}} - M^{-1}[\frac{a_n^{1/2}\tilde{A}(\hat{\theta} - \theta_0)}{n^{1/2}} \\ &+ \sum_{j=1}^{q}(\hat{\theta}_j - \theta_{j0})M_j(\hat{\theta} - \theta_0)/2 + \sum_{j=1}^{q}(\hat{\theta}_j - \theta_{j0})\frac{a_n^{1/2}\tilde{B}_j}{n^{1/2}}(\hat{\theta} - \theta_0)/2 \\ &+ \sum_{j,k=1}^{q}(\hat{\theta}_j - \theta_{j0})(\hat{\theta}_k - \theta_{k0})M_{jk}(\hat{\theta} - \theta_0)/6] + O_p(\frac{a_n^2}{n^2}). \end{aligned} \tag{37}$$

We consider the right-hand side terms in (37). First by Assumption d and a

$$\frac{a_n^{1/2}\tilde{A}(\hat{\theta} - \theta_0)}{n^{1/2}} = O_p(\frac{a_n}{n}). \tag{38}$$

By Assumption a

$$\sum_{j=1}^{q}(\hat{\theta}_j - \theta_{j0})M_j(\hat{\theta} - \theta_0)/2 = O_p(\frac{a_n}{n}). \tag{39}$$

By Assumptions a and d

$$\sum_{j=1}^{q}(\hat{\theta}_j - \theta_{j0})\frac{a_n^{1/2}\tilde{B}_j}{n^{1/2}}(\hat{\theta} - \theta_0)/2 = O_p(\frac{a_n^{3/2}}{n^{3/2}}). \tag{40}$$

Also by Assumption a

$$\sum_{j,k=1}^{q}(\hat{\theta}_j - \theta_{j0})(\hat{\theta}_k - \theta_{k0})M_{jk}(\hat{\theta} - \theta_0)/6 = O_p(\frac{a_n^{3/2}}{n^{3/2}}). \tag{41}$$

Then use (38)-(41) in (37) to have

$$(\hat{\theta} - \theta_0) = \frac{a_n^{1/2}\tilde{\psi}}{n^{1/2}} + O_p(\frac{a_n}{n}). \tag{42}$$

Then see that the last three terms (including the remainder term) on the right hand side of (37) is of order $O_p(\frac{a_n^{3/2}}{n^{3/2}})$ (specifically by (40)-(41)). So we replace $\hat{\theta} - \theta_0$ by $a_n^{1/2}\tilde{\psi}/n^{1/2}$ in the second and third right hand side terms of (37) (this also generates an error of $O_p(\frac{a_n^{3/2}}{n^{3/2}})$) to have

$$\begin{aligned}(\hat{\theta} - \theta_0) &= \frac{a_n^{1/2}\tilde{\psi}}{n^{1/2}} - M^{-1}[\frac{a_n \tilde{A}\tilde{\psi}}{n} + \frac{a_n \sum_{j=1}^q \tilde{\psi}_j M_j \tilde{\psi}}{2n}] + O_p(\frac{a_n^{3/2}}{n^{3/2}}) \\ &= \frac{a_n^{1/2}\tilde{\psi}}{n^{1/2}} + a_n Q_1(\tilde{\psi}, \tilde{a})/n + O_p(a_n^{3/2} n^{-3/2}),\end{aligned}$$

where $Q_1(.)$ is defined at the beginning of Lemma A.5.

Then replace as in Newey and Smith (2004) $\hat{\theta} - \theta_0$ in the second and third terms of the equation (37) by $a_n^{1/2}\tilde{\psi}/n^{1/2} + a_n Q_1(\tilde{\psi}, \tilde{a})/n$ and in the fourth and fifth terms replace $\hat{\theta} - \theta_0$ by $a_n^{1/2}\tilde{\psi}/n^{1/2}$ to have the desired result.**Q.E.D.**

Proof of Theorem 4. This is similar to the proof of Theorem 3.4 in Newey and Smith (2004). We have to verify assumptions of Lemma A.5 for the GEL case. Let $\theta = (\beta', \lambda')'$, $\theta_0 = (\beta_0', 0')'$, $\hat{\theta}$ is the GEL estimator. Then use (5) and we see that by Theorem 3, $\hat{\theta} = \theta_0 + O_p(n^{-1/2} a_n^{1/2})$ and by first order conditions (21)-(22) we have

$$\frac{a_n}{n} \sum_{i=1}^n m(x_i, \hat{\theta}) = 0,$$

wpa1 since $a_n = n^\kappa, 0 < \kappa < 1$. So Assumptions a and b of Lemma A.5 are easily satisfied in GEL with nearly singular design. Next we want to verify assumption c in Lemma A.5 for the GEL estimators with nearly singular design. By Lemma A.2 and using Assumption 6 and since $\rho_1(\nu)$ is three times continuously differentiable on T_n of Lemma A.5, following equation (A.17) of Newey and Smith (2004)

$$\|\frac{\partial^3 m(x_i, \theta)}{\partial\theta_k \partial\theta_j \partial\theta} - \frac{\partial^3 m(x_i, \theta_0)}{\partial\theta_k \partial\theta_j \partial\theta}\| \leq (Cl(x_i)^5)\|\theta - \theta_0\|. \tag{43}$$

We obtain (43) under Assumption 6. The only difference with the Assumption used in Newey and Smith (2004) is they use iid DGP and have $El(x_i)^5 < \infty$. In our case we need $\lim a_n/n \sum_{i=1}^n El(x_i)^{10+5\delta} < \infty$ in our Assumption 6. So Assumption c of Lemma A.5 is verified for GEL with nonsingular design, when $d(x_i) = l(x_i)^5$.

First of all, to verify Assumption d of Lemma A.5 in GEL case, we set the first order conditions (21)- (22) evaluated at θ_0 and to save from unecessary notation setting $g_i(\beta_0) = g_i$,

$$m(x_i, \theta_0) = -\begin{pmatrix} 0 \\ g_i \end{pmatrix}. \tag{44}$$

Then Assumption diii is satisfied by Lemma 1. To verify Assumption dii, we set up

$$A(x_i) = -\begin{pmatrix} 0 & G_i' - EG_i' \\ G_i - EG_i & g_i g_i' - Eg_i g_i' \end{pmatrix}, \tag{45}$$

which is equation (A.18) of Newey and Smith (2004), and can be obtained from (21)(22) easily applying $A(x_i)$ definition in Lemma A.5 with (5).

So Assumption dii is

$$\frac{a_n^{1/2}}{n^{1/2}} \sum_{i=1}^{n} vec(G_i - EG_i) = O_p(1), \tag{46}$$

$$\frac{a_n^{1/2}}{n^{1/2}} \sum_{i=1}^{n} vec(g_i g_i' - Eg_i g_i') = O_p(1). \tag{47}$$

First of all in the case of iid, nonsingular GEL of Newey and Smith (2004) (46)(47) amounts to moment bounds in partial derivatives of g_i. This is Assumption d of Lemma A.4 in Newey and Smith (2004). Our case is nonstandard and we consider nearly singular design with independent g_i. So to derive (46)(47) from primitive assumptions we need Assumption 6 as well as the near-singularity of the variance covariance for G_i and $g_i g_i'$ like Assumption 1. Since these become cumbersome we chose instead to keep (46), (47) and this is basically (6). To have $\tilde{b}_j = O_p(1)$ for $j = 1, \cdots, p, p+1, \cdots, p+m$, we specify by (A.19) of Newey and Smith (2004), for $j \le p$,

$$\frac{\partial^2 m(x_i, \theta_0)}{\partial \theta_j \partial \theta} = -\begin{pmatrix} 0 & G_i^{j'} \\ G_i^j & g_i^j g_i' + g_i g_i^{j'} \end{pmatrix}, \tag{48}$$

where $G_i^j = \partial^2 g_i(\beta_0)/\partial \beta_j \partial \beta$ and $g_i^j = \partial g_i(\beta_0)/\beta_j$. For $j > p$, $t = j - p$ and t denoting the t th element of a vector, e_t denotes the t th unit vector,

$$\frac{\partial^2 m(x_i, \theta_0)}{\partial \theta_j \partial \theta} = -\begin{bmatrix} -\partial^2 [e_t' g_i]/\partial \beta \partial \beta' & G_i' e_t g_i' + g_{it} G_i' \\ g_i e_t' G_i + G_i g_{it} & -\rho_3 g_{it} g_i g_i' \end{bmatrix}. \tag{49}$$

