Testing, Estimation and Higher Order Expansions in GMM with Semi-Weak Instruments

Mehmet Caner
University of Pittsburgh
Preliminary *

January 19, 2004

Abstract

In this paper we analyze GMM with semi-weak instruments. This case includes the standard GMM and the nearly-weak GMM. In the nearly weak-GMM, the correlation between the instruments and the first order conditions decline at a slower rate than root T. We find an important difference between the semi-weak case and the weak case. Inference with point estimates are possible with Wald, Likelihood Ratio and Lagrange Multiplier tests in GMM with semi-weak instruments. The limit is the standard χ^2 limit. This is important from an applied perspective since tests on the weak case do depend on the true value and can only test simple null. Even though we may have all nearly-weak instruments in GMM it is still possible to test various hypothesis of interest. We also find a difference between the two subcategories in the semi-weak case. We derive higher order expansions for test statistics in the semi-weak case, and we show that with declining quality of instruments finite sample bahavior of these tests get worser, so standard GMM finite sample behavior is always better than nearly-weak GMM. Unlike the standard GMM, in the nearly-weak GMM we can not eliminate the second order terms from these test statistics's expansions. Then analytically we show the power of these tests are not good in the nearly-weak case compared to standard GMM. We also see that if we want to test restrictions; Wald test in Continuous Updating Estimator has desirable properties in higher order expansions in the semi-weak case. Clearly, this paper shows that when we move away from the standard GMM towards nearly-weak case, finite sample behavior suffers but large sample theory remains intact. But if we further move away from the nearly-weak case to weakly identified GMM the large sample theory changes too.

*Department of Economics, WW Posvar Hall 4S01, Pittsburgh, PA 15260. email: caner@pitt.edu.

1 Introduction

In the recent years, there has been a surge of papers analyzing the weak instruments in GMM. One of the main reasons for this interest is the inability of GMM estimates to approach normal distributions even with large samples. This behavior is tied to the weak instrument asymptotics developed in a seminal paper by Stock and Wright (2000). In the case of weak instruments the authors derive a new large sample theory that is nonnormal and nonstandard. Furthermore they find the GMM estimates are inconsistent. Testing in the case of weak instruments in GMM can be done by the methods proposed by Stock and Wright (2000) and Kleibergen (2001). These tests only work under the true value of the parameter.

Another reason for the interest in weak instruments problem is the finite sample properties of GMM estimators and test statistics. These have finite sample bias and tests have size problems, and this behavior is affected by the quality of instruments as shown in Hansen, Heaton and Yaron (1996), and Stock, Wright and Yogo (2002). All these studies analyzed the case of "weak instruments" described by the low correlation between instruments and orthogonality restrictions. Specifically, this correlation is set to decline at the rate of root T, where T is the sample size.

In an important recent paper; Hahn and Kuersteiner (2002) analyzed the "nearly-weak" instruments in a linear IV structure. In their setup the correlation between instruments and the orthogonality restrictions decline at a slower rate than root T. This reults in consistent estimates but slower rate of convergence with normal limits. They also carefully examine two-stage least squares estimators in higher order expansions. So there is a distinction between the weak and nearly-weak cases in terms of large sample theory. In the nearly-weak case the two-stage least squares estimators have the same limit as in the standard strong two stage-least squares estimators.

In this paper, we consider "semi-weak" GMM which consists of "nearly-weak" and the standard GMM cases. This is not a trivial extension of the linear two-stage least squares case in Hahn and Kuersteiner (2002) paper since we use empirical process theory with changing classes to derive the limit theory in nonlinear GMM models including both two-step and CUE versions and various test statistics. The theoretical approach is also somewhat different than the theoretical perspective in Stock and Wright (2000) because it is not immediately clear how to model the semi-weak case in a system with weak instruments.

We show that in GMM context; standard, nearly-weak and weak cases are different. First, via higher order expansions we analytically show that Wald, LR, LM tests' finite sample behavior is affected by the quality of instruments. Hence the GMM estimates' finite sample behavior in the standard and nearly-weak cases are different. With declining quality of instruments, the higher order elements become more effective and distort the finite sample behavior of these tests. We also cannot correct the problems via second-order Edgeworth approximations in these tests. However, Wald test in CUE framework and K test of Kleibergen (2001) have few higher-order terms compared to other tests and less affected by the nearly-weak instruments problem. We also show analytically that test statistics donot have power against certain local alternatives in nearly-weak case unlike the standard GMM. Power against fixed alternatives also depend on the quality of the instruments even though the tests are consistent in large samples.

Next, we see that the distinction between the "nearly-weak" and the weak cases in GMM. In the case of "nearly-weak" instruments even though the correlation between the instruments and the orthogonality restrictions are not high we can get point estimates and use Wald, LR, LM tests. So testing various restrictions using point estimates in the nearly-weak case is possible unlike the weakly identified GMM. This is important from an applied perspective. It is then essential to distinguish between nearly-weak and weak cases if we want to test restrictions. We also provide an ad-hoc indicator of differentiating between the weak and the near-weak cases.

Apart from these findings we provide a rigorous theoretical treatment of a mixed weak, semi-weak case. Since the limits of the nearly-weak and standard GMM cases are the same they are analyzed under the category of GMM with semi-weak instruments. This mixed case shows the limit behavior of GMM estimates in various possible settings regarding the instruments, and we can see how they interact.

Section 2 provides the model and assumptions. Section 3 analyzes GMM with mixed weak and semi-weak instruments and provides the large sample estimates. Section 4 looks at the various test statistics limit behavior in the case of semi-weak instruments. Section 5 analyzes the power issues analytically. Section 6 considers the higher order expansion of these test statistics in GMM in the case of semi-weak instruments, hence extend the work of Kleibergen (2003). Section 7 concludes.

2 The Model and Assumptions

We benefit largely from the framework of Stock and Wright (2000). Let θ be a p-dimensional parameter vector, and θ_0 represents the population value which is in the interior of the compact parameter space $\Theta \subset R^p$. The population orthogonality conditions are of G dimension:

$$E[h(Y_t, \theta_0)|F_t] = 0, \tag{1}$$

where F_t is the information set. Let Z_t be a K-dimensional vector of instruments contained in F_t. The data are $\{(Y_t, Z_t) : t = 1, 2, \cdots, T\}$. The GMM estimator $\hat{\theta}_{GMM,T}$ minimizes the following over $\theta \in \Theta$

$$S_T(\theta; \bar{\theta}_T(\theta)) = [T^{-1/2} \sum_{t=1}^{T} \psi_t(\theta)]' W_T(\bar{\theta}_T(\theta))[T^{-1/2} \sum_{t=1}^{T} \psi_t(\theta)],$$

where $\psi_t(\theta) = h(Y_t, \theta) \bigotimes Z_t$ and $W_T(\bar{\theta}_T(\theta))$ is an $O_p(1)$ positive definite $GK \times GK$ weight matrix. In the case of efficient two-step GMM, $\bar{\theta}_T$ does not depend on θ and uses a preliminary estimator.

For the Continuous Updating Estimator (CUE) $\bar{\theta}_T = \theta$.

A very important notion in Stock and Wright (2000) is the introduction of weak instruments idea and the assumption related to that. In this paper we introduce the concept of semi-weak instruments in GMM framework and analyze estimation, testing and higher order expansions of the test statistics. In the linear case this is introduced by Hahn and Kuersteiner (2002) and they analyze the two-stage least squares estimator and the higher order expansion related to the estimator. Semi-weak instruments are not as weak as the ones in Stock and Wright (2000) but weaker than strong instruments in standard GMM literature as in Hansen (1982), Newey and McFadden (1994). In the following discussion and with the limit results we formalize this notion of semi-weak instruments.

Identification is an aspect of the moment condition $E\psi_t(\theta)$. In finite samples, as suggested in weak identification literature by Stock and Wright (2000) the value of the moment function might be small for a large set of θ, so the population objective function cannot provide information to discriminate among these values. In this respect, we can decompose the parameter set into weakly and semi-weakly identified parameters. This work extends and complements the work of Stock and Wright (2000).

First set $\theta = (\alpha', \beta')'$, where α is $p_1 \times 1$, β is $p_2 \times 1$. We treat α as weakly identified and β as semi-weakly identified. As in Stock and Wright (2000), let

$$ET^{-1}\sum_{t=1}^{T}\psi_t(\alpha,\beta) = \tilde{m}_T(\alpha,\beta).$$

Basically the above equation is written as identity of the following form:

$$\tilde{m}_T(\alpha,\beta) = \tilde{m}_T(\alpha_0,\beta_0) + \tilde{m}_{1T}(\alpha,\beta) + \tilde{m}_{2T}(\beta), \tag{2}$$

where $\tilde{m}_{1T}(\alpha,\beta) = \tilde{m}_T(\alpha,\beta) - \tilde{m}_T(\alpha_0,\beta)$, and $\tilde{m}_{2T}(\beta) = \tilde{m}_T(\alpha_0,\beta) - \tilde{m}_T(\alpha_0,\beta_0)$. Since $E\psi_t(\alpha_0,\beta_0) = 0$, $\tilde{m}_T(\alpha_0,\beta_0) = 0$ and $\tilde{m}_{2T}(\alpha_0,\beta_0) = 0$. We treat $\tilde{m}_{1T}(\alpha,\beta)$ as small for all α, β. The population objective function is nearly flat in α. Regarding β , the population objective function is also nearly flat in β but steeper in this parameter compared to α. This is one of the differences with Stock and Wright (2000) framework.

In Stock and Wright (2000), $\tilde{m}_{2T}$ assumed to be constant.(This is set up to save from notation in their case, actually $\tilde{m}_{2T}$ converges uniformly to a nonrandom function.) However, unlike the strong instruments case in Stock and Wright (2000) we link $\tilde{m}_{2T}(\beta)$ to the sample size T and this $\tilde{m}_{2T}$ goes to zero. We link the expectation of the moment condition to the sample size in both parameters, but by differing magnitudes. Specifically, as in Stock and Wright (2000), for weakly identified parameters the function $\tilde{m}_{1T}(\alpha,\beta)$ go to zero at the rate $T^{1/2}$. For the semi-weakly identified parameters; this again decays to zero but at a slower rate of T^{κ}, $0 \leq \kappa < 1/2$. We consider basically a system with weak and semi-weak instruments. Note that when $\kappa = 0$, the parameters are strongly identified. So the results of Stock and Wright (2000) is a special case of our results since they analyze mixed weak and standard cases.

We use the term "nearly-weak" instruments for $0 < \kappa < 1/2$ and "semi-weak" instruments for $0 \leq \kappa < 1/2$. The term "nearly-weak" is introduced by Hahn and Kuersteiner (2002). Semi-weak case includes both standard GMM and "nearly-weak" cases. We setup the model in this way because as we show below the large sample theory for semi-weak case can be explained as one case, rather than seperately analyzing the standard and nearly-weak cases. Note that the extension to mixed weak and semiweak case in GMM is not trivial compared to linear "nearly-weak" case of Hahn and Kuersteiner (2002). Now we can supply the identification assumption.

Assumption 1.

$$ET^{-1}\sum_{t=1}^{T}\psi_t(\theta) = \frac{m_{1T}(\theta)}{T^{1/2}} + \frac{m_{2T}(\beta)}{T^{\kappa}},$$

where $0 \leq \kappa < 1/2$, *and*

(i). $m_{1T}(\theta) \to m_1(\theta)$ *uniformly in* $\theta \in \Theta$, $m_1(\theta_0) = 0$, *and* $m_1(\theta)$ *is continuous in* θ *and is bounded on* Θ.

(ii). $m_{2T}(\beta) \to m_2(\beta)$ *uniformly in* β, $m_2(\beta)$ *is continuous and* $m_{2T}(\beta)$ *is continuously differentiable. Define* $R_T(\beta) = \frac{\partial m_{2T}(\beta)}{\partial \beta'}$ *as* $GK \times p_2$ *matrix. Assume also* $R_T(\beta) \to R(\beta)$ *uniformly in a* N *neighborhood of* β_0, $R(\beta_0)$ *is of full column rank. Note that* $R(\beta)$ *is continuous in* N.

Next we give two assumptions that are useful in providing the limit law.

Assumption 2.

(i). $\psi_t(\theta)$ *is m-dependent.*

(ii).

$$|\psi_t(\theta_1) - \psi_t(\theta_2)| \leq B_t|\theta_1 - \theta_2|,$$

where $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} EB_t^{2+\delta} < \infty$, *for some* $\delta > 0$.

(iii).

$$\sup_{\theta\in\Theta} E|\psi_t(\theta)|^{2+\delta} < \infty,$$

for some $\delta > \infty$.

Assumption 3. W_T *is positive definite and* $W_T(\theta) \xrightarrow{p} W(\theta)$ *uniformly in* θ, *where* $W(\theta)$ *is a symmetric nonrandom* $GK \times GK$ *matrix. That is continuous in* θ *and is positide definite for all* $\theta \in \Theta$.

Define the empirical process as

$$\Psi_T(\theta) = T^{-1/2}\sum_{t=1}^{T}\psi_t(\theta) - E\psi_t(\theta). \tag{3}$$

Under Assumption 2 we obtain the following result via Andrews (1994)

$$\Psi_T(\theta) \Longrightarrow \Psi(\theta). \tag{4}$$

where $\Psi(\theta)$ is a Gaussian Stochastic process on Θ with mean zero and covariance function $E\Psi(\theta_1)\Psi(\theta_2)' = \Omega_{\theta_1,\theta_2}$.

In this paper, we show that in semi-weak instruments case, the estimators are consistent. Furthermore, in a case of all semi-weak instruments in the coming sections the limit is asymptotically normal. Semi-weak case combines both the elements of strong and weak cases. Loosely speaking, like the weak instruments case the correlation between instruments and the orthogonality conditions goes to zero in large samples. However, since this is not decaying as fast as the weak instruments case we still have consistency as strong instruments case. These points are explained in detail in the following sections.

3 The Limit for Mixed Weak and Semi-Weak Instruments Case

In this section we analyze the model given by Assumption 1 (a mixed model), and specifically consider the limit of estimators of the parameters. Before the limit theory, let $\hat{\beta}(\alpha)$ solve $arg\min_{\beta\in B} S_T(\alpha,\beta;\bar{\theta}_T(\alpha,\beta))$. Let $\hat{\alpha}$ solve $arg\min_{\alpha\in A} S_T(\alpha,\hat{\beta}(\alpha);\bar{\theta}_T(\alpha,\hat{\beta}(\alpha))$, and let $\hat{\beta}=\hat{\beta}(\hat{\alpha})$.

