Journal of Multivariate Analysis 102 (2011) 1203–1216

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Moderate deviations of generalized method of moments and empirical likelihood estimators

Taisuke Otsu

Cowles Foundation & Department of Economics, Yale University, P.O. Box 208281, New Haven, CT 06520-8281, USA

ARTICLE INFO

Article history:
Received 25 November 2009
Available online 20 April 2011

AMS subject classifications:
60F10
62F12

Keywords:
Estimating equation
Empirical likelihood
Moderate deviation

ABSTRACT

This paper studies moderate deviation behaviors of the generalized method of moments and generalized empirical likelihood estimators for generalized estimating equations, where the number of equations can be larger than the number of unknown parameters. We consider two cases for the data generating probability measure: the model assumption and local contaminations or deviations from the model assumption. For both cases, we characterize the first-order terms of the moderate deviation error probabilities of these estimators. Our moderate deviation analysis complements the existing literature of the local asymptotic analysis and misspecification analysis for estimating equations, and is useful to evaluate power and robust properties of statistical tests for estimating equations which typically involve some estimators for nuisance parameters.

1. Introduction

This paper studies moderate deviation behaviors of the generalized method of moments (GMM) and generalized empirical likelihood (GEL) estimators for generalized estimating equations, where the number of equations can be larger than the number of unknown parameters.[1] We consider two cases for the data generating probability measure: the model assumption and local contaminations or deviations from the model assumption. For the model assumption or correct specification case, our moderate deviation analysis extends the conventional local asymptotic analysis for the GMM and GEL estimators focusing on $n^{-1/2}$-neighborhoods (see, [8,23]) toward moderate deviation regions focusing on c_n-neighborhoods with $c_n \to 0$ but $c_n n^{1/2} \to \infty$, where n is the sample size. For the local contamination or local misspecification case, our moderate deviation analysis extends the conventional misspecification analysis for estimating equations focusing on globally misspecified models (see, [29]) to locally misspecified models drifting to the model assumption as $n \to \infty$. For the model assumption and local contamination cases, we characterize the first-order terms of the moderate deviation error probabilities of the GMM and GEL estimators. Our moderate deviation analysis complements the existing literature of the local asymptotic analysis and misspecification analysis, and is useful to evaluate power and robust properties of statistical tests for estimating equations which typically involve some estimators for nuisance parameters.

Since Godambe [7] at least, there are numerous empirical applications and theoretical studies on estimating equations; see, e.g., [11,8] for a review. If the number of estimating equations is identical to the number of unknown parameters (called just-identification), we can apply the conventional method of moment estimator for point estimation, and its large and moderate deviation behaviors have been studied in the literature (e.g., [27,18,20,16,15,1]). However, particularly in econometrics and longitudinal data analysis, it is often the case that the number of estimating equations is larger than the number of unknown parameters (called over-identification). In this case the method of moments is not directly applicable

E-mail address: taisuke.otsu@yale.edu.
URL: http://cowles.econ.yale.edu/faculty/otsu.htm.

[1] See, e.g., [13,21] for a review on the GMM and GEL approaches on generalized estimating equations particularly in econometrics.

0047-259X/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2011.04.002

and several estimation methods have been proposed in the literature, such as the GMM [9] and GEL [28,23] which includes empirical likelihood [25,26], Euclidean likelihood [10], and exponential tilting [22,14] as special cases; see also [12]. These papers mostly focused on the local asymptotic properties of the GMM or GEL estimator under the model assumption, i.e., the local error probability $P\left(\sqrt{n}\left|\hat{\theta}-\theta_0\right|\geq z\right)$ for an estimator $\hat{\theta}$ of θ_0 with $z>0$ and a correctly specified P. On the other hand, Otsu [24] has investigated the large deviation properties of the GMM and GEL estimators, i.e., the large deviation error probability $P_n\left(\sqrt{n}\left|\hat{\theta}-\theta_0\right|\geq\sqrt{n}z\right)$ for a locally contaminated P_n. Otsu [24] showed that under some regularity conditions the GMM and GEL estimators have exponentially small large deviation error probabilities. The focus of this paper is on the moderate deviation error probability $P_n\left(\sqrt{n}\left|\hat{\theta}-\theta_0\right|\geq z_n\right)$ with $z_n\to\infty$ but $z_n=o\left(n^{1/2}\right)$. Compared to the literature on the method of moment estimator for the just-identified case, to our best knowledge, there is no theoretical work on moderate deviation analysis of the GMM and GEL estimators for the over-identified case.

The technical contribution of this paper is to derive the first-order terms of the moderate deviation error probabilities of the GMM and GEL estimators for over-identified estimating equations. The moderate deviation results are derived under two setups for the data generating probability measure: the model assumption and local contaminations. These setups are adopted by Inglot and Kallenberg [15] who derived moderate deviation results for some minimum contrast estimators. Our results can be considered as extensions of Inglot and Kallenberg [15] to over-identified estimating equations estimated by the GMM or GEL. It should be noted that although our results are extensions of the previous results to the over-identified case, theoretical arguments for these extensions are not trivial. The GMM estimator is defined as a minimizer of a quadratic form of the sample estimating equations and the GEL estimator is defined as a minimax solution of the GEL criterion function. Therefore, existing technical tools to analyze moderate deviation errors are not directly applicable to our context.

As the literature suggests (e.g., [18,19,15]), there are several reasons to investigate moderate deviation behaviors of estimators under the model assumption or local contaminations. First, moderate deviation analysis is a fundamental tool to assess the quality of estimators and plays a complementary role to the local asymptotic and large deviation analyses. Second, moderate deviation results are useful to evaluate power and robust properties of statistical tests which involve some estimators for nuisance parameters. In our context, the validity of the over-identified estimating equations is checked by the minimized GMM or GEL objective function, and parameter hypotheses are typically checked by likelihood ratio-type statistics using the GMM or GEL objective function. Both test statistics involve parameter estimators, and our moderate deviation results can be applied to evaluate power or robust properties of these tests when the data are generated from locally contaminated or misspecified measures. Third, moderate deviation analysis can provide some optimality criteria to evaluate statistical estimators or tests. For example, this paper shows asymptotic optimality results in a moderate deviation sense for the two-step GMM and GEL estimators over the GMM estimators with non-optimal weights; see Remarks 3.8 and 3.11.

This paper is organized as follows. Section 2 introduces our basic setup. Section 3 presents main results. Section 4 concludes. We use the following notation. Let $|A|=\text{trace}\left(A'A\right)$ be the Euclidean norm of a scalar, vector, or matrix A, B^c, $\text{int}(B)$, and $\text{cl}(B)$ be the complement, interior, and closure of a set B, respectively, C and c be generic positive constants that should be large and small enough, respectively, and "a.e." means "almost every".

2. Setup

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

$$E[g(X,\theta_0)]=\int g(x,\theta_0)\mathrm{d}P(x)=0, \tag{1}$$

where $g:\mathbb{X}\times\Theta\to\mathbb{R}^{d_g}$ is a vector of measurable functions with $d_g\geq d_\theta$. Except for the functional form of the estimating function g, we do not impose any parametric restriction on the distributional form of P. When $d_g=d_\theta$ (i.e., θ_0 is just-identified by the estimating equations), we can apply the method of moments to estimate θ_0 and there are several existing results on moderate deviation behaviors of the method of moment estimator (e.g., [18,15]). On the other hand, when $d_g>d_\theta$ (i.e., θ_0 is over-identified by the estimating equations), the method of moment estimator does not exist in general and we typically employ the GMM or GEL estimator or their variants to estimate θ_0. Although our results apply to the just-identified case as well, where the GMM and GEL estimators coincide with the method of moment estimator, this paper mainly focuses on the over-identified case. There are numerous empirical examples and theoretical studies of over-identified estimating equations. However, to our best knowledge, there is no paper which studies moderate deviation properties of the GMM or GEL estimator. This paper studies moderate deviation behaviors of these estimators under the model assumption (1) or local contaminations from the model assumption. More specifically, we consider the following data generating measure for 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 independently and identically distributed (i.i.d.) sample from the probability measure P_n having the density $\frac{\mathrm{d}P_n}{\mathrm{d}P}$ with respect to P satisfying

$$\frac{dP_n}{dP}(x) = 1 + a_n A_n(x),$$

where $\{a_n\}_{n\in\mathbb{N}}$ is a sequence of constants satisfying $a_n \to 0$ and $A_n : \mathbb{X} \to \mathbb{R}$ is a measurable function satisfying

$$\sup_{n\in\mathbb{N}} \sup_{x\in\mathbb{X}} |A_n(x)| < \infty, \qquad \int A_n(x) dP(x) = 0, \qquad \int A_n(x)^2 dP(x) = 1. \tag{2}$$

(ii) P is the probability measure under the model assumption and there exists a unique solution $\theta_0 \in \Theta \subseteq \mathbb{R}^{d_\theta}$ for the estimating equations $E[g(X, \theta_0)] = \int g(x, \theta_0) dP(x) = 0$.

Hereafter the mathematical expectations under P and P_n are denoted by $E[\cdot]$ and $E_n[\cdot]$, respectively. Assumption P is an adapted version of Inglot and Kallenberg [15, Assumption (A)] to the estimating equation context. This setup allows two cases for the data generating measure P_n.

(a) Model assumption ($a_n = 0$): the data are generated from $P_n = P$ and the estimating equations $E_n[g(X, \theta_0)] = 0$ are satisfied.
(b) Local contamination ($a_n \neq 0$): the data are generated from $P_n \neq P$ and the estimating equations $E_n[g(X, \theta_0)] = 0$ may or may not be satisfied. However, since $a_n \to 0$, the data generating measure P_n converges to the model assumption measure P as the sample size increases.

Note that except for the convergence of a_n to zero and some boundedness conditions in (2), we do not impose any additional restrictions on the way of deviations from the model assumption measure P. In this sense, our treatment on the local contamination is nonparametric. Since the generalized estimating equations are commonly applied to the case where the researcher does not have enough prior knowledge on the parametric distributional form of data, this nonparametric treatment on the local contaminations is suitable for our setup.

