

Hypothesis Testing with GMM and GEL Estimators

Fumihiko Goto

Contents

- 1 Introduction
- 2 Preliminaries
- 3 Hypothesis testing with GMM estimators
- 4 Hypothesis testing with GEL estimators

1 Introduction

Suppose that we have a set of observations $\{z_i\}$ and a vector function $g(z; \theta)$ which satisfies the moment condition

$$E[g(z; \theta_0)] = 0$$

at the true value θ_0 . Two major methods to estimate the above model are generalized method of moments (GMM) and generalized empirical likelihood (GEL). Ever since first introduced by Hansen (1982), GMM has been widely applied to analyze economic and financial data. See standard textbooks such as Hamilton (1994) and Hayashi (2000) for basics of GMM, and further, Hall (1999) for hypothesis testing. Owen (1988) proposed the empirical likelihood (EL) estimator, which is known to have better small sample properties than GMM. See Qin and lawless (1994) and Owen (2001) for theory and practice of EL. Later GEL, which includes EL as a special case, was presented by Newey and Smith (2004).

The purpose of this paper is to analyze the asymptotic power properties of hypothesis tests using GMM and GEL estimators. The approach we employ here is to investigate the asymptotic distributions of the test statistics when local alternatives exist as in Holly (1982), Newey (1985a, 1985b), and Newey and West (1988).

Section 2 provides the basic theorems necessary to construct and analyze GMM and GEL hypothesis tests, which include weak law for triangular array, uniform weak law of large numbers, consistency of M-estimators, the Lindeberg-Feller central limit theorem, and the Fisher-Cochran theorem.

Section 3 analyzes hypothesis testing with GMM estimators. Consistency and asymptotic normality of GMM estimators under the local alternatives are shown in a rigorous manner. Then asymptotic properties of GMM hypothesis tests are examined. Hypothesis testing about subsets of moment conditions is also considered.

Section 4 presents hypothesis testing with GEL estimators. We show consistency and asymptotic normality of GEL estimators under the local alternatives in a rigorous manner. Finally, asymptotic properties of GEL hypothesis tests are investigated.

2 Preliminaries

This section provides some basic theorems necessary to investigate the asymptotic properties of GMM and GEL hypothesis testing.

2.1 Weak law of large numbers for triangular array

Consider the sequence $X_{n1}, \dots, X_{nn}$ of random variables, where the probability space for the sequence may change with n . Such a collection is called a triangular array. In most cases we assume that for each n $X_{n1}, \dots, X_{nn}$ are independent. Theorem 1 and Lemma 1 correspond to Theorem 2.2.6 and Lemma 2.2.8 of Durrett (2010), respectively.

Theorem 1 *For each n let X_{nk} , $1 \leq k \leq n$, be independent. Let $b_n > 0$ with $b_n \rightarrow \infty$, and let $\bar{X}_{nk} = X_{nk} 1_{(|X_{nk}| \leq b_n)}$. Suppose that as $n \rightarrow \infty$*

(i) $\sum_{k=1}^n P(|X_{nk}| > b_n) \rightarrow 0$, and

(ii) $b_n^{-2} \sum_{k=1}^n E[\bar{X}_{nk}^2] \rightarrow 0$.

If we let $S_n = X_{n1} + \dots + X_{nn}$ and put $a_n = \sum_{k=1}^n E[\bar{X}_{nk}]$ then

$$\frac{S_n - a_n}{b_n} \xrightarrow{p} 0.$$

Proof: Let $\bar{S}_n = \bar{X}_{n1} + \cdots + \bar{X}_{nn}$. Clearly,

$$P\left(\left|\frac{S_n - a_n}{b_n}\right| > \varepsilon\right) \leq P(S_n \neq \bar{S}_n) + P\left(\left|\frac{\bar{S}_n - a_n}{b_n}\right| > \varepsilon\right).$$

To estimate the first term, we note that

$$P(S_n \neq \bar{S}_n) \leq P(\cup_{k=1}^n \{X_{nk} \neq \bar{X}_{nk}\}) \leq \sum_{k=1}^n P(|X_{nk}| > b_n) \rightarrow 0$$

by (i). For the second term, we note that Chebyshev's inequality, $a_n = E[\bar{S}_n]$, and $\text{var}(X) \leq E[X^2]$ imply

$$\begin{aligned} P\left(\left|\frac{\bar{S}_n - a_n}{b_n}\right| > \varepsilon\right) &\leq \varepsilon^{-2} E\left[\left(\frac{\bar{S}_n - a_n}{b_n}\right)^2\right] = \varepsilon^{-2} b_n^{-2} \text{var}(\bar{S}_n) \\ &= (\varepsilon b_n)^{-2} \sum_{k=1}^n \text{var}(\bar{X}_{nk}) \leq (\varepsilon b_n)^{-2} \sum_{k=1}^n E[\bar{X}_{nk}^2] \rightarrow 0 \end{aligned}$$

by (ii), and the proof is complete. ■

Lemma 1 *If $Y \geq 0$ and $p > 0$ then*

$$E[Y^p] = \int_0^\infty p y^{p-1} P(Y > y) dy.$$

Proof: Using the definition of expected value, Fubini's theorem (for nonnegative random variables), and then calculating the resulting integrals gives

$$\begin{aligned} \int_0^\infty p y^{p-1} P(Y > y) dy &= \int_0^\infty \int_\Omega p y^{p-1} 1_{(Y > y)} dP dy \\ &= \int_\Omega \int_0^\infty p y^{p-1} 1_{(Y > y)} dy dP \\ &= \int_\Omega \int_0^Y p y^{p-1} dy dP = \int_\Omega Y^p dP = E[Y^p], \end{aligned}$$

which is the desired result. ■

Theorem 1 combined with Lemma 1 yields Theorem 2, which plays an important role for weak law of large numbers under the local alternatives.

Theorem 2 For each n let Z_{nk} , $1 \leq k \leq n$, be independent and distributed according to a common probability density function $f_n(z)$ with respect to a σ -finite measure μ on R^p . Let $X_{nk} = g(Z_{nk})$, where g is a measurable function on R^p . It is assumed that there exists a measurable function $a(z)$ such that for almost all z , $f_n(z) < a(z)$ for all n with $\int |g(z)|a(z)d\mu(z) < \infty$. Then

$$\frac{1}{n} \sum_{k=1}^n X_{nk} - E[X_{n1}] \xrightarrow{p} 0.$$

Proof: We have $yP(|X_{n1}| > y) \rightarrow 0$ as $y \rightarrow \infty$. This follows from

$$\begin{aligned} yP(|X_{n1}| > y) &\leq \int |g(z)|1_{(|g(z)|>y)}f_n(z)d\mu(z) \\ &\leq \int |g(z)|1_{(|g(z)|>y)}a(z)d\mu(z) \end{aligned}$$

with dominated convergence theorem. Hence,

$$\sum_{i=1}^n P(|X_{nk}| > n) = nP(|X_{n1}| > n) \rightarrow 0 \text{ as } n \rightarrow \infty,$$

which satisfies (i) of Theorem 1. To check (ii) let $\bar{X}_{n1} = X_{n1}1_{(|X_{n1}| \leq n)}$. Then

$$E[\bar{X}_{n1}^2] = \int_0^\infty 2yP(|\bar{X}_{n1}| > y)dy \leq \int_0^n 2yP(|X_{n1}| > y)dy$$

by Lemma 1. Let $h(y) = 2yP(|X_{n1}| > y)$, then $0 \leq h(y) \leq 2y$ and $h(y) \rightarrow 0$ as $y \rightarrow \infty$, so we must have $M = \sup h(y) < \infty$. Let $h_n(y) = h(ny)$, then $h_n(y) \rightarrow 0$ as $n \rightarrow \infty$ for almost all y . Hence, we have

$$\frac{1}{n} \int_0^n h(y)dy = \int_0^1 h_n(x)dx \rightarrow 0$$

by dominated convergence theorem. We also have

$$\begin{aligned} |E[X_{n1}] - E[\bar{X}_{n1}]| &= \left| \int g(z)1_{(|g(z)|>n)}f_n(z)d\mu(z) \right| \\ &\leq \int |g(z)|1_{(|g(z)|>n)}a(z)d\mu(z) \rightarrow 0 \text{ as } n \rightarrow \infty, \end{aligned}$$

which completes the proof. ■

2.2 Uniform weak law of large numbers

This subsection considers uniform weak law of large numbers for triangular array, generalizing the discussions given by van de Geer (2000) for i.i.d. data. For each n let $X_{n1}, \dots, X_{nn}$ be independent copies of a random variable X_n in a measurable space $(\mathcal{X}, \mathcal{A})$ with distribution $P^{(n)}$, and let $\mathcal{G} = \{g\}$ be a class of functions on $\mathcal{X}$. We say that $(\mathcal{G}, P^{(n)})$ satisfies the uniform weak law of large numbers, if

$$\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{i=1}^n g(X_{ni}) - E[g(X_n)] \right| \xrightarrow{P} 0,$$

or equivalently for any $\varepsilon > 0$

$$\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{i=1}^n g(X_{ni}) - E[g(X_n)] \right| < \varepsilon \quad \text{w.p.a.1},$$

where w.p.a.1 is an abbreviation for “with probability approaching one.” Note that we say $(\mathcal{G}, P^{(n)})$ satisfies the weak law of large numbers, if

$$\left| \frac{1}{n} \sum_{i=1}^n g(X_{ni}) - E[g(X_n)] \right| \xrightarrow{P} 0.$$

Let P_n be the empirical distribution based on $X_{n1}, \dots, X_{nn}$. Then we may use the notation

$$\frac{1}{n} \sum_{i=1}^n g(X_{ni}) = \int g dP_n.$$

The difference between average and expectation is

$$\frac{1}{n} \sum_{i=1}^n g(X_{ni}) - E[g(X_n)] = \int g d(P_n - P^{(n)}).$$

In the new formulation, we say that $(\mathcal{G}, P^{(n)})$ satisfies the uniform weak law of large numbers if

$$\sup_{g \in \mathcal{G}} \left| \int g d(P_n - P^{(n)}) \right| < \varepsilon \quad \text{w.p.a.1}$$

for any $\varepsilon > 0$.

Let Q be a measure on $(\mathcal{X}, \mathcal{A})$ and $L(Q) = \{g : \mathcal{X} \rightarrow \mathbb{R} : \int |g| dQ < \infty\}$. For $g \in L(Q)$, write

$$\|g\|_Q = \int |g| dQ.$$

We refer to $\|\cdot\|_Q$ as the $L(Q)$ -norm, and call $\|g_1 - g_2\|_Q$ the $L(Q)$ -distance between g_1 and g_2 .

