GMM with Weak Identification and Near Exogeneity

Ying Fang*
Wang Yanan Institute for Studies in Economics
Xiamen University
Xiamen, Fujian Province, 361005, China
yifst1@pitt.edu

March 2006

Abstract

This chapter studies the asymptotic properties of estimation and inference with weak identification and near exogeneity in a GMM framework with instrumental variables. We obtained limiting results under weak identification and near exogeneity of general GMM estimators and some specific GMM estimators, such as one-step GMM estimator, two-step GMM estimator and continuous updating estimator. We also examine the asymptotic properties of the Anderson-Rubin type and the Kleibergen type tests under weak identification and near exogeneity.

JEL Classification: C12, C15
Key Words: Near Exogeneity, Weak Identification, GMM

*I would like to express my gratitude to my advisor Prof. Mehmet Caner and Professors Daniel Berkowitz, Oliver Board, Karen Clay, David DeJong, Taisuke Otsu, Jean-Francois Richard, and Nese Yildiz for their helpful conversations and suggestions. I also wish to thank Xin He and Yong Sui for their helpful discussions.

1 Introduction

This chapter studies the asymptotic properties of estimation and inference with weak identification and near exogeneity in a GMM framework with instrumental variables. GMM is a natural extension of a linear simultaneous equations model which allows a set of nonlinear and non-differentiable equations. The technique used in Chapter 1 which is mainly based on mean value theorem and the classic central limit theorem cannot be applied into a nonlinear and non-differentiable environment. We can benefit from empirical process theory and the functional central limit theorem to establish large sample properties. We obtained limiting results under weak identification and near exogeneity of general GMM estimators and some specific GMM estimators, such as one-step GMM estimator, two-step GMM estimator and continuous updating estimator. We also examine the asymptotic properties of the Anderson-Rubin type and the Kleibergen type tests under weak identification and near exogeneity.

This chapter is organized as follows. Section 2 decribes the model and assumptions. Section 3 examines the limiting results of GMM estimators under near exogeneity and weak identification. Section 4 studies inference under near exogeneity and weak identification, and Section 5 concludes.

2 The Model and Assumptions

In this chapter, we consider a GMM framework with instrumental variables under weak identification and near exogeneity. Let $\theta = (\alpha', \beta')$ be an m-dimensional unknown parameter vector with true value $\theta_0 = (\alpha_0', \beta_0')'$ in the interior of the compact parameter space Θ. The true value θ_0 satisfies some conditional moment restrictions which can be explicitly written as

$$E\phi_i(\theta_0) = E[h(Y_i, \theta_0) \otimes Z_i] = C/\sqrt{N}, \tag{1}$$

where $h(.)$ is a real valued $H \times 1$ vector of functions, Z_i is a $K \times 1$ vector of instrumental variables, and Y_i is the observation which possibly contains endogenous variables, for $i = 1, 2, ..., N$ and $HK \geq m$. The C is a $HK \times 1$ vector of constants. When C is a vector of zeros, this is the GMM model with instrumental variables defined by Stock and Wright (2000). When C is not all zeros, Equation (1) defines the GMM model with near exogeneity. The degree of near exogeneity is local to zero. When the sample size N grows to large, the correlation between $h(.)$ and the instruments Z_i tends to zero. The linear simultaneous equations model defined in Chapter 1 is a special case of Equation (1), where

$$E\phi_i(\theta_0) = E[Z_i'(y_i - Y_i\theta_0)] = C/\sqrt{N}. \tag{2}$$

So $h(.) = y_i - Y_i\theta_0$ is a linear function and $Y_i = (y_i, Y_i)$ contains only endogenous variables. But in this chapter, the $h(.)$ can be a set of general nonlinear functions with possible non-differentiability.

We follow Stock and Wright (2000)'s paper to consider a mixed case in which a subset of θ, say α, is weakly identified. Let $\Theta = A \times B$, where $\alpha \in A$ is an $m_1 \times 1$ vector, $\beta \in B$ is an $m_2 \times 1$ vector, and $m_1 + m_2 = m$. Also, let $\widetilde{m}_N(\alpha, \beta) = EN^{-1} \sum_{i=1}^{N} \phi_i(\alpha, \beta)$. Now, we can utilize the following identity,

$$\widetilde{m}_N(\alpha, \beta) = \widetilde{m}_N(\alpha_0, \beta_0) + \widetilde{m}_{1N}(\alpha, \beta) + \widetilde{m}_{2N}(\beta) \tag{3}$$

where

$$\widetilde{m}_{1N}(\alpha, \beta) = \widetilde{m}_N(\alpha, \beta) - \widetilde{m}_N(\alpha_0, \beta) \tag{4}$$

and

$$\widetilde{m}_{2N}(\beta) = \widetilde{m}_N(\alpha_0, \beta) - \widetilde{m}_N(\alpha_0, \beta_0) \tag{5}$$

The identification of θ requires whether the moment restrictions can be satisfied uniquely. If β is strictly identified, then $\widetilde{m}_{2N}(\beta)$ should be large when $\beta \neq \beta_0$. However, $\widetilde{m}_{1N}(\alpha, \beta)$ should be close to zero when $\alpha \neq \alpha_0$ and $\beta = \beta_0$ if α is weakly identified. We can use a local to zero model to define the weak identification of the α,

$$\widetilde{m}_N(\alpha, \beta) - \widetilde{m}_N(\alpha_0, \beta) = m_{1N}(\alpha, \beta)/\sqrt{N} \tag{6}$$

where $m_{1N}(\alpha, \beta) : A \times B \to R^{HK}$ is a set of continuous functions such that $m_{1N}(\theta) \to m_1(\theta)$ uniformly on Θ as N grows to large. The $m_1(\theta) : A \times B \to R^{HK}$ is a set of continuous functions and is bounded on Θ. Also, let $\widetilde{m}_{2N}(\beta) : B \to R^{HK}$ be a set of continuous functions such that $\widetilde{m}_{2N}(\beta) \to m_2(\beta)$ uniformly on B as N grows to large, where $m_2(\beta) : B \to R^{HK}$ is a set of continuous functions such that $m_2(\beta_0) = 0$ and $m_2(\beta) \neq 0$ for $\beta \neq \beta_0$. By taking into account a joint case of near exogeneity and weak identification, Equation (3) can be rewritten as

$$\begin{aligned} \widetilde{m}_N(\alpha, \beta) &= \widetilde{m}_N(\alpha_0, \beta_0) + \widetilde{m}_{1N}(\alpha, \beta) + \widetilde{m}_{2N}(\beta) \\ &= C/\sqrt{N} + m_{1N}(\alpha, \beta)/\sqrt{N} + \widetilde{m}_{2N}(\beta) \end{aligned} \tag{7}$$

because of Equation (1). When $C = 0$, we can obtain the result of Stock and Wright (2000), in which case they don't consider the problem of near exogeneity. Now, we can give assumptions that formally define near exogeneity and weak identification.

Assumption 1 The true parameter $\theta_0 = (\alpha_0', \beta_0')'$ is in the interior of the compact space $\Theta = A \times B$, $A \subset R^{m_1}$, $B \subset R^{m_2}$, and $m = m_1 + m_2$. The true parameter θ_0 satisfies the moment conditions defined by Equation (1).

Assumption 2

$$EN^{-1} \sum_{i=1}^{N} \phi_i(\alpha, \beta) = C/\sqrt{N} + m_{1N}(\alpha, \beta)/\sqrt{N} + \widetilde{m}_{2N}(\beta), \text{ where} \tag{8}$$

(2.1) $m_{1N}(\theta) \to m_1(\theta)$ uniformly on Θ, $m_1(\theta_0) = 0$, and $m_1(\theta)$ is continuous in θ and is bounded on Θ;

(2.2) $\widetilde{m}_{2N}(\beta) \to m_2(\beta)$ uniformly on Θ, $m_2(\beta) = 0$ if and only if $\beta = \beta_0$. Define $R(\beta) = \partial m_2(\beta)/\partial\beta'$ which is a $HK \times m_2$ matrix. $R(\beta)$ is continuous in β and $R(\beta_0)$ has a full column rank.