This paper considers three popular estimators for the generalized estimating equations: (i) the GMM estimator with some weight matrix, (ii) the optimally weighted two-step GMM estimator, and (iii) the GEL estimator. To deal with the over-identified estimating equations, where the method of moment estimator (a solution of $\frac{1}{n}\sum_{i=1}^n g(X_{in}, \theta) = 0$ with respect to θ) does not exist in general, the GMM estimator with the $d_g \times d_g$ weight matrix $\hat{W}$ minimizes the quadratic form of the sample estimating equations $\frac{1}{n}\sum_{i=1}^n g(X_{in}, \theta)$, i.e.,

$$\hat{\theta}_1 = \arg\min_{\theta\in\Theta} \left(\frac{1}{n}\sum_{i=1}^n g(X_{in}, \theta)\right)' \hat{W} \left(\frac{1}{n}\sum_{i=1}^n g(X_{in}, \theta)\right). \tag{3}$$

It is known that under the model assumption, $P_n = P$, mild regularity conditions guarantee that the GMM estimator $\hat{\theta}_1$ is consistent for θ_0 and asymptotically normal (see, e.g., [8]),

$$\sqrt{n}\left(\hat{\theta}_1 - \theta_0\right) \xrightarrow{d} N\left(0, V_W\right),$$

where $V_W = \left(G'WG\right)^{-1} G'W\Omega WG\left(G'WG\right)^{-1}$,

$$G = E\left[\frac{\partial g(X, \theta_0)}{\partial \theta'}\right], \qquad \Omega = E\left[g(X, \theta_0) g(X, \theta_0)'\right],$$

and W is the (probability) limit of $\hat{W}$. The asymptotic variance V_W depends on the limiting weight matrix W and is minimized (in the positive semi-definite sense) when $W = \Omega^{-1}$. Although the optimal weight Ω^{-1} is unknown, we can estimate it by using $\hat{\theta}_1$ as a preliminary estimator, i.e.,

$$\hat{\Omega}^{-1} = \left(\frac{1}{n}\sum_{i=1}^n g\left(X_{in}, \hat{\theta}_1\right) g\left(X_{in}, \hat{\theta}_1\right)'\right)^{-1}. \tag{4}$$

By using the estimated optimal weight matrix $\hat{\Omega}^{-1}$, the optimally weighted two-step GMM estimator is defined as

$$\hat{\theta}_2 = \arg\min_{\theta\in\Theta} \left(\frac{1}{n}\sum_{i=1}^n g(X_{in}, \theta)\right)' \hat{\Omega}^{-1} \left(\frac{1}{n}\sum_{i=1}^n g(X_{in}, \theta)\right). \tag{5}$$

Under the model assumption, $P_n = P$, mild regularity conditions guarantee the weak consistency of $\hat{\Omega}^{-1}$ to Ω^{-1} and the asymptotic normality of $\hat{\theta}_2$,

$$\sqrt{n}\left(\hat{\theta}_2 - \theta_0\right) \xrightarrow{d} N\left(0, \left(G'\Omega^{-1}G\right)^{-1}\right).$$

It is known that the two-step GMM estimator $\hat{\theta}_2$ attains the semiparametric efficiency (or information) bound under the model assumption [4,2].

As an alternative class of estimators to the two-step GMM, we consider the GEL estimator:

$$\hat{\theta}_3 = \arg\min_{\theta\in\Theta}\max_{\lambda\in\Lambda}\sum_{i=1}^{n}\rho\left(\lambda' g(X_{in},\theta)\right). \tag{6}$$

In contrast to the two-step GMM estimator $\hat{\theta}_2$, the GEL estimator does not require preliminary estimation for Ω^{-1}. Under suitable conditions this minimax problem can be interpreted as the dual problem of the minimum empirical discrepancy problem (see, [23, Theorem 2.2]),

$$\hat{\theta}_3 = \arg\min_{\theta\in\Theta}\min_{\{p_i\}_{i=1}^n}\sum_{i=1}^{n}h\left(p_i\right), \tag{7}$$

subject to

$$\sum_{i=1}^{n}p_i = 1, \qquad \sum_{i=1}^{n}p_i g(X_{in},\theta) = 0,$$

for some h. Thus, the GEL estimator $\hat{\theta}_3$ can be interpreted as a constrained maximum likelihood estimator by the nonparametric likelihood function $\sum_{i=1}^{n}h\left(p_i\right)$, which puts probability weights $\{p_i\}_{i=1}^n$ on the observed points of $\{X_{in}\}_{i=1}^n$ subject to the estimating equation constraints $\sum_{i=1}^{n}p_i g(X_{in},\theta) = 0$. Although the formulation in (7) is intuitive to understand the rationale of the GEL estimator, this formulation is not practical because of the n-variable optimization problem for $\{p_i\}_{i=1}^n$. We employ the dual formula in (6) to define the GEL estimator, which is used in practice to compute the GEL estimator.

To implement the GEL estimation, we need to specify the criterion function ρ (or h). The GEL estimator contains several existing estimators for generalized estimating equations as special cases:

- Empirical likelihood: $\rho(v) = \log(1-v)$ and $h(p) = -\log p$.
- Euclidean likelihood: $\rho(v) = -(1+v)^2/2$ and $h(p) = p^2$.
- Exponential tilting: $\rho(v) = -\exp(v)$ and $h(p) = p\log p$.
- Cressie and Read [5] divergence: $\rho(v) = -\frac{(1+\gamma v)^{(\gamma+1)/\gamma}}{1+\gamma}$ and $h(v) = \frac{p^{\gamma+1}-1}{\gamma(\gamma+1)}$ for $\gamma\in\mathbb{R}$.

Newey and Smith [23] showed that for a general class of the criterion functions ρ or h, the GEL estimator $\hat{\theta}_3$ has the same asymptotic distribution as the optimally weighted two-step GMM estimator $\hat{\theta}_2$ under the model assumption $P_n = P$, i.e., $\sqrt{n}\left(\hat{\theta}_3-\theta_0\right)\xrightarrow{d}N\left(0,\left(G'\Omega^{-1}G\right)^{-1}\right)$. Furthermore, Newey and Smith [23] investigated higher-order properties of the GEL estimator under the model assumption and found that the GEL estimator has better higher-order bias properties than the two-step GMM estimator.

The above asymptotic normality results approximate the local error probabilities $P\left(\sqrt{n}\left|\hat{\theta}_j-\theta_0\right|\geq z\right)$ for $z>0$ and $j=1,2,3$ based on the central limit theorems under the model assumption. On the other hand, Otsu [24] studied the large derivation error probabilities $P_n\left(\sqrt{n}\left|\hat{\theta}_j-\theta_0\right|\geq\sqrt{n}z\right)$ under P_n, which allows local contaminations, and showed that under some regularity conditions the GMM and GEL estimators have exponentially small large deviation error probabilities, i.e., $P_n\left(\sqrt{n}\left|\hat{\theta}_j-\theta_0\right|\geq\sqrt{n}z\right)\leq C\mathrm{e}^{-cn}$ for some $C,c>0$. The purpose of this paper is to bridge these two asymptotic results by characterizing the first-order terms of the moderate deviation error probabilities $P_n\left(\sqrt{n}\left|\hat{\theta}_j-\theta_0\right|\geq z_n\right)$ for $z_n\to\infty$ but $z_n = o\left(n^{1/2}\right)$.

We close this section by pointing out some differences with the existing moderate deviation results on the method of moments or minimum contrast estimators. The literature mostly focuses on the just-identified case and considers the method of moment estimator (i.e., a solution of $\frac{1}{n}\sum_{i=1}^{n}g(X_{in},\theta)=0$ with $d_g=d_\theta$) or the minimum contrast estimator (i.e., a minimizer of some objective function $\sum_{i=1}^{n}\gamma(X_{in},\theta)$ with respect to θ or a solution of $\sum_{i=1}^{n}\partial\gamma(X_{in},\theta)/\partial\theta=0$). It should be mentioned that our moderate deviation analysis is a non-trivial extension of the previous results at least in three senses. First, the GMM estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ are defined as minimizers of quadratic forms of the sample estimating functions $\frac{1}{n}\sum_{i=1}^{n}g(X_{in},\theta)$, instead of a single summation of some contrast function. Second, the two-step GMM estimator $\hat{\theta}_2$ contains the preliminary GMM estimator $\hat{\theta}_1$. Thus, we need to incorporate estimation errors of $\hat{\theta}_1$ to analyze the moderate deviation properties of $\hat{\theta}_2$. Third, the GEL estimator is defined as a minimax solution rather than a simple minimization solution. This minimax structure also complicates our moderate deviation analysis.

3. Main results

In this section, we present the moderate deviation properties of the GMM and GEL estimators. Hereafter denote $G(x,\theta) = \partial g(x,\theta)/\partial\theta'$. We first consider $\hat{\theta}_1$ in (3), the GMM estimator with the weight matrix $\hat{W}$. We impose the following assumptions.

Assumption G1.

(i) Θ is compact and $\theta_0 \in \text{int}(\Theta)$. There exist a measurable function $L : \mathbb{X} \to [0,\infty)$ and constants $\alpha, T_1 \in (0,\infty)$ 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 a constant $T_2 \in (0,\infty)$ satisfying $E[\exp(T_2|g(X,\theta)|)] < \infty$.

(ii) There exist a measurable function $H : \mathbb{X} \to [0,\infty)$, constants $\beta, T_3 \in (0,\infty)$, and a neighborhood $\mathcal{N}$ around θ_0 such that $|G(x,\theta) - G(x,\theta_0)| \leq H(x)|\theta - \theta_0|^\beta$ for all $\theta \in \mathcal{N}$ and a.e. x, and $E[\exp(T_3 H(X))] < \infty$. There exists a constant $T_4 \in (0,\infty)$ satisfying $E[\exp(T_4|G(X,\theta_0)|)] < \infty$. G has the full column rank. Ω is positive definite.

Assumption W. There exists a sequence of $d_g \times d_g$ matrices $\{W_n\}_{n\in\mathbb{N}}$ such that

$$P_n\left(\left|\hat{W} - W_n\right| \geq n^{-1/2} z_n\right) \leq \exp\left\{-\frac{z_n^2}{2} + O\left(\frac{z_n^3}{\sqrt{n}}\right) + O(\log z_n)\right\},$$

for any sequence $\{z_n\}_{n\in\mathbb{N}}$ satisfying $z_n \to \infty$ and $n^{-1/2} z_n \to 0$, and $W_n \to W$ with a positive definite matrix W.

Assumption G1 restricts the shape of the estimating function g. Assumption G1(i) is on the global shape of g over the parameter space Θ. Compared to the setups for the method of moment estimator (e.g., [16,15]), it is not easy to avoid the compactness assumption on Θ without imposing additional restrictions on the shape of g, such as concavity of the GMM objective function in θ. The Lipschitz-type condition on g is common in the literature and is satisfied with $\alpha = 1$ if g is differentiable on Θ for a.e. x and the derivative has an exponential moment. Boundedness conditions of exponential moments are required to control large and moderate deviation probabilities for the sum of the estimating functions. Assumption G1(ii) controls the local shape of the estimating functions g in a neighborhood of θ_0. The Lipschitz-type assumption on the derivative $G(x,\theta)$ is satisfied with $\beta = 1$ if g is second-order differentiable in a neighborhood of θ_0 for a.e. x and the derivative has an exponential moment.

The boundedness conditions for several exponential moments in Assumption G1 are restrictive and unnecessary to derive local asymptotic properties such as the asymptotic normality of the GMM estimator. However, to investigate the tail behaviors of the estimators, it is hard to proceed without these bounded exponential moments. For example, the conventional Cramér-type large and moderate deviation theorems for sums of random samples typically require existence of moment generating functions (see, e.g., [6]). Also note that even for the just-identified case, we need similar boundedness conditions for the moment functions and their derivatives to study large and moderate deviation properties of the method of moment estimator (see, [18,16,15]).