Let $N_B(\delta, \mathcal{G}, Q)$ be the smallest value of N for which there exist pairs of functions $\{[g_j^L, g_j^U]\}_{j=1}^N$ such that $\|g_j^U - g_j^L\|_Q \leq \delta$ for all $j = 1, \dots, N$, and such that for each $g \in \mathcal{G}$, there is a $j = j(g) \in \{1, \dots, N\}$ such that

$$g_j^L \leq g \leq g_j^U.$$

Then $H_B(\delta, \mathcal{G}, Q) = \log N_B(\delta, \mathcal{G}, Q)$ is called the δ -entropy with bracketing of $\mathcal{G}$.

Theorem 3 Suppose that $(g, P^{(n)})$ satisfies the weak law of large numbers for each $g \in \mathcal{G}$, and there exists a measure Q such that $\sup_n P^{(n)}(A) \leq Q(A)$ for all $A \in \mathcal{A}$ and

$$H_B(\delta, \mathcal{G}, Q) < \infty \text{ for all } \delta > 0.$$

Then $(\mathcal{G}, P^{(n)})$ satisfies the uniform weak law of large numbers.

Proof: Let $\delta > 0$ be arbitrary, $\{[g_j^L, g_j^U]\}_{j=1}^N$ be a δ -bracketing set for $\mathcal{G}$, i.e., $\|g_j^U - g_j^L\|_Q \leq \delta$, $j = 1, \dots, N$, and each $g \in \mathcal{G}$ can be squeezed between a pair $[g_j^L, g_j^U]$. Then

$$\begin{aligned} \int g d(P_n - P^{(n)}) &= \int g dP_n - \int g dP^{(n)} \\ &\leq \int g_j^U dP_n - \int g dP^{(n)} \\ &= \int g_j^U d(P_n - P^{(n)}) + \int (g_j^U - g) dP^{(n)} \\ &\leq \int g_j^U d(P_n - P^{(n)}) + \delta. \end{aligned}$$

Similarly,

$$\begin{aligned} \int g d(P_n - P^{(n)}) &\geq \int g_j^L d(P_n - P^{(n)}) - \int (g - g_j^L) dP^{(n)} \\ &\geq \int g_j^L d(P_n - P^{(n)}) - \delta. \end{aligned}$$

Since $\{[g_j^L, g_j^U]\}_{j=1}^N$ is finite, we have that eventually

$$\max_{j=1, \dots, N} \left| \int g_j^U d(P_n - P^{(n)}) \right| \leq \delta \quad \text{w.p.a.1,}$$

as well as

$$\max_{j=1, \dots, N} \left| \int g_j^L d(P_n - P^{(n)}) \right| \leq \delta \quad \text{w.p.a.1.}$$

So eventually,

$$\sup_{g \in \mathcal{G}} \left| \int g d(P_n - P^{(n)}) \right| \leq 2\delta \quad \text{w.p.a.1.}$$

This completes the proof. $\blacksquare$

Let

$$\mathcal{G} = \{g_\theta : \theta \in \Theta\},$$

where (Θ, τ) is a compact metric space, and where the map $\theta \mapsto g_\theta(x)$ is continuous for Q -almost all x . Moreover, we assume that the envelope

$$G = \sup_{\theta \in \Theta} |g_\theta|$$

is in $L(Q)$.

Theorem 4 *For the class $\mathcal{G} = \{g_\theta : \theta \in \Theta\}$ defined above, we have*

$$H_B(\delta, \mathcal{G}, Q) < \infty, \quad \text{for all } \delta > 0.$$

Proof: Write

$$w(\theta, \rho)(x) = \sup_{\tau(\theta, \tilde{\theta}) < \rho} |g_\theta(x) - g_{\tilde{\theta}}(x)|, \quad \theta \in \Theta, \quad \rho > 0.$$

Then $w(\theta, \rho)(x) \rightarrow 0$ as $\rho \rightarrow 0$, for Q -almost all x , since g_θ is uniformly continuous on Θ . By dominated convergence,

$$\lim_{\rho \rightarrow 0} \int w(\theta, \rho)(x) dQ = 0.$$

Let $\delta > 0$ be arbitrary. We can take, for each θ , ρ_θ in such a way that

$$\int w(\theta, \rho_\theta)(x) dQ \leq \delta.$$

Let $B_\theta = \{\tilde{\theta} : \tau(\tilde{\theta}, \theta) < \rho_\theta\}$ and let $B_{\theta_1}, \dots, B_{\theta_N}$ be a finite cover of Θ . This is possible since Θ is compact. Define $g_j^L = g_{\theta_j} - w(\theta_j, \rho_{\theta_j})$, and $g_j^U = g_{\theta_j} + w(\theta_j, \rho_{\theta_j})$, $j = 1, \dots, N$. Then $0 \leq \int (g_j^U - g_j^L) dQ \leq 2\delta$ and for $\theta \in B_{\theta_j}$, $g_j^L \leq g_\theta \leq g_j^U$. It follows that

$$H_B(2\delta, \mathcal{G}, Q) \leq \log N,$$

and the proof is complete. ■

2.3 Consistency of M-Estimators

The following theorem corresponds to Theorem 5.7 of van der Vaart (1998).

Theorem 5 *Let M_n be random functions and let M be a fixed function of θ such that for every $\varepsilon > 0$*

$$\sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)| \xrightarrow{P} 0, \quad (2.1)$$

and

$$\sup_{\theta: d(\theta, \theta_0) \geq \varepsilon} M(\theta) < M(\theta_0). \quad (2.2)$$

Then any sequence of estimators $\hat{\theta}_n$ with $M_n(\hat{\theta}_n) \geq M_n(\theta_0) - o_p(1)$ converges in probability to θ_0 .

Proof: By the property of $\hat{\theta}_n$, we have

$$M_n(\hat{\theta}_n) \geq M_n(\theta_0) - o_p(1).$$

Because the uniform convergence of M_n to M implies the convergence of $M_n(\theta_0) \xrightarrow{P} M(\theta_0)$, the right side equals $M(\theta_0) - o_p(1)$. It follows that

$$M_n(\hat{\theta}_n) \geq M(\theta_0) - o_p(1),$$

whence

$$\begin{aligned} M(\theta_0) - M(\hat{\theta}_n) &\leq M_n(\hat{\theta}_n) - M(\hat{\theta}_n) + o_p(1) \\ &\leq \sup_{\theta} |M_n - M|(\theta) + o_p(1) \xrightarrow{P} 0, \end{aligned}$$

by equation (2.1). By equation (2.2), there exists for every $\varepsilon > 0$ a number $\eta > 0$ such that $M(\theta) < M(\theta_0) - \eta$ for every θ with $d(\theta, \theta_0) \geq \varepsilon$. Thus, the event $\{d(\hat{\theta}_n, \theta_0) \geq \varepsilon\}$ is contained in the event $\{M(\hat{\theta}_n) < M(\theta_0) - \eta\}$. The probability of the latter event converges to 0, in view of the preceding display, which is the desired result. ■

2.4 Lindeberg-Feller Central Limit Theorem

Consider the triangular array $X_{n1}, \dots, X_{nn}$ of random variables. Put $S_n = X_{n1} + \dots + X_{nn}$. Suppose that

$$E[X_{nk}] = 0, \quad \sigma_{nk}^2 = E[X_{nk}^2], \quad s_n^2 = \sum_{k=1}^n \sigma_{nk}^2. \quad (2.3)$$

The following assumption is called the Lindeberg condition.

Assumption 1

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{s_n^2} \int_{|X_{nk}| \geq \varepsilon s_n} X_{nk}^2 dP = 0$$

for any $\varepsilon > 0$.

The next theorem is the well-known Lindeberg-Feller theorem.

Theorem 6 *Suppose that for each n the sequence $X_{n1}, \dots, X_{nn}$ is independent and satisfies conditions (2.3). If Assumption 1 is satisfied, then*

$$\frac{S_n}{s_n} \xrightarrow{d} N(0, 1).$$

Proof: See Billingsley (1986), p. 369. ■

2.5 Fisher-Cochran Theorem

The following theorem is part of the Fisher-Cochran theorem. See Rao (1973), p. 185 for the whole theorem. Let $\chi^2(r, \nu)$ denote a noncentral chi-square distribution with r degrees of freedom and a noncentrality parameter ν .

Theorem 7 Let $y \sim N(\mu, I_n)$ and A be an $n \times n$ -symmetric idempotent matrix of rank k , then

$$Q = y' Ay \sim \chi^2(k, \mu' A \mu).$$

Proof: Since A is symmetric and idempotent, there exists an orthogonal matrix P such that $PAP' = D$, where

$$D = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}.$$

Let $x = PAy$. Then

$$\text{Var}[x] = PAA'P' = PAP' = D,$$

and

$$E[x] = PA\mu = DP\mu = \begin{pmatrix} \hat{D}P\mu \\ 0 \end{pmatrix},$$

where $\hat{D}$ is a $k \times n$ matrix such that $\hat{D} = (I_k \ 0)$. Let $\hat{x} = \hat{D}x$. Then $x' = (\hat{x}' \ 0')'$ and

$$\hat{x} \sim N(\hat{D}P\mu, I_k).$$

Hence,

$$y' Ay = y' A' P' PAy = x' x = \hat{x}' \hat{x} \sim \chi^2(k, \delta),$$

where

$$\delta = (\hat{D}P\mu)' \hat{D}P\mu = \mu' P' DP\mu = \mu' A \mu,$$

and the proof is complete. ■

3 Hypothesis testing with GMM estimators

Let $F(z; \theta, \eta)$ be a p -variate probability distribution having a density function $f(z; \theta, \eta)$ with respect to a σ -finite measure μ on R^p , where θ is a q -dimensional Euclidean parameter of primary interest and η is a nuisance parameter which is finite or infinite dimensional. Let $g(z; \theta)$ be an r -vector of

functions with $r \geq q$. The model we consider is $f(z; \theta, \eta)$ which satisfies the moment condition

$$E_{\theta, \eta}[g(z; \theta)] = \int g(z; \theta) f(z; \theta, \eta) d\mu(z) = 0. \quad (3.1)$$

Let $\Omega = E[gg']$ and $G = E[\partial g(z; \theta_0)/\partial \theta]$, where $g = g(z; \theta_0)$ and $E[\cdot]$ indicates $E_{\theta_0, \eta_0}[\cdot]$. We call $g(z; \theta)$ the moment function. Suppose that the parameter θ of $g(z; \theta)$ is divided into a q_1 -vector β and a q_2 -vector γ , i.e., $\theta = (\beta', \gamma')'$. Let $\theta_0 = (\beta'_0, \gamma'_0)'$ and $\theta_n = (\beta'_n, \gamma'_n)'$. Consider the pair $(g(z; \theta), f(z; \theta, \eta))$ that satisfies equation (3.1). The observed data z_i ($i = 1, \dots, n$) consists of p -random vectors which are independently distributed with a common probability density function $f(z; \theta_n, \eta_0)$ with $\theta_n \rightarrow \theta_0$. Note that an extra n subscript on z_i has been suppressed for notational convenience. In the present section, we will consider a null hypothesis of the form

$$H_0 : \gamma_n = \gamma_0.$$

To measure the power of the test, we introduce the following local alternatives of the form

$$H_1 : \gamma_n = \gamma_0 + \delta/\sqrt{n},$$

where $\delta \in R^{q_2}$. Let $\hat{g}(\theta) = n^{-1} \sum_{i=1}^n g(z_i; \theta)$. The constrained GMM estimator is defined by

$$\bar{\theta}_{\text{GMM}} = \arg \min_{\theta \in \Theta} \hat{g}(\theta)' \hat{\Omega}^{-1} \hat{g}(\theta) \quad \text{subject to } \gamma = \gamma_0, \quad (3.2)$$

where $\hat{\Omega}$ is some preliminary estimator satisfying $\hat{\Omega} \xrightarrow{p} \Omega$, and is assumed to be symmetric and positive definite without loss of generality. On the other hand, the unconstrained GMM estimator is given by

$$\hat{\theta}_{\text{GMM}} = \arg \min_{\theta \in \Theta} \hat{g}(\theta)' \hat{\Omega}^{-1} \hat{g}(\theta). \quad (3.3)$$

