LARGE DEVIATIONS OF GENERALIZED METHOD OF MOMENTS AND EMPIRICAL LIKELIHOOD ESTIMATORS

By

Taisuke Otsu

February 2011

COWLES FOUNDATION DISCUSSION PAPER NO. 1783

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Large deviations of generalized method of moments and empirical likelihood estimators

Taisuke Otsu*

Cowles Foundation & Department of Economics

Yale University

February 2011

Abstract

This paper studies large deviation properties of the generalized method of moments and generalized empirical likelihood estimators for moment restriction models. We consider two cases for the data generating probability measure: the model assumption and local deviations from the model assumption. For both cases, we derive conditions where these estimators have exponentially small error probabilities for point estimation.

*E-mail: taisuke.otsu@yale.edu. Website: http://cowles.econ.yale.edu/faculty/otsu.htm. Address: P.O. Box 208281, New Haven, CT 06520-8281, USA. Phone: +1-203-432-9771. Fax: +1-203-432-6167.

The author thanks the managing editor and referees for helpful comments. Financial support from the National Science Foundation (SES-0720961) is gratefully acknowledged.

1 Introduction

This paper studies large deviation properties of the generalized method of moments (GMM) and generalized empirical likelihood (GEL) estimators for moment restriction models. Since Hansen (1982), there have been numerous empirical applications and theoretical studies on the GMM and related methods. If the model is just-identified, we can apply the conventional method of moments estimator. The large deviation properties of this estimator have been studied elsewhere (e.g., Jensen and Wood (1998) and Inglot and Kallenberg (2003)). If the model is over-identified, the method of moments is not directly applicable. Instead the GMM (Hansen (1982)) or GEL (Smith (1997) and Newey and Smith (2004)) should be applied. Special cases of GEL include empirical likelihood (Qin and Lawless (1994)), continuous updating GMM (Hansen, Heaton and Yaron (1996)), and exponential tilting (Kitamura and Stutzer (1997) and Imbens, Spady and Johnson (1998)).[1] In contrast with the literature on the method of moments estimator, and to the best of our knowledge, there is no theoretical work on the large deviation properties of the GMM and GEL estimators for the over-identified case.[2]

The purpose of this paper is to derive some regularity conditions that guarantee exponentially small large deviation error probabilities for the GMM and GEL estimators both when the model is correctly specified (we refer to this case as the model assumption) and also when there exist local deviations or contaminations from the model assumption. The first setup serves as a benchmark. The second setup is useful to evaluate robustness of the estimators under local

[1] See Kitamura (2007) for a review.

[2] Kitamura and Otsu (2006) proposed a large deviation minimax optimal estimator for moment restriction models, which is different from the existing GMM or GEL estimator. Our focus is on the large deviation properties of the conventional GMM and GEL estimators.

misspecification. It should be noted that although our large deviation results are extensions of the previous results on the method of moments estimator to the over-identified case, theoretical arguments for these extensions are not trivial because (i) the GMM estimator is defined as a minimizer of some quadratic form in the sample mean of the moment function; (ii) the objective function of the two-step GMM estimator contains the first-step estimator; and (iii) the GEL estimator is defined as a minimax solution of the GEL criterion function. Existing technical tools to analyze large deviation estimation errors are not directly applicable to these estimators. Finally, although our large deviation results are important in their own right, they can be employed as a building block for more detailed estimation error analysis. For example, Otsu (2009) used our large deviation results to derive moderate deviation rate functions for the GMM and GEL estimators.

2 Main Results

Suppose we observe a random sample $(X_{1n}, \ldots, X_{nn})$ with support $\mathbb{X} \subseteq \mathbb{R}^{d_x}$ and wish to estimate a vector of unknown parameters $\theta_0 \in \Theta \subseteq \mathbb{R}^{d_\theta}$ defined by moment restrictions

$$E\left[g\left(X, \theta_0\right)\right] = \int g\left(x, \theta_0\right) dP\left(x\right) = 0, \tag{1}$$

where $g : \mathbb{X} \times \Theta \to \mathbb{R}^{d_g}$ is a vector of measurable functions with $d_g \geq d_\theta$. Although our results apply to the just-identified case (i.e., $d_g = d_\theta$), we focus on the over-identified case (i.e., $d_g > d_\theta$). Let $\hat{g}\left(\theta\right) = \frac{1}{n}\sum_{i=1}^{n} g\left(X_{in}, \theta\right)$. We consider the following point estimators for θ_0:

- GMM estimator: $\hat{\theta}_1 = \arg\min_{\theta\in\Theta} \hat{g}\left(\theta\right)' W_n \hat{g}\left(\theta\right)$ with some weight matrix W_n,
- Two-step GMM estimator: $\hat{\theta}_2 = \arg\min_{\theta\in\Theta} \hat{g}\left(\theta\right)' \hat{\Omega}^{-1} \hat{g}\left(\theta\right)$ with $\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} g\left(X_{in}, \hat{\theta}_1\right) g\left(X_{in}, \hat{\theta}_1\right)'$,

- GEL estimator: $\hat{\theta}_3 = \arg\min_{\theta\in\Theta}\max_{\lambda\in\Lambda}\sum_{i=1}^{n}\rho\left(\lambda' g\left(X_{in},\theta\right)\right)$ for some $\rho(\cdot)$.[3]

This paper studies large deviation properties of these estimators under the model assumption (1) or local deviations from the model assumption. More specifically, we consider the following data generating measure on the triangular array $\{(X_{1n},\ldots,X_{nn})\}_{n\in\mathbb{N}}$.

Assumption P. ***(i)*** *For each $n\in\mathbb{N}$, $(X_{1n},\ldots,X_{nn})$ is an i.i.d. sample from the measure P_n with the density $\frac{dP_n}{dP} = 1 + a_n A_n(x)$ with respect to P for some $a_n \to 0$ and $A_n : \mathbb{X}\to\mathbb{R}$ satisfying $\sup_{n\in\mathbb{N}}\sup_{x\in\mathbb{X}}|A_n(x)| < \infty$ and $\int A_n(x)\,dP(x) = 0$.*

(ii) *There exists a unique solution $\theta_0\in\Theta$ for the moment restrictions $\int g(x,\theta_0)\,dP(x) = 0$.*

Hereafter the expectations under P and P_n are denoted by $E[\cdot]$ and $E_n[\cdot]$, respectively. Assumption P, adapted from Inglot and Kallenberg (2003) to the moment restriction setup, allows two cases for the data generating measure P_n: (a) model assumption (i.e., $a_n = 0$), where the data are generated from $P_n = P$ and the moment restrictions (1) are satisfied; and (b) local contamination (i.e., $a_n \neq 0$), where the data are generated from $P_n \neq P$ and the moment restrictions may or may not be satisfied for $n\in\mathbb{N}$ even though P_n converges to P as $n\to\infty$.

