Inference for Moment Inequalities with Nuisance Functions

Keren Fang*

December 8, 2025

Most Updated Version in Personal Website: https://kerenfang.github.io

Abstract

This paper develops a method to construct confidence set for partially identified finite-dimensional parameters of interest from a finite number of moment inequalities which involve point identified infinite-dimensional nuisance parameters. We first point identify the nuisance parameters that are mean square projections and then construct the confidence set for the parameters of interest in two-steps. The effect of nuisance parameters on the confidence set is characterized by a corresponding influence function of the sample mean of the moment functions through a reparametrized GMM condition and the general formula of the asymptotic variance of the sample mean is derived. The violation to the moment inequalities while approximating the nuisance parameters is addressed along with the two-step procedure where a Bonferroni-type correction is critical to the uniform asymptotic size. We prove the uniform asymptotic size control property of the constructed confidence set, while the uniform bootstrapping consistency when nuisance parameters are present is derived. We illustrate the method by developing a complete structural model of a static discrete incomplete information game with state-dependent interaction effects among radio stations considered in De Paula and Tang (2012), where there is a need to approximate the true conditional choice probabilities as the nuisance parameters.

KEYWORDS: partial identification, uniform asymptotic size, pathwise derivative, Bonferroni-type correction, two-step procedure, bootstrap, moment inequalities, conditional choice probability, structure model

*The author would like to thank Ruoyao Shi, Tae-Hwy Lee, Donggyu Kim, Siyang Xiong, Xiaoxia Shi, Patrik Guggenberger, Weixin Yao, Max Farrell and many other visitors at University of California at Riverside for the help in developing the earlier drafts. All the errors are mine.

1 Introduction

Under an unconditional moment inequality model, when the parameters of interest are partilly identified without any nuisance parameters, the usual sample mean and sample variance of the moment functions, as the estimator of the asymptotic variance of the sample mean, can be applied. However, when the nuisance parameters are present and are approximated by their estimators, the asymptotic variance might be different from the usual sample variance. The confidence set of interest regarding the test concerning the moment inequalities will then be affected by the inconsistent variance estimator. And the size of the test cannot be controlled.

This paper develops a non-standard testing method to conduct inference on partially identified finite-dimensional parameters of interest in finite number of unconditioanl moment inequalities when there are point-identified infinite-dimensional nuisance parameters by the confidence set of the true parameters of interest, which has not been generally resolved in the literature. We first point identify the nuisance parameters taking the form of the mean square projections by finding its estimators. Through a reparametrized GMM condition, we derive the general formula of the sample mean of the moment functions and the estimator of the asymptotic variance of the sample mean applying the results from Newey (1994a). These estimators given the estimator of the nuisance parameters appreciate the effect of the nuisance parameters on the confidence set of interest. The confidence set of interest is then formed by a two-step procedure proposed by Romano et al. (2014). The first step is to form the confidence set of population moment functions μ_M, in order to deal with the slackness parameter $\sqrt{n}\mu_M$ that cannot be consistently estimated. The second step is to construct the confidence set of interest given μ_M which takes the value of the lower-left vertex of its confidence set along with a Bonferroni-type correction. In this two-step procedure, the probability that the estimator of the nuisance parameters violates the moment inequalities when the moment inequalities hold is considered through the first step confidence set along with the aforementioned estimators. The test corresponding to the Bonferroni confidence set is then shown to satisfy the uniform asymptotic size control (UASC) property. The result relies on the uniform bootstrapping consistency when the nuisance parameters are present in the test statistics.

The method has general implication in industrial organization, microeconometrics, political economics, labor economics, finance, etc. We will illustrate the method and demonstrate the general interest through the Conditional Choice Probability (CCP) that are nuisance parameters occurring in the moment inequality model which has not been fully addressed with in the literature. We develop a novel complete structure model of a static discrete incomplete information game with state-dependent interaction effects among radio stations previously considered in De Paula and Tang (2012). The construction of the structure model for the payoff-related parameters of interest involved in strategic behaviors of radio stations consists of three stages. The first stage is to find the unique Bayesian Nash Equilibrium (BNE) upon the information structure. The second stage is to adopt the BNE to form a unique prediction model in terms of CCP. The third stage is to establish the econometric model in terms of moment inequalities via taking expectation of the prediction model written in terms of inequalities. The confidence set of interest in this case is then founded by grid search.

The main contribution of this paper is the reparameterized GMM condition provided for the general unconditional moment inequality model with partially identified finite-dimensional parameters of interest and point identified infinite-dimensional nuisance parameters and the consequential derivation of the asymptotic variance of the sample mean of the moment functions. The reparameterized GMM condition helps to transform the complex moment inequalities to moment equalities so that the tools of semiparametric analysis can play a role. The paper provides a general framework for the unconditional moment inequality model with partially identified finite-dimensional parameters of interest and point identified infinite-dimensional nuisance parameters under least restrictive assumptions. Besides, the paper contributes to statistics by the derivation of the uniform bootstrapping consistency when the estimator of nuisance parameters is present. And it also contributes to the applied research and industrial organization research by the structure model or the general solution to deal with the CCP as the infinite-dimensional nuisance parameters occurring in the moment inequality model. The main caveat however is we do not provide the identification condition of the nuisance parameters while we assume it to be point identified, which is not the focus on the current paper. Besides, the further reduction of computation burden will be an interesting topic in the future.

When we are facing with moment equality conditions focusing on the parameters of parametric part but with some nonparametric nuisance parameters, the semiparametric model will be the general method to infer the parameters of interest in the parametric part, with rich literature including the nonparametric GLS for unknown variance function in Robinson (1987), the partial linear model in Robinson (1988), the MINPIN estimation of a criteria function in Andrews (1994), the nonparametric propensity score for average treatment effect estimation in Hahn (1998) and Hirano et al. (2003), the two-step GMM with nuisance parameters identified by conditional moment restrictions in Ackerberg et al. (2014) and the locally robust semiparametric estimation by a debiased GMM in Chernozhukov et al. (2022), etc.