Then use definition of $B_j(x_i)$ and impose (7) to have $\tilde{b}_j = O_p(1)$,for $j = 1, 2, \cdots, p, p+1, \cdots, p+m$. So Assumption dii in Lemma A.5 is satisfied with (6)-(7) given the form of the second-order partial derivatives above. For verifying Assumption di in Lemma A.5 for GEL we need to write the third order partial derivative of $m(x_i, \theta_0)$ with either using first order conditions (21)(22) or by (A.20)-(A.22) of Newey and Smith (2004). Set $G_i^{jk} = \partial^3 g_i(\beta_0)/\partial \beta_k \partial \beta_j \partial \beta$ and $g_i^{jk} = \partial^2 g_i(\beta_0)/\partial \beta_k \partial \beta_j$, for $j = 1, \cdots, p, p+1, \cdots, p+m$. For the derivatives corresponding to β with $j \le p, k \le p$,

$$\frac{\partial^3 m(x_i, \theta_0)}{\partial \theta_k \partial \theta_j \partial \theta} = -\begin{pmatrix} 0 & G_i^{jk'} \\ G_i^{jk} & g_i^{jk} g_i' + g_i^j g_i^{k'} + g_i^k g_i^{j'} + g_i g_i^{jk} \end{pmatrix}$$

For the cross partial derivative between λ_t and β_j with $j \leq p, k > p, t = k - p$,

$$\frac{\partial^3 m(x_i,\theta_0)}{\partial\theta_k\partial\theta_j\partial\theta} = -\begin{pmatrix} \partial^3 g_{it}(\beta_0)/\partial\beta_j\partial\beta\partial\beta' & G_i' e_t g_i^{j'} + G_i^{j'} e_t g_i' + G_{itj} G_i' + g_{it} G_i^{j'} \\ g_i^j e_t G_i + g_i e_t G_i^j + G_{itj} G_i + g_{it} G_i^j & -\rho_3 (G_{itj} g_i g_i' + g_{it}(g_i^j g_i + g_i g_i^{j'})) \end{pmatrix}$$

For the partial derivative between λ_t and λ_u with $j > p, k > p, t = j - p, u = k - p$:

$$\begin{aligned} \frac{\partial^3 m(x_i,\theta_0)}{\partial\theta_k\partial\theta_j\partial\theta} &= \begin{pmatrix} -G_i' e_t e_u' G_i - G_i' e_u e_t' G_i & \rho_3 (g_{it} G_i' e_u + g_{iu} G_i' e_t) g_i' \\ \rho_3 g_i (g_{it} e_u' G_i + g_{iu} e_t' G_i) & \rho_4 g_{it} g_{iu} g_i g_i' \end{pmatrix} \\ &- \begin{pmatrix} g_{it}\partial^2 g_{iu}(\beta_0)/\partial\beta\partial\beta' + g_{iu}\partial^2 g_{it}(\beta_0)/\partial\beta\partial\beta' & -\rho_3 g_{it} g_{iu} G_i' \\ -\rho_3 g_{it} g_{iu} G_i & 0 \end{pmatrix} \end{aligned}$$

Use these third order partial derivatives and substitute them into definition of M_{jk} and $\partial^2 \hat{M}(\theta_0)/\partial\theta_j\partial\theta_k$. Then Assumption di is verified for GEL case. Note that we do not provide the primitive conditions for Assumption di but rather we provide instead (8) with the definitions of the third order derivatives used above in GEL.

To apply Assumption e of Lemma A.5 to nearly singular GEL note that

$$Em(x_i,\theta_0) = -\begin{pmatrix} 0 \\ Eg_i(\beta_0) \end{pmatrix} = 0$$

and since

$$\frac{\partial m(x_i,\theta_0)}{\partial\theta'} = -\begin{pmatrix} 0 & G_i' \\ G_i & g_i g_i' \end{pmatrix},$$

we have

$$M = -\begin{pmatrix} 0 & G' \\ G & D \end{pmatrix}, \tag{50}$$

using Assumptions 1 and 5. Then since all the Assumptions of Lemma A.5 are satisfied by the nearly singular GEL, stochastic expansion given in Lemma A.5 is satisfied. From the definition of $\tilde{\psi}$ in the statement of Lemma A.5 we have

$$\tilde{\psi} = -M^{-1} \frac{\sum_{i=1}^n m(x_i,\theta_0)}{r_n}, \tag{51}$$

where $r_n = a_n^{1/2}/n^{1/2}$. For GEL estimators with nearly singular design, then substitute (44) and (50) into (51) to have $\tilde{\psi}$. Q_1, Q_2 is given in the statement of Lemma A.5, and can be obtained by the expressions used in this proof and using $r_n = a_n^{1/2}/n^{1/2}$.**Q.E.D**

Proof of Theorem 5. We first derive an alternative expression for $Q_1(\tilde{\psi}, \tilde{a}, F_0)$ using Assumption 1. In Lemma A.5

$$Q_1(\tilde{\psi}, \tilde{a}, F_0) = -M^{-1}[\tilde{A}\tilde{\psi} + \sum_{j=1}^q \tilde{\psi}_j M_j \tilde{\psi}/2].$$

We analyze the following term

$$\sum_{j=1}^{q} \tilde{\psi}_j M_j \tilde{\psi}/2 = \sum_{j=1}^{q} M_j [\tilde{\psi}\tilde{\psi}]' e_j/2.$$

Next see that by independence of $\psi(x_i)$ and Assumption 2

$$\tilde{\psi}\tilde{\psi}' = \frac{a_n}{n}[\sum_{i=1}^{n} \psi(x_i)][\sum_{i=1}^{n} \psi(x_i)]' = \frac{a_n}{n}\sum_{i=1}^{n} \psi(x_i)\psi(x_i)' + o_p(1). \tag{52}$$

Then by Assumption 1 and $\psi(x_i)$ definition (on the nullspace of $\Omega(\beta_0)$)

$$\begin{aligned} \frac{a_n}{n}\sum_{i=1}^{n} \psi(x_i)\psi(x_i)' &= \begin{bmatrix} H(\frac{a_n}{n}\sum_{i=1}^{n} g_i g_i')H & H(\frac{a_n}{n}\sum_{i=1}^{n} g_i g_i')P \\ P(\frac{a_n}{n}\sum_{i=1}^{n} g_i g_i')H & P(\frac{a_n}{n}\sum_{i=1}^{n} g_i g_i')P \end{bmatrix} \\ &\xrightarrow{p} \begin{bmatrix} HDH & HDP \\ PDH & PDP \end{bmatrix} = \begin{bmatrix} \Sigma & 0 \\ 0 & P \end{bmatrix}, \end{aligned} \tag{53}$$

by $HDP = 0, HDH' = \Sigma, PDP = P$. So by adding and subtracting and by (53)-(52)

$$\sum_{j=1}^{q} \tilde{\psi}_j M_j \tilde{\psi}/2 = \sum_{j=1}^{q} M_j diag(\Sigma, P)' e_j/2 + o_p(1). \tag{54}$$

Now we can express

$$Q_1(\tilde{\psi}, \tilde{a}, F_0) = -M^{-1}[\tilde{A}\tilde{\psi} + \sum_{j=1}^{q} M_j diag(\Sigma, P)' e_j/2] + o_p(1).$$

So ignoring the asymptotically negligible term

$$EQ_1(\tilde{\psi}, \tilde{a}, F_0) = -M^{-1}E[\tilde{A}\tilde{\psi} + \sum_{j=1}^{q} M_j diag(\Sigma, P)' e_j/2]. \tag{55}$$

In this proof we want to specify these terms for GEL estimators. Let $g_{ij}(\beta)$ denote the jth element of $g_i(\beta)$ and e_j the jth unit vector. We benefit from Theorem 4 and the proof of Theorem 4. In our case by independence of x_i

$$E\tilde{A}\tilde{\psi} = E\left[\frac{a_n}{n}\sum_{i=1}^{n} A(x_i)\psi(x_i)\right],$$

and $\psi(x_i) = -[H', P]' g_i$,$H = \Sigma G' D^{-1}$, $A(x_i)$ is described in (45). Then use (45) with the definition of $\psi(x_i)$

$$E\tilde{A}\tilde{\psi} = \begin{pmatrix} \frac{a_n}{n}\sum_{i=1}^{n} E[G_i' P g_i] \\ \frac{a_n}{n}\sum_{i=1}^{n} E[G_i H g_i + g_i g_i' P g_i] \end{pmatrix}. \tag{56}$$

