Near Exogeneity and Weak Identification in Generalized Empirical Likelihood Estimators : Fixed and Many Moment Asymptotics

Mehmet Caner
University of Pittsburgh *

July 7, 2005

Abstract

This paper analyzes near exogeneity and weak identification in Generalized Empirical Likelihood Estimators. Near exogeneity and weak identification are related to the exogeneity and relevance of the instruments, respectively. These two issues are important from an applied perspective, such as empirical growth theory and labor economics. In the case of empirical growth and institutional economics literature a small number of moments/instruments are used in studies. First, we analyze the limit behavior of estimators and tests under fixed number of weak moments and near exogeneity. We show that Anderson-Rubin (1949) and Kleibergen (2002) type of tests' limits change when there is small correlation between the instruments and the structural equation error. The new limits are obtained under the null hypothesis at the true vale of the parameter. The test statistics are no longer asymptotically pivotal in the joint case of near exogeneity and weak instruments compared to the weak identification case. We use subsampling to get the new critical values. Subsampling is uniformly consistent in this specific case since tests are evaluated at true values rather than estimators. Contiguity concept also plays an important role in deriving our subsampling proof. Simulations show that among the class of GEL estimators and the tests we analyze, Anderson-Rubin (1949) type tests in Exponential Tilting and Continuous Updating framework have the best small sample properties. Empirical Likelihood based tests do not have good power. We also show that when used with the χ^2 critical values, which are not valid in the case of near exogeneity and weak instruments, the tests show large size distortions. Then we develop the limits of estimators and tests under many weak moments with near exogeneity. The results are different from the fixed moments case. Estimators are consistent, and test limits are simple, noncentral χ^2.

JEL Classification: C12, C15, C30.

Keywords: Valid Instruments, Subsampling, Many Weak Moments.

*Department of Economics, WW Posvar Hall 4S01, Pittsburgh, PA 15260. email: caner@pitt.edu. We thank Dan Berkowitz for discussions about the empirical ideas that led to this work.

1 Introduction

In empirical work, finding valid instrumental variable is very difficult. The instrument should satisfy two important criteria. The first is that the instrumental variable should be strongly correlated with endogenous explanatory variable. The second one is instrument should be uncorrelated with the structural errors. These two criteria may be conflicting as well. One instrument may be weak but exogenous, hence satisfying the second condition but not the first one, and vice versa. To give examples, in a study by Angrist and Krueger (1991) it is argued that quarter of birth may be used as instrument for educational attainment. Educational attainment is endogenous in an earnings equation. But several authors contend that quarter of birth is weak as an instrument, not correlated well with educational attainment. We can also see the case of instrument being not exogenous as well. In a study by Angrist (1990), the effect of serving in the Vietnam war on the earnings of men is considered. Clearly the participation in the military is endogenous. So Angrist (1990) uses the low draft lottery number as an instrument for participation. However, it is argued by Wooldridge (2002) that if we do not control for education in the earnings equation, lottery number may be endogenous.

Similar problems can also be seen in the empirical growth literature. One of the major questions is whether political institutions cause economic growth. There have been numerous papers analyzing this relation. Some of them Acemoglu et.al (2001, 2002) and Glaeser et. al (2004) use instrumental variable estimation. Acemoglu et.al (2001, 2002) show that settler mortality and population density in year 1500 predict institutional quality and the level of economic development today. The institutional quality then predicts economic growth in a structural equation. However, Glaeser et. al (2004) object to using settler mortality and population density in 1500 as instruments for political institutions. They show that settlement patterns influence growth through channels other than institutions, hence these instruments are correlated with error in the structural growth regression. This is the problem of near exogeneity we want to analyze. They also show that if you use the definition of political institutions as constitutional measures of checks and balances (i.e. plurality of the political system, proportional representation) then settler mortality and population density in 1500 is very weakly correlated with plurality and proportional representation. This is the problem of weak instruments. We think that these two problems are pervasive in the literature and hence methods robust to these problems should be useful. Importance of joint problem of near exogeneity and weak instruments are also mentioned in Stock, Wright, and Yogo (2002).

Bound, Jaeger, and Baker (1995) point out that even a small correlation between the instrument and the endogenous variable exacerbate the weak instruments problem in simultaneous equations setup. This small correlation may increase the finite sample bias and inconsistency of the estimators may be larger. Then they analyze Angrist and Krueger (1991) study and claim that quarter of birth used as an instrument for educational attainment is a weak instrument as described above.

But also, by using economic, educational, and psychiatric evidence they show that quarter of birth is not exogenous as an instrument and overidentification tests fail to detect this.

A somewhat related literature is the analysis of misspecified models. In the case of GMM, Newey (1985), and Hall and Inoue (2003) analyze local, and global misspecification of the model, respectively. They develop tests of detection of misspecification and inference procedures under misspecification. Our analysis considers validity of instruments rather than modeling issues. To give an example from linear structural equations model, we assume the linear structural equation is correctly specified but the instrument may be weakly correlated with structural error.

In this article we model near exogeneity as the local to zero correlation between the instruments and the structural error in the linear simultaneous equations case. In the nonlinear case, this is modeled as the local to zero population orthogonality conditions at the true value of the parameter vector. Weak identification is modeled as in Stock and Wright (2000). In our model, when sample size goes to infinity identification is not possible, but we have exogenous instruments. This way of modeling shows the tradeoff between the identification and exogeneity.

This paper analyzes the joint problem of weak identification and near exogeneity in Generalized Empirical Likelihood Estimators (GEL). These estimators have desirable properties and asymptotically pivotal tests are available in the case of weak instruments compared with two-step GMM, as can be seen from the articles by Stock and Wright (2000), Guggenberger and Smith (2005), Newey and Smith (2004). Also with this type of analysis we can learn which GEL estimator (Empirical Likelihood, Exponential Tilting, Continuous Updating) has better small sample properties when faced with these problems. Empirical Likelihood is introduced by Owen (1990), Imbens (1993) and analyzed in large samples by Qin and Lawless (1994). Asymptotic theory of Exponential Tilting estimator is analyzed by Imbens, Spady and Johnson (1998), and Kitamura and Stutzer (1997).

As discussed above, two strands of economic literature show different characteristics. Empirical growth and institutional economics work with smaller data sets. They have problems of near exogeneity and weak identification in their instruments, but the number of moments/instruments is small. On the other hand, labor economists work with large data sets. The models have many moments/instruments with near exogeneity and weak identification as discussed by Bound,Jaeger, and Baker (1995). We analyze asymptotics of these two cases. First one is near exogeneity and weak identification with fixed number of moments. Weak identification with fixed number of moments is analyzed by Stock and Wright (2000) for GMM and then Caner (2004), Guggenberger and Smith (2005) analyzed this for exponential tilting and GEL estimators, respectively. In this article we extend these former work to the important problem of joint analysis of near exogeneity and weak identification in GEL. The second case is near exogeneity with many weak moments. In the case of linear and nonlinear GMM papers by Chao and Swanson (2002), Bekker (1994), Han and Phillips (2004) analyze many weak moments problem and derive the limits of estimators. GEL with many weak moments is considered recently by Newey and Windmeijer (2005). We extend this work to near exogeneity. Furthermore we show that limits in the cases of fixed and many moment scenarios are different for estimators as well as test statistics. In addition to that because of near exogeneity, compared to exogenous cases analyzed before, the tests are no longer pivotal.

This study establishes the limits of GEL estimators under near exogeneity and weak identification. In the case of fixed number of moments we show that the bias of the estimators may be larger compared to only weak identification case of Guggenberger and Smith (2005). Then we analyze the limits of Anderson-Rubin (1949) and Kleibergen (2002) type of tests in the joint case of weak instruments and near exogeneity. Guggenberger and Smith (2005) show that the limits are χ^2 in the case of weak instruments. Here we show that the limits change when we also have near exogeneity, so these tests' limits are not robust to this problem. Next we subsample test statistics to have the data dependent critical values. We prove that subsampling is consistent in this joint case. We subsample the tests at θ_0 not $\hat{\theta}$. So subsampling works for Anderson-Rubin (1949) and Kleibergen (2002) type of tests. Simulations show that Anderson-Rubin (1949) type of tests in Exponential Tilting and Continuous Updating frameworks have the best size and power combination. We also introduce an overidentifying restrictions test in GEL framework. But we think that as long as there are constraints on our choice of instruments, it may be better to use subsampling in Anderson-Rubin (1949) type of tests. In the case of many weak moments with near exogeneity, Anderson-Rubin type tests are invalid, but we can subsample Kleibergen (2002) type of test.

Dufour (2003), Dufour and Taamouti (2004) analyze the exact version of excluded instruments setup. This excluded instruments setup is clearly different form ours, since we consider the case of instruments weakly correlated with the structural equation error, and the endogenous explanatory variables. Sections 2-4 analyze fixed number of moments case. Section 2 introduces the model and assumptions. Section 3 deals with estimation. Testing is analyzed in Section 4. Section 5 provides subsampling technique in our case. Section 6 introduces many weak moment asymptotics under near exogeneity. Section 7 introduces simulations that compare various test statistics and GEL estimators for the fixed number of moments case. Section 8 concludes. Appendix contains all the proofs. The following notation is used. $\stackrel{p}{\to}, \Longrightarrow, \stackrel{d}{\to}$ represent convergence in probability, weak convergence, and convergence in distribution respectively. $C^j(A)$ denotes j times continuous differentiable function.

2 Fixed Number of Moments: Model and Assumptions

First, we consider GEL estimators with fixed number of moments. We benefit from the definition in Guggenberger and Smith (2005). Let $\rho(.)$ be a real valued function, $\mathcal{V} \to \mathcal{R}$ where $\mathcal{V}$ is an open interval of the real line that contains 0 and $\hat{\Lambda}_n(\theta) = \{\lambda \in R^q : \lambda' g_i(\theta) \in \mathcal{V} \quad \text{for}\, i = 1, \cdots, n\}$. $\theta \in \Theta$ where Θ is a compact subset of R^p. Also define $\rho_j(\nu) = \partial^j \rho / \partial \nu^j$ and $\rho_j = \rho_j(0)$ for nonnegative integers j. Let $\rho_0 = \rho(0)$. We want to estimate the unknown θ_0, which is the true parameter vector. The GEL estimator is the solution to a saddle point problem

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \sup_{\lambda \in \hat{\Lambda}_n(\theta)} \hat{P}(\theta, \lambda),$$

where

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

In this section we model the weakly identified and nearly exoogenous GEL estimator. We introduce the assumptions that are used throughout the paper. These assumptions define "near exogeneity" and "weak identification". We also show examples that clarify these assumptions better. For each $n \in N$, let g_n be real-valued $q \times 1$ vector of functions, z_i represents the data set which is a $k \times 1$ real valued vector, for $i = 1, \cdots, n$. The following are the moment conditions that are satisfied at the true parameter value: θ_0

$$Eg_n(z_i, \theta_0) = \frac{C_1}{n^{1/2}}, \tag{1}$$

where C_1 is a constant vector. Instead of $g_n(z_i, \theta)$ we use $g_i(\theta)$ to save from notation. As in Stock and Wright (2000) specify g as a particular conditional moment restriction model: $g_i(\theta) = h(Y_i, \theta) \bigotimes x_i$, where x_i is the instrument vector.[1] So we have

$$Eg_i(\theta_0) = Eh(Y_i, \theta_0) \bigotimes x_i = \frac{C_1}{n^{1/2}}. \tag{2}$$

In the case of a simple linear system we can understand the above restriction much better. Let the structural equation be

$$y = Y\theta_0 + u, \tag{3}$$

and the reduced form is given by

$$Y = X\Pi + V, \tag{4}$$

where $y, u \in R^n, Y, V \in R^{n \times p}, X \in R^{n \times l}$, and $\Pi \in R^{l \times p}$. For simplicity let the matrix Y contains only endogenous variables. X is the set of instrumental variables. Let Y_i, V_i, X_i denote the rows of the matrices Y, V, X respectively. The data vector $z_i = (y_i, Y_i', X_i')'$ is $k = (l + p + 1) \times 1$. Then we can rewrite (2) for each $i = 1, 2, \cdots, n$ as

$$Eg_i(\theta_0) = Eu_i X_i = \frac{C_1}{n^{1/2}}.$$

So the instruments are weakly correlated with the structural error, but this converges to zero in the large samples. This is called "near exogeneity". Equation (2) defines "near exogeneity" in nonlinear moment restrictions.

[1] The results in this paper are also valid for $g_i(\theta) = h(Y_i, \theta) \bigotimes v(x_i)$ for any known function $v(.)$ of x_i.

Now we provide the assumptions.

Assumption 1. The true parameter $\theta_0 = (\alpha_0', \beta_0')'$ is in the interior of the compact parameter space $\Theta = A \times B$, $A \subset R^{p_a}, B \subset R^{p_b}$. θ is $p \times 1$, where $p = p_a + p_b$, and the data z_i is independent.

In order to understand the next assumption better, we benefit from the identity (equation (2.4)) in Stock and Wright (2000) but we take into account the "near exogeneity" problem as well as the weak identification. Define $\tilde{m}_n(\alpha, \beta) = En^{-1}\sum_{i=1}^n g_i(\theta)$. Then add and subtract to have the following identity:

$$\begin{aligned} \tilde{m}_n(\alpha, \beta) &= [\tilde{m}_n(\alpha, \beta) - \tilde{m}_n(\alpha_0, \beta)] \\ &+ [\tilde{m}_n(\alpha_0, \beta) - \tilde{m}_n(\alpha_0, \beta_0)] \\ &+ [\tilde{m}_n(\alpha_0, \beta_0)]. \end{aligned} \tag{5}$$

The first term in the square bracket above is set as in Stock and Wright (2000)

$$\tilde{m}_n(\alpha, \beta) - \tilde{m}_n(\alpha_0, \beta) = \frac{m_{1n}(\theta)}{n^{1/2}}.$$

Next, as in Stock and Wright (2000) set

$$\tilde{m}_n(\alpha_0, \beta) - \tilde{m}_n(\alpha_0, \beta_0) = m_{2n}(\beta).$$

Then in order to take into account near exogeneity we set

$$\tilde{m}_n(\alpha_0, \beta_0) = \frac{C_1}{n^{1/2}}. \tag{6}$$

Assumption 2.

$$En^{-1}\sum_{i=1}^n g_i(\theta) = \frac{m_{1n}(\theta)}{n^{1/2}} + m_{2n}(\beta) + \frac{C_1}{n^{1/2}}, \tag{7}$$

where $m_{1n}, m_1 : \Theta \to R^q$, $m_{2n}, m_2 : B \to R^q$ are continuous functions such that $m_{1n}(\theta) \to m_1(\theta)$ uniformly on Θ, $m_1(\theta_0) = 0$, and $m_{2n}(\beta) \to m_2(\beta)$ uniformly on B and $m_2(\beta) = 0$ iff $\beta = \beta_0$. Let $m_2 \in C^1(\mathcal{N})$ where $\mathcal{N} \subset \mathcal{B}$ denote an open neighborhood of β_0. Set $R_2(\beta) = \partial m_2(\beta)/\beta'$, which is a $q \times p_b$ matrix and $R_2(\beta_0)$ has full column rank, C_1 is a constant vector.

Now we consider how Assumption 2 works in linear case. So we analyze (3)-(4) with $Eu_iX_i = C_1/n^{1/2}$,

$$En^{-1}\sum_{i=1}^{n} g_i(\theta) = E(y_i - Y_i'\theta)X_i = Eu_iX_i + EX_iX_i'\Pi(\theta_0 - \theta),$$

where $\theta = (\alpha', \beta')$ and $\Pi = (\Pi_a, \Pi_b), \Pi_a = C_2/n^{1/2}, \Pi_b$ is a full-rank matrix. C_2 is $q \times p_a$ and Π_b is $q \times p_b$ matrix. So α is weakly identified, β is strongly identified. Assume $EX_iX_i' = Q_{xx} < \infty$. Then by using the terms in identity above assumption 2 to have

$$\frac{m_{1n}(\theta)}{n^{1/2}} = Q_{xx}\frac{C_2}{n^{1/2}}(\alpha_0 - \alpha),$$

$$m_{2n}(\beta) = Q_{xx}\Pi_b(\beta_0 - \beta),$$

$$\tilde{m}_n(\alpha_0, \beta_0) = Eu_iX_i = \frac{C_1}{n^{1/2}}.$$

Note that each one of these three terms above are right-hand side terms in (7). One thing to note is this formulation illustrates the tradeoff between near exogeneity and weak identification as well that we see in applications. When $n \to \infty$, $m_{1n}(\theta) \to 0$ rendering α unidentified in the large samples, but then the last term in (7) also converges to zero, achieving instrument exogeneity.

This formulation of Assumption 2 is different than the ones in Stock and Wright (2000) and Guggenberger and Smith (2003) in a way that we have the additional nearly exogenous term in (7), and this also clearly illustrates the tradeoff between weak instruments and near exogeneity problems in selection of instruments.

We need the following assumption for the consistency and limit distribution proof. This is taken from Guggenberger and Smith (2005). First define the following notation $\hat{\Omega}(\theta) = \sum_{i=1}^{n} g_i(\theta)g_i(\theta)'/n$ and $\Omega(\theta) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} Eg_i(\theta)g_i(\theta)'$. Note that Assumption 3 is Assumption M in Guggenberger and Smith (2005). A more primitive condition for Assumption 3i is mentioned in equation (2.4) of Guggenberger and Smith (2005) which involves the finiteness of certain moments of $g_i(\theta)$. The primitive conditions for the independent case in Assumption 3iii, which we consider is given in Pollard (1990) and given as a sub case of m-dependent one in Andrews (1994).

Assumption 3.

(i).

$$\max_i \sup_{\theta\in\Theta} \|g_i(\theta)\| = o_p(n^{1/2});$$

(ii). $\Omega(.)$ is in $C^0(A \times \{\beta_0\})$ and is bounded on Θ. $\Omega(\theta)$ is nonsingular for all $\theta \in A \times \{\beta_0\}$ and

$$\sup_{\theta\in\Theta} \|\hat{\Omega}(\theta) - \Omega(\theta)\| = o_p(1),$$

$$\sup_{A\times\mathcal{N}} \sum_{i=1}^{n} \frac{\|g_i(\theta)g_i(\theta)'\|}{n} = O_p(1).$$

(iii).

$$\Psi_n(\theta) \Longrightarrow \Psi(\theta),$$

where $\Psi_n(\theta) = n^{-1/2}\sum_{i=1}^n g_i(\theta) - Eg_i(\theta)$ and $\Psi(\theta)$ is a Gaussian process on Θ with mean zero and covariance function $E\Psi(\theta_1)\Psi(\theta_2)' = V(\theta_1, \theta_2)$. Specifically for each $\epsilon > 0$, there exists a $M_\epsilon < \infty$ such that $P(\sup_{\theta \in A \times \mathcal{N}} \|\Psi(\theta)\| \leq M_\epsilon) > 1 - \epsilon$.

Note that all the results in Assumption 3 are for triangular arrays (data), in order to save from notation we do not provide the additional subscript.

The following is standard in GEL literature, see Newey and Smith (2004), Guggenberger and Smith (2005).

Assumption 4.

(i). Function $\rho(.)$ is concave on $\mathcal{V}$.

(ii). $\rho(.)$ is C^2 in a neighborhood of 0 and $\rho_1 = \rho_2 = -1$, where ρ_1, ρ_2 are the first, and second partial derivatives of $\rho(.)$ evaluated at zero.

3 Estimation

In this section we derive the limit theory for the GEL estimators under weak identification and near exogeneity. First let $\hat{\lambda}$ denote

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

Then $\hat{\beta}(\alpha) = arg \min_{\beta \in B} \hat{P}(\alpha, \beta, \hat{\lambda}(\alpha, \beta))$, and $\hat{\alpha} = arg \min_{\alpha \in A} \hat{P}(\alpha, \hat{\beta}(\alpha), \hat{\lambda}(\alpha, \hat{\beta}(\alpha)))$, and $\hat{\beta} = \hat{\beta}(\hat{\alpha})$.

The following theorem extends the weak IV limit of GEL estimators in Guggenberger and Smith (2005) to joint weak IV and near exogeneity case. This is one of the main results of the paper.

Theorem 1. *Under Assumptions 1-4,*

$$(\hat{\alpha}, n^{1/2}(\hat{\beta} - \beta_0)) \Longrightarrow (\alpha^*, b^*),$$

where

$$b^* = -[R_2(\beta_0)'\Omega(\alpha^*, \beta_0)^{-1}R_2(\beta_0)]^{-1}R_2(\beta_0)'\Omega(\alpha^*, \beta_0)^{-1}[\Psi(\alpha^*, \beta_0) + m_1(\alpha^*, \beta_0) + C_1],$$

and $\alpha^* = arg \min_{\alpha \in A} P(\alpha)$, *where*

$$P(\alpha) = [\Psi(\alpha, \beta_0) + m_1(\alpha, \beta_0) + C_1]'M(\alpha, \beta_0)[\Psi(\alpha^*, \beta_0) + m_1(\alpha^*, \beta_0) + C_1],$$

and

$$M(\alpha, \beta_0) = \Omega(\alpha, \beta_0)^{-1} - \Omega(\alpha, \beta_0)^{-1}R_2(\beta_0)[R_2(\beta_0)'\Omega(\alpha, \beta_0)^{-1}R_2(\beta_0)]^{-1}R_2(\beta_0)'\Omega(\alpha, \beta_0)^{-1}.$$

Remark. The limit in Theorem 1 is similar to Theorem 2 in Guggenberger and Smith (2005) which is analagous to Theorem 1 in Stock and Wright (2000) for GMM. We obtain their limit when we set $C_1 = 0$. An interesting sub case here is the case of all weakly identified parameters with nearly exogenous moment restrictions. In that case the limit is (if we only have α which is weakly identified)

$$\hat{\alpha} \Longrightarrow \alpha^* = arg \min_{\alpha \in A} P(\alpha)$$

where

$$P(\alpha) = [\Psi(\alpha) + m_1(\alpha) + C_1]'\Omega(\alpha)^{-1}[\Psi(\alpha) + m_1(\alpha) + C_1].$$

This is an empirically important case since we observe there is a tradeoff between weak identification and near exogeneity. When the sample size gets large the assumptions show us there is no identification in large samples (i.e. $m_1(.)/n^{1/2} \to 0$) and there is exogeneity (i.e. $C_1/n^{1/2} \to 0$) in Assumption 2. This way of modeling reflects the idea that when we select the instruments, even though instrument may be weakly correlated with endogenous variable, it may not be correlated with the first order conditions.

Note that, estimation part does not change a lot of the results already established by Guggenberger and Smith (2005). However, it should be noted that under certain cases the bias term in Theorem 1 may be larger than the one found in their article. So the finite sample results as well as the large sample case may be affected when there is joint problem of weak instruments and near exogeneity.

Another important case to analyze is what if there are strong instruments and all the coefficients are strongly identified (i.e. Assume we have only βs in our system) but we have near exogeneity. In that case Assumption 2 is modified to

$$En^{-1}\sum_{i=1}^{n} g_i(\beta) = m_{2n}(\beta) + \frac{C_1}{n^{1/2}}.$$

Then the limit in Theorem 1 becomes

$$n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{d} b^*,$$

$$b^* = -[R_2(\beta_0)'\Omega(\beta_0)^{-1}R_2(\beta_0)]^{-1}R_2(\beta_0)'\Omega(\beta_0)^{-1}[\Psi(\beta_0) + C_1],$$

which is a normal limit with an additional drift C_1.