We can apply the above assumptions into the linear simultaneous equations model defined in Chapter 1. In Chapter 1, all parameters in θ are weakly identified. The identity defined by (3) can be rewritten as

$$\begin{aligned} \widetilde{m}_N(\theta) &= \widetilde{m}_N(\theta_0) + [\widetilde{m}_N(\theta) - \widetilde{m}_N(\theta_0)] \\ &= \widetilde{m}_N(\theta_0) + m_{1N}(\theta)/\sqrt{N} \end{aligned} \tag{9}$$

where $\widetilde{m}_N(\theta_0) = EN^{-1}\sum_{i=1}^N \phi_i(\theta_0) = C/\sqrt{N}$ by the near exogeneity in Assumption 2. In the linear simultaneous equations model,

$$EN^{-1}\sum_{i=1}^N \phi_i(\theta) = EN^{-1}\sum_{i=1}^N [Z_i'(y_i - Y_i\theta)] \tag{10}$$

$$= EN^{-1}\sum_{i=1}^N [Z_i'(y_i - Y_i\theta_0) - Z_i'Y_i(\theta - \theta_0)] \tag{11}$$

$$= EN^{-1}\sum_{i=1}^N \{[Z_i'(y_i - Y_i\theta_0)] - [Z_i'Z_i\Pi(\theta - \theta_0)]\} \tag{12}$$

By Equation (2), we obtain

$$EN^{-1}\sum_{i=1}^N [Z_i'(y_i - Y_i\theta_0)] = C/\sqrt{N} \tag{13}$$

Since $\Pi = \Pi_N = C_1/\sqrt{N}$ defined by Assumption ID in Chapter 1, we have

$$\widetilde{m}_N(\theta) = C/\sqrt{N} + m_{1N}(\theta)/\sqrt{N} \tag{14}$$

where $m_{1N}(\theta) = EN^{-1}\sum_{i=1}^N [Z_i'Z_iC_1(\theta - \theta_0)]$. The first term in (14) is due to near exogeneity and the second term is used to define the weak identification of θ.

Next, we consider the GMM estimator that minimizes the objective function $S_N(\theta, \overline{\theta}_N(\theta))$ for $\theta \in \Theta$, where

$$S_N(\theta, \overline{\theta}_N(\theta)) = [N^{-1/2}\sum_{i=1}^N \phi_i(\theta)]' W_N(\overline{\theta}_N(\theta))[N^{-1/2}\sum_{j=1}^N \phi_j(\theta)] \tag{15}$$

where $W_N(\overline{\theta}_N(\theta))$ is a positive definite $HK \times HK$ weighting matrix and bounded in probability. Different GMM estimators depend upon the adoption of different weighting matrix. For a one-step GMM estimator, the weighting matrix is usually an identity matrix so $W_N(\overline{\theta}_N(\theta))$ doesn't depend upon the data and the unknown parameter θ. For a two-step efficient GMM estimator (Hansen, 1982), the weighting matrix is computed by using a one-step GMM estimator. For a continuously updating GMM estimator (Hansen, Heaton and Yaron, 1996), the weighting matrix is changed with each choice of the unknown parameter θ, so $W_N(\overline{\theta}_N(\theta))$ can be written as $W_N(\theta)$. In order to establish the large sample properties of the GMM estimators, we need the uniform convergence of the weighting matrix $W_N(\theta)$. This is also the assumption used by Stock and Wright (2000).

Assumption 3 $W_N(\theta) \xrightarrow{p} W(\theta)$ uniformly on Θ, where $W(\theta)$ is a $HK \times HK$ symmetric positive definite matrix and is continuous in θ.

Next, following Andrews (1994) and Stock and Wright (2000), we define an empirical process $\Psi_N(\theta)$ by

$$\Psi_N(\theta) = N^{-1/2} \sum_{i=1}^{N} [\phi_i(\theta) - E\phi_i(\theta)] \quad \text{for } \theta \in \Theta \tag{16}$$

Note that $\phi_i(\theta) = \phi_i(Y_i, Z_i, \theta)$ where Y_i and Z_i are independent observations. $\phi_i(\theta)$ can be regarded as a class of R^{HK} valued functions defined on Y_i and Z_i indexed by $\theta \in \Theta$. Let "$\Rightarrow$" denote weak convergence of a sequence of empirical processes. By Andrews (1994) and Vaart and Wellner (1996), weak convergence of the empirical process in Equation (16) can be defined as

$$\Psi_N(\theta) \Rightarrow \Psi(\theta) \text{ if } E^* f(\Psi_N(.)) \rightarrow Ef(\Psi(.)) \tag{17}$$

for all bounded, uniformly continuous real functions f on $\mathbf{B}(\mathbf{\Theta})$, where $\mathbf{B}(\mathbf{\Theta})$ is the set of all continuous, bounded functions f: $\Theta \rightarrow R$. Note that "E^*" is the expectation over the empirical process. Let $\Omega(\theta_1, \theta_2) = \lim_{N\rightarrow\infty} E\Psi_N(\theta_1)\Psi_N(\theta_2)'$. The following assumption of weak convergence is mainly based on Pollard (1984, 1990), Andrews (1994) and Vaart and Wellner (1996). It's similar to Assumption A and B used in Stock and Wright (2000).

Assumption 4 $\Psi_N(\theta) \Rightarrow \Psi(\theta)$, where $\Psi(\theta)$ is a Gaussian limit stochastic process on Θ with zero mean and covariance $\Omega(\theta_1, \theta_2)$.

Assumption 4 is a kind of high level assumption which follows from three sufficient conditions (Andrews, 1994): (1) Θ is a totally bounded space; (2) finite dimensional convergence holds: $\forall(\theta_1, ..., \theta_J) \in \Theta$, $(\Psi_N(\theta_1)', ..., \Psi_N(\theta_J)')'$ converges in distribution; (3) $\Psi_N(\theta)$ is stochastic equicontinuity. Condition (1) is satisfied by Assumption 1 that Θ is a compact space. Condition (2) is easily to verified by multivariate central limit theorem. For example, we can use univariate triangular array central limit theorem (Liapunov Theorem, see Davidson, 1994) to obtain the normal limit of the stochastic process $\Psi_N(\theta)$ at $\theta = \theta_0$, say

$$\Psi_N(\theta_0) \xrightarrow{d} N(0, \Omega(\theta_0, \theta_0) \tag{18}$$

by imposing the moment condition such that $E \mid \phi_i(\theta_0) \mid^{2+\delta} < \Delta < \infty$ for some $\delta > 0$. For the finite dimensional convergence, we can assume a similar moment condition which holds uniformly on Θ. Condition (3) stochastic equicontinuity relies on a condition which is referred as entropy condition (Pollard, 1990). By Theorem 1 and 2 in Andrews (1994), $\phi_i(\theta)$ falls into a type II class of functions so that the Pollard's entropy condition follows from the Lipschitz continuity. To be summarized, Assumption 4 follows from the following primitive assumptions.

(i) Θ is a compact parameter space;
(ii) $\phi_i(\theta)$ is independent;
(iii) $E \mid \phi_i(\theta) \mid^{2+\delta} < \Delta < \infty$ uniformly over Θ for some $\delta > 0$;
(vi) Lipschitz in θ: $\mid \phi_i(\theta_1) - \phi_i(\theta_2) \mid \leq B_i(.) \|\theta_1 - \theta_2\| \; \forall \theta_1, \theta_2 \in \Theta$, and $B_i(.)$ satisfies $\lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} E[B_i(.)^{2+\delta}] < \infty$ for some $\delta > 0$.

Assumption (i) implies totally boundedness. Assumptions (ii) and (iii) imply finite dimensional convergence. Assumptions (i) and (vi) imply stochastic equicontinuity. It's very easy to verify that the $\phi_i(\theta)$ defined in the linear simultaneous equations model in Chapter 1 satisfies these assumptions.

3 Estimation: Limiting Results of GMM Estimators

In this section, we derive the asymptotic results of GMM estimators under near exogeneity and weak identification. We firstly derive general limiting results of GMM estimators and then derive limiting results of some specific GMM estimators, such as one-step estimator, two-step efficient estimator and continuously updating estimator. In each case, we examine the limiting results of the weakly identified parameter α and the well identified parameter β.

3.1 General Limiting Results of GMM Estimators

We derive the general asymptotic results of GMM estimators in this subsection. First, we examine the limiting results of the well identified parameter β. The following lemma shows that the GMM estimator $\widehat{\beta}$ is consistent under near exogeneity and the convergence rate is square root of the sample size N.

Lemma 1 $\sqrt{N}(\widehat{\beta} - \beta_0) = O_p(1)$.

All proofs are given in the appendix.