Usung the definition of M_j in Lemma A.5 and (48)-(49) we specify M_j in our case, for $j = 1, 2\cdots, p$ (for β)

$$M_j = -\begin{bmatrix} 0 & (\lim_{n\to\infty} E\frac{a_n}{n}\sum_{i=1}^n G_i^{j'}) \\ (\lim_{n\to\infty} E\frac{a_n}{n}\sum_{i=1}^n G_i^j) & (\lim_{n\to\infty} E\frac{a_n}{n}\sum_{i=1}^n g_i^j g_i' + g_i g_i^{j'}) \end{bmatrix}. \tag{57}$$

For $j = 1, 2\cdots, m$ (for λ)

$$M_{j+p} = -\begin{bmatrix} \lim_{n\to\infty} E\frac{a_n}{n}\sum_{i=1}^n \partial^2[e_j g_i(\beta_0)]/\partial\beta\partial\beta' & \lim_{n\to\infty} E\frac{a_n}{n}\sum_{i=1}^n G_i' e_j g_i' + g_{ij}' G_i' \\ \lim_{n\to\infty} E\frac{a_n}{n}\sum_{i=1}^n g_i e_j' G_i + G_i g_{ij} & -\rho_3 \lim_{n\to\infty} E\frac{a_n}{n}\sum_{i=1}^n g_{ij} g_i g_i' \end{bmatrix}. \tag{58}$$

Let us suppose we use (54)

$$\sum_{j=1}^q M_j diag(\Sigma, P) e_j/2 = \sum_{j=1}^p M_j [\Sigma, 0]' e_j/2 + \sum_{j=1}^m M_{j+p}[0, P]' e_j/2. \tag{59}$$

For the first term on the right-hand side of (59) by (57)

$$\sum_{j=1}^p M_j[\Sigma, 0]' e_j/2 = -\sum_{j=1}^p \begin{pmatrix} 0 \\ \lim_{n\to\infty} \frac{a_n}{n}\sum_{i=1}^n EG_i^j \Sigma e_j/2 \end{pmatrix}. \tag{60}$$

Then consider the second term on the right hand side of (59) by (58)

$$\sum_{j=1}^m M_{j+p}[0, P]' e_j/2 = -\sum_{j=1}^m \begin{bmatrix} (\lim_{n\to\infty} \frac{a_n}{n}\sum_{i=1}^n E[G_i' e_j g_i' + g_{ij} G_i']) P e_j/2 \\ -\rho_3 \left(\lim_{n\to\infty} \frac{a_n}{n}\sum_{i=1}^n E[g_{ij} g_i g_i']\right) P e_j/2 \end{bmatrix}. \tag{61}$$

We can simplify (61) by noting $\sum_{j=1}^m G_i' e_j g_i' P e_j = \sum_{j=1}^m G_i' e_j e_j' P g_i = G_i' P g_i$ and $\sum_{j=1}^m g_{ij} G_i' P e_j = \sum_{j=1}^m G_i' P e_j g_{ij} = G_i' P g_i$. So (61) can be rewritten

$$\begin{bmatrix} -\lim_{n\to\infty} \frac{a_n}{n}\sum_{i=1}^n EG_i' P g_i \\ \rho_3/2 \lim_{n\to\infty} \frac{a_n}{n}\sum_{i=1}^n E g_i g_i' P g_i \end{bmatrix}. \tag{62}$$

Now combine (62) (60) into (55) to have

$$E\sum_{j=1}^q \tilde{\psi}_j M_j \tilde{\psi}/2 = \begin{pmatrix} -\lim_{n\to\infty} \frac{a_n}{n}\sum_{i=1}^n EG_i' P g_i \\ (\rho_3/2)\lim_{n\to\infty} \frac{a_n}{n}\sum_{i=1}^n E g_i g_i' P g_i - \sum_{j=1}^p \lim_{n\to\infty} \frac{a_n}{n}\sum_{i=1}^n EG_i^j \Sigma e_j/2 \end{pmatrix}. \tag{63}$$

These are finite terms by Assumptions 2 and 6 and (7).

Now we can add and subtract to simplify (56)

$$\begin{aligned} E\tilde{A}\tilde{\psi} &= \begin{pmatrix} \lim_{n\to\infty} \frac{a_n}{n}\sum_{i=1}^n EG_i' P g_i \\ \lim_{n\to\infty} \frac{a_n}{n}\sum_{i=1}^n E[G_i H g_i + g_i g_i' P g_i] \end{pmatrix} \\ &+ \begin{pmatrix} \frac{a_n}{n}\sum_{i=1}^n EG_i' P g_i - \lim_{n\to\infty} \frac{a_n}{n}\sum_{i=1}^n EG_i' P g_i \\ \sum_{i=1}^n E[G_i H g_i + g_i g_i' P g_i] - \lim_{n\to\infty} \frac{a_n}{n}\sum_{i=1}^n E[G_i H g_i + g_i g_i' P g_i] \end{pmatrix} \\ &= \begin{pmatrix} \lim_{n\to\infty} \frac{a_n}{n}\sum_{i=1}^n EG_i' P g_i \\ \lim_{n\to\infty} \frac{a_n}{n}\sum_{i=1}^n E[G_i H g_i + g_i g_i' P g_i] \end{pmatrix} + o(1), \end{aligned} \tag{64}$$

by (6), Assumptions 2 and 6.

Then add (63)(64) to have

$$EQ_1(\tilde{\psi}, \tilde{a}, F_0) = -M^{-1}\left\{\left[\begin{array}{c} 0 \\ -c + (\lim_{n\to\infty} \frac{a_n}{n}\sum_{i=1}^n EG_iHg_i) + (1+\rho_3/2)(\lim_{n\to\infty}\frac{a_n}{n}\sum_{i=1}^n Eg_ig_i'Pg_i) \end{array}\right] + o(1)\right\}.$$

Use $[I_p, 0]M^{-1} = [\Sigma, -H]$, and the previous equation to have the result in Theorem 5. **Q.E.D.**

Proof of Theorem 6. Rewrite $S_n(\beta)$ by multiplying and dividing by a_n

$$S_n(\beta) = [\frac{a_n^{1/2}}{n^{1/2}}\sum_{i=1}^n g_i(\beta)]'[a_n\hat{V}(\hat{\beta}_1)]^{-1}[\frac{a_n^{1/2}}{n^{1/2}}\sum_{i=1}^n g_i(\beta)]. \qquad (65)$$

Then use $\Psi_n(\beta)$ definition to have

$$S_n(\beta) = [a_n^{1/2}\Psi_n(\beta) + a_n^{1/2}n^{-1/2}\sum_{i=1}^n Eg_i(\beta)]'[a_n\hat{V}(\hat{\beta}_1)]^{-1}[a_n^{1/2}\Psi_n(\beta) + a_n^{1/2}n^{-1/2}\sum_{i=1}^n Eg_i(\beta)].$$

Next multiply $S_n(\beta)$ by a_n/n to have

$$\frac{a_n}{n}S_n(\beta) = [\frac{a_n}{n^{1/2}}\Psi_n(\beta) + Ea_nn^{-1}\sum_{i=1}^n g_i(\beta)]'[a_n\hat{V}(\hat{\beta}_1)]^{-1}[\frac{a_n}{n^{1/2}}\Psi_n(\beta) + Ea_nn^{-1}\sum_{i=1}^n g_i(\beta)].$$

Then by Theorem 1, uniformly in β

$$\frac{a_n}{n^{1/2}}\Psi_n(\beta) \stackrel{p}{\to} 0.$$

Use Assumption 4 and (11) with Assumption 1 to have on the nullspace of $\Omega(\beta_0)$

$$\frac{a_n}{n}S_n(\beta) \Longrightarrow m_1(\beta)'D^{-1}m_1(\beta).$$

Use positive definiteness of D and Assumption 1ii, iii, Assumption 4 in argmax continuous mapping theorem, Corollary 3.2.3 of van der Vaart and Wellner (1996), to have the consistency result. Now we try to find the limit.