Assumption W is a high-level assumption on the weight matrix $\hat{W}$. This assumption should be checked for each specific choice of $\hat{W}$. If $\hat{W}$ is a constant positive definite matrix (i.e., $\hat{W} = W_n = W$), this assumption is trivially satisfied. If $\hat{W}$ is defined as a sample mean, the conventional moderate deviation theorems for i.i.d. sums, such as Book [3], Yurinskii [30], Jurečková et al. [17], and Dembo and Zeitouni [6], can be applied to verify this assumption.

Under these assumptions, we can characterize the moderate deviation behavior of the GMM estimator $\hat{\theta}_1$ with the weight matrix $\hat{W}$ as follows.

Theorem 3.1. *Suppose that Assumptions P, W and G1 hold.*

(i) *For all n large enough and $\delta \in (0,\infty)$ small enough, there exists a unique $\theta_{1n} \in \{\theta \in \Theta : |\theta - \theta_0| \leq \delta\}$ such that*

$$E_n[G(X,\theta_{1n})]' W_n E_n[g(X,\theta_{1n})] = 0,$$
$$\theta_{1n} = \theta_0 - a_n\left(G'WG\right)^{-1} G'WE[A_n(X)g(X,\theta_0)] + o(a_n). \tag{8}$$

(ii) *For any sequence $\{z_n\}_{n\in\mathbb{N}}$ satisfying $z_n \to \infty$ and $n^{-1/2} z_n \to 0$,*

$$P_n\left(\sqrt{n}\left|\left(G'W\Omega WG\right)^{-1/2} G'WG\left(\hat{\theta}_1 - \theta_{1n}\right)\right| \geq z_n\right) = \exp\left\{-\frac{z_n^2}{2} + O\left(a_n z_n^2\right) + O\left(\frac{z_n^3}{\sqrt{n}}\right) + O(\log z_n)\right\}.$$

Remark 3.1. Part (i) of this theorem shows the existence of a unique natural parameter θ_{1n}, which solves the population analogue of the first-order condition of the GMM estimator $\hat{\theta}_1$. Under the model assumption $P_n = P$, θ_{1n} becomes θ_0, the "true" parameter under correct specification. Under the local contamination $P_n \neq P$, it is more natural to employ θ_{1n} as a parameter to be estimated by $\hat{\theta}_1$. Using the terminology of misspecification analysis, θ_{1n} may be interpreted as a "pseudo-true value" [29] in our local contamination context. Also, θ_{1n} can be interpreted as a projection of the data generating measure P_n to the parameter space Θ using the quadratic distance based on the population analogue of the GMM objective function in (3), i.e., $\theta_{1n} = \arg\min_{\theta\in\Theta} E_n[g(X,\theta)]' W_n E_n[g(X,\theta)]$.

Remark 3.2. Part (ii) of this theorem says that even if the critical value z_n diverges, the tail probability of $\sqrt{n}\left(\hat{\theta}_1 - \theta_{1n}\right)$ can be still approximated by the normal distribution $N(0, V_W)$. The conventional local asymptotic theory based on a central limit theorem says that the GMM estimator $\hat{\theta}_1$ is asymptotically normal under the model assumption $P_n = P$, i.e., $P\left(\sqrt{n}\left|\left(G'W\Omega WG\right)^{-1/2} G'WG\left(\hat{\theta}_1 - \theta_{1n}\right)\right| \geq z\right) \to 1 - 2\Phi(z)$ with the standard normal distribution function Φ. On the other hand, under similar assumptions, Otsu [24] showed that the large deviation error probability of the GMM estimator is exponentially small, i.e., for every $z > 0$, there exist $C, c > 0$ such that $P_n\left(\sqrt{n}\left|\hat{\theta}_1 - \theta_{1n}\right| \geq \sqrt{n}z\right) \leq Ce^{-cn}$ for all n large enough. The moderate deviation result in Theorem 3.1(ii) bridges these two asymptotic results by focusing on the tail probabilities with the critical value $z_n \to \infty$ but $z_n = o\left(n^{1/2}\right)$.

Remark 3.3. By taking the limit $n \to \infty$ for the result in Theorem 3.1(ii), the moderate deviation rate function is obtained as

$$\lim_{n\to\infty} z_n^{-2} \log P_n\left(\sqrt{n}\left|\left(G'W\Omega WG\right)^{-1/2} G'WG\left(\hat{\theta}_1 - \theta_{1n}\right)\right| \geq z_n\right) = -\frac{1}{2}.$$

Remark 3.4. The statements in Theorem 3.1 hold even if we replace G, W, and Ω with $E_n[G(X, \theta_{1n})]$, W_n, and $E_n\left[g(X, \theta_{1n})\, g(X, \theta_{1n})'\right]$, respectively.

Remark 3.5. Although it is natural to consider the concentration of $\hat{\theta}_1$ around the natural parameter θ_{1n}, we can also derive an analogous moderate deviation result for the contrast $\hat{\theta}_1 - \theta_0$, i.e., if

$$\Delta_{1n} = n^{1/2} a_n z_n^{-1} \left(G'W\Omega WG\right)^{-1/2} G'WE_n[A_n(X)G(X, \theta_0)] \to \Delta_1, \tag{9}$$

with $|\Delta_1| \in [0, 1)$,[2] then

$$P_n\left(\sqrt{n}\left|\left(G'W\Omega WG\right)^{-1/2} G'WG\left(\hat{\theta}_1 - \theta_0\right)\right| \geq z_n\right)$$
$$= \exp\left\{-\frac{(1 - |\Delta_1|)^2 z_n^2}{2} + O\left(|\Delta_{1n} - \Delta_1| z_n^2\right) + O\left(a_n z_n^2\right) + O\left(\frac{z_n^3}{\sqrt{n}}\right) + O(\log z_n)\right\}.$$

We now analyze the two-step GMM estimator $\hat{\theta}_2$. The following assumption is imposed.

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

Assumption G2 is an additional boundedness condition on the estimating function g, which is used to control the moderate deviation behavior of the optimal weight matrix estimator $\hat{\Omega}^{-1}$. The moderate deviation properties of the two-step GMM estimator is obtained as follows.

Theorem 3.2. *Suppose that Assumptions P, W, G1 and G2 hold.*

(i) *For all n large enough and $\delta \in (0, \infty)$ small enough, there exists a unique $\theta_{2n} \in \{\theta \in \Theta : |\theta - \theta_0| \leq \delta\}$ such that*

$$E_n[G(X, \theta_{2n})]' E_n\left[g(X, \theta_{1n})\, g(X, \theta_{1n})'\right]^{-1} E_n[g(X, \theta_{2n})] = 0,$$
$$\theta_{2n} = \theta_0 - a_n\left(G'\Omega^{-1}G\right)^{-1} G'\Omega^{-1} E[A_n(X) g(X, \theta_0)] + o(a_n).$$

(ii) *For any sequence $\{z_n\}_{n\in\mathbb{N}}$ satisfying $z_n \to \infty$ and $n^{-1/2} z_n \to 0$,*

$$P_n\left(\sqrt{n}\left|\left(G'\Omega^{-1}G\right)^{1/2}\left(\hat{\theta}_2 - \theta_{2n}\right)\right| \geq z_n\right) = \exp\left\{-\frac{z_n^2}{2} + O\left(a_n z_n^2\right) + O\left(\frac{z_n^3}{\sqrt{n}}\right) + O(\log z_n)\right\}.$$

Remark 3.6. Similar remarks to Theorem 3.1 apply here. θ_{2n} is the natural parameter for the two-step GMM estimator $\hat{\theta}_2$, which solves the population analogue of the first-order condition of $\hat{\theta}_2$. The statements in Theorem 3.2 hold even if we replace G and Ω with $E_n[G(X, \theta_{2n})]$ and $E_n\left[g(X, \theta_{1n})\, g(X, \theta_{1n})'\right]$, respectively. The moderate deviation rate function is obtained as

[2] If $n^{1/2} a_n z_n^{-1} \to 0$, then the condition in (9) is satisfied with $\Delta_1 = 0$. If $n^{1/2} a_n z_n^{-1} \to c$, then we need to choose $A_n(X)$ to satisfy $\left|c\left(G'W\Omega WG\right)^{-1/2} G'WE_n[A_n(X)G(X, \theta_0)]\right| \to |\Delta_1| < 1$, which is guaranteed by assuming, e.g., $\sup_{n\in\mathbb{N}} \sup_{x\in\mathbb{X}} |A_n(x)| \leq \frac{\left|G'W\Omega WG\right|^{1/2}}{|c|\left|G'WE[|G(X,\theta_0)|]\right|} - \delta$ for some $\delta > 0$. Similar comments apply to the conditions in (10) and (13) (by setting $W = \Omega^{-1}$).

$$\lim_{n\to\infty} z_n^{-2} \log P_n\left(\sqrt{n}\left|\left(G'\Omega^{-1}G\right)^{1/2}\left(\hat{\theta}_2-\theta_{2n}\right)\right|\geq z_n\right)=-\frac{1}{2}.$$

Also, we can derive an analogous moderate deviation result for the estimation error $\hat{\theta}_2-\theta_0$ around θ_0, i.e.,

$$P_n\left(\sqrt{n}\left|\left(G'\Omega^{-1}G\right)^{1/2}\left(\hat{\theta}_2-\theta_0\right)\right|\geq z_n\right)$$
$$=\exp\left\{-\frac{(1-|\Delta_2|)^2 z_n^2}{2}+O\left(|\Delta_{2n}-\Delta_2|\, z_n^2\right)+O\left(a_n z_n^2\right)+O\left(\frac{z_n^3}{\sqrt{n}}\right)+O\left(\log z_n\right)\right\},$$

if

$$\Delta_{2n}=n^{1/2}a_n z_n^{-1}\left(G'\Omega^{-1}G\right)^{-1/2}G'\Omega^{-1}E_n\left[A_n(X)G(X,\theta_0)\right]\to\Delta_2, \tag{10}$$

with $|\Delta_2|\in[0,1)$.

Remark 3.7. A crucial difference with Theorem 3.1 is that now the moderate deviation probability of $\sqrt{n}\left(\hat{\theta}_2-\theta_{2n}\right)$ is approximated by the normal distribution $N\left(0,\left(G'\Omega^{-1}G\right)^{-1}\right)$ whose variance is always smaller or equal (in the positive semi-definite sense) to that of the GMM estimator $\hat{\theta}_1$ with some weight $\hat{W}$. In other words, the distribution of $\sqrt{n}\left(\hat{\theta}_2-\theta_{2n}\right)$ is more concentrated around zero than that of $\sqrt{n}\left(\hat{\theta}_1-\theta_{1n}\right)$. Let min eig$(A)$ be the minimum eigenvalue of a matrix A. From Theorems 3.1(ii) and 3.2 (ii), a similar argument to Inglot and Kallenberg [15, Corollary 3.3] implies

$$\frac{\log P_n\left(\sqrt{n}\left|\hat{\theta}_2-\theta_{2n}\right|\geq z_n\right)}{\log P_n\left(\sqrt{n}\left|\hat{\theta}_1-\theta_{1n}\right|\geq z_n\right)}\to\left(\frac{\min\text{eig}\left(G'\Omega^{-1}G\right)}{\min\text{eig}\left(V_W\right)}\right)^2\leq 1, \tag{11}$$

for any positive definite W.