In what follows, let $g(z; \theta) = g(z; \beta, \gamma)$ and $f(z; \theta, \eta) = f(z; \beta, \gamma, \eta)$ by abuse of notation. We call a model “restricted” if it has $g(z; \beta, \gamma_0)$, while “unrestricted” if it has $g(z; \beta, \gamma)$ for the moment condition. Estimators $\bar{\theta}_{\text{GMM}}$ and $\hat{\theta}_{\text{GMM}}$ correspond to the GMM estimators for the restricted and unrestricted models, respectively. Let $H = E[\partial g(z; \beta_0, \gamma_0)/\partial \beta]$ and $K = E[\partial g(z; \beta_0, \gamma_0)/\partial \gamma]$. Then $G = (H, K)$.

3.1 Consistency of estimators under local alternatives

The next assumption guarantees the consistency of the GMM estimator under the local alternatives.

Assumption 2 (a) $\theta_0 \in \Theta$ is the unique solution to $E[g(z; \theta)] = 0$; (b) Θ is compact; (c) $g(z; \theta)$ and $f(z; \theta, \eta_0)$ are continuous at each $\theta \in \Theta$ for almost all z ; (d) There exist measurable functions $b_1(z)$ and $b_2(z)$ such that for almost all z , $\|g(z; \theta)\| < b_1(z)$ and $f(z; \theta, \eta_0) < b_2(z)$ for all $\theta \in \Theta$ with $\int b_1(z)b_2(z)d\mu(z) < \infty$; (e) Ω is nonsingular.

Let $g_0(\theta) = E[g(z; \theta)]$ and let $E_n[\cdot]$ indicate $E_{\theta_n, \eta_0}[\cdot]$. Let $g^j(z; \theta)$ and $\hat{g}^j(\theta)$ denote the j -th elements of $g(z; \theta)$ and $\hat{g}(\theta)$, respectively, and $g_0^j(\theta) = E[g^j(z; \theta)]$.

Lemma 2 If Assumption 2 is satisfied, then under the local alternatives

$$\sup_{\theta \in \Theta} \|\hat{g}(\theta) - g_0(\theta)\| \xrightarrow{P} 0.$$

Proof: By Theorem 2 with (d) of Assumption 2, we have

$$|\hat{g}^j(\theta) - E_n[g^j(z; \theta)]| \xrightarrow{P} 0,$$

for each θ . Dominated convergence with (c) and (d) gives

$$|E_n[g^j(z; \theta)] - g_0^j(\theta)| \rightarrow 0.$$

Thus, we have

$$|\hat{g}^j(\theta) - g_0^j(\theta)| \xrightarrow{P} 0,$$

for each θ . Proceed by verifying the hypotheses of Theorems 3 and 4. Let $P^{(n)}(A) = \int_A f(z; \theta_n, \eta_0)d\mu(z)$ and $Q(A) = \int_A b_2(z)d\mu(z)$ for each measurable set $A \in R^p$. Then (d) gives $\sup_n P^{(n)}(A) \leq Q(A)$ for all A . Conditions of Theorem 4 follow by (b), (c), and (d). Hence, we have $\sup_{\theta \in \Theta} |\hat{g}^j(\theta) - g_0^j(\theta)| \xrightarrow{P} 0$ for all j , and clearly,

$$\sup_{\theta \in \Theta} \|\hat{g}(\theta) - g_0(\theta)\| \xrightarrow{P} 0.$$

This completes the proof. ■

Theorem 8 *If Assumption 2 is satisfied, then under the local alternatives $\bar{\theta}_{\text{GMM}} \xrightarrow{P} \theta_0$ and $\hat{\theta}_{\text{GMM}} \xrightarrow{P} \theta_0$.*

Proof: We show consistency of $\bar{\theta}_{\text{GMM}}$. Consistency of $\hat{\theta}_{\text{GMM}}$ can be shown in a similar way. To verify the conditions of Theorem 5, let $M_n(\theta) = -\hat{g}(\theta)' \hat{\Omega}^{-1} \hat{g}(\theta)$ and let $M(\theta) = -g_0(\theta)' \Omega^{-1} g_0(\theta)$. By (c) and (d) of Assumption 2 with dominated convergence $g_0(\theta)$ is continuous, and $M(\theta)$ is continuous, too. By (a) and (e) θ_0 is the unique solution to $M(\theta) = 0$, which is the maximum of $M(\theta)$ for $\theta \in \Theta$. Since the subset $\{\theta : d(\theta, \theta_0) \geq \varepsilon, \theta \in \Theta\}$ is compact, we have

$$\sup_{\theta: d(\theta, \theta_0) \geq \varepsilon} M(\theta) < M(\theta_0),$$

which is condition (2.2). By (b) $g_0(\theta)$ is bounded on Θ , and by the triangle and Cauchy-Schwarz inequalities,

$$\begin{aligned} |M_n(\theta) - M(\theta)| &\leq |[\hat{g}(\theta) - g_0(\theta)]' \hat{\Omega}^{-1} [\hat{g}(\theta) - g_0(\theta)]| + 2|g_0(\theta)' \hat{\Omega}^{-1} [\hat{g}(\theta) - g_0(\theta)]| \\ &\quad + |g_0(\theta)' (\hat{\Omega}^{-1} - \Omega^{-1}) g_0(\theta)| \\ &\leq \|\hat{g}(\theta) - g_0(\theta)\|^2 \|\hat{\Omega}^{-1}\| + 2\|g_0(\theta)\| \|\hat{g}(\theta) - g_0(\theta)\| \|\hat{\Omega}^{-1}\| \\ &\quad + \|g_0(\theta)\|^2 \|\hat{\Omega}^{-1} - \Omega^{-1}\|, \end{aligned}$$

so that $\sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)| \xrightarrow{P} 0$, and (2.1) holds. Obviously, $M_n(\bar{\theta}_{\text{GMM}}) \geq M_n(\theta_0)$. The conclusion follows. ■

3.2 Asymptotic normality of estimators under local alternatives

The following assumption combined with Assumption 2 guarantees the asymptotic normality of the GMM estimator under the local alternatives.

Assumption 3 (a) $\theta_0 \in \text{int}(\Theta)$; (b) $g(z; \theta)$ and $f(z; \theta, \eta_0)$ are continuously differentiable in a neighborhood $\mathcal{N}$ of θ_0 for almost all z ; (c) There exist measurable functions $b_1(z)$ and $b_2(z)$ such that for almost all z , $\|g(z; \theta)\|^2 < b_1(z)$, $\|\partial g(z; \theta)/\partial \theta\| < b_1(z)$, and $f(z; \theta, \eta_0) < b_2(z)$ for all $\theta \in \Theta$ with $\int b_1(z) b_2(z) d\mu(z) < \infty$; (d) $\text{rank}(G) = q$.

Lemma 3 *If Assumptions 2 and 3 are satisfied, then under the local alternatives*

$$\sup_{\theta \in \Theta} \left\| \frac{\partial \hat{g}(\theta)}{\partial \theta} - E \left[\frac{\partial g(z; \theta)}{\partial \theta} \right] \right\| \xrightarrow{p} 0.$$

Proof: The result follows similarly to the proof of Lemma 2. ■

For $g(z; \theta_n)$, we have expectation $E_n[g(z; \theta_n)] = 0$ by (3.1) and variance $\Omega_n = E_n[g(z; \theta_n)g(z; \theta_n)']$. By (c) of Assumption 2 and (c) of Assumption 3 with dominated convergence, we have

$$\lim \Omega_n = \Omega. \quad (3.4)$$

Let $\bar{\theta}_{\text{GMM}} = (\bar{\beta}', \gamma_0')'$ and $\hat{g}(\beta, \gamma) = n^{-1} \sum_{i=1}^n g(z_i; \beta, \gamma)$.

Theorem 9 *If Assumptions 2 and 3 are satisfied, then under the local alternatives*

$$\sqrt{n}(\bar{\beta} - \beta_n) = -(H' \Omega^{-1} H)^{-1} H' \Omega^{-1} \sqrt{n} \hat{g}(\beta_n, \gamma_0) + o_p(1), \quad (3.5)$$

$$\sqrt{n} \hat{g}(\bar{\theta}_{\text{GMM}}) = \{I_r - H(H' \Omega^{-1} H)^{-1} H' \Omega^{-1}\} \sqrt{n} \hat{g}(\beta_n, \gamma_0) + o_p(1), \quad (3.6)$$

$$\sqrt{n} \hat{g}(\beta_n, \gamma_0) = \sqrt{n} \hat{g}(\theta_n) - K \delta + o_p(1), \quad (3.7)$$

$$\sqrt{n} \hat{g}(\theta_n) \xrightarrow{d} N(0, \Omega), \quad (3.8)$$

$$\sqrt{n}(\hat{\theta}_{\text{GMM}} - \theta_n) = -(G' \Omega^{-1} G)^{-1} G' \Omega^{-1} \sqrt{n} \hat{g}(\theta_n) + o_p(1), \quad (3.9)$$

$$\sqrt{n} \hat{g}(\hat{\theta}_{\text{GMM}}) = \{I_r - G(G' \Omega^{-1} G)^{-1} G' \Omega^{-1}\} \sqrt{n} \hat{g}(\theta_n) + o_p(1). \quad (3.10)$$