When it comes to moment inequality conditions, even without the nuisance parameters, the usual M-estimation might not be applicable unless strong assumptions imposed while the parameters of interest are usually not point identified, i.e. the parameters of interest might not have a unique value when the moment inequalities hold. A moment inequality model under point identification has been addressed with the full column rank condition and the rich support condition, etc, in Tamer (2003). The parameters of interest therein could then be found by the semiparametric maximum likelihood estimator. Meanwhile, using any elements of the identified set to estimate the parameters of interest is impossible to have a consistent estimator since the value in the identified set cannot be distinguished from the true value of parameters of interest. And the general estimator of the identified set is hard to guarantee the set consistency, such as Hausdorff consistency. The usual way to deal with moment inequalities to conduct inference on partially identified parameters of interest thereby turns to focus on the confidence set under some hypothesis testing procedures so that the set can be found by inverting the test. The application of partial identification has been widely discussed in topics such as microeconomics, industrial organization, political economics, labor economics, finance, etc.

One approach to confidence set is the confidence set of the identified set. Manski and Tamer (2002) motivates it by a criteria function approach to find it as the estimator of the identified set and discusses the Haudorff consistency. Chernozhukov et al. (2007) extends it to a general framework of hypothesis testing procedure with the general criteria functions. Beresteanu and Molinari (2008) and Beresteanu et al. (2011) instead focus on a support functional approach based on the random set theory to perform the estimation and inference for convex identified sets. The estimator of the identified set or the confidence set of the identified set can be found by their corresponding sample analogs. Another approach focuses on the confidence set of the true parameter value, which is motivated by Imbens and Manski (2004). Andrews and Soares (2010a) and Andrews and Barwick (2012) extend it to a general framework where the slackness parameter is addressed with the general moment selection functions. Romano et al. (2014) on the other hand deals with it by a two-step procedure. In addition to unconditional moment inequality models, the conditional moment inequality models have been developed relatively afterwards. Andrews and Shi (2013) adopts an instrumental function approach to transform the original conditional moment inequalities into the equivalent unconditional moment inequalities. Another approach by Chernozhukov et al. (2013) utilizes a non-parametric estimator of the original conditional moment inequalities to estimate and to infer the parameters of interest. In this paper, we focus on the confidence set of the true parameter value that applies to more general identified sets which are not necessarily convex and that could be typically constructed smaller. In particular, we follow Romano et al. (2014) which can additionally control the probability that the estimator of the nuisance parameters violates the moment inequalities when the moment inequalities hold that involves with the UASC property.

When there are nuisance parameters in the moment inequalities, Shi and Shum (2015) develops a two-stage estimation method to concentrate out the point-identified finite dimensional nuisance parameters into moment equalities and then to estimate and infer the parameters of interest by a generalized minimum distance estimation procedure. Their size result, however, is pointwise in nuisance parameters and does not consider the effect of the approximation error of the nuisance parameters on the moment inequalities including that on the asymptotic variance of the sample mean of the moment functions and the violation of the moment inequalities while using the estimator of the nuisance parameters. For partially identified nuisance parameters, Kaido et al. (2019) considers the same moment inequality model as Andrews and Soares (2010b) and partially identify the subvector of parameters regarding all other parameters as nuisance parameters by a method of projection of confidence sets with calibrated critical values. Cox and Shi (2023) considers a conditional moment inequality model with nuisance parameters but only enter linearly and partially identifies the parameters of interest by a method of adaptive testing for subvectors using a quasi-likelihood ratio statistic through eliminating the nuisance parameters from moment inequalities which drastically reduce the computation burden. Andrews et al. (2023) considers a similar model to Cox and Shi (2023) with nuisance parameters enter linearly but uses a profiled studentized max statistics for a hybrid test combining the least-favorable and conditional methods. However, all these literatures for subvector parameters in the moment inequalities as the parameters of interest deal with finite-dimensional nuisance parameters only.

In this paper, we consider a general form of moment inequalities with infinite-dimensional nuisance parameters. We impose least restrictions regarding the nuisance parameters except some regularity conditions, the same way as Newey (1994a), such as Lipschitz continuity. And we impose least restrictions on the data generating process (DGP) except uniform integrability, similar to Romano et al. (2014). The way we get the confidence set of interest follows Romano et al. (2014), by a two-step procedure where Bonferroni-type correction is required. The confidence set of parameters of interest could thereby derived by inverting the test, in particular, bootstrapping the test statistics in two steps and then following a grid search procedure or some other proper algorithms.

The remainder of the paper is organized as follows. Section 2 discusses the moment inequality model and shows the uniform asymptotic size control. Section 3 provides the algorithm manual and the simulation study of an interval outcome regression model. Section 4 discusses an empirical application under a general game-theoretical structure model for radio stations in the U.S. Section 5 concludes.

