

Large Sample Properties of GMM and GEL Estimators: A Geometric Approach

Fumihito Goto

Contents

- 1 Introduction
- 2 The model and efficiency bounds
- 3 GMM and GEL estimators: Basic results
- 4 Hypothesis testing
- Appendix

1 Introduction

Consider the pair $(g(z; \theta), f(z; \theta, \eta))$ of functions, where $f(z; \theta, \eta)$ is a density function with a parameter of interest θ and a nuisance parameter η , while $g(z; \theta)$ is a vector function that satisfies the moment condition

$$\int g(z; \theta) f(z; \theta, \eta) d\mu(z) = 0.$$

We observe z and know only $g(z; \theta)$, the moment function. Then generalized method of moments (GMM) and generalized empirical likelihood (GEL) estimators are two major methods to estimate the parameter θ .

Large sample properties of these estimators have been studied by various researchers so far. First, Hansen (1982) proved consistency and asymptotic normality of GMM estimators. He also proposed the test of overidentifying restrictions. Hypothesis testing with GMM estimators was studied by Newey (1985a, b), Newey and West (1987), Hall (1999), and others. Chamberlain (1987) explored asymptotic efficiency of GMM estimators with conditional moment restrictions as well as moment restrictions. On the other hand, Qin

and Lawless (1994) investigated consistency, asymptotic normality, and asymptotic efficiency of empirical likelihood (EL) estimators first introduced by Owen (1988). Later, Donald, Imbens, and Newey (2003) and Newey and Smith (2004) proposed GEL, which includes EL, considering its consistency and asymptotic normality.

The purpose of this paper is to give a unified treatment of these issues by using the geometry of a Hilbert space composed of score and moment functions, and to shed new light on the large sample properties of GMM and GEL estimators. Some results might be familiar, but our proofs are more lucid and concise. For example, we show that GEL achieves the semiparametric efficiency bound, the proof of which is simple compared with the method of Qin and Lawless (1994). To the best of our knowledge, we completely analyze the optimal choice of moment functions for both estimation and hypothesis testing. Our results include those presented by Carrasco and Florens (2014). We also give a new interpretation of testing overidentifying restrictions. Newey (1985a) pointed out that the test of overidentifying restrictions can easily fail to detect a misspecified model. Such inconsistency of the test is widely known. See standard textbooks such as Hamilton (1994) and Hayashi (2000) for instance. An important finding in this paper is that the test is consistent against an alternative if and only if the corresponding unrestricted model is valid.

The outline of the paper is as follows. In Section 2 we describe the model and compute the semiparametric efficiency bounds. Section 3 proves consistency and asymptotic normality of GMM and GEL estimators. The optimal choice of moment functions is analyzed using the geometry of a Hilbert space. Conditional moment restrictions are treated as a special case. Section 4 investigates hypothesis testing with GMM and GEL estimators, providing a new interpretation of testing overidentifying restrictions.

2 The Model and Efficiency Bounds

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

$$E_{\theta, \eta}[g(z; \theta)] = \int g(z; \theta) f(z; \theta, \eta) d\mu(z) = 0. \quad (2.1)$$

We calculate the semiparametric efficiency bounds using the results of Bickel, Klaassen, Ritov, and Wellner (1993) and van der Vaart (1998). See also Newey (1990b) and Severini and Tripathi (2001) for related studies in econometrics. Suppose that we observe i.i.d. data z_i ($i = 1, \dots, n$) from $F_0 = F(z; \theta_0, \eta_0)$, the true distribution. Let P_0 be the probability measure corresponding to F_0 . Let $L_2^0(P_0)$ denote the set of measurable functions $h : R^p \mapsto R$ with $E[h^2] < \infty$ and $E[h] = 0$, where $E[\cdot]$ indicates the expectation at the truth. It is known that $L_2^0(P_0)$ is a Hilbert space. Let $\Pi(h|B)$ be the orthogonal projection of h on a closed linear space B in $L_2^0(P_0)$. If h is a k -vector $h = (h_1, h_2, \dots, h_k)'$ with $h_i \in L_2^0(P_0)$ ($i = 1, 2, \dots, k$), $\Pi(h|B)$ means $(\Pi(h_1|B), \Pi(h_2|B), \dots, \Pi(h_k|B))'$, and $\{h\} \equiv \{c'h : c \in R^k\} \subset L_2^0(P_0)$. Let ℓ_θ and a denote the score functions for θ and η at (θ_0, η_0) , respectively. Typically, $\ell_\theta = \partial \ln f(z; \theta_0, \eta_0)/\partial \theta$ and $a = \partial \ln f(z; \theta_0, \eta_t)/\partial t|_{t=0}$, where $f(z; \theta_0, \eta_t)$ is a one-dimensional submodel with respect to η at (θ_0, η_0) , although the scores are defined precisely in terms of Fréchet derivatives. Proposition 2.1.1 of Bickel, Klaassen, Ritov, and Wellner (1993) gives sufficient conditions for the validity of the definition in terms of pointwise derivatives. Throughout this paper we assume that $\ell_\theta = \partial \ln f(z; \theta_0, \eta_0)/\partial \theta$ and $a = \partial \ln f(z; \theta_0, \eta_t)/\partial t|_{t=0}$ hold as well as $E[\ell_\theta'] < \infty$ and $E[a^2] < \infty$.

The first assumption imposes dominance conditions which guarantee the interchange of differentiation and integration. Let $\|\cdot\|$ denote Euclidean norm.

Assumption 1 *There exist measurable functions $b_1(z)$ and $b_2(z)$ such that for almost all z , $\|g(z; \theta)\|$ and $\|g(z; \theta)/\partial \theta\|$ are each less than $b_1(z)$ for all θ and $f(z; \theta, \eta_t)$, $\|\partial f(z; \theta, \eta_t)/\partial \theta\|$, and $|\partial f(z; \theta, \eta_t)/\partial t|$ are each less than $b_2(z)$ for all θ and t . Further, it is the case that $\int b_2(z) d\mu(z) < \infty$ and $\int b_1(z) b_2(z) d\mu(z) < \infty$.*

Under Assumption 1, differentiation with respect to θ in equation (2.1) yields

$$E[g\ell_\theta'] = -E\left[\frac{\partial g(z; \theta_0)}{\partial \theta}\right], \quad (2.2)$$

where $g = g(z; \theta_0)$. Since any perturbation η_t of η_0 within the model must satisfy equation (2.1), differentiation with respect to t gives

$$E[ga] = 0. \quad (2.3)$$

Assumption 1 also ensures that $E[\ell_\theta] = 0$ and $E[a] = 0$, which leads to $\{\ell_\theta\} \subset L_2^0(P_0)$ and $a \in L_2^0(P_0)$. The function a runs through a finite or infinite-dimensional closed linear space A referred to as the tangent space for η . The full tangent space $T(P_0)$ for the model is the set $\{\delta'\ell_\theta + a : \delta \in R^q, a \in A\} = \{\ell_\theta\} + A$ in $L_2^0(P_0)$. Let $\phi(z_i; z) = 1(z_i < z) - F(z; \theta_0, \eta_0)$, $\Omega = E[gg']$, and $G = E[\partial g(z; \theta_0)/\partial \theta]$.

Assumption 2 (a) Ω is nonsingular; (b) $\text{rank}(G) = q$.

The following theorem and proposition give the semiparametric efficiency bounds for the model $f(z; \theta, \eta)$ with the moment condition (2.1).

Theorem 1 Suppose that the space $\{g\}$ is the orthogonal complement $A^\perp$ of the space A within $L_2^0(P_0)$. Then under Assumptions 1 and 2, the semiparametric variance bound Σ for θ is given by

$$\Sigma = (G'\Omega^{-1}G)^{-1},$$

and the semiparametric variance bound $W(z)$ for $F(z; \theta, \eta)$ is given by

$$\begin{aligned} W &= E[\phi^2] - E[\phi g']\Omega^{-1}E[g\phi] \\ &\quad + E[\phi g']\Omega^{-1}G\Sigma G'\Omega^{-1}E[g\phi]. \end{aligned}$$

Proof: The efficient score function $\tilde{\ell}_\theta$ is obtained by

$$\tilde{\ell}_\theta = \ell_\theta - \Pi(\ell_\theta|A) = \Pi(\ell_\theta|A^\perp).$$

Because $\{g\} = A^\perp$, we have

$$\tilde{\ell}_\theta = \Pi(\ell_\theta|\{g\}) = E[\ell_\theta g']\{E[gg']\}^{-1}g = -G'\Omega^{-1}g,$$

where the third equality follows from equation (2.2). The efficiency bound Σ is given by $\{E[\tilde{\ell}_\theta \tilde{\ell}_\theta']\}^{-1}$. The efficient influence function $\tilde{\phi}$ is calculated as

$$\begin{aligned} \tilde{\phi} &= \Pi(\phi|T(P_0)) \\ &= \Pi(\phi|A) + \Pi(\phi|\{\tilde{\ell}_\theta\}) \\ &= \phi - E[\phi g']\Omega^{-1}g + E[\phi g']\Omega^{-1}G\Sigma G'\Omega^{-1}g, \end{aligned}$$

where the second equality follows from Theorem A.4.5 of Bickel, Klaassen, Ritov, and Wellner (1993), and the efficiency bound W is given by $E[\tilde{\phi}^2]$. ■

Proposition 1 *If Assumptions 1 and 2 are satisfied, the semiparametric variance bound for θ is less than or equal to Σ , and the variance bound for $F(z; \theta, \eta)$ is less than or equal to $W(z)$.*

Proof: By (2.3), $\{g\} \subset A^\perp$ in $L_2^0(P_0)$. Hence, in view of the proof of Theorem 1 the efficiency bound for θ is less than or equal to Σ . Similarly, since $A \subset \{g\}^\perp$, the full tangent space $T(P_0)$ for the model is a subset of $\{\ell_\theta\} + \{g\}^\perp$ in $L_2^0(P_0)$. Hence, the efficiency bound for $F(z; \theta, \eta)$ is less than or equal to $W(z)$. ■

The above framework considering the pair $g(z; \theta)$ and $f(z; \theta, \eta)$ is essentially the same as in Newey (1985a, b) and Newey and West (1987), playing an important role later in Sections 3 and 4.