Remark 3.8. If we assume $P_n=P$, then the natural parameter becomes $\theta_{1n}=\theta_{2n}=\theta_0$ and the result obtained in (11) becomes $\lim_{n\to\infty}\frac{\log P\left(\sqrt{n}\left|\hat{\theta}_2-\theta_0\right|\geq z_n\right)}{\log P\left(\sqrt{n}\left|\hat{\theta}_1-\theta_0\right|\geq z_n\right)}\leq 1$. This result can be seen as an extension of the asymptotic optimality of the two-step GMM estimator in the local asymptotics to the moderate deviation zone.

Remark 3.9. An intuition for the results in Remarks 3.7 and 3.8 may be explained as follows. Similar to the local asymptotic analysis, dominant components to analyze the moderate deviation properties of $\sqrt{n}\left(\hat{\theta}_1-\theta_{1n}\right)$ and $\sqrt{n}\left(\hat{\theta}_2-\theta_{2n}\right)$ are still characterized by their score functions $\left(G'WG\right)^{-1}G'W\frac{1}{\sqrt{n}}\sum_{i=1}^n g\left(X_i,\theta_{1n}\right)$ and $\left(G'\Omega^{-1}G\right)^{-1}G'\frac{1}{\sqrt{n}}\sum_{i=1}^n g\left(X_i,\theta_{2n}\right)$, respectively. On the other hand, moderate deviation theorems for sums of independent random variables (e.g., [6]) guarantee that the moderate deviation properties for the sums (after normalization) can be characterized by the tail of the standard normal distribution. Thus, the asymptotic efficiency of $\hat{\theta}_2$ compared to $\hat{\theta}_1$ in the local asymptotics is maintained in the moderate deviation zone.

To derive the moderate deviation properties of the GEL estimator, we impose the following assumptions.

Assumption G3.

(i) Θ is compact and $\theta_0\in\text{int}(\Theta)$. $\rho(\cdot)$ is strictly concave and $\rho_1(0)=\rho_2(0)=-1$. Λ is compact and $\mathbf{0}\in\text{int}(\Lambda)$. For each $\theta\in\Theta$, the maximizer $\lambda_*(\theta)=\arg\max_{\lambda\in\Lambda}E\left[\rho\left(\lambda' g(X,\theta)\right)\right]$ satisfies $\lambda_*(\theta)\in\text{int}(\Lambda)$. $g(x,\theta)$ is differentiable on Θ for a.e. x. There exists a constant $T_8\in(0,\infty)$ satisfying $E\left[\exp\left(T_8|g(X,\theta_0)|\right)\right]<\infty$. For each $\theta\in\Theta$, there exist a constant $T_9\in(0,\infty)$ and neighborhoods $\mathcal{N}_\theta$ and $\mathcal{N}_{\lambda_*(\theta)}$ around θ and $\lambda_*(\theta)$, respectively, satisfying $E\left[\exp\left(T_9\sup_{\vartheta\in\mathcal{N}_\theta}\sup_{\lambda\in\mathcal{N}_{\lambda_*(\theta)}}\left|\rho_1\left(\lambda' g(X,\vartheta)\right)G(X,\vartheta)\right|\right)\right]<\infty$.

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

Assumption G3(i) is a replacement of Assumption G1(i). All examples of the GEL criterion function ρ listed in Section 2 are strictly concave and satisfy $\rho_1(0)=\rho_2(0)=-1$. Although technical arguments become more complicated, the compactness assumption on Λ may be avoided by adding a similar assumption to Inglot and Kallenberg [15, Assumption (R2')] which controls the global behaviors of the contrast function outside some compact set for λ. The last condition in Assumption G3(i), which corresponds to the bounded exponential moment 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 specific choices of ρ and g. Assumption G3(ii) contains additional conditions to control the local curvatures of the GEL objective function with respect to λ in a neighborhood of $\mathbf{0}$.

The boundedness conditions for exponential moments in Assumption G3 are typically more stringent and difficult to verify than the ones for the GMM estimator (Assumption G1) or the ones for the method of moment estimator [15]. For example, in the case of the empirical likelihood estimator (i.e., $\rho(v) = \log(1-v)$), the last condition in Assumption G3(i) becomes $E\left[\exp\left(T_9 \sup_{\vartheta\in\mathcal{N}_\theta} \sup_{\lambda\in\mathcal{N}_{\lambda_*(\theta)}} \left|\frac{1}{1-\lambda' g(X,\vartheta)} G(X,\vartheta)\right|\right)\right] < \infty$, and the condition in Assumption G3(ii) becomes $E\left[\exp\left(T_{10} \sup_{\theta\in\mathcal{N}_\rho} \sup_{\lambda\in\mathcal{N}'_\rho} \left|\frac{1}{(1-\lambda' g(X,\theta))^2} g(X,\theta) g(X,\theta)'\right|\right)\right] < \infty$. Such restrictions and complications are attributable to the fact that the GEL estimator is defined as a minimax solution using auxiliary parameters λ.

Under these assumptions, the moderate deviation properties of the GEL estimator is obtained as follows.

Theorem 3.3. *Suppose that Assumptions P and G1 (ii), and G3 hold.*

(i) *For all n large enough and $\delta \in (0,\infty)$ small enough, there exists a unique $\theta_{3n} \in \{\theta \in \Theta : |\theta - \theta_0| \le \delta\}$ such that*

$$E_n[G(X,\theta_{3n})]' E_n\left[g(X,\theta_{3n}) g(X,\theta_{3n})'\right]^{-1} E_n[g(X,\theta_{3n})] = 0,$$
$$\theta_{3n} = \theta_0 - a_n \left(G'\Omega^{-1}G\right)^{-1} G'\Omega^{-1} E[A_n(X) g(X,\theta_0)] + o(a_n). \tag{12}$$

(ii) *For any sequence $\{z_n\}_{n\in\mathbb{N}}$ satisfying $z_n \to \infty$ and $n^{-1/2} z_n \to 0$,*

$$P_n\left(\sqrt{n}\left|\left(G'\Omega^{-1}G\right)^{1/2}\left(\hat\theta_3 - \theta_{3n}\right)\right| \ge z_n\right) = \exp\left\{-\frac{z_n^2}{2} + O\left(a_n z_n^2\right) + O\left(\frac{z_n^3}{\sqrt{n}}\right) + O(\log z_n)\right\}.$$

Remark 3.10. Similar remarks to Theorems 3.1 and 3.2 apply here. θ_{3n} is the natural parameter for the GEL estimator $\hat\theta_3$. The statements in Theorem 3.3 hold even if we replace G and Ω with $E_n[G(X,\theta_{3n})]$ and $E_n\left[g(X,\theta_{3n}) g(X,\theta_{3n})'\right]$, respectively. The moderate deviation rate function is obtained as

$$\lim_{n\to\infty} z_n^{-2} \log P_n\left(\sqrt{n}\left|\left(G'\Omega^{-1}G\right)^{1/2}\left(\hat\theta_3 - \theta_{3n}\right)\right| \ge z_n\right) = -\frac{1}{2}.$$

Also, we can derive the moderate deviation result for the estimation error $\hat\theta_3 - \theta_0$ around θ_0, i.e.,

$$P_n\left(\sqrt{n}\left|\left(G'\Omega^{-1}G\right)^{1/2}\left(\hat\theta_3 - \theta_0\right)\right| \ge z_n\right)$$
$$= \exp\left\{-\frac{(1-|\Delta_3|)^2 z_n^2}{2} + O\left(|\Delta_{3n} - \Delta_3| z_n^2\right) + O\left(a_n z_n^2\right) + O\left(\frac{z_n^3}{\sqrt{n}}\right) + O(\log z_n)\right\},$$

if

$$\Delta_{3n} = n^{1/2} a_n z_n^{-1} \left(G'\Omega^{-1}G\right)^{-1/2} G'\Omega^{-1} E_n[A_n(X) G(X,\theta_0)] \to \Delta_3, \tag{13}$$

with $|\Delta_3| \in [0,1)$.

Remark 3.11. Similar to the two-step GMM estimator, the moderate deviation error probability of $\sqrt{n}\left(\hat\theta_3 - \theta_{3n}\right)$ is approximated by the normal distribution $N\left(0, \left(G'\Omega^{-1}G\right)^{-1}\right)$. From Theorem 3.3(i), we can see that the GEL estimator also enjoys the asymptotic optimality in the moderate deviation sense, i.e., $\frac{\log P_n\left(\sqrt{n}\left|\hat\theta_3 - \theta_{3n}\right| \ge z_n\right)}{\log P_n\left(\sqrt{n}\left|\hat\theta_2 - \theta_{2n}\right| \ge z_n\right)} \to 1$. This result can be seen as an extension of the asymptotic equivalence between the two-step GMM and GEL estimators under the local asymptotics to the moderate deviation region.

Remark 3.12. This paper mainly focuses on the case of over-identification, i.e., $d_g > d_\theta$. If the estimating equations are just-identified, i.e., $d_g = d_\theta$, then the above three estimators coincide with the method of moment estimator and the above theorems become variants of the moderate deviation results in [15].

4. Conclusion

This paper studies moderate deviation behaviors of the generalized method of moments (GMM) and generalized empirical likelihood (GEL) estimators for generalized estimating equations. As data generating probability measures, we consider the model assumption and locally contaminated measures. For both cases, we characterize the first-order terms of the moderate deviation error probabilities of the GMM and GEL estimators. There are several directions of the future

research. First, to compare the two-step GMM and GEL estimators which have the same moderate deviation rate function, it is important to study higher-order terms of those moderate deviation probabilities. For example, we can expect that the rate function of the GEL estimator depends on the criterion function ρ, and this rate function allows us to compare the competing members of the GEL estimators, such as the empirical likelihood and exponential tilting. Second, the GMM and GEL estimators are commonly applied to time series or panel data. Therefore, it is useful to extend the obtained results to more general data environments. Finally, it is interesting to extend the present results to more general models, such as non-compact parameter spaces and non-differentiable estimating functions (e.g., quantile restrictions).

Acknowledgment

Financial support from the National Science Foundation (SES-0720961) is gratefully acknowledged.

Appendix. Mathematical appendix

Hereafter let $\mathbf{x}_n = (x_{1n}, \ldots, x_{nn})$, $\hat{g}(\theta) = \frac{1}{n}\sum_{i=1}^n g(X_{in}, \theta)$, and $\hat{G}(\theta) = \frac{1}{n}\sum_{i=1}^n G(X_{in}, \theta)$.