The most important results of the paper are provided in the inference and simulation sections. In those sections we see how the existing inference procedures in weakly identified GEL change asymptotically and in finite samples when we also have near exogeneity problem at hand. We also consider the case of only near exogeneity in the system without the weak instruments problem.

4 Inference

This section develops tests in the case of weak identification and near exogeneity (Assumption 2). First we take a look at the robust tests to weak identification problem, and try to see whether they are robust to near exogeneity as well.

The following assumptions are weaker than the ones used for estimation, and needed for inference.

Assumption 2^*.

$$En^{-1}\sum_{i=1}^{n} g_i(\theta_0) = \frac{C_1}{n^{1/2}}.$$

This is a modification of Assumption 2. We use (5)-(6) at θ_0 to have Assumption 2^*.

Assumption 3^*.

(i).

$$\max_i \|g_i(\theta_0)\| = o_p(n^{1/2});$$

(ii).

$$\hat{\Omega}(\theta_0) \xrightarrow{p} \Omega(\theta_0);$$

(iii).

$$\Psi_n(\theta_0) \xrightarrow{d} \Psi(\theta_0) \equiv N(0, \Omega(\theta_0)).$$

The results in Assumption 3^* are for triangular arrays. To save from notation we do not include the additional subscript. Note that in Assumption 3^*iii the limit variance is $\Omega(\theta_0)$ which in this case is equal to $V(\theta_0) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}[g_i(\theta_0) - Eg_i(\theta_0)][g_i(\theta_0) - Eg_i(\theta_0)]'$. This is because of (2). This last assumption is a triangular array central limit theorem result.

In this respect we analyze first an Anderson-Rubin (1949) like test proposed by Guggenberger and Smith (2005). This is called Generalized Empirical Likelihood Ratio statistic (GELR) and is given by

$$GELR(\theta_0) = n\hat{P}(\theta_0, \lambda(\theta_0)).$$

We test $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$. The following theorem is one of the main results of the paper. This result shows what happens to $GELR(\theta_0)$ test when there is near exogeneity. In this respect we extend the limit theory in Guggenberger and Smith (2005).

Theorem 2. *Under the null hypothesis of* $H_0 : \theta = \theta_0$ *by Assumptions 1,*2^**,*3^**,4,*

$$n\hat{P}(\theta_0, \lambda(\theta_0)) \xrightarrow{d} \chi_q^2(\delta),$$

where the distribution is noncentral chi-square with q degrees of freedom and noncentrality parameter $\delta = C_1'\Omega(\theta_0)^{-1}C_1$

Remark. This test statistic is robust to identification problem. This can be seen from analyzing Assumption 2*

$$En^{-1}\sum_{i=1}^{n} g_i(\theta_0) = \frac{C_1}{n^{1/2}}. \tag{8}$$

Regardless of the identification problems we have the same limit in Theorem 2 because of (8). We obtain the limit obtained by Theorem 3 of Guggenberger and Smith (2005) in the only weak identification case when we set $C_1 = 0$ here. In that case the limit is χ^2_q. However, the test statistic is affected by endogeneity problem in instruments in our case. We see that even though we operate under the null hypothesis we have a nonstandard distribution. This new limit depends on nuisance parameters. This is unusual since this is not a power study and we are not deriving the limits under the alternative.

Using the standard χ^2 limit in Guggenberger and Smith (2005) in the case of near exogeneity will cause problems in inference about θ_0. We see that when Guggenberger and Smith (2005) find the limit of GELR (θ_0), under the alternative $(\theta \neq \theta_0)$ in the case of weak identification, they have a similar result. Their result is a typical power result, it is not unusual that the limit in their case depends on nuisance parameters.

Now we consider a different test statistic. We will not use Assumption 3* in its derivation. We provide an assumption that will be useful in deriving the limit theory of this test statistic $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$. Before that we provide notation that is useful. Set $\partial g_i(\theta_0)/\partial\theta' = G_i(\theta_0) = (G_{iA}(\theta_0), G_{iB}(\theta_0))$ where $G_i(\theta_0)$ is $q \times p$ and $G_{iA}(\theta_0), G_{iB}(\theta_0)$ are $q \times p_A, q \times p_B$ respectively. The following Assumption replaces Assumption 3* for the test statistic that follows.

Assumption 5.

(i). $\max_i \|g_i(\theta_0)\| = o_p(n^{1/2})$.

(ii).

$$\frac{1}{n}\sum_{i=1}^{n} g_i(\theta_0)g_i(\theta_0)' \xrightarrow{p} \Omega(\theta_0),$$

where $\Omega(\theta_0)$ is nonsingular.

$$\frac{1}{n}\sum_{i=1}^{n} \|g_i(\theta_0)g_i(\theta_0)'\| = O_p(1).$$

(iii).

$$\Psi_n(\theta_0) \xrightarrow{d} \Psi(\theta_0) \equiv N(0, \Omega(\theta_0)),$$

because of (2).

(iv). $\hat{g}(\theta)$ is differentiable at θ_0 almost surely, $\hat{g}(\theta_0)$ is integrable, and $\hat{G}(\theta_0) = \frac{1}{n}\sum_{i=1}^{n} \partial g_i(\theta_0)/\partial\theta'$ is integrable, $m_{1n} \in C^1(\theta_0)$ and $R_{1n}(\theta_0) = \partial m_{1n}(\theta_0)/\partial\theta'$ converges to some function $R_1(\theta_0)$.

(v).

$$\frac{1}{n}\sum_{i=1}^{n}(vecG_{iA}(\theta_0))g_i(\theta_0)' \xrightarrow{p} \Delta_A,$$

where Δ_A is defined in (vii);

$$\frac{1}{n}\sum_{i=1}^{n} G_{iB}(\theta_0) \xrightarrow{p} R_2(\beta_0).$$

(vi).

$$\frac{1}{n}\sum_{i=1}^{n} \|G_{iA}(\theta_0)\|\|g_i(\theta_0)\| = O_p(1),$$

$$\frac{1}{n^{3/2}}\sum_{i=1}^{n} \|G_{iB}(\theta_0)\|\|g_i(\theta_0)\| = o_p(1).$$

(vii).

$$\frac{1}{n^{1/2}}\sum_{i=1}^{n}[(vecG_{iA}(\theta_0) - EvecG_{iA}(\theta_0))', (g_i(\theta_0) - Eg_i(\theta_0))']' \xrightarrow{d} N(0, V_1(\theta_0)),$$

$$V_1(\theta_0) = \begin{pmatrix} \Delta_{AA} & \Delta_A \\ \Delta_A' & \Omega(\theta_0) \end{pmatrix},$$

and each block matrix above has full column rank, $V_1(\theta_0)$ is positive definite. Also,

$$V_1(\theta_0) = \lim_{n\to\infty} var\left[n^{-1/2}\sum_{i=1}^{n}((vecG_{iA}(\theta_0))', g_i(\theta_0)')'\right],$$

the dimension of $V_1(\theta_0) : (qp_A + q) \times (qp_A + q)$.

These Assumptions are providing CLT and WLLN type of results for the function $g(.)$ and its derivatives. Assumption 5iv allows us to change the order the of integration and differentiation in Assumption 2. So we have

$$E\hat{G}(\theta_0) = n^{-1/2}R_{1n}(\theta_0) + (0, R_2(\beta_0)) \to (0, R_2(\beta_0)). \tag{9}$$

The results in Assumption 5 are for triangular arrays (data), to save from notation we do not include the additional subscript. Assumption 5v-vii are the ones that are used in Assumption M_θ in Guggenberger and Smith (2005) and Assumption 1 in Kleibergen (2002). They are technical conditions that are helpful in deriving the limit. Now we start defining the tests that we consider. Kleibergen (2002) type of tests in weakly identified GEL is derived by Guggenberger and Smith (2005). They depend on the first order condition for θ in GEL,

$$\lambda(\hat{\theta})'\sum_{i=1}^{n} \rho_1(\lambda(\hat{\theta})'g_i(\hat{\theta}))G_i(\hat{\theta})/n = 0.$$

We can further rewrite the first order condition as

$$\lambda(\hat{\theta})'D(\hat{\theta}) = 0,$$

where

$$D(\hat{\theta}) = \sum_{i=1}^{n} \rho_1(\lambda(\hat{\theta})' g_i(\hat{\theta})) G_i(\hat{\theta})/n,$$

and it is $q \times p$ matrix. Since $\hat{\theta}$ is inconsistent usage of $\hat{\theta}$ in the test statistic does not result in nuisance parameter free limits. Kleibergen (2002) suggests usage of θ_0 in test statistics. In this way in weakly identified GEL and GMM we derive χ^2 limits. First define the normalized version of $D(\theta_0)$,

$$D^*(\theta_0) = D(\theta_0)\Lambda, \tag{10}$$

where $\Lambda = diag(n^{1/2}, n^{1/2}, \cdots, 1, 1, \cdots 1)$ is a $p \times p$ diagonal matrix with first p_A diagonal elements are $n^{1/2}$ and remaining p_B are 1. This is done because of weakly identified α_0 need a different normalization than β_0 in $D(\theta_0)$ definition. Guggenberger and Smith (2005) suggest the following test statistics:

$$S(\theta_0) = n\lambda(\theta_0)' D^*(\theta_0)[D^*(\theta_0)'\hat{\Omega}(\theta_0)^{-1} D^*(\theta_0)]^{-1} D^*(\theta_0)'\lambda(\theta_0), \tag{11}$$

and asymptotically equivalent

$$LM(\theta_0) = n\hat{g}(\theta_0)'\hat{\Omega}(\theta_0)^{-1} D^*(\theta_0)[D^*(\theta_0)'\hat{\Omega}(\theta_0)^{-1} D^*(\theta_0)]^{-1} D^*(\theta_0)'\hat{\Omega}(\theta_0)^{-1}\hat{g}(\theta_0). \tag{12}$$

To get (12) from (11) we use $\hat{\lambda}(\theta_0) = -\hat{\Omega}(\theta_0)^{-1}\hat{g}(\theta_0) + o_p(1)$ which can be seen from the proof of Theorem 1. Note that the form of the test statistics $S(\theta_0), LM(\theta_0)$ in (11)-(12) are needed to derive the limits. Practitioners can use the following form (exactly the same statistics because of (10)

$$S(\theta_0) = n\lambda(\theta_0)' D(\theta_0)[D(\theta_0)'\hat{\Omega}(\theta_0)^{-1} D(\theta_0)]^{-1} D(\theta_0)'\lambda(\theta_0), \tag{13}$$

and

$$LM(\theta_0) = n\hat{g}(\theta_0)'\hat{\Omega}(\theta_0)^{-1} D(\theta_0)[D(\theta_0)'\hat{\Omega}(\theta_0)^{-1} D(\theta_0)]^{-1} D(\theta_0)'\hat{\Omega}(\theta_0)^{-1}\hat{g}(\theta_0). \tag{14}$$

The following Theorem is one of the main results of the paper. We extend the limits of $S(\theta_0), LM(\theta_0)$ from the case of weakly identified GEL to the joint problem of nearly exogenous and weakly identified GEL.

Theorem 3. *Suppose that at $\theta_0 = (\alpha_0', \beta_0')'$ Assumptions 1,2*,4,5 hold. Then under the null hypotheses $H_0 : \theta = \theta_0$,*

$$S(\theta_0), LM(\theta_0) \xrightarrow{d} [W(C_1) + \zeta]'[W(C_1) + \zeta],$$

where the random $W(C_1)$ is defined in equation (34) of Appendix. $\zeta \equiv N(0, I_p)$ and $W(C_1)$ and ζ are independent.

Remarks. This limit shows that when there is no near exogeneity problem (i.e. $C_1 = 0$) we have the limit (Theorem 4) in Gugggenberger and Smith (2003): χ_p^2 distribution, since from (34) $W(C_1) = 0$ when $C_1 = 0$. As a more important remark, we observe that in the case of near exogeneity and weak identification the limit of $S(\theta_0), LM(\theta_0)$ tests are no longer nuisance parameter free. This is in contrast to only weak identification case in Guggenberger and Smith (2005). So the limits of the tests are affected by near exogeneity. So using χ^2_p critical values will be misleading if the first order conditions are affected by near exogeneity.

This result shows that under the null hypotheses, the limit is not pivotal. This is important because we do not assume to operate under the alternative $\theta \neq \theta_0$. In misspecification literature, this idea is used a lot, the limits are derived for the case of $\theta \neq \theta_0$. We also observe that the limits of $GELR(\theta_0)$ and $S(\theta_0), LM(\theta_0)$ tests differ under near exogeneity.

We now provide an overidentifying restrictions test. Caner (2004) provides an overidentifying restrictions test in Exponential Tilting framework with weak identification. A similar version that is provided here for GEL estimators may detect near exogeneity.

$$J(\theta_0) = n\hat{g}(\theta_0)'\hat{\Omega}(\theta_0)^{-1/2} M_{\hat{\Omega}(\theta_0)^{-1/2}D^*(\theta_0)} \hat{\Omega}(\theta_0)^{-1/2}\hat{g}(\theta_0),$$

where $M_X = I - X(X'X)^{-1}X'$.

Theorem 4. *Under Assumptions 1,2*,4,5 at* $\theta_0 = (\alpha_0', \beta_0')'$,

$$J(\theta_0) \xrightarrow{d} (\Xi + \Omega(\theta_0)^{-1/2}C_1)' M_{\Omega(\theta_0)^{-1/2}\bar{D}} (\Xi + \Omega(\theta_0)^{-1/2}C_1),$$

where $\Xi \equiv N(0, I_q)$, *and* $\bar{D}$ *is the limit for* $D^*(\theta_0)$ *which is shown in the proof of Theorem 3.*

Remark. Note that when $C_1 = 0$, (which is the null hypotheses for J-test) the limit becomes χ^2_{q-p}, since $n^{1/2}\hat{g}(\theta_0)$ is asymptotically independent from $D^*(\theta_0)$ as seen in the proof of Theorem 3. So J test detects near exogeneity when $C_1 \neq 0$. Note that J test is asymptotically independent of $LM(\theta_0)$ test . This is derived from the proof of Theorem 3.

Note that based on the J-test result we can go and select new instruments that will not have near exogeneity problem. Then to test $\theta = \theta_0$, we use $GELR(\theta_0), S(\theta_0), LM(\theta_0)$ tests. But pretesting with J-test may bring some problems into second stage of testing $H_0 : \theta = \theta_0$. Another possibility is sometimes it is not easy to find pure exogenous instruments. So what we provide and recommend is the usage of $GELR(\theta_0), S(\theta_0), LM(\theta_0)$ tests that will use critical values which take into account near exogeneity problem. So we do not have to change our data set. Subsampling approach to these test statistics provides data based critical values. Subsampling is also consistent even when there are weak exogeneity and weak identification. The next section reviews subsampling and derives the subsampling distribution of the test statistics $GELR(\theta_0), LM(\theta_0), S(\theta_0)$.

5 Subsampling

Note that $GELR(\theta_0), S(\theta_0), LM(\theta_0)$ test statistics are robust to identification problem, when there is no near exogeneity (i.e. $C_1 = 0$). In that case the limits are χ^2_q for $GELR(\theta_0)$, and χ^2_p for

$S(\theta_0), LM(\theta_0)$. This can be seen from our Theorems 2 and 3 (when $C_1 = 0$) or by Guggenberger and Smith (2005).

However, these test statistics are not robust to even small correlation between the instruments and the first order conditions. In a linear model this will show itself as the correlation between the structural equation error and the instrument selected. These tests are size-distorted in the realistic case of picking the "not so perfect" instrument for an empirical study. As we show below even when there is near exogeneity ($C_1 \neq 0$) the tests will have correct size when we subsample the test statistics. The subsampling approach gets the correct critical values regardless of $C_1 = 0$ or $C_1 \neq 0$ and also works in the case of identification problems. So we can use the critical values obtained from the empirical quantiles of subsample test statistics. Subsample test statistic is obtained by evaluating test statistic on blocks of data, when the block size is small compared to all of the data set.

The testing case is mentioned in Chapter 4.5 of Politis, Romano and Wolf (1999), this is not written formally there but it is an easy extension of their ideas. [2]. So we show this for our test statistics only. Our Theorem is specific to $GELR(\theta_0), S(\theta_0), LM(\theta_0)$ tests. It does not cover general testing via subsamples. So subsample tests we use will also be evaluated at θ_0. We do not use estimators $\hat{\theta}$ in subsampling tests here.

The following explanations and notation are largely borrowed from in subsampling via hypothesis testing in Politis Romano and Wolf (1999) . Let $z_{n1}, z_{n2}, \cdots, z_{nn}$ be a sample of n independent observations in a triangular array format The common unknown distribution is denoted by $\mathcal{P}$. This unknown law $\mathcal{P}$ is assumed to belong to a certain class of laws $\mathbf{P}$. The null hypothesis is $H_0 : \mathcal{P} \in \mathbf{P_0}$ and the alternative is $H_1 : \mathcal{P} \in \mathbf{P_1}$, where $\mathbf{P_0} \cup \mathbf{P_1} = \mathbf{P}$. Let $\mathcal{P}_n$ be contiguous to $\mathcal{P} \in \mathbf{P_0}$. $\mathcal{P}_n$ represents the probability distribution regarding $Eg_i(\theta_0) = C_1/n^{1/2}$ (sequences), and $\mathcal{P}$ represents the probability distribution regarding C_1 (nuisance parameter). Contiguity plays an important role in the proof. It enables us to work with $\mathcal{P}$ and then at the end using contiguity we show that results hold for $\mathcal{P}_n$.

We want to have a test which its asymptotic rejection probability under the null hypotheses is α (the exact level). Let T_n represents test statistics: $GELR(\theta_0), S(\theta_0), LM(\theta_0)$. Denote these tests more specifically as

$$T_n = \tau_n t_n(z_{n1}, z_{n2} \cdots, z_{nn}),$$

where τ_n can be thought as the rate of convergence, $\tau_n \to \infty$, as $n \to \infty$.

The corresponding cumulative distribution function

$$G_n(z, \mathcal{P}) = Prob_{\mathcal{P}}(T_n(z_{n1}, z_{n2}, \cdots, z_{nn}) \leq z).$$

[2]We thank Joseph Romano for pointing out that triangular array extension can be done with ease after the material in Politis Romano and Wolf (1999).

We now describe the subsampling test construction. First, let $Y_1, Y_2, \cdots, Y_{N_n}$ represent the $N_n = \begin{pmatrix} n \\ b \end{pmatrix}$ subsets of $\{z_{n1}, z_{n2}, \cdots, z_{nn}\}$ ordered in any fashion. Each Y_j, $j = 1, 2, \cdots, N_n$ represents a block of $z's$ with size b. Let $t_{n,b,j}$ be the "$t''_{n,b}$ evaluated at block Y_j. The sampling distribution of T_n is approximated by

$$\hat{G}_{n,b}(z) = N_n^{-1} \sum_{j=1}^{N_n} 1_{\{\tau_b t_{n,b,j} \leq z\}}.$$

As mentioned in section 2.4 of Politis, Romano, and Wolf (1999), instead of N_n number of subsamples we can do the following. We use "n-b+1" subsamples of size b of the form $\{z_{ni}, z_{ni+1}, \cdots, z_{ni+b-1}\}$ Order of the data is fixed and retained in the subsamples. This approach is used for computational purposes and also suggested for both iid and non iid contexts (p.52, Politis, Romano, and Wolf (1999)).

The critical value of the subsampling test is the $1-\alpha$ quantile of $\hat{G}_{n,b}$, specifically

$$c_{n,b}(1-\alpha) = \inf\{z : \hat{G}_{n,b}(z) \geq 1-\alpha\}. \tag{15}$$

Let $G_{b,j}(\mathcal{P}_n)$ be the sampling distribution of subsample test $T_{n,b,j} = \tau_b t_{n,b,j}$, $j = 1, 2, \cdots, n - b + 1$. Let $G_n(\mathcal{P}_n)$ be the sampling distribution of the test statistic itself.

The cumulative distribution function of $T_{n,b,j}$ is

$$G_{b,j}(z, \mathcal{P}_n) = Prob_{\mathcal{P}_n}\{\tau_b t_{n,b,j} \leq z\}. \tag{16}$$

Before writing the theorem we should note that the nominal level α test rejects H_0 if and only if

$$T_n > c_{n,b}(1-\alpha). \tag{17}$$

The $1-\alpha$ quantile of the limit of the tests is for $\mathcal{P} \in \mathbf{P_0}$

$$c(1-\alpha, \mathcal{P}) = \inf\{z : G(z, \mathcal{P}) \geq 1-\alpha\}.$$

The following Theorem benefits from Theorems 2.6.1 and 4.2.1 of Politis, Romano and Wolf (1999).

Now we provide one of the main contributions of the paper. Theorem 5 can be used to show that subsampling test is robust to near exogeneity problem. Since the test statistics, are robust to identification problems already, we can recover the limits in Theorems 2 and 3 via subsampling when there are near exogeneity and weak identification. So subsampling test statistics are important from the empirical point of view, since researchers can face these problems simultaneously in data. This is a new result in this literature. Subsampling provides a solution to an empirically very important problem.

Theorem 5.

(a). *Under Assumptions 1,2***,3***,4 and under the null hypothesis of* $H_0 : \theta = \theta_0$, *assuming* $b/n \to 0$, *as* $b \to \infty$, *when* $n \to \infty$, *and* $\tau_b/\tau_n \to 0$ *for* $GELR(\theta_0)$ *we have for* $\mathcal{P} \in \mathbf{P_0}$

$$c_{n,b}(1-\alpha) \xrightarrow{p} c(1-\alpha, \mathcal{P}), \quad (in\, \mathcal{P}_n\, probability)$$

and

$$Prob_{\mathcal{P}_n}\{T_n > c_{n,b}(1-\alpha)\} \to \alpha$$

as $n \to \infty$.

(b) The same result in (a) holds for , $S(\theta_0), LM(\theta_0)$ *under Assumptions 1, 2***, 4, 5.*

Remark. Theorem shows that the subsample tests are consistent under the null, whether there is near exogeneity or not. Furthermore, since these test work under various identification issues without subsampling them, they are also robust to identification problem. Guggenberger and Smith (2005) show that $GELR(\theta_0), S(\theta_0), LM(\theta_0)$ are robust to identification problem when there is no near exogeneity. However, as shown in Theorems 2 and 3 these test statistics' limit depend on C_1 when there is near exogeneity. By introducing subsampling method here we show that these tests are robust to weak exogeneity problem when use the subsample critical values. This is one of the contributions of this article. In this respect, Theorem 5 shows that even though there may be small correlation between the instruments and the first order conditions, and there also may be low correlation between instruments and the endogenous variables, we can still make inference.

We should also note that our problem is peculiar in a way that subsampling can be applied. It is clear from Andrews (2000) that we can not apply subsampling to the problems that involve parameter estimation, where the parameter is local to true value. An example is mean estimation. If true mean has the structure $\mu_n = \mu/n^{1/2}$ and we estimate μ_n by $\hat{\mu}_n$, the limit distribution of $n^{1/2}(\hat{\mu}_n - \mu_n)$ depends on μ. However, since this parameter μ is not consistently estimable, subsampling fails.