Chamberlain (1987) and Qin and Lawless (1994) studied asymptotic properties of GMM and EL estimators, respectively. They showed that each of the estimators achieves the semiparametric efficiency bound. Their framework deals with the pair $g(z; \theta)$ and totally unknown $f(z)$, but implicitly assumes that θ is a function of $f(z)$ through the equation

$$E_f[g(z; \theta)] = \int g(z; \theta(f))f(z)d\mu(z) = 0.$$

It seems that the model $f(z; \theta, \eta)$ we consider is slightly more restricted than the model $f(z)$ they considered. Theorem 1, however, shows that our efficiency bounds are equal to their bounds as long as $\{g\} = A^\perp$. This implies that our model can be a “least favorable” submodel (van der Vaart, 1998, p. 362) of theirs. As we will see in Example 2, in the multinomial distribution case $(\psi(z; \theta), G(z))$, one can construct a least favorable submodel $G(z; \theta, \eta)$ if choosing an appropriate nuisance parameter η . Here $\psi(z; \theta)$ is a moment function and $G(z)$ is a multinomial distribution. Moreover, for any pair $(\psi(z; \theta), F(z))$, where $F(z)$ is a general probability distribution, one can construct a corresponding multinomial model $G(z; \theta, \eta)$, too. Thus, it can be shown that all the results of this paper are valid for the general case by ana-

lyzing the corresponding multinomial case instead. An illustrative example for Theorem 1 is given in Example 2.

Next, let $z = (x', y')'$ and consider the conditional moment restrictions

$$E_{\theta, \eta}[g(z; \theta)|x] = 0,$$

which is equivalent to

$$\int g(z; \theta) f(z; \theta, \eta) d\mu(y) = 0. \quad (2.4)$$

The equations above are assumed to hold for almost all x . Differentiating the both sides of equation (2.4) yields

$$E[g\ell'_\theta|x] = -E\left[\frac{\partial g(z; \theta_0)}{\partial \theta} \mid x\right] \quad (2.5)$$

and

$$E[ga|x] = 0, \quad (2.6)$$

for almost all x . Differentiation under the integral sign in equation (2.4) is possible by the next assumption.

Assumption 3 *There exist measurable functions $b_1(z)$ and $b_2(z)$ such that for almost all z , $\|g(z; \theta)\|$ and $\|\partial g(z; \theta)/\partial \theta\|$ are each less than $b_1(z)$ for all θ and $f(z; \theta, \eta_t)$, $\|\partial f(z; \theta, \eta_t)/\partial \theta\|$, and $|\partial f(z; \theta, \eta_t)/\partial t|$ are each less than $b_2(z)$ for all θ and t . Further, it is the case that $\int b_1(z)b_2(z)d\mu(y) < \infty$ for almost all x .*

Equation (2.6) implies that a is orthogonal to all functions of the form $h(x)'g(z|\theta_0)$ in $L_2^0(P_0)$, where $h(x)$ is an r -vector function of x . The set of the functions is denoted by $\{h'g\}$. Let $\Omega(x) = E[gg'|x]$ and $G(x) = E[\partial g(z; \theta_0)/\partial \theta|x]$.

Assumption 4 (a) $\Omega(x)$ is nonsingular for almost all x ; (b) $\text{rank}[G(x)] = q$ for almost all x .

Theorem 2 and Proposition 2 give the semiparametric efficiency bounds for the model $f(z; \theta, \eta)$ with the conditional moment restrictions (2.4).

Theorem 2 Suppose that the space $\{h'g\}$ is the orthogonal complement $A^\perp$ of the space A within $L_2^0(P_0)$. Then under Assumptions 3 and 4, the semiparametric variance bound Σ_C for θ is given by

$$\Sigma_C = \left\{ E \left[G'(x) \Omega^{-1}(x) G(x) \right] \right\}^{-1},$$

and the semiparametric variance bound $W_C(z)$ for $F(z; \theta, \eta)$ is given by

$$\begin{aligned} W_C &= E[\phi^2] - E \left[E[\phi g' | x] \Omega^{-1}(x) E[g \phi | x] \right] \\ &\quad + E \left[E[\phi g' | x] \Omega^{-1}(x) G(x) \right] \Sigma_C E \left[G'(x) \Omega^{-1}(x) E[g \phi | x] \right]. \end{aligned}$$

Proof: The efficient score function $\tilde{\ell}_\theta$ is obtained by

$$\tilde{\ell}_\theta = \Pi(\ell_\theta | A^\perp).$$

Because $\{h'g\} = A^\perp$, we have

$$\tilde{\ell}_\theta = \Pi(\ell_\theta | \{h'g\}) = E[\ell_\theta g' | x] \{E[g g' | x]\}^{-1} g = -G'(x) \Omega^{-1}(x) g,$$

where the third equality follows from equation (2.5). The efficiency bound Σ_C is given by $\{E[\tilde{\ell}_\theta \tilde{\ell}_\theta']\}^{-1}$. The efficient influence function $\tilde{\phi}$ is calculated as

$$\begin{aligned} \tilde{\phi} &= \Pi(\phi | T(P_0)) \\ &= \Pi(\phi | A) + \Pi(\phi | \{\tilde{\ell}_\theta\}) \\ &= \phi - E[\phi g' | x] \Omega^{-1}(x) g \\ &\quad + E \left[E[\phi g' | x] \Omega^{-1}(x) G(x) \right] \Sigma_C G'(x) \Omega^{-1}(x) g, \end{aligned}$$

and the efficiency bound W_C is given by $E[\tilde{\phi}^2]$. ■

Proposition 2 If Assumptions 3 and 4 are satisfied, the semiparametric variance bound for θ is less than or equal to Σ_C , and the variance bound for $F(z; \theta, \eta)$ is less than or equal to $W_C(z)$.

Proof: Essentially the same as the proof of Proposition 1. ■

We conclude this section with a list of examples, most of which typically appear in the econometric literature.

Example 1. For a simple example, consider the probability density function

$$f(z; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(z - \mu)^2}{2\sigma^2} \right\},$$

where $z, \mu \in R$ and $\sigma^2 \in R^+$. Let μ be a parameter of primary interest, while σ^2 a nuisance parameter. The i.i.d. observations z_i are distributed according to $f(z; \mu_0, \sigma_0^2)$, the true density. The score function a for σ^2 is given by

$$a = -\frac{1}{2\sigma_0^2} + \frac{(z - \mu_0)^2}{2\sigma_0^4}.$$

By choosing the function $g(z; \mu) = z - \mu$ for the moment condition (2.1), we have $\{g\} \subset A^\perp$, the orthogonal complement of the tangent space A for σ^2 . In this case, however, $\{g\} = A^\perp$ does not hold. On the other hand, suppose that $f(z; \mu, \eta) = \eta(z - \mu)$, where $z, \mu \in R$, and η is a general density function with mean 0. The i.i.d. observations z_i are distributed according to $f(z; \mu_0, \eta_0)$, the true density. By choosing the function $g(z; \mu) = z - \mu$ for the moment condition, we have $\{g\}^\perp = A$, the tangent space for η . Hence, the efficiency bound for μ equals $\Sigma = E[(z - \mu_0)^2]$.

Example 2. Chamberlain (1987) considered the case in which an i.i.d. sequence $\{z_i\}$ has a multinomial distribution; i.e., the support of the true distribution G_0 is a finite set $\{\phi_1, \dots, \phi_r\}$. The ϕ_j are known; the unknown parameters are $G_0(\phi_j)$. These parameters are subject to the restriction that $\sum_{j=1}^r \psi(\phi_j, \theta_0) G_0(\phi_j) = 0$, where ψ is a q -vector function and $\theta_0 \in R^p$ ($p \leq q < r - 1$). Then by the implicit function theorem, one can construct a distribution function $G(\phi_j; \theta, \eta)$ which is continuously differentiable with respect to the parameter (θ, η) and satisfies $\sum_{j=1}^r \psi(\phi_j, \theta) G(\phi_j; \theta, \eta) = 0$ in a neighborhood of $(\theta_0, \eta_0) \in R^{p+r-q-1}$ with $G(\phi_j; \theta_0, \eta_0) = G_0(\phi_j)$ for some η_0 . It is shown that $\{\psi\}$ is the orthogonal complement of the tangent space A for η . Hence, the efficiency bound for θ equals Σ . This implies that if a distribution $G(z)$ with a finite support has a moment condition $\int \psi(z; \theta) dG(z) \equiv \sum_{j=1}^r \psi(\phi_j, \theta) G(\phi_j) = 0$, there always exists a representation $G(z) = G(z; \theta, \eta)$, where G is known up to the parameter (θ, η) . Furthermore, in the general case that $\int \psi(z; \theta) dF(z) = 0$ with the true parameter

(θ_0, F_0) , in view of Section 3 of Chamberlain (1987), by introducing an appropriate topology into the set $\{F\}$ of probability measures, one can construct a multinomial model $G(z; \theta, \eta)$ with G_0 arbitrarily close to F_0 and satisfying

$$\begin{aligned} \int \psi(z; \theta_0) \psi'(z; \theta_0) dF_0(z) &= \int \psi(z; \theta_0) \psi'(z; \theta_0) dG_0(z), \\ \int \frac{\partial \psi(z; \theta_0)}{\partial \theta} dF_0(z) &= \int \frac{\partial \psi(z; \theta_0)}{\partial \theta} dG_0(z) \\ &= - \int \psi(z; \theta_0) \frac{\partial \log G(z; \theta_0, \eta_0)}{\partial \theta} dG_0(z), \end{aligned}$$

which implies that the first assertion of Theorem 1 is applicable to the general pair $(\psi(z; \theta), F(z))$. Moreover, given z_0 , there exists some G_0 such that

$$F_0(z_0) = \int 1(z < z_0) dF_0(z) = \int 1(z < z_0) dG_0(z) = G_0(z_0),$$

$$\int \phi^2(z) dF_0(z) = \int \phi^2(z) dG_0(z),$$

$$\int \phi(z) \psi'(z; \theta_0) dF_0(z) = \int \phi(z) \psi'(z; \theta_0) dG_0(z),$$

where $\phi(z) = 1(z < z_0) - F_0(z_0)$. Thus the second assertion of Theorem 1 is also applicable to $(\psi(z; \theta), F(z))$.