Proof: Let $X_n(z) = \lambda' g(z; \theta_n)$ and $X_{ni} = X_n(z_i)$ ($i = 1, \dots, n$), where $z_i \sim f(\cdot; \theta_n, \eta_0)$ and λ is a nonzero element of R^r . Note that X_{ni} has expectation zero and variance $\lambda' \Omega_n \lambda$, which is positive for large enough n by equation (3.4) and (e) of Assumption 2. Note that for each n , X_{ni} ($i = 1, \dots, n$) are identically distributed, so that by (c) of Assumption 3, for any $\varepsilon > 0$,

$$\begin{aligned} & \frac{1}{n \lambda' \Omega_n \lambda} \sum_{i=1}^n \int_{|X_{ni}| \geq \varepsilon \sqrt{n \lambda' \Omega_n \lambda}} |X_{ni}|^2 f(z_i; \theta_n, \eta_0) d\mu(z_i) \\ &= \frac{1}{\lambda' \Omega_n \lambda} \int 1_{(|X_n(z)| \geq \varepsilon \sqrt{n \lambda' \Omega_n \lambda})} |X_n(z)|^2 f(z; \theta_n, \eta_0) d\mu(z) \\ &\leq \frac{1}{\lambda' \Omega_n \lambda} \int 1_{(\|\lambda\|^2 b_1(z) \geq \varepsilon^2 n \lambda' \Omega_n \lambda)} \|\lambda\|^2 b_1(z) b_2(z) d\mu(z). \end{aligned} \quad (3.11)$$

By (c) of Assumption 3 and equation (3.4) with dominated convergence, equation (3.11) implies that the conditions of Theorem 6 are satisfied, and

$$\frac{1}{\sqrt{n\lambda'\Omega_n\lambda}} \sum_{i=1}^n X_{ni} \xrightarrow{d} N(0, 1).$$

By $\lim \lambda'\Omega_n\lambda = \lambda'\Omega\lambda$, $\sum_{i=1}^n X_{ni}/\sqrt{n} \xrightarrow{d} N(0, \lambda'\Omega\lambda)$, and by the Cramer-Wold device, we have

$$\sum_{i=1}^n g(z_i, \theta_n)/\sqrt{n} \xrightarrow{d} N(0, \Omega),$$

which is equation (3.8). A mean value expansion gives

$$\sqrt{n}\hat{g}(\theta_n) = \sqrt{n}\hat{g}(\beta_n, \gamma_0) + \frac{\partial \hat{g}(\beta_n, \dot{\gamma})}{\partial \gamma} \delta,$$

where $\dot{\gamma}$ lies on the line joining γ_n and γ_0 ($\dot{\gamma}$ actually differs from row to row of $\partial \hat{g}(\beta_n, \dot{\gamma})/\partial \gamma$). By Lemma 3 with $\lim \beta_n = \beta_0$ and $\lim \dot{\gamma} = \gamma_0$ we have

$$\sqrt{n}\hat{g}(\theta_n) = \sqrt{n}\hat{g}(\beta_n, \gamma_0) + K\delta + o_p(1),$$

which is equivalent to equation (3.7). To show that equation (3.5) holds, a mean value expansion gives

$$\sqrt{n}\hat{g}(\bar{\theta}_{\text{GMM}}) = \sqrt{n}\hat{g}(\beta_n, \gamma_0) + \frac{\partial \hat{g}(\dot{\beta}, \gamma_0)}{\partial \beta} \sqrt{n}(\bar{\beta} - \beta_n), \quad (3.12)$$

where $\dot{\beta}$ lies on the line joining $\bar{\beta}$ and β_n ($\dot{\beta}$ actually differs from row to row of $\partial \hat{g}(\dot{\beta}, \gamma_0)/\partial \beta$), with $\text{plim } \dot{\beta} = \beta_0$ because $\text{plim } \bar{\theta}_{\text{GMM}} = \theta_0$ by Theorem 8. The first-order conditions of equation (3.2) are

$$\left[\frac{\partial \hat{g}(\bar{\theta}_{\text{GMM}})}{\partial \beta} \right]' \hat{\Omega}^{-1} \hat{g}(\bar{\theta}_{\text{GMM}}) = 0, \quad (3.13)$$

which holds w.p.a.1 by (a) of Assumption 3 with Theorem 8. Premultiplying the both sides of equation (3.12) by $[\partial \hat{g}(\bar{\theta}_{\text{GMM}})/\partial \beta]' \hat{\Omega}^{-1}$ yields,

$$\begin{aligned} & \left[\frac{\partial \hat{g}(\bar{\theta}_{\text{GMM}})}{\partial \beta} \right]' \hat{\Omega}^{-1} \frac{\partial \hat{g}(\dot{\beta}, \gamma_0)}{\partial \beta} \sqrt{n}(\bar{\beta} - \beta_n) \\ &= - \left[\frac{\partial \hat{g}(\bar{\theta}_{\text{GMM}})}{\partial \beta} \right]' \hat{\Omega}^{-1} \sqrt{n}\hat{g}(\beta_n, \gamma_0), \end{aligned}$$

by equation (3.13). Then equation (3.5) follows from Lemmas 2 and 3 with $\text{plim } \bar{\theta}_{\text{GMM}} = \theta_0$, $\text{plim } \hat{\beta} = \beta_0$, and $\sqrt{n}\hat{g}(\beta_n, \gamma_0) = O_p(1)$. Equation (3.6) now follows from equations (3.5) and (3.12) with $\text{plim } \partial\hat{g}(\hat{\beta}, \gamma_0)/\partial\beta = H$. Equations (3.9) and (3.10) can be proved similarly. ■

3.3 Hypothesis testing

Under these assumptions we construct the GMM hypothesis test in terms of the distance metric statistic DM_{GMM} . Let κ denote the noncentrality parameter of the noncentral chi-square distribution corresponding to the test.

Lemma 4 *If Assumptions 2 and 3 are satisfied, then under the local alternatives*

$$\sqrt{n}(\bar{\theta}_{\text{GMM}} - \hat{\theta}_{\text{GMM}})G'\Omega^{-1}G\sqrt{n}(\bar{\theta}_{\text{GMM}} - \hat{\theta}_{\text{GMM}}) \xrightarrow{d} \chi^2(q_2, \kappa),$$

where $\kappa = \delta'\{K'\Omega^{-1}K - K'\Omega^{-1}H(H'\Omega^{-1}H)^{-1}H'\Omega^{-1}K\}\delta$.

Proof: It follows from equations (3.5), (3.7), and (3.9) that

$$\begin{aligned} \sqrt{n}\left[\bar{\theta}_{\text{GMM}} - \begin{pmatrix} \beta_n \\ \gamma_0 \end{pmatrix}\right] &= \sqrt{n}\begin{pmatrix} \bar{\beta} - \beta_n \\ 0 \end{pmatrix} \\ &= -\begin{pmatrix} I_{q_1} \\ 0 \end{pmatrix}(H'\Omega^{-1}H)^{-1}H'\Omega^{-1}\sqrt{n}\hat{g}(\beta_n, \gamma_0) + o_p(1) \end{aligned}$$

and

$$\sqrt{n}\left[\hat{\theta}_{\text{GMM}} - \begin{pmatrix} \beta_n \\ \gamma_0 \end{pmatrix}\right] = -(G'\Omega^{-1}G)^{-1}G'\Omega^{-1}\sqrt{n}\hat{g}(\beta_n, \gamma_0) + o_p(1).$$

Thus, we have

$$\Omega^{-1/2}G\sqrt{n}(\bar{\theta}_{\text{GMM}} - \hat{\theta}_{\text{GMM}}) = A\Omega^{-1/2}\sqrt{n}\hat{g}(\beta_n, \gamma_0) + o_p(1),$$

where

$$A = \Omega^{-1/2}G(G'\Omega^{-1}G)^{-1}G'\Omega^{-1/2} - \Omega^{-1/2}H(H'\Omega^{-1}H)^{-1}H'\Omega^{-1/2}$$

with $A^2 = A$ and $\text{rank}(A) = q_2$. Further, we have

$$\delta'K'\Omega^{-1/2}A\Omega^{-1/2}K\delta = \delta'\{K'\Omega^{-1}K - K'\Omega^{-1}H(H'\Omega^{-1}H)^{-1}H'\Omega^{-1}K\}\delta.$$

Since $\Omega^{-1/2}\sqrt{n}\hat{g}(\beta_n, \gamma_0) \xrightarrow{d} N(-\Omega^{-1/2}K\delta, I_r)$ by equations (3.7) and (3.8), the result follows from Theorem 7 with the continuous-mapping theorem. ■

Theorem 10 *If Assumptions 2 and 3 are satisfied, then under the local alternatives*

$$\begin{aligned} \text{DM}_{\text{GMM}} &= n \left\{ \hat{g}(\bar{\theta}_{\text{GMM}})' \hat{\Omega}^{-1} \hat{g}(\bar{\theta}_{\text{GMM}}) - \hat{g}(\hat{\theta}_{\text{GMM}})' \hat{\Omega}^{-1} \hat{g}(\hat{\theta}_{\text{GMM}}) \right\} \\ &\xrightarrow{d} \chi^2(q_2, \kappa), \end{aligned}$$

where $\kappa = \delta' \{K' \Omega^{-1} K - K' \Omega^{-1} H (H' \Omega^{-1} H)^{-1} H' \Omega^{-1} K\} \delta$.

Proof: A mean value expansion of the sample moments for $\bar{\theta}_{\text{GMM}}$ about $\hat{\theta}_{\text{GMM}}$ gives

$$\sqrt{n} \hat{g}(\bar{\theta}_{\text{GMM}}) = \sqrt{n} \hat{g}(\hat{\theta}_{\text{GMM}}) + \frac{\partial \hat{g}(\hat{\theta})}{\partial \theta} \sqrt{n} (\bar{\theta}_{\text{GMM}} - \hat{\theta}_{\text{GMM}}),$$

where $\hat{\theta}$ lies on the line between $\bar{\theta}_{\text{GMM}}$ and $\hat{\theta}_{\text{GMM}}$. Since $\text{plim } \partial \hat{g}(\hat{\theta}_{\text{GMM}})/\partial \theta = \text{plim } \partial \hat{g}(\hat{\theta})/\partial \theta = G$, we have

$$\sqrt{n} \hat{g}(\bar{\theta}_{\text{GMM}}) = \sqrt{n} \hat{g}(\hat{\theta}_{\text{GMM}}) + \frac{\partial \hat{g}(\hat{\theta}_{\text{GMM}})}{\partial \theta} \sqrt{n} (\bar{\theta}_{\text{GMM}} - \hat{\theta}_{\text{GMM}}) + o_p(1).$$

Let $\hat{G} = \partial \hat{g}(\hat{\theta}_{\text{GMM}})/\partial \theta$. Then

$$\begin{aligned} \text{DM}_{\text{GMM}} &= 2\sqrt{n}(\bar{\theta}_{\text{GMM}} - \hat{\theta}_{\text{GMM}})' \hat{G}' \hat{\Omega}^{-1} \sqrt{n} \hat{g}(\hat{\theta}_{\text{GMM}}) \\ &\quad + \sqrt{n}(\bar{\theta}_{\text{GMM}} - \hat{\theta}_{\text{GMM}})' \hat{G}' \hat{\Omega}^{-1} \hat{G} \sqrt{n} (\bar{\theta}_{\text{GMM}} - \hat{\theta}_{\text{GMM}}) + o_p(1) \\ &= \sqrt{n}(\bar{\theta}_{\text{GMM}} - \hat{\theta}_{\text{GMM}})' G' \Omega^{-1} G \sqrt{n} (\bar{\theta}_{\text{GMM}} - \hat{\theta}_{\text{GMM}}) + o_p(1), \end{aligned}$$

since $\hat{G}' \hat{\Omega}^{-1} \sqrt{n} \hat{g}(\hat{\theta}_{\text{GMM}}) = 0$, the first order conditions of equation (3.3). The conclusion follows from Lemma 4. ■

(alternative proof) Combining equation (3.10) with equation (3.7) yields

$$\sqrt{n} \hat{g}(\hat{\theta}_{\text{GMM}}) = \{I_r - G(G' \Omega^{-1} G)^{-1} G' \Omega^{-1}\} \sqrt{n} \hat{g}(\beta_n, \gamma_0) + o_p(1).$$

Then, with equation (3.6) we have

$$\text{DM}_{\text{GMM}} = \sqrt{n} \hat{g}(\beta_n, \gamma_0)' \Omega^{-1/2} A \Omega^{-1/2} \sqrt{n} \hat{g}(\beta_n, \gamma_0) + o_p(1).$$

The conclusion follows. ■

Consider the test statistic J_{GMM} for overidentifying restrictions, which is expressed as

$$J_{\text{GMM}} = n \hat{g}(\bar{\theta}_{\text{GMM}})' \hat{\Omega}^{-1} \hat{g}(\bar{\theta}_{\text{GMM}}).$$

The next theorem gives the asymptotic distribution of J_{GMM} under the local alternatives.