The focus of this paper is on large deviation properties of the GMM and GEL estimators under the model assumption and local contaminations. In particular, we investigate whether these estimators have exponentially small estimation error probabilities for θ_0. Exponentially small probability events and estimation error probabilities are defined as follows.

Definition. ***(i)*** ***(ESP)*** *A sequence of events (or subsets in $\mathbb{X}^n$) $\{B_n\}_{n\in\mathbb{N}}$ has exponentially small probability under $\{P_n\}_{n\in\mathbb{N}}$ (we say "B_n has ESP") when* ***(a)*** *there exist $C, c > 0$ such that*

[3] For example, $\rho(v) = -(1+v)^2/2$ (continuous updating GMM), $\rho(v) = \log(1-v)$ (empirical likelihood), and $\rho(v) = -\exp(v)$ (exponential tilting).

$P_n(B_n) \leq Ce^{-cn}$ *for all* n *large enough; and* ***(b)*** *under the model assumption,* $P_n = P$, *there exist* $\tilde{C}, \tilde{c} > 0$ *such that* $P(B_n) \leq \tilde{C}e^{-\tilde{c}n}$ *for all* $n \in \mathbb{N}$.

(ii) (ESEP) *An estimator* $\hat{\theta}$ *for* θ_0 *has exponentially small error probability (we say "*$\hat{\theta}$ *has ESEP") when* ***(a)*** *the event* $\left\{\left|\hat{\theta} - \theta_0\right| > \epsilon\right\}$ *has ESP for each* $\epsilon > 0$*; and* ***(b)*** *there exists* $\bar{C} > 0$ *such that the event* $\left\{\left|\hat{\theta} - \theta_0\right| > \epsilon \text{ or } \hat{\theta} \text{ is not unique}\right\}$ *has ESP for each* $\epsilon \in \left(0, \bar{C}\right)$.

The following assumptions will imply that the GMM and GEL estimators have ESEP. Let $|A| = \text{trace}(A'A)$ be the Euclidean norm of a scalar, vector, or matrix A, $\text{int}(B)$ be the interior of a set B, "a.e." mean "almost every", $\rho_1(v) = \frac{d\rho(v)}{dv}$, and $\rho_2(v) = \frac{d^2\rho(v)}{dv^2}$.

Assumption G1. ***(i)*** Θ *is compact and* $\theta_0 \in \textit{int}(\Theta)$. *There exist* $L : \mathbb{X} \to [0, \infty)$ *and* $\alpha, T_1 > 0$ *such that* $|g(x, \theta_1) - g(x, \theta_2)| \leq L(x)|\theta_1 - \theta_2|^\alpha$ *for all* $\theta_1, \theta_2 \in \Theta$ *and a.e.* x*, and* $E[\exp(T_1 L(X))] < \infty$. *For each* $\theta \in \Theta$*, there exists* $T_2 > 0$ *satisfying* $E[\exp(T_2|g(X, \theta)|)] < \infty$.

(ii) *There exist* $H : \mathbb{X} \to [0, \infty)$*,* $\beta, T_3 > 0$*, and a neighborhood* $\mathcal{N}$ *around* θ_0 *such that* $\left|\frac{\partial g(x,\theta)}{\partial \theta'} - \frac{\partial g(x,\theta_0)}{\partial \theta'}\right| \leq H(x)|\theta - \theta_0|^\beta$ *for all* $\theta \in \mathcal{N}$ *and a.e.* x*, and* $E[\exp(T_3 H(X))] < \infty$. *There exists* $T_4 > 0$ *satisfying* $E\left[\exp\left(T_4\left|\frac{\partial g(X,\theta_0)}{\partial \theta'}\right|\right)\right] < \infty$. $E\left[\frac{\partial g(X,\theta_0)}{\partial \theta'}\right]$ *has the full column rank.*

Assumption W. *For each* $\epsilon > 0$*, the event* $\{|W_n - W| > \epsilon\}$ *has ESP for some positive definite symmetric matrix* W.

Assumption G2. *For each* $n \in \mathbb{N}$*, there exist* $T_5, T_6, T_7 > 0$ *such that* $E\left[\exp\left(T_5 L(X)^2\right)\right] < \infty$, $E[\exp(T_6 L(X)|g(X, \theta_0)|)] < \infty$*, and* $E\left[\exp\left(T_7\left|g(X, \theta_0)g(X, \theta_0)'\right|\right)\right] < \infty$. $E\left[g(X, \theta_0)g(X, \theta_0)'\right]$ *is positive definite.*

Assumption G3. ***(i)*** $\rho(v)$ *is strictly concave and* $\rho_1(0) = \rho_2(0) = -1$. Λ *is compact and* $\mathbf{0} \in \textit{int}(\Lambda)$. *For each* $\theta \in \Theta$*,* $\bar{\lambda}(\theta) = \arg\max_{\lambda \in \Lambda} E[\rho(\lambda' g(X, \theta))]$ *belongs to* $\textit{int}(\Lambda)$. $g(x, \theta)$ *is differentiable on* Θ *for a.e.* x. *There exists* $T_8 > 0$ *satisfying* $E[\exp(T_8|g(X, \theta_0)|)] < \infty$. *For* *each $\theta \in \Theta$, there exist $T_9 > 0$ and neighborhoods $\mathcal{N}_\theta$ and $\mathcal{N}'_{\bar{\lambda}(\theta)}$ around θ and $\bar{\lambda}(\theta)$, respectively, satisfying $E\left[\exp\left(T_9 \sup_{\vartheta \in \mathcal{N}_\theta} \sup_{\lambda \in \mathcal{N}'_{\bar{\lambda}(\theta)}} \left|\rho_1(\lambda' g(X,\vartheta)) \frac{\partial g(X,\vartheta)}{\partial \theta'}\right|\right)\right] < \infty$.*