Lemma 1 shows that near exogeneity doesn't affect the convergence of a well identified parameter. Intuitively, the drift term in Equation (1) shrinks toward zero as the sample size N grows to large. We have a similar story in the linear case. In the linear simultaneous equations model defined in Chapter 1, when there only exists the problem of near exogeneity, both the TSLS estimator and the LIML estimator are consistent. However, situations are a little complicated in this chapter. There are two parameters, of which one is weakly identified and the other is well identified. One natural question is whether the weakly identified parameter α affect the limiting results of the well identified parameter β. A joint limiting result of α and β is necessary to answer such a question. The following theorem gives the joint limits of both parameters under near exogeneity and weak identification for a general GMM estimator.

Theorem 1 Suppose that Assumptions 1-4 hold, then

$$(\widehat{\alpha}, \sqrt{N}(\widehat{\beta} - \beta_0)) \xrightarrow{d} (a^*, b^*) \tag{19}$$

where

$$a^* = \arg\min_{\alpha \in A} S^*(\alpha; \overline{\theta}(\alpha, \beta_0)) \tag{20}$$

$$\begin{aligned} b^* &= -[R(\beta_0)'W(\overline{\theta}(a^*, \beta_0)R(\beta_0)]'R(\beta_0)'W(\overline{\theta}(a^*, \beta_0) \\ &\quad \times[\Psi(a^*, \beta_0) + C + m_1(a^*, \beta_0)] \end{aligned} \tag{21}$$

$$\begin{aligned} S^*(\alpha; \overline{\theta}(\alpha, \beta_0)) &= [\Psi(\alpha, \beta_0) + C + m_1(\alpha, \beta_0)]'M(\alpha, \beta_0, \overline{\theta}(\alpha, \beta_0)) \\ &\quad \times[\Psi(\alpha, \beta_0) + C + m_1(\alpha, \beta_0)] \end{aligned} \tag{22}$$

where

$$\begin{aligned} M(\alpha, \beta_0, \overline{\theta}(\alpha, \beta_0)) &= W(\overline{\theta}(\alpha, \beta_0) - W(\overline{\theta}(\alpha, \beta_0)R(\beta_0) \\ &\quad \times[R(\beta_0)'W(\overline{\theta}(\alpha, \beta_0)R(\beta_0)]^{-1} \\ &\quad \times R(\beta_0)'W(\overline{\theta}(\alpha, \beta_0) \end{aligned} \tag{23}$$

The above theorem is similar to Theorem 1 in Caner (2005) and is analogous to Theorem 1 in Stock and Wright (2000) and Theorem 2 in Guggenberger and Smith (2005). We can obtain Stock and Wright's result by setting $C = 0$. It's not surprising that $\widehat{\alpha}$ is not consistent since α is a weakly identified parameter. Like the case of the linear simultaneous equations model, the estimator of the weakly identified parameter convergences to a nonstandard distribution a^*. The joint limits given in the above theorem can explain why the estimator $\widehat{\beta}$ of the well identified parameter also convergence to a nonstandard distribution b^*. The distribution of $\widehat{\beta}$ depends on a^* but we cannot estimate α consistently. When we set $C = 0$ and $\alpha = \alpha_0$, Equation (21) can be simplified as

$$\begin{aligned} b^* &= -[R(\beta_0)'W(\overline{\theta}(\alpha_0, \beta_0)R(\beta_0)]'R(\beta_0)'W(\overline{\theta}(\alpha_0, \beta_0)\Psi(\alpha_0, \beta_0) \\ &\xrightarrow{d} N(0, (R(\beta_0)'\Omega^{-1}(\alpha_0, \beta_0)R(\beta_0)) \end{aligned} \tag{24}$$

since $m_1(\alpha_0, \beta_0) = 0$ by Assumption 2 and $\Psi(\alpha_0, \beta_0) \xrightarrow{d} N(0, \Omega(\alpha_0, \beta_0))$ by triangular array central limit theorem. Near exogeneity doesn't affect the convergence rate of β but it shifts the distribution of the estimator. When the drift term $C \neq 0$, we have

$$b^* \xrightarrow{d} N(C, (R(\beta_0)'\Omega^{-1}(\alpha_0, \beta_0)R(\beta_0)) \tag{25}$$

To the weakly identified parameter α, near exogeneity can enlarge the bias term which is obtained by Stock and Wright (2000).

3.2 Limiting Results for Specific GMM Estimators

We first consider a one-step GMM estimator with an identity weighting matrix. Denote by $(\widehat{\alpha}_1, \widehat{\beta}_1)$ the one-step GMM estimator which minimizes the following objective function

$$S_{1N}(\theta) = [N^{-1/2}\sum_{i=1}^{N}\phi_i(\theta)]'[N^{-1/2}\sum_{j=1}^{N}\phi_j(\theta)]. \tag{26}$$

The following corollary gives the joint limits of $(\widehat{\alpha}_1, \sqrt{N}(\widehat{\beta}_1 - \beta_0))$ under near exogeneity and weak identification.

Corollary 1 Suppose that Assumptions 1, 2, 4 holds, then

$$(\widehat{\alpha}_1, \sqrt{N}(\widehat{\beta}_1 - \beta_0)) \xrightarrow{d} (a_1^*, b_1^*) \tag{27}$$

where

$$a_1^* = \arg\min_{\alpha \in A} S_1^*(\alpha, C) \tag{28}$$

$$b_1^* = -[R(\beta_0)'R(\beta_0)]^{-1}R(\beta_0)'[\Psi(a_1^*, \beta_0) + C + m_1(a_1^*, \beta_0)] \tag{29}$$

$$S_1^*(\alpha, C) = [\Psi(\alpha, \beta_0) + C + m_1(\alpha, \beta_0)]'M_1(\alpha)[\Psi(\alpha, \beta_0) + C + m_1(\alpha, \beta_0)] \tag{30}$$

where

$$M_1(\alpha) = I - R(\beta_0)[R(\beta_0)'R(\beta_0)]^{-1}R(\beta_0)'. \tag{31}$$

The two-step efficient GMM estimator is obtained by using the one-step GMM estimator $(\widehat{\alpha}_1, \widehat{\beta}_1)$ to establish an estimate of the weighting matrix. Denote by $(\widehat{\alpha}_2, \widehat{\beta}_2)$ the two-step efficient GMM estimator which minimizes the following objective function

$$S_{2N}(\theta) = [N^{-1/2}\sum_{i=1}^{N}\phi_i(\theta)]'W_N(\widehat{\alpha}_1, \widehat{\beta}_1)[N^{-1/2}\sum_{j=1}^{N}\phi_j(\theta)] \tag{32}$$

The following corollary establishes the joint limits of $(\widehat{\alpha}_2, \sqrt{N}(\widehat{\beta}_2 - \beta_0))$ under near exogeneity and weak identification.

Corollary 2 Suppose that Assumptions 1-4 hold, then

$$(\widehat{\alpha}_2, \sqrt{N}(\widehat{\beta}_2 - \beta_0)) \xrightarrow{d} (a_2^*, b_2^*) \tag{33}$$

where

$$a_2^* = \arg\min_{\alpha \in A} S_2^*(\alpha, a_1^*, C) \tag{34}$$

$$\begin{aligned} b_2^* &= -[R(\beta_0)'\Omega^{-1}(a_1^*, \beta_0)R(\beta_0)]^{-1}R(\beta_0)'\Omega^{-1}(a_1^*, \beta_0) \\ &\quad \times[\Psi(a_2^*, \beta_0) + C + m_1(a_2^*, \beta_0)] \end{aligned} \tag{35}$$

$$\begin{aligned} S_2^*(\alpha, a_1^*, C) &= [\Psi(\alpha, \beta_0) + C + m_1(\alpha, \beta_0)]'M_1(\alpha, a_1^*) \\ &\quad \times[\Psi(\alpha, \beta_0) + C + m_1(\alpha, \beta_0)] \end{aligned} \tag{36}$$

where

$$\begin{aligned} M_1(\alpha, a_1^*) &= \Omega^{-1}(a_1^*, \beta_0) \\ &\quad -\Omega^{-1}(a_1^*, \beta_0)R(\beta_0)[R(\beta_0)'\Omega^{-1}(a_1^*, \beta_0)R(\beta_0)]^{-1} \\ &\quad \times R(\beta_0)'\Omega^{-1}(a_1^*, \beta_0) \end{aligned} \tag{37}$$

In the two-step efficient GMM estimator, the weighting matrix $W_N(\widehat{\alpha}_1, \widehat{\beta}_1)$ is based on the one-step GMM estimator $\widehat{\alpha}_1$ and $\widehat{\beta}_1$, and so the weighting matrix converge to $\Omega^{-1}(a_1^*, \beta_0)$ in the limiting concentrated objective function $S_2^*(\alpha, a_1^*, C)$.