Theorem 11 *If Assumptions 2 and 3 are satisfied, then under the local alternatives*

$$J_{\text{GMM}} = n \hat{g}(\bar{\theta}_{\text{GMM}})' \hat{\Omega}^{-1} \hat{g}(\bar{\theta}_{\text{GMM}}) \xrightarrow{d} \chi^2(r - q_1, \kappa),$$

where $\kappa = \delta' \{K' \Omega^{-1} K - K' \Omega^{-1} H (H' \Omega^{-1} H)^{-1} H' \Omega^{-1} K\} \delta$.

Proof: By equation (3.6) we have

$$\Omega^{-1/2} \sqrt{n} \hat{g}(\bar{\theta}_{\text{GMM}}) = N \Omega^{-1/2} \sqrt{n} \hat{g}(\beta_n, \gamma_0) + o_p(1),$$

where $N = I_r - \Omega^{-1/2} H (H' \Omega^{-1} H)^{-1} H' \Omega^{-1/2}$ with $N^2 = N$ and $\text{rank}(N) = r - q_1$. Theorem 7 gives the desired result. ■

When $r = q$, $\hat{g}(\hat{\theta}_{\text{GMM}})' \hat{\Omega}^{-1} \hat{g}(\hat{\theta}_{\text{GMM}})$ is asymptotically equal to zero, or often exactly equal to zero since the number of moments equals the number of parameters. Thus the distance statistic DM_{GMM} coincides with the test statistic J_{GMM} for overidentifying restrictions.

Corollary 1 *Suppose that $r = q$. If Assumptions 1 and 2 are satisfied, then under the local alternatives*

$$J_{\text{GMM}} = n \hat{g}(\bar{\theta}_{\text{GMM}})' \hat{\Omega}^{-1} \hat{g}(\bar{\theta}_{\text{GMM}}) \xrightarrow{d} \chi^2(q_2, \kappa),$$

where $\kappa = \delta' \{K' \Omega^{-1} K - K' \Omega^{-1} H (H' \Omega^{-1} H)^{-1} H' \Omega^{-1} K\} \delta$.

3.4 Hypothesis testing about subsets of moment conditions

Suppose that the parameter β of $g(z; \beta, \gamma)$ is divided into an s_1 -vector $\beta^{(1)}$ and an s_2 -vector $\beta^{(2)}$. Let

$$g(z; \theta) = g(z; \beta, \gamma) = \begin{pmatrix} g_1(z; \beta^{(1)}) \\ g_2(z; \beta, \gamma) \end{pmatrix},$$

where $g_1(\cdot)$ is $r_1 \times 1$ with $r_1 \geq s_1$ and $g_2(\cdot)$ is $r_2 \times 1$ with $r_2 \geq s_2$. The condition $r_2 \geq s_2$ is essential because otherwise H is not of full rank. Let $\Omega_{11} = E[g_1 g_1']$, $H_1 = E[\partial g_1(z; \beta_0^{(1)}) / \partial \beta^{(1)}]$, and $K_1 = E[\partial g_1(z; \beta_0^{(1)}) / \partial \gamma]$, which is zero, of course.

The observed data z_i ($i = 1, \dots, n$) consists of p -random vectors which are independently distributed with a common probability density function $f(z; \theta_n, \eta_0)$ with $\theta_n \rightarrow \theta_0$. As in the preceding subsections, we will consider a null hypothesis of the form

$$H_0 : \gamma_n = \gamma_0,$$

and the local alternatives of the form

$$H_1 : \gamma_n = \gamma_0 + \delta/\sqrt{n},$$

where $\delta \in R^{q_2}$. Note that the moment condition

$$E_n[g_1(z; \beta_n^{(1)})] = \int g_1(z; \beta_n^{(1)}) f(z; \theta_n, \eta_0) d\mu(z) = 0$$

holds under the alternatives as well as the null hypothesis. This implies that only the subsets $g_2(z; \beta, \gamma)$ of the moment conditions are contaminated.

Let $\hat{g}(\theta) = n^{-1} \sum_{i=1}^n g(z_i; \theta)$ and $\hat{g}_1(\beta^{(1)}) = n^{-1} \sum_{i=1}^n g_1(z_i; \beta^{(1)})$. The first GMM estimator is defined by

$$\bar{\theta}_{\text{GMM}} = \arg \min_{\theta \in \Theta} \hat{g}(\theta)' \hat{\Omega}^{-1} \hat{g}(\theta) \quad \text{subject to } \gamma = \gamma_0,$$

where $\hat{\Omega}$ is some preliminary estimator satisfying $\hat{\Omega} \xrightarrow{p} \Omega$. Let

$$\mathcal{B} = \Theta \cap (R^{s_1} \times \{\beta_0^{(2)}\} \times \{\gamma_0\}).$$

The second GMM estimator is given by

$$\tilde{\beta}_{\text{GMM}}^{(1)} = \arg \min_{\beta^{(1)} \in \mathcal{B}} \hat{g}_1(\beta^{(1)})' \hat{\Omega}_{11}^{-1} \hat{g}_1(\beta^{(1)}),$$

where $\hat{\Omega}_{11}$ is some preliminary estimator satisfying $\hat{\Omega}_{11} \xrightarrow{p} \Omega_{11}$.

Assumption 4 (a) $\beta_0^{(1)} \in \mathcal{B}$ is the unique solution to $E[g_1(z; \beta^{(1)})] = 0$; (b) Ω_{11} is nonsingular; (c) $\text{rank}(H_1) = s_1$.

Lemma 5 If Assumptions 2, 3, and 4 are satisfied, then under the local alternatives $\tilde{\beta}_{\text{GMM}}^{(1)} \xrightarrow{p} \beta_0^{(1)}$ and

$$\sqrt{n} \hat{g}_1(\beta_n^{(1)}) \xrightarrow{d} N(0, \Omega_{11}), \quad (3.14)$$

$$\sqrt{n}(\tilde{\beta}_{\text{GMM}}^{(1)} - \beta_n^{(1)}) = -(H_1' \Omega_{11}^{-1} H_1)^{-1} H_1' \Omega_{11}^{-1} \sqrt{n} \hat{g}_1(\beta_n^{(1)}) + o_p(1), \quad (3.15)$$

$$\sqrt{n}\hat{g}_1(\tilde{\beta}_{\text{GMM}}^{(1)}) = \{I_{r_1} - H_1(H_1'\Omega_{11}^{-1}H_1)^{-1}H_1'\Omega_{11}^{-1}\}\sqrt{n}\hat{g}_1(\beta_n^{(1)}) + o_p(1). \quad (3.16)$$

Proof: Note that $\mathcal{B}$ is compact since Θ is compact by (b) of Assumption 2. It follows similarly to the proof of Theorem 8 that $\tilde{\beta}_{\text{GMM}}^{(1)} \xrightarrow{p} \beta_0^{(1)}$. Note that $\beta_0^{(1)} \in \text{int}(\mathcal{B})$ since $\theta_0 \in \text{int}(\Theta)$ by (a) of Assumption 3. Equations (3.14), (3.15), and (3.16) follow similarly to the proof of Theorem 9. ■

Eichenbaum, Hansen, and Singleton (1988) proposed the test statistic C_{GMM} , which aims to detect the contamination in the subsets $g_2(z; \beta, \gamma)$ of the moment conditions. The following theorem gives the asymptotic property of C_{GMM} under the local alternatives, which they did not give.

Theorem 12 *If Assumptions 2, 3, and 4 are satisfied, then under the local alternatives*

$$C_{\text{GMM}} = n \left\{ \hat{g}(\bar{\theta}_{\text{GMM}})' \hat{\Omega}^{-1} \hat{g}(\bar{\theta}_{\text{GMM}}) - \hat{g}_1(\tilde{\beta}_{\text{GMM}}^{(1)})' \hat{\Omega}_{11}^{-1} \hat{g}_1(\tilde{\beta}_{\text{GMM}}^{(1)}) \right\} \\ \xrightarrow{d} \chi^2(r_2 - s_2, \kappa),$$

where $\kappa = \delta' \{K' \Omega^{-1} K - K' \Omega^{-1} H (H' \Omega^{-1} H)^{-1} H' \Omega^{-1} K\} \delta$.

Proof: By equation (3.6) we have

$$\Omega^{-1/2} \sqrt{n} \hat{g}(\bar{\theta}_{\text{GMM}}) = N \Omega^{-1/2} \sqrt{n} \hat{g}(\beta_n, \gamma_0) + o_p(1),$$

where $N = I_r - \Omega^{-1/2} H (H' \Omega^{-1} H)^{-1} H' \Omega^{-1/2}$. Equation (3.16) yields

$$\Omega_{11}^{-1/2} \sqrt{n} \hat{g}_1(\tilde{\beta}_{\text{GMM}}^{(1)}) = M \Omega_{11}^{-1/2} \sqrt{n} \hat{g}_1(\beta_n^{(1)}) + o_p(1),$$

where $M = I_{r_1} - \Omega_{11}^{-1/2} H_1 (H_1' \Omega_{11}^{-1} H_1)^{-1} H_1' \Omega_{11}^{-1/2}$. Thus under the local alternatives, the test statistic C_{GMM} has the same asymptotic distribution as

$$\{\Omega^{-1/2} \sqrt{n} \hat{g}(\beta_n, \gamma_0)\}' A \{\Omega^{-1/2} \sqrt{n} \hat{g}(\beta_n, \gamma_0)\},$$

where

$$A = N - \Omega^{1/2} \begin{pmatrix} I_{r_1} \\ 0 \end{pmatrix} \Omega_{11}^{-1/2} M \Omega_{11}^{-1/2} (I_{r_1} \ 0) \Omega^{1/2}.$$

Straightforward calculations show that A is idempotent with rank equal to $r_2 - s_2$. Since $\Omega^{-1/2} \sqrt{n} \hat{g}(\beta_n, \gamma_0) \xrightarrow{d} N(-\Omega^{-1/2} K \delta, I_r)$ with $K_1 = 0$, the conclusion follows from Theorem 7. ■

The asymptotic power of GMM hypothesis tests is determined by the non-centrality parameter κ and the degrees of freedom, since the tail probability of a noncentral chi-square distribution is increasing in the noncentrality parameter and decreasing in the degrees of freedom.