The first order conditions for GMM objective function provides

$$0 = \left[\frac{a_n^{1/2}}{n^{1/2}}\sum_{i=1}^n G_i(\hat{\beta}_{GMM})\right]'\left[a_n\hat{V}(\hat{\beta}_1)\right]^{-1}\left[\frac{a_n^{1/2}}{n^{1/2}}\sum_{i=1}^n g_i(\hat{\beta}_{GMM})\right],$$

wpa1, by using also (65). Then take a Taylor series expansion for $g_i(\hat{\beta}_{GMM})$ around β_0 and after some simple algebra $\bar{\beta} \in (\beta_0, \hat{\beta}_{GMM})$

$$\begin{aligned}(\hat{\beta}_{GMM} - \beta_0) &= \left\{\left[\frac{a_n^{1/2}}{n^{1/2}}\sum_{i=1}^n G_i(\hat{\beta}_{GMM})\right]'\left[a_n\hat{V}(\hat{\beta}_1)\right]^{-1}\left[\frac{a_n^{1/2}}{n^{1/2}}\sum_{i=1}^n G_i(\hat{\beta}_{GMM})\right]\right\}^{-1} \\ &\times \left[\frac{a_n^{1/2}}{n^{1/2}}\sum_{i=1}^n G_i(\hat{\beta}_{GMM})\right]'\left[a_n\hat{V}(\hat{\beta}_1)\right]^{-1}\left[\frac{a_n^{1/2}}{n^{1/2}}\sum_{i=1}^n g_i(\beta_0)\right].\end{aligned}$$

Next multiply each side of the above equation by $n^{1/2}/a_n^{1/2}$, and use Assumptions 1, 2,4, 5 with consistency of $\hat{\beta}_1$, Lemma 1 to have the desired result. **Q.E.D.**

The following Lemma is needed for the higher order expansion for the first-step in GMM estimators. For standard GMM, this higher order expansion is established as Lemma A.5 in Newey and Smith (2004).

Lemma A.6.*Suppose that Assumptions 1-7 are satisfied and (6)-(8) are holding with m(.) given in (67). On the nullspace of W, let* $\Sigma_W = (G'K_W^{-1}G)^{-1}, H_W = \Sigma_W G' K_W^{-1}, P_W = K_W^{-1} - K_W^{-1} G H_W$, $\psi_i = -[H_W, P_W']g_i, G_i^j = E[\frac{\partial G_i(\beta_0)}{\partial \beta_j}]$

$$M_i = \frac{\partial m(x_i,\theta_0)}{\partial \theta} = -\begin{pmatrix} 0 & G_i' \\ G_i' & \frac{K_W}{a_n} + \xi(x_i) \end{pmatrix}.$$

$$M = -\begin{pmatrix} 0 & G' \\ G & K_W \end{pmatrix},$$

$$M^{-1} = -\begin{pmatrix} -\Sigma_W & H_W \\ H_W' & P_W \end{pmatrix},$$

$$M_j = -\begin{pmatrix} 0 & \lim_{n\to\infty} \frac{a_n}{n} \sum_{i=1}^n EG_i^{j'} \\ \lim_{n\to\infty} \frac{a_n}{n} \sum_{i=1}^n EG_i^{j'} & 0 \end{pmatrix}, \quad j \le p,$$

$$M_{p+j} = -\begin{pmatrix} \lim_{n\to\infty} \frac{a_n}{n} \sum_{i=1}^n E\frac{\partial^2 g_{ij}(\beta_0)}{\partial\beta\partial\beta'} & 0 \\ 0 & 0 \end{pmatrix}, \quad j \le m.$$

Then for the first step GMM estimates $\hat{\beta}_1$, and $\hat{\lambda}_1 = \hat{W}^{-1}\hat{g}(\hat{\beta}_1)$, $\hat{\theta}_1 = (\hat{\beta}_1', \hat{\lambda}_1')'$ and for $\tilde{\psi}, \tilde{a}$ and $Q_1(.)$ as in Lemma A.5 we have on the nullspace of W

$$\hat{\theta}_1 = \theta_0 + \tilde{\psi}/r_n + Q_1(\tilde{\psi},\tilde{a})/r_n^2 + O_p(r_n^{-2}),$$

where $r_n = (n/a_n)^{1/2} = n^{(1-\kappa)/2}, 0 < \kappa < 1$.

Proof of Lemma A.6. First see that

$$\frac{a_n^{1/2}}{n^{1/2}}\hat{\lambda}_1 = -[a_n\hat{W}]^{-1}\frac{a_n^{1/2}}{n^{1/2}}\sum_{i=1}^n g_i(\hat{\beta}_1). \tag{66}$$

Then apply Assumption 7, and a Taylor series expansion for $g_i(\hat{\beta}_1)$ in (66) around β_0 and using consistency and Lemma 1 with Assumption 5 supplies us

$$\frac{a_n^{1/2}}{n^{1/2}}\hat{\lambda}_1 = O_p(1).$$

Next by the result above and Theorem 6 we have $\hat{\theta}_1 = \theta_0 + O_p(n^{-1/2}a_n^{1/2})$. Furthermore set

$$m(x_i,\theta) = -(\lambda'\partial g(x_i,\beta)/\partial\beta, g(x_i,\beta) + \lambda'[\frac{K_W}{a_n} + \xi(x_i)])'. \tag{67}$$

By this choice of m(.) we obtain M_i, M_j.

Then by $\hat{m}(\hat{\theta}_1) = \frac{a_n}{n}\sum_{i=1}^n m(x_i, \hat{\theta}_1)$, and first order conditions for GMM with $\hat{\lambda}_1$ definition

$$\begin{aligned} 0 &= \hat{m}(\hat{\theta}_1) + [0, -\hat{\lambda}_1'(a_n(\hat{W} - W) - K_W - \tilde{\xi})]' \\ &= \hat{m}(\hat{\theta}_1) + O_p(a_n^{3/2} n^{-3/2}), \end{aligned}$$

since W is singular, and when $\hat{\lambda}_1'$ is in the nullspace of W, $\hat{\lambda}_1' W = 0$. Next expand $\hat{m}(\hat{\theta}_1)$ as in the proof of Lemma A.5 to have the desired result. **Q.E.D**

The following Lemma is nearly singular counterpart of Lemma A.6 of Newey and Smith (2004). This provides an expansion for the weight matrix in GMM for the second step of GMM.

Lemma A.7. *Suppose that Assumptions 1-7 are satisfied with the addition to Assumption 1ii*

$$a_n(\hat{\Omega}(\beta_0) - \Omega(\beta_0)) - D = O_p(n^{-1/2} a_n^{1/2}),$$

and (6)-(8) holding for the m (.) function given for GMM case (67), then on the nullspace of $\Omega(\beta_0)$

$$\begin{aligned} a_n[\hat{\Omega}(\hat{\beta}_1) - \Omega(\beta_0)] &= D + [a_n \hat{\Omega}(\beta_0) - a_n \Omega(\beta_0) - D] \\ &+ \frac{\sum_{j=1}^p \bar{\Omega}_{\beta_j} e_j' \tilde{\psi}^W}{(n^{1/2}/a_n^{1/2})} + \frac{\sum_{j=1}^p \bar{\Omega}_{\beta_j} e_j' Q_1^W}{(n/a_n)} \\ &+ \frac{\sum_{j=1}^p \tilde{\Omega}_{\beta_j} e_j' \tilde{\psi}^W}{(n/a_n)} + 1/2 \frac{\sum_{j,k=1}^p \bar{\Omega}_{\beta_k,\beta_j} e_j' \tilde{\psi}^W e_k' \tilde{\psi}^W}{(n/a_n)} + O_p(n^{-3/2} a_n^{3/2}), \end{aligned}$$

with $\tilde{\psi}^W, Q_1^W$ *is coming from the Conclusion of Lemma A.6. (i.e.* $\tilde{\psi}, Q_1(,)$ *in Lemma A.6 respectively)*

We should remind that in Lemma A.7 we consider the case when both steps in GMM face nearly singular design. The limit theory associated with those designs are:

$$\frac{n^{1/2}}{a_n^{1/2}}(\hat{\beta}_1 - \beta_0) \xrightarrow{d} N(0, (G' K_W^{-1} G)^{-1}),$$

$$\frac{n^{1/2}}{a_n^{1/2}}(\hat{\beta}_{GMM} - \beta_0) \xrightarrow{d} N(0, \Sigma),$$

$\Sigma = (G' D^{-1} G)^{-1}$. The notation in Lemma A.7 is as follows:

$$\bar{\Omega}_{\beta_j} = \lim_{n\to\infty} \sum_{i=1}^n a_n n^{-1} E\Omega_{i\beta_j},$$

$$\Omega_{i\beta_j} = \partial[g_i(\beta_0) g_i(\beta_0)']/\partial\beta_j.$$

$$\tilde{\Omega}_{\beta_j} = \frac{a_n^{1/2}}{n^{1/2}} \sum_{i=1}^n \Omega_{i\beta_j} - E[\partial g_i(\beta_0) g_i(\beta_0)']/\partial\beta_j,$$

$$\bar{\Omega}_{\beta_j\beta_k} = \lim_{n\to\infty} \sum_{i=1}^n a_n n^{-1} E \frac{\partial^2 [g_i(\beta_0) g_i(\beta_0)']}{\partial\beta_j \partial\beta_k}.$$