(ii) *There exist $T_{10} > 0$ and neighborhoods $\mathcal{N}_\rho$ and $\mathcal{N}'_\rho$ around θ_0 and $\mathbf{0}$, respectively, satisfying $E\left[\exp\left(T_{10} \sup_{\theta \in \mathcal{N}_\rho} \sup_{\lambda \in \mathcal{N}'_\rho} \left|\rho_2(\lambda' g(X,\theta)) g(X,\theta) g(X,\theta)'\right|\right)\right] < \infty$. $E\left[g(X,\theta_0) g(X,\theta_0)'\right]$ is positive definite.*

Assumption G1 (i) restricts the global shape of the moment function g over the parameter space Θ. The Lipschitz-type condition on g is common in the literature (e.g. Jensen and Wood (1998)) and is satisfied with $\alpha = 1$ if g is differentiable on Θ for a.e. x and $E\left[\tilde{T}_1 \sup_{\theta \in \Theta} \left|\frac{\partial g(x,\theta)}{\partial \theta'}\right|\right] < \infty$ for some $\tilde{T}_1 > 0$. The conditions for exponential moments are typically required to control large deviation probabilities. Assumption G1 (ii), which controls the local shape of the moment function around θ_0, is required only to guarantee the uniqueness of $\hat{\theta}_1$. Assumption W, a high-level assumption on the weight matrix W_n, should be checked for each choice of W_n. Assumption G2 is required only for the two-step GMM estimator $\hat{\theta}_2$ to guarantee that Assumption W holds for $W_n = \hat{\Omega}^{-1}$. Assumption G3 is used for the GEL estimator. Assumption G3 (i) replaces Assumption G1 (i). The conditions on the GEL criterion function ρ are satisfied by the examples listed in Section 2. Although technical arguments become more complicated, the compactness assumption on Λ may be avoided by adding an assumption similar to the one used by Inglot and Kallenberg (2003, Assumption (R2')) which controls the global behavior of the objective function outside some compact set for λ. The last condition in Assumption G3 (i), which corresponds to the condition for $L(X)$ in Assumption G1 (i), restricts the slope of the GEL objective function with respect to θ. This condition needs to be checked for each choice of ρ. Assumption G3 (ii) contains additional assumptions to guarantee the uniqueness of the GEL estimator $\hat{\theta}_3$. This assumption restricts the local curvature of the GEL objective function with respect to λ in a neighborhood of $\mathbf{0}$.

Based on these assumptions, our main theorem is presented as follows.

Theorem. ***(i)*** *Under Assumptions P, G1, and W, the GMM estimator* $\hat{\theta}_1$ *has ESEP.*

(ii) *Under Assumptions P, G1, W, and G2, the two-step GMM estimator* $\hat{\theta}_2$ *has ESEP.*

(iii) *Under Assumptions P, G1 (ii), and G3, the GEL estimator* $\hat{\theta}_3$ *has ESEP.*

Remarks: 1. Based on Definition (ii), this theorem says that (a) under the model assumption the error probabilities of the GMM and GEL estimators are exponentially small for all sample sizes, and under local contaminations the convergence rates of the estimation error probabilities to zero are exponentially fast; and (b) the probabilities for multiple solutions of the GMM and GEL minimization problems are also exponentially small. If one wants to guarantee only Definition (ii)-(a), Assumptions G1 (ii) and G3 (ii) are unnecessary.

2. For the GMM estimator $\hat{\theta}_1$ (i.e., Part (i) of this theorem), Definition (ii)-(a) is shown by verifying conditions for a general lemma to establish the ESEP property for extremum estimators (Lemma A.2 below). Lemma A.2 is a modification of general consistency theorems for extremum estimators (e.g., Newey and McFadden (1994, Theorem 3.1)) to our large deviation context, and thus can be applied to other contexts. On the other hand, for the GEL estimator $\hat{\theta}_3$ (i.e., Part (iii) of this theorem), the minimax form of the estimator prevents us from applying directly the general lemma. Thus, we followed the proof strategy of Newey and Smith (2004, Theorem 3.1), which effectively utilized the minimax form of the estimator. Part (ii) of this theorem is shown by verifying that the optimal weight $\hat{\Omega}^{-1}$ satisfies Assumption W.

3. Although this theorem is important in its own right, it can be used as a building block for more detailed estimation error analysis, such as formal derivations of the large deviation rate functions for the GMM and GEL estimators. To this end, we need to derive not only concrete forms of the constants in the definition of ESEP, but also lower bounds for the large deviation error probabilities (we conjecture that the lower bounds will be characterized by the Kullback-Leibler divergence between P and the set of measures satisfying the moment restrictions). Otsu (2009) used the above theorem to derive moderate deviation rate functions for the GMM and GEL estimators.

A Mathematical Appendix

We repeatedly use the following lemma to show that events associated with means have ESP.

Lemma A.1. *Let $f : \mathbb{X} \to \mathbb{R}$ be a measurable function. Suppose that Assumption P (i) holds and there exists $T > 0$ satisfying $E\left[\exp\left(Tf\left(X\right)\right)\right] < \infty$. Then the event $\left\{\frac{1}{n}\sum_{i=1}^{n} f\left(X_{in}\right) > z\right\}$ has ESP for each $z \in \left(E\left[f\left(X\right)\right], \infty\right)$.*

Proof. Pick any $n \in \mathbb{N}$ and $z \in \left(E\left[f\left(X\right)\right], \infty\right)$. Let $M\left(t\right) = E\left[\exp\left\{t\left(f\left(X\right) - z\right)\right\}\right]$. Since $M\left(0\right) = 1$, $\left.\frac{dM(t)}{dt}\right|_{t=0} = E\left[f\left(X\right)\right] - z < 0$, and $M\left(t\right)$ is continuous at each $t \in [0, T]$, there exists $t^* = \arg\min_{t \in [0,T]} M\left(t\right)$ with $M\left(t^*\right) < 1$. The Markov inequality and Assumption P (i) imply

$$P_n\left(\frac{1}{n}\sum_{i=1}^{n} f\left(X_{in}\right) > z\right) \leq \left(E_n\left[\exp\left\{t^*\left(f\left(X_{in}\right) - z\right)\right\}\right]\right)^n \leq \left\{\left(1 + C_A a_n\right) M\left(t^*\right)\right\}^n,$$

where $C_A = \sup_{n \in \mathbb{N}} \sup_{x \in \mathbb{X}} \left|A_n\left(x\right)\right| < \infty$. Since $a_n \to 0$ and $M\left(t^*\right) < 1$, it holds $\left(1 + C_A a_n\right) M\left(t^*\right) < 1$ for all n large enough. Thus Definition (i)-(a) is satisfied. The same argument with setting $a_n = 0$ guarantees Definition (i)-(b). $\square$