The ordinary test statistic J_{GMM} for overidentifying restrictions may work in order to detect the contamination in the subsets $g_2(z; \beta, \gamma)$ of the moment conditions. However, the degrees of freedom of the test is $r - q_1$ according to Theorem 11, while that of C_{GMM} is $r_2 - s_2$. Since $(r - q_1) - (r_2 - s_2) = r_1 - s_1 \geq 0$, the power of C_{GMM} is greater.

4 Hypothesis testing with GEL estimators

Let $\rho(v)$ be a function of a scalar v that is concave on its domain, an open interval $\mathcal{V}$ containing zero, and twice continuously differentiable in a neighborhood of zero. Let $\hat{\Lambda}_n(\theta) = \{\lambda : \lambda' g(z_i; \theta) \in \mathcal{V}, i = 1, \dots, n\}$. The GEL estimator is the solution to a saddle point problem

$$\hat{\theta}_{\text{GEL}} = \arg \min_{\theta \in \Theta} \sup_{\lambda \in \hat{\Lambda}_n(\theta)} \sum_{i=1}^n \rho(\lambda' g(z_i; \theta)).$$

Let $\hat{\lambda} = \arg \max_{\lambda \in \hat{\Lambda}_n(\hat{\theta}_{\text{GEL}})} \sum_{i=1}^n \rho(\lambda' g(z_i; \hat{\theta}_{\text{GEL}}))$ and $\rho_j(v) = \partial^j \rho(v) / \partial v^j$ ($j = 1, 2, \dots$). We normalize so that $\rho_1(0) = \rho_2(0) = -1$. EL is a special case with $\rho(v) = \ln(1 - v)$ and $\mathcal{V} = (-\infty, 1)$. Associated with each GEL estimator empirical probabilities are given by

$$\hat{\pi}_i = \frac{\rho_1(\hat{\lambda}' g(z_i; \hat{\theta}_{\text{GEL}}))}{\sum_{j=1}^n \rho_1(\hat{\lambda}' g(z_j; \hat{\theta}_{\text{GEL}}))} \quad (i = 1, 2, \dots, n),$$

which yield an estimate of the distribution function F as

$$\hat{F}_n(z) = \sum_{i=1}^n \hat{\pi}_i 1_{(z_i < z)}.$$

It can be shown that $\hat{F}_n(z)$ is consistent and asymptotically normal. The proof is omitted in this paper. See Lemma 2 of Goto (2021).

The observed data z_i ($i = 1, \dots, n$) consists of p -random vectors which are independently distributed with a common probability density function

$f(z; \theta_n, \eta_0)$ with $\theta_n \rightarrow \theta_0$. As in the previous section, we will consider a null hypothesis of the form

$$H_0 : \gamma_n = \gamma_0.$$

To measure the power of the test, we introduce the following local alternatives of the form

$$H_1 : \gamma_n = \gamma_0 + \delta/\sqrt{n},$$

where $\delta \in R^{q_2}$. In this section, we consider the hypothesis test using the GEL estimators. The constrained GEL estimator is defined by

$$\bar{\theta}_{\text{GEL}} = \arg \min_{\theta \in \Theta} \sup_{\lambda \in \hat{\Lambda}_n(\theta)} \sum_{i=1}^n \rho(\lambda' g(z_i; \theta)) \quad \text{subject to } \gamma = \gamma_0. \quad (4.1)$$

Let $\bar{\lambda} = \arg \max_{\lambda \in \hat{\Lambda}_n(\bar{\theta}_{\text{GEL}})} \sum_{i=1}^n \rho(\lambda' g(z_i; \bar{\theta}_{\text{GEL}}))$. On the other hand, the unconstrained GEL estimator is given by

$$\hat{\theta}_{\text{GEL}} = \arg \min_{\theta \in \Theta} \sup_{\lambda \in \hat{\Lambda}_n(\theta)} \sum_{i=1}^n \rho(\lambda' g(z_i; \theta)).$$

Estimators $\bar{\theta}_{\text{GEL}}$ and $\hat{\theta}_{\text{GEL}}$ correspond to the GEL estimators for the restricted and unrestricted models, respectively.

Proofs of theorems in this section are based on those in Newey and Smith (2004) with slight modifications for non-i.i.d. data. Throughout this section, C will denote a generic positive constant that may be different in different uses, and CS the Cauchy-Schwarz inequality. Also, UWL will denote a uniform weak law of large numbers, and CLT will refer to the Lindeberg-Feller central limit theorem.

4.1 Consistency of estimators under local alternatives

Here, we show consistency of $\bar{\theta}_{\text{GEL}}$ under the local alternatives. Consistency of $\hat{\theta}_{\text{GEL}}$ follows similarly.

Assumption 5 (a) $\theta_0 \in \Theta$ is the unique solution to $E[g(z; \theta)] = 0$; (b) Θ is compact; (c) $g(z; \theta)$ and $f(z; \theta, \eta_0)$ are continuous at each $\theta \in \Theta$ for almost all z ; (d) There exist measurable functions $b_1(z)$ and $b_2(z)$ such that for almost all z , $\|g(z; \theta)\|^\alpha < b_1(z)$ ($\alpha > 2$) and $f(z; \theta, \eta_0) < b_2(z)$ for all $\theta \in \Theta$ with $\int b_1(z)b_2(z)d\mu(z) < \infty$; (e) Ω is nonsingular.

Let $\hat{P}(\theta, \lambda) = n^{-1} \sum_{i=1}^n \rho(\lambda' g(z_i; \theta))$ and $a_i = \sup_{\theta \in \Theta} \|g(z_i; \theta)\|$.

Lemma 6 *If Assumption 5 is satisfied, then for any ζ with $1/\alpha < \zeta < 1/2$ and $\Lambda_n = \{\lambda : \|\lambda\| \leq n^{-\zeta}\}$, $\sup_{\theta \in \Theta, \lambda \in \Lambda_n, 1 \leq i \leq n} |\lambda' g(z_i; \theta)| \xrightarrow{P} 0$ and w.p.a.1, $\Lambda_n \subset \hat{\Lambda}_n(\theta)$ for all $\theta \in \Theta$.*

Proof: By (d) of Assumption 5,

$$\begin{aligned} P\left(\max_{1 \leq i \leq n} a_i^\alpha > n\varepsilon\right) &\leq nP(a_1^\alpha > n\varepsilon) \\ &\leq \frac{1}{\varepsilon} \int 1_{[a_1^\alpha > n\varepsilon]} a_1^\alpha f(z_1; \theta_n, \eta_0) d\mu(z_1) \\ &\leq \frac{1}{\varepsilon} \int 1_{[b_1(z_1) > n\varepsilon]} b_1(z_1) b_2(z_1) d\mu(z_1) < \infty. \end{aligned}$$

By dominated convergence, we have $\max_{1 \leq i \leq n} a_i = o_p(n^{1/\alpha})$. Then by CS,

$$\sup_{\theta \in \Theta, \lambda \in \Lambda_n, 1 \leq i \leq n} |\lambda' g(z_i; \theta)| \leq n^{-\zeta} \max_{1 \leq i \leq n} a_i = o_p(n^{-\zeta+1/\alpha}) \xrightarrow{P} 0,$$

giving the first conclusion, so w.p.a.1 $\lambda' g(z_i, \theta) \in \mathcal{V}$ for all $\theta \in \Theta$ and $\|\lambda\| \leq n^{-\zeta}$. ■

Let $\hat{\Omega}(\theta) = n^{-1} \sum_{i=1}^n g(z_i; \theta) g(z_i; \theta)'$.

Lemma 7 *If Assumption 5 is satisfied, $\check{\theta} \in \Theta$, $\check{\theta} \xrightarrow{P} \theta_0$, and $\hat{g}(\check{\theta}) = O_p(n^{-1/2})$, then $\check{\lambda} = \arg \max_{\lambda \in \hat{\Lambda}_n(\check{\theta})} \hat{P}(\check{\theta}, \lambda)$ exists w.p.a.1, $\check{\lambda} = O_p(n^{-1/2})$, and*

$$\sup_{\lambda \in \hat{\Lambda}_n(\check{\theta})} \hat{P}(\check{\theta}, \lambda) \leq \rho(0) + O_p(n^{-1}).$$

Proof: It follows similarly to the proof of Lemma 2 that UWL for $\hat{\Omega}(\theta)$ holds, giving $\check{\Omega} = \hat{\Omega}(\check{\theta}) \xrightarrow{P} \Omega$. Then by nonsingularity of Ω the smallest eigenvalue of $\check{\Omega}$ is bounded away from zero w.p.a.1. Let Λ_n be as defined in Lemma 6. By Lemma 6 and twice continuous differentiability of $\rho(v)$ in a neighborhood of zero, $\hat{P}(\check{\theta}, \lambda)$ is twice continuously differentiable on Λ_n w.p.a.1. Since $\hat{P}(\check{\theta}, \lambda)$ is continuous with Λ_n compact, $\check{\lambda} = \arg \max_{\lambda \in \Lambda_n} \hat{P}(\check{\theta}, \lambda)$ exists w.p.a.1. Furthermore, for $\check{g}_i = g(z_i; \check{\theta})$ and any $\dot{\lambda}$ on the line joining $\check{\lambda}$ and 0, by Lemma 6 and $\rho_2(0) = -1$,

$$\max_{1 \leq i \leq n} \rho_2(\dot{\lambda}' \check{g}_i) < -1/2 \quad \text{w.p.a.1.}$$

Then by a Taylor expansion around $\lambda = 0$ with Lagrange remainder, there is $\check{\lambda}$ on the line joining $\tilde{\lambda}$ and 0 such that for $\check{g} = \hat{g}(\check{\theta})$,

$$\begin{aligned}\rho(0) &= \hat{P}(\check{\theta}, 0) \leq \hat{P}(\check{\theta}, \tilde{\lambda}) = \rho(0) - \tilde{\lambda}'\check{g} + (1/2)\tilde{\lambda}' \left[\sum_{i=1}^n \rho_2(\check{\lambda}'\check{g}_i)\check{g}_i\check{g}_i'/n \right] \tilde{\lambda} \\ &\leq \rho(0) - \tilde{\lambda}'\check{g} - (1/4)\tilde{\lambda}'\check{\mathcal{G}}\tilde{\lambda} \leq \rho(0) + \|\tilde{\lambda}\|\|\check{g}\| - C\|\tilde{\lambda}\|^2.\end{aligned}$$