Proof of Lemma A.7. First have the following Taylor series expansion with the order of the remainder explained immediately below. Note that $\hat{\Omega}(.)$ matrices are defined as in section 2.

$$\begin{aligned} a_n\hat{\Omega}(\hat{\beta}_1) &= a_n\hat{\Omega}(\beta_0) + \sum_{j=1}^{p} \bar{\Omega}_{\beta_j}(\hat{\beta}_{1j} - \beta_{j0}) \\ &+ \frac{a_n^{1/2}}{n^{1/2}} \sum_{j=1}^{p} \tilde{\Omega}_{\beta_j}(\hat{\beta}_{1j} - \beta_{j0}) \\ &+ \sum_{j,k=1}^{p} \bar{\Omega}_{\beta_k,\beta_j}(\hat{\beta}_{1j} - \beta_{j0})(\hat{\beta}_{k1} - \beta_{k0})/2 + O_p(a_n^{3/2} n^{-3/2}). \end{aligned} \tag{68}$$

Now we give the proof of the order of the remainder in (68). By Assumption 6, as in the analysis of (33)-(35), and Theorem 6

$$\|\frac{a_n}{n}\sum_{i=1}^{n}\frac{\partial^2 g_i(\bar{\beta})g_i(\bar{\beta})'}{\partial\beta_k\partial\beta_j} - \frac{a_n}{n}\sum_{i=1}^{n}\frac{\partial^2 g_i(\beta_0)g_i(\beta_0)'}{\partial\beta_k\partial\beta_j}\| \leq (\frac{a_n}{n}\sum_{i=1}^{n} l(x_i))\|\bar{\beta} - \beta_0\| = O_p(n^{-1/2}a_n^{1/2}). \tag{69}$$

Then (7) implies

$$\|\frac{a_n}{n}\sum_{i=1}^{n}\frac{\partial^2 g_i(\beta_0)g_i(\beta_0)'}{\partial\beta_k\partial\beta_j} - E\frac{a_n}{n}\sum_{i=1}^{n}\frac{\partial^2 g_i(\beta_0)g_i(\beta_0)'}{\partial\beta_k\partial\beta_j}\| = O_p(n^{-1/2}a_n^{1/2}). \tag{70}$$

Then by (69)-(70)

$$\|\frac{a_n}{n}\sum_{i=1}^{n}\frac{\partial^2 g_i(\bar{\beta})g_i(\bar{\beta})'}{\partial\beta_k\partial\beta_j} - E\frac{a_n}{n}\sum_{i=1}^{n}\frac{\partial^2 g_i(\beta_0)g_i(\beta_0)'}{\partial\beta_k\partial\beta_j}\| = O_p(n^{-1/2}a_n^{1/2}). \tag{71}$$

Since the remainder term in (68) is (71) multiplied by $(\hat{\beta}_{1j} - \beta_{j0})(\hat{\beta}_{1k} - \beta_{k0})$, by Theorem 6, the order of the remainder is the one shown as the last term on the right-hand side of (68).

Then add and subtract D and $a_n\Omega(\beta_0)$

$$\begin{aligned} a_n[\hat{\Omega}(\hat{\beta}_1) - \Omega(\beta_0)] &= D + [a_n\hat{\Omega}(\beta_0) - a_n\Omega(\beta_0) - D] + \sum_{j=1}^{p} \bar{\Omega}_{\beta_j}(\hat{\beta}_{1j} - \beta_{j0}) \\ &+ \frac{a_n^{1/2}}{n^{1/2}} \sum_{j=1}^{p} \tilde{\Omega}_{\beta_j}(\hat{\beta}_{1j} - \beta_{j0}) \\ &+ \sum_{j,k=1}^{p} \bar{\Omega}_{\beta_k,\beta_j}(\hat{\beta}_{1j} - \beta_{j0})(\hat{\beta}_{k1} - \beta_{k0})/2 + O_p(a_n^{3/2} n^{-3/2}). \end{aligned} \tag{72}$$

Via the stochastic expansion for GMM in Lemma A.6

$$\begin{aligned} \hat{\beta}_{1j} - \beta_{j0} &= \frac{e_j'\tilde{\psi}^W}{n^{1/2}a_n^{-1/2}} + O_p(n^{-1}a_n) \\ &= \frac{e_j'\tilde{\psi}^W}{n^{1/2}a_n^{-1/2}} + \frac{e_j'Q_1^W}{na_n^{-1}} + O_p(n^{-3/2}a_n^{3/2}). \end{aligned} \tag{73}$$

Then substitute the first equality for the last two terms in equation (72). For the last term on the right-hand side of (72)

$$\sum_{j,k=1}^{p} \bar{\Omega}_{\beta_k\beta_j}(\hat{\beta}_{1j}-\beta_{j0})(\hat{\beta}_{1k}-\beta_{k0})/2 = \frac{1}{2}\sum_{j,k=1}^{p} \bar{\Omega}_{\beta_k\beta_j}\frac{e_j'\tilde{\psi}^W e_k'\tilde{\psi}^W}{n^{-1}a_n} + O_p(n^{-2}a_n^2). \tag{74}$$

Then operate on the following term on the right hand side of (72)

$$\frac{a_n^{1/2}}{n^{1/2}}\sum_{j=1}^{p} \tilde{\Omega}_{\beta_j}(\hat{\beta}_{1j}-\beta_{j0}) = \sum_{j=1}^{p} \frac{\tilde{\Omega}_{\beta_j} e_j'\tilde{\psi}^W}{na_n^{-1}} + O_p(n^{-3/2}a_n^{3/2}). \tag{75}$$

Substitute second equality in (73) for the third term on the right hand side of (72)

$$\sum_{j=1}^{p} \bar{\Omega}_{\beta_j}(\hat{\beta}_{1j}-\beta_{j0}) = \sum_{j=1}^{p} \frac{\bar{\Omega}_{\beta_j} e_j'\tilde{\psi}^W}{n^{1/2}a_n^{-1/2}} + \sum_{j=1}^{p} \frac{\bar{\Omega}_{\beta_j} e_j' Q_1^W}{na_n^{-1}} + O_p(n^{-3/2}a_n^{3/2}). \tag{76}$$

Use (74)-(76) in (72) to have the desired result. **Q.E.D**

Proof of Theorem 7. This is similar to the proof of Theorem 3.3 of Newey and Smith (2004). Note also that $\tilde{\psi}^W = \sum_{i=1}^n \psi_i^W/(n^{1/2}a_n^{-1/2})$, $\psi_i^W = -[H_W, P_W']' g_i$ as in Lemma A.6.

First for the second step in GMM

$$m(x_i,\theta) = -[\lambda' G_i(\beta), g_i(\beta)' + \lambda'(g_i g_i' - \Omega(\beta_0) + \sum_{j=1}^{p} \bar{\Omega}_{\beta_j} e_j' \psi_i^W)]'. \tag{77}$$

Then use $\hat{m}(\hat{\theta}) = \frac{a_n}{n}\sum_{i=1}^n m(x_i,\hat{\theta})$, and adding and subtracting D

$$\hat{m}(\hat{\theta}) = -[\hat{\lambda}'\frac{a_n}{n}\sum_{i=1}^{n} G_i(\hat{\beta}_{GMM}), \frac{a_n}{n}\sum_{i=1}^{n} g_i(\hat{\beta}_{GMM})' + \hat{\lambda}'(D + \tilde{\xi}^D/n^{1/2}a_n^{-1/2})]',$$

where

$$\frac{\tilde{\xi}^D}{n^{1/2}a_n^{-1/2}} = a_n\hat{\Omega}(\beta_0) - a_n\Omega(\beta_0) - D + \frac{\sum_{j=1}^p \bar{\Omega}_{\beta_j} e_j'\tilde{\psi}^W}{n^{1/2}a_n^{-1/2}}.$$

Matrices, M, M_i, M_j will be as in the first step GMM expansion of Lemma A.6 but substitute D for K_W, H for H_W, P for P_W, Σ for Σ_W. Furthermore, see that $\psi_i = -[H, P']' g_i, \tilde{\psi} = \sum_{i=1}^n \psi_i/(n^{1/2}a_n^{-1/2})$. $Q_1(.)$ is defined as in Lemma A.5.