Example 3. Let $(x_1, y_1, w_1), (x_2, y_2, w_2), \dots, (x_n, y_n, w_n)$ be i.i.d. and consider the instrumental variable model

$$y = x'\theta + \varepsilon \quad E[w\varepsilon] = 0,$$

where $x, \theta \in R^q$ and $w \in R^r$ with $q \leq r$. Then $z = (x', y, w')'$ has a density $f(z; \theta, \eta) = \eta(x, y - x'\theta, w)$ where η is only restricted by the relations $\int w\varepsilon\eta(x, \varepsilon, w) dx d\varepsilon dw = 0$. The observations $z_i = (x'_i, y_i, w'_i)'$ follow the true density $f(z; \theta_0, \eta_0)$. Choosing $g(z; \theta) = w(y - x'\theta)$ for the moment condition, we have $\{g\}^\perp = A$, the tangent space for η , and the efficiency bound for θ is equal to Σ which is given by

$$\Sigma = \left\{ E[xw'](E[ww'\varepsilon^2])^{-1} E[wx'] \right\}^{-1}.$$

Example 4. Consider the i.i.d. observations $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ and the regression model

$$y = x'\theta + \varepsilon \quad E[\varepsilon|x] = 0,$$

where $x, \theta \in R^q$. Then $z = (x', y)'$ has a density $f(z; \theta, \eta) = \eta(x, y - x'\theta)$ and the observations z_i follow the true density $f(z; \theta_0, \eta_0)$. The nuisance parameter η is only restricted by the relations $\int \varepsilon \eta(x, \varepsilon) d\varepsilon = 0$, which implies that the score function for η is orthogonal to any function of the form $\varepsilon h(x)$. Hence, by choosing $g(z; \theta) = x(y - x'\theta)$ for the moment condition, we have $\{g\} \subset A^\perp$ (but $\{g\} \neq A^\perp$) and the efficiency bound for θ is not greater than Σ , which is given by

$$\Sigma = \left\{ E[xx'] (E[xx'\varepsilon^2])^{-1} E[xx'] \right\}^{-1}.$$

In fact, the efficiency bound for θ is obtained by $\{E[xx'/E[\varepsilon^2|x]]\}^{-1}$ as we shall see later.

Example 5. Let $\{(y_i, x_i)\}$ be i.i.d. observations. Amemiya (1977) considered the nonlinear simultaneous equations model defined by

$$\psi(y, x; \theta) = \varepsilon \quad E[\varepsilon|x] = 0,$$

where ψ is an m -vector function, y is an m -vector of endogenous variables, x is an l -vector of exogenous variables, and θ is a q -vector of unknown parameters. We assume that (x, ε) possesses a density η which satisfies the relations $\int \varepsilon \eta(x, \varepsilon) d\varepsilon = 0$ and otherwise unrestricted. Under some regularity conditions, the observation (y, x) has a density $f(y, x; \theta, \eta) = k(y, x, \theta) \eta(x, \psi(y, x; \theta))$ via the change of variables formula. Choose the function $g(y, x; \theta) = \psi(y, x; \theta) \otimes x$ for the moment condition and set the optimal weighting matrix $\{E[\varepsilon \varepsilon' \otimes xx']\}^{-1}$, then the GMM estimator coincides with the three-stage least squares estimator using x for the instrumental variable under conditional homoscedasticity $E[\varepsilon \varepsilon'|x] = \Sigma$. See equation (5.1) of Amemiya (1977). However, the above three-stage least squares estimator is not efficient. By choosing

$$\begin{aligned} g(y, x; \theta) &= E[\partial \psi(y, x; \theta_0) / \partial \theta' | x] \\ &\quad \times \{E[\psi(y, x; \theta_0) \psi(y, x; \theta_0)' | x]\}^{-1} \psi(y, x; \theta) \end{aligned}$$

for the moment condition, the GMM estimator achieves the efficiency bound, which will be shown in Section 3.

3 GMM and GEL Estimators: Basic Results

Let $\theta_0 \in \Theta$ be the true parameter of the model satisfying the moment condition

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

where the expectation is taken with respect to the distribution of z_i , i.e., $F(z; \theta_0, \eta_0)$. Let $\hat{g}(\theta) = n^{-1} \sum_{i=1}^n g(z_i; \theta)$. The GMM estimator we consider is

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

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

Assumption 5 (a) $\theta_0 \in \Theta$ is the unique solution to $E[g(z; \theta)] = 0$; (b) Θ is compact; (c) $g(z; \theta)$ is continuous at each $\theta \in \Theta$ with probability one; (d) $E[\sup_{\theta \in \Theta} \|g(z; \theta)\|] < \infty$; (e) Ω is nonsingular.

Assumption 6 (a) $\theta_0 \in \text{int}(\Theta)$; (b) $g(z; \theta)$ is continuously differentiable in a neighborhood $\mathcal{N}$ of θ_0 with probability one and $E[\sup_{\theta \in \mathcal{N}} \|\partial g(z; \theta)/\partial \theta\|] < \infty$; (c) $\text{rank}(G) = q$.

The next lemma shows consistency and asymptotic normality of GMM estimators.

Lemma 1 *If Assumption 5 is satisfied, then $\hat{\theta}_{\text{GMM}} \xrightarrow{P} \theta_0$. If Assumptions 5 and 6 are satisfied, then*

$$\sqrt{n}(\hat{\theta}_{\text{GMM}} - \theta_0) \xrightarrow{d} N(0, \Sigma).$$

Proof: See Theorems 2.6 and 3.4 of Newey and McFadden (1994). ■

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 $\rho_1(0) = \rho_2(0) = -1$. Associated with each GEL estimator empirical probabilities are given by

$$\hat{\pi}_i = \frac{\rho_1(\hat{\lambda}' g(z_i; \hat{\theta}_{\text{GEL}}))}{\sum_{j=1}^n \rho_1(\hat{\lambda}' g(z_j; \hat{\theta}_{\text{GEL}}))} \quad (i = 1, 2, \dots, n), \quad (3.1)$$

which yield an estimate of the distribution function F as

$$\hat{F}_n(z) = \sum_{i=1}^n \hat{\pi}_i 1(z_i < z). \quad (3.2)$$

Let $P = \Omega^{-1} - \Omega^{-1} G \Sigma G' \Omega^{-1}$ and $W(z) = F(z)(1 - F(z)) - B'(z) P B(z)$, where $F(z) = F(z; \theta_0, \eta_0)$ and $B(z) = E[g(z_i; \theta_0) \phi(z_i; z)]$.

Assumption 7 (a) $\theta_0 \in \Theta$ is the unique solution to $E[g(z; \theta)] = 0$; (b) Θ is compact; (c) $g(z; \theta)$ is continuous at each $\theta \in \Theta$ with probability one; (d) $E[\sup_{\theta \in \Theta} \|g(z; \theta)\|^\alpha] < \infty$ for some $\alpha > 2$; (e) Ω is nonsingular.

The following lemma gives consistency and asymptotic normality of GEL estimators.

Lemma 2 *If Assumption 7 is satisfied, then $\hat{\theta}_{\text{GEL}} \xrightarrow{P} \theta_0$. If Assumptions 6 and 7 are satisfied, then*

$$\sqrt{n} \begin{pmatrix} \hat{\theta}_{\text{GEL}} - \theta_0 \\ \hat{\lambda} \end{pmatrix} \xrightarrow{d} N(0, \text{diag}(\Sigma, P)),$$

$$\sqrt{n}(\hat{F}_n(z) - F(z)) \xrightarrow{d} N(0, W(z)).$$

Proof: The first and second relations follow from Theorems 3.1 and 3.2 of Newey and Smith (2004), respectively. For the third relation, by equations

(3.1) and (3.2) we have

$$\begin{aligned}
 \sqrt{n}(\hat{F}_n(z) - F(z)) &= \frac{\frac{\sqrt{n}}{n} \sum_{i=1}^n \{\rho_1(\hat{\lambda}' g(z_i; \hat{\theta}_{\text{GEL}})) - \rho_1(0)\} 1(z_i < z)}{\frac{1}{n} \sum_{i=1}^n \rho_1(\hat{\lambda}' g(z_i; \hat{\theta}_{\text{GEL}}))} \\
 &\quad - \frac{\frac{\sqrt{n}}{n} \sum_{i=1}^n 1(z_i < z)}{\frac{1}{n} \sum_{i=1}^n \rho_1(\hat{\lambda}' g(z_i; \hat{\theta}_{\text{GEL}}))} - \sqrt{n}F(z) \\
 &= -\sqrt{n}\hat{\lambda}' \frac{1}{n} \sum_{i=1}^n \rho_2(\bar{\lambda}' g(z_i; \hat{\theta}_{\text{GEL}})) g(z_i; \hat{\theta}_{\text{GEL}}) 1(z_i < z) \\
 &\quad + \frac{\sqrt{n}}{n} \sum_{i=1}^n (1(z_i < z) - F(z)) + o_p(1) \\
 &= \sqrt{n}\hat{\lambda}' E[g(z_i; \theta_0) 1(z_i < z)] \\
 &\quad + \frac{\sqrt{n}}{n} \sum_{i=1}^n (1(z_i < z) - F(z)) + o_p(1),
 \end{aligned}$$

where $\bar{\lambda}$ lies on the line joining $\hat{\lambda}$ and 0. The second equality follows from the fact $\max_{i \leq n} |\rho_1(\hat{\lambda}' g(z_i; \hat{\theta}_{\text{GEL}})) + 1| \xrightarrow{p} 0$ and the mean value theorem. The third equality follows from $\max_{i \leq n} |\rho_2(\bar{\lambda}' g(z_i; \hat{\theta}_{\text{GEL}})) + 1| \xrightarrow{p} 0$ and the uniform weak law of large numbers. The conclusion follows from the above equation with the fact $\hat{\lambda} = -P \sum_{i=1}^n g(z_i; \theta_0)/n + o_p(n^{-1/2})$. ■

Note that Lemmas 1 and 2 also hold for a general distribution $F(z)$ only satisfying $\int g(z; \theta) dF(z) = 0$. In other words no specification $F(z; \theta, \eta)$ needed for the proofs.