In our problem, the test statistics depend on true value θ_0. There is no estimation of θ in our subsampling. We just compute the test statistic on subsamples. Then our limit is continuous on C_1. Also we do not need to estimate C_1. Our test statistics are continous on C_1. The results that we have is true both pointwise and uniformly over C_1. These points are illlustrated in the proof of Theorem 5, especially in equations (35)-(38). Politis, Romano and Wolf (1999) analyze the power property of test against a local alternative in Theorem 2.6.1iii. They also show that subsampling is consistent in that scenario. The bootstrap fails in our case, since it uses $\hat{\theta}$ for θ_0 which is inconsistent due to weak identification.

5.1 Choice of Data Dependent Block Size

An important part of the subsampling method is the choice of the block size. In this part, we mainly benefit from section 9.4.1 in Politis, Romano and Wolf (1999). We propose using two different methods to choose the block size. The first one is calibration method. The second one is minimum volatility method.

We use calibration method an idea proposed by Loh (1987). This is called "calibration by adjusting the block size" in sections 9.3.1 and 9.4.1 in Politis, Romano and Wolf (1999). This is used in subsampling linear hypotheses in the case of identification failures for GMM in Guggenberger and Wolf (2004). We briefly give the intuition behind the method and then provide a formal description of the algorithm. The nominal size of the test is α. The actual size is γ and different from α. Calibration method adjusts the block size such that actual size γ will be close to α. The calibration function "h" maps the block size into actual size of the test. So if we knew "h" than we could have adjusted block size "b" to obtain approximately α. However, we do not know "h". Since we also do not know the probability distribution $\mathcal{P}_n$ we can not simulate "h". By generating pseudo data from a suitable distribution P_n^* that imposes the null (i.e. $H_0 : \theta = \theta_0$), we can simulate "h" we can find the block size "b" that makes h(b) near α as much as possible by trying various b levels. Note that P_n^* may be a parametric model imposing the null as described in Remark 9.4.2 in Politis, Romano, and Wolf (1999). So we generate the pseudo data from a parametric model imposing the null: $H_0 : \theta = \theta_0$. This is plausible since the test statistics depend on θ_0. Similar idea for linear inference in identification failure problems in GMM is proposed in section 4.1 of Guggenberger and Wolf (2004). Now we provide the algorithm.

1. Have a simple grid of block sizes of b, i.e. $b \in [b_{small}, b_{big}]$, where b_{small}, b_{big} indicate the smallest and largest values of the grid respectively.

2. Generate K pseudo sequences $z_1^{*k}, \cdots, z_n^{*k}$, $k = 1, \cdots, K$ according to a a suitable distribution P_n^* that imposes the null hypothesis. For each sequence k, and for each block size "b" in the grid, carry out a test of significance level α, as in (17). Denote each rejection as $\phi_{k,b} = 1$, and $\phi_{k,b} = 0$ otherwise.

3. Set $\hat{h}(b) = K^{-1} \sum_{k=1}^{K} \phi_{k,b}$.

4. Use the block size $b^* = \arg\min_b |\hat{h}(b) - \alpha|$.

In practice $K \geq 1000$ recommended.

Remarks.

1. This is not the algorithm in section 9.4.1 in Politis, Romano, and Wolf (1999) but a modification of it as described in Remark 9.4.1 in Politis, Romano, and Wolf (1999).

2. This is not the algorithm used in Guggenberger and Wolf (2004) but this variant is mentioned in section 4.1 of their paper. We do not use bootstrap distribution for P_n^* since in the case of weak identification bootstrap is not consistent, see Dufour (1997). Instead we benefit from Remark 9.4.2 in Politis, Romano, and Wolf (1999) and have P_n^* as a parametric model imposing the null hypothesis.

3. Since the block size is data dependent, this may affect asymptotics described in the previous theorem. However, as it can be seen from the proof of Theorem 2.7.1 of Politis, Romano, and Wolf

(1999) if $b_{small} \to \infty$, $b_{big} = o(n)$, then asymptotics remain intact. The critical issue in that proof is the usage of Hoeffding's exponential inequality as in (40) which is also valid for independent random variables.

One of the reasons of the usage of calibration method is certain optimality properties (p.197, Politis, Romano, and Wolf (1999)). A much general but a heuristic way is the minimum volatility method. This is described in section 9.4.2 of Politis, Romano and Wolf (1999) as an alternative method to calibration. We first give a simple description. The main idea is for a range of block sizes "b" we see that rejection probability for the null hypothesis will not change so much, then according to some arbitrary criterion we pick the best possible b among the reasonable block sizes. The algorithm is as follows.

1. For a grid of values for b, from b_{small} to b_{big} compute a subsampling quantile $c_{n,b}(1-\alpha)$ as defined in (15) for the desired level α.

2. Smooth the quantiles using a running mean of span m. So we replace $c_{n,b}(1-\alpha)$ by the average of the values $\{c_{n,b-m}(1-\alpha), \cdots, c_{n,b+m}(1-\alpha)\}$.

3. For each "b" compute a volatility index VI_b as the standard deviation of the quantiles in a neighborhood of b. More specifically, for a small integer "s", let VI_b be equal to the standard deviation of the values $\{c_{n,b-s}(1-\alpha), \cdots, c_{n,b+s}(1-\alpha)\}$ (remember that these critical values are already smoothed in step 2). VI_b is the volatility index.

4. Pick b^* corresponding to the smallest volatility index and use $c_{n,b^*}(1-\alpha)$ as the critical value of the test.

Remarks.

1. To make the algorithm computationally efficient we include every other b or every third b between the b_{small} and b_{big}.

2. Simulations in Politis, Romano, and Wolf (1999) show that the algorithm is insensitive to the choice of s and m, we employ $s = m = 1$.

6 Many Weak Moment Asymptotics

In this part we derive the limit for GEL estimators under many weak moment asymptotics subject to near exogeneity. Now define GEL estimator in the following way as in Newey and Windmeijer (2005).

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \sup_{\lambda \in \hat{\Lambda}_n(\theta)} \sum_{i=1}^{n} \rho(\lambda' g_i(\theta))/q_n,$$

where $g_i(\theta) : q_n \times 1$ unlike section 2, and q_n increases with n, and the relationship between q_n and n will be made specific below in Assumptions. But $q_n/n \to 0$. So q_n will grow slower than n. This is the approach taken by Newey and Windmeijer (2005) to control variance. The data Z_i is iid. The domain of λ is the same as in section 2, but $\lambda : q_n \times 1$. Θ is a compact subset in R^p. We normalize $\rho(\nu)$, $\nu \in \mathcal{V}$ such that $\rho(0) = 0, \partial\rho(0)/\partial\nu = 1, \partial^2\rho(0)/\partial\nu^2 = -1$. $\mathcal{V}$ is explained at the beginning of section 2.

Denote

$$\hat{Q}(\theta) = \sup_{\lambda \in \hat{\Lambda}_n(\theta)} \sum_{i=1}^{n} \rho(\lambda' g_i(\theta))/q_n.$$

Then we can write

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \hat{Q}(\theta).$$

We define $\hat{\theta}$ slightly differently in section 2. That form helps us in inference in the case of Anderson-Rubin (1949) tests. Since we do not have Anderson-Rubin (1949) type of tests in this section we use the more convenient form in this section. In both sections note that $\hat{\theta}$ is exactly the same.

Many weak moment asymptotics in GEL is derived by Newey and Windmeijer (2005). We try to show what happens to many weak moment asymptotics under near exogeneity. Since near exogeneity with many weak moments is a realistic setup in empirical work in labor economics as discussed in Bound et. al (1994) we pursue this here. Stock, Wright and Yogo (2002) also discuss the need to address the same problem.

Newey and Windmeijer (2005) have detailed explanations why GEL estimator is consistent under many weak asymptotics whereas GMM may not be. The limit of the objective function in GMM consists of a "noise" term and "signal" term. Noise term consists of weight matrix multiplied by $\Omega(\theta) = Eg_i(\theta)g_i(\theta)'$. This noise does not disappear and contaminates the limit and hence leads to inconsistency. This is shown first by Han and Phillips (2004) and then by Newey and Windmeijer (2005). However, CUE in GEL does not have a noise term in the limit since weight matrix is $\Omega(\theta)^{-1}$, and noise disappears. CUE estimator is consistent under many weak moment asymptotics. Since Newey and Windmeijer (2005) show GEL objective function is well approximated by CUE, any GEL estimator is also consistent. Compared to fixed weak moment asymptotics as in Guggenberger and Smith (2005), Stock and Wright (2000), and Caner (2004), even though the moments decay to zero at $n^{1/2}$, their numbers (q_n) increase. Hence information increases and we obtain consistency.

Here we provide the near exogeneity assumption first.

Assumption M1.

$$Eg_i(\theta_0) = \frac{C_1}{n^{1/2}},$$

where C_1 is a $q_n \times 1$ vector, and $C_1 = (0'_{q_n-l}, C'_l)'$, C_l is an $l \times 1$ vector of nonzero scalars, l is a fixed number, it does not increase with n.

This Assumption mainly shows that the near exogeneity manifests itself in fixed number of moments not growing with n. At the end of this section we also discuss the case of C_1 a vector of $q_n \times 1$ with all nonzero cells. Assumption M1 is in line with empirical studies such as Angrist and

Krueger (1991), Bound et. al (1994). Bound et.al (1994) show that Angrist and Krueger (1991) paper suffers from near exogeneity problem.

We provide high level assumptions that are used in Newey and Windmeijer (2005). These are Assumptions 2, 3, and 3a in their article.

Assumption M2. *There is a continuous function $\Delta(a)$ such that $\Delta(a) > 0$ for all $a \neq 0$, and $S_n(\theta) \geq \Delta(\|\theta - \theta_0\|)$ where*

$$S_n(\theta) = \frac{n}{q_n}[Eg_i(\theta)]'\Omega(\theta)^{-1}[Eg_i(\theta)].$$

This assumption is an identification, and uniqueness condition for θ_0. This is Assumption 2 in Newey and Windmeijer (2005). The following Assumption M3i-v, and Assumption M3v is Assumption 3 and Assumption 3a respectively in Newey and Windmeijer (2005). Let "Eig" represent the eigenvalue.

Assumption M3.

i) $\theta_0 \in \Theta$ with Θ compact, there is a constant C with $Eig_{min}(\Omega(\theta)) \geq \frac{1}{C}$ and $Eig_{max}(\Omega(\theta)) \leq C$ for each θ.

ii)

$$E[g_i(\theta)'g_i(\theta)]^2/q_n^2 n \to 0,$$

for each $\theta \in \Theta$.

iii)

$$\sup_{\theta \in \Theta} \|\hat{\Omega}(\theta) - \Omega(\theta)\| \xrightarrow{p} 0,$$

for $\hat{\Omega}(\theta) = \frac{1}{n}\sum_{i=1}^n g_i(\theta)g_i(\theta)'$.

iv) $S_n(\theta)$ is equicontinuous, $n/q_n\hat{g}(\theta)'\Omega(\theta)^{-1}\hat{g}(\theta)$ is stochastically equicontinuous, $\hat{g}(\theta) = \sum_{i=1}^n g_i(\theta)/n$.

v) $\rho(\nu)$ is three times continuously differentiable on $\mathcal{V}$, and there is $\gamma > 2$ such that

$$n^{1/\gamma}E[\sup_{\theta} \|g_i(\theta)\|^{\gamma}]^{1/\gamma}\frac{q_n^{1/2}}{n^{1/2}} \to 0.$$

Assumption M3 is similar to Assumptions 2-4 in fixed weak moment asymptotics case analyzed in this paper. The primitives in linear case are shown in Newey and Windmeijer (2005). For the nonlinear case both Newey and Windmeijer (2005) and Han and Phillips (2004) use these high level conditions. Stochastic equicontinuity primitives can be provided by using section 2.11.3 in van der Vaart and Wellner (1996).

The key assumption here is Assumption M3iv. The equicontinuity property of $S_n(\theta)$ defines basically the many weak moment idea. Given the definition of $S_n(\theta)$ in Assumption M2, the moments decay at rate $n^{1/2}$ like the fixed weak moment case, however since their number increases with sample size we have to divide $Eg_i(\theta)'\Omega(\theta)^{-1}Eg_i(\theta)$ by q_n to control that. Assumption M3iv is also needed for the uniform convergence results, as in Newey and Windmeijer (2005) and Han and Phillips (2004). Assumptions M3i, M3iii are standard assumptions. Assumption M3iii can give us the relation between q_n and n in certain cases. For example if $Eg_{ij}(\theta)^2$ is uniformly bounded in j, q_n, and θ, then sufficient condition for Assumption to hold is $q_n^2/n \to 0$. Assumption M3ii is needed to control the growth of $E(g_i(\theta)'g_i(\theta))^2$. Assumption M3v is standard in GEL literature, see Newey and Smith (2004). The only difference in many weak moment case of Newey and Windmeijer (2005) is multiplication by square root of q_n. This is needed to get consistency and rate of concvergence of lagrange multiplier.

Now we provide the consistency result.

Theorem 6. *Under Assumptions M1-M3,*

$$\hat{\theta} \xrightarrow{p} \theta_0.$$

This shows that even with a near exogeneity problem we can achieve consistency. In Theorem 1 in this paper, this is not the case when there are fixed number of weak moments. The main reason that we have consistency is the availability of many moment conditions which provide information. Also corruption of information (near exogeneity) in Assumption M1 is limited.

We now analyze the limit of GEL estimator under many weak moments and near exogeneity. We want to provide some intuition about the results. Near exogeneity is defined in Assumption M1. The limit depends on the partial derivative of the objective function

$$q_n^{1/2}(\hat{\theta} - \theta_0) = -\left[\frac{\partial^2 \hat{Q}(\bar{\theta})}{\partial\theta\partial\theta'}\right]^{-1} q_n^{1/2}\frac{\partial \hat{Q}(\theta_0)}{\partial\theta}, \tag{18}$$

where $\bar{\theta} \in (\theta_0, \hat{\theta})$. The crucial part in deriving the limit is the following expansion

$$q_n^{1/2}\frac{\partial \hat{Q}(\theta_0)}{\partial\theta} = [G_n'\hat{\Omega}^{-1}n^{1/2}\hat{g}] + [n^{1/2}\hat{U}'\hat{\Omega}^{-1}\frac{n^{1/2}\hat{g}}{q_n^{1/2}}], \tag{19}$$

which is proven in the appendix. To remind notation $G_n = \frac{n^{1/2}}{q_n^{1/2}}E\frac{\partial g_i(\theta_0)'}{\partial\theta}$: $q_n \times p$ matrix, $\hat{\Omega} = \sum_{i=1}^n g_i(\theta_0)g_i(\theta_0)'/n$, $\hat{g} = \sum_{i=1}^n g_i(\theta_0)/n$. $\hat{U}$ is the matrix of residuals from regressing the derivatives on the moments. The expression for this is supplied in the appendix, before Lemma A.7.

These two terms on the right hand side of (19) converge to a normal distribution. They are independent from each other. We see in the proofs that only the first term is affected by near exogeneity, the second term is unaffected by this important problem in data. This is due to increasing number of moments (i.e.$q_n \to \infty$) and C_1 containing only fixed number of nonzero covariances as described in Assumption M1. The following Theorem, in that sense brings a new result into the literature and shows the value of increasing moment conditions as long as dimension of near exogeneity is fixed (i.e. $C_1 = (0'_{q_n - l}, C'_l)'$). Now we provide assumptions that are needed to provide the limits of estimators.

These are used in many weak moment asymptotics GEL in Newey and Windmeijer (2005).

Assumption M4.

(i). $g(\theta)$ is twice continuously differentiable in a neighborhood N of θ_0.

(ii).

$$E\|g_i(\theta_0)\|^4(\frac{q_n}{n} + \frac{1}{q_n n^{1/2}}) \to 0,\; E\|\partial g_i(\theta_0)/\partial\theta_0\|^4(\frac{q_n}{n} + \frac{1}{q_n n^{1/2}}) \to 0.$$

(iii). For all $\theta \in \mathcal{N}$ we have

$$Eig_{max}(E[\frac{\partial g_i(\theta)}{\partial\theta_j}\frac{\partial g_i(\theta)}{\partial\theta_j}'] \leq C,$$

$$Eig_{max}(E[\left(\frac{\partial^2 g_i(\theta)}{\partial\theta_j\partial\theta_k}\right)\left(\frac{\partial^2 g_i(\theta)}{\partial\theta_j\partial\theta_k}\right)']) \leq C,$$

for a constant $C > 0$.

Assumption M5.*For all $\theta \in \mathcal{N}$ of θ_0, $\hat{g}(\theta)$ is defined in Assumption M3iv.*

(i).$n^{1/2}/q_n^{1/2}\sup_{\theta\in\mathcal{N}}\|\hat{g}(\theta)\|, n^{1/2}/q_n^{1/2}\sup_{\theta\in\mathcal{N}}\|\partial\hat{g}(\theta)/\partial\theta_j\|$, and $n^{1/2}/q_n^{1/2}\sup_{\theta\in\mathcal{N}}\|\partial^2\hat{g}(\theta)/\partial\theta_j\partial\theta_k\|$ are bounded in probability.

(ii). Each of $E\|g_i(\theta)\|^4/n, E\|\partial g_i(\theta)/\partial\theta_j\|^4/n, E\|\partial^2 g_i(\theta)/\partial\theta_j\partial\theta_k\|^4$ converge to zero.

(iii). $\sup_{\theta\in\mathcal{N}}\|\hat{\Omega}(\theta) - \Omega(\theta)\| \xrightarrow{p} 0$, $\sup_{\theta\in\mathcal{N}}\|\partial\hat{\Omega}(\theta)/\partial\theta_j - \partial\Omega(\theta)/\partial\theta_j\| \xrightarrow{p} 0$, and $\sup_{\theta\in\mathcal{N}}\|\partial^2\hat{\Omega}(\theta)/\partial\theta_j\partial\theta_k - \partial^2\Omega(\theta)/\partial\theta_j\partial\theta_k\| \xrightarrow{p} 0$.

Assumption M4 restricts the rate of growth of moment conditions. Assumption M5 contains high level conditions. Condition (i) is essential in obtaining the limit. Primitives for linear case are provided in Newey and Windmeijer (2005). Also in nonlinear case, the primitives can be supplied using Chapter 19 in van der Vaart (2000). The following Assumption explains the nature of many weak moment asymptotics very well, and used as Assumption 1 in Newey and Windmeijer (2005). Note that with this kind of asymptotics the individual cells in $E[\partial g_i(\theta_0)/\theta]$ shrink in magnitude (weak identification, at rate of $n^{1/2}$) and the number of moments grow with sample size (many moments).

Assumption M6.

$$G'_n\Omega^{-1}G_n \to H,$$

and H is nonsingular.

Then we need the following Assumption for inference as well as limits. Denote $\tilde{Q}(\theta) = \frac{n}{2q_n}\hat{g}(\theta_0)'\Omega^{-1}\hat{g}(\theta)$, and $\hat{g} = \sum_{i=1}^n g_i(\theta), \Omega = Eg_i(\theta_0)g_i(\theta_0)'$.

Assumption M7.*$\partial^2\tilde{Q}(\theta)/\partial\theta\partial\theta'$ is stochastically equicontinuous.*

Now we set up the following notation. Define $\lim_{n\to\infty}\Lambda_n = \lim_{n\to\infty} nE[\tilde{U}'\Omega^{-1}\tilde{U}]/q_n = \Lambda^*$, where $\tilde{U}$ is defined in (56)-(58) and basically the same as $\hat{U}$ in (19) except that it uses population values of covariance matrices. Λ^* is additional component to the variance of estimators. This is basically the variance that comes from second right hand side term in (19).

Theorem 7.*If Assumptions M1-M7 are satisfied and* $\Lambda_n \to \Lambda^*$ *then*

$$q_n^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{d} N(\tau, V),$$

where $\tau = \lim_{n\to\infty} G_n' \Omega^{-1} C_1$ *and* $V = H^{-1} + H^{-1}\Lambda^* H^{-1}$.

In the case of fixed number of weak moments with near exogeneity (Theorem 1) the estimator is inconsistent and the limits are very different from the one here. Increasing number of moment conditions here bring back consistency because they provide valuable information to system. Also as discussed before there is no drift associated with the second term in (19) since $q_n \to \infty$. This plays a big role in getting inference results below. Compared to Theorem 3 of Newey and Windmeijer (2005), we have an additional drift due to first right hand side term in (19).

Next we need to define notation for the inference on our parameters. Define $\hat{D}_j(\theta)$ as the Jacobian estimated efficiently by using the probabilities in Smith (1997), Brown and Newey (1992).

$$\hat{D}_j(\theta) = \left[\frac{n}{q_n}\right]^{1/2} \left[\frac{\partial \hat{g}(\theta)}{\partial \theta_j} - A^j(\theta)\hat{g}(\theta)\right], \; j = 1, \cdots, p$$

and $\hat{D}(\theta) = [\hat{D}_1(\theta), \cdots, \hat{D}_p(\theta)]$, $q_n \times p$ matrix.

$$A^j(\theta) = \left[\frac{\sum_{i=1}^n g_i^j(\theta) g_i(\theta)'}{n}\right] \Omega(\theta)^{-1},$$

and $g_i^j(\theta) = \partial g_i(\theta)/\partial\theta_j$, $\hat{\Omega}(\theta) = \sum_{i=1}^n g_i(\theta)g_i(\theta)'/n$.

To test $H_0 : \theta = \theta_0$, we benefit from Kleibergen (2002) type test. This is generalized to many weak moments setting in GEL by Newey and Windmeijer (2005). We use an asymptotically equivalent form that shows the driving force behind our result.

$$K(\theta_0) = q_n \left[\frac{\partial \hat{Q}(\theta_0)}{\partial \theta}\right]' [\hat{D}(\theta_0)'\hat{\Omega}^{-1}\hat{D}(\theta_0)]^{-1} \left[\frac{\partial \hat{Q}(\theta_0)}{\partial \theta}\right],$$

where $\hat{\Omega} = \frac{1}{n}\sum_{i=1}^n g_i(\theta_0)g_i(\theta_0)'$. This test has χ_p^2 limit in the case of many weak moments (Theorem 5, Newey and Windmeijer 2005). Next Theorem extends their result to near exogeneity case defined by Assumption M1. This Theorem also extends Theorem 3 here from fixed number of moment equations to many moment setup.

Theorem 8.*If Assumptions M1-M7 are satisfied and* $\Lambda_n \to \Lambda^*$,

$$K(\theta_0) \xrightarrow{d} \chi_p^2(\tau' V^{-1}\tau),$$

where $\tau' V^{-1} \tau$ *is the noncentrality parameter.*

Remarks. There are two interesting connections to this limit. First when we compare this with fixed number of weak moments (q) case (Theorem 3 here) we see that limit is entirely different from here. The main reason for that comes from second term on the right hand side of (19). When $q_n \to \infty$ that term converges to zero and drift is only introduced through the first term on the right hand side of (19). This is shown in the proof of Lemma A9. So having more information simplifies the drift in the limit. Also note that in Theorem 3 here we use $S(\theta_0), LM(\theta_0)$ which are asymptotically equivalent to $K(\theta_0)$ when the number of moment equations is fixed (q).