Also

$$B_j^1(x_i) = -\begin{pmatrix} 0 & G_i^{j'} - EG_i^{j'} \\ G_i^j - EG_i^j & 0 \end{pmatrix}, \quad (j \le p),$$

$$B_{p+j}^1(x_i) = -\begin{pmatrix} \frac{\partial^2 g_{ij}(\beta_0)}{\partial\beta'\partial\beta'} - E\frac{\partial^2 g_{ij}(\beta_0)}{\partial\beta'\partial\beta'} & 0 \\ 0 & 0 \end{pmatrix} \quad (j \le m).$$

Set $\hat{\lambda} = -(\hat{\Omega}(\hat{\beta}_1))^{-1}\hat{g}(\hat{\beta}_{GMM})$. Then rewrite that

$$\frac{n^{1/2}}{a_n^{1/2}}\hat{\lambda} = -[a_n\hat{\Omega}(\hat{\beta}_1)]^{-1}\left[\frac{a_n^{1/2}}{n^{1/2}}\sum_{i=1}^{n} g_i(\hat{\beta}_{GMM})\right].$$

Then by Assumption 1, and a Taylor series expansion around β_0 for $g_i(\hat{\beta}_{GMM})$, coupled with Theorem 6, Assumption 5 and Lemma 1 shows that $\hat{\lambda} = O_p(a_n^{1/2} n^{-1/2})$. As in Newey and Smith (2004)

$$Q_2 = Q_{21}(\tilde{\psi}, \tilde{a}, \tilde{b}) + M^{-1} diag(0, \tilde{Q}_1^D)/(n/a_n),$$

where $Q_{21}(.)$ above is $Q_2(.)$ in Lemma A.5 and

$$\tilde{Q}_1^D = \sum_{j=1}^p \tilde{\Omega}_{\beta_j} e_j' \tilde{\psi}^W + \sum_{j=1}^p \bar{\Omega}_{\beta_j} e_j' Q_1^W + 1/2 \sum_{j,k=1}^p \bar{\Omega}_{\beta_k, \beta_j} e_j' \tilde{\psi}^W e_k' \tilde{\psi}^W.$$

Q_1^W is described in Lemma A.6, and the other terms are described in the statement of Lemma A.7. We can proceed as in the proof of Theorem 3.3 in Newey and Smith (2004) to have the result by seeing $r_n = n^{1/2}/a_n^{1/2}$. **Q.E.D**

Proof of Theorem 8. By the proof of Theorem 7 and $r_n = n^{1/2}/a_n^{1/2}$,

$$\frac{\tilde{\xi}^D}{r_n} = [a_n \hat{\Omega}(\beta_0) - a_n \Omega(\beta_0) - D] + \frac{\sum_{j=1}^p \bar{\Omega}_{\beta_j} e_j' \tilde{\psi}^\omega}{r_n}.$$

Use Lemma A.7 to claim

$$\begin{aligned} a_n(\hat{\Omega}(\hat{\beta}_1) - \Omega(\beta_0)) &= D + [a_n \hat{\Omega}(\beta_0) - a_n \Omega(\beta_0) - D] \\ &+ \frac{\sum_{j=1}^p \bar{\Omega}_{\beta_j} e_j' \tilde{\psi}^\omega}{r_n} + O_p(r_n^{-2}) \\ &= D + \frac{\tilde{\xi}^D}{r_n} + O_p(r_n^{-2}). \end{aligned}$$

Then using $\hat{W} = \hat{\Omega}(\hat{\beta}_1)$, substituting $W = \Omega(\beta_0), K_W = D$, and also benefiting from the addition to Assumption 1ii

$$a_n \hat{\Omega}(\beta_0) - a_n \Omega - D = O_p(r_n^{-1}),$$

we have $\tilde{\xi}^D / r_n = O_p(r_n^{-1})$ is satisfied. So Assumption 7 is satisfied for the second step in GMM as well. By (55)

$$Bias = EQ_1(\tilde{\psi}, \tilde{a})/r_n^2,$$

where

$$EQ_1(\tilde{\psi}, \tilde{a}) = -M^{-1}[\tilde{A}\tilde{\psi} + \sum_{j=1}^q M_j diag(\Sigma, P)' e_j/2].$$

See that

$$\begin{aligned} E\tilde{A}\tilde{\psi} &= E\frac{\sum_{i=1}^n A(x_i)\psi(x_i)}{r_n^2} \\ &= \sum_{i=1}^n E\left\{ \begin{bmatrix} 0 & G_i' \\ G_i & g_i g_i' - E g_i g_i' + \sum_{j=1}^p \bar{\Omega}_{\beta_j} e_j' \psi_i^W \end{bmatrix} \begin{bmatrix} H \\ P \end{bmatrix} g_i \right\} / r_n^2, \end{aligned}$$

by using the independence, $A(x_i) = \frac{\partial m(x_i,\theta_0)}{\partial \theta'} - E\frac{\partial m(x_i,\theta_0)}{\partial \theta'}$, and $m(x_i,\theta)$ in (77), $\psi_i = -(H', P)'g_i$, $\psi_i^W = -(H_W', P_W)'g_i$. Given the M_j, M_{p+j} definitions in Lemma A.6 and following the proof of Theorem 4.1 in Newey and Smith (2004) using "r_n" instead of "$n^{1/2}$" gives us the desired result.**Q.E.D.**

REFERENCES

ANDREWS, D.W.K. (1994): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics*, Vol.4, ed. by R.Engle and D.McFadden. Amsterdam: North-Holland, 2247-2294.

DAVIDSON, J. (1994): *Stochastic Limit Theory.* Oxford University Press.

GUGGENBERGER, P., AND R.J. SMITH (2003): "Generalized Empirical Likelihood Estimators and Tests Under Partial, Weak and Strong Identification", Working Paper, UCLA.

HANSEN, L.P., J.HEATON, AND A. YARON (1996): "Finite Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics,* 14, 262-280.

IMBENS, G.W., R.H. SPADY, AND P. JOHNSON (1998): "Information Theoretic Approaches to Inference in Moment Condition Models," *Econometrica*, 66, 333-357.

KITAMURA, Y., AND M. STUTZER (1997): "An Information-Theoretic Alternative to Generalized Methods of Moments Estimation," *Econometrica*, 861-874.

KNIGHT, K., AND W. FU (2000): "Asymptotics For Lasso Type Estimators," *Annals of Statistics,* 28, 1356-1378.

KOCHERLAKOTA, N. (1990): "On Tests of Representative Consumer Asset Pricing Models," *Journal of Monetary Economics,* 26, 285-304.

NEWEY, W.K., AND R.J. SMITH (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," 72, 219-255.

OWEN, A. (2001): *Empirical Likelihood.* New York: Chapman and Hall.

PAGAN, A., AND J.ROBERTSON (1997): "GMM and Its Problems," Manuscript, Australian National University.

PFANZAGL, J., AND W. WEFELMEYER (1978): "A Third-Order Optimum Property of the

Maximum Likelihood Estimator," *Journal of Multivariate Analysis,* 8, 1-29.

POLLARD, D., (1990): *Empirical Processes: Theory and Applications.* NSF-CBMS Regional Conference Series in Probability and Statistics, Vol.2.

QIN, J. AND J. LAWLESS (1994): "Empirical Likelihood and General Estimating Equations," *Annals of Statistics,* 22, 300-325.

SMITH, R.J. (1997): "Alternative Semi-Parametric Likelihood Approaches to Generalized Method of Moments Estimation," *Economic Journal,* 107, 503-519.

STOCK, J.H., AND J.H. WRIGHT (2000): "GMM with Weak Identification," *Econometrica*, 68, 1055-1096.

TAUCHEN, G. (1986): "Statistical Properties of Generalized Method of Moments Estimators of Structural Parameters Obtained from Financial Market Data," *Journal of Business and Economic Statistics*, 4, 397-425.

TECHNICAL APPENDIX:

This appendix deals with the proofs for section 5.1. We provide now a Lemma that is useful in deriving the higher order expansion for the peculiar GMM estimator.