2 Moment Inequalities

In this paper, we consider the true value θ_0 and h_0 that satisfy the moment inequalities:

$$E_{F_0}[m_j(W_i, \theta_0, h_0)] \geq 0, j = 1, \dots, k \tag{1}$$

where $\theta_0(\in \Theta \subset \mathbb{R}^d)$ is the finite-dimensional parameter of interest and $h_0(\in \mathcal{H} \subset \mathcal{L}^2(x))$ is the infinite-dimensional nuisance parameter. $X = x$ is a subvector of W that will be defined accordingly under a specific moment inequality model. $\{m_j(\cdot, \theta, h) : j = 1, \dots k\}$ are known real-valued moment functions. The observed random sample of random vectors $W_i(i = 1, \dots n)$ are assumed to be i.i.d. throughout the paper with the true joint distribution $F_0 \in \mathbf{F}$ on $\mathbb{R}^p$, where $\mathbf{F}$ is a class of nonparametric distributions. We consider W_i defined based on Lebesgue measure. The moment inequalities (1) are general enough since it could be adopted to form moment conditions for moment functions that less or equal to zero and thereby for moment equalities. Define the generics $(\theta, h, F) \in \mathcal{F}$. The key feature of this model is that θ_0 need not be point-identified, whereas h_0 is assumed to be point-identified under the form of mean square projections and the identification of h_0 is not the focus of this paper. That is, given the true DGP F_0, we could pin down $h_{0,}$ however we could not pin down θ_0 unless additional sufficient information is provided.

In this model, there are two main tasks in inferring θ_0. The first one comes from the nuisance parameter h_0 that will affect the estimation of θ_0, so that we need to estimate it first. In this paper, we are interest in point identified h_0 and estimating it by mean square projection, i.e. $h_0(x) = E(Y|X)$, where both X and Y are subvectors of W that will be determined under the specific models. We then characterize the impact of first stage estimation of the nuisance parameter h_0 on the second stage inference of the parameter of interest θ_0 by the pathwise derivative under an reparametrized GMM condition of (1) following Newey (1994a).

The second one comes from the inequality of the moment conditions (1), which implies that θ_0 is not necessarily point identified. In this paper, we are interested in partially identified θ_0 by getting a confidence set covering the true parameter θ_0 in the end, rather than forming additional conditions to point identify θ_0. We will achieve it in a two-step procedure following Romano et al. (2014) where the first step is to get the confidence set of population mean of the moment functions or μ_M in (1) and the second step is to get the confidence set of θ_0. However, the critical values are data-dependent and thereby depend on the true DGP, therefore the uniform asymptotic size control (UASC) regarding the confidence set of θ_0 is crucial and thereby demanded.

In the first stage, the consistent and asymptotic normal (CAN) estimator of h_0, i.e. $\widehat{h}_0$, could be obtained by the estimation of the mean square projection of the subvector Y on another subvector X, i.e. $\widehat{E(Y|X)}$, through the kernel density estimation or series estimation such that $\gamma_n(\widehat{h}_0 - h_0) \xrightarrow{d} \mathcal{N}(0, V_h)$. As the Assumption 3 in the appendix, it is sufficient to require that the convergence rate of $\widehat{h}_0$ is faster than $n^{-1/4}$ in terms of norm $\parallel h \parallel$ in this paper. Through the inference based on the sub-model of the DGP F_τ in the direction of τ, we could get the pathwise derivative of the population mean of the moment functions or μ_M in (1) denoted as a $k \times 1$ vector d(W) which measures the influence of $\widehat{h}_0$ on the second stage confidence set of θ_0 through the estimator of μ_M.

Given $\widehat{h}_0$, we want to conduct inference on partially identified θ_0 such that the system (1) is satisfied. In order to avoid the potential Hausdorff inconsistency issue of the identified set (IDS), that is $\Theta_0 = \{\theta \in \Theta : (1)\}$, and to fit more general IDS which might not be convex, we focus on the inference of θ_0 using the confidence set of θ_0 in the presence of the potential partial identification:

$$CS_n = \{\theta \in \Theta : T_n(\theta, \widehat{h}_0)) \leq C_n(1-\alpha, \theta, \widehat{h}_0)\}, \tag{2}$$

while we consider testing

$$H_{0:}\, \theta_0 = \theta \, s.t. \, E_{F_0}[m_j(W_i, \theta_0, h_0)] \geq 0, j = 1 \dots k \tag{3}$$

where the alternative is $H_1 : \theta_0 = \theta \, s.t. \, E_{F_0}[m_j(W_i, \theta_0, h_0)] < 0, j = 1 \dots k$ and $C_n(1-\alpha, \theta, \widehat{h}_0)$ is the $1-\alpha$ quantile of the asymptotic distribution of $T_n(\theta, \widehat{h_0}) \equiv T_n(\theta)$, where $\alpha \in (0,1)$. The idea of the confidence set derives from the duality of hypothesis testing and confidence set.

As pointed out by Andrews and Soares (2010a), there is a general class of statistics in the form of $T_n = S(\sqrt{n}\overline{m}_n(\theta), \widehat{\Sigma_n}(\theta))$ which includes

$$T_n^{max} = \underset{1 \leq j \leq k}{max} - \frac{\sqrt{n}\overline{m}_{j,n}}{S_{j,n}}, \tag{4}$$

$$T_n^{qlr} = \underset{t \in \mathbb{R}^k_{+,\infty}}{inf} \; Z_n(t)' \widehat{\Omega}_n^{-1} Z_n(t), \tag{5}$$

where $Z_n(t) = (\frac{\sqrt{n}(\overline{m}_{1,n}-t)}{S_{1,n}}, \dots, \frac{\sqrt{n}(\overline{m}_{k,n}-t)}{S_{k,n}})$,

$$T_n^{qlr,ad} = \underset{t \in \mathbb{R}^k_{+,\infty}}{inf} \; Z_n(t)' \widetilde{\widehat{\Omega}_n}^{-1} Z_n(t), \tag{6}$$

where $\widetilde{\Omega_n} = \widehat{\Omega_n} + max\{\epsilon - det(\widehat{\Omega_n}), 0\} \cdot I_k$, and

$$T_n^{mmm} = \Sigma_{j=1}^{k}\big(\frac{\sqrt{n}\overline{m}_{j,n}}{S_{j,n}}\big)^2 \cdot \mathbb{1}\{\overline{m}_{j,n} < 0\}. \tag{7}$$