Now we consider the basic properties of the estimators $\hat{\theta}_{\text{GMM}}$ and $\hat{\theta}_{\text{GEL}}$. All the calculations are done with respect to the true parameter value (θ_0, η_0) . We suppress the subscript “0” for notational simplicity.

Proposition 3 *Under Assumptions 1-2 and 5-7, the estimators $\hat{\theta}_{\text{GMM}}$, $\hat{\theta}_{\text{GEL}}$, and $\hat{F}_n(z)$ achieve the semiparametric efficiency bounds if the tangent space A for the nuisance parameter η is equal to $\{g\}^\perp$ in $L_2^0(P_0)$.*

Proof: It follows from Lemmas 1 and 2 with Theorem 1. ■

To examine the choice of vector functions $g(z; \theta)$ for the moment condition, we see how it influences the efficiency of the estimators. Since the asymptotic

variances of $\sqrt{n}(\hat{\theta}_{\text{GMM}} - \theta_0)$ and $\sqrt{n}(\hat{\theta}_{\text{GEL}} - \theta_0)$ are the same, we omit the GMM cases. Throughout the rest of this section it is assumed that all the moment functions and the underlying density function satisfy Assumptions 1-7. We consider the two moment functions $g(z; \theta)$ and $g^{(1)}(z; \theta)$ with $\{g^{(1)}\} \subset \{g\}$ in $L_2^0(P_0)$. For example, let $g^{(1)}(z; \theta) = C(\theta)g(z; \theta)$, where $C(\theta)$ is an $r_1 \times r$ matrix with $r_1 \leq r$ and $\text{rank}[C(\theta)] = r_1$. Let $\|w\|_P = (E[w^2])^{1/2}$ for $w \in L_2^0(P_0)$.

Theorem 3 *The asymptotic variance Σ of $\sqrt{n}(\hat{\theta}_{\text{GEL}} - \theta_0)$ satisfies $\delta' \Sigma^{-1} \delta = \|\Pi(\delta' \ell_\theta | \{g\})\|_P^2$ for every $\delta \in R^q$. Furthermore, the asymptotic variance $W(z)$ of $\sqrt{n}(\hat{F}_n(z) - F(z))$ is equal to $\|\Pi(\phi | \{\ell_\theta\} + \{g\}^\perp)\|_P^2$.*

Proof: The first assertion follows from the fact that

$$\Pi(\delta' \ell_\theta | \{g\}) = \delta' E[\ell_\theta g'] \{E[g g']\}^{-1} g$$

and that $E[\partial g(z; \theta_0) / \partial \theta] = -E[g \ell_\theta']$. For the second assertion, straightforward calculations yield the desired results as in the proof of Theorem 1. ■

The above theorem associates the asymptotic variance of GEL estimators with the geometric structure of the space $L_2^0(P_0)$. The following lemma gives some basic projection formulas.

Lemma 3 *Suppose that the two moment functions $g(z; \theta)$ and $g^{(1)}(z; \theta)$ satisfy $\{g^{(1)}\} \subset \{g\}$ in $L_2^0(P_0)$. Then*

- (a) $\|\Pi(w | \{g^{(1)}\})\|_P \leq \|\Pi(w | \{g\})\|_P \leq \|w\|_P$ for every $w \in L_2^0(P_0)$;
- (b) $\|\Pi(w | \{\ell_\theta\} + \{g\}^\perp)\|_P \leq \|\Pi(w | \{\ell_\theta\} + \{g^{(1)}\}^\perp)\|_P \leq \|w\|_P$ for every $w \in L_2^0(P_0)$;
- (c) $\{\ell_\theta\} + \{g\}^\perp = L_2^0(P_0)$ if $\{\Pi(\ell_\theta | \{g\})\} = \{g\}$.

Proof: Generally, if H_1 and H_2 are subspaces of a Hilbert space H with $H_1 \subset H_2$, then $\Pi(h | H_1) = \Pi(\Pi(h | H_2) | H_1)$ for $h \in H$. Combined with the fact that $\|\Pi(h | H_3)\|_H \leq \|h\|_H$ for $h \in H$ and $H_3 \subset H$, we have $\|\Pi(h | H_1)\|_H \leq \|\Pi(h | H_2)\|_H \leq \|h\|_H$, where $\|\cdot\|_H$ denotes a norm of H . Then claims (a) and (b) immediately follow from this. For (c),

$$\{\ell_\theta\} + \{g\}^\perp = \{\Pi(\ell_\theta | \{g\})\} + \{g\}^\perp = \{g\} + \{g\}^\perp = L_2^0(P_0).$$

This completes the proof. ■

Let Σ and Σ_1 denote the asymptotic variances of $\sqrt{n}(\hat{\theta}_{\text{GEL}} - \theta_0)$ with $g(z; \theta)$ and $g^{(1)}(z; \theta)$, respectively. Also let $W(z)$ and $W_1(z)$ denote those of $\sqrt{n}(\hat{F}_n(z) - F(z))$ with $g(z; \theta)$ and $g^{(1)}(z; \theta)$. Theorem 3 combined with Lemma 3 yields the next proposition, which shows the efficiency comparison between the two sets of moment conditions.

Proposition 4 *Suppose that the two moment functions $g(z; \theta)$ and $g^{(1)}(z; \theta)$ satisfy $\{g^{(1)}\} \subset \{g\}$ in $L_2^0(P_0)$. Then*

- (a) $\Sigma \leq \Sigma_1$;
- (b) $W(z) \leq W_1(z)$;
- (c) $W(z) = F(z)(1 - F(z))$ if $r = q$.

Proof: Claim (a) follows from the first assertion of Theorem 3 and Lemma 3 (a). Claim (b) follows from the second assertion of Theorem 3 and Lemma 3 (b). For claim (c), if $r = q$, then $\{\Pi(\ell_\theta|\{g\})\} = \{g\}$, which leads to $W(z) = \|\phi\|_P^2$ by the second assertion of Theorem 3 and Lemma 3 (c). ■

Now, we choose the q -vector function $\ell_\theta(\theta) = \partial \ln f(z; \theta, \eta_0)/\partial \theta$ for the moment condition. In this case the underlying density function $f(z; \theta, \eta)$ is assumed to be known up the parameter θ . Let Σ_0 denote the asymptotic variance of $\sqrt{n}(\hat{\theta}_{\text{GEL}} - \theta_0)$. Note that Σ_0 is the inverse of the Fisher information matrix with η known. Let $W_0(z)$ denote that of $\sqrt{n}(\hat{F}_n(z) - F(z))$.

Proposition 5 *For any Σ and $W(z)$,*

- (a) $\Sigma_0 \leq \Sigma$;
- (a') $\Sigma_0 = \Sigma$ if and only if $\{\ell_\theta\} \subset \{g\}$;
- (b) $W(z) \leq W_0(z) = F(z)(1 - F(z))$.

Proof: Claim (a) follows from the first assertion of Theorem 3 and Lemma 3 (a). For claim (a'), we have

$$\|\delta' \ell_\theta\|_P^2 = \|\Pi(\delta' \ell_\theta|\{g\})\|_P^2 + \|\Pi(\delta' \ell_\theta|\{g\}^\perp)\|_P^2,$$

and for any $\delta \neq 0$ $\|\Pi(\delta' \ell_\theta|\{g\}^\perp)\|_P^2 = 0$ holds if and only if $\{\ell_\theta\} \subset \{g\}$. Claim (b) follows from the second assertion of Theorem 3 and Lemma 3 (b) and (c), which completes the proof. ■

Proposition 5 implies that $\ell_\theta(\theta)$ is the optimal moment function for estimating θ . At the same time, a moment function g is optimal if and only if $\{\ell_\theta\} \subset \{g\}$. This result corresponds to Proposition 4.1 of Carrasco and Florens (2014). For estimating the distribution $F(z)$, however, the GEL estimator with $\ell_\theta(\theta)$ is as inefficient as the empirical distribution function.