The next interesting result is comparison of this with Theorem 5 in Newey and Windmeijer (2005). They have a nuisance parameter free limit. However, here we see that near exogeneity introduces a drift, and the limit is not nuisance parameter free. This clearly shows that if we do not take into account the near exogeneity problem and use χ^2_p values we can overreject true null.

Subsampled $K(\theta_0)$ test also converges to its limit. Theorem 5 applies here as well since $C_1 = (0'_{q_n-l}, C'_l)'$ (i.e. C_1 has fixed amount of nonzero cells , not changing with n). Of course, n, q_n are provided then we choose block size "b" as described in subsampling section. However, q_b must be chosen by using $n/q_n = b/q_b$ after choosing b.

If C_1 is an $q_n \times 1$ vector of all nonzero entries, then $\hat{\theta}$ is inconsistent. This can be easily seen analyzing (49). In that case $S_n(\theta_0) = O(1)$ when $q_n \to \infty$. So in (48) $S_n(\hat{\theta}) \leq O(1)$ and hence there is no consistency. This can also be established using argmax continuous mapping theorem more directly. This approach is used in showing inconsistency of two-step GMM estimator in many weak moments by Han and Phillips (2004).

Since the estimator is inconsistent in this case, we have to see whether $K(\theta_0)$ converge to some limit that can be used either by resampling techniques or directly. Note that $K(\theta_0)$ test statistic depends on the limit behavior of $q_n^{1/2} \frac{\partial \hat{Q}(\theta_0)}{\partial \theta}$. The limit of this term is derived in Lemma A9. Analysis is done by decomposing that expression into various terms as in (62). The last term in (62) is

$$n^{1/2} E\tilde{U}'\Omega^{-1}(n^{1/2}\frac{E\hat{g}}{q_n^{1/2}}) = \frac{C_A'\Omega^{-1}C_1}{q_n^{1/2}},$$

by (72), where $C_A = [A^1 C_1, \cdots, A^p C_1]$. However if C_1 is $q_n \times 1$ nonzero vector for all cells then

$$\frac{C_A'\Omega^{-1}C_1}{q_n^{1/2}} \leq C\frac{C_1'C_1}{q_n^{1/2}} \to \infty.$$

So $K(\theta_0) \xrightarrow{p} \infty$ if Assumption M1 is changed in a way that C_1 is a vector of nonzeros. This is an all near exogenous system. The system is full of bad information. This distorts the test immensely. So test is useless in this scenario. Only in the case of Assumption M1, with some limited near exogeneity we can benefit from inference using subsampling.

7 Simulation

Our simulations consider fixed number of moments case. There are two reasons for that. First, this involves more complicated and differing limits in tests. Then, this is less costly in computer time.

We consider the linear model in (3)-(4).

$$y = Y\theta_0 + u, \tag{20}$$

$$Y = X\Pi + V, \tag{21}$$

where $Y : n \times p$, $X : n \times l$, the number of instruments "l"can be one (just-identified system) or two (over-identified system). We set $p = 1$. There is only one structural parameter and $\theta_0 = 0$. The sample size "n" is 100. Π vector $(l \times 1)$ takes the value of 0, .1, 1 in all cells of the vector. Π vector determines the strength of instruments.

(X_i, u_i, V_i) is iid and jointly distributed as $N(0, \Omega)$, When $l = 1$

$$\Omega = \begin{pmatrix} 1 & covX_iu_i & 0 \\ covX_iu_i & 1 & 0.25 \\ 0 & 0.25 & 1 \end{pmatrix}.$$

When $l = 2$

$$\Omega = \begin{pmatrix} 1 & 0 & covX_{1i}u_i & 0 \\ 0 & 1 & covX_{2i}u_i & 0 \\ covX_{1i}u_i & covX_{2i}u_i & 1 & 0.25 \\ 0 & 0 & 0.25 & 1 \end{pmatrix},$$

where $covX_iu_i$ can take the values of $0.1, 0.05, 0$ in $l = 1$ case. Note that zero covariance represents standard assumption of exogeneity of instruments. The other values considered here are consistent with weak exogeneity issue. Also we set $CovX_{1i}u_i = CovX_{2i}u_i$ in $l = 2$ case and they take the values of 0.1, 0.05, 0.

When Π vector takes the value of 1 in all cells and $covX_iu_i = 0$, this is the standard instrumental variable estimation setup. There the instruments are strong and exogenous.The case of $covX_iu_i = 0.1$ and $\Pi = .1$ represent the more problematic case of weak instruments and near exogeneity. In our simulation both $covX_iu_i$ and Π varies. As far as we know, such a joint analysis of weak instruments and near exogeneity is new, important, and helps us to understand what may happen in empirical studies.

We test

$$H_0 : \theta_0 = 0,$$

against

$$H_1 : \theta_0 \neq 0.$$

We analyze three specific GEL estimators: Empirical Likelihood (EL), Exponential Tilting (ET), Continuous Updating (CUE). EL, ET, CUE correspond to $\rho(v) = ln(1 - v), -e^v, -(1 + v)^2/2$ respectively in section 2. The simulation exercise considers three test statistics and analyzes their size and power. These are $GELR(\theta_0), S(\theta_0), LM(\theta_0)$ that are introduced in the former sections.

We use subsampling critical values and in the case of block size choice we benefit from the minimum volatility method. We also tried using calibration algorithm in section 4.1 for block size but minimum volatility method gave better results. The block sizes that are considered here $b = \{4, 6, 8, 10, 12, 16, 20\}$, and we set the smoothing parameters $s = m = 1$. In this exercise we consider three test statistics for each GEL estimator. So across test statistics in each estimator we learn which has better size and power. Then we analyze each test statistic across three estimators (EL,ET, CUE). We try to find which estimator has better size and power given a certain test statistic. This exercise helps us understand not only which test statistic performs well but also under which GEL estimator it has good properties. Another important point that is considered is the size of the test statistics when we use asymptotic critical values. This is important because even though theoretically we prove that the limits are changing, in actual simulations the size may not change that much. So we want to analyze this possibility.

7.1 Size

We analyze the size of the tests that are discussed in various GEL estimators. We run 1000 iterations and consider the case of 2 instruments at 10 % nominal level. We also analyzed a just-identified system but the results are very similar so we do not report those. Tables 1 and 2 report the actual percentage of the rejections of the null hypothesis where the data is generated under the null model with $\theta_0 = 0$. We want to answer several questions in this size exercise. The limits of GELR(θ_0), $S(\theta_0)$, $LM(\theta_0)$ tests are different when there is near exogeneity ($C_1 \neq 0$), and when there is exogeneity ($C_1 = 0$). In the case of weak/strong/no identification Guggenberger and Smith (2005) show that the limits of these test statistics are χ^2 distributed. However, here Theorems 2 and 3 show that combined with near exogeneity, these limits change. So we want to observe the magnitude of this change in small samples. Table 1 tries to answer this question.

The second question concerns how well the test statistics fare compared with each other when we use the subsampling correction for the critical values given EL, ET, CUE frameworks. Next, we want to see whether we can observe better size properties across various GEL estimators (EL, ET, CUE) given a specific test statistic. We also want to analyze specific cases of near exogeneity ($cov(X_i, u_i) = 0.10, cov(X_i, u_i) = 0.05$ for both instruments) with strong instruments ($\pi = 1$). This is more of an empirically relevant case. We also want to analyze joint problems of near exogeneity ($cov(X_i, u_i) = 0.10, cov(X_i, u_i) = 0.05$) with weak instruments ($\pi = 0.1$). In the simulations that we conducted $S(\theta_0)$ fared badly across all estimators and various simulations, both in size and power, so we do not provide it in our results. We only focus on $GELR(\theta_0), LM(\theta_0)$ test statistics.

When we look at Table 1, clearly we see that GELR(θ_0), LM (θ_0) test statistics have large size distortions when there is near exogeneity (i.e. $cov(X_i, u_i) = 0.10, 0.05$ columns). So using χ^2 values without taking into account the possibility of small correlation between instrument and the first order conditions lead to important mistakes in inference. These are true for GEL type estimators we consider. For example, analyzing LM (θ_0) test in the case of weak identification $(\pi = 0.1)$ and near exogeneity $(cov(X_i, u_i) = 0.10)$: we observe that size of the test in EL framework is 31.7%, 28.9% in ET, 41.5% in CUE at 10% level. The size distortions are very large when there is no weak identification problem $(\pi = 1)$ but there is near exogeneity $(cov(X_i, u_i) = 0.10)$. We also observe from Table 1 that when there is no near exogeneity $(cov(X_i, u_i) = 0)$ then the actual size in those two test statistics are around 10% level. This last result, the case of exogeneity, is in line with the simulation reported in weakly identified GEL of Guggenberger and Smith (2005). We also observe that even with mild level of correlation between the instrument and the structural error $(cov(X_i, u_i) = 0.05)$ there are size distortions associated with the test statistics across GEL estimators.

In Table 2, we consider each GEL estimator and analyze how two test statistics behave in small samples. We benefit from the subsampling critical values for the limits derived in Theorems 2 and 3. This is explained already in subsampling section. In EL, we see that $GELR(\theta_0)$ test is undersized compared with $LM(\theta_0)$ at 10% level. For example, when there is small correlation between the instrument and the structural error $(cov(X_i, u_i) = 0.05)$ $GELR(\theta_0)$ has 3-5% size, whereas $LM(\theta_0)$ has 10-11% size at the nominal 10% level. Considering ET, in Table 2, $GELR(\theta_0)$ does better than the $LM(\theta_0)$ in the case of near exogeneity $(cov(X_i, u_i) = 0.10)$. In that case, $GELR(\theta_0)$ has 14-16% size compared with $LM(\theta_0)$ which has 19-32% size at 10% level. In the case of CUE both test statistics have large size distortions when $cov(X_i, u_i) = 0.10$.

Now we analyze Table 2 from a different perspective. Given each test statistic we want to consider how the test fares across EL, ET, and CUE. $GELR(\theta_0)$ test statistic has the best size in EL framework at $cov(X_i, u_i) = 0.10$, the size is between 8-10% at different identification issues. $LM(\theta_0)$ test statistic also has the best size in EL framework when $cov(X_i, u_i) = 0.10$, the size is between 10-13% at 10% level. When $cov(X_i, u_i) = 0.05$ both $GELR(\theta_0), LM(\theta_0)$ do well in CUE as well as EL frameworks. In Table 2 we also analyze specifications of near exogeneity $(cov(X_i, u_i) = 0.10, 0.05)$ with strong instruments $(\pi = 1)$, since this is an empirically relevant case. In this setup, we see that $LM(\theta_0)$, $GELR(\theta_0)$ tests in EL framework give the best size results, with sizes ranging 11-13%, and 4-8% respectively at 10% nominal level. Another interesting case is the joint analysis of $cov(X_i, u_i) = 0.1$ with $\pi = 0.1$. Again in this case, $GELR(\theta_0), LM(\theta_0)$ tests in EL have very good size.

To summarize all the findings in Table 2, we find the size of $LM(\theta_0)$ in EL framework to be the best. Also $GELR(\theta_0)$ tests in EL, ET frameworks also have desirable size properties across various combinations in Table 2.

7.2 Power

We analyze the power properties of $GELR(\theta_0), LM(\theta_0)$ test statistics under various GEL estimators using the subsampling critical values. The data is generated according to (20)-(21) where θ_0 takes the values of $\{-1, -0.8, -0.4, 0.4, 0.8, 1\}$. We run 1000 iterations, and report the rejection rates. We have two setups. In the first one near exogeneity with strong instruments are analyzed ($cov(X_i, u_i) = 0.1, \pi = 1$). The second setup involves joint analysis of near exogeneity and weak instruments ($cov(X_i, u_i) = 0.1, \pi = 0.1$). The results of the exercise are presented in Figures 1-5. At each figure part "a" represents strong instruments case (setup 1), part "b" represents the weak instruments case (setup 2). We want to answer two basic questions in this power part of the simulation study.

First given a test statistic we want to know which GEL framework produces the better power. Second, given a GEL framework, which test statistic ($GELR(\theta_0), LM(\theta_0)$) provides the better power. Figures 1a and 1b provide an analysis of $GELR(\theta_0)$ test in near exogeneity with strong (Figure 1a) and weak instruments (Figure 1b). $GELR(\theta_0)$ test in EL framework has the worst power compared with $GELR(\theta_0)$ tests in ET, CUE frameworks. All the tests have low power in Figure 1b, where there are near exogeneity and weak instruments. Figures 2a and 2b analyze $LM(\theta_0)$ tests in EL, ET, CUE frameworks under setups 1 and 2. Again we see that $LM(\theta_0)$ test in EL framework has low power compared with the LM (θ_0) tests in ET, CUE frameworks. In Figure 2a we see that $LM(\theta_0)$ tests in ET, CUE frameworks have good power when there is near exogeneity and strong instruments. To answer the first question; we observe from Figures 1-2 that in EL framework test statistics provide the lowest power. Tests do well in the case of near exogeneity coupled with strong instruments in ET, CUE frameworks.

Figures 3-5 answer the second question. In Figures 3a and 3b we compare $GELR(\theta_0), LM(\theta_0)$ tests in an EL framework. $GELR(\theta_0)$ has much better power than the $LM(\theta_0)$ test in Figure 3a. Figures 4-5 show that in ET, CUE frameworks $GELR(\theta_0), LM(\theta_0)$ have similar power properties. These tests have good power when there are near exogeneity with strong instruments (Figures 4a,5a), but the power suffers dramatically when there are both problems (Figures 4b, 5b).

We think that ET, CUE frameworks have high power in the case of near exogeneity since their objective functions increase the magnitude of "near exogeneity" problem by taking exponentials, quadratics of the sample moments.

To summarize the power results, in the case of setup 1 (near exogeneity, strong instruments) $GELR(\theta_0)$ tests in ET, CUE frameworks provide very good power. Also this is true for $LM(\theta_0)$ tests in ET, CUE. $LM(\theta_0)$ tests even do slightly better than the GELR (θ_0) tests (Figures 4a, 5a). In the case of setup 2 all the tests do a poor job regardless of the GEL estimator type. But if we were to choose a test in that framework, these are $GELR(\theta_0), LM(\theta_0)$ tests in CUE framework. So there is no single test and framework that gives the best result in power exercise. However, when we look at the tests that have done well in power exercise, one of them has a better size than the others. This is $GELR(\theta_0)$ test in ET framework (Table 2). Another test that has good size with good power is GELR (θ_0) test in CUE framework. Unfortunately, LM tests in ET, CUE frameworks come with a large size problem in certain cases (Table 2).

8 Conclusion

This article extends the literature on GEL estimators to the joint analysis of near exogeneity and weak instruments. This joint problem is important from an applied perspective. We show that Anderson-Rubin (1949) and Kleibergen (2002) type of tests' limits change when we have near exogeneity of instruments addition to the weak instruments problem. This is true in spite of the fact that these tests are evaluated under the null of true parameter values. To get correct critical values we use subsampling approach, and we provide a theoretical proof. Simulations show that Anderson-Rubin (1949) type of tests in Exponential Tilting and Continuous Updating frameworks perform the best in terms of size and power. Tests in Empirical Likelihood model do not perform well. An interesting further study may be the analysis of this problem in many instruments framework.

APPENDIX

In the appendix we begin with three lemmata that are helpful in deriving the consistency results for the GEL estimators. These lemmata are Lemmata 7-9 in Guggenberger and Smith (2005). These are designed for weak identification problem in GEL, however these clearly apply to near exogeneity without any change. Addition of an extra local to zero drift term in Assumption 2 does not change any of the algebra in Lemma 7-9 in Guggenberger and Smith (2005). This can be provided from the author on demand. Before providing these results , we introduce some notation as in Guggenberger and Smith (2005). For $n \in N$, let $\Theta_n \subset \Theta$. Let $c_n = n^{-1/2} \max_i \sup_{\theta \in \Theta_n} \|g_i(\theta)\|$, $\Lambda_n = \{\lambda \in R^q : \|\lambda\| \le n^{-1/2} c_n^{-1/2}\}$ if $c_n > 0$, $\Lambda_n = R^q$ otherwise. "uwpa1" denote uniformly over $\theta \in \Theta_n$ with probability approaching 1. Note that $\hat{g}(\theta) = 1/n \sum_{i=1}^{n} g_i(\theta)$.

Lemma A.1 is Lemma 7 of Guggenberger and Smith (2005).

Lemma A.1.*Assume* $\max_i \sup_{\theta \in \Theta_n} \|g_i(\theta)\| = o_p(n^{1/2})$, *then*

$$\sup_{\theta \in \Theta_n, \lambda \in \Lambda_n} |\lambda' g_i(\theta)| \xrightarrow{p} 0,$$

and $\Lambda_n \subset \hat{\Lambda}_n(\theta)$ *uwpa1.* $\hat{\Lambda}_n(\theta)$ *is defined at the beginning of section 2.*

Lemma A.2 is Lemma 8 of Guggenberger and Smith (2005).

Lemma A.2.*Suppose*

- $$\max_i \sup_{\theta \in \Theta_n} \|g_i(\theta)\| = o_p(n^{1/2});$$

- $$\lambda_{min}(\hat{\Omega}(\theta)) \ge \epsilon$$

uwpa1 for some $\epsilon > 0$*;*

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

uniformly over $\theta \in \Theta_n$*;*

- *Assumption 4 holds, then* $\lambda(\theta) \in \hat{\Lambda}_n(\theta)$ *satisfying* $\hat{P}(\theta, \lambda(\theta)) = \sup_{\lambda \in \hat{\Lambda}_n(\theta)} \hat{P}(\theta, \lambda)$ *exists uwpa1.*

$\lambda(\theta) = O_p(n^{-1/2})$ *and* $\sup_{\lambda \in \hat{\Lambda}_n(\theta)} \hat{P}(\theta, \lambda) = O_p(n^{-1})$ *uniformly over* Θ_n*.*

The next Lemma is Lemma 9 of Guggenberger and Smith (2005). We first provide notation for that. Suppose $\Theta_1 \times \Theta_2 \subset \Theta, \Theta_i \subset R^{p_i}$, $p_1 + p_2 = p$. Partition $\theta_0 = (\theta_{01}', \theta_{02}')'$ and $\theta_{02} \in \Theta_2$. For $d_1 \in \Theta_1$ define

$$\hat{\theta}_2(d_1) = arg \min_{d_2 \in \Theta_2} \sup_{\lambda \in \hat{\Lambda}_n((d_1', d_2')')} \hat{P}((d_1', d_2')', \lambda) \in R^{p_2},$$

$$\hat{\theta}(d_1) = (d_1', \hat{\theta}_2(d_1)')' \in R^p,$$

and $\theta_{d_1} = (d_1', \theta_{02}')' \in R^p$.

Lemma A.3.*Suppose*

- $$\max_i \sup_{\theta \in \Theta_1 \times \Theta_2} \|g_i(\theta)\| = o_p(n^{1/2});$$
- $$\lambda_{max}(\hat{\Omega}(\hat{\theta}_{d_1})) \leq \kappa,$$

uwpa1 for some $\kappa < \infty$*;*

- $$\sup_{\lambda \in \hat{\Lambda}_n(\theta_{d_1})} \hat{P}(\theta_{d_1}, \lambda) = O_p(n^{-1}),$$

uniformly over $d_1 \in \Theta_1$*;*

- *Assumption 4 holds then*

$$\hat{g}(\hat{\theta}_{d_1}) = O_p(n^{-1/2})$$

uniformly over $d_1 \in \Theta_1$*.*

Proof of Theorem 1. First we show the consistency. Uniformly on Θ

$$\frac{1}{n}\sum_{i=1}^{n} g_i(\theta) - Eg_i(\theta) \xrightarrow{p} 0,$$

by Assumption 3iii. Then uniformly on θ

$$En^{-1}\sum_{i=1}^{n} g_i(\theta) \to m_2(\beta),$$

by Assumption 2 and $m_2(\beta) = 0$ iff $\beta = \beta_0$. So we need to show $\hat{g}(\hat{\theta}) = o_p(1)$ for consistency. In this respect apply Lemma A.2 to $\Theta_n = \theta_0$ to have

$$\sup_{\lambda \in \hat{\Lambda}_n(\theta_0)} \hat{P}(\theta_0, \lambda) = O_p(n^{-1}).$$

Then by Assumption 3ii $\lambda_{max}(\hat{\Omega}(\hat{\theta})) \leq \kappa$ wpa1 for some $\kappa < \infty$. Then have $p_1 = 0, p_2 = p$ (i.e. $\Theta_2 = \Theta$) and using Lemma A.3 we have the result $\hat{g}(\hat{\theta}) = o_p(1)$. Rate of convergence proof follows from the proof of Theorem 2i of Guggenberger and Smith (2005) (rate of convergence proof for the strongly identified parameter in system with both weakly and strongly identified parameters, equations (A.4)-(A.6), p.29-30) . So we have

$$n^{1/2}(\hat{\beta} - \beta_0) = O_p(1).$$

Now we start the proof of the limits for the GEL estimators. By Assumption 3iii and the rate of convergence (uniformly over α, b)

$$\Psi_n(\alpha, \beta_0 + \frac{b}{n^{1/2}}) \Longrightarrow \Psi(\alpha, \beta_0). \tag{22}$$

Then by Assumption 2, uniformly over $\alpha \times b \in A \times B_M$ where $A \times B_M$ is a compact subset of R^p.

$$En^{-1/2}\sum_{i=1}^{n} g_i(\alpha,\beta_0+\frac{b}{n^{1/2}}) = m_{1n}(\alpha,\beta_0+\frac{b}{n^{1/2}})+n^{1/2}m_{2n}(\beta_0+\frac{b}{n^{1/2}})+C_1 \to m_1(\alpha,\beta_0)+R_2(\beta_0)b+C_1. \quad (23)$$

To save notation set $(\alpha, \beta_0 + \frac{b}{n^{1/2}}) = \theta_{\alpha,b}$. Combine (22)-(23) to have

$$\hat{g}(\theta_{\alpha,b}) = n^{-1/2}[\Psi_n(\theta_{\alpha,b}) + En^{-1/2}\sum_{i=1}^{n} g_i(\theta_{\alpha,b})] = O_p(n^{-1/2}). \quad (24)$$

Then use Lemma A.2 to have

$$\lambda(\theta_{\alpha,b}) = O_p(n^{-1/2}),$$

uniformly over α, b. Also

$$\hat{P}(\theta_{\alpha,b}, \lambda(\theta_{\alpha,b})) = \sup_{\lambda \in \hat{\Lambda}_n(\theta_{\alpha,b})} \hat{P}(\theta_{\alpha,b}, \lambda(\theta_{\alpha,b})),$$

exists uwpa1. This implies that the first order conditions for λ

$$n^{-1}\sum_{i=1}^{n} \rho_1(\lambda' g_i(\theta))g_i(\theta) = 0,$$

has to hold at $\lambda = \lambda(\theta_{\alpha,b})$ uniformly over α, b with probability approaching one.