In the case of the linear simultaneous equations model defined in Chapter 1, when the conditional homoskedasticity of the errors is assumed, the objective function of the two-step efficient GMM estimator can be rewritten as

$$S_{2N}(\theta) = (y - Y\theta)'P_Z(y - Y\theta)/\widehat{\Sigma}_{hh}(\widehat{\theta}_1) \tag{38}$$

where

$$\begin{aligned} \widehat{\Sigma}_{hh}(\widehat{\theta}_1) &= N^{-1}\sum_{i=1}^{N} E\{[h_i(\widehat{\theta}_1) - Eh_i(\widehat{\theta}_1)] \\ &\quad \times[h_i(\widehat{\theta}_1) - Eh_i(\widehat{\theta}_1)]'\} \end{aligned} \tag{39}$$

and

$$P_Z = Z(Z'Z)^{-1}Z'. \tag{40}$$

In the linear simultaneous equations model, $h_i(\theta) = y_i - Y_i\theta$ and all parameters in θ are weakly identified. Since θ is quadratic in $S_{2N}(\theta)$, we can derive an analytical solution from Equation (38), which yields

$$\widehat{\theta} = (Y'P_ZY)^{-1}(Y'P_Zy) \tag{41}$$

We know this is just the TSLS estimator.

The continuously updating estimator is obtained when the weighting matrix is continuously updated at the parameter value θ. Denote by $(\widehat{\alpha}_c, \widehat{\beta}_c)$ the continuously updating estimator that minimizes the following objective function

$$S_{cN}(\theta) = [N^{-1/2}\sum_{i=1}^{N}\phi_i(\theta)]'W_N(\theta)[N^{-1/2}\sum_{j=1}^{N}\phi_j(\theta)] \tag{42}$$

The following corollary establishes the joint limits of the continuously updating estimator $(\widehat{\alpha}_c, \widehat{\beta}_c)$ under near exogeneity and weak identification.

Corollary 3 Suppose that Assumptions 1-4 hold, then

$$(\widehat{\alpha}_c, \sqrt{N}(\widehat{\beta}_c - \beta_0)) \xrightarrow{d} (a_c^*, b_c^*) \tag{43}$$

where

$$a_c^* = \arg\min_{\alpha\in A} S_c^*(\alpha, C) \tag{44}$$

$$\begin{aligned} b_c^* &= -[R(\beta_0)'\Omega^{-1}(a_c^*,\beta_0)R(\beta_0)]^{-1}R(\beta_0)'\Omega^{-1}(a_c^*,\beta_0) \\ &\quad \times[\Psi(a_c^*,\beta_0) + C + m_1(a_c^*,\beta_0)] \end{aligned} \tag{45}$$

$$\begin{aligned} S_c^*(\alpha, C) &= [\Psi(\alpha,\beta_0) + C + m_1(\alpha,\beta_0)]'\Omega^{-1}(a_c^*,\beta_0) \\ &\quad \times\{I - R(\beta_0)[R(\beta_0)'\Omega^{-1}(a_c^*,\beta_0)R(\beta_0)]^{-1}R(\beta_0)'\Omega^{-1}(a_c^*,\beta_0)\} \\ &\quad \times[\Psi(\alpha,\beta_0) + C + m_1(\alpha,\beta_0)]. \end{aligned} \tag{46}$$

Consider a special case of Corollary 3: the linear simultaneous equations model with all weakly identified parameters and conditional homoskedasticity defined in Chapter 1. Since

$$\phi_i(\theta) = Z_i'(y_i - Y_i\theta) \tag{47}$$

and

$$W_N(\theta) = [N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_i(\theta)\phi_j(\theta)']^{-1}, \tag{48}$$

the objective function $S_{cN}(\theta)$ defined in (42) can be simplified as

$$\begin{aligned} S_{cN}(\theta) &= [N^{-1/2}\sum_{i=1}^{N}\phi_i(\theta)]'[N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_i(\theta)\phi_j(\theta)']^{-1} \\ &\quad \times[N^{-1/2}\sum_{j=1}^{N}\phi_j(\theta)] \\ &= (y - Y\theta)'Z(Z'Z)^{-1}Z'(y - Y\theta)/u(\theta)'u(\theta) \\ &= N[1 + \kappa^{-1}(\theta)]^{-1} \end{aligned} \tag{49}$$

where

$$u(\theta) = y - Y\theta$$

$$\kappa(\theta) = (y - Y\theta)' P_z (y - Y\theta)/(y - Y\theta)' M_z (y - Y\theta) \tag{50}$$

and

$$M_Z = I - P_Z. \tag{51}$$

Note that Equation (49) is obtained since we have

$$T^{-1}(y - Y\theta)' M_z (y - Y\theta) \xrightarrow{p} u(\theta)' u(\theta) \tag{52}$$

The continuously updating estimator in the linear case is identical to minimize $\kappa(\theta)$, which is just the LIML estimator; see Davidson and MacKinnon (1993).

4 Inference with Near Exogeneity and Weak Identification

In a GMM framework with instrumental variables, we want to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ under near exogeneity and weak identification. Staiger and Wright (2000) examined several conventional test statistics under weak identification, such as Wald statistic and likelihood ratio statistic. These conventional test statistics do not work in general under weak identification. The exogeneity tests of instruments, like J-test (Hansen, 1982; Newey, 1985), cannot be valid in general under weak identification either.

In this section, we firstly consider some robust test statistics which have been recently developed against weak identification in the literature, and then examine their performance under near exogeneity.

We first consider an Anderson-Rubin type test proposed by Stock and Wright (2000). The test is given by

$$S_N(\theta_0; \theta_0) = [N^{-1/2} \sum_{i=1}^{N} \phi_i(\theta_0)]' W_N(\theta_0) [N^{-1/2} \sum_{j=1}^{N} \phi_j(\theta_0)]. \tag{53}$$

Since the moment function is generally nonlinear, it's easier to work on the objective function rather than on the estimator as we did in the case of the linear simultaneous equations model. The Anderson-Rubin type test given by Equation (53) is just the objective function $S_{cN}(\theta)$ of the continuously updating estimator when $\theta = \theta_0$. Since it utilizes the objective function $S_{cN}(.)$, it was called "S statistic" by Stock and Wright (2000). The S statistic is robust to weak identification because the test itself is asymptotically pivotal and convergence in distribution to a chi-square distribution under the null hypothesis. Note that we cannot establish an Anderson-Rubin type test based on the objective function of the two-step GMM estimator. The objective function of the two-step GMM estimator is not asymptotically pivotal because the weighting matrix in the objective function is derived through the one-step estimator, which is not consistent under weak identification.

To examine the asymptotic property of the S statistic under near exogeneity, we can work under a much weaker assumption described in Equation (18) than Assumption 4. The following theorem summarizes the asymptotic result of the S statistic under near exogeneity.

Theorem 2 Suppose Assumptions 1-3 and Equation (18) hold under the null hypothesis of $\theta = \theta_0$, then

$$S_N(\theta_0; \theta_0) \xrightarrow{d} \chi^2_{HK}(C'\Omega^{-1}(\theta_0; \theta_0)C) \tag{54}$$

where $\chi^2_{HK}(C'\Omega^{-1}(\theta_0; \theta_0)C)$ is a noncentral chi-square distribution with noncentral parameter $C'\Omega^{-1}(\theta_0; \theta_0)C$ and the degree of freedom HK.

Theorem 2 shows that the S statistic is not asymptotically pivotal under near exogeneity. The limit of the test statistic depends on the nuisance unknown parameter C which comes from near exogeneity. We obtain a chi-square distribution with degree of freedom HK when we set $C = 0$. It leads to a size distortion under near exogeneity when we use critical values from the chi-square distribution. In empirical practice, it'll overreject a true hypothesis.