The function $S : \mathbb{R}^k \times \mathbb{R}^{k\times k} \to \mathbb{R}$ is continuous in both arguments and weakly decreasing in each element of its first arguments. Besides, $S(m, \Sigma) = S(Dm, D\Sigma D)$ for all $m \in \mathbb{R}^k$, $\Sigma \in \mathbb{R}^{k\times k}$ and positive definite $D \in \mathbb{R}^{k\times k}$. In this paper, let $\mu_{M_j} \equiv \mu_{M_j}(\theta, h_0, F) \equiv E_F[m_j(W_i, \theta, h_0(x))]$, $j = 1, \ldots, k$, $\Sigma \equiv \Sigma_M(\theta, h_0, F) \equiv V_F[m(W_i, \theta, h_0(x)])$, $D \equiv Diag(\Sigma) = D(\theta, h_0, F)$, $\Omega \equiv \Omega(\theta, h_0, F) \equiv D^{-\frac{1}{2}}\Sigma D^{-\frac{1}{2}}$, $\sigma_j^2(\theta, h_0, F)$ denotes the j-th diagonal element of Σ, $\widehat{\mu}$ denotes the estimator of μ_M, $\widehat{\Sigma}$ denotes the estimator of Σ and similarly for $\widehat{D}$ and $\widehat{\Omega}$ and $S_{j,n}^2 \equiv \sigma_j^2(\theta, \widehat{h_0}, \widehat{F_n})$ where $\widehat{F_n}$ denotes the empirical distribution of the $W_i (i = 1, \ldots, n)$, that is $\widehat{F_n}(w) = \frac{1}{n}\Sigma_{i=1}^{n}\mathbb{1}\{\{W_i \le w\}$,where $\mathbb{1}\{\cdot\}$ denotes the indicator function. Hence, $\widehat{\mu_j} = \widehat{\mu_{j,n}} = \mu_{M_j}(\theta, \widehat{h_0}, \widehat{F_n}) = \overline{m}_{j,n}$ and $\widehat{\Sigma} = \Sigma_M(\theta, \widehat{h_0}, \widehat{F_n})$.

$$\begin{aligned} T_n(\theta) &= S(\sqrt{n}\widehat{\mu}(\theta), \widehat{\Sigma}(\theta)) \\ &= S\big(\sqrt{n}\widehat{D}^{-\frac{1}{2}}(\theta)(\widehat{\mu}(\theta) - \mu_M(\theta)) + \sqrt{n}\widehat{D}^{-\frac{1}{2}}(\theta)\mu_M(\theta), \widehat{\Omega}(\theta)\big) \end{aligned} \tag{8}$$

When we try to find the asymptotic distribution of $T_n(\theta)$, under suitable conditions, the inner part $\sqrt{n}\widehat{D}^{-\frac{1}{2}}(\theta)(\widehat{\mu}(\theta) - \mu_M(\theta))$ is convergent in distribution to a normal distribution while $\widehat{\Omega}(\theta)$ is convergent in probability to a correlation matrix. However, since $\sqrt{n}\mu_M$ cannot be consistently estimated, we may want to replace it by some statistics such that our ultimate goal that the test defined by the critical function ϕ_n later using (27) is uniformly valid and the UASC regarding the confidence set of θ_0 w.r.t. the approximated critical value $C_n(\widehat{1-\alpha, \theta}, \widehat{h}_0)$ of the test is achieved, that is,

$$\limsup_{n\to\infty} \sup_{(\theta,h,F)\in\mathcal{F}} E_F[\phi_n] \le \alpha. \tag{9}$$

We will obtain the critical value by bootstrapping the test statistics after the replacement demonstrated by the two-step procedure. As pointing out by Andrews (2000), we cannot simply bootstrap $T_n(\theta)$ to get its critical value and to form the test. This is because the bootstrapping may fail when the true DGP is on the boundary of the DGP space while we need a uniform consistency in levels that is defined in this paper as (9). The UASC is crucial because $C_n(\widehat{1-\alpha, \theta}, \widehat{h}_0)$

is data-dependent and thereby depend on the true DGP so that the objective size need to be correct for the sample size regardless of the DGP in order to have meaningful asymptotic properties. In another words, the limit of the pointwise size will be very misleading since it does not rule out the case, for example, that for any sample size n, in particular for any $n < inf\{N_1, \dots, N_I\}$ where I is the cardinality of the DGP space and N_j is the infimum of the set of the natural numbers of sample size of the test w.r.t. the j-th DGP that makes the size controlled, there is a F such that $E_F[\phi_n] \gg \alpha$. The counter-example can be easily formed using a t-test. For more details regarding the importance of the uniformity, Imbens and Manski (2004) discussed an example of finite sample sample size distortion and Andrews and Guggenberger (2009) discussed the possible discontinuity of the asymptotic distribution of $T_n(\theta)$ under drifting sequence of parameters.