Expand this first order condition and for $\tilde{\lambda} \in (0, \lambda_{\alpha,b})$ after some simple algebra

$$\lambda(\theta_{\alpha,b}) = -[\sum_{i=1}^{n} \rho_2(\tilde{\lambda}' g_i(\theta_{\alpha,b}))g_i(\theta_{\alpha,b})g_i(\theta_{\alpha,b})'/n]^{-1}\hat{g}(\theta_{\alpha,b}),$$

where $g_i(\theta_{\alpha,b}) = g_i(\alpha, \beta_0 + \frac{b}{n^{1/2}})$. Then substitute this into second order Taylor series expansion

$$\begin{aligned}
\hat{P}(\theta_{\alpha,b}, \lambda(\theta_{\alpha,b})) &= 2\hat{g}(\theta_{\alpha,b})'[\sum_{i=1}^{n} \rho_2(\tilde{\lambda}' g_i(\theta_{\alpha,b}))g_i(\theta_{\alpha,b})g_i(\theta_{\alpha,b})'/n]^{-1}\hat{g}(\theta_{\alpha,b}) \\
&- [\hat{g}(\theta_{\alpha,b})[\sum_{i=1}^{n} \rho_2(\tilde{\lambda}' g_i(\theta_{\alpha,b}))g_i(\theta_{\alpha,b})g_i(\theta_{\alpha,b})'/n]^{-1} \\
&\times [\sum_{i=1}^{n} \rho_2(\lambda^{*'} g_i(\theta_{\alpha,b}))g_i(\theta_{\alpha,b})g_i(\theta_{\alpha,b})'/n] \\
&\times [\sum_{i=1}^{n} \rho_2(\tilde{\lambda}' g_i(\theta_{\alpha,b}))g_i(\theta_{\alpha,b})g_i(\theta_{\alpha,b})'/n]^{-1}\hat{g}(\theta_{\alpha,b})], \quad (25)
\end{aligned}$$

where $\lambda^* \in (0, \lambda(\theta_{\alpha,b}))$. Then by Lemma A.1 and Assumption 4

$$\sup_{\theta} |\rho_2(\tilde{\lambda}' g_i(\theta)) + 1| \xrightarrow{p} 0,$$

$$\sup_{\theta} |\rho_2(\lambda^{*'} g_i(\theta)) + 1| \xrightarrow{p} 0,$$

So by Assumption 3ii and above equations uniformly over α, b

$$[\sum_{i=1}^{n} \rho_2(\tilde{\lambda}' g_i(\theta_{\alpha,b}))g_i(\theta_{\alpha,b})g_i(\theta_{\alpha,b})'/n] \xrightarrow{p} \Omega(\alpha, \beta_0), \tag{26}$$

$$[\sum_{i=1}^{n} \rho_2(\lambda^{*'} g_i(\theta_{\alpha,b}))g_i(\theta_{\alpha,b})g_i(\theta_{\alpha,b})'/n] \xrightarrow{p} \Omega(\alpha, \beta_0). \tag{27}$$

So use (22)-(24) to get

$$n^{1/2}\hat{g}(\theta_{\alpha,b}) \Longrightarrow \Psi(\alpha, \beta_0) + m_1(\alpha, \beta_0) + R_2(\beta_0)b + C_1. \tag{28}$$

Next by (25) and (28)

$$n\hat{P}(\theta_{\alpha,b}, \lambda(\theta_{\alpha,b})) \Longrightarrow [\Psi(\alpha,\beta_0)+m_1(\alpha,\beta_0)+R_2(\beta_0)b+C_1]'\Omega(\alpha,\beta_0)^{-1}[\Psi(\alpha,\beta_0)+m_1(\alpha,\beta_0)+R_2(\beta_0)b+C_1] \equiv P(\alpha, b).$$

Take the partial derivative with respect to b given α to have

$$b^*(\alpha) = -[R_2(\beta_0)'\Omega(\alpha, \beta_0)^{-1}R_2(\beta_0)]^{-1}R_2(\beta_0)'\Omega(\alpha, \beta_0)^{-1}[\Psi(\alpha, \beta_0) + m_1(\alpha, \beta_0) + C_1].$$

Then use Lemma 3.2.1. of van der Vaart and Wellner (1996) with $\alpha^* = arg\min_\alpha P(\alpha, b^*(\alpha))$ and $b^* = b^*(\alpha^*)$ to have the limits. **Q.E.D.**

Proof of Theorem 2. First, use $\hat{g}(\theta_0) = \frac{1}{n}\sum_{i=1}^{n} g_i(\theta_0)$

$$n^{1/2}\hat{g}(\theta_0) = [n^{-1/2}\sum_{i=1}^{n} g_i(\theta_0) - Eg_i(\theta_0)] + En^{-1/2}\sum_{i=1}^{n} g_i(\theta_0).$$

Then by Assumption 3*iii

$$n^{-1/2}\sum_{i=1}^{n} g_i(\theta_0) - Eg_i(\theta_0) \xrightarrow{d} \Psi(\theta_0) \equiv N(0, \Omega(\theta_0)), \tag{29}$$

where we have $\Omega(\theta_0)$ instead of $V(\theta_0)$ since $Eg_i(\theta_0) = \frac{C_1}{n^{1/2}}$ by (2).

Next

$$En^{-1/2}\sum_{i=1}^{n} g_i(\theta_0) \to C_1, \tag{30}$$

Combine (29)-(30) to have

$$n^{1/2}\hat{g}(\theta_0) = n^{-1/2}\sum_{i=1}^{n} g_i(\theta_0) \xrightarrow{d} N(C_1, \Omega(\theta_0)). \tag{31}$$

Then use (25)-(27) at θ_0 with (31) to have the desired result.**Q.E.D**

Proof of Theorem 3. The proof is very similar to the proof of Theorem 4 in Guggenberger and Smith (2005). Assumption 2* here allows also for near exogeneity compared to Guggenberger and Smith (2005). We only consider $LM(\theta_0)$ test statistic since $S(\theta_0)$ is asymptotically equivalent to that. Denote $1/n\sum_{i=1}^{n} g_i(\theta_0)$ by $\hat{g}$. By following the proof of Theorem 4 (equation (A.8) and the equation immediately after that) in Guggenberger and Smith (2005) at $\theta = \theta_0$, using (9), Assumption 5

$$vec(D^*(\theta_0), n^{1/2}\hat{g}) = \omega_1 + M\nu + o_p(1), \tag{32}$$

where $\omega_1 = vec(0, -R_2(\beta_0), 0) \in R^{qp_A+qp_B+q}$ and

$$M = \begin{pmatrix} -I_{qp_A} & \Delta_A\Omega(\theta_0)^{-1} \\ 0 & 0 \\ 0 & I_q \end{pmatrix},$$

$$\nu = \begin{pmatrix} n^{-1/2}\sum_{i=1}^{n} vecG_{iA}(\theta_0) \\ n^{-1/2}\sum_{i=1}^{n} g_i(\theta_0) \end{pmatrix}.$$

M and ν have dimensions of $(qp_A + qp_B + q) \times (qp_A + q)$ and $(qp_A + q) \times 1$ respectively. Our equation (32) is the same as in the proof of Theorem 4 in Guggenberger and Smith (2005). The reason that equation (32) is the same as in Guggenberger and Smith (2005) is the derivation does not use any near exogeneity assumption. The near exogeneity becomes an issue in the limit for $n^{1/2}\hat{g}$ which is the subsequent step.

By Assumption 2*, 5iv,vii and equation (9)

$$\nu \xrightarrow{d} N(\omega_2, V_1(\theta_0)),$$

where

$$\omega_2 = [(vecR_{1A})', C_1']'.$$

R_{1A} is matrix that is formed by same rows but only first p_A columns of $R_1(\theta_0)$.

Now we derive the joint distribution of $D^*(\theta_0)$ and $n^{1/2}\hat{g}$:

$$vec(D^*(\theta_0), n^{1/2}\hat{g}) \xrightarrow{d} N(\omega_1 + M\omega_2, V_2(\theta_0)),$$

where

$$V_2(\theta_0) = \begin{pmatrix} \Psi_A & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Omega(\theta_0) \end{pmatrix},$$

$\Psi_A = \Delta_{AA} - \Delta_A\Omega(\theta_0)^{-1}\Delta_A'$ has full column rank. The last equation shows that $D^*(\theta_0)$ and $n^{1/2}\hat{g}$ are asymptotically independent. The limits of $D^*(\theta_0), n^{1/2}\hat{g}$ are respectively $\bar{D}, \bar{g}$. Then

$$\begin{aligned}(D^*(\theta_0)'\hat{\Omega}(\theta_0)^{-1}D^*(\theta_0))^{-1/2}D^*(\theta_0)\hat{\Omega}(\theta_0)^{-1}n^{1/2}\hat{g} &\xrightarrow{d} (\bar{D}'\Omega(\theta_0)^{-1}\bar{D})^{-1/2}(\bar{D}'\Omega(\theta_0)^{-1}\bar{g}) \\ &\equiv W(C_1) + \zeta, \end{aligned} \tag{33}$$

where

$$W(C_1) = (\bar{D}'\Omega(\theta_0)^{-1}\bar{D})^{-1/2}\bar{D}'\Omega(\theta_0)^{-1}C_1, \tag{34}$$

and $\zeta \equiv N(0, I_p)$. $W(C_1)$ and ζ are independent. The result follows from the test statistic and (33) proceeding as in the proof of Theorem 4 in Guggenberger and Smith (2005). **Q.E.D**

Proof of Theorem 4. First use (31) to have

$$\hat{\Omega}(\theta_0)^{-1/2} n^{1/2} \hat{g} \xrightarrow{d} \Xi + \Omega(\theta_0)^{-1/2} C_1,$$

where $\Xi \equiv N(0, I_q)$. Then use the above result with $D^*(\theta_0) \xrightarrow{d} \bar{D}$ from the proof of Theorem 3 to have the desired result.**Q.E.D**.

Proof of Theorem 5.

a) First we show a different way of writing the test statistic that will be helpful in understanding the continuity of the limit in C_1 (near exogeneity parameter) in Assumption 2*. This is also true in the test itself.

Rewrite

$$\begin{aligned} n\hat{P}(\theta_0, \lambda(\theta_0)) &= GELR(\theta_0) \\ &= [n^{1/2}\hat{g}(\theta_0)]'\Omega(\theta_0)^{-1}[n^{1/2}\hat{g}(\theta_0)] + o_p(1), \end{aligned} \tag{35}$$

by (25)-(27). The asymptotically negligible remainder term is derived from (26)-(27). Then in Assumption 3* iii we have empirical process evaluated at θ_0

$$\Psi_n(\theta_0) = n^{-1/2} \sum_{i=1}^{n} g_i(\theta_0) - E g_i(\theta_0). \tag{36}$$

Note that

$$\begin{aligned} n^{1/2}\hat{g}(\theta_0) &= n^{-1/2} \sum_{i=1}^{n} g_i(\theta_0) - E g_i(\theta_0) + E n^{-1/2} \sum_{i=1}^{n} g_i(\theta_0) \\ &= \Psi_n(\theta_0) + C_1, \end{aligned} \tag{37}$$

by (36) and Assumption 2*. By (37) and (35)

$$GELR(\theta_0) = [\Psi_n(\theta_0) + C_1]'\Omega(\theta_0)^{-1}[\Psi_n(\theta_0) + C_1] + o_p(1). \tag{38}$$

It is clear from (38) that by Assumption 3* iii

$$GELR(\theta_0) \xrightarrow{d} \chi^2_q(\delta), \tag{39}$$

where $\delta = C_1'\Omega(\theta_0)^{-1}C_1$. The noncentrality parameter is continuous in C_1 and the noncentral χ^2 distribution is continuous in δ. The limit in (39) is therefore continuous in C_1.

We derive the result for $\mathcal{P} \in \mathbf{P_0}$, then we use contiguity to prove the result for $\mathcal{P}_n$.

Now we derive the result. First, $0 \leq \hat{G}_{n,b}(z) \leq 1$. $\hat{G}_{n,b}(z)$ is a U-statistic of degree b with $E\hat{G}_{n,b}(z) = \frac{1}{n-b+1} \sum_{j=1}^{n-b+1} G_{b,j}(z, \mathcal{P})$ where $G_{b,j}(z, \mathcal{P})$ is defined in (16). By Hoeffding's inequality for independent stochastic processes (Proposition A.6.1 in van der Vaart and Wellner (1996)) for $t > 0$

$$Prob_{\mathcal{P}}\{\hat{G}_{n,b}(z) - \frac{1}{n-b+1}\sum_{j=1}^{n-b+1} G_{b,j}(z,\mathcal{P}) \geq t\} \leq exp\{-2[n/b]t^2\}, \tag{40}$$

same things hold for $t < 0$. So

$$\hat{G}_{n,b}(z) - \frac{1}{n-b+1}\sum_{j=1}^{n-b+1} G_{b,j}(z,\mathcal{P}) \xrightarrow{p} 0. \tag{41}$$

Note that since Theorem 2 is for triangular arrays, and by $b \to \infty, n \to \infty, b/n \to 0$, so cumulative distribution function of our subsampled tests $T_{n,b,j}$ averaged over all subsamples converges to the cumulative distribution function of the limit in Theorem 2:

$$\frac{1}{n-b+1}\sum_{j=1}^{n-b+1} G_{b,j}(z,\mathcal{P}) \to G(z,P), \tag{42}$$

where $G(z,P)$ is the cumulative distribution function in the limit of Theorem 2. Then by (42) and (41) to have

$$\hat{G}_{n,b}(z) \xrightarrow{p} G(z,\mathcal{P}).$$

Contiguity of $\mathcal{P}_n to \mathcal{P} \in \mathbf{P_0}$ forces these to hold under $\mathcal{P}_n$ (Politis, Romano and Wolf 1999, proof of Theorem 2.6.1.iii). So it follows that $c_{n,b}(1-\alpha) \to c(1-\alpha,\mathcal{P})$ in probability $\mathcal{P}$. Then by Slutsky's theorem, the asymptotic rejection probability of the event $T_n > c_{n,b}(1-\alpha)$ is α. Uniformity of the results ((40)-(42)) in C_1 can be seen by (38). Every C_1 in test statistic is mapped to its counterpart in the limit, C_1 is a constant vector and not a sequence.

b) Given (37) and since the limit and tests in Theorem 3 are also continuous in C_1, same approach as the proof of Theorem 5a yields the result. **Q.E.D**

The rest of the Appendix provides proofs for the many weak moment asymptotics with near exogeneity. We need to introduce some notation. Let $\Omega(\theta) = Eg_i(\theta)g_i(\theta)'$, and

$$S_n(\theta) = \frac{n}{2q_n}[Eg_i(\theta)]'\Omega(\theta)^{-1}[Eg_i(\theta)].$$

$S_n(\theta)$ is related to the limit of CUE objective function. First we approach the consistency issue. The consistency proof consists of two main steps. First it has to be shown that

$$\sup_{\theta\in\Theta}|\hat{Q}(\theta) - \bar{Q}(\theta)| \xrightarrow{p} 0, \tag{43}$$

where $\bar{Q}(\theta) = 1/2 + S_n(\theta)$.

To show (43), some intermediate steps are needed. Newey and Windmeijer (2005) show that GEL objective function converges uniformly in θ to CUE objective function. Then CUE objective function converges uniformly over θ to its limit $1/2 + S_n(\theta)$.

The second step in consistency proof is identifiability condition and this is analyzed in detail in this paper.

The following Lemmata are used to get (43). Note that these lemmata do not warrant the usage of moment conditions at the true value of the parameter θ_0. Neither they use the sample moments at the true value. These are already obtained in Newey and Windmeijer (2005) and Assumption M1 is not used in derivation. These are Lemmata A.3, A3a, A3b in Newey and Windmeijer (2005) respectively. Let "wpa1" denote with probability approaching one.

Lemma A.4. *If Assumptions M2 and M3 are satisfied, then*

$$\sup_{\theta \in \Theta} |\tilde{Q}(\theta) - \bar{Q}(\theta)| \xrightarrow{p} 0,$$

where $\tilde{Q}(\theta) = \frac{n}{2q_n}\hat{g}(\theta)'\Omega(\theta)^{-1}\hat{g}(\theta)$.

Lemma A.5. *If Assumptions M2, M3, M4 are satisfied then for wpa1,*

(i).

$$\hat{\lambda}(\theta) = \arg \max_{\lambda \in \hat{\Lambda}_n(\theta)} \sum_{i=1}^{n} \rho(\lambda' g_i(\theta))/q_n,$$

exists for all $\theta \in \Theta$.

(ii).

$$\sup_{\theta \in \Theta} \|\hat{\lambda}(\theta)\| = O_p(q_n^{1/2}/n^{1/2}).$$

Lemma A.6.*If Assumptions M2, M3, are satisfied then*

$$\sup_{\theta \in \Theta} |\hat{Q}(\theta) - \tilde{Q}^*(\theta)| \xrightarrow{p} 0,$$

where $\tilde{Q}^*(\theta) = \frac{n}{2q_n}\hat{g}(\theta)'\hat{\Omega}(\theta)^{-1}\hat{g}(\theta)$, $\hat{\Omega}(\theta) = \sum_{i=1}^{n} g_i(\theta)g_i(\theta)'/n$.

Proof of Theorem 6. Proceeding as in the proof of Theorem 1 in Newey and Windmeijer (2005) using Lemmata A.4-A.6 we obtain

$$\sup_{\theta \in \Theta} |\hat{Q}(\theta) - \bar{Q}(\theta)| \xrightarrow{p} 0,$$

which is equation (43). We should note that from Lemmata A.4-A.6 derivation of (43) is not trivial but the moment functions at true value of the parameter (i.e. $Eg_i(\theta_0)$) or the sample version at the true value does not play any role in achieving this via Lemmata A.4-A.6. Following Newey and Windmeijer (2005) after Lemmata A.4-A.6 provides (43).

However, $Eg_i(\theta_0)$ plays a crucial role in the second step of the proof of consistency: identifiability condition. By (43), for any $\zeta > 0$, wpa1

$$\bar{Q}(\hat{\theta}) \leq \hat{Q}(\hat{\theta}) + \zeta, \tag{44}$$

$$\hat{Q}(\theta_0) \leq \bar{Q}(\theta_0) + \zeta, \tag{45}$$

wpa1. Then from the definition of $\hat{\theta}$

$$\hat{Q}(\hat{\theta}) \leq \hat{Q}(\theta_0). \tag{46}$$

So, wpa1 by (44)-(46)

$$\bar{Q}(\hat{\theta}) \leq \hat{Q}(\hat{\theta}) + \zeta \leq \hat{Q}(\theta_0) + \zeta \leq \bar{Q}(\theta_0) + 2\zeta. \tag{47}$$

Then since $\bar{Q}(\theta) = 1/2 + S_n(\theta)$ by definition in Newey and Windmeijer (2005) we can rewrite (47) as wpa1

$$S_n(\hat{\theta}) \leq S_n(\theta_0) + 2\zeta. \tag{48}$$

By Assumption M1 we have wpa1

$$S_n(\theta_0) = \frac{n}{2q_n}[Eg_i(\theta_0)]'\Omega(\theta_0)^{-1}[Eg_i(\theta_0)] = \frac{C_1'\Omega(\theta_0)^{-1}C_1}{2q_n}. \tag{49}$$

Note that $S_n(\theta_0) = o(1)$ as $q_n \to \infty$ by Assumption M1, and M3. Then by Assumption M2, (48) and the above fact we have wpa1

$$\Delta(\|\hat{\theta} - \theta_0\|) \leq S_n(\hat{\theta}) \leq 2\zeta.$$

Since $\zeta > 0$ it follows that $\Delta(\|\hat{\theta} - \theta_0\|) \xrightarrow{p} 0$ which implies $\hat{\theta} \xrightarrow{p} \theta_0$.**Q.E.D**

Now we want to prove the limit theory for GEL estimators with many weak moments and near exogeneity. To that end we need the following notation, used in Newey and Windmeijer (2005). For each $j = 1, \cdots, p$, $g_i = g_i(\theta_0), g_i^j = \frac{\partial g_i(\theta_0)}{\partial \theta_j}, \hat{g} = \frac{1}{n}\sum_{i=1}^n g_i : q_n \times 1 vectors$

$$\hat{\Omega} = \sum_{i=1}^n g_i g_i'/n, \quad q_n \times q_n$$

$$\hat{G}^j = \sum_{i=1}^n g_i^j/n, \quad q_n \times 1$$

$$\hat{A}^j = (\sum_{i=1}^n g_i^j g_i'/n)\hat{\Omega}^{-1}, \quad q_n \times q_n$$

$$\hat{U}^j = \hat{G}^j - Eg_i^j - \hat{A}^j\hat{g}, \quad q_n \times 1.$$

Basically for each "j", $\hat{U}^j$ represents the residuals from projection of derivatives on the moment functions.

By the proof of Lemma A3a in Newey and Windmeijer (2005)

$$\sup_\theta |\rho_2(\bar{\lambda}' g_i(\theta)) + 1| \xrightarrow{p} 0, \tag{50}$$

where $\bar{\lambda} \in (0, \hat{\lambda})$. Also this can be shown by using $\hat{\lambda} = O_p(\sqrt{q_n/n})$, Lemma A5 and Assumption M3v and $\rho_2 = -1$ in this paper.

Then we also need the following result which is in the proof of Lemma A3b in Newey and Windmeijer (2005)

$$\hat{\lambda}(\theta) = \hat{\Omega}(\theta)^{-1}[\hat{g}(\theta) + o_p(\sqrt{q_n/n})], \tag{51}$$

where $\hat{g}(\theta) = \frac{1}{n}\sum_{i=1}^n g_i(\theta)$. Then note that we have the mean value expansion for $\hat{\lambda} = \hat{\lambda}(\theta_0)$,

$$\rho_1(\hat{\lambda}' g_i) = \rho_1 + \hat{\lambda}' g_i \rho_2(\bar{\lambda}' g_i) = 1 + \hat{\lambda}' g_i \rho_2(\bar{\lambda}' g_i). \tag{52}$$

Then using (50) we can rewrite (52) as

$$\rho_1(\hat{\lambda}' g_i) - [1 - \hat{\lambda}' g_i] \xrightarrow{p} 0. \tag{53}$$

Use the envelope theorem, (51),(53)

$$\begin{aligned}
\sqrt{q_n}\frac{\partial \hat{Q}(\theta_0)}{\partial \theta_j} &= \sqrt{q_n}\frac{\partial\left[\sum_{i=1}^n \rho(\hat{\lambda}' g_i(\theta)/q_n\right]|_{\theta=\theta_0}}{\partial \theta_j} \\
&= \sum_{i=1}^n \rho_1(\hat{\lambda}' g_i)\hat{\lambda}' g_i^j/\sqrt{q_n} = \sum_{i=1}^n (1-\hat{\lambda}' g_i)\hat{\lambda}' g_i^j/\sqrt{q_n} + o_p(1) \\
&= \frac{n^{1/2}}{q_n^{1/2}}(\hat{G}^j - \hat{A}^j\hat{g})\hat{\Omega}^{-1} n^{1/2}\hat{g} + \frac{n^{1/2}}{q_n^{1/2}} o_p(\frac{q_n^{1/2}}{n^{1/2}}) + o_p(1) \\
&= \frac{n^{1/2}}{q_n^{1/2}}(Eg_i^j)'\hat{\Omega}^{-1} n^{1/2}\hat{g} + n^{1/2}\hat{U}^j\hat{\Omega}^{-1} n^{1/2}\hat{g}/q_n^{1/2} + o_p(1),
\end{aligned} \tag{54}$$

where in the last step we add and subtract $\frac{n^{1/2}}{q_n^{1/2}}(Eg_i^j)'\hat{\Omega}^{-1} n^{1/2}\hat{g}$ and use the definition of $\hat{U}^j$ above.