A.1. Proof of Theorem 3.1

Proof of (i). First, we show the continuity of

$$Q_n(\theta) = E_n[g(X,\theta)]' W_n E_n[g(X,\theta)] - E_n[g(X,\theta_0)]' W_n E_n[g(X,\theta_0)],$$

in $\theta \in \mathcal{N}$, where the neighborhood $\mathcal{N}$ is defined in Assumption G1(ii). By Assumption P, $Q_n(\theta)$ is well defined on Θ. Pick any $\vartheta, \theta \in \mathcal{N}$. By an expansion of $E_n[g(X,\vartheta)]$ around $\vartheta = \theta$,

$$|Q_n(\vartheta) - Q_n(\theta)| \le 2\left|E_n\left[G\left(X,\bar{\vartheta}\right)\right]' W_n E_n[g(X,\theta)]\right| |\vartheta - \theta| + \left|E_n\left[G\left(X,\bar{\vartheta}\right)\right]' W_n E_n\left[G\left(X,\bar{\vartheta}\right)\right]\right| |\vartheta - \theta|^2, \tag{14}$$

where $\bar{\vartheta}$ is a point on the line joining ϑ and θ. From Assumptions P and G1,

$$\begin{aligned} &|E_n[g(X,\theta)]| \le |E[g(X,\theta)]| + a_n |E[A_n(X) g(X,\theta)]| < \infty, \\ &\left|E_n\left[G\left(X,\bar{\vartheta}\right)\right]\right| \le \left|E\left[G\left(X,\bar{\vartheta}\right)\right]\right| + a_n \left|E\left[A_n(X) G\left(X,\bar{\vartheta}\right)\right]\right| < \infty, \end{aligned} \tag{15}$$

for each $n \in \mathbb{N}$, where the last inequality follows from $\left|E\left[G\left(X,\bar{\vartheta}\right)\right]\right| \le E[H(X)]\left|\bar{\vartheta} - \theta_0\right|^\beta + E[|G(X,\theta_0)|] < \infty$ using Assumption G1(ii). From (14) and (15), $Q_n(\theta)$ is continuous on $\mathcal{N}$ for each $n \in \mathbb{N}$.

Second, we show the differentiability of $Q_n(\theta)$ in $\theta \in \mathcal{N}$. Pick any $\theta \in \mathcal{N}$ and $\varepsilon \neq 0$ small enough so that $\theta + \varepsilon e_j \in \mathcal{N}$, where $e_j = (0, \ldots, 0, 1, 0, \ldots, 0)$ is the jth unit vector. Let $G_j(X,\theta)$ be the jth column of $G(X,\theta)$. By a Taylor expansion of $E_n\left[g\left(X, \theta + \varepsilon e_j\right)\right]$ around $\varepsilon = 0$ combined with Assumptions P and G1 and (15),

$$\left|\varepsilon^{-1}\left\{Q_n\left(\theta + \varepsilon e_j\right) - Q_n(\theta)\right\} - 2E_n\left[G_j(X,\theta)\right]' W_n E_n[g(X,\theta)]\right| \le C\left(|\bar{\varepsilon}|^\beta + |\varepsilon|^\alpha\right),$$

where $\bar{\varepsilon}$ is a point between ε and 0. Thus, by taking $\varepsilon \to 0$ (so, $\bar{\varepsilon} \to 0$ as well), we obtain the differentiability of $Q_n(\theta)$ in $\theta \in \mathcal{N}$ for each $n \in \mathbb{N}$ with the derivative $D_n(\theta) = 2E_n[G(X,\theta)]' W_n E_n[g(X,\theta)]$.

Third, we show the existence of θ_{1n} defined in (8). Let $Q(\theta) = E[g(X,\theta)]' W_n E[g(X,\theta)]$ and $\mathcal{N}_Q = \{\theta \in \Theta : |\theta - \theta_0| < \delta, Q(\theta) < \epsilon\}$. Pick any $\delta, \epsilon \in (0,\infty)$ small enough so that $\mathrm{cl}\left(\mathcal{N}_Q\right) \subset \{\theta \in \Theta : |\theta - \theta_0| \le \delta\} \subset \mathcal{N}$. For any $\theta \in \mathcal{N}_Q^c \setminus \{\theta \in \Theta : |\theta - \theta_0| \le \delta\}$, Assumption P implies

$$\begin{aligned} Q_n(\theta) &= Q(\theta) + 2a_n E[A_n(X) g(X,\theta)]' W_n E[g(X,\theta)] + a_n^2 \left(E[A_n(X) g(X,\theta)]' W_n E[A_n(X) g(X,\theta)]\right. \\ &\quad \left. - E[A_n(X) g(X,\theta_0)]' W_n E[A_n(X) g(X,\theta_0)]\right) \\ &> \epsilon/2, \end{aligned} \tag{16}$$

for all n large enough. From $Q_n(\theta_0) = 0$, the point $\theta_{1n} = \arg\min_{\theta \in \mathrm{cl}(\mathcal{N}_Q)} Q_n(\theta)$ (which always exists by the Weierstrass theorem) is a global minimizer of $Q_n(\theta)$ on $\{\theta \in \Theta : |\theta - \theta_0| \le \delta\}$. Also, since $\theta_{1n} \notin \mathcal{N}_Q^c$, the minimizer θ_{1n} belongs to $\mathcal{N}_Q$ (i.e., an interior solution of $\min_{\theta \in \mathrm{cl}(\mathcal{N}_Q)} Q_n(\theta)$), which implies that θ_{1n} satisfies the first-order condition $D_n(\theta_{1n}) = 0$.

Fourth, we show the uniqueness of θ_{1n}. To this end, it is sufficient to show that $D_n(\theta)$ is one-to-one on the set $\{\theta \in \Theta : |\theta - \theta_0| \le \delta\}$ for sufficiently small δ. Pick any $\theta, \theta + \vartheta \in \{\theta \in \Theta : |\theta - \theta_0| < \delta\} \subset \mathcal{N}$ (taking δ small enough) with $\vartheta \neq 0$. From the triangle inequality,

$$|D_n(\theta + \vartheta) - D_n(\theta)| \ge \left|2G'WG\vartheta\right| - \left|D_n(\theta + \vartheta) - D_n(\theta) - 2G'WG\vartheta\right|. \tag{17}$$

Since G is full rank and W is positive definite (Assumption G1(ii) and W), the first term $\left|2G'WG\vartheta\right|$ is a positive constant. Also the second term of (17) satisfies

$$\frac{1}{2}\left|D_n(\theta+\vartheta)-D_n(\theta)-2G'WG\vartheta\right| \le C|\vartheta|a_n + C|W_n-W| + C\left(E_n[H(X)]\delta^\beta + a_n\right),$$

where the inequality follows from an expansion of $E_n[g(X,\theta+\vartheta)]$ around $\vartheta=0$ and

$$|E_n[G(X,\theta)]-G| \le E_n[H(X)]|\theta-\theta_0|^\beta + Ca_n, \tag{18}$$

for each $\theta \in \{\theta\in\Theta : |\theta-\theta_0|<\delta\}$ (by Assumptions P and G1(ii)). Since the first term of (17) is positive and the second term of (17) can be arbitrary small for sufficiently small δ and large n, we obtain $|D_n(\theta+\vartheta)-D_n(\theta)|>0$ for all δ small enough and n large enough. Therefore, θ_{1n} exists uniquely for all n large enough.

Finally, we show (8). By expanding $D_n(\theta_{1n})=0$ around $\theta_{1n}=\theta_0$ with Assumption P and (18),

$$\begin{aligned}0 = {} & G'W\left\{a_nE[A_n(X)g(X,\theta_0)] + G'(\theta_{1n}-\theta_0)\right\} + O\left((a_n+|\theta_{1n}-\theta_0|)|W_n-W| + |\theta_{1n}-\theta_0|^{1+\beta}\right.\\ & \left. + |\theta_{1n}-\theta_0|^2 + a_n|\theta_{1n}-\theta_0|^\beta + a_n|\theta_{1n}-\theta_0| + a_n^2\right).\end{aligned}$$

Solving this equation for θ_{1n} yields (8). □

Proof of (ii). Let

$$\begin{aligned}&B_{1n} = \left\{\left|\hat\theta_1-\theta_0\right| \le \epsilon, \hat G\left(\hat\theta_1\right)'\hat W\hat g\left(\hat\theta_1\right)=0\right\}, \qquad BG_n = \left\{\left|\hat G(\theta_0)-G\right| \le c_G n^{-1/2}z_n\right\},\\ &BH_n = \left\{\frac{1}{n}\sum_{i=1}^n H(X_i) \le E[H(X)]+1\right\}, \qquad BW_n = \left\{\left|\hat W-W_n\right| \le c_W n^{-1/2}z_n\right\},\\ &Y_{in} = I_{1n}^{-1/2}G_{1n}'W_n g(X_{in},\theta_{1n}), \qquad T_{1n} = I_{1n}^{-1/2}G_{1n}'W_nG_{1n}\left(\hat\theta_1-\theta_{1n}\right),\\ &I_{1n} = G_{1n}'W_n\Omega_{1n}W_nG_{1n}, \qquad I_1 = G'W\Omega WG, \qquad t_{1n} = |\theta_{1n}-\theta_0|^\beta + n^{-1/2}z_n,\\ &G_{1n} = E_n[G(X,\theta_{1n})], \qquad \Omega_{1n} = E_n\left[g(X,\theta_{1n})g(X,\theta_{1n})'\right],\end{aligned}$$

for $\epsilon, c_G, c_W \in (0,\infty)$. Note that since $|I_{1n}-I_1|\to 0$ and I_1 is positive definite (by Assumption G1(ii) and W), $I_{1n}^{-1/2}$ exists for all n large enough. For a.e. $\mathbf{x}_n \in B_{1n}$ and all n large enough, an expansion of $\hat G\left(\hat\theta_1\right)'\hat W\hat g\left(\hat\theta_1\right)=0$ around $\hat\theta_1=\theta_{1n}$ yields

$$\begin{aligned}0 = {} & \frac{1}{n}\sum_{i=1}^n Y_{in} + T_{1n} + I_{1n}^{-1/2}\left\{\left(\hat G\left(\hat\theta_1\right)-G_{1n}\right)'\hat W + G_{1n}'\left(\hat W-W_n\right)\right\}\hat g(\theta_{1n})\\ & + I_{1n}^{-1/2}\begin{Bmatrix}\left(\hat G\left(\hat\theta_1\right)-G_{1n}\right)'\hat W\hat G\left(\bar\theta_1\right)\\ +G_{1n}'\hat W\left(\hat G\left(\bar\theta_1\right)-G_{1n}\right)+G_{1n}'\left(\hat W-W_n\right)G_{1n}\end{Bmatrix}\left(\hat\theta_1-\theta_{1n}\right),\end{aligned} \tag{19}$$

where $\bar\theta_1$ is a point between $\hat\theta_1$ and θ_{1n}. Observe that for a.e. $\mathbf{x}_n \in B_{1n}\cap BG_n\cap BH_n\cap BW_n$ and all n large enough and ϵ small enough so that $\{\theta\in\Theta : |\theta-\theta_0|<\epsilon\}\subset\mathcal{N}$, Assumptions P and G1 guarantee

$$\begin{aligned}&\left|\hat G\left(\hat\theta_1\right)-G_{1n}\right| \le C\left(\left|\hat\theta_1-\theta_{1n}\right|^\beta + t_{1n}\right), \qquad |G_{1n}| \le C|\theta_{1n}-\theta_0|^\beta + |G|,\\ &\left|\hat W-W_n\right| \le c_W n^{-1/2}z_n, \qquad |I_{1n}| \le C\left(|\theta_{1n}-\theta_0| + |\theta_{1n}-\theta_0|^\beta + |W_n-W|\right) + |I_1|.\end{aligned} \tag{20}$$