The next lemma, an adaptation of Newey and McFadden (1994, Theorem 3.1), provides conditions where an extremum estimator $\hat{\theta} = \arg\min_{\theta \in \Theta} Q_n\left(\theta\right)$ has ESEP.

Lemma A.2. *Suppose that (i) Θ is compact; (ii) the event $\{\sup_{\theta\in\Theta}|Q_n(\theta)-Q_0(\theta)|>\epsilon_1\}$ has ESP for each $\epsilon_1>0$; (iii) the limiting objective function $Q_0(\theta)$ is continuous at each $\theta\in\Theta$ and is uniquely minimized at $\theta_0\in\Theta$. Then the event $\left\{\left|\hat{\theta}-\theta_0\right|>\epsilon\right\}$ has ESP for each $\epsilon>0$.*

Proof. Pick any $n\in\mathbb{N}$ and $\epsilon>0$. Let $\delta=\inf_{\theta\in\Theta,|\theta-\theta_0|\geq\epsilon}Q_0(\theta)-Q_0(\theta_0)>0$ (by conditions (i) and (iii)). Set inclusion relations imply

$$
\begin{aligned}
& P_n\left(\left|\hat{\theta}-\theta_0\right|>\epsilon\right)\leq P_n\left(Q_0\left(\hat{\theta}\right)\geq Q_0(\theta_0)+\delta\right)\\
\leq\;& P_n\left(Q_0\left(\hat{\theta}\right)\geq Q_0(\theta_0)+\delta,\ \sup_{\theta\in\Theta}|Q_n(\theta)-Q_0(\theta)|\leq\frac{\delta}{3}\right)+P_n\left(\sup_{\theta\in\Theta}|Q_n(\theta)-Q_0(\theta)|>\frac{\delta}{3}\right).
\end{aligned}
$$

Since the first term is dominated by $P_n\left(Q_n\left(\hat{\theta}\right)\geq Q_n(\theta_0)+\frac{\delta}{3}\right)=0$, condition (ii) implies the conclusion. $\square$

Hereafter, let $\mathbf{x}_n=(x_{1n},\ldots,x_{nn})$, $\hat{G}(\theta)=\frac{1}{n}\sum_{i=1}^n\frac{\partial g(X_{in},\theta)}{\partial\theta'}$, $\hat{L}=\frac{1}{n}\sum_{i=1}^n L(X_{in})$, $\hat{H}=\frac{1}{n}\sum_{i=1}^n H(X_{in})$, $G=E\left[\frac{\partial g(X,\theta_0)}{\partial\theta'}\right]$, and $\Omega=E\left[g(X,\theta_0)\,g(X,\theta_0)'\right]$.

Proof of Theorem. Proof of (i). Verification of Definition (ii)-(a). To this end, we check the conditions of Lemma A.2 with $Q_n(\theta)=\hat{g}(\theta)'W_n\hat{g}(\theta)$ and $Q_0(\theta)=E[g(X,\theta)]'WE[g(X,\theta)]$. From Assumptions P (ii), G1 (i), and W, condition (iii) of Lemma A.2 is satisfied. Since Θ is compact, it remains to check condition (ii) of Lemma A.2. Now pick any $\epsilon>0$ and $n\in\mathbb{N}$. Since the compact set Θ is covered by a finite sequence of balls $\{\Theta_j\}_{j=1}^J$ with radius $c_\epsilon>0$ and centers $\{\theta_j\}_{j=1}^J$, the triangle inequality yields

$$
\begin{aligned}
& \sup_{\theta\in\Theta}|Q_n(\theta)-Q_0(\theta)|\\
\leq\;& \max_{1\leq j\leq J}\sup_{\theta\in\Theta_j}|Q_n(\theta)-Q_n(\theta_j)|+\max_{1\leq j\leq J}|Q_n(\theta_j)-Q_0(\theta_j)|+\max_{1\leq j\leq J}\sup_{\theta\in\Theta_j}|Q_0(\theta_j)-Q_0(\theta)|\\
=\;& \max_{1\leq j\leq J}T_{1j}+\max_{1\leq j\leq J}T_{2j}+\max_{1\leq j\leq J}T_{3j},
\end{aligned}
$$

where T_{1j}, T_{2j}, and T_{3j} are implicitly defined. Define the event

$$B_{1n} = \left\{ \max_{1 \le j \le J} \left| \hat{g}(\theta_j) - E[g(X, \theta_j)] \right| \le c_\epsilon, \ \ |W_n - W| \le c_\epsilon, \ \hat{L} \le E[L(X)] + 1 \right\}.$$

From the triangle inequality and Assumptions G1 (i) and W, there exist $C_1, C_2 > 0$ such that

$$\begin{aligned} T_{1j} &\le \sup_{\theta \in \Theta_j} \left| (\hat{g}(\theta) - \hat{g}(\theta_j))' W_n (\hat{g}(\theta) - \hat{g}(\theta_j)) \right| + \sup_{\theta \in \Theta_j} \left| \hat{g}(\theta_j)' W_n (\hat{g}(\theta) - \hat{g}(\theta_j)) \right| \le C_1 \left(c_\epsilon^{2\alpha} + c_\epsilon^{\alpha} \right), \\ T_{2j} &\le \left| (\hat{g}(\theta_j) - E[g(X, \theta_j)])' W_n (\hat{g}(\theta_j) - E[g(X, \theta_j)]) \right| + 2 \left| E[g(X, \theta_j)]' W_n (\hat{g}(\theta_j) - E[g(X, \theta_j)]) \right| \\ &\quad + \left| E[g(X, \theta_j)]' (W_n - W) E[g(X, \theta_j)] \right| \le C_2 \left(c_\epsilon^2 + 3 c_\epsilon \right), \end{aligned}$$