Subtracting $\rho(0) - C\|\tilde{\lambda}\|^2$ from both sides and dividing by $\|\tilde{\lambda}\|$ we find that $C\|\tilde{\lambda}\| \leq \|\check{g}\|$, w.p.a.1. By assumption, $\check{g} = O_p(n^{-1/2})$, and hence $\|\tilde{\lambda}\| = O_p(n^{-1/2}) = o_p(n^{-\zeta})$. Therefore, w.p.a.1 $\tilde{\lambda} \in \text{int}(\Lambda_n)$ and hence $\partial \hat{P}(\check{\theta}, \tilde{\lambda})/\partial \lambda = 0$, the first order conditions for an interior maximum. By Lemma 6, w.p.a.1 $\tilde{\lambda} \in \hat{\Lambda}_n(\check{\theta})$, so by concavity of $\hat{P}(\check{\theta}, \lambda)$ and convexity of $\hat{\Lambda}_n(\check{\theta})$ it follows that $\hat{P}(\check{\theta}, \tilde{\lambda}) = \sup_{\lambda \in \hat{\Lambda}_n(\check{\theta})} \hat{P}(\check{\theta}, \lambda)$, giving the first and second conclusions with $\check{\lambda} = \tilde{\lambda}$. Then by the last inequality of the above equation, $\|\check{g}\| = O_p(n^{-1/2})$, and $\|\tilde{\lambda}\| = O_p(n^{-1/2})$, we obtain $\hat{P}(\check{\theta}, \check{\lambda}) \leq \rho(0) + \|\check{\lambda}\|\|\check{g}\| - C\|\check{\lambda}\|^2 = \rho(0) + O_p(n^{-1})$. This completes the proof. ■

Lemma 8 *If Assumption 5 is satisfied, then $\|\hat{g}(\bar{\theta}_{\text{GEL}})\| = O_p(n^{-1/2})$.*

Proof: Let $\bar{g}_i = g(z_i; \bar{\theta}_{\text{GEL}})$, $\bar{g} = \hat{g}(\bar{\theta}_{\text{GEL}})$, and for ζ in Lemma 6, $\check{\lambda} = -n^{-\zeta}\bar{g}/\|\bar{g}\|$. By Lemma 6, $\max_{i \leq n} |\check{\lambda}'\bar{g}_i| \xrightarrow{P} 0$ and $\check{\lambda} \in \hat{\Lambda}_n(\bar{\theta}_{\text{GEL}})$ w.p.a.1. Thus for any $\check{\lambda}$ on the line joining $\check{\lambda}$ and 0, w.p.a.1 $\rho_2(\check{\lambda}'\bar{g}_i) \geq -C$ ($i = 1, \dots, n$). Also, by CS and UWL, $\sum_i \bar{g}_i\bar{g}_i'/n \leq (\sum_i a_i^2/n)I_r \xrightarrow{P} CI_r$, so that the largest eigenvalue of $\sum_i \bar{g}_i\bar{g}_i'/n$ is bounded above w.p.a.1. An expansion then gives

$$\begin{aligned}\hat{P}(\bar{\theta}_{\text{GEL}}, \check{\lambda}) &= \rho(0) - \check{\lambda}'\bar{g} + (1/2)\check{\lambda}' \left[\sum_{i=1}^n \rho_2(\check{\lambda}'\bar{g}_i)\bar{g}_i\bar{g}_i'/n \right] \check{\lambda} \\ &\geq \rho(0) + n^{-\zeta}\|\bar{g}\| - C(1/2)\check{\lambda}' \left[\sum_{i=1}^n \bar{g}_i\bar{g}_i'/n \right] \check{\lambda} \\ &\geq \rho(0) + n^{-\zeta}\|\bar{g}\| - Cn^{-2\zeta}\end{aligned}$$

w.p.a.1. The hypotheses of Lemma 7 are satisfied by $\check{\theta} = \bar{\theta}_n \equiv (\beta'_n, \gamma'_0)'$ using the CLT as in the proof of Theorem 9. By $\bar{\theta}_{\text{GEL}}$ and $\tilde{\lambda}$ being a saddle point, this

equation and Lemma 7 give

$$\begin{aligned} \rho(0) + n^{-\zeta} \|\bar{g}\| - Cn^{-2\zeta} &\leq \hat{P}(\bar{\theta}_{\text{GEL}}, \hat{\lambda}) \leq \hat{P}(\bar{\theta}_{\text{GEL}}, \bar{\lambda}) \\ &\leq \sup_{\lambda \in \hat{\Lambda}_n(\bar{\theta}_n)} \hat{P}(\bar{\theta}_n, \lambda) \leq \rho(0) + O_p(n^{-1}). \end{aligned} \quad (4.2)$$

Also, by $\zeta < 1/2$, $\zeta - 1 < -1/2 < -\zeta$. Solving equation (4.2) for $\|\bar{g}\|$ then gives

$$\|\bar{g}\| \leq O_p(n^{\zeta-1}) + Cn^{-\zeta} = O_p(n^{-\zeta}). \quad (4.3)$$

Now, consider any $\varepsilon_n \rightarrow 0$. Let $\hat{\lambda} = -\varepsilon_n \bar{g}$. Note that $\hat{\lambda} = o_p(n^{-\zeta})$ by equation (4.3), so that $\hat{\lambda} \in \Lambda_n$ w.p.a.1. Then as in equation (4.2),

$$\begin{aligned} \rho(0) + \varepsilon_n \|\bar{g}\|^2 - C\varepsilon_n^2 \|\bar{g}\|^2 &= \rho(0) - \hat{\lambda}' \bar{g} - C\|\hat{\lambda}\|^2 \\ &\leq \hat{P}(\bar{\theta}_{\text{GEL}}, \hat{\lambda}) \leq \rho(0) + O_p(n^{-1}). \end{aligned}$$

Since, for all n large enough, $1 - \varepsilon_n C$ is bounded away from zero, it follows that $\varepsilon_n \|\bar{g}\|^2 = O_p(n^{-1})$, which is equivalent to $n\varepsilon_n \|\bar{g}\|^2/w_n \xrightarrow{P} 0$ for all $w_n \rightarrow \infty$. Then for all $w_n^* \rightarrow \infty$, $n\|\bar{g}\|^2/w_n^* \xrightarrow{P} 0$ holds by putting $\varepsilon_n = 1/\sqrt{w_n^*}$ and $w_n = \sqrt{w_n^*}$. Hence, we have $\|\bar{g}\|^2 = O_p(n^{-1})$, too. ■

Theorem 13 *If Assumption 5 is satisfied, then under the local alternatives $\bar{\theta}_{\text{GEL}} \xrightarrow{P} \theta_0$ with $\bar{\lambda} = O_p(n^{-1/2})$ and $\hat{\theta}_{\text{GEL}} \xrightarrow{P} \theta_0$ with $\hat{\lambda} = O_p(n^{-1/2})$.*

Proof: To see $\bar{\theta}_{\text{GEL}} \xrightarrow{P} \theta_0$, let $M_n(\theta) = -\|\hat{g}(\theta)\|$ and $M(\theta) = -\|g_0(\theta)\|$, where $M(\theta) = 0$ has a unique solution at θ_0 . Similarly to the proof of Lemma 2, $\sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)| \xrightarrow{P} 0$ holds by Assumption 5. Also $M_n(\bar{\theta}_{\text{GEL}}) \geq M_n(\theta_0) - o_p(1)$ holds by Lemma 8. The conclusion follows as in the proof of Theorem 8. By Lemma 8 and thus putting $\check{\theta} = \bar{\theta}_{\text{GEL}}$ in Lemma 7, we obtain $\bar{\lambda} = O_p(n^{-1/2})$. Corresponding results with respect to $\hat{\theta}_{\text{GEL}}$ can be shown similarly. ■

4.2 Asymptotic normality of estimators under local alternatives

The following assumption combined with Assumption 5 guarantees the asymptotic normality of the GEL estimator under the local alternatives.

Assumption 6 (a) $\theta_0 \in \text{int}(\Theta)$; (b) $g(z; \theta)$ is continuously differentiable in a neighborhood $\mathcal{N}$ of θ_0 for almost all z ; (c) There exist measurable functions $b_1(z)$ and $b_2(z)$ such that for almost all z , $\|\partial g(z; \theta)/\partial \theta\| < b_1(z)$ and $f(z; \theta, \eta_0) < b_2(z)$ for all $\theta \in \Theta$. Further, it is the case that $\int b_1(z)b_2(z)d\mu(z) < \infty$; (d) $\text{rank}(G) = q$.

Let $\bar{\theta}_{\text{GEL}} = (\bar{\beta}', \gamma_0')'$.

Theorem 14 If Assumptions 5 and 6 are satisfied, then under the local alternatives

$$\sqrt{n} \begin{pmatrix} \bar{\beta} - \beta_n \\ \bar{\lambda} \end{pmatrix} = - \begin{pmatrix} (H' \Omega^{-1} H)^{-1} H' \Omega^{-1} \\ \Omega^{-1} - \Omega^{-1} H (H' \Omega^{-1} H)^{-1} H' \Omega^{-1} \end{pmatrix} \sqrt{n} \hat{g}(\beta_n, \gamma_0) + o_p(1), \quad (4.4)$$

$$\sqrt{n} \hat{g}(\bar{\theta}_{\text{GEL}}) = \{I_r - H (H' \Omega^{-1} H)^{-1} H' \Omega^{-1}\} \sqrt{n} \hat{g}(\beta_n, \gamma_0) + o_p(1), \quad (4.5)$$

$$\sqrt{n} \hat{g}(\beta_n, \gamma_0) = \sqrt{n} \hat{g}(\theta_n) - K \delta + o_p(1), \quad (4.6)$$

$$\sqrt{n} \hat{g}(\theta_n) \xrightarrow{d} N(0, \Omega), \quad (4.7)$$

$$\sum_{i=1}^n \{ \rho(\bar{\lambda}' g(z_i; \bar{\theta}_{\text{GEL}})) - \rho(0) \} = \frac{n}{2} \hat{g}(\bar{\theta}_{\text{GEL}})' \Omega^{-1} \hat{g}(\bar{\theta}_{\text{GEL}}) + o_p(1), \quad (4.8)$$

$$\sqrt{n} \begin{pmatrix} \hat{\theta}_{\text{GEL}} - \theta_n \\ \hat{\lambda} \end{pmatrix} = - \begin{pmatrix} (G' \Omega^{-1} G)^{-1} G' \Omega^{-1} \\ \Omega^{-1} - \Omega^{-1} G (G' \Omega^{-1} G)^{-1} G' \Omega^{-1} \end{pmatrix} \sqrt{n} \hat{g}(\theta_n) + o_p(1), \quad (4.9)$$

$$\sqrt{n} \hat{g}(\hat{\theta}_{\text{GEL}}) = \{I_r - G (G' \Omega^{-1} G)^{-1} G' \Omega^{-1}\} \sqrt{n} \hat{g}(\theta_n) + o_p(1), \quad (4.10)$$

$$\sum_{i=1}^n \{ \rho(\hat{\lambda}' g(z_i; \hat{\theta}_{\text{GEL}})) - \rho(0) \} = \frac{n}{2} \hat{g}(\hat{\theta}_{\text{GEL}})' \Omega^{-1} \hat{g}(\hat{\theta}_{\text{GEL}}) + o_p(1). \quad (4.11)$$