Next, consider the efficient score function $\tilde{\ell}_\theta = \Pi(\ell_\theta|A^\perp)$. Let $\tilde{\ell}_\theta(\theta)$ be the efficient score function with the parameter θ_0 replaced by θ . Let Σ_A denote the asymptotic variance of $\sqrt{n}(\hat{\theta}_{\text{GEL}} - \theta_0)$ using $\tilde{\ell}_\theta(\theta)$ for the moment condition. Also let $W_A(z)$ denote that of $\sqrt{n}(\hat{F}_n(z) - F(z))$. Let $G_A = E[\partial \tilde{\ell}_\theta(\theta_0)/\partial \theta]$ and $\Omega_A = E[\tilde{\ell}_\theta \tilde{\ell}_\theta']$. Since $G_A = E[\partial \tilde{\ell}_\theta(\theta_0)/\partial \theta] = E[\tilde{\ell}_\theta \ell'_\theta] = E[\tilde{\ell}_\theta \tilde{\ell}'_\theta] = \Omega_A$, we have $\Sigma_A = (G'_A \Omega_A^{-1} G_A)^{-1} = G_A^{-1} = \Omega_A^{-1}$.

Proposition 6 For any Σ and $W(z)$,

- (a) $\Sigma_0 \leq \Sigma_A \leq \Sigma$;
- (b) $W(z) \leq W_0(z) = W_A(z) = F(z)(1 - F(z))$.

Proof: Note that

$$\delta' \Sigma_A^{-1} \delta = \|\Pi(\delta' \ell_\theta | \{\tilde{\ell}_\theta\})\|_P^2 = \|\Pi(\delta' \ell_\theta | A^\perp)\|_P^2.$$

Since $\{g\} \subset A^\perp \subset L_0^2(P_0)$ we have $\Sigma_0 \leq \Sigma_A \leq \Sigma$. For (b), Lemma 3 (c) yields $W_A(z) = \|\Pi(\phi | \{\ell_\theta\} + \{\tilde{\ell}_\theta\}^\perp)\|_P^2 = \|\phi\|_P^2 = F(z)(1 - F(z))$ since $\{\Pi(\ell_\theta | \{\tilde{\ell}_\theta\})\} = \{\tilde{\ell}_\theta\}$. ■

The function $\tilde{\ell}_\theta(\theta)$ satisfies the moment condition

$$E[\tilde{\ell}_\theta(\theta_0)] = 0. \quad (3.3)$$

An important application of equation (3.3) is the case of conditional moment restrictions. See Section 2 for the detailed settings. Let $\theta_0 \in \Theta$ be the true parameter of the model satisfying the conditional moment restrictions

$$E[g(z; \theta_0) | x] = 0 \quad \text{for almost all } x, \quad (3.4)$$

where $z = (x', y')'$. Then (3.4) is equivalent to

$$E[h(x)' g(z; \theta_0)] = 0,$$

where $h(x)$ is any r -vector function of x , and it is the case that $\{h'g\} \subset A^\perp$. If the nuisance parameter η is only restricted by equation (2.4), then $\{h'g\} = A^\perp$

holds. See Examples 4 and 5. Assuming that $\{h'g\} = A^\perp$, the efficient score function $\tilde{\ell}_\theta = \Pi(\ell_\theta|A^\perp)$ is given by

$$\tilde{\ell}_\theta = -G'(x)\Omega^{-1}(x)g$$

and $\tilde{\ell}_\theta(\theta)$ is obtained by

$$\tilde{\ell}_\theta(\theta) = -E_{\theta, \eta_0} \left[\frac{\partial g(z; \theta)}{\partial \theta'} \mid x \right] \left\{ E_{\theta, \eta_0} [g(z; \theta)g'(z; \theta)|x] \right\}^{-1} g(z; \theta). \quad (3.5)$$

The asymptotic variance Σ_A of $\sqrt{n}(\hat{\theta}_{\text{GEL}} - \theta_0)$ then becomes

$$\Omega_A^{-1} = \{E[\tilde{\ell}_\theta \tilde{\ell}_\theta']\}^{-1} = \left\{ E \left[G'(x)\Omega^{-1}(x)G(x) \right] \right\}^{-1},$$

which implies that $\hat{\theta}_{\text{GEL}}$ achieves the semiparametric efficiency bound Σ_C given in Theorem 2. This result leads to the following theorem.

Theorem 4 *Define the instrumental variable matrix $B(x)$ by $-G'(x)\Omega^{-1}(x)$. If we choose $\tilde{\ell}_\theta(\theta) = B(x)g(z; \theta)$ for the moment condition, then $\hat{\theta}_{\text{GEL}}$ achieves the semiparametric efficiency bound under the assumption that $\{h'g\} = A^\perp$.*

Proof: Since $E[g(z; \theta_0)|x] = 0$,

$$E \left[\frac{\partial (B(x)g(z; \theta_0))}{\partial \theta} \right] = E \left[\frac{\partial \tilde{\ell}_\theta(\theta_0)}{\partial \theta} \right].$$

We have also $E[(B(x)g(z; \theta_0))(B(x)g(z; \theta_0))'] = E[\tilde{\ell}_\theta \tilde{\ell}_\theta']$. Hence, the asymptotic variance of $\sqrt{n}(\hat{\theta}_{\text{GEL}} - \theta_0)$ is Ω_A^{-1} . Direct calculations also yield the desired results. ■

Theorem 4 is basically the same as the one proposed by Chamberlain (1987) about the optimal instrumental variables for conditional moment restrictions. See also Newey (1990a). Our proof is a simpler alternative one. The condition $\{h'g\} = A^\perp$ is not always satisfied. Even so, GMM and GEL estimators for θ using $B(x)g(z; \theta)$ are more efficient than those using only $g(z; \theta)$ for the moment condition, due to the fact that $\{g\} \subset \{h'g\}$.

Example 4 (Continued). The score function ℓ_θ for the parameter θ is $-\{\partial \eta(x, \varepsilon)/\partial \varepsilon\}/\eta(x, \varepsilon)x$, and $A^\perp$ is a set of all the functions of the form

$\varepsilon h(x)$. As in Example 25.28 of van der Vaart (1998), the efficient score function is calculated as

$$\tilde{\ell}_\theta = \Pi(\ell_\theta | A^\perp) = \frac{(y - x'\theta_0)}{E[\varepsilon^2|x]} E[\varepsilon \ell_\theta | x] = \frac{(y - x'\theta_0)}{E[\varepsilon^2|x]} x,$$

which gives $\tilde{\ell}_\theta(\theta) = B(x; \theta)(y - x'\theta)$, where $B(x; \theta) = x/E_{\theta, \eta_0}[(y - x'\theta)^2|x]$. By choosing $\tilde{\ell}_\theta(\theta)$ for the moment condition, $\hat{\theta}_{\text{GMM}}$ achieves the efficient variance bound $\Sigma_A = \{E[\tilde{\ell}_\theta \tilde{\ell}_\theta']\}^{-1} = \{E[xx'/E[\varepsilon^2|x]]\}^{-1}$. In view of the fact that $E[\partial B(x; \theta_0)(y - x'\theta_0)/\partial \theta] = E[B(x; \theta_0)\partial(y - x'\theta_0)/\partial \theta]$, choosing $\bar{\ell}_\theta(\theta) = B(x; \theta_0)(y - x'\theta)$ for the moment condition also achieves the efficiency bound, which is no other than the generalized least squares method.

Example 5 (Continued). When the model restriction is $E[\varepsilon|x] = 0$, $A^\perp$ is a set of all the functions of the form $\varepsilon'h(x)$, where $h(x)$ is an m -vector function of x . The efficient score function $\tilde{\ell}_\theta$ is given by

$$\begin{aligned} \Pi(\ell_\theta | A^\perp) &= E[\ell_\theta \varepsilon' | x] \{E[\varepsilon \varepsilon' | x]\}^{-1} \varepsilon \\ &= -E[\partial \psi(y, x; \theta_0)/\partial \theta' | x] \\ &\quad \times \{E[\psi(y, x; \theta_0)\psi(y, x; \theta_0)' | x]\}^{-1} \psi(y, x; \theta_0), \end{aligned}$$

where we use the relation $E[\psi(y, x; \theta_0)\ell'_\theta | x] = -E[\partial \psi(y, x; \theta_0)/\partial \theta | x]$. Choose $\tilde{\ell}_\theta(\theta) = B(x; \theta)\psi(y, x; \theta)$ for the moment condition, where

$$B(x; \theta) = -E_{\theta, \eta_0}[\partial \psi(y, x; \theta)/\partial \theta' | x] \{E_{\theta, \eta_0}[\psi(y, x; \theta)\psi(y, x; \theta)' | x]\}^{-1}.$$

Then the GEL asymptotic variance for θ coincides with

$$\begin{aligned} \Sigma_A &= \left\{ E \left[E \left[\partial \psi(y, x; \theta_0)/\partial \theta' | x \right] \{ E[\psi(y, x; \theta_0)\psi(y, x; \theta_0)' | x] \}^{-1} \right. \right. \\ &\quad \left. \left. \times E[\partial \psi(y, x; \theta_0)/\partial \theta | x] \right] \right\}^{-1}, \end{aligned}$$

which is equal to the efficiency bound. For the same reason as in Example 4, choosing $B(x; \theta_0)\psi(y, x; \theta)$ for the moment condition also achieves the efficiency bound Σ_A .

4 Hypothesis Testing

Suppose that we have an r -vector of functions $g(z; \theta)$ for the moment condition, where θ is a q -vector of parameters and is divided into a q_1 -vector β

and a q_2 -vector γ , i.e., $\theta = (\beta', \gamma')'$. Consider the pair $(g(z; \theta), f(z; \theta, \eta))$ that satisfies equation (2.1). 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

$$H_0 : \gamma_n = \gamma_0. \quad (4.1)$$

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

$$H_1 : \gamma_n = \gamma_0 + \delta/\sqrt{n}, \quad (4.2)$$

where $\delta \in R^{q_2}$. The constrained GMM estimator is defined by

$$\bar{\theta}_{\text{GMM}} = \arg \min_{\theta \in \Theta} \hat{g}(\theta)' \hat{\Omega}^{-1} \hat{g}(\theta) \quad \text{subject to } \gamma = \gamma_0, \quad (4.3)$$

where $\hat{\Omega}$ is some preliminary estimator satisfying $\hat{\Omega} \xrightarrow{P} \Omega$. On the other hand, the unconstrained GMM estimator is given by

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

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 $G = (H, K)$. Let $\ell_\beta = \partial \ln f(z; \beta_0, \gamma_0, \eta_0)/\partial \beta$ and $\ell_\gamma = \partial \ln f(z; \beta_0, \gamma_0, \eta_0)/\partial \gamma$. Then $\ell_\theta = (\ell'_\beta, \ell'_\gamma)'$. We have $E[g \ell'_\beta] = -H$ and $E[g \ell'_\gamma] = -K$ under Assumption 1.