for a.e. $\mathbf{x}_n \in B_{1n}$ and all $j = 1, \ldots, J$. Similarly, we obtain $T_{3j} \le C_3 (c_\epsilon^{2\alpha} + c_\epsilon^{\alpha})$ for some $C_3 > 0$. By choosing c_ϵ small enough to satisfy $C_1 (c_\epsilon^{2\alpha} + c_\epsilon^{\alpha}) + C_2 (c_\epsilon^2 + 3c_\epsilon) + C_3 (c_\epsilon^{2\alpha} + c_\epsilon^{\alpha}) < \epsilon$, we obtain $P_n(\{\sup_{\theta \in \Theta} |Q_n(\theta) - Q_0(\theta)| > \epsilon\} \cap B_{1n}) = 0$, which implies $P_n(\sup_{\theta \in \Theta} |Q_n(\theta) - Q_0(\theta)| > \epsilon) \le P_n(B_{1n}^c)$. Since B_{1n}^c has ESP from Assumption W and Lemma A.1 (setting $f(x) = L(x)$ and $\pm g_l(x, \theta_j)$ for $l = 1, \ldots, d_g$ and $j = 1, \ldots, J$), condition (ii) of Lemma A.2 is satisfied.

Verification of Definition (ii)-(b). Pick any $n \in \mathbb{N}$. Let $B_{2n} = \left\{ \inf_{\theta \in \Theta, |\theta - \theta_0| > \epsilon} Q_n(\theta) > Q_n(\theta_0) \right\}$ and $B_{3n} = \left\{ |\hat{g}(\theta_0)| \le \epsilon, \ \left| \hat{G}(\theta_0) - G \right| \le \epsilon, \ |W_n - W| \le \epsilon, \ \hat{L} \le E[L(X)] + 1, \ \hat{H} \le E[H(X)] + 1 \right\}$ for $\epsilon > 0$. By a similar argument to the proof of Lemma A.2, we see that B_{2n}^c has ESP for each $\epsilon > 0$. Also, by Assumption W and Lemma A.1, B_{3n}^c has ESP for each $\epsilon > 0$. Therefore, it is sufficient for the conclusion to show that there exists $\bar{C}_1 > 0$ such that $B_{2n} \cap B_{3n} \subseteq \left\{ \left| \hat{\theta}_1 - \theta_0 \right| \le \epsilon \text{ and } \hat{\theta}_1 \text{ is unique} \right\}$ for all $\epsilon \in (0, \bar{C}_1)$. Since $\theta_0 \in \operatorname{int}(\Theta)$, we can find $\bar{C}_1' > 0$ such that $\{\theta \in \Theta : |\theta - \theta_0| \le \epsilon\} \subset \operatorname{int}(\Theta)$ for all $\epsilon \in (0, \bar{C}_1')$. Note that for each $\epsilon \in (0, \bar{C}_1')$ and a.e. $\mathbf{x}_n \in B_{2n}$, there exists a minimum $\hat{\theta}_1$ which solves the first-order condition $S_n(\theta) = \hat{G}(\theta)' W_n \hat{g}(\theta) = 0$ with respect to θ. Thus, it is sufficient to show that there exists $\bar{C}_1 \in (0, \bar{C}_1')$ satisfying $B_{2n} \cap B_{3n} \subseteq \{S_n(\theta) \text{ is one-to-one in } \{\theta \in \Theta : |\theta - \theta_0| \le \epsilon\}\}$ for all $\epsilon \in (0, \bar{C}_1)$.

Now, pick any $\epsilon \in \left(0, \bar{C}_1'\right)$ and then pick any θ and $\vartheta \neq 0$ to satisfy $\theta, \theta+\vartheta \in \{\theta \in \mathcal{N} : |\theta - \theta_0| \leq \epsilon\}$, where $\mathcal{N}$ appears in Assumption G1 (ii). By the triangle inequality,

$$
\begin{aligned}
& |S_n(\theta+\vartheta) - S_n(\theta)| \\
\geq \; & |G'WG\vartheta| - \left|\hat{G}(\theta_0)' W_n \hat{G}(\theta_0)\vartheta - G'WG\vartheta\right| - \left|S_n(\theta+\vartheta) - S_n(\theta) - \hat{G}(\theta_0)' W_n \hat{G}(\theta_0)\vartheta\right| \\
= \; & |G'WG\vartheta| - |A_1| - |A_2|,
\end{aligned}
$$

where A_1 and A_2 are implicitly defined. For a.e. $\mathbf{x}_n \in B_{2n}$, there exists $C_1 > 0$ such that $|A_1| \leq C_1\epsilon$. By Taylor expansions and Assumptions G1 (ii) and W,

$$
\begin{aligned}
|A_2| \leq \; & \left|\left(\hat{G}(\theta+\vartheta) - \hat{G}(\theta)\right)' W_n \hat{g}(\theta_0)\right| + \left|\left(\hat{G}(\theta+\vartheta) - \hat{G}(\theta)\right)' W_n \hat{G}\left(\tilde{\theta}\right)\right| |\theta - \theta_0| \\
& + \left|\hat{G}(\theta+\vartheta)' W_n \hat{G}\left(\theta + \tilde{\vartheta}\right) - \hat{G}(\theta_0)' W_n \hat{G}(\theta_0)\right| |\vartheta| \leq C_2\left(\epsilon + \epsilon^{\beta}\right),
\end{aligned}
$$

for some $C_2 > 0$, where $\tilde{\theta}$ is a point on the line joining θ and θ_0, and $\tilde{\vartheta}$ is a point on the line joining ϑ and 0. Combining these results with $|G'WG\vartheta| > 0$ (because G has the full column rank and W is positive definite), for a.e. $\mathbf{x}_n \in B_{2n} \cap B_{3n}$ we can find a constant $\bar{C}_1 \in \left(0, \bar{C}_1'\right)$ such that $\{\theta \in \Theta : |\theta - \theta_0| \leq \epsilon\} \subseteq \mathcal{N}$ and $|S_n(\theta+\vartheta) - S_n(\theta)| > 0$ for all $\epsilon \in \left(0, \bar{C}_1\right)$, which implies Definition (ii)-(b). □