First we provide the consistency and the rate of convergence result. This generalizes Lemma A.1 of Stock and Wright (2000) to the semi-weak case from the pure strong case (i.e. strong instrument case is a subcase of ours). Stock and Wright (2000) show that the estimates of well identified parameters in their case are consistent, and the rate of convergence of the estimator is $T^{1/2}$. In this paper, even though the correlation between the instruments and orthogonality conditions goes to zero in large samples, since the decay rate is slow we are able to achieve consistency in the semi-weak case. This result is new in GMM literature and shows that the rate of decay in correlation between the instruments and the orthogonality conditions is important, not the decay itself. Hahn and Kuersteiner (2002) derived the same result for two-stage least squares estimator in a linear case with a system of only "nearly-weak" instruments. This following result also generalizes their result to GMM (both two-step and Continuous Updating Estimator) and to a mixed weak and semi-weak instruments case.

Theorem 1. *Under Assumptions 1-3,*

$$T^{1/2-\kappa}(\hat{\beta}-\beta_0) = O_p(1),$$

where $0 \leq \kappa < 1/2$.

Remark. This clearly shows that the link between the instrument quality and the rate of convergence . If κ is near zero, then correlation between instruments and the orthogonality restrictions decline very slowly, hence the rate of convergence "$T^{1/2-\kappa}$" will be near the standard rate of $T^{1/2}$. If on the other hand κ is near $1/2$ but $\kappa < 1/2$, the rate of convergence is very slow, we are near the setup of weak instruments of Stock and Wright (2000). We can also observe several things about the theorem, if it were to happen that $\kappa = 1/2$ (Stock and Wright (2000) weak instruments case) then β is weakly identified and there is no consistency. When $\kappa = 0$, the convergence rate is $T^{1/2}$ we are in the strong instruments case as in Lemma A.1 of Stock and Wright (2000). If $0 < \kappa < 1/2$, we see that still we have consistency, but rate of convergence is affected. From the theorem it is clear that strong instruments case (Lemma A.1 of Stock and Wright (2000)) is a special case of ours (i.e. $\kappa = 0$).

We can now provide the limit theory for both estimators, it is clear from the above theorem that α will not be consistent. The following theorem shows the limit law for the estimators of GMM in a mixed system of weak and semi-weak instruments. This generalizes the mixed weak and standard instruments case of Stock and Wright (Theorem 1, 2000).

Theorem 2. *Under Assumptions 1-3, and $\bar{\theta}_T(\theta) \Longrightarrow \bar{\theta}$ uniformly in θ. Then*

$$(\hat{\alpha}', T^{1/2-\kappa}(\hat{\beta} - \beta_0)') \xrightarrow{d} (\alpha^{*'}, b^{*'}),$$

where

$$\alpha^* = arg \min_{\alpha \in A} S^*(\alpha; \bar{\theta}(\alpha, \beta_0)),$$

$$b^* = -[R(\beta_0)' W(\bar{\theta}(\alpha^*, \beta_0)) R(\beta_0)]^{-1} R(\beta_0)' W(\bar{\theta}(\alpha^*, \beta_0))[\Psi(\alpha^*, \beta_0) + m_1(\alpha^*, \beta_0)],$$

$$S^*(\alpha; \bar{\theta}(\alpha, \beta_0)) = [\Psi(\alpha, \beta_0) + m_1(\alpha, \beta_0)]' M(\alpha, \beta_0; \bar{\theta}(\alpha, \beta_0))[\Psi(\alpha, \beta_0) + m_1(\alpha, \beta_0)],$$

$$M(\alpha, \beta_0; \bar{\theta}(\alpha, \beta_0)) = W(\bar{\theta}(\alpha, \beta_0)) - W(\bar{\theta}(\alpha, \beta_0)) R(\beta_0)[R(\beta_0)' W(\bar{\theta}(\alpha, \beta_0)) R(\beta_0)]^{-1} R(\beta_0)' W(\bar{\theta}(\alpha, \beta_0)).$$

Set $\bar{\psi}(\theta) = T^{-1} \sum_{t=1}^{T} \psi_t(\theta)$. To obtain the limits for the efficient two-step GMM and CUE we need the following Assumption about the heteroskedasticity robust weighting matrices

Assumption 4. *Let*

$$V_T(\theta) = T^{-1} \sum_{t=1}^{T} [\psi_t(\theta) - \bar{\psi}(\theta)][\psi_t(\theta) - \bar{\psi}(\theta)]'. \tag{5}$$

Then uniformly in θ

$$V_T(\theta) \Longrightarrow \Omega_{\theta,\theta}.$$

The one-step GMM estimator $\hat{\theta}_1$ uses I_{GK} as weight matrix. For efficient two-step GMM estimator $\hat{\theta}_2$ the weight matrix is : $W_T(\bar{\theta}(\theta)) = V_T(\hat{\theta}_1)^{-1}$. For efficient CUE the weight matrix is : $W_T(\bar{\theta}(\theta)) = V_T(\theta)^{-1}$.

In order to understand the limits in the following theorem, we need the following notation. Let $\mu(\alpha) = \Omega_{\alpha,\beta_0}^{-1/2'} m_1(\alpha, \beta_0)$ and $z(\alpha) = \Omega_{\alpha,\beta_0}^{-1/2'} \Psi(\alpha, \beta_0)$ so $z(\alpha)$ is a mean-zero GK dimensional Gaussian process in α with covariance function $Ez(\alpha_1) z(\alpha_2)' = \Omega_{\alpha_1,\beta_0}^{-1/2'} \Omega_{(\alpha_1,\beta_0)',(\alpha_2,\beta_0)'} \Omega_{\alpha_2 \beta_0}^{-1/2}$. Before two-step result note that for one-step estimator (i.e. where we chose $W_T = I_{GK}$) simple applying Theorem 2 with the given weight matrix for this case:

$$\alpha_1^* = \arg\min_{\alpha}[z(\alpha)+\mu(\alpha)]' Q_1(\alpha)[z(\alpha)+\mu(\alpha)],$$

and

$$Q_1(\alpha) = \Omega^{1/2}_{\alpha,\beta_0}\{I - R(\beta_0)[R(\beta_0)'R(\beta_0)]^{-1}R(\beta_0)'\}\Omega^{1/2}_{\alpha,\beta_0}.$$

Simple application of Theorem 2 gives the following Corollaries.

Corollary 1 . *Under Assumptions 1,2,4 for the efficient two-step GMM estimator,* $\hat\theta_2$

$$(\hat\alpha_2', T^{1/2-\kappa}(\hat\beta_2-\beta_0)') \Longrightarrow (\alpha_2^{*'}, b_2^{*'}),$$

where

$$\alpha_2^* = \arg\min_{\alpha\in A} S_2^*(\alpha),$$

$$S_2(\alpha) = [z(\alpha)+\mu(\alpha)]' Q_2(\alpha)[z(\alpha)+\mu(\alpha)],$$

$$Q_2(\alpha) = \Omega^{1/2}_{\alpha,\beta_0}\{\Omega^{-1}_{\alpha_1^*,\beta_0} - \Omega^{-1}_{\alpha_1^*,\beta_0}R(\beta_0)[R(\beta_0)'\Omega^{-1}_{\alpha_1^*,\beta_0}R(\beta_0)]^{-1}R(\beta_0)'\Omega^{-1}_{\alpha_1^*,\beta_0}\}\Omega^{1/2}_{\alpha,\beta_0}.$$

and

$$b_2^* = -[R(\beta_0)'\Omega^{-1}_{\alpha_1^*,\beta_0}R(\beta_0)]^{-1}R(\beta_0)'\Omega^{-1}_{\alpha_1^*,\beta_0}\Omega^{1/2'}_{\alpha_2^*,\beta_0}[z(\alpha_2^*)+\mu(\alpha_2^*)].$$

Similarly for the efficient CUE, $\hat\theta_c$, setting $F(\alpha) = \Omega^{-1/2}_{\alpha,\beta_0}R(\beta_0)$, and

Corollary 2.

$$(\hat\alpha_c', T^{1/2-\kappa}(\hat\beta_c-\beta_0)') \Longrightarrow (\alpha_c^{*'}, b_c^{*'}),$$

where

$$\alpha_c^* = \arg\min S_c^*(\alpha),$$

$$S_c^*(\alpha) = [z(\alpha)+\mu(\alpha)]'[I - F(\alpha)(F(\alpha)'F(\alpha))^{-1}F(\alpha)'][z(\alpha)+\mu(\alpha)],$$

$$b_c^* = -[R(\beta_0)'\Omega^{-1}_{\alpha_c^*,\beta_0}R(\beta_0)]^{-1}R(\beta_0)'\Omega^{-1/2}_{\alpha_c^*,\beta_0}[z(\alpha_c^*)+\mu(\alpha_c^*)].$$

Remark. Theorem 2 and Corollaries extend the limit theory in Stock and Wright for the mixed system with strong and weak identified cases (Theorem 1, Corollary 4, 2000) to the weak and semi-weak cases. Note that semi-weak case corresponds to both strong $\kappa = 0$ and nearly weak case of $0 < \kappa < 1/2$. Compared to Stock and Wright (2000) one major difference is the rate of convergence of near-weakly identified parameter estimates ($\hat\beta$) have slower convergence rate (i.e. $T^{1/2-\kappa}$, and when $0 < \kappa < 1/2$) than the standard rate $T^{1/2}$. But the limits of the estimates are the same as in Stock and Wright (2000). The limit of the weakly identified parameter estimate affects the limit of the semi-weak one, and $\hat\beta$ limit is non-normal, because of inconsistency of $\hat\alpha$.

This points out to the fact that we need a large data set if we have nearly-weak instruments. Large sample approximation of the finite sample distribution of the estimates of β can be very poor with declining instrument quality (for example $0 < \kappa < 1/2$, κ being near $1/2$) . This is analyzed in detail in the following sections of all semi-weak instruments. We now briefly deal with the estimation in the case of only semi-weak instruments. This is a subcase of our Theorem 2 and Corollaries 1 and 2. Now assume $\theta = \beta$, all semi-weak instruments. In Assumption 1 the moment condition is

$$ET^{-1} \sum_{t=1}^{T} \psi_t(\beta) = \frac{m_{2T}(\beta)}{T^{\kappa}}, \tag{6}$$

where $0 \leq \kappa < 1/2$. Basically to get the limit in this case, we need to analyze the proof of Theorem 2. We now provide the limit for the efficient two-step GMM and CUE in the case of only semi-weak instruments.

Corollary 3. *Under (6), Assumptions 1ii, 2, 4 we have*

$$T^{1/2-\kappa}(\hat{\beta} - \beta_0) \xrightarrow{d} N[0, (R(\beta_0)'\Omega^{-1}_{\beta_0,\beta_0} R(\beta_0))^{-1}].$$

Remark. This corollary is one of the main results of this paper. This clearly shows that even though we may have instruments weakly correlated with the orthogonality restrictions (i.e. equation (6) we have the standard normal limit for efficient two-step GMM and CUE. This illustrates that correlation decaying to zero is not important, we may have semi-weak instruments, but again have the same limit as in the case of strong instruments case (standard GMM). The key observation is the rate of the decay in correlation. This result complements and extends the literature in weak instruments asymptotics and introduces the semi-weak case. If $0 < \kappa < 1/2$ we have the same limit as strong instruments case, but the estimators suffer from the slow rate of convergence. This rate depends on the quality of instruments and with declining quality (if κ is near $1/2$) the rate of convergence slows down (the rate of convergence is $T^{1/2-\kappa}$). This has been previously observed by Hahn and Kuersteiner (2002) in two-stage least squares estimator.

When $\kappa = 0$, in the case of all strong instruments we achive the same result as in standard GMM. So the strong case is a subcase of ours. Corollary 3 is one of the reasons we gave the name semi-weak instruments to the setup in (6). The limit is exactly the same as in standard GMM case, the estimators are consistent again as in standard GMM case, but decaying correlation between instruments and the orthogonality restrictions is a property of weak instruments and this results in slower rate of convergence for the estimators. Case of only weak instruments can be seen by the analysis of Stock and Wright (2000) ($\kappa = 1/2$), the estimators are not consistent and the limit is nonstandard.

4 Testing

In this section we analyze testing in the all semi-weak case in (6). The testing for mixed weak, semi-weak case results exactly in the same limit as in Stock and Wright (2000). So this is not pursued here. As can be seen from Thorems 2 and Corollary 2 and 3 the Likelihood ratio and J test for overidentifying restrictions will have nonstandard limits decribed in Corollary 4 of Stock and Wright (2000). These tests are not asymptotically pivotal, hence not usable in practice. In our case, since these limits are the same as in Corollary 4 of Stock and Wright (2000) and follows simply by Theorems 2 and Corollaries 1 and 2. But we try to answer the important question that whether having semi-weak instruments makes a difference in testing compared to weak case. Only considering the case of all semi-weak instruments and comparing with weak instruments case we can answer that question.

4.1 S and K tests

To test a simple null as $H_0 : \theta = \theta_0$, or build a confidence interval using a test statistic based on that Stock and Wright (2000) suggested S-test for GMM based on Anderson-Rubin (1949) test. This is the GMM objective function evaluated at θ_0. Under weaker assumptions than used in estimation we can get a nuisance parameter free limit. Under Assumption 2

$$\Psi_T(\theta_0) \to \Psi(\theta_0) \equiv N(0, \Omega_{\theta_0,\theta_0}). \tag{7}$$

Set the efficient weight matrix as $W(\theta_0) = \Omega^{-1}_{\theta_0,\theta_0}$.

Theorem 3. *Under Assumption 2 and* $V_T(\theta_0) \to \Omega_{\theta_0,\theta_0}$, *we derive*

$$S_T(\theta_0; \theta_0) \xrightarrow{d} \chi^2_{GK}.$$

This result is exactly the same as in the weakly identified case of Stock and Wright (2000). The result is robust to identification problems such as weak, semi-weak instruments. This is not surprising since the moment function at true value is zero regardless of problems in the instruments.