Now stack over "j" in (54) and use $G_n = \frac{n^{1/2}}{q_n^{1/2}} E\frac{\partial g_i(\theta_0)}{\partial \theta'}$.

$$q_n^{1/2}\frac{\partial \hat{Q}(\theta_0)}{\partial \theta'} = G_n'\hat{\Omega}^{-1} n^{1/2}\hat{g} + n^{1/2}\hat{U}'\hat{\Omega}^{-1}\frac{n^{1/2}\hat{g}}{q_n^{1/2}} + o_p(1), \tag{55}$$

where $\hat{U} = [\hat{U}^1, \cdots, \hat{U}^p]$, $\hat{U}^j$, $j = 1, \cdots p$ is defined above in notation explanation before (50).

We need the following Lemma, which is Lemma A.10 in Newey and Windmeijer (2005).

Lemma A.7.*If Assumption M4 is satisfied, then for each* $j = 1, \cdots, p$

$$q_n^{1/2}\|\hat{\Omega} - \Omega\| \xrightarrow{p} 0,$$

$$q_n^{1/2}\|\hat{A}^j - A^j\| \xrightarrow{p} 0,$$

where $\Omega = Eg_i g_i'$, $A^j = (Eg_i^j g_i')\Omega^{-1}$.

Lemma A.10 in Newey and Windmeijer (2005) stays the same even when there is near exogeneity, since the proof does not depend on either $\hat{g}$, or Eg_i.

We need to simplify the expression in (55) so that we can benefit it from the limit of estimators. In that respect define

$$U_i^j = \frac{\partial g_i(\theta_0)}{\partial \theta_j} - E\frac{\partial g_i(\theta_0)}{\partial \theta_j} - A^j g_i, \tag{56}$$

and

$$\tilde{U}^j = \sum_{i=1}^{n} U_i^j / n, \tag{57}$$

for $j = 1, \cdots, p$. $\tilde{U}^j : q_n \times 1$ vector.

$$\tilde{U} = [\tilde{U}^1, \cdots, \tilde{U}^p],\ q_n \times p. \tag{58}$$

Then use Lemma A.7, and Assumption M5i

$$\begin{aligned} |n^{1/2}(\hat{U}^{j\prime}\hat{\Omega}^{-1} - \tilde{U}^{j\prime}\Omega^{-1})n^{1/2}\hat{g}/q_n^{1/2}| &\leq |n^{1/2}\hat{g}'(\hat{A}^j - A^j)'\hat{\Omega}^{-1}n^{1/2}\hat{g}/q_n^{1/2}| + |n^{1/2}\hat{g}'A^{j\prime}\Omega^{-1}(\hat{\Omega} - \Omega)\hat{\Omega}^{-1}n^{1/2}\hat{g}/q_n^{1/2}| \\ &\leq C(n\|\hat{g}\|^2/q_n)q_n^{1/2}\|\hat{A}^j - A^j\| + C(n/q_n)\|\hat{g}'A^{j\prime}\Omega^{-1}\|\|\hat{g}\|q_n^{1/2}\|\hat{\Omega} - \Omega\| \\ &\xrightarrow{p} 0. \end{aligned} \tag{59}$$

Also similarly

$$G_n'\hat{\Omega}^{-1}n^{1/2}\hat{g} - G_n'\Omega^{-1}n^{1/2}\hat{g} \xrightarrow{p} 0. \tag{60}$$

Then use (59)(60) in (55) we have

$$q_n^{1/2}\frac{\partial \hat{Q}(\theta_0)}{\partial \theta} = G_n'\Omega^{-1}n^{1/2}\hat{g} + n^{1/2}\tilde{U}'\Omega^{-1}n^{1/2}\hat{g}/q_n^{1/2} + o_p(1), \tag{61}$$

where $\tilde{U} = [\tilde{U}^1, \cdots, \tilde{U}^p] : q_n \times p$ matrix. Rewrite that adding and subtracting

$$\begin{aligned} q_n^{1/2}\frac{\partial \hat{Q}(\theta_0)}{\partial \theta} &= [G_n'\Omega^{-1}n^{1/2}(\hat{g} - E\hat{g})] + [n^{1/2}(\tilde{U} - E\tilde{U})'\Omega^{-1}n^{1/2}(\hat{g} - E\hat{g})/q_n^{1/2}] \\ &+ [G_n'\Omega^{-1}n^{1/2}E\hat{g}] + [n^{1/2}E\tilde{U}'\Omega^{-1}n^{1/2}(\hat{g} - E\hat{g})/q_n^{1/2}] \\ &+ [n^{1/2}(\tilde{U} - E\tilde{U})'\Omega^{-1}n^{1/2}E\hat{g}/q_n^{1/2}] + [n^{1/2}E\tilde{U}'\Omega^{-1}n^{1/2}E\hat{g}/q_n^{1/2}] + o_p(1). \end{aligned} \tag{62}$$

The third, fourth, fifth, sixth terms on the right hand side are nonzero because of the near exogeneity. These terms are zero in many weak moment asymptotics of Newey and Windmeijer (2005). In (62), note that

$$n^{1/2}E\tilde{U} = n^{1/2}[E\tilde{U}^1, \cdots, E\tilde{U}^p],$$

where $n^{1/2}E\tilde{U}^j = \sum_{i=1}^n E\tilde{U}_i^j/n = A^j\sum_{i=1}^n Eg_i/n = A^jC_1$ by Assumption M1. So

$$n^{1/2}E\tilde{U} = [A^1C_1, \cdots, A^pC_1] \equiv C_A, \tag{63}$$

where this is a $q_n \times p$ matrix.

We need the following central limit theorem for our estimators. This is Lemma A.2 in Newey and Windmeijer (2005).

Note that W_i, F_i are $q_n \times 1$ iid random vectors with fourth moments (these can depend on n, but suppressed for notational convenience) in Lemma A.8, and be specified for our case in the proof of Lemma A.9.

Lemma A.8.*Define* $EW_iW_i' = \Psi$, *if* $EW_i = EF_i = 0, EF_iF_i' = I_{q_n}, EF_iW_i' = 0_{q_n}$, $na_n'a_n \to H, n^2 tr(\Psi) \to \Lambda^*, n^3 a_n'\Psi a_n \to 0, tr(\Psi^2)/(tr\Psi)^2 \to 0$, $nE[|a_n'F_i|]^4 \to 0, n^{-1}E[|F_1'W_2|^4]/[tr\Psi]^2 \to 0, nE[|F_i'W_i|^2] \to 0$, *then*

$$\sum_{i=1}^n a_n'F_i + \sum_{i=1}^n \sum_{j=1}^n W_i'F_j \xrightarrow{d} N(0, H + \Lambda^*).$$

Now we provide one of the main results of this study, we provide the limit for the score of the objective function in the case of near exogeneity, and many weak moment asymptotics.

Lemma A.9.*If Assumptions M1, M4-M6 are satisfied then*

$$q_n^{1/2} \frac{\partial \hat{Q}(\theta_0)}{\partial \theta} \xrightarrow{d} N(\tau, H + \Lambda^*),$$

where $G_n'\Omega^{-1}C_1 \to \tau$, *and* $\frac{\xi' E(U_i'\Omega^{-1}U_i)\xi}{q_n} \to \xi'\Lambda^*\xi$.

Proof of Lemma A.9. We derive the limit using (62), and Lemma A.8. First, take any nonzero $p \times 1$ vector ξ. Set

$$a_n' = \frac{\xi' G_n' \Omega^{-1/2}}{n^{1/2}},$$

$$F_i = \Omega^{-1/2}(g_i - Eg_i),$$

$$W_i = \frac{\Omega^{-1/2}(U_i - EU_i)\xi}{nq_n^{1/2}}$$

$$\bar{H} = \xi' H \xi,$$

also note that

$$U_i^j - EU_i^j = g_i^j - Eg_i^j - A^j(g_i - Eg_i)$$

and

$$U_i - EU_i = (U_i^1 - EU_i^1, \cdots, U_i^p - EU_i^p),$$

which is a $q_n \times p$ matrix, and

$$EU_i = [A^1 Eg_i, \cdots, A^p Eg_i]. \tag{64}$$

Note that F_i, W_i definitions are slightly different here compared with Newey and Windmeijer (2005) since we demean the random variables here because of near exogeneity.

We try to satisfy the conditions of Lemma A.8 for the first two terms on the right hand side of (62). Clearly $EW_i = 0, EF_i = 0$, then

$$EF_iF_i' = \Omega^{-1/2}E(g_i - Eg_i)(g_i - Eg_i)'\Omega^{-1/2} = \Omega^{-1/2}[\Omega - \frac{C_1C_1'}{n}]\Omega^{-1/2} \to I_{q_n},$$

by Assumption M1.

Next

$$EF_iW_i' = \frac{E[\Omega^{-1/2}(g_i - Eg_i)\xi'(U_i - EU_i)'\Omega^{-1/2}]}{nq_n^{1/2}} \to 0_{q_n},$$

by the form of $U_i - EU_i$, and $g_i - Eg_i$. These terms are asymptotically uncorrelated since U_i is the matrix of residuals from the regression of moment functions on the partial derivatives of these moment functions.

Now consider

$$na_n' a_n = n\xi' G_n' \Omega^{-1} G_n \xi \to \xi' H \xi = \bar{H},$$

via Assumption M6.

Another condition to be satisfied in Lemma A.8 is

$$\begin{aligned} n^2 tr\Psi &= n^2 tr EW_i W_i' = \frac{\xi' E(U_i - EU_i)'\Omega^{-1}(U_i - EU_i)\xi}{q_n} \\ &= \frac{\xi' E(U_i'\Omega^{-1}U_i)\xi}{q_n} - \frac{\xi'(EU_i)'\Omega^{-1}(EU_i)\xi}{q_n} \\ &= \frac{\xi' E(U_i'\Omega^{-1}U_i)\xi}{q_n} - \frac{\xi' C_1'\Omega^{-1}C_1\xi}{q_n n}. \end{aligned}$$

Note that in the above equation the last term is obtained by using EU_i definition in (64) and Assumption M1. Then $C_1'\Omega^{-1}C_1 \leq C$ by $Eig_{max}(\Omega^{-1}) \leq C$ and the definition of C_1 in Assumption M1. So

$$\frac{\xi' C_1'\Omega^{-1}C_1\xi}{q_n n} \to 0.$$

Then since

$$\frac{\xi' E(U_i'\Omega^{-1}U_i)\xi}{q_n} \to \xi'\Lambda^*\xi,$$

we obtain

$$n^2 tr\Psi \to \xi'\Lambda^*\xi.$$

Next consider the following condition in Lemma A.8,

$$\begin{aligned} n^3 a_n'\Psi a_n &= \frac{\xi' G_n'\Omega^{-1}E[(U_i - EU_i)\xi\xi'(U_i - EU_i)']\Omega^{-1}G_n\xi}{q_n} \\ &= \xi' G_n'\Omega^{-1}E[U_i\xi\xi'U_i']\Omega^{-1}G_n\xi/q_n + o(1), \end{aligned}$$

by the definition of EU_i in (64) and Assumption M1. Then

$$EU_i\xi\xi'U_i' \leq C\sum_{j=1}^{p} Eg_i^j g_i^{j'} \leq C. \tag{65}$$

by Assumption M4. So use these in the equation above to have

$$\xi' G_n'\Omega^{-1}EU_i\xi\xi'U_i'\Omega^{-1}G_n\xi/q_n \leq C\xi' G_n'\Omega^{-1}G_n\xi/q_n \to 0,$$

by Assumption M6. So $n^3 a_n'\Psi a_n \to 0$. Similar to this as in Newey and Windmeijer (2005) we obtain $tr(\Psi\Psi)/[tr\Psi]^2 \to 0$. Next, since $\|G_n'\Omega^{-1}\| \leq C$, by Assumption M6, then use Assumption M4 and M1 to have

$$nE[|a_n'F_i|^4] \leq CE\|G_n'\Omega^{-1}(g_i - Eg_i)\|^4/n \leq CE\|g_i\|^4/n \to 0.$$

Then in the same way by Assumption M4 and Assumption M1 we have

$$n^{-1}E|F_1'W_2|^4/[tr\Psi]^2 \to 0,$$

$$nE|F_i'W_i|^2 \to 0.$$

So all the conditions of Lemma A.8 are satisfied for the first two terms on the right hand side of (62)

$$G_n'\Omega^{-1}n^{1/2}(\hat{g}-E\hat{g}) + n^{1/2}(\tilde{U}-E\tilde{U})'\Omega^{-1}n^{1/2}(\hat{g}-E\hat{g})/q_n^{1/2} \xrightarrow{d} N(0, H+\Lambda^*). \tag{66}$$

Now we consider the third term on the right hand side of (62).

$$G_n'\Omega^{-1}n^{1/2}E\hat{g} = G_n'\Omega^{-1}n^{1/2}\sum_{i=1}^{n} En^{-1}g_i = G_n'\Omega^{-1}C_1,$$

by Assumption M1. Then we know that

$$G_n'\Omega^{-1}C_1 \leq \|G_n'\Omega^{-1}\|\|C_1\| \leq C.$$

So define τ as the limit of that expression, where

$$G_n'\Omega^{-1}C_1 \to \tau. \tag{67}$$

This term is the drift due to near exogeneity. This happens because of nonzero mean of the sample moment function (i.e.$E\hat{g} \neq 0$)in the third term. The remaining terms on the right hand side of (62) are more complicated. The interaction between the moments and the residuals that are obtained by regressing the moments on partial derivatives play an important role.

Now consider the fourth term on the right hand side of (62). Rewrite that by (63)

$$n^{1/2}\xi' E\tilde{U}'\Omega^{-1}\frac{n^{1/2}(\hat{g}-E\hat{g})}{q_n^{1/2}} = \frac{\xi' C_A'\Omega^{-1}}{q_n^{1/2}}\frac{\sum_{i=1}^{n}(g_i - Eg_i)}{n^{1/2}}. \tag{68}$$

Then define the following $a_n' = \xi' C_A'\Omega^{-1/2}/(q_n^{1/2}n^{1/2})$, and $F_i = \Omega^{-1/2}(g_i - Eg_i)$. We can think of (68) as $a_n'\sum_{i=1}^{n} F_i$.

Clearly $Ea_n'F_i = 0$, then note that

$$Ea_n'F_iF_i'a_n = a_n'a_n + o(1),$$

by F_i definition and $Eg_i = C_1/n^{1/2}$. Next we analyze the variance

$$\sum_{i=1}^{n} E(a_n'F_i)^2 = na_n'a_n + o(1) = n\frac{\xi' C_A'\Omega^{-1}C_A\xi}{nq_n} \to 0, \tag{69}$$

since $C_A = [A^1C_1, \cdots, A^pC_1]$ and $C_1 = (0_{q_n-l}', C_l')'$. So clearly the fourth term on the right hand side of (62) converges in probability to zero.

$$\frac{n^{1/2}E\tilde{U}'\Omega^{-1}n^{1/2}(\hat{g}-E\hat{g})}{q_n^{1/2}} \to 0. \tag{70}$$

Remark. There is no drift occurring due to fourth term. This happened because the nonzero C_l is fixed in number where as the number of moment restrictions q_n is growing. So here in this sense, quality information (i.e. $0_{q_n - l}$ in C_1) dominates bad information (i.e.C_l in C_1). This can be seen from (69) clearly if q_n were to be not changing with n , the fourth term clearly would not be converging in probability to zero. Also fifth and sixth terms on the right hand side of (62) share this property when q_n is not changing with n.

Now consider the fifth term on the right hand side of (62). We want to rewrite that term. We use the terms in deriving the limit for the first two terms on the right hand side of (62). Define

$$W_i = \frac{\Omega^{-1/2}(U_i - EU_i)\xi}{nq_n^{1/2}}.$$

Note that this notation is used at the beginning of the proof of Lemma A.9. Then we can rewrite

$$n^{1/2}(\tilde{U} - E\tilde{U})'\Omega^{-1}n^{1/2}E\hat{g}/q_n^{1/2} = n^{1/2}\sum_{i=1}^{n} W_i'\Omega^{-1/2}C_1.$$

To get the above equation we benefit from the definition of $U_i^j - EU_i^j$ at the beginning of the proof of Lemma A.9 and $E\hat{g} = C_1/n^{1/2}$. Note that $W_i'\Omega^{-1/2}C_1$ is scalar so

$$n^{1/2}E(C_1'\Omega^{-1/2}W_i) = 0,$$

by the definition of W_i. Then we analyze

$$\begin{aligned}
\sum_{i=1}^{n} E[n^{1/2}C_1'\Omega^{-1/2}W_i]^2 &= n^2 C_1'\Omega^{-1/2}E\left[\frac{\Omega^{-1/2}(U_i - EU_i)\xi\xi'(U_i - EU_i)'}{n^2 q_n}\right]\Omega^{-1/2}C_1 \\
&= \frac{C_1'\Omega^{-1}E[(U_i - EU_i)\xi\xi'(U_i - EU_i)']\Omega^{-1}C_1}{q_n} \\
&\leq \frac{C_1'\Omega^{-1}E[U_i\xi\xi'U_i']\Omega^{-1}C_1}{q_n}
\end{aligned}$$

by the definition of EU_i in (64) by Assumption M1.

Then note that

$$EU_i\xi\xi'U_i' \leq C,$$

by (65). Then by $Eig_{max}(\Omega^{-1}) \leq C$,and the inequality above

$$\frac{C_1'\Omega^{-1}E[U_i\xi\xi'U_i']\Omega^{-1}C_1}{q_n} \leq C\frac{C_1'C_1}{q_n} \to 0,$$

where we use C_1 definition in Assumption M1. This clearly shows that the random variable $n^{1/2}\sum_{i=1}^{n} C_1'\Omega^{-1/2}W_i \xrightarrow{p} 0$. So

$$n^{1/2}(\tilde{U} - E\tilde{U})'\Omega^{-1}n^{1/2}E\hat{g}/q_n^{1/2} = n^{1/2}\sum_{i=1}^{n} W_i'\Omega^{-1/2}C_1 \xrightarrow{p} 0. \tag{71}$$

Now analyze the sixth term on the right hand side of (62). Use (63) and Assumption M1

$$n^{1/2}E\tilde{U}'\Omega^{-1}n^{1/2}E\hat{g}/q_n^{1/2} = C_A'\Omega^{-1}C_1/q_n^{1/2} \to 0, \tag{72}$$

since $C_1 = (0'_{q_n-l}, C_l')'$ and C_l is fixed dimension "l" subvector of nonzero constants. So by applying (66),(67),(70),(71),(72) to the right hand side of (62) we have the result. **Q.E.D**

We need the following Lemma from Newey and Windmeijer (Lemma A.1, 2005). This is used in the subsequent proofs, specifically in the proof of Lemma A.11 below. We use notation in Lemma A.8. Set general $q_n \times 1$ random variables. They are iid, may depend on n, but the additional subscript is avoided in this Lemma for notational convenience. A is a $q_n \times q_n$ matrix. Both F_i, W_i, A will be specified in subsequent Lemma A.11. Set also

$$\bar{F} = \frac{\sum_{i=1}^n F_i}{n},\ \mu_F = EF_i,\ \Sigma_{FF} = EF_iF_i',\ \Sigma_{FW} = EF_iW_i',\ \Sigma_{WW} = EW_iW_i'.$$

Lemma A.10.*If* $Eig_{max}(AA') \le C, Eig_{max}(A'A) \le CEig_{max}(\Sigma_{FF}) \le C, Eig_{max}(\Sigma_{WW}) \le C, E(F_i'F_i)^2/(nq_n^2) \to 0, E(W_i'W_i)^2/(nq_n^2) \to 0, n(\mu_F)'\mu_F/q_n \le C, n(\mu_W)'\mu_W/q_n \le C$ *then*

$$n\bar{F}'A\bar{W}/q_n = tr(A\Sigma_{FW}')/q_n + n(\mu_F)'A\mu_W/q_n + o_p(1).$$

We now provide another Lemma which is useful in deriving the limits for estimators.

Lemma A.11. *If Assumptions M1, M4-M7 are satisfied for any* $\bar{\theta} \xrightarrow{p} \theta_0$

$$\frac{\partial^2\hat{Q}(\bar{\theta})}{\partial\theta\partial\theta'} \xrightarrow{p} H.$$

The result here is the same as in Newey and Windmeijer (2005). So under near exogeneity, we do not observe change in the limit of the derivative of the score.

Proof of Lemma A.11. First of all, we can consider CUE objective function ($\tilde{Q}^*(\theta)$) rather than GEL ($\hat{Q}(\theta)$) by Lemma A6 here or the proof of Lemma A.12 of Newey and Windmeijer (2005). Then even with near exogeneity (Assumption M1) we can obtain

$$\|\frac{\partial\tilde{Q}^*(\bar{\theta})}{\partial\theta\partial\theta'} - \frac{\partial\tilde{Q}(\theta_0)}{\partial\theta\partial\theta'}\| \xrightarrow{p} 0, \tag{73}$$

where $\tilde{Q}(\theta) = \frac{n}{2q_n}\hat{g}(\theta)'\Omega(\theta)^{-1}\hat{g}(\theta)$, $\Omega(\theta) = \sum_{i=1}^n g_i(\theta)g_i(\theta)'/n$. This is equation (7.7) in Newey and Windmeijer (2005). This is not affected by our Assumption M1. To obtain (73) Assumption M7 and consistency is used. Next we want to prove the following

$$\frac{\partial^2\tilde{Q}(\theta_0)}{\partial\theta\partial\theta'} \xrightarrow{p} H.$$

Let $\tilde{Q}_{kj} = \frac{\partial^2\tilde{Q}(\theta_0)}{\partial\theta_k\partial\theta_j}$ so we have to prove for $j, k = 1, \cdots, p$

$$\tilde{Q}_{kj} \xrightarrow{p} H_{kj}. \tag{74}$$

We now introduce some notation. Let $\partial g_i(\theta_0)/\partial\theta_k = g_i^k, \bar{g}_k = \sum_{i=1}^n g_i^k/n, g_i^{kj} = \partial^2 g_i(\theta_0)/\partial\theta_k\partial\theta_j, \bar{g}_{kj} = \sum_{i=1}^n g_i^{kj}/n$. Also we set

$$\Gamma_k = Eg_i^k g_i', \; \Gamma_{k,j} = E[g_i^k - \bar{g}_k][g_i^j - \bar{g}_j]', \; \Gamma_{kj} = E[(g_i^{kj} - \bar{g}_{kj})g_i'].$$

Also denote the first order partial derivative of $\Omega(\theta)$ with respect to θ_k evaluated at true value as Ω_k, and the second order partial derivative with respect to θ_k, θ_j evaluated at true values as $\Omega_{k,j}$. Then

$$\Omega_k = \Gamma_k + \Gamma_k',$$

$$\Omega_{k,j} = \Gamma_{kj} + \Gamma_{k,j} + \Gamma_{k,j}' + \Gamma_{kj}' + o(1),$$

since $Eg_i = C_1/n^{1/2}$, and $Eg_i \to 0$, by Assumption M1. Then also denote the partial derivative of $\tilde{g}(\theta) = n^{1/2}\hat{g}(\theta)/q_n^{1/2}$ with respect to θ_k at true value as $\tilde{g}_k$ and second order partial derivative of $\tilde{g}(\theta)$ with respect to θ_k, θ_j evaluated at true values as $\tilde{g}_{kj}$. Set also $\tilde{g} = \tilde{g}(\theta_0)$ for ease of notation.