Assumption 8 (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.

Assumption 9 (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 , $\|g(z; \theta)\|^2 < b_1(z)$, $\|\partial g(z; \theta)/\partial \theta\| < b_1(z)$, 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) $\text{rank}(G) = q$.

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

Lemma 4 *If Assumption 8 is satisfied, then $\bar{\theta}_{\text{GMM}} \xrightarrow{p} \theta_0$ under the local alternatives. If Assumptions 8 and 9 are satisfied, then under the local alternatives*

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

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

Proof: [Sketch]¹⁾ The proof of the first assertion is essentially the same as in Newey and McFadden (1994, Theorem 2.6) with slight modification for non-i.i.d. data using an appropriate uniform weak law of large numbers given in Newey (1985b, Lemma A1), for instance. Let $\bar{\theta}_{\text{GMM}} = (\bar{\beta}', \gamma_0')'$ and $\hat{g}(\beta, \gamma) = n^{-1} \sum_{i=1}^n g(z_i; \beta, \gamma)$. For the second assertion, we have

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

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

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

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

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

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

Above relations are derived as in the proof of Lemma 2.1 in Newey (1985b). The second assertion of the lemma follows from these results. ■

1) A complete proof will be given in Goto (2021).

Next, we consider the hypothesis test using the GEL estimators. The constrained GMM estimator is defined by

$$\bar{\theta}_{\text{GEL}} = \arg \min_{\theta \in \Theta} \sup_{\lambda \in \hat{\Lambda}_n(\theta)} \sum_{i=1}^n \rho(\lambda' g(z_i; \theta)) \quad \text{subject to } \gamma = \gamma_0. \quad (4.4)$$

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.

Assumption 10 (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 $\bar{\lambda} = \arg \max_{\lambda \in \hat{\Lambda}_n(\bar{\theta}_{\text{GEL}})} \sum_{i=1}^n \rho(\lambda' g(z_i; \bar{\theta}_{\text{GEL}}))$. The next lemma gives the power property of GEL hypothesis testing.

Lemma 5 If Assumption 10 is satisfied, then $\bar{\theta}_{\text{GEL}} \xrightarrow{P} \theta_0$ under the local alternatives. If Assumptions 9 and 10 are satisfied, then under the local alternatives

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

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

Proof: [Sketch]²⁾ The proof of the first assertion is essentially the same as in Newey and Smith (2004, Theorem 3.1) with slight modification for non-i.i.d. data. Let $\bar{\theta}_{\text{GEL}} = (\bar{\beta}', \gamma_0')'$. For the second assertion, as in the proof of Theorem 3.2 of Newey and Smith (2004), we have

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

2) A complete proof will be given in Goto (2021).

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

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

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

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

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

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

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

The second assertion follows from these results. ■

Lemmas 4 and 5 imply that the test statistics with GMM and GEL estimators are asymptotically equivalent to each other. Note that these lemmas also hold for a general distribution $F(z)$ only satisfying $\int g(z; \theta) dF(z) = 0$ as in Section 3. Throughout the rest of this section we assume that all the moment functions and the underlying density functions satisfy Assumptions 1-4 and 8-10. Let $\Pi(\{\ell_\beta\}|\{g\}) = \{c'\Pi(\ell_\beta|\{g\}) : c \in R^{q_1}\}$. The following theorem associates the noncentrality parameter κ with the geometric structure of $L_2^0(P_0)$.

Theorem 5 *The noncentrality parameter κ is expressed as*

$$\kappa = \|\Pi(\delta'\ell_\gamma|\{g\}) - \Pi(\delta'\ell_\gamma|\Pi(\{\ell_\beta\}|\{g\}))\|_P^2 = \|\Pi(\delta'\ell_\gamma|\{g\} \cap \{\ell_\beta\}^\perp)\|_P^2.$$

Proof: The first equality follows from straightforward computation. The second equality follows from Proposition A.2.4 of Bickel, Klaassen, Ritov, and Wellner (1993). ■

We examine the optimal choice of moment functions as in Section 3. The asymptotic power of hypothesis tests is determined by the non-centrality parameter and the degrees of freedom. If the latter is fixed, the larger the non-centrality parameter, the more efficient the test. Let κ and κ_1 denote the non-centrality parameters of the tests using $g(z; \theta)$ and $g^{(1)}(z; \theta)$ for the moment condition, respectively. The following proposition implies that adding elements to a moment function increases the power of the hypothesis test.

Proposition 7 *Suppose that the two moment functions $g(z; \theta)$ and $g^{(1)}(z; \theta)$ satisfy $\{g^{(1)}\} \subset \{g\}$ in $L_2^0(P_0)$. Then $\kappa_1 \leq \kappa$.*

Proof: By Theorem 5

$$\kappa_1 = \|\Pi(\delta' \ell_\gamma | \{g^{(1)}\} \cap \{\ell_\beta\}^\perp)\|_P^2 \leq \|\Pi(\delta' \ell_\gamma | \{g\} \cap \{\ell_\beta\}^\perp)\|_P^2 = \kappa,$$

which is the desired result. ■

Now, we choose the function $\ell_\theta(\theta) = \partial \ln f(z; \theta, \eta_0) / \partial \theta$ for the moment condition. Let κ_0 denote the noncentrality parameter of the test.

Proposition 8 *For any κ , $\kappa \leq \kappa_0$ holds with $\kappa = \kappa_0$ if $\{\ell_\theta\} \subset \{g\}$.*

Proof: For the first assertion, by Theorem 5

$$\begin{aligned} \kappa_0 &= \|\Pi(\delta' \ell_\gamma | \{\ell_\theta\}) - \Pi(\delta' \ell_\gamma | \Pi(\{\ell_\beta\} | \{\ell_\theta\}))\|_P^2 \\ &= \|\delta' \ell_\gamma - \Pi(\delta' \ell_\gamma | \{\ell_\beta\})\|_P^2 \\ &= \|\Pi(\delta' \ell_\gamma | \{\ell_\beta\}^\perp)\|_P^2 \\ &\geq \|\Pi(\delta' \ell_\gamma | \{g\} \cap \{\ell_\beta\}^\perp)\|_P^2 = \kappa. \end{aligned}$$

For the second assertion, by Theorem 5

$$\begin{aligned} \kappa &= \|\Pi(\delta' \ell_\gamma | \{g\}) - \Pi(\delta' \ell_\gamma | \Pi(\{\ell_\beta\} | \{g\}))\|_P^2 \\ &= \|\delta' \ell_\gamma - \Pi(\delta' \ell_\gamma | \{\ell_\beta\})\|_P^2 = \kappa_0, \end{aligned}$$

where the second equality follows from $\{\ell_\theta\} \subset \{g\}$. ■

Note that the only-if part of the second assertion in Proposition 8 does not hold as in Proposition 5. For example, if $\{\ell_\gamma\} \subset \{g\}$ and $\{\ell_\gamma\} \perp \{\ell_\beta\}$, then $\kappa = \kappa_0$. This is because

$$\kappa = \|\Pi(\delta' \ell_\gamma | \{g\} \cap \{\ell_\beta\}^\perp)\|_P^2 \geq \|\Pi(\delta' \ell_\gamma | \{\ell_\gamma\} \cap \{\ell_\beta\}^\perp)\|_P^2 = \|\delta' \ell_\gamma\|_P^2 = \kappa_0.$$

Consider the efficient score function $\tilde{\ell}_\theta = \Pi(\ell_\theta|A^\perp)$. Let $\tilde{\ell}_\theta(\theta)$ be the efficient score function with the parameter θ_0 replaced by θ . Choose the function $\tilde{\ell}_\theta(\theta)$ for the moment condition, and let κ_A denote the noncentrality parameter of the test.

Proposition 9 *For any κ , $\kappa \leq \kappa_A \leq \kappa_0$.*

Proof: By Theorem 5

$$\begin{aligned}\kappa_A &= \|\Pi(\delta' \ell_\gamma | \{\tilde{\ell}_\theta\}) - \Pi(\delta' \ell_\gamma | \Pi(\{\ell_\beta\} | \{\tilde{\ell}_\theta\}))\|_P^2 \\ &= \|\Pi(\delta' \ell_\gamma | A^\perp) - \Pi(\delta' \ell_\gamma | \Pi(\{\ell_\beta\} | A^\perp))\|_P^2 \\ &= \|\Pi(\delta' \ell_\gamma | A^\perp \cap \{\ell_\beta\}^\perp)\|_P^2.\end{aligned}$$

Since $\{g\} \subset A^\perp \subset L_2^0(P_0)$, we have $\kappa \leq \kappa_A \leq \kappa_0$. ■

In this connection, we examine the case of the conditional moment restrictions $E_{\theta,\eta}[g(z; \theta)|x] = 0$. The moment function $\tilde{\ell}_\theta(\theta)$ is given in equation (3.5) under the assumption that $\{h'g\} = A^\perp$. It is easy to find the noncentrality parameter κ_A is the same as that of the test using the moment function $\tilde{\ell}_\theta(\theta)$, which is defined in Theorem 4. Let $H(x) = E[\partial g(z; \theta_0)/\partial \beta|x]$ and $K(x) = E[\partial g(z; \theta_0)/\partial \gamma|x]$. The following theorem gives the power of hypothesis testing with conditional moment restrictions.