Next we try to setup a test statistic that may result in higher power than the Anderson-Rubin like test. This is Kleibergen's (2002) test statistic for weakly identified GMM . We analyze whether this statistic works under sem-weak instruments. We need the notation below before the following assumption. Denote

$$\bar{p}_T(\theta_0) = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial \psi_t(\theta)}{\partial \theta'}|_{\theta=\theta_0},$$

and $\bar{p}_{Ti}(\theta_0)$ represents ith column of $\bar{p}_T(\theta_0)$ matrix, $i = 1, \cdots, p$. And let

$$\tilde{\Psi}_T(\theta_0) = T^{-1}\sum_{t=1}^{T} \psi_t(\theta_0).$$

Assumption 5.*The $GK \times 1$ dimensional derivative of $\psi_t(\theta_0)$ with respect to $\theta_i, i = 1, 2, \cdots, p$:*

$$p_{i,t}(\theta_0) = \frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_0 i}$$

is such that

$$p_{i,t}(\theta_0) - E[p_{i,t}(\theta_0)] = A_i \left(q_{i,t}(\theta_0) - E[q_{i,t}(\theta_0)]\right)$$

with $q_{i,t}(\theta_0) : l_i \times 1$ and A_i a deterministic full rank $GK \times l_i$ dimensional matrix $l_i \leq GK$. The joint limiting behavior of the sums of martingale difference series $\psi_t(\theta_0)$ and $q_{i,t}(\theta_0) - E[q_{i,t}(\theta_0)|I_t]$ satisfy the following Central Limit Theorem:

$$T^{1/2} \begin{bmatrix} \tilde{\Psi}_T(\theta_0) \\ \bar{q}_T(\theta_0) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \Psi(\theta_0) \\ \Psi_q(\theta_0) \end{bmatrix} \equiv N(0, V(\theta_0)). \tag{8}$$

where

$$\bar{q}_T(\theta_0) = T^{-1} \sum_{t=1}^{T} q_t(\theta_0) - E[q_t(\theta_0)].$$

$\bar{q}_T(\theta_0)$ is of dimension $\sum_{i=1}^{p} l_i \times 1$.

$$V(\theta_0) = \begin{bmatrix} \Omega_{\theta_0,\theta_0} & \Omega_{\theta_0,q} \\ \Omega_{q,\theta_0} & \Omega_{q,q} \end{bmatrix}$$

where dimensions of the sub matrices are $\Omega_{\theta_0,\theta_0} : GK \times GK$, $\Omega_{q,\theta_0} : (\sum_{i=1}^{p} l_i) \times GK$, $\Omega_{q,q} : (\sum_{i=1}^{p} l_i) \times (\sum_{i=1}^{p} l_i)$ and $\Omega_{q,\theta_0} = \Omega'_{\theta_0,q}$. Explicitly the sub matrices are

$$\Omega_{q,\theta_0} = \lim_{T\to\infty} E[\frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{T} [q_t(\theta_0) - E[q_t(\theta_0)]][\psi_j(\theta_0)]']$$

$$\Omega_{q,q} = \lim_{T\to\infty} E[\frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{T} [q_t(\theta_0) - E[q_t(\theta_0)]][q_j(\theta_0) - E[q_j(x_j, \theta_0)]]'$$

The instruments span only the part of the information set which is relevant to estimation of θ so $E[q_t(\theta)] = E[q_t(\theta)|I_t]$. Assumption 5 assumes the existence of a simple central limit theorem for martingale difference sequences. This can be satisfied under weaker conditions. The limit of the derivative of $\tilde{\Psi}_T(\theta_0)$ only holds for the part of the derivative with respect to θ_i which lies in the span of A_i. The degeneracy of the limit can happen when the derivative of h(.) in the moment condition in (1) is completely spanned by Z_t. By choosing $A_i = 0$, this can be avoided. In that case $\bar{q}_T(\theta_0)$ does not exist. Another possible degenarate case is when the derivative of several elements of $h(.)$ with respect to θ_i are identical. By specifying appropriate $\bar{q}_T(\theta_0)$ we can avoid this. These are why we need a limit for $\bar{q}_T(\theta_0)$ rather than $\bar{p}_T(\theta_0)$.

Note that Assumption 5 most closely resembles Assumption 2. In Kleibergen test, we show that there is no need for identification Assumption 1, and weight matrix assumption is not used here because that issue is auxilary, the extra assumption is the CLT for partial derivative compared to Assumptions 1-3.

Kleibergen (2001) uses the first order derivative of the objective function for CUE evaluated at true value with efficient weight limit as a basis for his test. First rewrite the CUE objective function as follows

$$S_T(\theta_0;\theta_0) = [T^{1/2}\tilde{\Psi}_T(\theta_0)]' V_T(\theta_0)^{-1}[T^{1/2}\tilde{\Psi}_T(\theta_0)],$$

where $V_T(\theta_0)$ is defined in (5), and the K-test statistic is

$$K(\theta_0) = T[\tilde{\Psi}_T(\theta_0)'\Omega_{\theta_0,\theta_0}^{-1}\bar{D}_T(\theta_0)][\bar{D}_T(\theta_0)'\Omega_{\theta_0,\theta_0}^{-1}\bar{D}_T(\theta_0)]^{-1}[\tilde{\Psi}_T(\theta_0)'\Omega_{\theta_0,\theta_0}^{-1}\bar{D}_T(\theta_0)]',$$

where $\bar{D}_T(\theta_0)$ is a $GK \times p$ matrix and each column $i = 1, 2, \cdots, p$, is:

$$\bar{D}_{Ti}(\theta_0) = \bar{p}_{Ti}(\theta_0) - A_i \Omega_{q_i,\theta_0}\Omega_{\theta_0,\theta_0}^{-1}\tilde{\Psi}_T(\theta_0). \tag{9}$$

Ω_{q_i,θ_0} represents $l_i \times GKmatrix$ This is the $l_i \times GK$ sub matrix of matrix Ω_{q,θ_0} defined in Assumption 5.

Now define

$$J(\theta_0) = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} E(p_t(\theta_0)|I_t),$$

This matrix plays a critical role in K-test. Kleibergen shows that under $J(\theta_0)$ having full rank, under $J(\theta_0) = C/T^{1/2}$ where C is of full rank , finite matrix (weak identification), and when $J(\theta_0) = 0$ (complete unidentification) K-test converges to a χ^2_p distribution under the null when testing the null of $H_0 : \theta = \theta_0$. This is Lemma 2 and Theorem 1 in Kleibergen (2001). First note that regardless of the form of J matrix Lemma 1 of Kleibergen shows that

$$T^{1/2}[\bar{D}_T(\theta_0) - J(\theta_0)] = O_p(1). \tag{10}$$

In this paper we analyze the case of semi-weak identification which corresponds to

$$J(\theta_0) = \frac{C}{T^{\kappa}}, \tag{11}$$

where $0 \leq \kappa < 1/2$ and C is of full rank and finite. The case of $\kappa = 0$ covers standard GMM and $0 < \kappa < 1/2$ covers the nearly-weak case. So semi-weak case $(0 \leq \kappa < 1/2)$ covers both standard and the nearly-weak cases. We need the following lemma to derive the limit.

Lemma 1. *Under Assumption 5 and semi-weak identification (equation (11))* $T^{\kappa}\bar{D}_T(\theta_0)$ *is independent from* $T^{1/2}\tilde{\Psi}_T(\theta_0) = \Psi_T(\theta_0)$.

This shows that Lemma 2 in Kleibergen (2001) works under semi-weak identification.

Now the theorem about K-test in the case of semi-weak instruments.

Theorem 4. *Under Assumption 5, and (11)*

$$K(\theta_0) \xrightarrow{d} \chi^2_p.$$

This clearly shows that K-test has the same limit in the case of semi-weak instruments as in the earlier cases of weak,strong and completely unidentified systems.

4.2 Wald, LM, LR tests

We know that in a mixed case of weak and semi-weak instruments, the tests are not asymptotically pivotal. Here we consider the case of only semi-weak instruments. We specifically show that tests are asymptotically pivotal.

We define Wald, LM and LR tests in two-step GMM case as in Newey and McFadden (1994). First we test the following null:

$$H_0 : a(\theta_0) = 0,$$

where a(.) is $r \times 1$ dimensional.

We need the following variant of Assumption 1 for partial derivative matrix estimation and identification in GMM estimates.

Assumption 1'. *For* $0 \leq \kappa < 1/2$*;*

$$ET^{-1}\sum_{t=1}^{T}\psi_t(\theta) = \frac{m_{2T}(\theta)}{T^\kappa},$$

$m_{2T}(\theta) \to m_2(\theta)$ *uniformly in* θ*,* $m_2(\theta)$ *is continuous. Furthermore* $\psi_t(.)$ *is continuously differentiable in* θ *in N, a neighborhood of* θ_0*, and*

$$\frac{T^\kappa}{T}\sum_{t=1}^{T}\frac{\partial\psi_t(\theta)}{\partial\theta'} \xrightarrow{p} R(\theta),$$

uniformly in N, $R(\theta_0)$ *is of full column rank.* $R(\theta)$ *is* $GK \times p$ *matrix. Note that* $R(\theta)$ *is continuous.* Note that the similarity between Assumption 1' and equation (11).

The following puts some structure on restrictions.

Assumption 6. $a(\theta)$ *is continuously differentiable with* $A(\theta) = \partial a(\theta)/\partial\theta'$*,* $A(\theta_0)$ *is of full rank r, where* $A(\theta_0) = A$ *is* $r \times p$*.*

First we give theory for the Wald test. The Wald test is defined as follows:

$$Wald = Ta(\hat{\theta}_2)' \left[A(\hat{\theta}_2)\left\{\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\psi_t(\hat{\theta}_2)}{\partial\theta'}\right)' V_T(\hat{\theta}_2)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\psi_t(\hat{\theta}_2)'}{\partial\theta}\right)\right\}^{-1} A(\hat{\theta}_2)'\right]^{-1} a(\hat{\theta}_2).$$

$V_T(\hat{\theta}_2)$ represents the consistent estimate of the variance-covariance matrix in efficient two-step GMM, and can be seen in equation (5). $\hat{\theta}_2$ represents the efficient two-step GMM estimator.

Now we also show the result for LM and LR tests. For the LM test define a constrained two-step GMM estimate in the following way:

$$\tilde{\theta}_2 = \arg\min S_T(\theta;\hat{\theta}_1) \quad \text{subject to} \quad a(\theta) = 0.$$

It should be clear that $\hat{\theta}_1$ represent first step estimates in two-step GMM,and we form the weight matrix using these estimates. Define the lagrangian multiplier corresponding to the constraint as γ: $r \times 1$vector. Let $V_T(\tilde{\theta}_2)^{-1}$ be the consistent estimate of the efficient limit weight matrix. Let $\tilde{\gamma}$ be the estimate of the multiplier. Because we scale the objective function differently than the standard GMM case of Newey and McFadden, our LM test takes the following form:

$$LM = \frac{1}{T}\tilde{\gamma}' A(\tilde{\theta}_2) \left\{ \left(\frac{1}{T} \frac{\sum_{t=1}^{T} \partial \psi_t(\tilde{\theta}_2)'}{\partial \theta} \right)' V_T(\tilde{\theta}_2)^{-1} \left(\frac{1}{T} \frac{\sum_{t=1}^{T} \partial \psi_t(\tilde{\theta}_2)'}{\partial \theta} \right) \right\}^{-1} A(\tilde{\theta}_2)'\tilde{\gamma}.$$

For the LR test, define it as follows:

$$LR = S_T(\hat{\theta}_2; \hat{\theta}_2) - S_T(\tilde{\theta}_2; \tilde{\theta}_2).$$

Theorem 5. *Under Assumption 1',2,4,6, and under the null hypothesis*
(a) $Wald \xrightarrow{d} \chi^2_r$.
(b) $LM \xrightarrow{d} \chi^2_r$.
(c)$LR \xrightarrow{d} \chi^2_r$.

Remark. We show that even though the correlation of instruments with the orthogonality restrictions decline to zero as sample size increases, Wald, LM,LR tests have the χ^2 distribution. This is surprising result since the limit is the same as in standard GMM. We also show that even though the estimator converges slower compared to strong instruments (standard GMM), the Wald, LM, LR tests are not affected by this. These results are new and show that large sample results regarding inference has not been affected by low correlation between the instruments and the orthogonality restrictions as long as decay in correlation is moderate. As it is said earlier only when the correlation declines at $T^{1/2}$ rate, these classical tests are not asymptotically pivotal. This also shows that the limit for the standard case ($\kappa = 0$) and the nearly-weak case ($0 < \kappa < 1/2$) is the same.

If we are interested in testing the general hypothesis (such as nonlinear or various linear ones) about parameters in neraly-weak case, these tests are usable, since they are asymptotically pivotal. This shows the importance of analyzing the nearly-weak case since with slowly declining correlation (i.e. $0 < \kappa < 1/2$) between the instruments and orthogonality restrictions we can test various hypotheses. This is unlike the weak case where we can only test $H_0 : \theta = \theta_0$. At the end of section 4 we also propose an ad-hoc indicator of differentiating between weak and semi-weak case.

Wald, LR, LM tests have the same limits under CUE. This has been discussed in the appendix, after the proof of Theorem 5. Note that K test in section 3.1 does not work when we test general restrictions such as $H_0 : a(\theta_0) = 0$ since then you have to use constrained estimator $\tilde{\theta}$ in the test statistic. In that case

$$\sum_{t=1}^{T} \frac{\psi_t(\tilde{\theta})}{T^{1/2}} \neq \sum_{t=1}^{T} \frac{\psi_t(\theta_0)}{T^{1/2}} + o_p(1).$$

which does not result in the χ^2 limit.

4.3 J test

In this subsection we derive the limit for overidentifying restrictions. In the weakly identified case of Stock and Wright (2000) we see that the limit is not asymptotically pivotal, and also we can derive the same limit in Stock and Wright (2000) in our mixed weak-semiweak case. But if the reality is that all the instruments are semi-weak, then the limit is the standard χ^2_{GK-p} limit. The following theorem shows that the limit for the two-step case, for CUE the differences are minor in derivation but the limit is the same, so we skip CUE case. We show below that the limit for the semi-weak case is the same for both standard GMM and the nearly-weak GMM. This can be easliy seen since they are subcases of semi-weak GMM and the limit in the semi-weak case is χ^2_{GK-p}. This result is new and shows that the limit for J test is entirely different in the weak case of Stock and Wright (2000) and the semi-weak case considered here. J-test for testing the validity of moment restrictions are

$$J_{2s}(\hat{\theta}_2) = \left[\frac{1}{T^{1/2}}\sum_{t=1}^{T}\psi_t(\hat{\theta}_2)\right]' V_T(\hat{\theta}_2)^{-1}\left[\frac{1}{T^{1/2}}\sum_{t=1}^{T}\psi_t(\hat{\theta}_2)\right].$$

Theorem 6. *Under Assumptions 1',2,4*

$$J_{2s}(\hat{\theta}_2) \xrightarrow{d} \chi^2_{GK-p}$$

So again this show that declining correlation between instruments and orthogonality restrictions does not affect the limit of J-test, the key issue is the correlation should not decay at rate $T^{1/2}$. This is true for both two-step and CUE based J-tests. Theorem 6 is valid also for CUE based J-test.