Kleibergen (2005) proposes a GMM version K statistic. The K statistic is also based on the objective function of the continuously updating GMM estimator. To establish the limits of the K statistic, we need two more assumptions. Denote by $q_i(\theta_0)$ the first order derivative of $\phi_i(\theta)$ with respect to θ which is evaluated at $\theta = \theta_0$, and let

$$J_\theta(\theta_0) = \lim_{N\to\infty} E[N^{-1} \sum_{i=1}^{N} q_i(\theta_0)] \tag{55}$$

Assumption 5 Let

$$q_{i,j}(\theta_0) = \partial\phi_i(\theta)/\partial\theta'_j \mid_{\theta=\theta_0} \qquad j = 1, 2, ..., m. \tag{56}$$

and $q_i(\theta_0) = (q'_{i,1}(\theta_0), q'_{i,2}(\theta_0), ..., q'_{i,m}(\theta_0))'$. We assume the following limits hold jointly

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \begin{pmatrix} \phi_i(\theta_0) - E[\phi_i(\theta_0)] \\ q_i(\theta_0) - E[q_i(\theta_0)] \end{pmatrix} \xrightarrow{d} \left(\Psi'_\phi, \Psi'_q\right)' \tag{57}$$

where

$$\begin{pmatrix} \Psi_\phi \\ \Psi_q \end{pmatrix} \sim N(0, V(\theta)) \tag{58}$$

and $V(\theta)$ is a positive semi-definite symmetric $(HK + mHK) \times (HK + mHK)$ matrix

$$V(\theta) = \begin{pmatrix} V_{\phi\phi} & V_{\phi q} \\ V_{q\phi} & V_{qq} \end{pmatrix}$$

and

$$V(\theta) = \lim_{N\to\infty} EN^{-1} \sum_{i=1}^{N}\sum_{l=1}^{N} \begin{pmatrix} \phi_i(\theta_0) - E[\phi_i(\theta_0)] \\ q_i(\theta_0) - E[q_i(\theta_0)] \end{pmatrix} \begin{pmatrix} \phi_l(\theta_0) - E[\phi_i(\theta_0)] \\ q_l(\theta_0) - E[q_i(\theta_0)] \end{pmatrix}'. \quad (59)$$

Assumption 6 Assume that the estimator of the covariance matrix $V(\theta_0)$ and the estimator of the derivative of $W(\theta_0) = V_{\phi\phi}^{-1}(\theta_0)$ with respect to θ have the limits that hold jointly

$$\widehat{V}(\theta_0) \xrightarrow{p} V(\theta_0) \quad (60)$$

and

$$\partial vec(\widehat{V}_{\phi\phi}(\theta_0))/\partial\theta' \xrightarrow{p} \partial vec(V_{\phi\phi}(\theta_0))/\partial\theta' \quad (61)$$

where

$$V_{\phi\phi}(\theta_0) = \lim_{N\to\infty} E\{N^{-1}\sum_{i=1}^{N}\sum_{l=1}^{N}(\phi_i(\theta_0) - E[\phi_i(\theta_0)])(\phi_i(\theta_0) - E[\phi_i(\theta_0)])'\}. \quad (62)$$

The K statistic is based on the first order derivative of Equation (53) with respect to θ. The K statistic is given by

$$\begin{aligned} K(\theta_0) &= \frac{1}{4N}(\partial S_N(\theta_0;\theta_0)/\partial\theta)[\widehat{D}_N(\theta_0)'\widehat{V}_{\phi\phi}^{-1}(\theta_0)\widehat{D}_N(\theta_0)]^{-1} \\ &\quad \times(\partial S_N(\theta_0;\theta_0)/\partial\theta)' \end{aligned} \quad (63)$$

where

$$\frac{1}{2}\partial S_N(\theta_0;\theta_0)/\partial\theta = \phi_N(\theta_0)'\widehat{V}_{\phi\phi}^{-1}(\theta_0)\widehat{D}_N(\theta_0) \quad (64)$$

$$\begin{aligned} \widehat{D}_N(\theta_0) &= [q_{N,1}(\theta_0) - \widehat{V}_{q\phi,1}(\theta_0)\widehat{V}_{\phi\phi}^{-1}(\theta_0)\phi_N(\theta_0) \ ... \\ &\quad ...q_{N,m}(\theta_0) - \widehat{V}_{q\phi,m}(\theta_0)\widehat{V}_{\phi\phi}^{-1}(\theta_0)\phi_N(\theta_0)] \end{aligned} \quad (65)$$

and $\widehat{V}_{q\phi}(\theta_0) = (\widehat{V}_{q\phi,1}(\theta_0)', \widehat{V}_{q\phi,2}(\theta_0)', ..., \widehat{V}_{q\phi,m}(\theta_0))'$.

Note that $\widehat{D}_N(\theta_0)$ is a consistent estimator of $J_\theta(\theta_0)$ even in the case of weak identification. Either under strong identification or weak identification, the K statistic is an asymptotically pivotal distribution conditional on $\widehat{D}_N(\theta_0)$. Because of the asymptotic independence between $\widehat{D}_N(\theta_0)$ and Ψ_ϕ, the K statistic converges unconditionally to a chi-square distribution with degree of freedom m under weak identification. The following theorem summarizes the asymptotic results of the K statistic under near exogeneity and weak identification.

Theorem 3 Suppose that Assumptions 1, 2, 5 and 6 hold under the null hypothesis of $\theta = \theta_0$, then

$$K(\theta_0) \xrightarrow{d} (\xi + \Xi(C))'(\xi + \Xi(C))$$

where

$$\xi \sim N(0, I_{HK}) \tag{66}$$

$$\Xi(C) = [D'V_{\phi\phi}^{-1}(\theta_0)D]^{-1/2}D'V_{\phi\phi}^{-1}(\theta_0)C \tag{67}$$

and D *is the limit of* $\gamma(N)\widehat{D}_N(\theta_0)$*, and further* D *varies when*

(i) θ *is well identified,* $D \xrightarrow{d} C_q$

(ii) θ *is weakly identified,* $D \xrightarrow{d} C_q + \Psi_{q.\phi}$

(iii) θ *is nonidentified,* $D \xrightarrow{d} \Psi_{q.\phi}$

$C_q = J_\theta(\theta_0)$ *which has a fixed full rank value, and* $\Psi_{q.\phi}$ *is a limiting distribution such that*

$$N^{-1/2}vec[\widehat{D}_N(\theta_0) - J_\theta(\theta_0)] \xrightarrow{d} \Psi_{q.\phi}. \tag{68}$$

Theorem 3 shows that the K statistic converges to a nonstandard distribution under near exogeneity. The nonstandard distribution is a quadratic form of the sum of a standard normal variable ξ and the drift term $\Xi(C)$ which comes from near exogeneity. When the identification condition varies, we obtain different limits of $\Xi(C)$. We can obtain a chi-square distribution with degree of freedom m when $C = 0$. So Theorem 3 provides a general result. Theorem 3 also implies that inference based on the critical value from chi-square distribution can result in a large size distortion.

5 Conclusions

This chapter studies the asymptotic properties of estimation and inference under near exogeneity and weak identification in a GMM framework with instrumental variables. We derive the limits of the one-step GMM estimator, the efficient two-step GMM estimator and the continuously updating estimator under near exogeneity and weak identification. We consider a mixed case where some parameters are weakly identified and others are well identified. The GMM estimators of the well identified parameters are consistent but converge to a nonstandard distribution. In all cases, near exogeneity can bring a relatively large asymptotic bias for GMM estimators compared to the case where only weak identification occurs. We show that the Anderson-Rubin type S statistic and the Kleibergen type K statistic are no longer asymptotically pivotal under near exogeneity. It leads to a serious size distortion when using critical values from chi-square distribution.