Proof: Equations (4.6) and (4.7) follow similarly to the proof of Theorem 9. To see equation (4.4) hold, we will compute the GEL first order conditions in equation (4.1). The first order conditions for $\bar{\lambda}$ are

$$\sum_{i=1}^n \rho_1(\bar{\lambda}' \bar{g}_i) \bar{g}_i / n = 0. \quad (4.12)$$

Equation (4.12) holds since $\bar{\lambda} \in \hat{\Lambda}_n(\bar{\theta}_{\text{GEL}})$ with $\hat{\Lambda}_n(\bar{\theta}_{\text{GEL}})$ open by definition. Also, $\tilde{\Omega} = \sum_{i=1}^n \rho_2(\bar{\lambda}' \bar{g}_i) \bar{g}_i \bar{g}_i' / n \xrightarrow{P} \rho_2(0) \Omega$, so that $\tilde{\Omega}$ is nonsingular w.p.a.1. Then by the implicit function theorem there is a neighborhood of $\bar{\beta}$ where the solution $\lambda(\beta)$ to $\sum_{i=1}^n \rho_1(\lambda' g(z_i; \beta, \gamma_0)) g(z_i; \beta, \gamma_0) / n = 0$ exists and is continuously differentiable. Let $g_i(\beta) = g(z_i; \beta, \gamma_0)$ and $H_i(\beta) = \partial g(z_i; \beta, \gamma_0) / \partial \beta$. By the envelope theorem the first order conditions for $\bar{\beta}$ are

$$\sum_{i=1}^n \rho_1(\bar{\lambda}' \bar{g}_i) H_i(\bar{\beta})' \bar{\lambda} / n = 0. \quad (4.13)$$

Equation (4.13) follows w.p.a.1 from (a) of Assumption 6 with Theorem 13. By a mean value expansion of these first order conditions we have, for $\bar{\phi} = (\bar{\beta}', \bar{\lambda}')'$ and $\phi_n = (\beta_n', 0)'$,

$$0 = \begin{pmatrix} 0 \\ -\hat{g}(\beta_n, \gamma_0) \end{pmatrix} + \bar{M}(\bar{\phi} - \phi_n), \quad (4.14)$$

$$\bar{M} = \begin{pmatrix} 0 & \sum_{i=1}^n \rho_1(\bar{\lambda}' \bar{g}_i) H_i(\bar{\beta})' / n \\ \sum_{i=1}^n \rho_1(\bar{\lambda}' \bar{g}_i) H_i(\bar{\beta})' / n & \sum_{i=1}^n \rho_2(\bar{\lambda}' \bar{g}_i) \bar{g}_i \bar{g}_i' / n \end{pmatrix},$$

where $\bar{\beta}$ and $\bar{\lambda}$ are mean values that actually differ from row to row of the matrix $\bar{M}$. By $\bar{\lambda} = O_p(n^{-1/2})$, it follows as in Lemma 6 that $\max_{i \leq n} |\bar{\lambda}' \bar{g}_i| \xrightarrow{P} 0$. Therefore, $\max_{i \leq n} |\rho_1(\bar{\lambda}' \bar{g}_i) + 1| \xrightarrow{P} 0$ and $\max_{i \leq n} |\rho_2(\bar{\lambda}' \bar{g}_i) + 1| \xrightarrow{P} 0$. It then follows from UWL that $\bar{M} \xrightarrow{P} M$. Here

$$M = - \begin{pmatrix} 0 & H' \\ H & \Omega \end{pmatrix}, \quad M^{-1} = - \begin{pmatrix} -\Sigma & J \\ J' & P \end{pmatrix},$$

where $\Sigma = (H' \Omega^{-1} H)^{-1}$, $J = \Sigma H' \Omega^{-1}$, and $P = \Omega^{-1} - \Omega^{-1} H \Sigma H' \Omega^{-1}$. Inverting and solving in equation (4.14) then gives

$$\begin{aligned} \sqrt{n}(\bar{\phi} - \phi_n) &= -\bar{M}^{-1}(0, -\sqrt{n}\hat{g}(\beta_n, \gamma_0)')' \\ &= -M^{-1}(0, -\sqrt{n}\hat{g}(\beta_n, \gamma_0)')' + o_p(1) \\ &= -(J', P)' \sqrt{n}\hat{g}(\beta_n, \gamma_0) + o_p(1). \end{aligned}$$

Then equation (4.4) follows. Equation (4.5) follows from the mean value theorem and equation (4.4). By equations (4.4) and (4.5) we have

$$\hat{g}(\bar{\theta}_{\text{GEL}}) = -\Omega \bar{\lambda} + o_p(n^{-1/2}).$$

Then

$$\begin{aligned} \hat{P}(\bar{\theta}_{\text{GEL}}, \bar{\lambda}) &= \rho(0) - \bar{\lambda}' \hat{g}(\bar{\theta}_{\text{GEL}}) + \bar{\lambda}' \left[\sum_{i=1}^n \rho_2(\bar{\lambda}' \bar{g}_i) \bar{g}_i \bar{g}_i' / n \right] \bar{\lambda} / 2 \\ &= \rho(0) - \bar{\lambda}' \hat{g}(\bar{\theta}_{\text{GEL}}) - \bar{\lambda}' \Omega \bar{\lambda} / 2 + o_p(n^{-1}) \\ &= \rho(0) + \hat{g}(\bar{\theta}_{\text{GEL}})' \Omega^{-1} \hat{g}(\bar{\theta}_{\text{GEL}}) / 2 + o_p(n^{-1}), \end{aligned}$$

which gives equation (4.8). Equations (4.9)-(4.11) can be shown similarly. ■

4.3 Hypothesis testing

We will construct the GEL hypothesis test in terms of the distance metric statistic DM_{GEL} . Let κ denote the noncentrality parameter of the noncentral chi-square distribution corresponding to the test.

Theorem 15 *If Assumptions 5 and 6 are satisfied, then under the local alternatives*

$$\text{DM}_{\text{GEL}} = 2 \left\{ \sum_{i=1}^n \rho(\bar{\lambda}' g(z_i; \bar{\theta}_{\text{GEL}})) - \sum_{i=1}^n \rho(\hat{\lambda}' g(z_i; \hat{\theta}_{\text{GEL}})) \right\} \xrightarrow{d} \chi^2(q_2, \kappa),$$

where $\kappa = \delta' \{K' \Omega^{-1} K - K' \Omega^{-1} H (H' \Omega^{-1} H)^{-1} H' \Omega^{-1} K\} \delta$.

Proof: By equations (4.8) and (4.11) we have

$$\text{DM}_{\text{GEL}} = n \left\{ \hat{g}(\bar{\theta}_{\text{GEL}})' \Omega^{-1} \hat{g}(\bar{\theta}_{\text{GEL}}) - \hat{g}(\hat{\theta}_{\text{GEL}})' \Omega^{-1} \hat{g}(\hat{\theta}_{\text{GEL}}) \right\} + o_p(1).$$

Theorems 9 and 14 give

$$\sqrt{n} \hat{g}(\bar{\theta}_{\text{GEL}}) = \sqrt{n} \hat{g}(\bar{\theta}_{\text{GMM}}) + o_p(1) = O_p(1)$$

and

$$\sqrt{n} \hat{g}(\hat{\theta}_{\text{GEL}}) = \sqrt{n} \hat{g}(\hat{\theta}_{\text{GMM}}) + o_p(1) = O_p(1),$$

which yields $\text{DM}_{\text{GEL}} = \text{DM}_{\text{GMM}} + o_p(1)$. The conclusion follows from Theorem 10. ■

REFERENCES

- Billingsley, P. (1986): *Probability and Measure* (2nd ed.). New York: John Wiley and Sons.
- Durrett, R. (2010): *Probability: Theory and Examples* (4th ed.). Cambridge: Cambridge University Press.
- Eichenbaum, M. S., L. P. Hansen, and K. J. Singleton (1988): “A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice Under Uncertainty,” *The Quarterly Journal of Economics*, 103, 51–78.
- Goto, F. (2021): “Large Sample Properties of GMM and GEL Estimators: A Geometric Approach,” *The Aoyama Journal of Economics*, Vol. 73, No. 1, 1–33.
- Hall, A. R. (1999): “Hypothesis Testing in Models Estimated by GMM,” in *Generalized Method of Moments Estimation*, ed. by L. Mátyás. Cambridge: Cambridge University Press, 96–127.
- Hamilton, J. (1994): *Time Series Analysis*. Princeton: Princeton University Press.
- Hansen, L. P. (1982): “Large Sample Properties of Generalized Method of Moments Estimators,” *Econometrica*, 50, 1029–1054.
- Hayashi, F. (2000): *Econometrics*. Princeton: Princeton University Press.
- Holly, A. (1982): “A Remark on Hausman’s Specification test,” *Econometrica*, 50, 749–759.
- Newey, W. K. (1985a): “Generalized Method of Moments Specification Testing,” *Journal of Econometrics*, 29, 229–256.
- Newey, W. K. (1985b): “Maximum Likelihood Specification Testing and Conditional Moment Tests,” *Econometrica*, 53, 1047–1070.
- Newey, W. K., and D. McFadden (1994): “Large Sample Estimation and Hypothesis Testing,” in *Handbook of Econometrics*, Vol. 4, ed. by R. Engel and D. McFadden. New York: North Holland, 2111–2245.
- Newey, W. K., and R. J. Smith (2004): “Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators,” *Econometrica*, 72, 219–255.
- Newey, W. K., and K. D. West (1987): “Hypothesis Testing with Efficient Method of Moments Estimation,” *International Economic Review*, 28, 777–787.
- Owen, A. B. (1988): “Empirical Likelihood Ratio Confidence Intervals for a Single Functional,” *Biometrika*, 75, 237–249.
- Owen, A. B. (2001): *Empirical Likelihood*. London: Chapman and Hall.
- Qin, J., and J. Lawless (1994): “Empirical Likelihood and General Estimating Equations,” *Annals of Statistics*, 22, 300–325.
- Rao, C. R. (1973): *Linear Statistical Inference and Its Applications* (2nd ed.). New York: John Wiley and Sons.
- van de Geer, S. (2000): *Empirical Processes in M-Estimation*. Cambridge: Cambridge University Press.
- van der Vaart, A. W. (1998): *Asymptotic Statistics*. Cambridge: Cambridge University Press.