Proof of (ii). Omitted. It is obtained by showing that $\hat{\Omega}^{-1}$ satisfies Assumption W. A detailed proof is available from the author upon request. □

Proof of (iii). Verification of Definition (ii)-(a). Let $\hat{P}(\lambda, \theta) = \frac{1}{n}\sum_{i=1}^{n} \rho(\lambda' g(X_{in}, \theta))$ and $\hat{\lambda}(\theta) = \arg\max_{\lambda \in \Lambda} \hat{P}(\lambda, \theta)$. We first derive some properties of $\hat{\lambda}(\theta)$. Pick any $n \in \mathbb{N}$ and $\theta \in \Theta$. Also pick $\epsilon > 0$ small enough to satisfy $\left\{\lambda : \left|\lambda - \bar{\lambda}(\theta)\right| \leq \epsilon\right\} \subset \Lambda$. Define the event $BR_{1n}(\theta) = \left\{\sup_{\lambda \in \Lambda, \left|\lambda - \bar{\lambda}(\theta)\right| > \epsilon} \hat{P}(\lambda, \theta) < \hat{P}\left(\bar{\lambda}(\theta), \theta\right)\right\}$. By applying Inglot and Kallenberg (2003, Theorem 2.1), $BR_{1n}^c(\theta)$ has ESP. Since $\rho(\cdot)$ is strictly concave, the maximizer $\hat{\lambda}(\theta)$ exists uniquely and satisfies $\left|\hat{\lambda}(\theta) - \bar{\lambda}(\theta)\right| \leq \epsilon$ for a.e. $\mathbf{x}_n \in BR_{1n}(\theta)$.

Now define $Q_{\rho n}(\theta) = \sup_{\lambda\in\Lambda} \hat{P}(\lambda,\theta)$ and $B_{4n} = \left\{\inf_{\theta\in\Theta,|\theta-\theta_0|>\epsilon} Q_{\rho n}(\theta) > Q_{\rho n}(\theta_0)\right\}$. Pick any $\epsilon > 0$ and $n \in \mathbb{N}$ again. Using a finite cover of $\{\theta \in \Theta : |\theta-\theta_0| > \epsilon\}$ by a sequence of balls $\{\Theta_j\}_{j=1}^J$ with centers $\{\theta_j\}_{j=1}^J$ and radius c_ϵ, a set inclusion relation implies

$$\inf_{\theta\in\Theta,|\theta-\theta_0|>\epsilon} Q_{\rho n}(\theta) \geq \min_{1\leq j\leq J} \inf_{\theta\in\Theta_j,|\theta-\theta_0|>\epsilon} \left\{\hat{P}\left(\hat{\lambda}(\theta_j),\theta\right) - \hat{P}\left(\hat{\lambda}(\theta_j),\theta_j\right)\right\} + \hat{P}\left(\hat{\lambda}(\theta_j),\theta_j\right).$$

Let $\rho_{1ij}^g = \sup_{\theta\in\mathcal{N}_{\theta_j}} \sup_{\lambda\in\mathcal{N}'_{\bar{\lambda}(\theta_j)}} \left|\rho_1(\lambda' g(X_{in},\theta)) \frac{\partial g(X_{in},\theta)}{\partial\theta'}\right|$ and $BR_{2n} = \left\{\frac{1}{n}\sum_{i=1}^n \rho_{1ij}^g \leq E\left[\rho_{1ij}^g\right] + 1\right\}$. For a.e. $\mathbf{x}_n \in BR_{1n}(\theta_j) \cap BR_{2n}$, an expansion around $\theta = \theta_j$ yields

$$\sup_{\theta\in\Theta_j,|\theta-\theta_0|>\epsilon} \left|\hat{P}\left(\hat{\lambda}(\theta_j),\theta\right) - \hat{P}\left(\hat{\lambda}(\theta_j),\theta_j\right)\right| \leq C_1 c_\epsilon,$$

for some $C_1 > 0$. Also, for a.e. $\mathbf{x}_n \in BR_{1n}(\theta_j)$, $\hat{\lambda}(\theta_j)$ uniquely maximizes $\hat{P}(\lambda,\theta_j)$ with respect to $\lambda \in \Lambda$, i.e., $\delta = \hat{P}\left(\hat{\lambda}(\theta_j),\theta_j\right) - \rho(0) > 0$. On the other hand, for a.e. $\mathbf{x}_n \in BR_{1n}(\theta_0)$,

$$\begin{aligned} Q_{\rho n}(\theta_0) &= \rho(0) - \hat{\lambda}(\theta_0)'\hat{g}(\theta_0) + \frac{1}{2}\hat{\lambda}(\theta_0)'\left[\frac{1}{n}\sum_{i=1}^n \rho_2\left(\tilde{\lambda}' g(X_{in},\theta_0)\right) g(X_{in},\theta_0) g(X_{in},\theta_0)'\right]\hat{\lambda}(\theta_0) \\ &\leq \rho(0) + C_2|\hat{g}(\theta_0)|. \end{aligned}$$

for some $C_2 > 0$, where $\tilde{\lambda}$ is a point on the line joining $\hat{\lambda}(\theta_0)$ and 0, the equality follows from an expansion around $\hat{\lambda}(\theta_0) = 0$, and the inequality follows from the concavity of ρ and $\left|\hat{\lambda}(\theta_0) - \bar{\lambda}(\theta_0)\right| \leq \epsilon$ for a.e. $\mathbf{x}_n \in BR_{1n}(\theta_0)$. Combining these results and choosing c_ϵ small enough, there exists $C_3 > 0$ such that $|\hat{g}(\theta_0)| \geq C_3$ for a.e. $\mathbf{x}_n \in B_{4n}^c \cap \left(\cap_{j=1}^J BR_{1n}(\theta_j) \cap BR_{1n}(\theta_0)\right) \cap BR_{2n}$. Thus, Lemma A.1 implies that $B_{4n}^c \cap \left(\cap_{j=1}^J BR_{1n}(\theta_j) \cap BR_{1n}(\theta_0)\right) \cap BR_{2n}$ has ESP. Since $BR_{1n}^c(\theta_0)$ and $BR_{1n}^c(\theta_j)$ $(j = 1,\ldots,J)$ have ESP by Inglot and Kallenberg (2003, Theorem 2.1) and BR_{2n}^c has ESP by Lemma A.1, B_{4n}^c has ESP, which implies Definition (ii)-(a).