In this paper, we will follow Romano et al. (2014) to adopt a two-step procedure to conduct uniform inference on the partially identified θ_0 through bootstrapping. Andrews and Soares (2010a) considers a method where $\sqrt{n}D^{-\frac{1}{2}}\mu_M$ is replaced by a moment selection function $\varphi(\xi, \Omega)$ defined there to avoid the consistent estimation for $\sqrt{n}\mu_M$ and to find the critical value $C_n(\widehat{1-\alpha}, \theta)$ so that the UASC property is achieved. Similar to Andrews and Soares (2010a), the key replacement applying the method in Romano et al. (2014) for $\sqrt{n}\mu_M$ is thereby the lower-left vertex of the confidence set of μ_M multiplied by root n, denoted as $\sqrt{n}\lambda \in \mathbb{R}^k$. This confidence set about $\sqrt{n}\mu_M$ will be the first step confidence set corresponding to the test $H_{0'}$ defined by (12). Let $C_n(\widehat{1-\alpha}, \theta, \widehat{h}_0)$ be the $1-\alpha$ quantile of the asymptotic distribution of $T_n^{TS}(\theta)$ which denotes the $T_n(\theta)$ after the replacement. The second step confidence set corresponding to the test H_0 will then be that of θ_0 given μ_M defined by (23). The UASC corresponding to the Bonferroni confidence set will then be shown under these two step testings. This non-standard test procedure is also applied in Staiger and Stock (1997) under a weak IV context. Both $T_n(\theta)$ and $T_n^{TS}(\theta)$ depend on $\widehat{h}_0$ but ignored for notational conciseness.

Therefore, the first step is to construct a confidence set of μ_M under $H_{0'}$ and then a confidence set of θ_0 under H_0 in the second step. Without loss of generality, we will use

$$T_n^{min} = \underset{1 \le j \le k}{min} \frac{\sqrt{n}(\mu_{M_j}(F) - \widehat{\mu}_{j,n})}{S_{j,n}} \tag{10}$$

as the first step test-statistic and we could use any type of $T_n(\theta)$ as described above as the second step one. Those statistics will typically have asymptotic distributions that are continuous and strictly increasing, which is easy to establish consistency upon getting the critical values through bootstrapping. The uniform bootstrapping consistency of $T_n(\theta)$ and $C_n(\widehat{1-\alpha, \theta}, \widehat{h}_0)$ will be shown upon the Lemma 9, which essentially applies the Theorem 2.4 of Romano and Shaikh (2012a).

Suppose μ_M is the true value $E_{F_0}[m_j(W_i, \theta_0, h_0)]$ and

$$H_{0'} : \mu_M = \mu \, s.t. \, E_{F_0}[m_j(W_i, \theta_0, h_0)] \geq 0, j = 1 \dots k \tag{11}$$

where $H'_{0'} : \mu_M = \mu \, s.t. \, E_{F_0}[m_j(W_i, \theta_0, h_0)] < 0, j = 1 \dots k$, which enables us to form the test of H_0 conditional on μ_M characterized by (3). The ultimate test of interest defined by (27) will then be constructed along with the test of H_0 and that of $H_{0'}$. The test cannot be rejected when there exists a null of H_0 and $H_{0'}$ is true. The type I error will then be the probability of both H_0 and $H_{0'}$ are rejected simultaneously when there is at least one null is true. Hence, similar to Bonferroni correction for controlling the probability of one or more false rejections not exceed a given level $\alpha \in (0, 1)$ by $\beta \in (0, \alpha)$, we need to introduce a Bonferroni-type correction β here to account for the possible case that μ_M does not lie in $M_n(\beta)$ defined by (12), which may lead to the non-convergence of the first step confidence set as $\mu_M \to \infty$, and to adjust it in the second step. And the test ϕ_n needs to be defined accordingly. In this case, the corrected level β for $H_{0'}$ need to be strictly less than α. Given the Bonferroni-type correction β in the test of $H_{0'}$ and H_0, the probability that the moment inequalities are violated while approximating the nuisance parameters by $\widehat{h_0}$ when the moment inequalities actually hold will be naturally contained in the size of the test of interest. As long as the UASC property holds, this special approximation error will be controlled.

For more generous construction of Bonferroni critical value, see McCloskey (2017). Similar to Andrews and Barwick (2012), β is essentially a tuning parameter under this context and could be found accordingly by maximizing the weighted average power, albeit not the focus of this paper. The naive choice of β is $\frac{\alpha}{10}$ that leads to good power properties as pointed out by Romano et al. (2014). Ideally, the size would be less conservative and the power would be better if β is closer to zero which could be demonstrated by the simulation study in the next section.

The idea of using the lowest λ concerns the least favorable case such that the power is least, that is when $T_n^{TS}(\theta)$ is highest. The first step confidence set is thereby defined by

$$M_n(\beta) = \{\mu \in \mathbb{R}^k : T_n^{min} \geq K_n^{-1}(\beta, \hat{F}_n)\}, \tag{12}$$

where $\beta \in (0, \alpha)$ and the distribution of T_n^{min} is defined by

$$K_n(x, F) = P\left\{ \underset{1 \leq j \leq k}{min} \frac{\sqrt{n}(\mu_{M_j}(F) - \widehat{\mu}_{j,n})}{S_{j,n}} \leq x \right\}. \tag{13}$$

However, in order to find $M_n(\beta)$, we need first find $\widehat{\mu}$ and $\widehat{\Sigma}$. In this paper, h_0 is assumed to be $E(Y|X)$, which allows general ways to estimate it such as kernel or series estimators as discussed in Newey (1994a). Given the consistent nonparametric estimator $\widehat{h} \equiv \widehat{h}_0 = \widehat{E(Y|X)}$, we could find the $\widehat{\mu}$ and $\widehat{\Sigma}$ by a reparameterized GMM condition (15) with the method provided by Newey (1994a). Thereby consider the reparameterized moment functions under H_0

$$g(W, \mu_M, h_0) \equiv m(W, \theta, h_0) - E[m(W, \theta, h_0)] = m(W, \theta, h_0) - \mu_M, \tag{14}$$

where

$$E[g(W, \mu_M, h_0)] = 0 \tag{15}$$

holds trivially. Then

$$\widehat{\mu} = \frac{1}{n}\Sigma_{i=1}^n m(W_i, \theta, \widehat{h}) \tag{16}$$