Note that by simple differentiation we have

$$\begin{aligned} \tilde{Q}_{kj} &= \tilde{g}_{kj}'\Omega^{-1}\tilde{g} + \tilde{g}_k'\Omega^{-1}\tilde{g}_j \\ &- \tilde{g}_k'\Omega^{-1}\Omega_j\Omega^{-1}\tilde{g} - \tilde{g}_j'\Omega^{-1}\Omega_k\Omega^{-1}\tilde{g} \\ &+ \tilde{g}'\Omega^{-1}\Omega_j\Omega^{-1}\Omega_k\Omega^{-1}\tilde{g} - \frac{1}{2}\tilde{g}'\Omega^{-1}\Omega_{kj}\Omega^{-1}\tilde{g} \end{aligned}$$

First we need to consider the asymptotic behavior of the following terms in order to evaluate $\tilde{Q}_{kj}$.

$$\tilde{g}_{kj}'\Omega^{-1}\tilde{g} = \frac{n}{q_n}\left[\frac{1}{n}\sum_{i=1}^n \frac{\partial^2 g_i(\theta_0)}{\partial\theta_k\partial\theta_j}\right]'\Omega^{-1}\left[\frac{1}{n}\sum_{i=1}^n g_i(\theta_0)\right].$$

Then we use Lemma A.10 for the term above. First, $Eig_{max}(\Omega^{-1}\Omega^{-1}) \leq C$ by Assumption M5ii. Then $Eig_{max}(Eg_i^{kj}g_i^{kj'}) \leq C$ by Assumption M4iii, $Eig_{max}(\Omega) \leq C$. Also we have $E[(g_i^{kj'}g_i^{kj})^2]/(nq_n^2) \to 0$ by Assumption M5ii, $E[(g_i'g_i)^2]/(nq_n^2) \to 0$ by Assumption M5ii as well. Next $nEg_i^{kj'}Eg_i^{kj}/q_n \leq C$ by Assumption M5i, $nEg_i'Eg_i/q_n = C_1'C_1/q_n \leq C$ by Assumption M1.

Next via Assumption M5, Assumption M1

$$\left(\frac{n^{1/2}}{q_n^{1/2}}Eg_i^{kj}\right)'\Omega^{-1}\left(\frac{n^{1/2}}{q_n^{1/2}}Eg_i\right) = \left(\frac{n^{1/2}}{q_n^{1/2}}Eg_i^{kj}\right)'\Omega^{-1}\left(\frac{C_1}{q_n^{1/2}}\right) \to 0.$$

So we satisfied all the conditions in Lemma A.10 for the term we analyzed, via Lemma A.10

$$\tilde{g}_{kj}'\Omega^{-1}\tilde{g} = tr(\Omega^{-1}\Sigma_{kj}')/q_n + o_p(1). \tag{75}$$

Then exactly as in Newey and Windmeijer (2005)

$$\tilde{g}_k'\Omega^{-1}\tilde{g}_j = \left(\frac{n^{1/2}}{q_n^{1/2}}Eg_i^k\right)'\Omega^{-1}\left(\frac{n^{1/2}}{q_n^{1/2}}Eg_i^j\right) + tr(\Omega^{-1}\Gamma_{k,j}')/q_n + o_p(1).$$

Next see that

$$\left(\frac{n^{1/2}}{q_n^{1/2}}Eg_j\right)'\Omega^{-1}\Omega_k\Omega^{-1}\left(\frac{n^{1/2}}{q_n^{1/2}}E\hat{g}\right) = \left(\frac{n^{1/2}}{q_n^{1/2}}Eg_j\right)'\Omega^{-1}\Omega_k\Omega^{-1}\left(\frac{C_1}{q_n^{1/2}}\right) \to 0,$$

$$\left(\frac{n^{1/2}}{q_n^{1/2}}E\hat{g}\right)\Omega^{-1}\Omega_j\Omega^{-1}\Omega_k\Omega^{-1}\left(\frac{n^{1/2}}{q_n^{1/2}}E\hat{g}\right) = \left(\frac{C_1}{q_n^{1/2}}\right)'\Omega^{-1}\Omega_j\Omega^{-1}\Omega_k\Omega^{-1}\left(\frac{C_1}{q_n^{1/2}}\right) \to 0,$$

by Assumptions M5 and M7.

Then use same analysis as in (75) and apply the results immediately above to have

$$\tilde{g}_j'\Omega^{-1}\Omega_k\Omega^{-1}\tilde{g} = tr(\Omega^{-1}\Omega_k\Omega^{-1}\Gamma_j')/q_n + o_p(1),$$

$$\tilde{g}'\Omega^{-1}\Omega_{kj}\Omega^{-1} = tr(\Omega^{-1}\Omega_{kj})/q_n + o_p(1),$$

$$\tilde{g}'\Omega^{-1}\Omega_j\Omega^{-1}\Omega_k\Omega^{-1}\tilde{g} = tr(\Omega^{-1}\Omega_j\Omega^{-1}\Omega_k)/q_n + o_p(1).$$

Collecting all these results in the expression for $\tilde{Q}_{k,j}$ above

$$\tilde{Q}_{k,j} = \left(\frac{n^{1/2}}{q_n^{1/2}}Eg_i^k\right)'\Omega^{-1}\left(\frac{n^{1/2}}{q_n^{1/2}}Eg_i^j\right) + o_p(1). \qquad (76)$$

Then clearly by (76), Assumption M6 and $\partial^2\tilde{Q}(\theta_0)/\partial\theta\partial\theta'$ definition we have

$$\frac{\partial^2\tilde{Q}(\theta_0)}{\partial\theta\partial\theta'} = G_n'\Omega^{-1}G_n + o_p(1) \to H.$$

Q.E.D

Proof of Theorem 7. The proof follows from Lemma A9, Lemma A11 and (18).**Q.E.D**

We need the following Lemma to establish limits for test statistics. This extends Lemma A13 of Newey and Windmeijer (2005) to nearly exogenous systems. The limit result is the same as in Newey and Windmeijer (2005).

Lemma A.12. *If Assumptions M1,M3-M7 are satisfied,* $\Lambda_n \to \Lambda^*$

$$\hat{D}(\theta_0)'\hat{\Omega}^{-1}\hat{D}(\theta_0) \xrightarrow{p} H + \Lambda^*.$$

Proof of Lemma A.12. Using definition of $\hat{D}(\theta_0)$ after Theorem 7 we can show easily that

$$\|\hat{D}(\theta_0)'\hat{\Omega}^{-1}\hat{D}(\theta_0) - \tilde{D}(\theta_0)'\Omega^{-1}\tilde{D}(\theta_0)\| \xrightarrow{p} 0,$$

where

$$\tilde{D}_j(\theta_0) = [\frac{n}{q_n}]^{1/2}\left[\frac{\partial\hat{g}}{\partial\theta_j} - A^j\hat{g}\right],$$

for $j = 1, 2, \cdots, p$. $\tilde{D}(\theta_0)$ is a $q_n \times p$ matrix where columns are $\tilde{D}_j(\theta_0)$. A^j is defined in the statement of Lemma A7.

We now apply Lemma A10 to prove the result.

$$\tilde{D}(\theta_0)'\Omega^{-1}\tilde{D}(\theta_0) \xrightarrow{p} H + \Lambda^*.$$

Set $A = \Omega^{-1}$, $F_i = \frac{\partial g_i(\theta_0)}{\partial \theta_j} - A^j g_i = g_i^j - A^j g_i$, $W_i = \frac{\partial g_i(\theta_0)}{\partial \theta_k} - A^k g_i = g_i^k - A^k g_i$. We show that with these choices, conditions of Lemma A10 are satisfied.

$Eig_{max}(\Omega^{-1}\Omega^{-1}) \leq C$ is satisfied by Assumption M3. Use $A^j = (Eg_i^j g_i')(Eg_i g_i')^{-1}$ definition to have $Eig_{max}(\Sigma_{FF}) = Eig_{max}(Eg_i^j g_i^{j'} - A^j Eg_i g_i^{j'}) \leq C$ by Assumption M4. Same is true for $Eig_{max}(\Sigma_{WW}) \leq C$.

Now we have to show in Lemma A10

$$E(F_i'F_i)^2/(nq_n^2) \to 0.$$

First in our case

$$E(F_i'F_i)^2/(nq_n^2) = E[g_i^{j'} g_i^j - g_i'(A^j)'g_i^j - g_i^{j'} A^j g_i + g_i'(A^j)'A^j g_i]/(nq_n^2) \to 0,$$

by Assumption M5ii. In the same way $E(W_i'W_i)^2/(nq_n^2) \to 0$. Consider the following condition in Lemma A10

$$\begin{aligned} n(\mu_F)'\mu_F/q_n &= n(Eg_i^j - A^j Eg_i)'(Eg_i^j - A^j Eg_i)/q_n \\ &= n(Eg_i^j)'(Eg_i^j)/q_n - n(Eg_i^j)'A^j Eg_i/q_n \\ &- nEg_i'(A^j)'Eg_i^j/q_n + n(Eg_i)'(A^j)'A^j Eg_i/q_n, \end{aligned}$$

where $Eg_i \neq 0$ because of Assumption M1. This is the main difference between the proof of Lemma A13 in Newey and Windmeijer (2005) and this one. Consider the last term above, using Assumption M1

$$n(Eg_i)'(A^j)'A^j(Eg_i)/q_n = \frac{C_1' A^{j'} A^j C_1}{q_n} \leq C,$$

by Assumption M4iii and Cauchy-Schwartz inequality we have $Eig_{max}(A^{j'}A^j) \leq C$. Then the first term $n(Eg_i^j)'(Eg_i^j)/q_n \leq C$ by Assumption M6. The second and third terms are bounded by a constant using the analysis for first and last terms. Now we can apply Lemma A10, since all conditions of that Lemma are satisfied.

First in Lemma A10,

$$n(\mu_F)'A\mu_W/q_n = \frac{n}{q_n}[Eg_i^j - A^j Eg_i]'\Omega^{-1}[Eg_i^k - A^k Eg_i]. \tag{77}$$

Note that in Newey and Windmeijer (2005) $Eg_i = 0$, so terms in (77) simplify. In our case, we have Assumption M1, so $Eg_i = C_1/n^{1/2}$. Analyze following term in (77)

$$\frac{n}{q_n}(Eg_i)'A^{j'}\Omega^{-1}Eg_i^k = (\frac{C_1}{q_n^{1/2}})'A^{j'}\Omega^{-1}(\frac{n^{1/2}Eg_i^k}{q_n^{1/2}}).$$

Note that $Eig_{max}(A^{j'}\Omega^{-1}) \leq C$ by Assumption M3 and M4iii. So

$$\frac{n}{q_n}(Eg_i)'A^{j'}\Omega^{-1}Eg_i^k \leq C(\frac{C_1}{q_n^{1/2}})'(\frac{n^{1/2}Eg_i^k}{q_n^{1/2}}) \to 0, \tag{78}$$

by Assumption M1 and M6 (i.e. Definition of G_n). Next

$$\frac{n}{q_n}(Eg_i)'A^{j'}\Omega^{-1}A^j(Eg_i) \to 0, \tag{79}$$

by the same analysis in (78). So using definition of G_n and (78)(79) we can rewrite (77)

$$\frac{n(\mu_F)'A\mu_W}{q_n} = e_j'G_n'\Omega^{-1}G_ne_k + o_p(1), \tag{80}$$

where e_j represents the j th unit vector. Next,

$$tr(A\Sigma_{FW})/q_n = tr(\Omega^{-1}EU_i^kU_i^{j'})/q_n + o_p(1) = e_j'\Lambda_ne_k/q_n + o_p(1), \tag{81}$$

by Assumption M1 and $\Lambda_n = nE[\tilde{U}'\Omega^{-1}\tilde{U}]/q_n$. Then combine (80)(81) to have

$$\tilde{D}_j(\theta_0)'\Omega^{-1}\tilde{D}_k(\theta_0) = e_j'G_n'\Omega^{-1}G_ne_k + e_j'\Lambda_ne_k/q_n + o_p(1) = H_{jk} + e_j'\Lambda_ne_k/q_n + o_p(1).$$

Conclusion follows by Assumption M6 and $\Lambda_n \to \Lambda^*$.**Q.E.D**

Proof of Theorem 8. Apply Lemma A9 with Lemma A12.**Q.E.D**

References

ACEMOGLU, D., JOHNSON, S., ROBINSON, J.A (2001): "The colonial origins of comparative development: An empirical investigation," *American Economic Review* 91, 1369-1401.

ACEMOGLU, D., JOHNSON, S., ROBINSON, J.A. (2002): "Reversal of fortune: Geography and development in the making of the modern world income distribution," *Quarterly Journal of Economics* 117, 1231-1294.

ANDERSON, T.W., RUBIN H. (1949): "Estimators of the parameters of a single equation in a complete set of stochastic equations," *The Annals of Mathematical Statistics* 21, 570-582.

ANDREWS, D.W.K. (1994): "Empirical process methods in econometrics," Handbook of Econometrics, vol.4. North-Holland.

ANGRIST, J.D. (1990): "Lifetime earnings and the Vietnam era draft lottery: Evidence from Social Security Administrative Records," *American Economic Review* 80, 313-336.

ANGRIST, J.D., KRUEGER, A.B. (1991): "Does compulsory school attendance affect schooling and earnings?," *Quarterly Journal of Economics* 106, 979-1014.

BEKKER, P.A. (1994): "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 63, 657-681.

BOUND, J., JAEGER, D.A., BAKER, R.M. (1995): "Problems with instrumental variable estimation when the correlation between the instruments and the endogenous explanatory variable is weak," *Journal of the American Statistical Association* 90, 443-450.

BROWN, B.W. AND W.K. NEWEY (1992): "Bootstrapping for GMM," Working paper, Department of Economics, MIT.

CANER, M. (2004): "Exponential Tilting with Weak Instruments: Estimation and Testing," University of Pittsburgh, Working Paper.

CHAO, J.C. AND N.R. SWANSON (2002): "Consistent Estimation with a Large Number of Weak Instruments," Working paper, Rutgers University.

DUFOUR, J.M. (1997): "Some impossibility theorems in econometrics with applications to structural and dynamic models," *Econometrica* 65, 1365-1387.

DUFOUR, J.M. (2003): "Identification, weak instruments and statistical inference in econometrics," Presidential address to the Canadian Economics Association. Canadian Journal of Economics 36, 767-808.

DUFOUR, J.M., TAAMOUTI, M. (2004): "Further results on projection-based inference in IV regressions with weak, collinear or missing instruments," University of Montreal, Working Paper.

HALL, A.R., INOUE, A. (2003): "The large sample behaviour of the generalized method of moments estimator in misspecified models," *Journal of Econometrics* 114, 361-394.

GLAESER, E.L., LA PORTA, R., LOPEZ-DE SILANES, F., SHLEIFER, A. (2004): "Do Institutions cause growth?," *Journal of Economic Growth* 9, 271-303.

GUGGENBERGER, P., SMITH, R.J. (2005): "Generalized empirical likelihood estimators and tests under partial, weak and strong identification," *Econometric Theory*, Forthcoming.

GUGGENBERGER, P., WOLF, M. (2004): "Subsampling tests of parameter hypotheses and overidentifying restrictions with possible failure of identification," UCLA, Working Paper.

HAN, C. AND P.C.B. PHILLIPS (2004): "GMM With Many Moment Conditions," Working Paper, Yale University.

IMBENS, G. (1993): " A New Approach to Generalized Method on Moments Estimation". Harvard Institute of Economic Research Working Papers 1633, Harvard - Institute of Economic Research.

IMBENS, G.W., SPADY, R.H., JOHNSON, P. (1998): "Information theoretic approaches to inference in moment condition models," *Econometrica* 66, 333-357.

KITAMURA, Y., STUTZER, M. (1997): "An information theoretic alternative to generalized method of moments estimation," *Econometrica* 65, 861-875.

KLEIBERGEN, F. (2002): "Testing in parameters in GMM without assuming that they are identified," Brown University, Working Paper.

LOH, W.Y. (1987): "Calibrating confidence coefficients," *Journal of the American Statistical Association* 82, 155-162.

NEWEY , W.K. (1985): "Generalized Method of Moments Specification Testing," *Journal of Econometrics* 29, 229-256.

NEWEY, W.K., SMITH, R.J. (2004): "Higher order properties of GMM and generalized empirical likelihood estimators," *Econometrica* 72, 219-255.

NEWEY, W.K. AND F. WINDMEIJER (2005): "Many Weak Moment Asymptotics for Generalized Empirical Likelihood Estimators," Working Paper, Department of Economics, MIT.

OWEN0, A. (1990): Empirical Likelihood. Chapman-Hall.

POLITIS, D.N., ROMANO, J.P., WOLF M., (1997): "Subsampling for heteroskedastic time series," *Journal of Econometrics* 81, 281-319.

POLITIS, D.N., ROMANO, J.P., WOLF M. (1999): Subsampling, Springer Verlag. New York.

POLLARD, D. (1990): Empirical Processes: Theory and applications. Institute of Mathematical Statistics Publication.NSF-CBMS Regional Conference Series, vol.2.

QIN, J., LAWLESS, J. (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 Monets Estimation," *Economic Journal*, 107, 503-519.

STOCK, J.H., WRIGHT, J.H. (2000): *GMM with weak identification* Econometrica 68, 1055-1096.

STOCK, J.H., WRIGHT, J.H., YOGO, M. (2002): "A survey of weak instruments and weak identification in generalized method of moments," *Journal of Business and Economic Statistics* 20, 518-529.

VAN DER VAART, A.W (2000): Asymptotic Statistics. Paperback. 2nd. Ed. Cambridge University Press.

VAN DER VAART, A.W., WELLNER, J. (1996): Weak convergence and empirical processes. Springer-Verlag, New York.

WOOLDRIDGE, J.W. (2002): Econometric Analysis of Cross Section and Panel Data. MIT press.

Table 1: Size of Tests using χ^2 critical values

GELR (θ_0), EL			
	$cov(X_i, u_i) = 0.1$	$cov(X_i, u_i) = 0.05$	$cov(X_i, u_i) = 0.00$
$\pi = 1$	36.2	16.0	13.0
$\pi = 0.1$	35.5	17.4	11.7
$\pi = 0$	35.5	18.8	12.4
GELR (θ_0), ET			
$\pi = 1$	35.3	17.1	10.9
$\pi = 0.1$	34.9	17.8	12.6
$\pi = 0$	38.7	19.5	11.6
GELR (θ_0), CUE			
$\pi = 1$	33.9	17.3	11.2
$\pi = 0.1$	33.0	16.2	10.3
$\pi = 0$	33.4	17.1	9.0
LM (θ_0), EL			
$\pi = 1$	40.1	20.0	10.7
$\pi = 0.1$	31.7	18.7	10.6
$\pi = 0$	28.6	15.6	9.8
LM (θ_0), ET			
$\pi = 1$	39.6	20.9	9.0
$\pi = 0.1$	28.9	16.9	9.3
$\pi = 0$	27.0	15.5	11.5
LM (θ_0), CUE			
$\pi = 1$	45.3	28.0	15.5
$\pi = 0.1$	41.5	18.4	9.3
$\pi = 0$	38.4	18.0	9.9

For $GELR(\theta_0)$ test we use χ^2_1 distribution at 10% level. The critical value is 2.71. For $LM(\theta_0)$ test we use χ^2_2 at 10% level. The critical value is 4.61. The columns refer to covariance between the instruments and the structural error. Both instruments have the same covariance. The reduced form coefficient π comes from the model (20)-(21), and determines the strength of instruments.

Table 2: Size of Tests using Subsampling Approach

GELR (θ_0), EL			
	$cov(X_i, u_i) = 0.1$	$cov(X_i, u_i) = 0.05$	$cov(X_i, u_i) = 0.00$
$\pi = 1$	8.2	3.9	2.9
$\pi = 0.1$	8.3	3.6	2.5
$\pi = 0$	9.5	4.7	1.8
GELR (θ_0), ET			
$\pi = 1$	15.1	5.7	4.7
$\pi = 0.1$	16.0	6.3	4.3
$\pi = 0$	13.9	6.1	4.5
GELR (θ_0), CUE			
$\pi = 1$	20.9	10.4	6.7
$\pi = 0.1$	22.8	10.1	4.9
$\pi = 0$	19.8	9.0	6.6
LM (θ_0), EL			
$\pi = 1$	13.4	11.3	10.2
$\pi = 0.1$	9.7	10.8	10.5
$\pi = 0$	11.4	10.8	9.5
LM (θ_0), ET			
$\pi = 1$	31.5	15.7	7.9
$\pi = 0.1$	20.7	10.6	7.3
$\pi = 0$	19.3	10.6	8.3
LM (θ_0), CUE			
$\pi = 1$	21.4	10.2	6.2
$\pi = 0.1$	24.7	9.9	6.7
$\pi = 0$	24.1	11.1	7.3

The columns refer to covariance between the instruments and the structural error. Both instruments have the same covariance. The reduced form coefficient π comes from the model (20)-(21), and determines the strength of instruments.