Technical Lemma A.1. *Suppose that our GMM estimator follows (9)-(10) and Assumptions 1-6 and 8 are satisfied and with the additional Assumption to 1ii of*

$$a_n(\hat{\Omega}(\beta_0) - \Omega(\beta_0)) = D + O_p(n^{1/2}a_n^{-1/2}),$$

then on the nullspace of $\Omega(\beta)$

$$\begin{aligned} a_n[\hat{\Omega}(\hat{\beta}_1) - \Omega(\beta_0)] &= D + [a_n\hat{\Omega}(\beta_0) - a_n\Omega(\beta_0) - D] \\ &+ \frac{\sum_{j=1}^p \bar{\Omega}_{\beta_j} e_j' \tilde{\psi}^\omega}{n^{1/2}} + \frac{\sum_{j=1}^p \tilde{\Omega}_{\beta_j} e_j' \tilde{\psi}^\omega}{n/a_n^{1/2}} \\ &+ \frac{\sum_{j=1}^p \bar{\Omega}_{\beta_j} e_j' Q_1^\omega}{n} + \frac{1}{2}\frac{\sum_{j,k=1}^p \bar{\Omega}_{\beta_k,\beta_j} e_j' \tilde{\psi}^\omega e_k' \tilde{\psi}^\omega}{n} + O_p(n^{-3/2}a_n^{1/2}), \end{aligned}$$

where $\bar{\Omega}_{\beta_j}, \tilde{\Omega}_{\beta_j}, \bar{\Omega}_{\beta_k,\beta_j}$ *are described in Lemma A.7. Also* $\tilde{\psi}^\omega = \sum_{i=1}^n \psi_i/n^{1/2}$, $\psi_i^\omega = -[H_\omega', P_\omega]' g_i, H_\omega = \Sigma_\omega G' W^{-1}, P_\omega = W^{-1} - W^{-1} G H_\omega, \Sigma_\omega = (G'W^{-1}G)^{-1}$.

$$Q_1^\omega = -M_\omega^{-1}[\tilde{A}^\omega \tilde{\psi}^\omega + \sum_{j=1}^q \tilde{\psi}_j^\omega M_j^\omega \tilde{\psi}^\omega/2],$$

where $M_\omega = \lim_{n\to\infty} En^{-1}\sum_{i=1}^n \frac{\partial m(x_i,\theta_0)}{\partial\theta'}$, $\tilde{A}^\omega = \frac{1}{n^{1/2}}\sum_{i=1}^n A(x_i)$, $M_j^\omega = \lim_{n\to\infty} En^{-1}\sum_{i=1}^n \frac{\partial^2 m(x_i,\theta_0)}{\partial\theta_j\partial\theta'}$ *and* $m(x_i,\theta) = -(\lambda'\partial g(x_i,\beta)/\partial\beta', g(x_i,\beta) + \lambda[W + \xi(x_i)])$. *These terms are from the standard GEL of Newey and Smith (2004), since now the first step of GMM has no near-singularity problem.*

Remark. The terms in the expansion are different than the ones in Lemma A.7 . Except from the first two terms on the right hand side of the expansion the other terms on the right hand side of Technical Lemma A.1 converges to zero at a faster rate than in Lemma A.8. This is because of the standard behavior of the first-step GMM estimator here in Technical Lemma A.1 compared to nearly singular behavior of both first and second steps in GMM in Lemma A.7.

Proof of Technical Lemma A.1. Instead of the Taylor Series expansion that is used in the proof of Lemma A.7, (72) we have

$$\begin{aligned} a_n[\hat{\Omega}(\hat{\beta}_1) - \Omega(\beta_0)] &= D + [a_n\hat{\Omega}(\beta_0) - a_n\Omega(\beta_0) - D] + \sum_{j=1}^p \bar{\Omega}_{\beta_j}(\hat{\beta}_{1j} - \beta_{j0}) \\ &+ \frac{a_n^{1/2}}{n^{1/2}}\sum_{j=1}^p \tilde{\Omega}_{\beta_j}(\hat{\beta}_{1j} - \beta_{j0}) \\ &+ \sum_{j,k=1}^p \bar{\Omega}_{\beta_k,\beta_j}(\hat{\beta}_{1j} - \beta_{j0})(\hat{\beta}_{1k} - \beta_{k0})/2 + O_p(a_n^{1/2}n^{-3/2}). \end{aligned} \tag{78}$$

The difference between (72) and (78) is in the remainder term. The order of the remainder term here is determined by the first step GMM estimators $(\hat{\beta}_{1j}-\beta_{j0}) = O_p(n^{-1/2}), (\hat{\beta}_{1k}-\beta_0) = O_p(n^{-1/2})$ (equation (9)) and by (70)-(71).

Then use independent data version of Lemma A.5 of Newey and Smith (2004) to have

$$\hat{\beta}_{1j} - \beta_{j0} = \frac{e_j' \tilde{\psi}^\omega}{n^{1/2}} + O_p(n^{-1}), \tag{79}$$

$$\hat{\beta}_{1j} - \beta_{j0} = \frac{e_j' \tilde{\psi}^\omega}{n^{1/2}} + \frac{e_j' Q_1^\omega}{n} + O_p(n^{-3/2}). \tag{80}$$

Substitute (79) for the last two terms on the right-hand side of (78) to have

$$\frac{1}{2}\sum_{j,k=1}^{p} \Omega_{\beta_k,\beta_j}(\hat{\beta}_{1j} - \beta_{j0})(\hat{\beta}_{1k} - \beta_{k0}) = \frac{1}{2}\frac{\sum_{j,k=1}^{p} \bar{\Omega}_{\beta_k \beta_j} e_j' \tilde{\psi}^\omega e_k' \tilde{\psi}^\omega}{n} + O_p(n^{-2}), \tag{81}$$

and

$$\frac{a_n^{1/2}}{n^{1/2}} \sum_{j=1}^{p} \tilde{\Omega}_{\beta_j}(\hat{\beta}_{1j} - \beta_{j0}) = \frac{\sum_{j=1}^{p} \tilde{\Omega}_{\beta_j} e_j' \tilde{\psi}^\omega}{n/a_n^{1/2}} + O_p(n^{-3/2} a_n^{1/2}). \tag{82}$$

Next substitute the expansion (80) into the following term in (78)

$$\sum_{j=1}^{p} \bar{\Omega}_{\beta_j}(\hat{\beta}_{1j} - \beta_{j0}) = \frac{\sum_{j=1}^{p} \bar{\Omega}_{\beta_j} e_j' \tilde{\psi}^\omega}{n^{1/2}} + \frac{\sum_{j=1}^{p} \bar{\Omega}_{\beta_j} e_j' Q_1^\omega}{n} + O_p(n^{-3/2}), \tag{83}$$

Then combine (81)-(83) into (78) to have the result.**Q.E.D.**

Proof of Theorem 9. Let

$$m(x_i, \theta) = -[\lambda' G_i(\beta), g_i(\beta)' + \lambda'(\frac{D}{a_n} + g_i g_i' - \Omega + \sum_{j=1}^{p} \bar{\Omega}_{\beta_j} e_j' \psi_i^\omega / a_n)]'.$$

Then using $\hat{m}(\hat{\theta}) = \frac{a_n}{n}\sum_{i=1}^n m(x_i, \hat{\theta})$, by adding and subtracting D

$$\hat{m}(\hat{\theta}) = -[\hat{\lambda}' \frac{a_n}{n} \sum_{i=1}^{n} G_i(\hat{\beta}_{GMM}), \frac{a_n}{n} \sum_{i=1}^{n} g_i(\hat{\beta}_{GMM}) + \hat{\lambda}'(D + \tilde{\xi}^D / n^{1/2} a_n^{-1/2})]',$$

where

$$\frac{\tilde{\xi}^D}{n^{1/2} a_n^{-1/2}} = \frac{\sum_{i=1}^{n} \xi(x_i)}{n a_n^{-1}} = a_n \hat{\Omega}(\beta_0) - a_n \Omega(\beta_0) - D + \frac{\sum_{j=1}^{p} \bar{\Omega}_{\beta_j} e_j' \tilde{\psi}^\omega}{n^{1/2}}.$$

Note that $\tilde{\psi}^\omega = \sum_{i=1}^n \psi_i^\omega / n^{1/2}$, since the first step GMM estimator is not subject to nearly singular design in this peculiar case.