Theorem 6 *Under the assumption that $\{h'g\} = A^\perp$, the noncentrality parameter κ_A is given by*

$$\begin{aligned}\kappa_A &= \delta' \{E[K'(x)\Omega^{-1}(x)K(x)] \\ &\quad - E[K'(x)\Omega^{-1}(x)H(x)](E[H'(x)\Omega^{-1}(x)H(x)])^{-1} \\ &\quad \times E[H'(x)\Omega^{-1}(x)K(x)]\} \delta.\end{aligned}$$

Proof: For $\tilde{\ell}_\theta(\theta) = -G'(x)\Omega^{-1}(x)g(z; \theta)$, we have

$$H = E[\partial \tilde{\ell}_\theta(\theta_0)/\partial \beta] = -E[G'(x)\Omega^{-1}(x)H(x)],$$

$$K = E[\partial \tilde{\ell}_\theta(\theta_0)/\partial \gamma] = -E[G'(x)\Omega^{-1}(x)K(x)],$$

and

$$\Omega = E[\tilde{\ell}_\theta(\theta_0)\tilde{\ell}_\theta'(\theta_0)] = E[G'(x)\Omega^{-1}(x)G(x)].$$

Since

$$E[G'(x)\Omega^{-1}(x)G(x)] = \begin{pmatrix} E[H'(x)\Omega^{-1}(x)H(x)] & E[H'(x)\Omega^{-1}(x)K(x)] \\ E[K'(x)\Omega^{-1}(x)H(x)] & E[K'(x)\Omega^{-1}(x)K(x)] \end{pmatrix},$$

we have

$$K'\Omega^{-1}H = E[K'(x)\Omega^{-1}(x)H(x)],$$

for instance. ■

Next, we discuss the general hypothesis testing with respect to θ ,

$$H_0 : a(\theta_n) = 0$$

and

$$H_1 : a(\theta_n) = \delta/\sqrt{n},$$

where $a(\theta)$ is a q_2 -vector of functions of the parameter θ , and δ is a nonzero q_2 -vector. This test can be reduced to the hypothesis testing given in equations (4.1) and (4.2) by some reparameterization. Hence, the above propositions with respect to the noncentrality parameter hold in the general setting, too. See the Appendix for the detail.

Finally, an important special case of the hypothesis testing in (4.1) and (4.2) is when $r = q$, i.e., the number of the equations is equal to that of the parameters in the moment condition. Since $n\hat{g}(\hat{\theta}_{\text{GMM}})'\hat{\Omega}^{-1}\hat{g}(\hat{\theta}_{\text{GMM}}) = o_p(1)$ then, the hypothesis test is asymptotically equivalent to the test of overidentifying restrictions

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

This idea is presented in Section 9 of Newey and McFadden (1994), too. For GEL, $2\{\sum_{i=1}^n \rho(\hat{\lambda}'g(z_i; \hat{\theta}_{\text{GEL}})) - n\rho(0)\} = o_p(1)$ holds. Hence, the hypothesis test is also asymptotically equivalent to the test of overidentifying restrictions

$$2\left\{\sum_{i=1}^n \rho(\bar{\lambda}'g(z_i; \bar{\theta}_{\text{GEL}})) - n\rho(0)\right\} \xrightarrow{d} \chi^2(q_2, \kappa).$$

We say that the unrestricted model is “valid” if $\text{rank}(G) = q$, and is “invalid” otherwise. If the model is invalid, the identification problem occurs. Since

$\Pi(\ell_\theta|\{g\}) = -G'\Omega^{-1}g$, the dimension of $\Pi(\{\ell_\theta\}|\{g\})$ is q in $L_2^0(P_0)$ if the model is valid. Assumptions 8 and 9 guarantee $\text{rank}(G) = q$ for the usual hypothesis testing, but we only assume $\text{rank}(H) = q_1$, i.e., the validity of the restricted model here. The following theorem gives a new interpretation of testing overidentifying restrictions.

Theorem 7 *If the unrestricted model is valid in the test of overidentifying restrictions, then $\kappa = 0$ if and only if $\delta = 0$. If the unrestricted model is invalid, then there exists some $\delta \neq 0$ such that $\kappa = 0$.*

Proof: For the first assertion, by Theorem 5

$$\kappa = \|\Pi(\delta'\ell_\gamma|\{g\}) - \Pi(\delta'\ell_\gamma|\Pi(\{\ell_\beta\}|\{g\}))\|_P^2.$$

If $\kappa = 0$, it is the case that $\delta'\Pi(\ell_\gamma|\{g\}) \in \Pi(\{\ell_\beta\}|\{g\})$. Since the dimension of $\Pi(\{\ell_\theta\}|\{g\})$ is q , δ must be zero. For the second assertion, since the dimension of $\Pi(\{\ell_\theta\}|\{g\})$ is less than q , there exists a nonzero constant vector $(\delta', \alpha')'$ such that $\delta'\Pi(\ell_\gamma|\{g\}) = \alpha'\Pi(\ell_\beta|\{g\})$. Because the dimension of $\Pi(\{\ell_\beta\}|\{g\})$ is q_1 , δ must be nonzero. ■

Newey (1985a) pointed out that the test of overidentifying restrictions can easily fail to detect a misspecified model since there are directions for which the noncentrality parameter κ is zero when general misspecification is present. Such inconsistency of the test is widely known. See standard textbooks such as Hamilton (1994) and Hayashi (2000) for instance. See also Guggenberger (2012) for recent studies. However, Theorem 7 argues that as long as the unrestricted model is valid, the test does not fail to detect misspecification of the moment condition. If the noncentrality parameter κ is zero, there is no misspecification or the unrestricted model is invalid.

We can easily extend the results of Theorem 7 to the cases $r > q$ and $r < q$. Theorem 7 is also applicable to the case $r > q$, but the test is inefficient because the degrees of freedom $r - q_1$ is greater than q_2 . One should rather use the hypothesis tests given in Lemmas 4 and 5, which have q_2 degrees of freedom. If $r < q$, then for any alternative the corresponding unrestricted model is automatically invalid, and the test is inconsistent. The above results are summarized in the following proposition.

Proposition 10 *The test of overidentifying restrictions is consistent against an alternative if and only if the corresponding unrestricted model is valid.*

Proof: The result follows similarly to the proof of Theorem 7. ■

A practical implication of Theorem 7 and Proposition 10 is as follows. When choosing an r -vector moment function, applied researchers should imagine the unrestricted model with r parameters that corresponds to an alternative they want to detect. As long as the unrestricted model is valid, the test of overidentifying restrictions is consistent against the alternative. Here are some simple examples.

Example 6. Consider the linear regression model

$$y = \alpha + \beta x + \varepsilon$$

and the instrumental variable z , which is correlated with x but uncorrelated with ε , where all the variables are one-dimensional. We choose the 3-vector function g_R for the moment condition under the null hypothesis $E[x\varepsilon] = 0$,

$$g_R = \begin{pmatrix} y - \alpha - \beta x \\ z(y - \alpha - \beta x) \\ x(y - \alpha - \beta x) \end{pmatrix}.$$

Under the alternative hypothesis $E[x\varepsilon] \neq 0$, we have the unrestricted model g_U ,

$$g_U = \begin{pmatrix} y - \alpha - \beta x \\ z(y - \alpha - \beta x) \\ x(y - \alpha - \beta x) - \gamma \end{pmatrix}.$$

Then

$$G = \begin{pmatrix} -1 & -E[x] & 0 \\ -E[z] & -E[zx] & 0 \\ -E[x] & -E[x^2] & -1 \end{pmatrix},$$

and $|G| = -\text{Cov}[z, x] \neq 0$. Since the unrestricted model is valid, the test of overidentifying restrictions using g_R is consistent against the alternative $E[x\varepsilon] \neq 0$.

Example 7. Consider the linear regression model

$$y = \alpha + \beta x + \gamma z + \varepsilon,$$

where all the variables are one dimensional and $E[x\varepsilon] = E[z\varepsilon] = 0$. We choose the 3-vector function g_R for the moment condition under the null hypothesis $\beta = \gamma$,

$$g_R = \begin{pmatrix} y - \alpha - \beta(x + z) \\ x(y - \alpha - \beta(x + z)) \\ z(y - \alpha - \beta(x + z)) \end{pmatrix}.$$

Suppose that there is multicollinearity, say $z = cx$, where c is a constant. Under the alternative hypothesis $\beta \neq \gamma$, we have the unrestricted model g_U ,

$$g_U = \begin{pmatrix} y - \alpha - \beta x - \gamma z \\ x(y - \alpha - \beta x - \gamma z) \\ z(y - \alpha - \beta x - \gamma z) \end{pmatrix}.$$

Then

$$G = \begin{pmatrix} -1 & -E[x] & -E[z] \\ -E[x] & -E[x^2] & -E[xz] \\ -E[z] & -E[zx] & -E[z^2] \end{pmatrix},$$

and $|G| = 0$. Since the unrestricted model is invalid, the test of overidentifying restrictions using g_R is inconsistent against the alternative $\beta \neq \gamma$.