Thus, for a.e. $\mathbf{x}_n \in B_{1n}\cap BG_n\cap BH_n\cap BW_n$ and all n large enough and ϵ small enough, the norms of the third and fourth terms of (19) are bounded by $C\left(|T_{1n}|^\beta+t_{1n}\right)\left|\frac{1}{n}\sum_{i=1}^n Y_{in}\right|$ and $C\left(|T_{1n}|^\beta+t_{1n}\right)|T_{1n}|$, respectively. Combining these results, for a.e. $\mathbf{x}_n \in B_{1n}\cap BG_n\cap BH_n\cap BW_n$ and all n large enough and ϵ small enough,

$$\begin{aligned}&\left|\frac{1}{n}\sum_{i=1}^n Y_{in}\right| \ge \frac{1-C\left\{|T_{1n}|^\beta+t_{1n}+\left(|T_{1n}|^\beta+t_{1n}\right)^2\right\}}{1+C\left(|T_{1n}|^\beta+t_{1n}\right)}|T_{1n}|,\\ &\left|\frac{1}{n}\sum_{i=1}^n Y_{in}\right| \le \frac{1+C\left\{|T_{1n}|^\beta+t_{1n}+\left(|T_{1n}|^\beta+t_{1n}\right)^2\right\}}{1-C\left(|T_{1n}|^\beta+t_{1n}\right)}|T_{1n}|.\end{aligned}$$

Let $\tilde B_{1n} = B_{1n}\cap BG_n\cap BH_n\cap BW_n$. Since $t_{1n}\to 0$ by $\theta_{1n}-\theta_0\to 0$ (from Part (i) of this theorem) and $n^{-1/2}z_n\to 0$, it holds that for all n large enough and ϵ small enough, and some sequence $c_n\to 0$,

$$P_n\left(|T_{1n}| \ge n^{-1/2}z_n\right) \le P_n\left(\left|n^{-1/2}\sum_{i=1}^n Y_{in}\right| \ge (1-c_n)z_n\right) + P_n\left(\tilde B_{1n}^c\right), \tag{21}$$

$$P_n\left(z_n n^{-1/2} \leq |T_{1n}|\right) \geq P_n\left(\left|n^{-1/2}\sum_{i=1}^{n} Y_{in}\right| \geq (1+c_n)\, z_n\right) - P_n\left(\tilde{B}_{1n}^c\right).$$

From [24], which establishes the large deviation results $P_n\left(B_{1n}^c\right) \leq Ce^{-cn}$ and $P_n\left(BH_n^c\right) \leq Ce^{-cn}$, and Assumption W,

$$\begin{aligned} P_n\left(\tilde{B}_{1n}^c\right) &\leq P_n\left(B_{1n}^c\right) + P_n\left(BG_n^c\right) + P_n\left(BH_n^c\right) + P_n\left(BW_n^c\right) \\ &\leq Ce^{-cn} + P_n\left(BG_n^c\right) + \exp\left\{-\frac{c_W^2 z_n^2}{2} + O\left(\frac{z_n^3}{\sqrt{n}}\right) + O\left(\log z_n\right)\right\}. \end{aligned} \tag{22}$$

Now consider the moderate deviation probability $P_n\left(BG_n^c\right)$. From Assumptions P and G1(ii), we have $E_n\left[\exp\left(T_4\left|G(X,\theta_0)\right|\right)\right] < C$ for all $n \in \mathbb{N}$. Then for each $v \in \mathbb{R}^{d_g}$, $j = 1, \ldots, d_\theta$, and $k \in \mathbb{N}$,

$$\begin{aligned} E_n\left[\left(v' G_j(X,\theta_0)\right)^k\right] &\leq |v|^k E_n\left[\left|G_j(X,\theta_0)\right|^k\right] \leq |v|^k T_4^{-k} k! E_n\left[(k!)^{-1}\left|G_j(X,\theta_0)\right|^k\right] \\ &\leq |v|^k T_4^{-k} k! E_n\left[\exp\left(T_4\left|G_j(X,\theta_0)\right|\right)\right] < \infty. \end{aligned}$$

Therefore, we can apply Yurinskii [30, Theorem 3.1], which implies

$$P_n\left(BG_n^c\right) \leq \exp\left\{-\frac{c_G^2 z_n^2}{2} + O\left(\frac{z_n^3}{\sqrt{n}}\right) + O\left(\log z_n\right)\right\}. \tag{23}$$

From (22), (23), and taking c_W and c_G small enough, there exists some $\bar{c} \in (0, 1)$ satisfying

$$P_n\left(\tilde{B}_{1n}^c\right) \leq \exp\left\{-\frac{\bar{c} z_n^2}{2} + O\left(\frac{z_n^3}{\sqrt{n}}\right) + O\left(\log z_n\right)\right\}. \tag{24}$$

Also, since $E_n[Y_{in}] = 0$, $E_n\left[Y_{in} Y_{in}'\right]$ equals the identity matrix, and $E_n\left[\exp\left(\bar{T}\left|Y_{in}\right|\right)\right] < C$ for some $\bar{T} \in (0, \infty)$ (by $|G_{1n}| \leq C$, $|W_n| \leq C$, $|\Omega_{1n}| \leq C$, and Assumption G1(i)), we can apply the same argument as the proof of Inglot and Kallenberg [15, Lemma 4.2] which yields

$$P_n\left(\left|n^{-1/2}\sum_{i=1}^{n} Y_{in}\right| \geq z_n\right) = \exp\left\{-\frac{z_n^2}{2} + O\left(\frac{z_n^3}{\sqrt{n}}\right) + O\left(\log z_n\right)\right\}. \tag{25}$$

Combining (21), (24) and (25), we obtain the conclusion. □

A.2. Proof of Theorem 3.2

Based on Theorem 3.1, it is sufficient to show that Assumption W is satisfied with $\hat{W} = \hat{\Omega}^{-1}$, $W_n = \Omega_{1n}^{-1}$, and $W = \Omega^{-1}$. A detailed proof is available from the author upon request.

A.3. Proof of Theorem 3.3

Proof of (i). First, we show the continuity of

$$Q_{\rho n}(\theta) = E_n\left[\rho\left(\lambda_n(\theta)' g(X,\theta)\right)\right] - E_n\left[\rho\left(\lambda_n(\theta_0)' g(X,\theta_0)\right)\right],$$

in $\theta \in \mathcal{N}$, where the neighborhood $\mathcal{N}$ around θ_0 appears in Assumption G1(ii) and $\lambda_n(\theta) = \arg\max_{\lambda \in \Lambda} E_n\left[\rho\left(\lambda' g(X,\theta)\right)\right]$. Note that the maximizer $\lambda_n(\theta)$ exists for each $\theta \in \Theta$ and $n \in \mathbb{N}$ by Assumption G3(i) and the Weierstrass theorem. Therefore, $Q_{\rho n}(\theta)$ is well defined for each $\theta \in \Theta$ and $n \in \mathbb{N}$. Since $\rho(\cdot)$ is strictly concave and Λ is compact, the maximum theorem guarantees that $\lambda_n(\theta)$ is continuous in $\theta \in \mathcal{N}$ for each $n \in \mathbb{N}$. Pick any $\vartheta, \theta \in \mathcal{N}$. By expansions of $\rho\left(\lambda_n(\vartheta)' g(X,\vartheta)\right)$ and $g(X,\vartheta)$ around $\lambda_n(\vartheta) = \lambda_n(\theta)$ and $\vartheta = \theta$, respectively,

$$\begin{aligned} \left|Q_{\rho n}(\vartheta) - Q_{\rho n}(\theta)\right| &\leq \left|E_n\left[\rho_1\left(\lambda_n(\theta)' g(X,\tilde{\vartheta})\right) G(X,\tilde{\vartheta})\right]\right| \left|\lambda_n(\theta)\right| \left|\vartheta - \theta\right| \\ &\quad + \left|E_n\left[\rho_1\left(\tilde{\lambda}_n' g(X,\vartheta)\right) g(X,\vartheta)\right]\right| \left|\lambda_n(\vartheta) - \lambda_n(\theta)\right|, \end{aligned} \tag{26}$$

for each $n \in \mathbb{N}$, where $\tilde{\lambda}_n$ is a point on the line joining $\lambda_n(\vartheta)$ and $\lambda_n(\theta)$ and $\tilde{\vartheta}$ is a point on the line joining ϑ and θ. From Assumptions P and G1(ii), and G3(i),

$$\left|E_n\left[\rho_1\left(\lambda_n(\theta)' g(X,\tilde{\vartheta})\right) G(X,\tilde{\vartheta})\right]\right| < \infty, \qquad \left|E_n\left[\rho_1\left(\tilde{\lambda}_n' g(X,\vartheta)\right) g(X,\vartheta)\right]\right| < \infty, \tag{27}$$

for each $n \in \mathbb{N}$. From (26), (27), and continuity of $\lambda_n(\theta)$ in $\theta \in \mathcal{N}$, $Q_{\rho n}(\theta)$ is continuous in $\theta \in \mathcal{N}$ for each $n \in \mathbb{N}$.