The following J tests are analyzed by Kleibergen (2003) in the case of weak instruments ($\kappa = 1/2$).

$$J_{2s}(\theta_0) = S_T(\theta_0 : \theta_0) - W_{2s}(\theta_0),$$

$$J_{CUE}(\theta_0) = S_T(\theta_0;\theta_0) - W_{CUE}(\theta_0),$$

$$J_{LM}(\theta_0) = S_T(\theta_0;\theta_0) - LM(\theta_0),$$

$$J_K(\theta_0) = S_T(\theta_0;\theta_0) - K(\theta_0),$$

where $S_T(.)$ represent the S test analyzed in section 3.1.

We consider the limits of the above J tests in the case of semi-weak instruments ($0 \leq \kappa < 1/2$). Under the joint hypotheses of $H_0 : \theta = \theta_0$, and the hypotheses that $H_E : E\psi_t(\theta) = 0$, we show that these four test statistics have χ^2_{GK-p} distribution.

Lemma 2. *Under Assumptions 1',2,4 and Theorem 5*

$$J_{2s}(\theta_0) \xrightarrow{d} \chi^2_{GK-p},$$

$$
\begin{aligned}
J_{CUE}(\theta_0) &\xrightarrow{d} \chi^2_{GK-p}, \\
J_{LM}(\theta_0) &\xrightarrow{d} \chi^2_{GK-p}, \\
J_{K}(\theta_0) &\xrightarrow{d} \chi^2_{GK-p}.
\end{aligned}
$$

5 Power

In this section we analyze the power of the Wald, LR, LM tests in the semi-weak case analytically. We only analyze Wald two-step case, the others have the same results and thus omitted. First we consider the local alternative of the form used in standard GMM (Newey and McFadden, 1994)

$$H_1 : a(\theta_0) = l/T^{1/2}.$$

The following expansion will be used in the analysis of power:

$$
\begin{aligned}
T^{1/2-\kappa}a(\hat{\theta}) &= T^{1/2-\kappa}a(\theta_0) + A(\bar{\theta})T^{1/2-\kappa}(\hat{\theta}-\theta_0) \\
&= \frac{T^{1/2}}{T^{\kappa}}\frac{l}{T^{1/2}} + A(\bar{\theta})T^{1/2-\kappa}(\hat{\theta}-\theta_0) \\
&= A(\bar{\theta})T^{1/2-\kappa}(\hat{\theta}-\theta_0) + o_p(1),
\end{aligned}
$$

This expression is the same as (28) in the proof of Theorem 5a. Using this result in (25) of the proof of Theorem 5a we again converge to χ^2_r limit like the null. So there is no power at this locality unlike standard GMM case when $0 < \kappa < 1/2$. If the alternative is fixed

$$H_1 : a(\theta_0) = l \neq 0.$$

Then

$$T^{1/2-\kappa}a(\hat{\theta}) = T^{1/2-\kappa}l + A(\bar{\theta})T^{1/2-\kappa}(\hat{\theta}-\theta_0).$$

So

$$T^{1/2-\kappa}a(\hat{\theta}) \to \infty,$$

but at a slower rate when $0 < \kappa < 1/2$. Again by the proof of Theorem 5a below the test statistics are consistent but in small samples they lose power when the correlation between instruments and the othogonality restrictions are slowly decaying. This shows that compared to standard GMM $\kappa = 0$, in the near-weak case of $0 < \kappa < 1/2$, the power is smaller and declines with the weakness of the instrument. Now we can see analytically why the instrument quality is important in small samples.

6 Higher Order Expansions

As in Kleibergen (2003) we analyze higher order properties of various test statistics, but we consider the semi-weak case of $0 \leq \kappa < 1/2$ instead of $\kappa = 1/2$ which is analyzed in Kleibergen (2003). Note again that standard GMM case corresponds to $\kappa = 0$ and the nearly-weak GMM corresponds to $0 < \kappa < 1/2$. We have seen so far that these tests have the same limit in both standard and nearly-weak GMM. The interesting question is: whether their higher order expansions are the same? We see below that these are different in standard GMM and nearly-weak GMM. We also cannot correct these by using Edgeworth approximations in the nearly-weak case. So the finite sample behavior of test statistics is different at $\kappa = 0$ from the case $0 < \kappa < 1/2$. This is one of the main findings of the paper. We can clearly see the problem in asymptotic approximations to the finite sample behavior of these test when there are nearly-weak instruments. The quality of the instruments is very important in understanding the finite sample behavior of these tests. As we move away from $\kappa = 0$ towards $\kappa = 1/2 - \epsilon$, where ϵ is a small positive number, capturing the finite sample properties by asymptotics become a major challenge.

The following Assumption is used in Kleibergen (2003) to get the higher order terms related to variance-covariance matrix. Since $\Omega_{\theta_0,\theta_0}$ is unknown we use an estimator $V(\theta_0)$.

Assumption 7. *The convergence of the covariance matrix estimator is such that*

$$T^{\mu/2} vec(V(\theta_0) - \Omega_{\theta_0,\theta_0}) = \Psi_u.$$

The details of the limit is given in Kleibergen (2003). We now write the test statistics when we test $H_0 : \theta = \theta_0$. By choice of this simple null, we simplify a lot of the notation in the proofs. These test statistics can be derived from the test statistics that are given in section 3. Let $\hat{\theta}_2, \hat{\theta}_c$ denote the efficient two-step and CUE estimators respectively.

$$\begin{aligned}
W_{2s}(\theta_0) &= T(\hat{\theta}_2 - \theta_0)' \left[\left(\frac{1}{T} \sum_{t=1}^{T} \frac{\partial \psi_t(\hat{\theta}_2)}{\partial \theta'} \right)' V_T(\hat{\theta}_2)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} \frac{\partial \psi_t(\hat{\theta}_2)}{\partial \theta'} \right) \right] (\hat{\theta}_2 - \theta_0) \\
&= \Psi_T(\theta_0)' V_T(\hat{\theta}_2)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} \frac{\partial \psi_t(\hat{\theta}_2)}{\partial \theta'} \right) \left[\left(\frac{1}{T} \sum_{t=1}^{T} \frac{\partial \psi_t(\hat{\theta}_2)}{\partial \theta'} \right)' V_T(\hat{\theta}_2)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} \frac{\partial \psi_t(\hat{\theta}_2)}{\partial \theta'} \right) \right]^{-1} \\
&\times \left(\frac{1}{T} \sum_{t=1}^{T} \frac{\partial \psi_t(\hat{\theta}_2)}{\partial \theta'} \right) V_T(\hat{\theta}_2)^{-1} \Psi_T(\theta_0) \\
&+ o_p(1),
\end{aligned}$$

where the second equality uses the asymptotic equivalent expression for estimators distribution which can be found from the proof of Theorem 1 with only semi-weak instruments. $\Psi_T(\theta_0) = T^{-1/2} \sum_{t=1}^{T} \psi_t(\theta_0)$ and general definition is given immediately after Assumption 3. Wald test for

CUE is similar to two-step case except from the Jacobian term:

$$
\begin{aligned}
W_{CUE}(\theta_0) &= T(\hat{\theta}_c-\theta_0)'\left[\frac{1}{T}D_T(\hat{\theta}_c)'V_T(\hat{\theta}_c)^{-1}\frac{1}{T}D_T(\hat{\theta}_c)\right](\hat{\theta}_c-\theta_0) \\
&= \Psi_T(\theta_0)'V_T(\hat{\theta}_c)^{-1}\frac{1}{T}D_T(\hat{\theta}_c)\left[\frac{1}{T}D_T(\hat{\theta}_c)'V_T(\hat{\theta}_c)^{-1}\frac{1}{T}D_T(\hat{\theta}_c)\right]^{-1} \\
&\times \frac{1}{T}D_T(\hat{\theta}_c)'V_T(\hat{\theta}_c)^{-1}\Psi_T(\theta_0) \\
&+ o_p(1),
\end{aligned}
$$

$D_T(\hat{\theta}_c)$ is $D_T(\theta_0)$ term described in section 3.1 but evaluated at $\hat{\theta}_c$ and using the estimators for the covariance matrices.

LM test can be obtained from Newey and McFadden (1994) or can be obtained using the LM test given here and using (30)

$$
\begin{aligned}
LM(\theta_0) &= \Psi_T(\theta_0)'V_T(\theta_0)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\psi_t(\theta_0)}{\partial\theta'}\right) \\
&\times \left[\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\psi_t(\theta_0)}{\partial\theta'}\right)'V_T(\theta_0)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\psi_t(\theta_0)}{\partial\theta'}\right)\right]^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\psi_t(\theta_0)}{\partial\theta'}\right)'V_T(\theta_0)^{-1}\Psi_T(\theta_0).
\end{aligned}
$$

K statistic can be written using $T^{1/2}\tilde{\Psi}_T(\theta_0)=\Psi_T(\theta_0)=T^{-1}\psi_t(\theta_0)$,

$$
K(\theta_0)=\Psi_T(\theta_0)'V_T(\theta_0)^{-1}\hat{D}_T(\theta_0)[\hat{D}_T(\theta_0)'V_T(\theta_0)^{-1}\hat{D}_T(\theta_0)]^{-1}\hat{D}_T(\theta_0)'V_T(\theta_0)^{-1}\Psi_T(\theta_0)',
$$

where $\hat{D}_T(\theta_0)$ denotes $D_T(\theta_0)$ term in section 3.1 but with estimated covariance matrices. The following theorem is one of the main results of this paper. Theorem 7 and explanations below show that there are important differences between the standard GMM ($\kappa=0$) and nearly weak case ($0<\kappa<1/2$) in terms of finite sample behavior. This result also extends the Kleibergen (2003) from the weak case ($\kappa=1/2$) to semi-weak case ($0\leq\kappa<1/2$).

Theorem 7. *Under Assumptions 5 and 7 in a setup of semi-weak instruments (i.e. equation (11), $0\leq\kappa<1/2$)*

(a).

$$
\begin{aligned}
W_{2s}(\theta_0) &= \eta_0+\frac{\eta_1}{T^{1/2-\kappa}}+\frac{\eta_2}{T^{\delta/2}}+\frac{\eta_3}{T^{(1/2-\kappa)+\delta/2}}+\frac{\eta_4}{T^{1-2\kappa}} \\
&+ \frac{\eta_5}{T^{\delta}}+\frac{\eta_6}{T^{1-2\kappa+\delta/2}}+\frac{\eta_7}{T^{1/2-\kappa+\delta}}+\frac{\eta_8}{T^{3/2-3\kappa}}+o_p(T^{3\kappa-3/2}),
\end{aligned}
$$

where $\delta=\min(1-2\kappa,\mu)$.

(b).

$$
\begin{aligned}
W_{CUE}(\theta_0) &= \eta_0+\frac{\eta_2}{T^{\delta/2}}+\frac{\eta_3}{T^{(1/2-\kappa)+\delta/2}} \\
&+ \frac{\eta_5}{T^{\delta}}+\frac{\eta_7}{T^{1/2-\kappa+\delta}}+o_p(T^{3\kappa-3/2}),
\end{aligned}
$$

(c).

$$\begin{aligned} LM(\theta_0) &= \eta_0 + \frac{\eta_1}{T^{1/2-\kappa}} + \frac{\eta_2}{T^{\mu/2}} + \frac{\eta_3}{T^{(1/2-\kappa)+\mu/2}} + \frac{\eta_4}{T^{1-2\kappa}} \\ &+ \frac{\eta_5}{T^{\mu}} + \frac{\eta_6}{T^{1-2\kappa+\mu/2}} + \frac{\eta_7}{T^{1/2-\kappa+\mu}} + \frac{\eta_8}{T^{3/2-3\kappa}} + o_p(T^{3\kappa-3/2}), \end{aligned}$$

(d).

$$\begin{aligned} K(\theta_0) &= \eta_0 + \frac{\eta_2}{T^{\mu/2}} + \frac{\eta_3}{T^{(1/2-\kappa)+\mu/2}} \\ &+ \frac{\eta_5}{T^{\mu}} + \frac{\eta_7}{T^{1/2-\kappa+\mu}} + + o_p(T^{3\kappa-3/2}), \end{aligned}$$

All η_j terms in (a)(b)(c)(d) are $O_p(1)$, specifically

$$\eta_0 \xrightarrow{d} \chi^2_p.$$

Remarks. These expansions in Kleibergen (2003) is only valid for $\kappa = 1/2 and 0$. This theorem extends Kleibergen's (2003) paper to $0 \leq \kappa < 1/2$. Since we have $0 \leq \kappa < 1/2$, the higher order terms depend on the quality of instruments. Unlike standard asymptotics when $\kappa = 0$, the higher order terms in nearly-weak case converge to zero slower . This is the main reason why we see in simulation studies asymptotic approximation was not doing a good job in capturing finite sample behavior of test statistics in GMM. With low quality of instruments (i.e. when κ is near 1/2) the convergence to zero is very slow and large sample approximation to the finite sample is not very good. Apparently W_{CUE} and $K(\theta_0)$ tests do a better job in approximating the finite sample distributions of tests, since they do not have terms $\eta_1, \eta_4, \eta_6, \eta_8$. However, since K tests are only relevant for simple tests of the null, in semi-weak case for testing general null W_{CUE} may be preferable.