Verification of Definition (ii)-(b). Let $\rho_{1i}^g = \sup_{\theta\in\mathcal{N}_\rho} \sup_{\lambda\in\mathcal{N}'_\rho} \left|\rho_1(\lambda' g(X_{in},\theta)) \frac{\partial g(X_{in},\theta)}{\partial\theta'}\right|$, $\rho_{2i}^g = \sup_{\theta\in\mathcal{N}_\rho} \sup_{\lambda\in\mathcal{N}'_\rho} \left|\rho_2(\lambda' g(X_{in},\theta)) g(X_{in},\theta) g(X_{in},\theta)'\right|$, and

$$B_{5n} = B_{4n} \cap BR_{1n}(\theta_0) \cap \left\{\begin{array}{c} \left|\hat{G}(\theta_0) - G\right| \leq \epsilon,\ \hat{H} \leq E[H(X)] + 1, \\ \frac{1}{n}\sum_{i=1}^n \rho_{1i}^g \leq E\left[\rho_{1i}^g\right] + 1,\ \frac{1}{n}\sum_{i=1}^n \rho_{2i}^g \leq E\left[\rho_{2i}^g\right] + 1 \end{array}\right\}.$$

Pick any $n \in \mathbb{N}$. Since we have already seen that B_{4n}^c has ESP, it is sufficient to show that there exists $\bar{C}_2 > 0$ satisfying $B_{5n} \subseteq \left\{\left|\hat{\theta}_3 - \theta_0\right| \leq \epsilon \text{ and } \hat{\theta}_3 \text{ is unique}\right\}$ for all $\epsilon \in \left(0, \bar{C}_2\right)$. Observe that: (a) by the strict concavity of $\rho\,(v)$ and compactness of Λ (Assumption G3 (i)), the maximum theorem implies that the maximizer $\hat{\lambda}\,(\theta)$ is continuous in $\theta \in \mathcal{N}_\rho$, and (b) Assumptions P (ii) and G3 (i) guarantee that $\bar{\lambda}\,(\theta_0) = 0 \in \text{int}\,(\Lambda)$ and $\left|\hat{\lambda}\,(\theta_0) - \bar{\lambda}\,(\theta_0)\right| = \left|\hat{\lambda}\,(\theta_0)\right| \leq \epsilon$ for a.e. $\mathbf{x}_n \in B_{5n}$ (by Inglot and Kallenberg (2003, Theorem 2.1)). Thus, for a.e. $\mathbf{x}_n \in B_{5n}$, we can pick a constant $\bar{C}_2' > 0$ such that for each $\epsilon \in \left(0, \bar{C}_2'\right)$, $\left|\hat{\lambda}\,(\theta)\right| \leq \left|\hat{\lambda}\,(\theta) - \hat{\lambda}\,(\theta_0)\right| + \left|\hat{\lambda}\,(\theta_0)\right| \leq C_1\epsilon$ for some $C_1 > 0$ and all $\theta \in \{\theta \in \mathcal{N}_\rho : |\theta - \theta_0| \leq \epsilon\}$. On the other hand, from $\theta_0 \in \text{int}\,(\Theta)$, we can find a constant $\bar{C}_2'' > 0$ such that $\{\theta \in \mathcal{N}_\rho : |\theta - \theta_0| \leq \epsilon\} \subset \text{int}\,(\Theta)$ for all $\epsilon \in \left(0, \bar{C}_2''\right)$. Combining these results, for each $\epsilon \in \left(0, \min\left\{\bar{C}_2', \bar{C}_2''\right\}\right)$ and a.e. $\mathbf{x}_n \in B_{5n}$, there exists a minimum $\hat{\theta}_3$ which solves the first-order condition $S_{\rho n}\,(\theta) = \frac{1}{n}\sum_{i=1}^n \rho_1\left(\hat{\lambda}\,(\theta)' g\,(X_{in}, \theta)\right)\left(\frac{\partial g(X_{in},\theta)}{\partial \theta'}\right)' \hat{\lambda}\,(\theta) = 0$ with respect to θ. Thus, it is sufficient for the conclusion to show that there exists $\bar{C}_2 \in \left(0, \min\left\{\bar{C}_2', \bar{C}_2''\right\}\right)$ satisfying $B_{5n} \subseteq \{S_{\rho n}\,(\theta) \text{ is one-to-one in } \{\theta \in \Theta : |\theta - \theta_0| \leq \epsilon\}\}$ for all $\epsilon \in \left(0, \bar{C}_3\right)$.

Now, pick any $\epsilon \in \left(0, \min\left\{\bar{C}_2', \bar{C}_2''\right\}\right)$ and then pick any θ and $\vartheta \neq 0$ to satisfy $\theta, \theta + \vartheta \in \{\theta \in \mathcal{N}_\rho : |\theta - \theta_0| \leq \epsilon\}$. Since $G'\Omega^{-1}G$ is positive definite (Assumption G3 (ii)) and $|S_{\rho n}\,(\theta + \vartheta) - S_{\rho n}\,(\theta)| \geq |G'\Omega^{-1}G\vartheta| - |S_{\rho n}\,(\theta + \vartheta) - S_{\rho n}\,(\theta) + G'\Omega^{-1}G\vartheta|$, it is sufficient to show that $|S_{\rho n}\,(\theta + \vartheta) - S_{\rho n}\,(\theta) + G'\Omega^{-1}G\vartheta| \leq C_2\epsilon$ for some $C_2 > 0$. Observe that