which minimizes the squared error loss function $\widehat{Q}_n(\mu_M) \equiv [\frac{1}{n}\Sigma_{i=1}^n g(W_i, \mu_M, \widehat{h})]'\widehat{\Gamma}[\frac{1}{n}\Sigma_{i=1}^n g(W_i, \mu_M, \widehat{h})]$, where a weighting matrix $\widehat{\Gamma} \xrightarrow{p} \Gamma$ along with the identification of μ_M are assumed by Assumption 7 in the appendix A. Then by the Theorem 2.1 of Newey (1994a), we could find the influence function of $\widehat{\mu}$ under F by the pathwise derivative $d(W)$ along a path of $F_\tau \in \mathbf{F}$, a.k.a. a one-dimensional subfamily of $\mathbf{F}$, under some regularity conditions (RC) described in the appendix, so that the consistent estimator $\widehat{\Sigma}$ can be found as follows. Let $\| f \|$ denote the $\mathcal{L}^2$ norm for any generic function f.

Lemma 1. *Under RC as displayed in the appendix, if:*

(i) h_0 *is point identified and* $\| \widehat{h} - h_0 \| \xrightarrow{p} 0$,

(ii) $\frac{1}{n}\Sigma_{i=1}^{n} \parallel \widehat{\alpha}(W_i) - \alpha(W_i) \parallel^2 \xrightarrow{p} 0$, *where* $\widehat{\alpha_j}(W_i) = E\big(\widehat{\frac{\partial m_j(W,\theta_0,h)}{\partial h}\big|_{h=h_0(x)}}\Big|X = x_i)(y_i - \widehat{h}_0(x_i))$ *and* $\alpha_j(W_i) = E\big(\frac{\partial m_j(W,\theta_0,h)}{\partial h}\big|_{h=h_0(x)}\Big|X = x_i)(y_i - h_0(x_i))$*,*

(iii) $\parallel g(W,\mu,h) - g(W,\mu_M,h_0) \parallel \leq b(W)(\parallel \mu - \mu_M \parallel + \parallel h - h_0 \parallel)$ *while* $E[b^2(W)] < \infty$*,*

then

$$\widehat{\Sigma} \xrightarrow{p} \Sigma, \tag{17}$$

where

$$\widehat{\Sigma} = \frac{1}{n}\Sigma_{i=1}^{n}\widehat{\iota} \times \widehat{\iota}' \tag{18}$$

$$\widehat{\iota_j} \equiv \widehat{\iota_j}(W_i,\theta_0,\widehat{h_0}(x_i)) = m_j(W_i,\theta_0,\widehat{h}_0(x_i)) - \frac{1}{n}\Sigma_{i=1}^{n} m_j(W_i,\theta_0,\widehat{h}_0(x_i)) + \widehat{\alpha_j}(W_i)$$

and

$$\Sigma = \mathbb{E}(\iota \times \iota') \tag{19}$$

$$\iota_j \equiv \iota_j(W,\theta_0,h_0) = m_j(W,\theta_0,h_0(x)) - \mu_{M_j} + \alpha_j(W_i)$$

Proof. First observe that $E[g(W,\mu_M,h_0)] = 0$ is the GMM moment condition, where g is defined as (14). We fix a θ_0 value at θ under H_0 so that only μ_M and h_0 are the unknown parameters. Then under the point-identification of μ_M and h_0 , the moment function g has randomness coming from $W \sim F$ only while the two parameters μ_M and h_0 will be determined once the DGP F is given. Then by Theorem 2.1 of Newey (1994a), under Assumption 2, there exists a pathwise derivative of μ_M denoted as $d(W)$ that equals to the influence function of $\widehat{\mu}$ denoted as $\psi(W)$ such that

$$\sqrt{n}(\widehat{\mu}_M - \mu_M) = n^{-\frac{1}{2}}\Sigma_{i=1}^{n}\psi(W_i) + op(1). \tag{20}$$

Under RC Assumption 3 and 5 that guarantees the functional $E[g(W,\mu_M,h_\tau)]$ is continuous and linear so that Riesz representation theorem is applicable , by Propositions 1,4 and 5 of Newey (1994a),

$$\psi(W) = -G^{-1}[g(W,\mu_M,h_0) + \alpha(W)], \tag{21}$$

a $k \times 1$ vector, where $G \equiv \frac{\partial E[g(W,\mu,h_0)]}{\partial \mu}|_{\mu=\mu_M}$, $\alpha(W) = \delta(x)(y - h_0(x))$, $\delta(x) = E[\Delta(W)|x]$ and $\Delta(W) = \frac{\partial g(W,\mu_M,h)}{\partial h}|_{h=h_0(x)}$. Hence $G = -I_k$ by dominated convergence theorem given Assumption 1, where I_k is a k-dimensional identity matrix. Therefore, as $\Delta(W) = \frac{\partial}{\partial h} m(W_i, \theta_0, h)\big|_{h=h_0(x)}$,

$$\psi(W) = m(W, \theta_0, h_0) - \mu_M + E\Big(\frac{\partial m(W, \theta_0, h)}{\partial h}\Big|_{h=h_0(x)}\Big|x\Big)(y - h_0(x)), \tag{22}$$

where $E(\psi(W)) = \mathbf{0}$. Hence, the asymptotic variance of $\widehat{\mu}_M$ is

$$\Sigma = var(\psi(W)) = E(\psi^2(W)) = E[\psi(W)\psi'(W)],$$

which is (19). The consistency result of the plug-in estimator $\widehat{\Sigma}$ is a direct consequence of Lemma 5.2 and Lemma 5.4 of Newey (1994a) under Assumption 3-8. □