The higher order terms in Theorem 7 affect the approximation of finite sample distributions by large sample theory. In the case of standard GMM Edgeworth approximations remove some of these higher order terms, see equation (33) of Kleibergen (2003) for this point. However, the second order Edgeworth approximations donot perform well when $0 < \kappa < 1/2$. Since following Theorem 7 here equation (39) of Kleibergen (2003) and the argument immediately after Corollary 3 of Kleibergen (2003)applies ; Edgeworth approximations depend on the value of κ which is unknown. Specifically, $E(\eta_4|\eta_0)$ is proportional to $(D_0'\Omega^{-1}_{\theta_0,\theta_0} D_0)^{-1}$ which can be estimated by

$$\left(\frac{1}{T^{2-2\kappa}} D_T(\theta_0)' \Omega^{-1}_{\theta_0,\theta_0} D_T(\theta_0) \right)^{-1},$$

but even writing that as;

$$\left[\frac{1}{T} \left(T^{\kappa} D_T(\theta_0)\right)' \Omega^{-1}_{\theta_0,\theta_0} \left(T^{\kappa} D_T(\theta_0)\right) \right]^{-1},$$

does not help since without knowing κ , the second order approximation is not useful. We cannot estimate $T^{\kappa} D_T(\theta_0)$ without knowing κ. Only in standard case $\kappa = 0$, as shown in Kleibergen (2003) this is possible and hence this approximation is useful in correcting the test statistics asymptotic approximation. Note that Kleibergen (2003) uses term ν which is the concentration parameter, in the proof of Theorem 7 we show that $\nu = 1 - 2\kappa$ and hence use κ.

Corollary 4. *Under Assumptions 5,7 and Theorem 5 in a setup of semi-weak instruments $(0 \leq \kappa < 1/2)$ (a).*

$$
\begin{aligned}
J_{2s}(\theta_0) &= (\omega_0 - \eta_0) + \frac{\omega_\mu}{T^{\mu/2}} \\
&- (\frac{\eta_1}{T^{1/2-\kappa}} + \frac{\eta_2}{T^{\delta/2}} + \frac{\eta_3}{T^{(1/2-\kappa)+\delta/2}}) \\
&- (\frac{\eta_4}{T^{1-2\kappa}}\frac{\eta_5}{T^{\delta}} + \frac{\eta_6}{T^{1-2\kappa+\delta/2}}) \\
&- (\frac{\eta_7}{T^{1/2-\kappa+\delta}} + \frac{\eta_8}{T^{3/2-3\kappa}} + o_p(T^{3\kappa-3/2})),
\end{aligned}
$$

(b).

$$
\begin{aligned}
J_{CUE}(\theta_0) &= (\omega_0 - \eta_0) + \frac{\omega_\mu}{T^{\mu/2}} \\
&- (\frac{\eta_2}{T^{\delta/2}} + \frac{\eta_3}{T^{(1/2-\kappa)+\delta/2}}) \\
&- (\frac{\eta_5}{T^{\delta}} + \frac{\eta_7}{T^{1/2-\kappa+\delta}} + o_p(T^{3\kappa-3/2})),
\end{aligned}
$$

(c).

$$
\begin{aligned}
J_{LM}(\theta_0) &= (\omega_0 - \eta_0) + \frac{\omega_\mu}{T^{\mu/2}} \\
&- (\frac{\eta_1}{T^{1/2-\kappa}} + \frac{\eta_2}{T^{\mu/2}} + \frac{\eta_3}{T^{(1/2-\kappa)+\mu/2}} + \frac{\eta_4}{T^{1-2\kappa}}) \\
&- (\frac{\eta_5}{T^{\mu}} + \frac{\eta_6}{T^{1-2\kappa+\mu/2}} + \frac{\eta_7}{T^{1/2-\kappa+\mu}} + \frac{\eta_8}{T^{3/2-3\kappa}} + o_p(T^{3\kappa-3/2})),
\end{aligned}
$$

(d).

$$
\begin{aligned}
J_{K}(\theta_0) &= (\omega_0 - \eta_0) + \frac{\omega_\mu}{T^{\mu/2}} \\
&- (\frac{\eta_2}{T^{\mu/2}} + \frac{\eta_3}{T^{(1/2-\kappa)+\mu/2}}) \\
&- (\frac{\eta_5}{T^{\mu}} + \frac{\eta_7}{T^{1/2-\kappa+\mu}} + o_p(T^{3\kappa-3/2})),
\end{aligned}
$$

where $\omega_0 - \eta_0 \xrightarrow{d} \chi^2_{GK-p}$ and all the other ω and η terms are $O_p(1)$. The details are in the proof of Corollary 4.

Let us consider the standard J test in two-step GMM and with semi-weak instruments.

$$
J_{2s}(\hat{\theta}_2) = [T^{-1/2} \sum_{t=1}^{T} \psi_t(\hat{\theta}_2)]' V_T(\hat{\theta}_2)^{-1} [T^{-1/2} \sum_{t=1}^{T} \psi_t(\hat{\theta}_2)].
$$

Then by (44)

$$
\begin{aligned}
J_{2s}(\hat{\theta}_2) &= [T^{-1/2}\sum_{t=1}^{T}\psi_t(\theta_0)]'V_T(\hat{\theta}_2)^{-1}[T^{-1/2}\sum_{t=1}^{T}\psi_t(\theta_0)] \\
&- [(T^{-1/2}\sum_{t=1}^{T}\psi_t(\theta_0))'V_T(\hat{\theta}_2)^{-1}G_T(\hat{\theta}_2)][G_T(\hat{\theta}_2)'V_T(\hat{\theta}_2)^{-1}G_T(\hat{\theta}_2)]^{-1}[G_T(\hat{\theta}_2)'V_T(\hat{\theta}_2)^{-1}(T^{-1/2}\sum_{t=1}^{T}\psi_t(\theta_0))] \\
&= S_T(\theta_0;\theta_0) - W_{2s}(\theta_0) \\
&+ [T^{-1/2}\sum_{t=1}^{T}\psi_t(\theta_0)]'[V_T(\hat{\theta})^{-1} - V_T(\theta_0)^{-1}][T^{-1/2}\sum_{t=1}^{T}\psi_t(\theta_0)], \qquad (12)
\end{aligned}
$$

where the last equality used in $W_{2s}(\theta_0)$ definition and add and subtract $S_T(\theta_0;\theta_0)$ from the first equation.

In (12) we know the higher order expansion of $S_T(\theta_0;\theta_0) - W_{2s}(\theta_0)$ by Corollary 4. Then the remainder term is

$$
[T^{-1/2}\sum_{t=1}^{T}\psi_t(\theta_0)]'[V_T(\hat{\theta})^{-1} - V_T(\theta_0)^{-1}][T^{-1/2}\sum_{t=1}^{T}\psi_t(\theta_0)] = T^{\kappa-1/2}a_1,
$$

where

$$
a_1 = [T^{-1/2}\sum_{t=1}^{T}\psi_t(\theta_0)]'T^{1/2-\kappa}[V_T(\hat{\theta})^{-1} - V_T(\theta_0)^{-1}][T^{-1/2}\sum_{t=1}^{T}\psi_t(\theta_0)],
$$

$a_1 = O_p(1)$ by rate of convergence of $\hat{\theta}$ and Assumption 5. So

$$
J_{2s}(\hat{\theta}) = J_{2s}(\theta_0) + T^{\kappa-1/2}a_1.
$$

In the same manner we can obtain for CUE

$$
J_{CUE}(\hat{\theta}) = J_{CUE}(\theta_0) + O_p(T^{\kappa-1/2}).
$$

So J tests that are used in standard GMM literature has additional term compared with $J_{2s}(\theta_0), J_{CUE}(\theta_0)$ in semi-weak case. This is one of the main reasons why J tests have finite sample problems.

Note that as long as there are only semi-weak instruments ($0 \leq \kappa < 1/2$), by Theorem 5 and Theorem 2.1 of Horowitz (2001) or Beran and Ducharme (1991) bootstrap versions of Wald, LM tests are consistent. But in reality we do not know what κ may be, so in practice it is not a good idea to use bootstrap in uncertain situations. Only in the case of K tests we can use bootstrap regardless of the nature of instruments, see Kleibergen (2003).

However, in order to understand whether the first order asymptotic theory is correct in the case of semi-weak instruments case; we can use bootstrap critical values of W_{CUE} test and compare it with asymptotic χ^2 values. Even though this is ad-hoc we can have at least some basic understanding whether we face semi-weak or weak case. Since by Theorem 7b, W_{CUE} contains the least possible higher order terms and these terms go to zero faster compared with the other Wald and

LM test, we can make use of this test. To have an idea about whether we have semi-weak or weak case is important, since in the semi-weak case we can test nonlinear restrictions and use estimators unlike the weak case. In large samples we can have good inference by using the $Wald_{CUE}$ test.

7 Conclusion

We analyzed GMM with a mixed system with weak and semi-weak instruments. Semi-weak instruments case include standard GMM and GMM with nearly weak instruments. These two groups are put together and form semi-weak case since their limit behavior is the same for estimators and test statistics. In order to understand the understand the differences between the weakly identified GMM of Stock and Wright (2000) and the semi-weak case proposed here, we considered a system with weak and semi-weak instruments. Estimators of semi-weakly identified parameters are consistent unlike the inconsistent estimates for the weakly identified ones.

We also analyzed the limit behavior of various test statistics with semi-weak instruments in GMM and we find that their limits are χ^2, unlike the weakly identified GMM case. Furthermore, by deriving higher order expansions we also show that even nearly-weak and standard GMM is different, but difference is in the finite sample behavior. Finite sample behavior of test statistics are adversely affected in the case of nearly-weak instruments case. For the future work, an interesting idea is to look at the case of many instruments in the semi-weak GMM framework and extending linear IV with many instruments case of Chao and Swanson (2003).

APPENDIX

Proof of Theorem 1. First we provide the consistency result. Let

$$m_T(\theta) = ET^{-1/2}\sum_{t=1}^{T}\psi_t(\theta). \tag{13}$$

Then

$$S_T(\theta;\bar{\theta}_T(\theta)) = [\Psi_T(\theta) + m_T(\theta)]' W_T(\bar{\theta}_T(\theta))[\Psi_T(\theta) + m_T(\theta)].$$

Then by (4) and $0 \leq \kappa < 1/2$,

$$T^{\kappa - 1/2}\Psi_T(\theta) \xrightarrow{p} 0. \tag{14}$$

Then via Assumption 1 and (13)

$$\begin{aligned} T^{\kappa-1/2} m_T(\theta) &= T^{\kappa} ET^{-1}\sum_{t=1}^{T}\psi_t(\theta) \\ &= T^{\kappa-1/2} m_{1T}(\theta) + m_{2T}(\beta) \\ &\xrightarrow{p} m_2(\beta). \end{aligned} \tag{15}$$

So combine (14)(15) in the $S_T(\theta;\bar{\theta}_T(\theta))$ and by Assumption 3 , uniformly in θ

$$T^{2\kappa-1} S_T(\theta;\bar{\theta}_T(\theta)) \xrightarrow{p} m_2(\beta)' W(\theta) m_2(\beta).$$

Then since $m_2(\beta) = 0$ iff $\beta = \beta_0$ and W is positive definite by Corollary 3.2.3 of van der Vaart and Wellner (1996) we have the consistency. In the rate of convergence proof we follow closely the proof in Stock and Wright (2000). Since $\hat{\theta}_T$ minimizes S_T and by Assumption 3, Assumption 1

$$\begin{aligned} S_T(\hat{\theta};\bar{\theta}_T(\theta))) - S_T(\theta_0;\bar{\theta}_T(\theta_0)) &= [\Psi_T(\hat{\theta}) + m_{1T}(\hat{\theta}) + T^{1/2-\kappa} m_{2T}(\hat{\beta})]' W_T(\bar{\theta}_T(\hat{\theta})) \\ &\times \ [\Psi_T(\hat{\theta}) + m_{1T}(\hat{\theta}) + T^{1/2-\kappa} m_{2T}(\hat{\beta})] \\ &- \ \Psi_T(\theta_0)' W_T(\bar{\theta}_T(\theta_0)\Psi_T(\theta_0) \\ &\leq \ 0. \end{aligned}$$

Then after some simple algebra as following equations (A.1)-(A.2) of Stock and Wright (2000) , and noting that the only difference between equations (A.1)-(A.2) in Stock and Wright (2000) and our following equation is, $0 \leq \kappa < 1/2$ in our case and

$$\|T^{1/2-\kappa} m_{2T}(\hat{\beta})\|^2 - 2d_{2T}\|T^{1/2-\kappa} m_{2T}(\hat{\beta})\| + d_{3T} \leq 0, \tag{16}$$

where

$$d_{2T} = \|W_T(\hat{\theta})[\Psi_T(\hat{\theta}) + m_{1T}(\hat{\theta})]\| / meval W_T(\bar{\theta}_T(\hat{\theta}).$$

$$d_{3T} = d_{1T}(\hat{\theta}) / meval W_T(\bar{\theta}_T(\hat{\theta}).$$

$$
\begin{aligned}
d_{1T}(\hat{\theta}) &= [\Psi_T(\hat{\theta}) + m_{1T}(\hat{\theta})]' W_T(\bar{\theta}(\hat{\theta}))[\Psi_T(\hat{\theta}) + m_{1T}(\hat{\theta})] \\
&- \Psi_T(\theta_0)' W_T(\bar{\theta}(\hat{\theta}))\Psi_T(\theta_0)
\end{aligned}
$$

where $meval(.)$ denotes the minimum eigenvalue of the matrix in the parantheses..

Now take the roots of (16) and write

$$
T^{1/2-\kappa} m_{2T}(\hat{\beta}) = R_T(\tilde{\beta}) T^{1/2-\kappa}(\hat{\beta} - \beta_0),
$$

where we used $m_{2T}(\beta_0) = 0$, and Assumption 1 and $\tilde{\beta} \in (\beta_0, \hat{\beta})$. Then for (16)to hold

$$
\| R_T(\tilde{\beta}) T^{1/2-\kappa}(\hat{\beta} - \beta_0) \| \leq d_{2T} + (d_{2T}^2 - d_{3T})^{1/2}. \tag{17}
$$

But via Assumption 1 and consistency $R_T(\tilde{\beta}) \xrightarrow{p} R(\beta_0)$. Note that $d_{2T} = O_p(1)$ and $d_{3T} = O_p(1)$ by (4) and Assumption 3 exactly as in equations (A.4) and (A.5) of Stock and Wright (2000). The desired result follows since $d_{2T} = O_p(1)$, $d_{3T} = O_p(1)$.■

Proof of Theorem 2. To derive the limit we need to find the limit for the following uniformly over the local parameter space $(\alpha', b')' \in (A, \bar{B})'$ where A and $\bar{B}$ is compact $S_T(\alpha, \beta_0 + \frac{b}{T^{1/2-\kappa}})$ where $0 \leq \kappa < 1/2$. To evaluate that use (3)

$$
\begin{aligned}
T^{-1/2} \sum_{t=1}^{T} \psi_t(\alpha, \beta_0 + \frac{b}{T^{1/2-\kappa}}) &= \Psi_T(\alpha, \beta_0 + \frac{b}{T^{1/2}}) + E T^{-1/2} \sum_{t=1}^{T} \psi_t(\alpha, \beta_0 + \frac{b}{T^{1/2}}) \\
&= \Psi_T(\alpha, \beta_0 + \frac{b}{T^{1/2-\kappa}}) + m_{1T}(\alpha, \beta_0 + \frac{b}{T^{1/2-\kappa}}) + T^{1/2-\kappa} m_{2T}(\beta_0 + \frac{b}{T^{1/2-\kappa}}).
\end{aligned}
$$

First by using the triangle array emprical process limit in Andrews (1994) via Assumption 2 (or equation (4)) and the consistency result (Theorem 1) uniformly over $A \times \bar{B}$

$$
\Psi_T(\alpha, \beta_0 + \frac{b}{T^{1/2-\kappa}}) \Longrightarrow \Psi(\alpha, \beta_0). \tag{18}
$$

Note that we can obtain (18) in an alternative way this is discussed at the end of the proof.