Figure 1a: Near Exogeneity and Strong Instruments: $cov(X,u)=0.1, \pi=1$

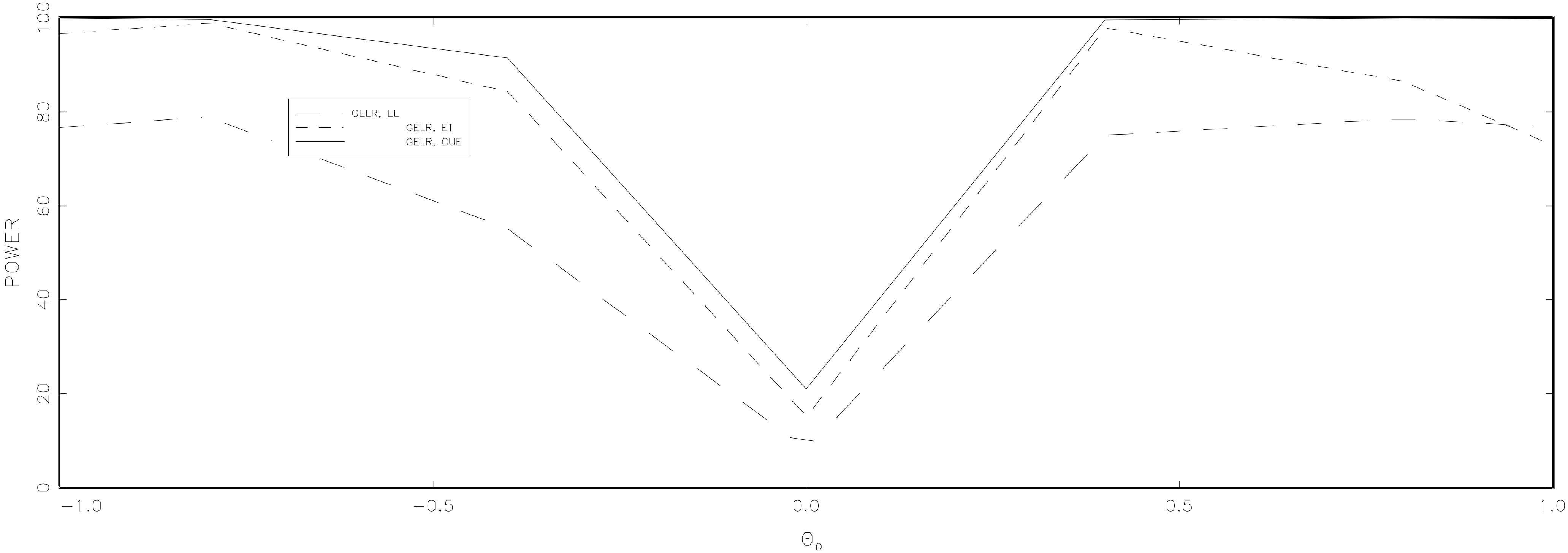

Figure 1b: Near Exogeneity and Weak Instruments: $cov(X,u)=0.1, \pi=0.1$

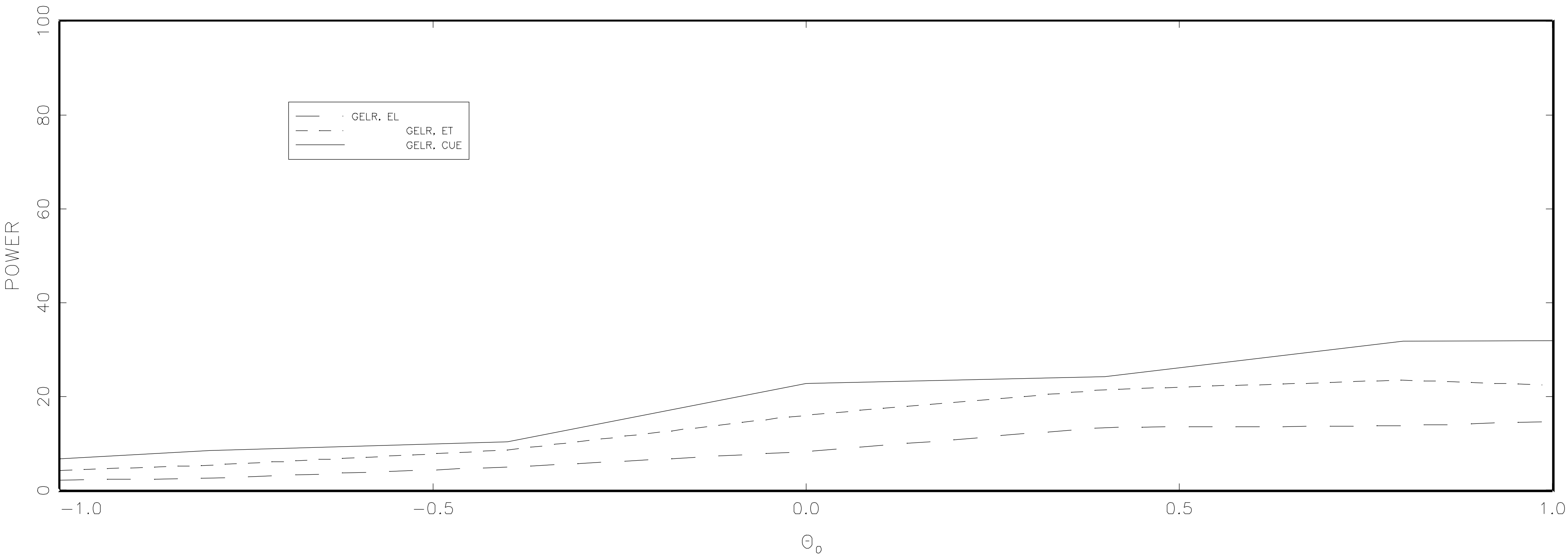

Figure 2a: Near Exogeneity and Strong Instruments: cov(X,u)=0.1,π =1

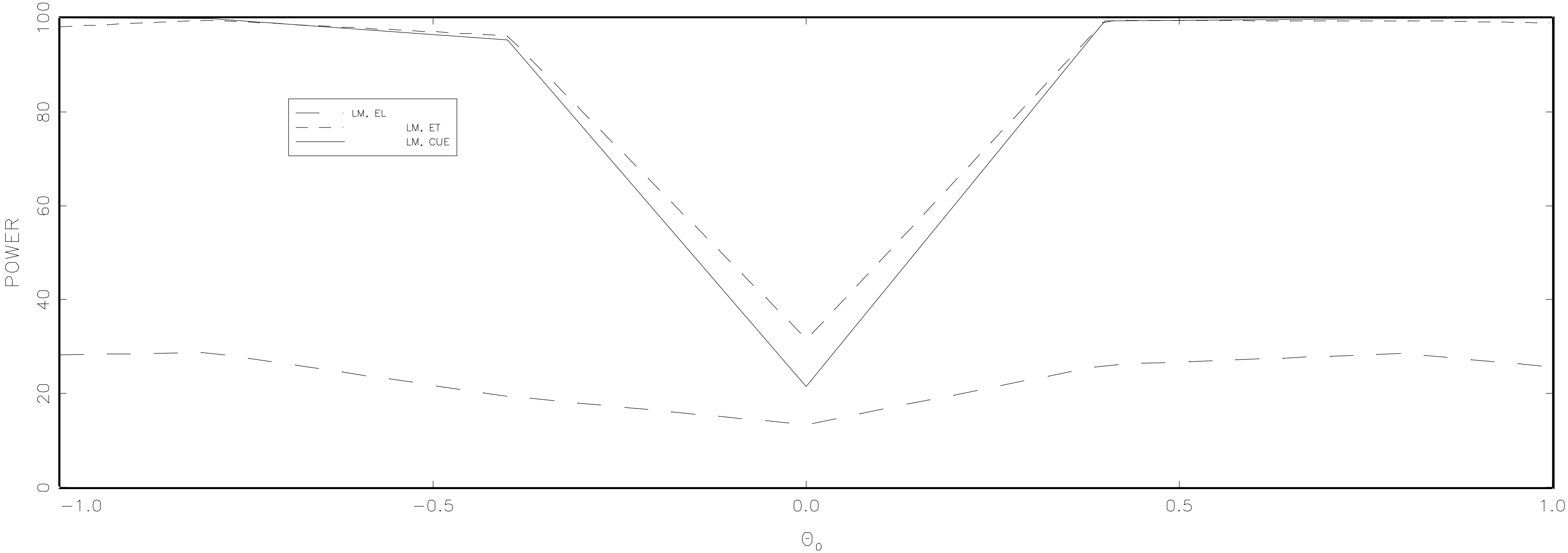

Figure 2b: Near Exogeneity and Weak Instruments: cov(X,u)=0.1,π =0.1

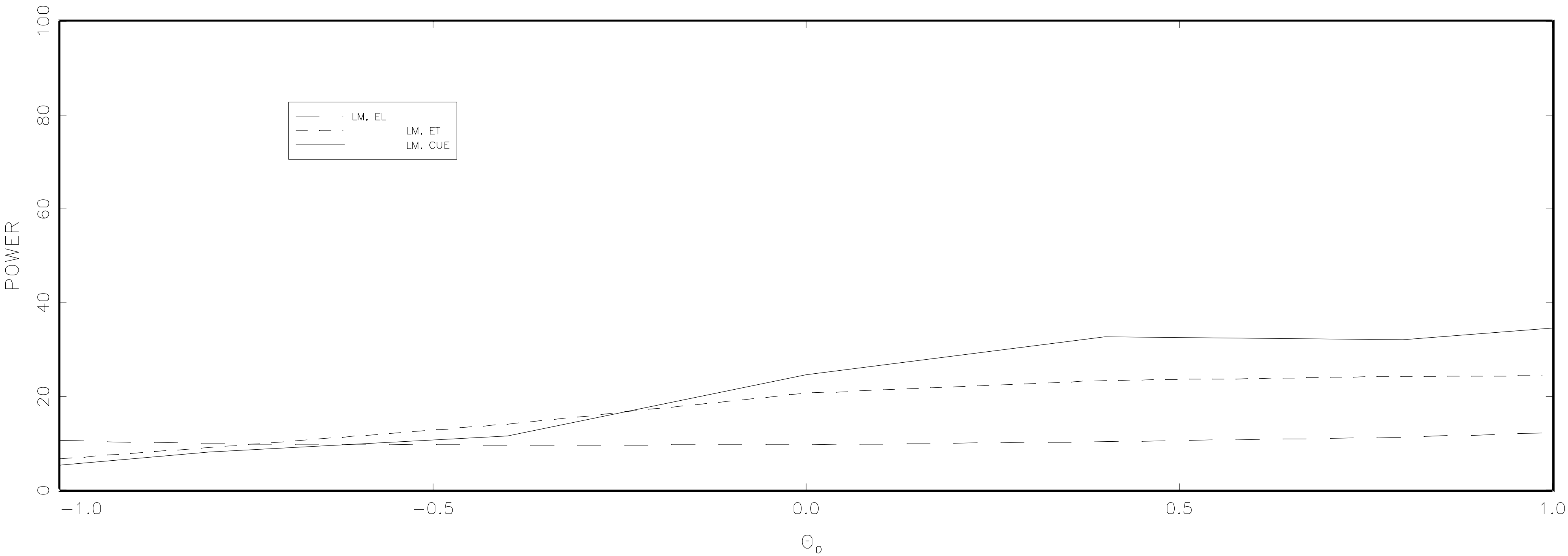

Figure 3A: Near Exogeneity and Strong Instruments: cov(X,u)=0.1,π =1

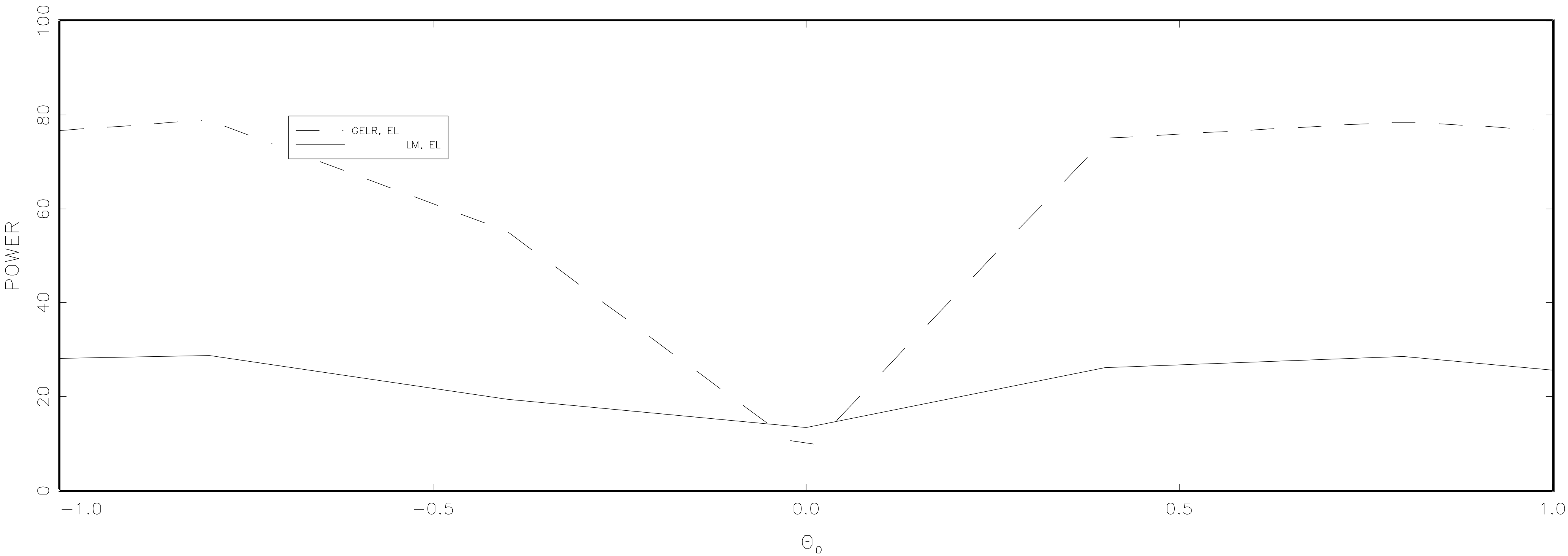

Figure 3b: Near Exogeneity and Weak Instruments: cov(X,u)=0.1,π =0.1

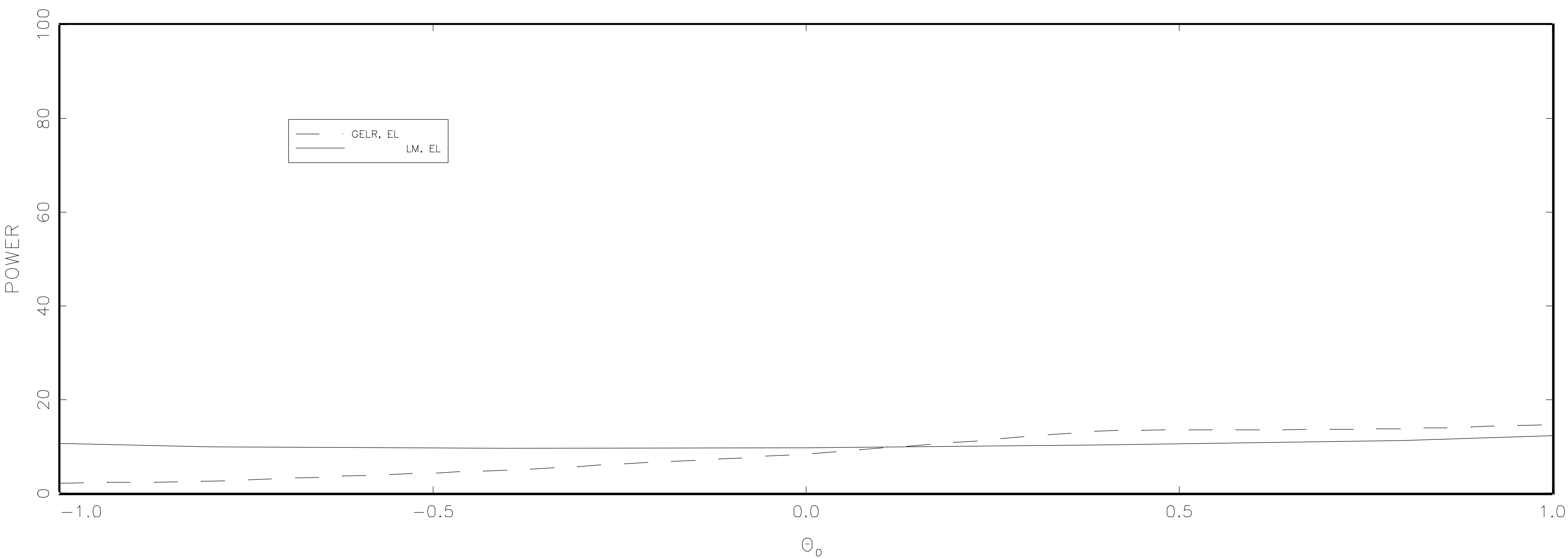

Figure 4A: Near Exogeneity and Strong Instruments: cov(X,u)=0.1,π =1

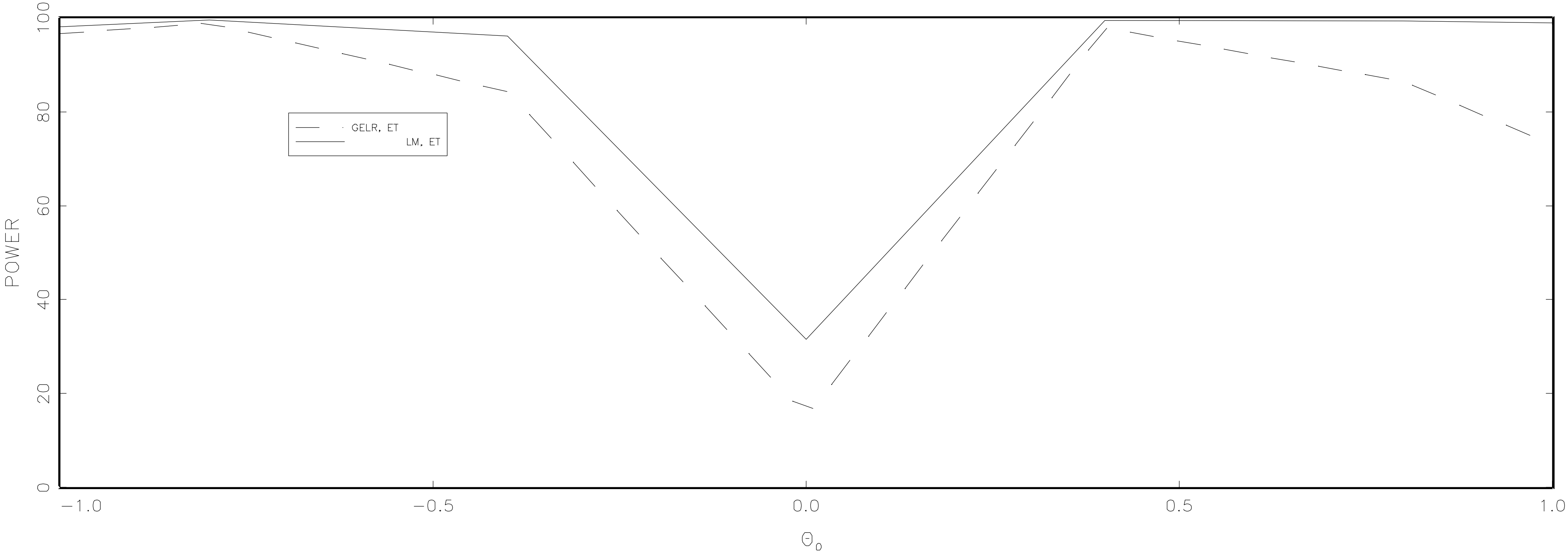

Figure 4b: Near Exogeneity and Weak Instruments: cov(X,u)=0.1,π =0.1

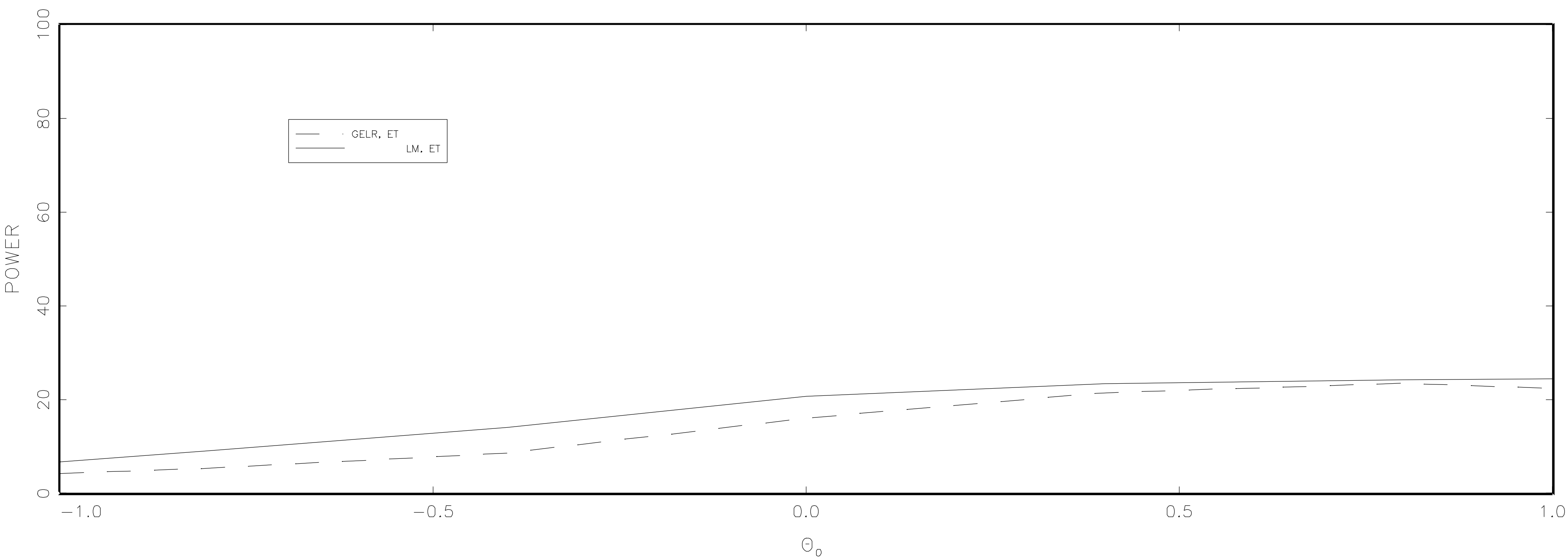

Figure 5A: Near Exogeneity and Strong Instruments: cov(X,u)=0.1,π =1

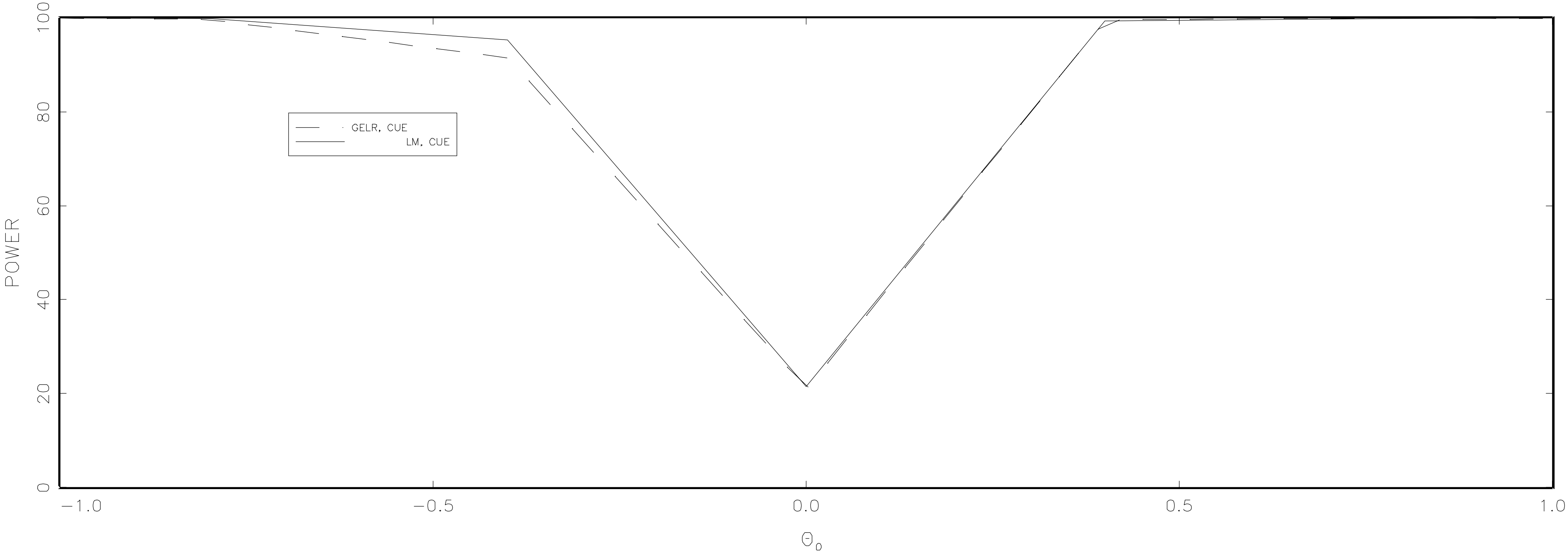

Figure 5b: Near Exogeneity and Weak Instruments: cov(X,u)=0.1,π =0.1