Remark 1. $E\big(\frac{\partial m_j(W,\theta_0,h)}{\partial h}|_{h=h_0(x)}\big|X = x\big)$, $j = 1, \ldots, k$, is the projection of $\frac{\partial m_j(W,\theta_0,h)}{\partial h}|_{h=h_0(x)}$ on the plane of all square integrable functions of X. $E\big(\widehat{\frac{\partial m_j(W,\theta_0,h)}{\partial h}|_{h=h_0(x)}}\big|X = x\big)$ will be determined using a smooth function, such as kernel function or series function under some additional conditions for the nonparametric estimation. And the key feature of $\widehat{\Sigma}$, comparing to the normal one-step GMM is that we have an adjustment term regarding $\widehat{h}_0(x)$, which is

$$E\Big(\widehat{\frac{\partial m_j(W, \theta_0, h)}{\partial h}\Big|_{h=h_0(x)}}\Big|X = x\Big)(y_i - \widehat{h}_0(x_i)),$$

$j = 1, \ldots, k$, in this paper. This adjustment term characterizes the effect of estimator $\widehat{h}_0$ on the asymptotic distribution of $\widehat{\mu}$ through the channel of the asymptotic normality of $\sqrt{n}E[g(\widehat{W, \mu_M}, h_0)]$, where $E[g(\widehat{W, \mu_M}, h_0)] \equiv \widehat{g}_n(\mu_M) \equiv \frac{1}{n}\Sigma_{i=1}^n m(W_i, \theta, \widehat{h}) - \mu_M$. Through $\widehat{\mu}$, which further affect $\widehat{\Sigma}$, along with $\widehat{\Sigma}$, the confidence sets in two steps, i.e. (12) and (23), will be characterized under bootstrapping. Apparently, if the choice of $\widehat{\mu}$ and $\widehat{\Sigma}$ is problematic, the validity of the confidence set of parameters of interest will fail. If $\alpha(W) = 0$, $\widehat{h}_0$ will have no impact on it.

The estimator of $\alpha(W)$ could be formed by plug-in without knowing the actual function form of $\alpha(W)$ which could be shown immediately follows the proof of Lemma 4 in the Appendix B. The corresponding construction for kernel estimator for $\widehat{h}$ has been discussed in Newey (1994b) and the series estimators has been discussed in Newey (1994a).

Remark 2. For the case of N nuisance functions under

$$E[g(W, \mu_M, h_{10}, \dots, h_{N0})] = 0,$$

the corresponding result in Lemma 1 remains with the following formula for the pathwise derivative.

$$\psi(W) = -G^{-1}[g(W, \mu_M, h_{10}, \dots, h_{N0}) + \alpha(W)],$$

where

$$G \equiv \frac{\partial E[g(W, \mu, h_{10}, \dots, h_{N0})]}{\partial \mu}|_{\mu=\mu_M},$$

$$\alpha(W) = \alpha_1(W) + \cdots + \alpha_N(W),$$

$$\alpha_j(W) = E\big(\frac{\partial g(W, \mu_M, h_1, \dots, h_N)}{\partial h_j}|_{h_j=h_{j0}(x)}\Big|X = x_i)(y_{ji} - h_{j0}(x_i)),$$

and

$$h_{j0}(x) = E(Y_j|X = x)$$

for $j = 1, \dots, N$.

The second confidence set is then defined by

$$CS_n(\alpha - \beta) = \big\{\theta \in \Theta : T_n(\theta, \widehat{h}_0) \leq C_n(1 - \widehat{\alpha + \beta}, \theta, \widehat{h}_0)\big\}, \tag{23}$$

where

$$C_n(1 - \widehat{\alpha + \beta}, \theta, \widehat{h}_0) = \underset{\lambda \in M_n(\beta) \bigcap \mathbb{R}^k_+}{sup} J_n^{-1}(1 - \alpha + \beta, \lambda, \widehat{F_n}),$$

and

$$J_n(x,\lambda,F) = P\Big(S\big(\widehat{D}^{-1/2}(\sqrt{n}(\widehat{\mu}-\mu(F))) + \widehat{D}^{-1/2}\sqrt{n}\lambda, \widehat{\Omega_n}\big) \leq x\Big) \tag{24}$$

represents the distribution of $T_n^{TS}(\theta)$ while $J_n(x,\mu,F) \equiv J_n(x,F)$ represents the distribution of $T_n(\theta)$. Under the null,

$$\lambda_j^* = max(\widehat{\mu_j} + \frac{S_j K_n^{-1}(\beta, \widehat{F_n})}{\sqrt{n}}, 0) \tag{25}$$

and

$$C_n(1\widehat{-\alpha+}\beta, \theta, \widehat{h}_0) = J_n^{-1}(1-\alpha+\beta, \lambda^*, \widehat{F_n}).$$

Then our task will be getting the asymptotic distribution of $T_n(\theta)$ and its approximated critical value $C_n(1\widehat{-\alpha+}\beta,\theta)$ using bootstrapping, ignoring $\widehat{h}$ hereafter, at level of α so that we could discuss the UASC which leads to the following main result. The algorithm of the two-step procedure with nuisance functions will be discussed in the next section.

Theorem 1. *Suppose $W_i, i = 1,\dots n$, is an i.i.d. sequence of random vectors with distribution $F \in \mathbf{F}$ and $\mathbf{F}$ satisfies*

$$\underset{\lambda\to\infty}{lim} \underset{(\theta,h,F)\in\mathcal{F}}{sup} \mathbb{E}_F\Big[\Big(\frac{\psi_j(W)}{\sigma_j(F)}\Big)^2 \mathbb{1}\Big\{\Big|\frac{\psi_j(W)}{\sigma_j(F)}\Big| > \lambda\Big\}\Big] = 0 \tag{26}$$

for all $1 \leq j \leq k$, where $\psi(W)$ is defined by (22). If RC as displayed in the appendix and the assumptions of Lemma 1 satisfied, then the test defined by the critical function

$$\phi_n \equiv \phi_n(\alpha,\beta) = 1 - \mathbb{1}\big\{\{M_n(\beta) \subseteq \mathbb{R}_+^k\} \cup \{T_n(\theta) \leq C_n(1\widehat{-\alpha+}\beta,\theta)\}\big\} \tag{27}$$

of H_0 satisfies the UASC property as defined by (9) for statistics (4), (6) and(7).