Then by Assumption 1 and Theorem 1

$$
m_{1T}(\alpha, \beta_0 + \frac{b}{T^{1/2-\kappa}}) \to m_1(\alpha, \beta_0).
$$

By using Assumption 1 and $m_{2T}(\beta_0) = 0$,

$$
\begin{aligned}
T^{1/2-\kappa} m_{2T}(\beta_0 + \frac{b}{T^{1/2-\kappa}}) &= T^{1/2-\kappa} m_{2T}(\beta_0) + T^{1/2-\kappa} R_T(\alpha, \beta_0 + \frac{\bar{b}}{T^{1/2-\kappa}}) \frac{b}{T^{1/2-\kappa}} \\
&\to R(\beta_0) b,
\end{aligned}
$$

where $\bar{b} \in (0, b)$. Using these results via Assumption 3

$$
\begin{aligned}
S_T(\alpha, \beta_0 + \frac{b}{T^{1/2-\kappa}}; \bar{\theta}_T(\alpha, \beta_0)) &\Longrightarrow [\Psi(\alpha, \beta_0) + m_1(\alpha, \beta_0) + R(\beta_0) b]' \\
&\times W(\bar{\theta}(\alpha, \beta_0)[\Psi(\alpha, \beta_0) + m_1(\alpha, \beta_0) + R(\beta_0) b] \\
&\equiv \bar{S}(\alpha, b; \bar{\theta}(\alpha, \beta_0))
\end{aligned}
$$

So this is the same limit of the objective found in Theorem 1i in Stock and Wright (2000). But in our case we have semi-weak instruments and weak instruments in the model (i.e. $0 \leq \kappa < 1/2$) compared to Stock and Wright's (2000) weak and strong instruments case ($\kappa = 0$). So the limit in that theorem generalizes to semi-weak instruments as well.

If $\bar{S}(\alpha, b; \bar{\theta}(\alpha, \beta_0))$ has a unique minimum following then following the proof of Theorem 1ii in Stock and Wright (2000) for the rate of convergence of $T^{1/2-\kappa}$ gives the desired result. This is a simple application of argmin continuous mapping theorem and the envelope theorem.

Now we discuss how to derive (18) under an alternative way. Remembering the original parameter space is $\Theta = A \times B$ where A and B is compact we can get the same result for (18) under when the original parameter space B is open, but with slightly stronger assumptions. So we add $\psi_t(.)$ is continuously differentiable and

$$\frac{T^\kappa}{T} \sum_{t=1}^{T} \partial \psi_t(\alpha, \beta_0)/\partial \beta' - E \partial \psi_t(\alpha, \beta_0)/\partial \beta' \xrightarrow{p} 0, \tag{19}$$

uniformly over α. Setting $r_T = T^{1/2-\kappa}$ we want to show, uniformly over $A \times barB$ (note that the local parameter space is still compact, so b is still in a compact)

$$\Psi_T(\alpha, \beta_0 + b/r_T) - \Psi_T(\alpha, \beta_0) \xrightarrow{p} 0.$$

Rewrite this as

$$\frac{1}{r_T} \left\{ r_T [\Psi_T(\alpha, \beta_0 + b/r_T) - \Psi_T(\alpha, \beta_0)] \right\}. \tag{20}$$

Then by Assumption 2, Theorem 1 and $\psi_t(.)$ assumed to be continuously differentiable using Lemma 19.31 of van der Vaart (2000) (also valid for m-dependent data under Assumption 2)

$$r_T(\Psi_T(\alpha, \beta_0 + b/r_T) - \Psi_T(\alpha, \beta_0)) = \left(\frac{1}{T^{1/2}} \sum_{t=1}^{T} \partial \psi_t(\alpha, \beta_0)/\partial \beta' - E \partial \psi_t(\alpha, \beta_0)/\partial \beta' \right) b + o_p(1). \tag{21}$$

Substitute (21) into (20) we have

$$\Psi_T(\alpha, \beta_0 + b/r_T) - \Psi_T(\alpha, \beta_0) = \left(\frac{1}{T^{1-\kappa}} \sum_{t=1}^{T} \partial \psi_t(\alpha, \beta_0)/\partial \beta' - E \partial \psi_t(\alpha, \beta_0)/\partial \beta' \right) b + o_p((r_T)^{-1}). \tag{22}$$

Then by Assumption 1 and the added assumption (19) we achieve (18). ■.

Proof of Corollary 3. In all semi-weak instruments case (i.e. only β case) the consistency and the rate of convergence proofs are the same given Assumptions 1ii, 2, 4. The limit proof is similar to Theorem 2, so we go over some steps that are different. Then we substitute the efficient weight matrices which gives the desired result. First in our case, following the proof of Theorem 2, by Assumption 2 and the consistency

$$\Psi_T(\beta_0 + \frac{b}{T^{1/2-\kappa}}) \Longrightarrow \Psi(\beta_0),$$

where with the efficient weight matrix $\Psi(\beta_0) \equiv N(0, \Omega_{\beta_0,\beta_0})$. Note that as at the end of the proof of Theorem 2, via equations (19)-(22) we can obtain the above result in an alternative way.

There are no $m_1(.)$ terms now so as in the proof of Theorem 2,

$$T^{1/2-\kappa} m_{2T}(\beta_0 + \frac{b}{T^{1/2-\kappa}}) \to R(\beta_0)b.$$

Then

$$S_T(\beta_0 + \frac{b}{T^{1/2-\kappa}}; \bar{\beta}_T(\beta_0)) \Longrightarrow [\Psi(\beta_0) + R(\beta_0)b]'\Omega^{-1}_{\beta_0,\beta_0}[\Psi(\beta_0) + R(\beta_0)b].$$

Then differentiating the above limit and finding the argmin gives us the desired result. ■

Proof of Theorem 3. Since $E\psi_t(\theta_0) = 0$ Assumption 1 is not used. Then set $W_T(\theta_0) = V_T(\theta_0)^{-1}$,

$$S_T(\theta_0; \theta_0) = \Psi_T(\theta_0)'V_T(\theta_0)^{-1}\Psi_T(\theta_0),$$

then the result follows from Assumption 2 and the weight matrix assumption at θ_0, and equation (8). ■.

Proof of Lemma 1.

$$T^{\kappa}\bar{D}_T(\theta_0) = T^{\kappa}J(\theta_0) + \frac{T^{\kappa}}{T^{1/2}}\{T^{1/2}[\bar{D}_T(\theta_0) - J(\theta_0)]\}. \tag{23}$$

By (10) the second term on the right-hand side of equation (23) converges in probability to zero. Then by (11)

$$T^{\kappa}J(\theta_0) \to C.$$

So

$$T^{\kappa}\bar{D}_T(\theta_0) \to C. \tag{24}$$

This shows that the limit of $T^{\kappa}\bar{D}_T(\theta_0)$ (C) is independent from the limit of the zero mean normal limit for $T^{1/2}\tilde{\Psi}_T(\theta_0)$ which is in Assumption 5.■.

Proof of Theorem 4.

First

$$T^{1/2}\tilde{\Psi}_T(\theta_0)'\Omega^{-1}_{\theta_0,\theta_0}[T^{\kappa}\bar{D}_T(\theta_0)] \xrightarrow{d} \Psi(\theta_0)'\Omega^{-1}_{\theta_0,\theta_0}C.$$

The conditional distribution of $\Psi(\theta_0)'\Omega^{-1}_{\theta_0,\theta_0}C$ given C is $N(0, C'\Omega^{-1}_{\theta_0,\theta_0}C)$. Since C is independent of $\Psi(\theta_0)$ we obtain unconditional result.

Then

$$\begin{aligned}
[T^{1/2}\tilde{\Psi}_T(\theta_0)'\Omega^{-1}_{\theta_0,\theta_0}\bar{D}_T(\theta_0)] \quad &\times \quad [\bar{D}_T(\theta_0)'\Omega^{-1}_{\theta_0,\theta_0}\bar{D}_T(\theta_0)]^{-1/2} \\
&= \quad [(T^{1/2}\tilde{\Psi}_T(\theta_0))'\Omega^{-1}_{\theta_0,\theta_0}(T^{\kappa}\bar{D}_T(\theta_0))][(T^{\kappa}\bar{D}_T(\theta_0))'\Omega^{-1}_{\theta_0,\theta_0}(T^{\kappa}\bar{D}_T(\theta_0))]^{-1/2} \\
&\xrightarrow{d} \quad N(0, I_p).
\end{aligned}$$

Then the desired result follows from the quadratic nature of K-test.■.

Proof of Theorem 5a. First rewrite the test statistics as multiplying and dividing the Wald test by $T^{2\kappa}$:

$$Wald = \left(\frac{T^{1/2}}{T^{\kappa}}a(\hat{\theta}_2)\right)' \left[A(\hat{\theta}_2)\left\{G_T(\hat{\theta}_2)'V_T(\hat{\theta}_2)^{-1}G_T(\hat{\theta}_2)\right\}^{-1}A(\hat{\theta}_2)'\right]^{-1}\left(\frac{T^{1/2}}{T^{\kappa}}\right)a(\hat{\theta}_2), \tag{25}$$

where

$$G_T(\hat{\theta}_2) = \frac{T^{\kappa}}{T}\sum_{t=1}^{T}\partial\psi_t(\hat{\theta}_2)/\partial\theta'. \tag{26}$$

Note that by Assumption 1' and consistency of $\hat{\theta}_2$

$$G_T(\theta) \xrightarrow{p} R(\theta), \tag{27}$$

uniformly in N a neighborhood of θ_0.

Now have a Taylor series expansion of $a(\hat{\theta}_2)$ around θ_0 and use the null, also see $\bar{\theta} \in (\theta_0, \hat{\theta}_2)$,

$$\begin{aligned}\frac{T^{1/2}}{T^{\kappa}}a(\hat{\theta}_2) &= A(\bar{\theta}_2)\frac{T^{1/2}}{T^{\kappa}}(\hat{\theta}_2-\theta_0)\\ &\xrightarrow{d} N(0, A(\theta_0)(R(\theta_0)'\Omega^{-1}_{\theta_0,\theta_0}R(\theta_0))^{-1}A(\theta_0)').\end{aligned} \tag{28}$$

by Corollary 3. Then use Assumption 1' with (28) and the consistency of the estimates to have the result. Note that Corollary 3 also works under Assumption 1'. ■.

Proof of Theorem 5b. Set $\tilde{\theta}_1$ as the any consistent first step GMM estimator. First of all in Theorem 9.1. of Newey and McFadden (1994) set up the objective function as $T^{2\kappa-1}S_T(\theta : \tilde{\theta}_1)$ and then proceed exactly as in the proof of Theorem 9.1 in Newey and McFadden (1994) to have the consistency result. $\tilde{\theta}_2 \xrightarrow{p} \theta_0$. See that the objective function $S_T(.)$ is defined in section 1. $\tilde{\theta}_2$ represents the restricted two-step GMM estimator.

First see that

$$T^{\kappa-1}\sum_{t=1}^{T}\psi_t(\tilde{\theta})_2 = T^{\kappa-1/2}[T^{-1/2}\sum_{t=1}^{T}\psi_t(\tilde{\theta}_2) - E\psi_t(\tilde{\theta}_2)] + T^{\kappa}ET^{-1}\sum_{t=1}^{T}\psi_t(\tilde{\theta}_2).$$

Then by (4) Assumption 1' and consistency of $\tilde{\theta}_2$ and $E\psi_t(\theta_0) = 0$,

$$T^{\kappa-1}\sum_{t=1}^{T}\psi_t(\tilde{\theta}_2) \xrightarrow{p} 0. \tag{29}$$

Using the consistency and Assumption 1' and (29)

$$\begin{aligned}T^{2\kappa-1}\frac{\partial S_T(\tilde{\theta}_2;\tilde{\theta}_1)}{\partial\theta} &= \left(T^{\kappa-1}\frac{\sum_{t=1}^{T}\partial\psi_t(\tilde{\theta}_2)}{\partial\theta'}\right)' V_T(\tilde{\theta}_1)^{-1}\left(T^{\kappa-1}\sum_{t=1}^{T}\psi_t(\tilde{\theta}_2)\right)\\ &\xrightarrow{p} 0.\end{aligned}$$

Since by first order conditions

$$T^{2\kappa-1}\frac{\partial S_T(\tilde{\theta}_2;\tilde{\theta}_1)}{\partial\theta} = -A(\theta_0)'T^{2\kappa-1}\tilde{\gamma} + o_p(1). \tag{30}$$

by consistency of the constrained GMM estimator. Since $A(\theta_0)$ has full rank we have

$$T^{2\kappa-1}\tilde{\gamma} \xrightarrow{p} 0.$$

Then Under Assumption 2 central limit Theorem implies:

$$\Omega_{\theta_0,\theta_0}^{-1/2}T^{1/2}(\frac{\sum_{t=1}^T \psi_t(\theta_0)}{T}) = U_T \xrightarrow{d} N(0,I). \tag{31}$$