Example 8. This example corresponds to Example 9.1 of Newey and McFadden (1994). Consider the linear regression model

$$y = x'\beta + \varepsilon \quad \text{with } E[\varepsilon|x] = 0 \text{ and } E[\varepsilon^2|x] = \sigma^2,$$

where $x, \beta \in R^q$ and $E[xx']$ is assumed to be nonsingular. To test normality of ε in addition to estimating β and σ^2 , we set up the $(q + 3)$ -vector moment function g_R under the null hypothesis,

$$g_R = \begin{pmatrix} x(y - x'\beta) \\ (y - x'\beta)^2 - \sigma^2 \\ (y - x'\beta)^3/\sigma^3 \\ (y - x'\beta)^4/\sigma^4 - 3 \end{pmatrix}.$$

Under the alternative hypothesis, i.e., nonnormality of ε , we have the unrestricted model g_U ,

$$g_U = \begin{pmatrix} x(y - x'\beta) \\ (y - x'\beta)^2 - \sigma^2 \\ (y - x'\beta)^3/\sigma^3 - \gamma_1 \\ (y - x'\beta)^4/\sigma^4 - 3 - \gamma_2 \end{pmatrix}.$$

Then

$$G = \begin{pmatrix} -E[xx'] & 0 & 0 & 0 \\ \dots & -1 & 0 & 0 \\ \dots & \cdot & -1 & 0 \\ \dots & \cdot & \cdot & -1 \end{pmatrix},$$

and $|G| = (-1)^{q+3}|E[xx']| \neq 0$. Since the unrestricted model is valid, the test of overidentifying restrictions using g_R is consistent against the alternative, nonnormality of ε .

Example 9. Consider the linear regression model

$$y = x'\beta + \varepsilon$$

and the instrumental variable $z \in R^r$, where $x, \beta \in R^{q_1}$ and $z = (z'_1, z'_2)'$ with $z_1 \in R^{(r-q_2)}$ and $z_2 \in R^{q_2}$. The rank of $E[zx']$ is assumed to be q_1 . Under the null hypothesis we have $E[z\varepsilon]=0$. Under the alternative we have $E[z_1\varepsilon] = 0$ but $E[z_2\varepsilon] \neq 0$. We choose the r -vector function g_R for the moment condition under the null hypothesis,

$$g_R = \begin{pmatrix} z_1(y - x'\beta) \\ z_2(y - x'\beta) \end{pmatrix}.$$

Under the alternative hypothesis, we have the unrestricted model g_U ,

$$g_U = \begin{pmatrix} z_1(y - x'\beta) \\ z_2(y - x'\beta) - \gamma \end{pmatrix}.$$

Suppose that $q_1 + q_2 = r$ and $|E[z_1x']| \neq 0$. Then

$$G = \begin{pmatrix} -E[z_1x'] & 0 \\ -E[z_2x'] & -I_{q_2} \end{pmatrix},$$

and $|G| = (-1)^r |E[z_1 x']| \neq 0$. Since the unrestricted model is valid, the test of overidentifying restrictions using g_R is consistent against the alternative $E[z_2 \varepsilon] \neq 0$. Next, suppose that $q_1 + q_2 > r$. Then the rank of G is less than $q_1 + q_2$. Hence, the test is inconsistent against the alternative $E[z_2 \varepsilon] \neq 0$. When $q_1 + q_2 < r$, this is a special case of the one considered by Eichenbaum, Hansen, and Singleton (1988). Suppose that $\text{rank}(E[z_1 x']) = q_1$. It is not difficult to show that $\text{rank}(G) = q_1 + q_2$. The test of overidentifying restrictions is consistent against the alternative $E[z_2 \varepsilon] \neq 0$ with $r - q_1$ degrees of freedom. Hypothesis tests given in Lemmas 4 and 5 are more efficient since they have q_2 degrees of freedom. They are as efficient as the test defined by equation (C.1) of Eichenbaum, Hansen, and Singleton (1988), which has q_2 degrees of freedom, too.

APPENDIX

In this Appendix we consider the general hypothesis testing. Let $a(\theta)$ be a q_2 -vector of continuously differentiable functions of the parameter θ in a neighborhood of θ_0 . Assume that $A = \partial a(\theta_0)/\partial \theta$ has rank q_2 . Consider the general hypothesis testing with respect to θ

$$H_0 : a(\theta_n) = 0$$

and

$$H_1 : a(\theta_n) = \delta/\sqrt{n},$$

where $\theta_n \rightarrow \theta_0$ and $\delta \in R^{q_2}$. Since A has rank q_2 , there exists a $q_1 \times q$ matrix Φ such that the $q \times q$ matrix

$$M = \begin{pmatrix} \Phi \\ A \end{pmatrix}$$

is nonsingular. Define the reparameterization

$$\lambda(\theta) = \begin{pmatrix} \beta(\theta) \\ \gamma(\theta) \end{pmatrix} = \begin{pmatrix} \Phi \theta \\ a(\theta) + \gamma_0 \end{pmatrix}.$$

Then the reparameterization is one to one and has a continuously differentiable inverse in a neighborhood of θ_0 by the inverse function theorem. Let

$\bar{g}(z; \lambda(\theta)) = g(z; \theta)$ and $\lambda_0 = \lambda(\theta_0)$. Let $\bar{G} = E[\partial \bar{g}(z; \lambda_0)/\partial \lambda]$, $\bar{H} = E[\partial \bar{g}(z; \lambda_0)/\partial \beta]$, and $\bar{K} = E[\partial \bar{g}(z; \lambda_0)/\partial \gamma]$. Then $\bar{G} = (\bar{H}, \bar{K})$. Since $G = E[\partial g(z; \theta_0)/\partial \theta]$, we have $G = \bar{G}M$, where $M = \partial \lambda(\theta_0)/\partial \theta$. Let $(\beta_n, \gamma_n) = (\beta(\theta_n), \gamma(\theta_n))$. Note that $(\beta_n, \gamma_n) \rightarrow (\beta_0, \gamma_0)$ as $\theta_n \rightarrow \theta_0$. Consider the constrained GMM and GEL estimators

$$\bar{\theta}_{\text{GMM}} = \arg \min_{\theta \in \Theta} \hat{g}(\theta)' \hat{\Omega}^{-1} \hat{g}(\theta) \quad \text{subject to} \quad a(\theta) = 0, \quad (\text{A.1})$$

where $\hat{\Omega}$ is some preliminary estimator satisfying $\hat{\Omega} \xrightarrow{p} \Omega$, and

$$\bar{\theta}_{\text{GEL}} = \arg \min_{\theta \in \Theta} \sup_{\lambda \in \hat{\Lambda}_n(\theta)} \sum_{i=1}^n \rho(\lambda' g(z_i; \theta)) \quad \text{subject to} \quad a(\theta) = 0. \quad (\text{A.2})$$

Since equations (A.1) and (A.2) correspond to equations (4.3) and (4.4), respectively, through the reparameterization, the following lemmas are given as the corollaries of Lemmas 4 and 5.

Lemma A.1: *If Assumptions 8 and 9 are satisfied, then under the local alternatives*

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

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

Lemma A.2: *If Assumptions 9 and 10 are satisfied, then under the local alternatives*

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

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

Proof: These two lemmas have a common proof. The above reparameterization gives definitions $\bar{g}(z; \lambda) = g(z; \theta)$ and $\bar{f}(z; \lambda, \eta_0) = f(z; \theta, \eta_0)$. It is not difficult to show that Assumptions 8-10 are also satisfied with respect to $\bar{g}$ and $\bar{f}$ in a compact neighborhood of λ_0 . Hence, Lemmas 4 and 5 are applicable to them, and thus yield

$$\begin{aligned} \kappa &= \delta' \{ \bar{K}' \Omega^{-1} \bar{K} - \bar{K}' \Omega^{-1} \bar{H} (\bar{H}' \Omega^{-1} \bar{H})^{-1} \bar{H}' \Omega^{-1} \bar{K} \} \delta \\ &= \delta' \{ \bar{A} (\bar{G}' \Omega^{-1} \bar{G})^{-1} \bar{A}' \}^{-1} \delta, \end{aligned}$$

where $\bar{A} = [0, I_{q_2}]$, a $q_2 \times q$ matrix. Since $\bar{A} = AM^{-1}$ and $\bar{G} = GM^{-1}$, we have

$$\begin{aligned}\kappa &= \delta' \{AM^{-1}[(GM^{-1})' \Omega^{-1}(GM^{-1})]^{-1}(AM^{-1})'\}^{-1} \delta \\ &= \delta' \{A(G' \Omega^{-1} G)^{-1} A'\}^{-1} \delta.\end{aligned}$$

This completes the proof. ■

Above results coincide with the one given by Theorem 9.2 of Newey and McFadden (1993).

REFERENCES

- Amemiya, T. (1977): “The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model,” *Econometrica*, 45, 955–968.
- Bickel, P. J., C. A. J. Klaassen, Y. Ritov, and J. A. Wellner (1993): *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins University Press.
- Carrasco, M., and J. P. Florens (2014): “On the Asymptotic Efficiency of GMM,” *Econometric Theory*, 30, 372–406.
- Chamberlain, G. (1987): “Asymptotic Efficiency in Estimation with Conditional Moment Restrictions,” *Journal of Econometrics*, 34, 305–334.
- Donald, S. G., G. W. Imbens, and W. K. Newey (2003): “Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions,” *Journal of Econometrics*, 117, 55–93.
- 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): “Hypothesis Testing with GMM and GEL Estimators,” *The Aoyama Journal of Economics*, forthcoming.
- Guggenberger, P. (2012): “A Note on the (In)consistency of the Test of Overidentifying Restrictions and the Concepts of True and Pseudo-True Parameters,” *Economics Letters*, 117, 901–904.
- 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.
- 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. (1990a): "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 58, 809–837.
- Newey, W. K. (1990b): "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics*, 5, 99–135.
- 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.
- Qin, J., and J. Lawless (1994): "Empirical Likelihood and General Estimating Equations," *Annals of Statistics*, 22, 300–325.
- Severini, T. A., and G. Tripathi (2001): "A Simplified Approach to Computing Efficiency Bounds in Semiparametric Models," *Journal of Econometrics*, 102, 23–66.
- van der Vaart, A. W. (1998): *Asymptotic Statistics*. Cambridge: Cambridge University Press.