Appendix

Proof of Lemma 1 First, we show that β is consistent. Consider the objective function $S_N(\theta, \overline{\theta}_N(\theta))$ given by (15). The first term can be rewritten as

$$N^{-1/2}\sum_{i=1}^{N}\phi_i(\theta) = N^{-1/2}\sum_{i=1}^{N}[\phi_i(\theta) - E\phi_i(\theta)] + N^{-1/2}\sum_{i=1}^{N}E\phi_i(\theta). \tag{69}$$

The first term converges to $\Psi(\theta)$ by Assumption 4 and the second term can be rewritten as

$$\begin{aligned} N^{-1/2}\sum_{i=1}^{N}E\phi_i(\theta) &= \sqrt{N}EN^{-1}\sum_{i=1}^{N}\phi_i(\theta) \qquad (70)\\ &\rightarrow C + m_1(\alpha,\beta) + \sqrt{N}m_2(\beta) \end{aligned}$$

by Assumption 2. By Assumption 3, we have

$$\begin{aligned} &S_N(\theta, \overline{\theta}_N(\theta)) \xrightarrow{p} [\Psi(\theta) + C + m_1(\alpha,\beta) + \sqrt{N}m_2(\beta)]'W(\overline{\theta}(\theta)) \qquad (71)\\ &\times[\Psi(\theta) + C + m_1(\alpha,\beta) + \sqrt{N}m_2(\beta)]. \end{aligned}$$

Scale Equation (71) by N^{-1}, we obtain

$$N^{-1}S_N(\theta, \overline{\theta}_N(\theta)) \xrightarrow{p} m_2(\beta)'W(\overline{\theta}(\theta))m_2(\beta) \tag{72}$$

uniformly in β. Since $W(\overline{\theta}(\theta))$ is positive definite by Assumption 3 and $m_2(\beta) = 0$ if and only if $\beta = \beta_0$, the consistency of β follows by the continuity of the arg min operator. The rate of convergence follows from the proof of Lemma A1 in Stock and Wright(2000). *Q.E.D.*

Proof of Theorem 1 To derive the limiting results in the theorem, we work on the objective function $S_N(\alpha, \beta, \overline{\theta}_N(\theta))$ directly. First, we define

$$b = \sqrt{N}(\beta - \beta_0). \tag{73}$$

By Lemma 1, we know that $b = O_p(1)$. The objective function then can be written as

$$\begin{aligned} S_N(\alpha,\beta,\overline{\theta}_N(\theta)) &= S_N(\alpha, \beta_0 + b/\sqrt{N}, \overline{\theta}_N(\theta)) \qquad (74)\\ &= [N^{-1/2}\sum_{i=1}^{N}\phi_i(\theta)]'W_N(\overline{\theta}_N(\theta))[N^{-1/2}\sum_{j=1}^{N}\phi_j(\theta)]. \end{aligned}$$

The first and last terms in Equation (74) can be written as

$$\begin{aligned} &N^{-1/2}\sum_{i=1}^{N}\phi_i(\alpha, \beta_0 + b/\sqrt{N}) \qquad (75)\\ = \;&N^{-1/2}\sum_{i=1}^{N}[\phi_i(\alpha,\beta_0 + b/\sqrt{N}) - E\phi_i(\alpha,\beta_0 + b/\sqrt{N})] + N^{-1/2}\sum_{i=1}^{N}E\phi_i(\alpha,\beta_0 + b/\sqrt{N}). \end{aligned}$$

By Assumption 4 and Lemma 1, we have

$$N^{-1/2}\sum_{i=1}^{N}[\phi_i(\alpha,\beta_0+b/\sqrt{N})-E\phi_i(\alpha,\beta_0+b/\sqrt{N})]\Rightarrow\Psi(\alpha,\beta_0).$$

The second term in Equation (75) can be written as

$$\begin{aligned}N^{-1/2}\sum_{i=1}^{N}E\phi_i(\alpha,\beta_0+b/\sqrt{N}) &= \sqrt{N}EN^{-1}\sum_{i=1}^{N}\phi_i(\alpha,\beta_0+b/\sqrt{N}) \qquad (76)\\ &= C+m_{1N}(\alpha,\beta_0+b/\sqrt{N})+\sqrt{N}m_{2N}(\beta_0+b/\sqrt{N})\end{aligned}$$

which follows from Assumption 2. Note that $m_{1N}(\theta)\to m_1(\theta)$ uniformly in θ and by Lemma 1, we have

$$m_{1N}(\alpha,\beta_0+b/\sqrt{N})\xrightarrow{p}m_1(\alpha,\beta_0).$$

We apply the mean value theorem to the last term in Equation (76). We can obtain

$$\sqrt{N}m_{2N}(\beta_0+b/\sqrt{N})=\sqrt{N}m_{2N}(\beta_0)+R(\widetilde{\beta})b \qquad (77)$$

where $\widetilde{\beta}\in[\beta_0,\beta_0+b/\sqrt{N}]$ and $R(\beta)=\partial m_2(\beta)/\partial\beta'$ which is defined in Assumption 2. By Assumption 2, $m_{2N}(\beta_0)\to m_2(\beta_0)=0$ and $\widetilde{\beta}\xrightarrow{p}\beta$ by Lemma 1. So we have

$$\sqrt{N}m_{2N}(\beta_0+b/\sqrt{N})\to R(\beta_0)b. \qquad (78)$$

By Assumption 3, we have

$$W_N(\overline{\theta}_N(\theta))\xrightarrow{p}W(\overline{\theta}(\alpha,\beta_0)).$$

So the objective function has the following limits

$$\begin{aligned}S_N(\alpha,\beta,\overline{\theta}_N(\theta)) \;\Rightarrow\; &[\Psi(\alpha,\beta_0)+C+m_1(\alpha,\beta_0)+R(\beta_0)b]' \qquad (79)\\ &\times W(\overline{\theta}(\alpha,\beta_0))[\Psi(\alpha,\beta_0)+C+m_1(\alpha,\beta_0)+R(\beta_0)b].\end{aligned}$$

Next, we fix α in Equation (79) and differentiate it with respect to b. By solving the first order condition, we denote the solution by b^*,

$$\begin{aligned}b^*(\alpha) &= -[R(\beta_0)'W(\overline{\theta}(\alpha,\beta_0))R(\beta_0)]^{-1}R(\beta_0)'W(\overline{\theta}(\alpha,\beta_0)) \qquad (80)\\ &\quad\times[\Psi(\alpha,\beta_0)+C+m_1(\alpha,\beta_0)]\end{aligned}$$

Plug $b^*(\alpha)$ into Equation (79) to yield the concentrated limiting objective function $S^*(\alpha;\overline{\theta}(\alpha,\beta_0))$. To see this, note that

$$\begin{aligned}R(\beta_0)b^* &= -R(\beta_0)[R(\beta_0)'W(\overline{\theta}(\alpha,\beta_0))R(\beta_0)]^{-1}R(\beta_0)'W(\overline{\theta}(\alpha,\beta_0)) \qquad (81)\\ &\quad\times[\Psi(\alpha,\beta_0)+C+m_1(\alpha,\beta_0)].\end{aligned}$$

So we have

$$\begin{aligned}
& \Psi(\alpha,\beta_0)+C+m_1(\alpha,\beta_0)+R(\beta_0)b^* \qquad (82)\\
= & [I-R(\beta_0)(R(\beta_0)'W(\overline{\theta}(\alpha,\beta_0))R(\beta_0))^{-1}R(\beta_0)'W(\overline{\theta}(\alpha,\beta_0))]\\
& \times[\Psi(\alpha,\beta_0)+C+m_1(\alpha,\beta_0)].
\end{aligned}$$

Plug Equation (82) into Equation (79),

$$\begin{aligned}
& S^*(\alpha;\overline{\theta}(\alpha,\beta_0)) \qquad (83)\\
= & [\Psi(\alpha,\beta_0)+C+m_1(\alpha,\beta_0)]'\\
& \times[I-R(\beta_0)(R(\beta_0)'W(\overline{\theta}(\alpha,\beta_0))R(\beta_0))^{-1}R(\beta_0)'W(\overline{\theta}(\alpha,\beta_0))]'\\
& \times W(\overline{\theta}(\alpha,\beta_0))\\
& \times[I-R(\beta_0)(R(\beta_0)'W(\overline{\theta}(\alpha,\beta_0))R(\beta_0))^{-1}R(\beta_0)'W(\overline{\theta}(\alpha,\beta_0))]\\
& \times[\Psi(\alpha,\beta_0)+C+m_1(\alpha,\beta_0)].
\end{aligned}$$

Note that

$$\begin{aligned}
& [I-R(R'WR)^{-1}R'W]'W[I-R(R'WR)^{-1}R'W]\\
= & [I-R(R'WR)^{-1}R'W]'[W-WR(R'WR)^{-1}R'W]\\
= & [I-R(R'WR)^{-1}R'W]'[I-WR(R'WR)^{-1}R']W\\
= & [I-WR(R'WR)^{-1}R']W\\
= & M(\alpha,\beta_0,\overline{\theta}(\alpha,\beta_0)).
\end{aligned}$$

So we obtain that

$$\begin{aligned}
S^*(\alpha;\overline{\theta}(\alpha,\beta_0)) = & [\Psi(\alpha,\beta_0)+C+m_1(\alpha,\beta_0)]' \qquad (84)\\
& \times M(\alpha,\beta_0,\overline{\theta}(\alpha,\beta_0))[\Psi(\alpha,\beta_0)+C+m_1(\alpha,\beta_0)].
\end{aligned}$$

and $\alpha^*=\arg\min_{\alpha\in A}S^*(\alpha;\overline{\theta}(\alpha,\beta_0))$. Substituting α^* into $b^*(\alpha)$ in Equation (80), we can obtain $b^*(\alpha^*)$ defined in the theorem.