$$
\begin{aligned}
& \left|S_{\rho n}\,(\theta + \vartheta) - S_{\rho n}\,(\theta) + G'\Omega^{-1}G\vartheta\right| \\
\leq\ & \left|\frac{1}{n}\sum_{i=1}^n \rho_1\left(\hat{\lambda}\,(\theta + \vartheta)' g\,(X_{in}, \theta + \vartheta)\right)\left(\frac{\partial g\,(X_{in}, \theta + \vartheta)}{\partial \theta'}\right)'\left(\hat{\lambda}\,(\theta + \vartheta) - \hat{\lambda}\,(\theta)\right) + G'\Omega^{-1}G\vartheta\right| \\
& + \left|\hat{\lambda}\,(\theta)\right|\left|\frac{1}{n}\sum_{i=1}^n \rho_1\left(\hat{\lambda}\,(\theta + \vartheta)' g\,(X_{in}, \theta + \vartheta)\right)\frac{\partial g\,(X_{in}, \theta + \vartheta)}{\partial \theta'}\right| \\
& + \left|\hat{\lambda}\,(\theta)\right|\left|\frac{1}{n}\sum_{i=1}^n \rho_1\left(\hat{\lambda}\,(\theta)' g\,(X_{in}, \theta)\right)\frac{\partial g\,(X_{in}, \theta)}{\partial \theta'}\right| = A_1 + A_2 + A_3,
\end{aligned}
$$

where A_1, A_2, and A_3 are implicitly defined. Note that there exists $C_3 > 0$ such that $A_2 + A_3 \leq$

$C_3\epsilon$ for a.e. $\mathbf{x}_n \in B_{5n}$. Also for a.e. $\mathbf{x}_n \in B_{5n}$, the triangle inequality implies that

$$\begin{aligned} A_1 &\leq \left| \frac{1}{n} \sum_{i=1}^{n} \rho_1 \left(\hat{\lambda} (\theta + \vartheta)' g (X_{in}, \theta + \vartheta) \right) \left(\frac{\partial g (X_{in}, \theta + \vartheta)}{\partial \theta'} \right)' + G' \right| \left| \hat{\lambda} (\theta + \vartheta) - \hat{\lambda} (\theta) \right| \\ &\quad + \left| G' \left(\hat{\lambda} (\theta + \vartheta) - \hat{\lambda} (\theta) \right) + G' \Omega^{-1} G \vartheta \right| \leq |G| \left| \hat{\lambda} (\theta + \vartheta) - \hat{\lambda} (\theta) + \Omega^{-1} G \vartheta \right| + C_4 \epsilon, \end{aligned}$$

for some $C_4 > 0$. On the other hand, for a.e. $\mathbf{x}_n \in B_{5n}$, $\hat{\lambda}(\theta)$ is an interior solution and satisfies the first-order condition, which is expanded as

$$0 = -\hat{g}(\theta) - \Omega \hat{\lambda}(\theta) + \left\{ \Omega + \frac{1}{n} \sum_{i=1}^{n} \rho_2 \left(\tilde{\lambda}' g (X_{in}, \theta) \right) g (X_{in}, \theta) g (X_{in}, \theta)' \right\} \hat{\lambda}(\theta)$$

with a point $\tilde{\lambda}$ on the line joining $\hat{\lambda}(\theta)$ and 0. We can obtain a similar expansion for $\hat{\lambda}(\theta + \vartheta)$. Thus, by an expansion around $\vartheta = 0$, there exist $C_5, C_6, C_7 > 0$ such that

$$\begin{aligned} & \left| \hat{\lambda} (\theta + \vartheta) - \hat{\lambda} (\theta) + \Omega^{-1} G \vartheta \right| \\ \leq\ & \left| -\Omega^{-1} \left\{ \hat{g} (\theta + \vartheta) - \hat{g} (\theta) \right\} + \Omega^{-1} G \vartheta \right| + C_5 \epsilon \leq \left| \Omega^{-1} \right| \left| -\hat{G} \left(\theta + \tilde{\vartheta} \right) + G \right| |\vartheta| + C_5 \epsilon \\ \leq\ & C_6 \left\{ \left| -\hat{G}(\theta_0) + G \right| + \left| \hat{H} \right| \epsilon^{\beta} + \epsilon \right\} \leq C_7 \left(\epsilon^{\beta} + \epsilon \right), \end{aligned}$$

for a.e. $\mathbf{x}_n \in B_{5n}$, where $\tilde{\vartheta}$ is a point on the line joining ϑ and 0. Combining these results, we verify Definition (ii)-(b). □

References

[1] Hansen, L. P. (1982) Large sample properties of generalized method of moments estimators, *Econometrica*, 50, 1029-1054.

[2] Hansen, L. P., Heaton, J. and A. Yaron (1996) Finite-sample properties of some alternative GMM estimators, *Journal of Business & Economic Statistics*, 14, 262-280.

[3] Imbens, G. W., Spady, R. H. and P. Johnson (1998) Information theoretic approaches to inference in moment condition models, *Econometrica*, 66, 333-357.

[4] Inglot, T. and W. C. M. Kallenberg (2003) Moderate deviations of minimum contrast estimators under contamination, Annals of Statistics, 31, 852-879.

[5] Jensen, J. L. and A. T. A. Wood (1998) Large deviation and other results for minimum contrast estimators, *Annals of Institute of Statistical Mathematics*, 50, 673-695.

[6] Kitamura, Y. (2007) Empirical likelihood methods in econometrics: theory and practice, in Blundell, R., Newey, W. K. and T. Persson (eds.), *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, vol. 3, 174-237, Cambridge University Press.

[7] Kitamura, Y. and T. Otsu (2006) Minimax estimation and testing for moment condition models via large deviations, Working paper.

[8] Kitamura, Y. and M. Stutzer (1997) An information-theoretic alternative to generalized method of moments estimation, *Econometrica*, 65, 861-874.

[9] Newey, W. K. and D. L. McFadden (1994) Large sample estimation and hypothesis testing, in Engle R. F. and D. L. McFadden (eds.), *Handbook of econometrics*, vol. IV, 2213-2245, Elsevier, Amsterdam.

[10] Newey, W. K. and R. J. Smith (2004) Higher order properties of GMM and generalized empirical likelihood estimators, *Econometrica*, 72, 219-255.

[11] Otsu, T. (2009) Moderate deviations of generalized method of moments and empirical likelihood estimators, Working paper.

[12] Qin, J. and J. Lawless (1994) Empirical likelihood and general estimating equations, *Annals of Statistics*, 22, 300-325.

[13] Smith, R. J. (1997) Alternative semi-parametric likelihood approaches to generalized method of moments estimation, *Economic Journal*, 107, 503-519.