Second, we show the differentiability of $Q_{\rho n}(\theta)$ in $\theta \in \mathcal{N}$. Pick any $\theta \in \mathcal{N}$ and $\varepsilon \neq 0$. By expansions of $\rho\left(\lambda_n\left(\theta+\varepsilon e_j\right)' g\left(X,\theta+\varepsilon e_j\right)\right)$ and $g\left(X,\theta+\varepsilon e_j\right)$ around $\lambda_n\left(\theta+\varepsilon e_j\right)=\lambda_n(\theta)$ and $\varepsilon=0$, respectively,

$$\begin{aligned}
&\left|\varepsilon^{-1}\left\{Q_{\rho n}\left(\theta+\varepsilon e_j\right)-Q_{\rho n}(\theta)\right\}-E_n\left[\rho_1\left(\lambda_n(\theta)'g(X,\theta)\right)G_j(X,\theta)'\right]\lambda_n(\theta)\right| \\
&\leq\left|E_n\left[\rho_1\left(\lambda_n\left(\theta+\dot{\varepsilon}e_j\right)' g\left(X,\theta+\dot{\varepsilon}e_j\right)\right)g\left(X,\theta+\dot{\varepsilon}e_j\right)'\right]\frac{\mathrm{d}\lambda_n\left(\theta+\dot{\varepsilon}e_j\right)}{\mathrm{d}\theta_j}\right| \\
&\quad+\left|E_n\left[\rho_1\left(\lambda_n(\theta)'g\left(X,\theta+\dot{\varepsilon}e_j\right)\right)G_j\left(X,\theta+\dot{\varepsilon}e_j\right)-\rho_1\left(\lambda_n(\theta)'g(X,\theta)\right)G_j(X,\theta)\right]\right|\ |\lambda_n(\theta)| \qquad (28)
\end{aligned}$$

for any ε small enough, where $\dot{\lambda}_n$ is a point between $\lambda_n\left(\theta+\varepsilon e_j\right)$ and $\lambda_n(\theta)$, and $\dot{\varepsilon}$ is a point between ε and 0. The implicit function theorem guarantees the existence of $\frac{\mathrm{d}\lambda_n\left(\theta+\dot{\varepsilon}e_j\right)}{\mathrm{d}\theta_j}$ for any ε small enough. Also, since $\lambda_n(\theta)\to\bar{\lambda}(\theta)\in\operatorname{int}(\Lambda)$ for each $\theta\in\mathcal{N}$, $\lambda_n(\theta)$ satisfies the first-order condition

$$E_n\left[\rho_1\left(\lambda_n(\theta)'g(X,\theta)\right)g(X,\theta)\right]=0, \qquad (29)$$

for each $\theta\in\mathcal{N}$, which implies that the first term of (28) is zero. So, by taking $\varepsilon\to 0$ (so, $\dot{\varepsilon}\to 0$ as well) with Assumptions P and G1(ii), and G3(i) and (27), we obtain the differentiability of $Q_{\rho n}(\theta)$ on $\mathcal{N}$ for each $n\in\mathbb{N}$ with the derivative $D_{\rho n}(\theta)=E_n\left[\rho_1\left(\lambda_n(\theta)'g(X,\theta)\right)G(X,\theta)'\right]\lambda_n(\theta)$.

Third, we show the existence of θ_{3n}. Let $Q_\rho(\theta)=E\left[\rho\left(\bar{\lambda}(\theta)'g(X,\theta)\right)\right]$ and $\mathcal{N}_3=\left\{\theta\in\Theta:|\theta-\theta_0|<\delta, Q_\rho(\theta)<\epsilon\right\}$. Pick $\delta,\epsilon\in(0,\infty)$ small enough so that $\operatorname{cl}(\mathcal{N}_3)\subset\{\theta\in\Theta:|\theta-\theta_0|\leq\delta\}\subset\mathcal{N}$. For any $\theta\in\mathcal{N}_3^c\setminus\{\theta\in\Theta:|\theta-\theta_0|\leq\delta\}$, expansions around $\lambda_n(\theta)=\bar{\lambda}(\theta)$ and $\lambda_n(\theta_0)=0$ with Assumption P yield

$$\begin{aligned}
Q_{\rho n}(\theta)&=Q_\rho(\theta)+E\left[\rho\left(\dot{\lambda}_n'g(X,\theta)\right)g(X,\theta)'\right]\left(\lambda_n(\theta)-\bar{\lambda}(\theta)\right)+a_nE\left[A_n(X)\rho\left(\lambda_n(\theta)'g(X,\theta)\right)\right] \\
&\quad-E\left[\rho\left(\ddot{\lambda}_n'g(X,\theta_0)\right)g(X,\theta_0)'\right]\lambda_n(\theta_0)-a_nE\left[A_n(X)\rho\left(\lambda_n(\theta_0)'g(X,\theta_0)\right)\right]>\epsilon/2, \qquad (30)
\end{aligned}$$

for all n large enough, where $\dot{\lambda}_n$ is a point on the line joining $\lambda_n(\theta)$ and $\bar{\lambda}(\theta)$ and $\ddot{\lambda}_n$ is a point on the line joining $\lambda_n(\theta_0)$ and 0. From $Q_{\rho n}(\theta_0)=0$, the point $\theta_{3n}=\arg\max_{\theta\in\operatorname{cl}(\mathcal{N}_3)}Q_{\rho n}(\theta)$ (which always exists by the Weierstrass theorem) is a global maximizer of $Q_{\rho n}(\theta)$ on $\{\theta\in\Theta:|\theta-\theta_0|\leq\delta\}$. Also, since $\theta_{3n}\notin\mathcal{N}_3^c$, the maximizer θ_{3n} belongs to $\mathcal{N}_3$ (i.e., θ_{3n} is a interior solution of $\max_{\theta\in\operatorname{cl}(\mathcal{N}_3)}Q_{\rho n}(\theta)$), which implies that θ_{3n} satisfies the first-order condition $D_{\rho n}(\theta_{3n})=0$.

Fourth, we show the uniqueness of θ_{3n}. To this end, it is sufficient to show that $D_{\rho n}(\theta)$ is one-to-one on the set $\{\theta\in\Theta:|\theta-\theta_0|\leq\delta\}$ for sufficiently small δ. Pick any $\theta,\theta+\vartheta\in\{\theta\in\Theta:|\theta-\theta_0|<\delta\}\subset\mathcal{N}$ (taking δ small enough) with $\vartheta\neq 0$. From the triangle inequality,

$$\left|D_{\rho n}(\theta+\vartheta)-D_{\rho n}(\theta)\right|\geq\left|G'\Omega^{-1}G\vartheta\right|-\left|D_{\rho n}(\theta+\vartheta)-D_{\rho n}(\theta)-G'\Omega^{-1}G\vartheta\right|. \qquad (31)$$

From Assumption G1(ii), $\left|G'\Omega^{-1}G\vartheta\right|$ is a positive constant. By the triangle inequality,

$$\begin{aligned}
&\left|D_{\rho n}(\theta+\vartheta)-D_{\rho n}(\theta)-G'\Omega^{-1}G\vartheta\right|\leq\left|G'\left(\lambda_n(\theta+\vartheta)-\lambda_n(\theta)-\Omega^{-1}G\vartheta\right)\right| \\
&\quad+\left|E_n\left[\rho_1\left(\lambda_n(\theta+\vartheta)'g(X,\theta+\vartheta)\right)G(X,\theta+\vartheta)'\right]-G'\right|\ |\lambda_n(\theta+\vartheta)-\lambda_n(\theta)| \\
&\quad+\left|E_n\left[\rho_1\left(\lambda_n(\theta+\vartheta)'g(X,\theta+\vartheta)\right)G(X,\theta+\vartheta)'\right]\right|\ |\lambda_n(\theta)|+\left|E_n\left[\rho_1\left(\lambda_n(\theta)'g(X,\theta)\right)G(X,\theta)\right]\right|\ |\lambda_n(\theta)| \\
&=A_1+A_2+A_3+A_4.
\end{aligned}$$

By expanding (29) around $\lambda_n(\theta)=0$ and solving for $\lambda_n(\theta)$,

$$\lambda_n(\theta)=\hat{\Omega}_\rho(\theta)^{-1}E_n[g(X,\theta)], \qquad (32)$$

where $\hat{\Omega}_\rho(\theta)=E_n\left[\rho_2\left(\tilde{\lambda}_n'g(X,\theta)\right)g(X,\theta)g(X,\theta)'\right]$ and $\tilde{\lambda}_n$ is a point on the line joining $\lambda_n(\theta)$ and 0 (note that by Assumption G1(ii) and G3, $\hat{\Omega}_\rho(\theta)$ is invertible for any δ small enough and n large enough). Thus, an expansion around $\theta=\theta_0$ combined with Assumptions P and G3(ii) and (15) yields $|\lambda_n(\theta)|\leq C(a_n+\delta)$. Similarly, we have $|\lambda_n(\theta+\vartheta)|\leq C(a_n+\delta)$. Thus, from Assumption G3(i), we have $A_2,A_3,A_4\leq C(a_n+\delta)$. We now consider A_1. From (32) (which also holds for $\lambda_n(\theta+\vartheta)$),

$$\begin{aligned}
\lambda_n(\theta+\vartheta)-\lambda_n(\theta)-\Omega^{-1}G\vartheta&=\left(E_n\left[\rho_2\left(\dot{\lambda}_n'g(X,\theta+\vartheta)\right)g(X,\theta+\vartheta)g(X,\theta+\vartheta)'\right]^{-1}-\Omega^{-1}\right)E_n[g(X,\theta+\vartheta)] \\
&\quad+\left(\Omega^{-1}-\hat{\Omega}_\rho(\theta)^{-1}\right)E_n[g(X,\theta)]+\Omega^{-1}\left(E_n[g(X,\theta+\vartheta)]-E_n[g(X,\theta)]-G\vartheta\right)=A_{11}+A_{12}+A_{13},
\end{aligned}$$

where $\dot{\lambda}_n$ is a point between $\lambda_n(\theta+\vartheta)$ and 0. An expansion around $\theta=\theta_0$ and Assumption G1(ii) imply $|E_n[g(X,\theta)]|\leq\left|E_n\left[G\left(X,\tilde{\theta}\right)\right]\right|\ |\theta-\theta_0|\leq C\delta$ and thus $|A_{12}|\leq C\delta$ (using Assumption G3(ii)). Similarly, we have $|E_n[g(X,\theta+\vartheta)]|\leq C\delta$

and $|A_{11}| \leq C\delta$. For A_{13}, an expansion around $\vartheta = 0$ combined with Assumptions P and G1(ii) yield $|A_{13}| \leq C\left(\delta^{\beta} + a_n\right)$. Combining these results, we can see that the first term of (31) is positive and the second term of (31) can be arbitrary small for sufficiently small δ and large n. Therefore, we obtain $\left|D_{\rho n}(\theta + \vartheta) - D_{\rho n}(\theta)\right| > 0$ for all δ small enough and n large enough, which implies that θ_{3n} exists uniquely for all n large enough.