Since arg min is a continuous mapping and α^* is a unique minimum over A, by Theorem 3.2.2 of Vaart and Wellner (1996), it follows that $(\widehat{\alpha},\sqrt{N}(\widehat{\beta}-\beta_0))\xrightarrow{d}(a^*,b^*)$. *Q.E.D.*

Proof of Corollary 1 The result in the corollary follows by Theorem 1 when we replace the general objective function $S_N(\alpha,\beta,\overline{\theta}_N(\theta))$ by the one-step objective function $S_{1N}(\theta)$ defined in (26). *Q.E.D.*

Proof of Corollary 2 The two-step efficient GMM estimator depends on an estimate of the weighting matrix which utilizes the first-step GMM estimator. By Assumption 3, Lemma 1, and the definition of the two-step efficient GMM estimator, we have

$$W_N(\widehat{\alpha}_1,\widehat{\beta}_1)\xrightarrow{p}\Omega^{-1}(a_1^*,\beta_0). \qquad (85)$$

Following Theorem 1 by replacing the general objective function $S_N(\alpha,\beta,\overline{\theta}_N(\theta))$ by the two-step objective function$S_{2N}(\theta)$ defined in (32), we can obtain the results in the corollary. Note that in this case the b_2^* depends on both the one-step estimator a_1^* and the two-step estimator a_2^*. *Q.E.D.*

Proof of Corollary 3 The continuously updating estimator depends on a weighting matrix which is continuously updated by the value of the estimator. But, we can simplify the limiting weighting matrix by Lemma 1 and Assumption 3,

$$\begin{aligned} W_N(\alpha, \beta) &= W_N(\alpha, \beta_0 + b/\sqrt{N}) \\ &\xrightarrow{p} \Omega^{-1}(a, \beta_0). \end{aligned} \tag{86}$$

The limiting weighting matrix doesn't depend on b. Then we can follow Theorem 1 by replacing the general objective function $S_N(\alpha, \beta, \overline{\theta}_N(\theta))$ by the continuously updating objective function $S_{cN}(\theta)$ defined in (42). *Q.E.D.*

Proof of Theorem 2 By Equation (53), we have

$$S_N(\theta_0; \theta_0) = [N^{-1/2} \sum_{i=1}^{N} \phi_i(\theta_0)]' W_N(\theta_0) [N^{-1/2} \sum_{j=1}^{N} \phi_j(\theta_0)]$$

The first and the last terms can be rewritten as

$$\begin{aligned} N^{-1/2} \sum_{i=1}^{N} \phi_i(\theta_0) &= N^{-1/2} \sum_{i=1}^{N} [\phi_i(\theta) - E\phi_i(\theta)] + \sqrt{N} E N^{-1} \sum_{i=1}^{N} \phi_i(\theta) \\ &\Rightarrow \Psi(\theta_0) + C + m_1(\theta_0) + \sqrt{N} m_2(\beta_0) \end{aligned}$$

by Assumptions 2 and 4. Since $m_1(\theta_0) = 0$ and $m_2(\beta_0) = 0$ from Assumption 2, we have

$$N^{-1/2} \sum_{i=1}^{N} \phi_i(\theta_0) \xrightarrow{d} \varrho = N(C, \Omega(\theta_0, \theta_0)). \tag{87}$$

By Assumption 3, we have

$$W_N(\theta_0) \xrightarrow{p} \Omega^{-1}(\theta_0, \theta_0) \tag{88}$$

So we obtain that

$$\begin{aligned} & S_N(\theta_0; \theta_0) \\ & \xrightarrow{d} \varrho' \Omega^{-1}(\theta_0, \theta_0) \varrho \\ & \xrightarrow{d} \chi^2_{HK}(C' \Omega^{-1}(\theta_0, \theta_0) C). \qquad Q.E.D. \end{aligned}$$

Proof of Theorem 3 We follow Kleibergen's (2005) idea to construct two asymptotically independent variables. By Assumption 5, we have

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \begin{pmatrix} \phi_i(\theta_0) - E[\phi_i(\theta_0)] \\ q_i(\theta_0) - E[q_i(\theta_0)] \end{pmatrix} \xrightarrow{d} \left(\Psi'_\phi, \Psi'_q \right)'$$

where

$$\begin{pmatrix} \Psi_\phi \\ \Psi_q \end{pmatrix} \sim N(0, V(\theta)).$$

Pre-multiplying it by

$$\left(\begin{array}{cc} I_{HK} & 0 \\ -\widehat{V}_{q\phi}(\theta_0)\widehat{V}_{\phi\phi}(\theta_0)^{-1} & I_{mHK} \end{array} \right), \tag{89}$$

and by Assumption 6, we have

$$\begin{aligned} & \left(\begin{array}{cc} I_{HK} & 0 \\ -\widehat{V}_{q\phi}(\theta_0)\widehat{V}_{\phi\phi}(\theta_0)^{-1} & I_{mHK} \end{array} \right) \\ & \xrightarrow{p} \left(\begin{array}{cc} I_{HK} & 0 \\ -V_{q\phi}(\theta_0)V_{\phi\phi}(\theta_0)^{-1} & I_{mHK} \end{array} \right). \end{aligned}$$

Let

$$\overline{\phi}_N(\theta_0) = \sum_{i=1}^{N} \{\phi_i(\theta_0) - E[\phi_i(\theta_0)]\}$$

and

$$\overline{q}_N(\theta_0) = \sum_{i=1}^{N} \{q_i(\theta_0) - E[q_i(\theta_0)]\}.$$

Then, we can obtain that

$$\begin{aligned} & \sqrt{N} \left(\begin{array}{cc} I_{HK} & 0 \\ -\widehat{V}_{q\phi}(\theta_0)\widehat{V}_{\phi\phi}(\theta_0)^{-1} & I_{mHK} \end{array} \right) \begin{pmatrix} N^{-1}\overline{\phi}_N(\theta_0) \\ N^{-1}\overline{q}_N(\theta_0) \end{pmatrix} \qquad (90) \\ = & \sqrt{N} \begin{pmatrix} N^{-1}\overline{\phi}_N(\theta_0) \\ N^{-1}\overline{q}_N(\theta_0) - N^{-1}\widehat{V}_{q\phi}(\theta_0)\widehat{V}_{\phi\phi}(\theta_0)^{-1}\overline{\phi}_N(\theta_0) \end{pmatrix} \\ & \xrightarrow{d} \begin{pmatrix} \Psi_\phi \\ \Psi_{q.\phi} \end{pmatrix} \end{aligned}$$

where

$$\Psi_{q.\phi} = \Psi_q - V_{q\phi}(\theta_0)V_{\phi\phi}(\theta_0)^{-1}\Psi_\phi$$

and

$$\begin{pmatrix} \Psi_\phi \\ \Psi_{q.\phi} \end{pmatrix} \sim N(0, (\begin{array}{cc} V_{\phi\phi}(\theta_0) & 0 \\ 0 & V_{qq.\phi}(\theta_0) \end{array})) \tag{91}$$

Note that

$$V_{qq.\phi}(\theta_0) = V_{qq}(\theta_0) - V_{q\phi}(\theta_0)V_{\phi\phi}(\theta_0)^{-1}V_{\phi q}(\theta_0)$$

So $(\Psi'_\phi, \Psi'_{q.\phi})'$ has a joint normal distribution with zero correlation which means the asymptotic independence between Ψ_ϕ and $\Psi_{q.\phi}$.