For the constrained sample moment use the following Taylor series expansion:

$$T^{1/2}(\frac{1}{T}\sum_{t=1}^T \psi_t(\tilde{\theta}_2)) = \frac{1}{T^{1/2}}\sum_{t=1}^T \psi_t(\theta_0) + (\frac{T^\kappa}{T}\frac{\sum_{t=1}^T \partial\psi_t(\bar{\theta})}{\partial\theta'})\frac{T^{1/2}}{T^\kappa}(\tilde{\theta}_2-\theta_0), \tag{32}$$

where $\bar{\theta} \in (\theta_0, \tilde{\theta}_2)$. Then have the following Taylor series expansion for the restrictions:

$$\frac{T^{1/2}}{T^\kappa}a(\tilde{\theta}_2) = A(\bar{\theta})\frac{T^{1/2}}{T^\kappa}(\tilde{\theta}_2-\theta_0). \tag{33}$$

Now use the first order condition for θ in the constrained case and benefit from (32) and multiply by $T^{\kappa-1/2}$ to have

$$\begin{aligned} 0 &= \left(\frac{T^\kappa}{T}\frac{\sum_{t=1}^T \partial\psi_t(\tilde{\theta}_2)}{\partial\theta'}\right)' V_T(\tilde{\theta}_1)^{-1}\left(\frac{1}{T^{1/2}}\sum_{t=1}^T \psi_t(\tilde{\theta}_2)\right) - A(\theta_0)'\frac{T^\kappa}{T^{1/2}}\tilde{\gamma} + o_p(1) \\ &= \left(\frac{T^\kappa}{T}\frac{\sum_{t=1}^T \partial\psi_t(\tilde{\theta}_2)}{\partial\theta'}\right)' V_T(\tilde{\theta}_1)^{-1}\left(\frac{1}{T^{1/2}}\sum_{t=1}^T \psi_t(\theta_0)\right) \\ &+ \left(\frac{T^\kappa}{T}\frac{\sum_{t=1}^T \partial\psi_t(\tilde{\theta}_2)}{\partial\theta'}\right)' V_T(\tilde{\theta}_1)^{-1}\left(\frac{T^\kappa}{T}\frac{\sum_{t=1}^T \partial\psi_t(\bar{\theta})}{\partial\theta'}\right)\left(\frac{T^{1/2}}{T^\kappa}(\tilde{\theta}_2-\theta_0)\right) \\ &- A(\theta_0)'\frac{T^\kappa}{T^{1/2}}\tilde{\gamma} + o_p(1). \end{aligned} \tag{34}$$

where $\bar{\theta} \in (\theta_0, \tilde{\theta})$. By (33) and Assumption 6 and by the first order conditions for γ,

$$0 = A(\theta_0)\frac{T^{1/2}}{T^\kappa}(\tilde{\theta}-\theta_0) + o_p(1). \tag{35}$$

Use these first order conditions (34)(35) with simple matrix algebra and Assumption 1',2,4; as in equations (9.4)(9.5) of Newey and McFadden (1994) we have the following version of (9.6) in Newey and McFadden (1994):

$$T^{1/2-\kappa}(\tilde{\theta}-\theta_0) = B(\theta_0)^{-1/2}M(\theta_0)B(\theta_0)^{-1/2}R(\theta_0)'\Omega_{\theta_0,\theta_0}^{-1/2}U_T + o_p(1), \tag{36}$$

and

$$T^{\kappa-1/2}\tilde{\gamma} = (A(\theta_0)B(\theta_0)^{-1}A(\theta_0)')^{-1}A(\theta_0)B(\theta_0)^{-1}R(\theta_0)'\Omega_{\theta_0,\theta_0}^{-1/2}U_T + o_p(1), \quad (37)$$

where $B(\theta_0) = R(\theta_0)'\Omega_{\theta_0,\theta_0}^{-1}R(\theta_0)$ and

$$M(\theta_0) = I - B(\theta_0)^{-1/2}A(\theta_0)'(A(\theta_0)B(\theta_0)^{-1}A(\theta_0)')^{-1}A(\theta_0)B(\theta_0)^{-1/2}.$$

Note also that the limit variance matrix for $\tilde{\gamma}$ is $(A(\theta_0)B(\theta_0)^{-1}A(\theta_0)')^{-1}$.

Then rewrite the LM test by multiplying and dividing by $T^{2\kappa}$

$$LM = (\frac{T^\kappa}{T^{1/2}}\tilde{\gamma})'A(\tilde{\theta}_2)\left\{\left(\frac{T^\kappa}{T}\frac{\sum_{t=1}^T \partial\psi_t(\tilde{\theta}_2)'}{\partial\theta}\right)' V_T(\tilde{\theta}_2)^{-1}\left(\frac{T^\kappa}{T}\frac{\sum_{t=1}^T \partial\psi_t(\tilde{\theta}_2)'}{\partial\theta}\right)\right\}^{-1} A(\tilde{\theta}_2)'(\frac{T^\kappa}{T^{1/2}}\tilde{\gamma}).$$

First note that in the above test

$$A(\tilde{\theta}_2)\left\{\left(\frac{T^\kappa}{T}\frac{\sum_{t=1}^T \partial\psi_t(\tilde{\theta}_2)'}{\partial\theta}\right)' V_T(\tilde{\theta}_2)^{-1}\left(\frac{T^\kappa}{T}\frac{\sum_{t=1}^T \partial\psi_t(\tilde{\theta}_2)'}{\partial\theta}\right)\right\}^{-1} A(\tilde{\theta}_2)' \xrightarrow{p} A(\theta_0)B(\theta_0)^{-1}A(\theta_0)'.$$

Use consistency of the constrained estimator with Assumption 6 and (31), (37) to have the desired result. Also note that in LM test we use $V_T(\tilde{\theta}_2)$ instead of $V_T(\tilde{\theta}_1)$, this does not make any impact on asymptotics.

Note that our LM test is slightly different than that exists in Newey and McFadden (1994) the main reason is how we scale the objective function if we had scaled our objective function like

$$S_T^* = (\theta;\theta) = (1/T\sum_{t=1}^T \psi_t(\theta))'\Omega_{\theta_0,\theta_0}^{-1}(1/T\sum_{t=1}^T \psi_t(\theta)),$$

then we could have found $T^{1/2+\kappa}\tilde{\gamma} = O_p(1)$. And our test would have the same form as in Newey and McFadden LM test.■.

Proof of Theorem 5c. This proof is basically the same as in the proof of Theorem 9.2 in Newey and McFadden. The only difference is the following Taylor series expansion:

$$\frac{1}{T^{1/2}}\sum_{t=1}^T \psi_t(\hat{\theta}) = \frac{1}{T^{1/2}}\sum_{t=1}^T \psi_t(\tilde{\theta}) + \left(\frac{T^\kappa}{T}\frac{\sum_{t=1}^T \partial\psi_t(\tilde{\theta})'}{\partial\theta}\right)\left(\frac{T^{1/2}}{T^\kappa}(\hat{\theta}-\tilde{\theta})\right) + o_p(1).$$

Then simply apply Corollary 3, (36) and Assumption 1' and follow exactly the same steps in Thoerm 9.2 in Newey and McFadden to reach the result.■.

■.

We now briefly discuss why the limits of Wald, LR and LM tests in CUE framework are the same with two-step counterparts. The main difference between Wald, LR, LM tests in two-step and CUE cases is the Jacobian of the objective function. Take the case of LM test and use (30)

$$LM = \frac{1}{T}\frac{\partial S_T(\tilde{\theta};\tilde{\theta})}{\partial\theta'}[A(\theta_0)'(A(\theta_0)B(\theta_0)^{-1}A(\theta_0)')^{-1}A(\theta_0)']^{-}\frac{\partial S_T(\tilde{\theta};\tilde{\theta})}{\partial\theta},$$

where $[.]^{-}$ represents the Moore-Penrose inverse of the given matrix in square brackets. In that term actually there is only one difference, for CUE type test

$$\frac{\partial S_T(\tilde{\theta}_c;\tilde{\theta}_c)}{\partial \theta_c'} = T^{\kappa}\bar{D}_T(\tilde{\theta}_c)'V_T(\tilde{\theta}_c)^{-1}T^{1/2}\Psi_T(\tilde{\theta}_c).$$

For two-step case from (26)

$$\frac{\partial S_T(\tilde{\theta}_2;\tilde{\theta}_2)}{\partial \theta_2'} = G_T(\tilde{\theta}_2)'V_T(\tilde{\theta}_2)^{-1}T^{1/2}\Psi_T(\tilde{\theta}_2).$$

Note that $\tilde{\theta}_2$ represents the two-step estimator, $\tilde{\theta}_c$ represents the CUE. Both of them are constrained estimators and consistent. This is mentioned in Theorem 5b and this is similar in CUE case. By Assumption 4 the efficient weight matrices in those cases go the same limit weight matrix. Now benefiting from K-test terminology in section 3.1 write the following:

$$T^{\kappa}\bar{D}_T(\tilde{\theta}_c) = [T^{\kappa}\bar{D}_{T1}(\tilde{\theta}_c),\cdots,T^{\kappa}\bar{D}_{Tp}(\tilde{\theta}_c)],$$

where the vectors are, by benefiting from (26), and (9)

$$T^{\kappa}\bar{D}_{Ti}(\tilde{\theta}_c) = G_{Ti}(\tilde{\theta}_c) - A_iV_{qi}(\tilde{\theta}_c)V_T(\tilde{\theta}_c)^{-1}T^{\kappa}\sum_{t=1}^{T}\frac{\psi_t(\tilde{\theta}_c)}{T}.$$

$V(.)$ matrices are the estimators of Ω matrices described in section 3.1. By the continuity of $\psi_t(\theta)$ and consistency of $\tilde{\theta}_c$ with Assumption 1' and (27)

$$G_T(\tilde{\theta}_c) \xrightarrow{p} R(\theta_0). \tag{38}$$

Then again by consistency and (3) and Assumption 1' and since $E\psi_t(\theta_0) = 0$,

$$\begin{aligned} T^{\kappa}\sum_{t=1}^{T}\frac{\psi_t(\tilde{\theta}_c)}{T} &= \frac{T^{\kappa}}{T}\sum_{t=1}^{T}\psi_t(\tilde{\theta}_c) - E\psi_t(\tilde{\theta}_c) \\ &+ \frac{T^{\kappa}}{T}\sum_{t=1}^{T}E\psi_t(\tilde{\theta}_c) \\ &\xrightarrow{p} 0, \end{aligned} \tag{39}$$

So by Assumption 4 and (38)(39)

$$T^{\kappa}\bar{D}_T(\tilde{\theta}_c) = G_T(\tilde{\theta}_c) + o_p(1). \tag{40}$$

By (40) and the consistency of the estimators $\tilde{\theta}_c, \tilde{\theta}_2$; the jacobian terms in both test statistics are of equivalent asymptotically and the remainder terms do not play any role asymptotically.

Proof of Theorem 6.

First by Taylor series expansion

$$T^{-1/2}\sum_{t=1}^{T}\psi_t(\hat\theta_2)=T^{-1/2}\sum_{t=1}^{T}\psi_t(\theta_0)+(T^{\kappa-1}\sum_{t=1}^{T}\frac{\partial\psi_t(\bar\theta)}{\partial\theta'})(T^{1/2-\kappa}(\hat\theta_2-\theta_0)), \tag{41}$$

where $\bar\theta\in(\theta_0,\hat\theta_2)$. Then seeing that $\hat\theta_2$ is the argmin of the objective function in two-step GMM and using (41) we have

$$\begin{aligned}
T^{1/2-\kappa}(\hat\theta_2-\theta_0) &= -\left\{\left[T^{\kappa-1}\sum_{t=1}^{T}\frac{\partial\psi_t(\hat\theta_2)}{\partial\theta'}\right]' V_T(\hat\theta_2)^{-1}\left[T^{\kappa-1}\sum_{t=1}^{T}\frac{\partial\psi_t(\hat\theta_2)}{\partial\theta'}\right]\right\}^{-1}\\
&\times \left\{\left[T^{\kappa-1}\sum_{t=1}^{T}\frac{\partial\psi_t(\hat\theta_2)}{\partial\theta'}\right] V_T(\hat\theta_2)^{-1}(T^{-1/2}\sum_{t=1}^{T}\psi_t(\theta_0))\right\}\\
&+ o_p(1).
\end{aligned} \tag{42}$$

Note that we used $V_T(\hat\theta_2)$ instead of $V_T(\hat\theta_1)$. This does not make any difference since both two-step and first step estimators are consistent and we use Assumption 4. Then rewrite (41) by (42) and by using $G_T(\theta)$ definition in (26) to have

$$T^{-1/2}\sum_{t=1}^{T}\psi_t(\hat\theta_2)=T^{-1/2}\sum_{t=1}^{T}\psi_t(\theta_0)-G_T(\bar\theta)[G_T(\hat\theta_2)'V_T(\hat\theta_2)^{-1}G_T(\hat\theta_2)]^{-1}[G_T(\hat\theta_2)'V_T(\hat\theta_2)^{-1}(T^{-1/2}\sum_{t=1}^{T}\psi_t(\theta_0))]. \tag{43}$$

Then use (43) in the test statistic given in Theorem 6 to have

$$\begin{aligned}
J_{2s}(\hat\theta_2) &= [T^{-1/2}\sum_{t=1}^{T}\psi_t(\theta_0)]'V_T(\hat\theta_2)^{-1}[T^{-1/2}\sum_{t=1}^{T}\psi_t(\theta_0)]\\
&- [(T^{-1/2}\sum_{t=1}^{T}\psi_t(\theta_0))'V_T(\hat\theta_2)^{-1}G_T(\hat\theta_2)][G_T(\hat\theta_2)'V_T(\hat\theta_2)^{-1}G_T(\hat\theta_2)]^{-1}[G_T(\hat\theta_2)'V_T(\hat\theta_2)^{-1}(T^{-1/2}\sum_{t=1}^{T}\psi_t(\theta_0))]+o_p(1)
\end{aligned} \tag{44}$$

Then by Assumptions 1', 2 and 4, (27) we get the desired result.■.

Proof of Lemma 2. We first go over $J_{2s}(\theta_0)$ proof. Others are very similar; we use $W_{CUE}(\theta_0)$ and $LM(\theta_0)$ test statistics in the proof instead of $W_{2s}(\theta_0)$ that is used here.