Finally, we show (12). From (32),

$$\begin{aligned} 0 = D_{\rho n}(\theta_{3n}) &= \left(E_n\left[\rho_1\left(\lambda_n(\theta_{3n})' g(X, \theta_{3n})\right) G(X, \theta_{3n})\right] + G\right)' \hat{\Omega}_{\rho}(\theta_{3n})^{-1} E_n[g(X, \theta_{3n})] \\ &\quad - G'\left(\hat{\Omega}_{\rho}(\theta_{3n})^{-1} - \Omega^{-1}\right) E_n[g(X, \theta_{3n})] - G'\Omega^{-1} E_n[g(X, \theta_{3n})] \\ &= A_5 - A_6 - A_7. \end{aligned} \tag{33}$$

For A_5, the triangle inequality, Assumption G3(ii), and an expansion around $\lambda_n(\theta_{3n}) = 0$ imply

$$\begin{aligned} |A_5| &\leq C\,|E_n[G(X, \theta_{3n})] - G|\;|E_n[g(X, \theta_{3n})]| \\ &\quad + C\left|E_n\left[\rho_2\left(\tilde{\lambda}_n' g(X, \theta_{3n})\right) G(X, \theta_{3n})' g(X, \theta_{3n})\right]\right|\;|\lambda_n(\theta_{3n})|\;|E_n[g(X, \theta_{3n})]| \\ &\leq C\left(|\theta_{3n} - \theta_0|^{\beta} + a_n\right)(|\theta_{3n} - \theta_0| + a_n) + (|\theta_{3n} - \theta_0| + a_n)^2, \end{aligned}$$

where the second inequality follows from Assumptions P and G1(ii), and G3(i), an expansion of $E_n[g(X, \theta_{3n})]$ around $\theta_{3n} = \theta_0$, and (32) (which guarantees $|\lambda_n(\theta_{3n})| \leq C(|\theta_{3n} - \theta_0| + a_n)$). Similarly, A_6 satisfies $|A_6| \leq C(|\theta_{3n} - \theta_0| + a_n)^2$. For A_7, an expansion around $\theta_{3n} = \theta_0$ yields

$$A_7 = G'\Omega^{-1} E_n[g(X, \theta_0)] + G'\Omega^{-1} G(\theta_{3n} - \theta_0) + O\left(|\theta_{3n} - \theta_0|^{1+\beta} + a_n|\theta_{3n} - \theta_0|\right),$$

where $\bar{\theta}_3$ is a point between θ_{3n} and θ_0. Inserting these results to (33) and solving for θ_{3n}, we have (12). □

Proof of (ii). Pick any $n \in \mathbb{N}$. Let $\hat{\lambda}(\theta) = \arg\max_{\lambda \in \Lambda} \frac{1}{n}\sum_{i=1}^{n} \rho\left(\lambda' g(X_i, \theta)\right)$,

$$B_{3n} = \left\{\left|\hat{\theta}_3 - \theta_0\right| \leq \epsilon, \hat{D}_{\rho}\left(\hat{\theta}_3\right) = 0\right\}, \qquad B\Omega_n = \left\{\left|\frac{1}{n}\sum_{i=1}^{n} g(X_{in}, \theta_0) g(X_{in}, \theta_0)' - \Omega\right| \leq c_{\Omega} n^{-1/2} z_n\right\},$$

$$\hat{D}_{\rho}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \rho_1\left(\hat{\lambda}(\theta)' g(X_{in}, \theta)\right) G(X_{in}, \theta)'\hat{\lambda}(\theta), \qquad Y_{3in} = -I_{3n}^{-1/2} G_{3n}' \Omega_{3n}^{-1} g(X_{in}, \theta_{3n}),$$

$$T_{3n} = -I_{3n}^{1/2}\left(\hat{\theta}_3 - \theta_{3n}\right), \qquad I_{3n} = G_{3n}'\Omega_{3n}^{-1} G_{3n}, \qquad I_3 = G'\Omega^{-1} G,$$

$$G_{3n} = E_n[G(X, \theta_{3n})], \qquad \Omega_{3n} = E_n\left[g(X, \theta_{3n}) g(X, \theta_{3n})'\right], \qquad t_{3n} = |\theta_{3n} - \theta_0|^{\beta} + n^{-1/2} z_n,$$

for $\epsilon \in (0, \infty)$. Since $|I_{3n} - I_3| \to 0$ and I_3 is positive definite (by Assumption G1(ii)), $I_{3n}^{-1/2}$ exists for all n large enough. For a.e. $\mathbf{x}_n \in B_{3n}$, the condition $\hat{D}_{\rho}\left(\hat{\theta}_3\right) = 0$ for $\hat{\theta}_3$ satisfies

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{i=1}^{n} Y_{3in} + T_{3n} - I_{3n}^{-1/2}\{(\hat{G}(\hat{\theta}_3) - G_{3n})'\hat{\Omega}_{\rho}(\hat{\theta}_3)^{-1} - G_{3n}'(\hat{\Omega}_{\rho}(\hat{\theta}_3)^{-1} - \Omega_{3n}^{-1})\}\hat{g}(\theta_{3n}) \\ &\quad - I_{3n}^{-1/2}\begin{Bmatrix} (\hat{G}(\hat{\theta}_3) - G_{3n})'\hat{\Omega}_{\rho}(\hat{\theta}_3)^{-1}\hat{G}(\bar{\theta}_3) \\ -G_{3n}'(\hat{\Omega}_{\rho}(\hat{\theta}_3)^{-1} - \Omega_{3n}^{-1})\hat{G}(\bar{\theta}_3) - G_{3n}'\Omega_{3n}^{-1}(\hat{G}(\bar{\theta}_3) - G_{3n}) \end{Bmatrix}(\hat{\theta}_3 - \theta_{3n}), \end{aligned}$$

for all n large enough, where the second equality follows from (32) and an expansion around $\hat{\lambda}\left(\hat{\theta}_3\right) = 0$ ($\bar{\lambda}_3$ is a point on the line joining $\hat{\lambda}\left(\hat{\theta}_3\right)$ and 0), and the third equality follows from an expansion around $\hat{\theta}_3 = \theta_{3n}$ ($\bar{\theta}_3$ is a point on the line joining $\hat{\theta}_3$ and θ_{3n}). Note that for a.e. $\mathbf{x}_n \in B_{3n} \cap BG_n \cap BH_n \cap B\Omega_n$ with any ϵ small enough so that $\{\theta \in \Theta : |\theta - \theta_0| < \epsilon\} \subset \mathcal{N}$, a similar argument to (20) yields

$$\left|\hat{G}\left(\hat{\theta}_3\right) - G_{3n}\right| \leq C|\theta_{3n} - \theta_0|^{\beta} + |G|, \qquad \left|\hat{\Omega}_{\rho}\left(\hat{\theta}_3\right)^{-1} - \Omega_{3n}^{-1}\right| \leq C\left(\left|\hat{\theta}_3 - \theta_{3n}\right|^{\beta} + t_{3n}\right),$$

$$\left|\Omega_{3n}^{-1}\right| \leq C|\theta_{3n} - \theta_0|^{\beta} + \left|\Omega^{-1}\right|, \qquad |I_{3n}| \leq C\left(|\theta_{3n} - \theta_0| + |\theta_{3n} - \theta_0|^{\beta}\right) + |I_3|.$$

Thus, for a.e. $\mathbf{x}_n \in B_{3n} \cap BG_n \cap BH_n \cap B\Omega_n$ with any ϵ small enough and n large enough,

$$\left|\frac{1}{n}\sum_{i=1}^{n} Y_{3in} + T_{3n}\right| \leq C\left(|T_{3n}|^{\beta} + t_{3n}\right)\left|\frac{1}{n}\sum_{i=1}^{n} Y_{3in}\right| + C\left\{|T_{3n}|^{\beta} + t_{3n} + \left(|T_{3n}|^{\beta} + t_{3n}\right)^2\right\}|T_{3n}|.$$

Therefore, the same argument to the proof of Theorem 3.1 yields the conclusion. □

References

[1] M.A. Arcones, Large deviations for M-estimators, Annals of the Institute of Statistical Mathematics 58 (2006) 21–52.
[2] P.J. Bickel, C.A.J. Klaassen, Y. Ritov, J.A. Wellner, Efficient and Adaptive Estimation for Semiparametric Models, Springer, 1993.
[3] S.A. Book, The Cramér-Feller-Petrov large deviation theorem for triangular arrays, Working Paper, California State College, 1976.
[4] G. Chamberlain, Asymptotic efficiency in estimation with conditional moment restrictions, Journal of Econometrics 34 (1987) 305–334.
[5] N. Cressie, T.R.C. Read, Multinomial goodness-of-fit tests, Journal of the Royal Statistical Society: Series B 46 (1984) 440–464.
[6] A. Dembo, O. Zeitouni, Large Deviation Techniques and Applications, Springer, 1998.
[7] V.P. Godambe, An optimum property of regular maximum likelihood estimation, Annals of Mathematical Statistics 31 (1960) 1208–1212.
[8] A.R. Hall, Generalized Method of Moments, Oxford University Press, 2005.
[9] L.P. Hansen, Large sample properties of generalized method of moments estimators, Econometrica 50 (1982) 1029–1054.
[10] L.P. Hansen, J. Heaton, A. Yaron, Finite-sample properties of some alternative GMM estimators, Journal of Business & Economic Statistics 14 (1996) 262–280.
[11] J.W. Hardin, J.M. Hilbe, Generalized Estimating Equations, Chapman & Hall/CRC, 2002.
[12] G.W. Imbens, One-step estimators for over-identified generalized method of moments models, Review of Economic Studies 64 (1997) 359–383.
[13] G.W. Imbens, Generalized method of moments and empirical likelihood, Journal of Business and Economic Statistics 20 (2002) 493–506.
[14] G.W. Imbens, R.H. Spady, P. Johnson, Information theoretic approaches to inference in moment condition models, Econometrica 66 (1998) 333–357.
[15] T. Inglot, W.C.M. Kallenberg, Moderate deviations of minimum contrast estimators under contamination, Annals of Statistics 31 (2003) 852–879.
[16] J.L. Jensen, A.T.A. Wood, Large deviation and other results for minimum contrast estimators, Annals of the Institute of Statistical Mathematics 50 (1998) 673–695.
[17] J. Jurečková, W.C.M. Kallenberg, N. Veraverbeke, Moderate and Cramér-type large deviation theorems for M-estimators, Statistics & Probability Letters 6 (1988) 191–199.
[18] W.C.M. Kallenberg, On moderate deviation theory in estimation, Annals of Statistics 11 (1983) 498–504.
[19] W.C.M. Kallenberg, Efficiency, intermediate or Kallenberg, in: S. Kotz, C.B. Read, D.L. Banks (Eds.), Encyclopedia of Statistical Sciences, Wiley, New York, 1999, pp. 192–197.
[20] A.D.M. Kester, W.C.M. Kallenberg, Large deviations of estimators, Annals of Statistics 14 (1986) 648–664.
[21] Y. Kitamura, Empirical likelihood methods in econometrics: theory and practice, in: R.W. Blundell, W.K. Newey, T. Persson (Eds.), Advances in Economics and Econometrics, in: Theory and Applications: Ninth World Congress of the Econometric Society, vol. 3, Cambridge University Press, 2007, pp. 174–237. Chapter 8.
[22] Y. Kitamura, M. Stutzer, An information-theoretic alternative to generalized method of moments estimation, Econometrica 65 (1997) 861–874.
[23] W.K. Newey, R.J. Smith, Higher order properties of GMM and generalized empirical likelihood estimators, Econometrica 72 (2004) 219–255.
[24] T. Otsu, Large deviations of generalized method of moments and empirical likelihood estimators, Working Paper, 2009.
[25] A.B. Owen, Empirical likelihood ratio confidence intervals for a single functional, Biometrika 75 (1988) 237–249.
[26] J. Qin, J. Lawless, Empirical likelihood and general estimating equations, Annals of Statistics 22 (1994) 300–325.
[27] G.L. Sievers, Estimates of location: a large deviation comparison, Annals of Statistics 6 (1978) 610–618.
[28] R.J. Smith, Alternative semi-parametric likelihood approaches to generalized method of moments estimation, Economic Journal 107 (1997) 503–519.
[29] H. White, Estimation, Inference and Specification Analysis, Cambridge University Press, 1994.
[30] V.V. Yurinskii, Exponential inequalities for sums of random vectors, Journal of Multivariate Analysis 6 (1976) 473–499.