Next, note that

$$\begin{aligned} & N^{-1}\overline{q}_N(\theta_0) - N^{-1}\widehat{V}_{q\phi}(\theta_0)\widehat{V}_{\phi\phi}(\theta_0)^{-1}\overline{\phi}_N(\theta_0) \\ = & [N^{-1}q_N(\theta_0) - N^{-1}\widehat{V}_{q\phi}(\theta_0)\widehat{V}_{\phi\phi}(\theta_0)^{-1}\overline{\phi}_N(\theta_0)] - EN^{-1}q_N(\theta_0) \\ = & N^{-1}\widehat{D}_N(\theta_0) - J_\theta(\theta_0). \end{aligned}$$

So we have

$$\sqrt{N}\begin{pmatrix} N^{-1}\overline{\phi}_N(\theta_0) \\ vec(N^{-1}\widehat{D}_N(\theta_0) - J_\theta(\theta_0)) \end{pmatrix}$$
$$\xrightarrow{d} \begin{pmatrix} \Psi_\phi \\ \Psi_{q.\phi} \end{pmatrix}.$$

Now, consider the K statistic given by (63),

$$\begin{aligned} K(\theta_0) &= \frac{1}{4N}(\partial S_N(\theta_0;\theta_0)/\partial\theta)[\widehat{D}_N(\theta_0)'\widehat{V}_{\phi\phi}^{-1}(\theta_0)\widehat{D}_N(\theta_0)]^{-1} \\ &\quad \times \times (\partial S_N(\theta_0;\theta_0)/\partial\theta)' \\ &= N^{-1/2}\phi_N(\theta_0)'\widehat{V}_{\phi\phi}^{-1}(\theta_0)\widehat{D}_N(\theta_0)[\widehat{D}_N(\theta_0)'\widehat{V}_{\phi\phi}^{-1}(\theta_0)\widehat{D}_N(\theta_0)]^{-1} \\ &\quad \times \widehat{D}_N(\theta_0)'\widehat{V}_{\phi\phi}^{-1}(\theta_0)N^{-1/2}\phi_N(\theta_0). \end{aligned}$$

Let

$$\widehat{\xi} = [\widehat{D}_N(\theta_0)'\widehat{V}_{\phi\phi}^{-1}(\theta_0)\widehat{D}_N(\theta_0)]^{-1/2}\widehat{D}_N(\theta_0)'\widehat{V}_{\phi\phi}^{-1}(\theta_0)N^{-1/2}\overline{\phi}_N(\theta_0). \quad (92)$$

and

$$\begin{aligned} \widehat{\Xi} &= [\widehat{D}_N(\theta_0)'\widehat{V}_{\phi\phi}^{-1}(\theta_0)\widehat{D}_N(\theta_0)]^{-1/2}\widehat{D}_N(\theta_0)'\widehat{V}_{\phi\phi}^{-1}(\theta_0) \\ &\quad \times\sqrt{N}EN^{-1}\sum_{i=1}^{N}\phi_i(\theta_0). \end{aligned} \quad (93)$$

By Assumption 2 and Assumption 4, we have

$$\widehat{\xi} \Rightarrow \xi \sim N(0, I_{HK})$$

and

$$\widehat{\Xi} \xrightarrow{p} \Xi[C]$$

where

$$\Xi[C] = [D'V_{\phi\phi}^{-1}(\theta_0)D]^{-1/2}D'V_{\phi\phi}^{-1}(\theta_0)C$$

which is defined by (67) and $\gamma(N)\widehat{D}_N(\theta_0) \xrightarrow{d} D$.

When θ is well identified, $J_\theta(\theta_0)$ has full rank. We set $\gamma(N) = 1/N$, then

$$\begin{aligned} N^{-1}\widehat{D}_N(\theta_0) &= \frac{1}{\sqrt{N}}\{\sqrt{N}[N^{-1}\widehat{D}_N(\theta_0) - J_\theta(\theta_0)]\} + J_\theta(\theta_0) \\ &\xrightarrow{p} C_q \end{aligned}$$

because $\sqrt{N}[vec(N^{-1}\widehat{D}_N(\theta_0) - J_\theta(\theta_0))] \xrightarrow{d} \Psi_{q.\phi}$.

When θ is weakly identified, $J_\theta(\theta_0) = J_{\theta,N}(\theta_0) = C_q/\sqrt{N}$. We set $\gamma(N) = 1/\sqrt{N}$, then

$$\begin{aligned} N^{-1/2}\widehat{D}_N(\theta_0) &= \sqrt{N}[N^{-1}\widehat{D}_N(\theta_0) - J_\theta(\theta_0)] + \sqrt{N}J_\theta(\theta_0) \\ &\xrightarrow{d} C_q + \Psi_{q.\phi} \end{aligned}$$

When is totally nonidentified, $J_\theta(\theta_0) = 0$. We set $\gamma(N) = 1/\sqrt{N}$, then

$$N^{-1/2} \widehat{D}_N(\theta_0) \xrightarrow{d} \Psi_{q.\phi}. \qquad Q.E.D.$$

References

[1] Andersen, T.W., Rubin, H., (1949) "Estimators of the Parameters of a Single Equation in a Complete Set of Stochastic Equations," *The Annals of Mathematical Statistics*, 21, 570-582.

[2] Andrews, D.W.K., (1994) "Empirical Process Methods in Econometrics," Ch.37 in : Griliches and Intriligator, eds., *Handbook of Econometrics,* Amsterdam: North-Holland. .

[3] Andrews, D.W.K., Moreira, M., Stock, J., (2004) "Optimal Invariant Similar Tests for Instrumental Variables Regression," working paper, Harvard University and Yale University.

[4] Bound, J., Jaeger, D.A., Baker, R.M., (1995) "Problems with Instrumental Variable Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak," *Journal of the American Statistical Association*, 90, 443–450.

[5] Caner, M., (2005) "Near Exogeneity and Weak Indentification in Generalized Empirical Likelihood Estimators: Fixed and Many Moment Asymptotics," working paper, University of Pittsburgh.

[6] Davidson, J., (1994) *Stochastic Limit Theory: An Introduction for Econometricians,* Cambridge: Oxford University Press.

[7] Dufour, J.M., (1997) "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365-1387.

[8] Dufour, J.M., (2003) "Identification, Weak Instruments, and Statistical Inference in Econometrics," *Canadian Journal of Economics*, 36, 767-808.

[9] Fang, Y., (2005) "Instrumental Variables Regression with Weak Instruments and Near Exogeneity," working paper, University of Pittsburgh.

[10] Guggenberger, P., Smith, R.J., (2005) "Generalized Empirical Likelihood Estimators and Tests under Partial, Weak and Strong Identification," *Econometric Theory*, 21, 667-709.

[11] Guggenberger, P., Wolf, M., (2004) "Subsampling Tests of Parameter Hypotheses and Overidentifying Restrictions with Possible Failure of Identification," working paper, UCLA.

[12] Hall, A.R., Inoue, A., (2003) "The Large Sample Behaviour of the Generalized Method of Moments Estimator in Misspecified Models," *Journal of Econometrics*, 114, 361-394.

[13] Hansen, Lars Peter, (1982) "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

[14] Hansen, Lars Peter, Heaton, J., and Yaron, A., (1996) "Finite Sample Properties of Some Alternative GMM estimators," *Journal of Business and Economic Statistics*, 14, 262-280.

[15] Kleibergen, F., (2005) "Testing Parameters in GMM without Assuming that They Are Identified," *Econometrica,* 73, 1103-1123.

[16] Moreira, M.J., (2003) "A Conditional Likelihood Ratio Tese for Structural Models," *Econometrica*, 71, 1027-1048.

[17] Newey, W.K., (1985) "Generalized Method of Moments Specification Testing," *Journal of Econometrics*, 29, 229-256.

[18] Phillips, Peter C.B., (1984) "Exact Small Sample Theory in the Simultaneous Equations Model," Ch8 in: Griliches and Intriligator, eds., *Handbook of Econometrics*, Amsterdam: North-Holland.

[19] Pollard, D. (1984) *Convergence of Stochastic Process*, New York: Springer Verlag.

[20] Pollard, D. (1990) *Empirical Process: Theory and Applications*, Hayward CA: Institute of Mathematical Statistics.

[21] Sargan, J.D., (1958) "Estimation of Economic Relationship Using Instrumental Variables" *Econometrica*,26, 393-514.

[22] Staiger, D., Stock, J.H., (1997) "Instrumental Variables Regression with Weak Instruments," *Econometrica,* 65, 557-586.

[23] Stock, J.H., Wright, J.H., (2000) "GMM with Weak Identification," *Econometrica,* 68, 1055-1096.

[24] Stock, J.H., Wright, J.H., Yogo, M., (2000) " A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, 20, 518-529.

[25] Van der Vaart, A.W., Wellner, J., (1996) *Weak Convergence and Empirical Process*, New York: Springer-Verlag.

[26] Wang, J. and Zivot, E., (1998) "Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments," *Econometrica*, 66, 1389-1404.

[27] Wooldridge, J.W., (2002) *Econometric Analysis of Cross Section and Panel Data,* Cambridge: MIT Press.

[28] Zivot, E., Startz, R., and Nelson, C.R., (1998) "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," *International Economic Review*, 39, 1119-1144.