Write the test statistic as follows:

$$\begin{aligned}
J_{2s}(\theta_0) &= \left[\frac{1}{T^{1/2}}\sum_{t=1}^{T}\psi_t(\theta_0)\right]' V_T(\theta_0)^{-1}\left[\frac{1}{T^{1/2}}\sum_{t=1}^{T}\psi_t(\theta_0)\right]\\
&- \left[(T^{-1/2}\sum_{t=1}^{T}\psi_t(\theta_0))'V_T(\hat\theta_2)^{-1}G_T(\hat\theta_2)\right]\\
&\times [G_T(\hat\theta_2)'V_T(\hat\theta_2)^{-1}G_T(\hat\theta_2)]^{-1}[G_T(\hat\theta_2)'V_T(\hat\theta_2)^{-1}(T^{-1/2}\sum_{t=1}^{T}\psi_t(\theta_0))]
\end{aligned}$$

where the first term on the right hand side is $S_T(.)$ test statistic and the second one is the asymptotically equivalent form of $W_{2s}(\theta_0)$. $G_T(\hat\theta_2)$ is defined in (26). We can rewrite these terms by using (31)(27) and consistency with Assumption 1', 4

$$J_{2s}(\theta_0) = U_T'[I - \Omega_{\theta_0,\theta_0}^{-1/2} R(\theta_0)[R(\theta_0)'\Omega_{\theta_0,\theta_0}^{-1} R(\theta_0)]^{-1} R(\theta_0)'\Omega_{\theta_0,\theta_0}^{-1/2}]U_T + o_p(1), \tag{45}$$

where the middle term is idempotent with rank $GK - p$.Then we have the desired result. The limit for $J_K(\theta_0)$ test can be obtained easily from section 3.1 ■.

Proof of Theorem 7. We benefit heavily from the proof of Theorem 1 in Kleibergen (2003). That proof basically depends on Lemma 1, Corollaries 1 and 2 in Kleibergen (2003), the rest of the proof of Theorem 1 in Kleibergen (2003) is simple tedious algebra regardless of the properties of instruments. So we only prove Lemma 1, Corollaries 1 and 2 in Kleibergen (2003) in the case of semi-weak instruments, then since the rest of the proof is exactly the same as in the proof of Theorem 1 in Keibergen (2003) reader may refer to that article.

We use Assumption 5 and the setup of semi-weak case as defined in equation (11). First set

$$T\bar{D}_T(\theta_0) = D_T(\theta_0),$$

where $\bar{D}_T(\theta_0)$ is explained in detail in section 3.1. Define ν in the range of $0 < \nu \leq 1$ which is the concentration parameter in Kleibergen (2003), then by Assumption 5

$$T^{-1/2(1+\nu)} D_T(\theta_0) = D_0 + o_p(T^{-\nu/2}), \tag{46}$$

where

$$D_0 = T^{1/2(1-\nu)} \frac{1}{T} \sum_{t=1}^{T} E(p_t(\theta_0)|I_t) + T^{-\nu/2} \Psi_{q\perp\theta_0}, \tag{47}$$

with the matrix $\Psi_{q\perp\theta_0}$ is $GK \times p$ matrix with each ith column is

$$\Psi_{qi\perp}(\theta_0) = \Psi_{qi}(\theta_0) - A_i \Omega_{qi} \Omega_{\theta_0,\theta_0}^{-1} \Psi(\theta_0).$$

This result in (46) (47) is directly obtained from Assumption 5, (8) and the definition of $D_T(\theta_0)$, and $p_{it}(\theta_0) - Ep_{it}(\theta_0) = A_i(q_{it}(\theta_0) - Eq_{it}(\theta_0))$. Equations (46)(47) extend Lemma 1 of Kleibergen (2003) which is written only for $\nu = 0$ and $\nu = 1$. Our result also holds when $\nu = 0$ but that case is not our concern. In the remaining parts of the proof we relate ν to κ and hence show that $0 < \nu < 1$ corresponds to nearly-weak instruments framework, whereas $\nu = 0$ corresponds to weak IV and $\nu = 1$ corresponds to strong instruments.

First we present a Lemma from Kleibergen (2001) which is a simple algebraic transformation of Assumption 5 and holds regardless of the identification issues. $(0 \leq \nu \leq 1)$

Lemma A.1 (Lemma 1, Kleibergen, 2001) *Under Assumption 5,*

$$T^{1/2} vec(\frac{1}{T} D_T(\theta_0) - J(\theta_0)) \xrightarrow{d} A\Psi_{q\perp\theta_0},$$

$$T^{1/2} \tilde{\Psi}_T(\theta_0) \xrightarrow{d} \Psi(\theta_0),$$

where $A = diag(A_1, \cdots, A_p)$ *and* $\Psi_{q\perp\theta_0} = \Psi_q(\theta_0) - \Omega_{q,\theta_0}\Omega^{-1}_{\theta_0,\theta_0}\Psi(\theta_0)$ *and* $\Psi_{q\perp\theta_0}$ *is normal zero mean and is independent of* $\Psi(\theta_0)$ *.*

We now use (10), Lemma A.1 and (46) to have

$$T^{1/2} vec\left[\frac{D_T(\theta_0)}{T} - J(\theta_0)\right] = T^{1/2} vec[T^{\frac{-1}{2}(1-\nu)} D_0 - J(\theta_0)] \xrightarrow{d} A\Psi_{q,\perp}(\theta_0). \tag{48}$$

where as described in section 3.1 $J(\theta_0) = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E(p_t(\theta_0)|I_t)$. Note that (48) works regardless of $\nu = 0, \nu = 1, 0 < \nu < 1$. This is clear from Lemma A.1 and (46). This shows that Corollary 1 of Kleibergen (2003) holds also for $0 < \nu \leq 1$ which is the case of concern here.

Now we have to relate ν to κ and show that Corollary 2 of Kleibergen (2003) also holds when $0 < \nu \leq 1$.

To find why $0 < \nu \leq 1$ correspond to semi-weak case and the relationship between ν and κ "index of the quality of the instruments", proceed by (46)(47) and using $J(\theta_0)$ definition to have

$$D_0 = T^{\frac{1}{2}(1-\nu)} J(\theta_0) + O_p(T^{-\nu/2}). \tag{49}$$

In the case of semi-weak identification it is assumed that

$$J(\theta_0) = \frac{C}{T^\kappa}, \tag{50}$$

where C is a full rank, finite matrix and $0 \leq \kappa < 1/2$. Note that Corollary 2 of Kleibergen (2003) establishes $D_0 = O_p(1)$ which is crucial to higher order expansions when $\nu = 0, \nu = 1$. We want to show that Corollary 2 of Kleibergen (2003) holds also under $0 < \nu \leq 1$. Then substitute (50) into (49) to have

$$D_0 = T^{\frac{1}{2}(1-\nu)}\frac{C}{T^\kappa} + o_p(1).$$

Then

$$D_0 \xrightarrow{p} C, \tag{51}$$

when $\kappa = \frac{1}{2}(1-\nu)$. So since $0 \leq \kappa < 1/2$, we verify $0 < \nu \leq 1$. So with (10) and $\kappa = 1/2(1-\nu)$ we have

$$D_0'\Omega^{-1}_{\theta_0,\theta_0} D_0 \xrightarrow{p} C'\Omega^{-1}_{\theta_0,\theta_0} C. \tag{52}$$

Equation (52) extends the Corollary 2 of Kleibergen (2003) to semi-weak case of $0 \leq \kappa < 1/2$ (i.e. $0 < \nu \leq 1$). Then exactly following the proof of Theorem 1 in Kleibergen (2003) gives the higher-order expansions.

The expansions are as follows:

$$\begin{aligned} W_{2s}(\theta_0) &= \eta_0 + T^{-\nu/2}\eta_1 + T^{-\delta/2}\eta_2 + T^{-(\nu+\delta)/2}\eta_3 \\ &+ T^{-\nu}\eta_4 + T^{-\delta}\eta_5 + T^{-1/2(2\nu+\delta)}\eta_6 + T^{-1/2(\nu+2\delta)}\eta_7 + T^{-3/2\nu}\eta_8 + o_p(T^{-3/2\nu}), \end{aligned}$$

where $\delta = min(\mu, \nu)$ and $0 < \nu \leq 1$. All η_j terms are $O_p(1)$. Now we give the result for CUE based Wald test:

$$
\begin{aligned}
W_{CUE}(\theta_0) &= \eta_0 + T^{-\delta/2}\eta_2 + T^{-(\nu+\delta)/2}\eta_3 \\
&+ T^{-\delta}\eta_5 + T^{-1/2(\nu+2\delta)}\eta_7 + o_p(T^{-3/2\nu}),
\end{aligned}
$$

So $W_{CUE}(\theta_0)$ is the same as $W_{2s}(\theta_0)$ except that we donot have $\eta_1, \eta_4, \eta_6, \eta_8$.

LM test results are the same as $W_{2s}(\theta_0)$ except we set μ instead of δ in the expansions. K test higher order expansion is the same as $W_{CUE}(\theta_0)$ but again we substitute μ instead of δ.

Note that following the proof of Theorem 1 in Kleibergen (2003)

$$
\eta_0 = D_0'\Omega_{\theta_0,\theta_0}^{-1}\Psi_T(\theta_0)[D_0'\Omega_{\theta_0,\theta_0}^{-1}D_0]^{-1}\Psi_T(\theta_0)'\Omega_{\theta_0,\theta_0}^{-1}D_0. \tag{53}
$$

By (51) and (7) we obtain

$$
\eta_0 \xrightarrow{d} \chi_p^2. \tag{54}
$$

So in large samples all these tests converge to χ_p^2 limit and the other higher order terms converge to zero. This limit distribution is also found in general testing for Wald, LM in section 3.2, for for simple null of $H_0 : \theta = \theta_0$ for the K test in section 3.1.

Then in these expansions since we try to relate them to the "quality of instruments" (i.e. κ), we use our finding $\kappa = 1/2(1-\nu)$ and substitute that into higher order expansions to get the results in the statement of Theorem 6. ■.

Proof of Corollary 4. We start with the proof of $J_{2s}(\theta_0)$. Since

$$
J_{2s}(\theta_0) = S_T(\theta_0; \theta_0) - W_{2s}(\theta_0)
$$

as shown before Lemma 2, we analyze the higher order expansion for $S_T(.)$ test.

$$
\begin{aligned}
S_T(\theta_0; \theta_0) &= \Psi_T(\theta_0)'V_T(\theta_0)^{-1}\Psi_T(\theta_0) \\
&= \Psi_T(\theta_0)'[V_T(\theta_0)^{-1} - \Omega_{\theta_0,\theta_0}^{-1}]\Psi_T(\theta_0) \\
&+ \Psi_T(\theta_0)'\Omega_{\theta_0,\theta_0}^{-1}\Psi_T(\theta_0)
\end{aligned}
$$

which can be rewritten as:

$$
\begin{aligned}
S_T(\theta_0; \theta_0) &= \Psi_T(\theta_0)'\Omega_{\theta_0,\theta_0}^{-1}\Psi_T(\theta_0) \\
&+ T^{-\mu/2}[\Psi_T(\theta_0)'T^{\mu/2}[V_T(\theta_0)^{-1} - \Omega_{\theta_0,\theta_0}^{-1}]\Psi_T(\theta_0)] \\
&+ o_p(T^{-\mu/2}) \\
&= \omega_0 + T^{-\mu/2}\omega_\mu + o_p(T^{\mu/2}),
\end{aligned}
$$

where

$$\omega_0 = \Psi_T(\theta_0)'\Omega^{-1}_{\theta_0,\theta_0}\Psi_T(\theta_0).$$

$$\omega_\mu = T^{\mu/2}\Psi_T(\theta_0)'[V_T(\theta_0)^{-1} - \Omega^{-1}_{\theta_0,\theta_0}]\Psi_T(\theta_0).$$

Rewrite J_{2s} and use (53)(54) with ω_0 definition to have

$$\omega_0 - \eta_0 \xrightarrow{d} \chi^2_{GK-p}.$$

Then using $J_{2s}(\theta_0)$ definition in terms of $S_T(\theta_0;\theta_0)$ and $W_{2s}(\theta_0)$, and using the higher order expansions for these terms we have

$$\begin{aligned} J_{2s}(\theta_0) &= (\omega_0 - \eta_0) + T^{-\mu/2}\omega_\mu \\ &- (\frac{\eta_1}{T^{1/2-\kappa}} + \frac{\eta_2}{T^{\delta/2}} + \frac{\eta_3}{T^{(1/2-\kappa)+\delta/2}} + \frac{\eta_4}{T^{1-2\kappa}}) \\ &- (\frac{\eta_5}{T^{\delta}} + \frac{\eta_6}{T^{1-2\kappa+\delta/2}} + \frac{\eta_7}{T^{1/2-\kappa+\delta}} + \frac{\eta_8}{T^{3/2-3\kappa}} + o_p(T^{3\kappa-3/2})), \end{aligned}$$

Other expansions for J test statistics involving the true parameter are derived exactly in the same way since $S_T(\theta_0;\theta_0)$ test expansion is the same for all four J test statistics.

REFERENCES

Anderson, T. W., Rubin, H., 1949. Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations. Annals of Mathematical Statistics, 20, 46-63.

Andrews, D. W. K., 1994. Empirical Process Methods in Econometrics. Handbook of Econometrics, vol.4, 2247-2294.

Beran, R., Ducharme, G.R., 1991. Asymptotic Theory for Bootstrap Methods in Statistics.

Chao, J.C., Swanson, N.R., 2003. Consistent Estimation with a Large Number of Weak Instruments. Unpublished Manuscript.

Hahn, J. Kuersteiner,G., 2002. Discontinuities of Weak Instrument Limiting Distributions. Economics Letters, 75, 325-331.

Hansen, L.P, 1982. Large Sample Properties of Generalized Method of Moments Estimators. Econometrica, 50, 1029-1054.

Hansen, L. P., Heaton,J., Yaron, A., 1996. Finite Sample Properties of Some Alternative GMM Estimators. Journal of Business and Economic Statistics 14, 262-280.

Horowitz, J., 2001. The Bootstrap. Handbook of Econometrics, vol.5, 3160-3212.

Kleibergen, F. 2001. Testing Parameters in GMM Without Assuming That They Are Identified. Manuscript, University of Amsterdam.

Kleibergen, F. 2003. Expansions of GMM Statistics that Indicate Their Properties under Weak and or many Instrumennts and The Bootstrap. Unpublished Manuscript.

Newey, W.K., McFadden, D., 1994. Large Sample Estimation and Hypothesis Testing. Handbook of Econometrics, Vol.4, ed. by R.Engle and D.McFadden. Amsterdam: North Holland, 2113-2247.

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 Instrumnets 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. First Paperback Edition.

Van Der Vaart, A. W., Wellner ,J.A. 1996. Weak Convergence and Empirical Processes. New York: Springer.