We have to introduce more notation. To form $A(x_i) = M_i - EM_i$, where $M_i = \frac{\partial m(x_i, \theta_0)}{\partial \theta'}$ we use

$$M_i = -\begin{pmatrix} 0 & G_i' \\ G_i & g_i g_i' - \Omega(\beta_0) + \sum_{j=1}^{p} \bar{\Omega}_{\beta_j} e_j' \psi_i^\omega / a_n \end{pmatrix}. \tag{84}$$

Also see that by Assumption 1

$$M = \lim_{n\to\infty} \frac{a_n}{n} \sum_{i=1}^{n} E\frac{\partial m(x_i,\theta_0)}{\partial \theta'} = -\begin{pmatrix} 0 & G' \\ G & D \end{pmatrix}.$$

M_j, M_{p+j} are given in Lemma A.6.

$$B_j^1(x_i) = -\begin{pmatrix} 0 & G_i^{j'} - EG_i^{j'} \\ G_i^j - EG_i^j & 0 \end{pmatrix} \quad j \le p,$$

$$B_{p+j}^1(x_i) = -\begin{pmatrix} \frac{\partial^2 g_{ij}(\beta_0)}{\partial\beta\partial\beta'} - E\frac{\partial^2 g_{ij}(\beta_0)}{\partial\beta\partial\beta'} & 0 \\ 0 & 0 \end{pmatrix} \quad j \le m.$$

Let $\hat{\lambda} = -(\hat{\Omega}(\hat{\beta}_1))^{-1}\hat{g}(\hat{\beta}_{GMM})$. Then by Assumption 1, and a Taylor series expansion around $g_i(\hat{\beta}_{GMM})$ at β_0, coupled with limit for $\hat{\beta}_{GMM}$ and Assumption 5 we have $\hat{\lambda} = O_p(a_n^{1/2} n^{-1/2})$.

Next define

$$\tilde{Q}_{11}^{\Omega} = \sum_{j=1}^{p} \tilde{\Omega}_{\beta_j} e_j' \tilde{\psi}^{\omega},$$

$$\tilde{Q}_{12}^{\Omega} = \sum_{j=1}^{p} \bar{\Omega}_{\beta_j} e_j' Q_1^{\omega} + 1/2 \sum_{j,k=1}^{p} \bar{\Omega}_{\beta_k,\beta_j} e_j' \tilde{\psi}^{\omega} e_k' \tilde{\psi}^{\omega}.$$

By first order conditions for GMM and Technical Lemma A.1

$$\begin{aligned} 0 &= \hat{m}(\hat{\theta}) + [0, -\hat{\lambda}'(\frac{\tilde{Q}_{11}^{\Omega}}{n a_n^{-1/2}} + \frac{\tilde{Q}_{12}^{\Omega}}{n} + O_p(n^{-3/2} a_n^{1/2}))] \\ &= \hat{m}(\hat{\theta}) + [0, -\hat{\lambda}'\frac{\tilde{Q}_{11}^{\Omega}}{n a_n^{-1/2}} - \hat{\lambda}'\frac{\tilde{Q}_{12}^{\Omega}}{n}] + O_p(n^{-2} a_n). \end{aligned} \tag{85}$$

We need to expand $\hat{m}(\hat{\theta})$ as in Lemma A.5. However, there will be differences here, because of the nature of the first step GMM estimators in this peculiar case. So $Q_1(.)$ is defined as in Lemma A.5

$$Q_1 = -M^{-1}[\tilde{A}\tilde{\psi} + \sum_{j=1}^{q} \tilde{\psi}_j M_j \tilde{\psi}/2],$$

where $\tilde{A} = \frac{a_n^{1/2}}{n^{1/2}} \sum_{i=1}^{n} A(x_i)$, by (84) and $A(x_i) = M_i - EM_i$,

$$A(x_i) = -\begin{pmatrix} 0 & G_i' - EG_i' \\ G_i - EG_i & g_i g_i' - E g_i g_i' + \sum_{j=1}^{p} \bar{\Omega}_{\beta_j} e_j' \psi_i^{\omega}/a_n \end{pmatrix}.$$

Then we can decompose $A(x_i)$ into two

$$A(x_i) = A_1(x_i) + A_2(x_i)/a_n,$$

where $A_1(x_i)$ is

$$A_1(x_i) = -\begin{pmatrix} 0 & G_i' - EG_i' \\ G_i - EG_i & g_i g_i' - E g_i g_i' \end{pmatrix},$$

and $A_2(x_i)$ is

$$A_2(x_i) = -\begin{pmatrix} 0 & 0 \\ 0 & \sum_{j=1}^{p} \bar{\Omega}_{\beta_j} e_j' \psi_i^{\omega} \end{pmatrix}.$$

Then

$$\begin{aligned} \tilde{A} &= \frac{a_n^{1/2}}{n^{1/2}} \sum_{i=1}^{n} A(x_i) \\ &= \left(\frac{\sum_{i=1}^{n} A_1(x_i)}{n^{1/2} a_n^{-1/2}} \right) + \left(\frac{\sum_{i=1}^{n} A_2(x_i)}{n^{1/2} a_n^{1/2}} \right) \\ &= \tilde{A}_1 + \tilde{A}_2^{\omega} / a_n^{1/2}, \end{aligned}$$

where $\tilde{A}_2^{\omega} = \sum_{i=1}^{n} A_2(x_i)/n^{1/2}$, and $\tilde{A}_2^{\omega} = O_p(1)$ since the first step GMM is not subject to nearly singular design. $\tilde{A}_1 = O_p(1)$ by (6)-(8). Using this decomposition we can rewrite Q_1 as

$$Q_1 = Q_{11}(\tilde{\psi}, \tilde{a}) + Q_{12}/a_n^{1/2},$$

where

$$Q_{11}(\tilde{\psi}, \tilde{a}) = -M^{-1}[\tilde{A}_1 \tilde{\psi} + \sum_{j=1}^{q} \tilde{\psi}_j M_j \tilde{\psi}/2],$$

$$Q_{12}(\tilde{\psi}, \tilde{a}^{\omega}) = -M^{-1} \tilde{A}_2^{\omega} \tilde{\psi}.$$

Now we proceed as in Lemma A.5 and have the Taylor series expansion for $\hat{m}(\hat{\theta})$, and noting that $Q_{21}(.)$ is Q_2 in Lemma A.5, so by (85)

$$\begin{aligned} \hat{\theta} &= \theta_0 + \frac{\tilde{\psi}}{n^{1/2} a_n^{-1/2}} + \frac{Q_{11}(\tilde{\psi}, \tilde{a})}{n a_n^{-1}} + \frac{Q_{12}(\tilde{\psi}, \tilde{a}^{\omega})}{n a_n^{-1/2}} + \frac{Q_{21}(\tilde{\psi}, \tilde{a}, b^1)}{n^{3/2} a_n^{-3/2}} \\ &+ M^{-1}[0, \hat{\lambda}' \frac{\tilde{Q}_{11}^{\Omega}}{n a_n^{-1/2}} + \hat{\lambda}' \frac{\tilde{Q}_{12}^{\Omega}}{n}] + O_p(n^{-2} a_n). \end{aligned}$$

Since $\hat{\lambda}' \tilde{Q}_{11}^{\Omega} / (n a_n^{-1/2}) = O_p(n^{-3/2} a_n)$, and by $\hat{\lambda} = [0, I_m] \tilde{\psi} / (n a_n^{-1})^{1/2}$ substituting that in the above expression

$$\begin{aligned} \hat{\theta} &= \theta_0 + \frac{\tilde{\psi}}{n^{1/2} a_n^{-1/2}} + \frac{Q_{11}(\tilde{\psi}, \tilde{a})}{n a_n^{-1}} + \frac{Q_{12}(\tilde{\psi}, \tilde{a}^{\omega})}{n a_n^{-1/2}} + \frac{Q_{21}(\tilde{\psi}, \tilde{a}, b^1)}{n^{3/2} a_n^{-1/2}} \\ &+ M^{-1} diag[0, \tilde{Q}_{11}^{\Omega}] \tilde{\psi} / (n^{3/2} a_n^{-1}) + M^{-1} diag[0, \tilde{Q}_{12}^{\Omega}] \tilde{\psi} / (n^{3/2} a_n^{-3/2}) + O_p(n^{-2} a_n). \end{aligned}$$

Then arranging gives us the desired result.**Q.E.D**

Proof of Theorem 10. The proof is the same as in Theorem 8 given the new bias definition. However, from analyzing the proof of Theorem 8 and the proof of Theorem 9, when GMM is subject to nearly singular design only, we donot use $Q_{12}()$ term which basically resulted in B_W in Theorem 8. The relationship between B_W and Q_{12} is also very well illustrated in the proof of Theorem 4.1 of Newey and Smith (2004).**Q.E.D.**