Remark 3. The uniform integrability condition requires uniformity over DGP of integral while the normal asymptotic tools such as the weak LLN and Lindeberg-Levy CLT only require for a fixed DGP. This is the only assumption we would need for UASC in comparing to that of the literature. Besides, the UASC in this paper is achieved more general compared to the Andrews and Soares (2010a) in the sense of weaker conditions demanded that, for example, the CLT we apply is Lindeberg CLT while they apply Lyapunov CLT. All other assumptions are regular for the sake of the approximation of the infinite dimensional parameter h_0 and the finite dimensional parameter μ_M. Therefore, the space of DGP $\mathcal{F}$ is of least restrictive compared to the literature.

Remark 4. The test (27) takes into account that the estimator $\widehat{h_0}$ of h_0 could take any value in the neighbour of h_0 including the case $E_{F_0}[m_j(W_i,\theta_0,h_0)] \geq 0, j = 1\ldots k$ would be rejected as long as $\widehat{h_0}$ converges in probability to h_0 under an appropriate rate. That is we could tolerate the violation of the nulls $H_{0'}$ and H_0, i.e. there exists a $j \in 1,\ldots,k$ such that$E_{F_0}[m_j(W_i,\theta_0,h_0)] < 0$, but the frequency of the violation cannot be more than the levels defined in each tests. The size of each individual tests of $H_{0'}$ and H_0 will be shown by the Lemma 7 and Theorem 1. Therefore, the confidence set CS_n of interest allows to some extend the sample noises overall. However, in order to uniformly control the false positives including this, we need to introduce the Bonferroni-type correction for the two-step procedure amounts to the one in Romano et al. (2014) as described in details above.

The proof of this theorem demands the asymptotic distribution of the test statistics and thereby relys on the uniform bootstrapping consistency result motivated by the Theorem 2.4 of Romano and Shaikh (2012a). To save space, we provide the proof in the appendix.

3 Algorithm and Simulation Study

3.1 Algorithm for Partially Identified Parameters of Interest

(1) Find nonparametric estimator $\widehat{h} = \widehat{E(Y|X)}$ and $E\big(\frac{\partial m_j\widehat{(W,\theta_0,h_0(x))}}{\partial h_0(x)}|x\big)\,\forall 1 \leq j \leq k$ by a kernel function or series function using the original sample;

(2) Find $\widehat{\mu}$ by (16) and $\widehat{\Sigma}$ by (18);

(3) Find T_n w.r.t. H_0 and the bootstrap critical value $K_n^{-1}(\beta,\hat{F}_n)$ derived by (13);

(4) Use $\sqrt{n}\lambda_j^*$ by (25) to replace $\sqrt{n}\mu_M$ in (8);

(5) Find the bootstrap critical value $C_n\widehat{(1-\alpha+\beta,\theta,\widehat{h}_0)} = J_n^{-1}(1-\alpha+\beta,\lambda^*,\widehat{F_n})$ derived by (24);

(6) Find the set of values of θ as defined by the confidence set (23).

During the process, we implicitly utilize the first step confidence set $M_n(\beta)$ through its lower left cortex value which will be regarded as the value of λ_n to replace μ_M as described by (25) so as to find the critical value $C_n(1-\widehat{\alpha+\beta,\theta},\widehat{h}_0)$ of second step. One can fix $\beta = 0.005$ w.r.t. $\alpha = 0.05$ for simplicity. The confidence set (23) can be obtained by grid search or some other algorithms subject to the moment inequality model therein. If applying the grid search, the step (6) will be repeating step (1) to (5) for all θ values in Θ to find the confidence set of interest.

The bootstrapping statistics is defined as such as followed for $T_n^{min} = \underset{1\leq j\leq k}{min} \frac{\sqrt{n}(\mu_{M_j}(F)-\widehat{\mu}_j)}{S_j}$ along with $T_n = \underset{1\leq j\leq k}{max} - \frac{\sqrt{n}\widehat{\mu_j}}{S_j}$ for the two hypothesis tests as defined by $H_{0'}$ and H_0 correspondently with $R = (R_1, R_2)$ repetitions for two separate bootstrapping resampling procedures. As shown in Lemma 7,

$$\underset{n\to\infty}{liminf} \underset{F\in\mathbf{F}}{inf} P\{J_n^{-1}(\alpha_1, \widehat{F_n}) \leq T_n^{min} \leq J_n^{-1}(1-\alpha_2, \widehat{F_n})\} \geq 1-\alpha_1-\alpha_2$$

which enables us to construct the critical value of the distribution of T_n^{min} behaving uniformly over $\mathbf{F}$. Therefore, the bootstrapping statistics for $K_n^{-1}(\beta, \hat{F}_n)$ under R_1 repetitions will be

$$\underset{1\leq j\leq k}{min} \frac{\sqrt{n}(\widehat{\mu_j} - \widehat{\mu^*}_j)}{S_j^*} \tag{28}$$

where $\widehat{\mu_j^*}$ and S_j^* denote the bootstrap sample mean and the bootstrap sample variance respectively generated by the nonparametric i.i.d. bootstrapping samples $\{W_{i,r}^* : i = 1,\ldots,n\}$ for $r = 1,\ldots,R_1$, each with sample size n, which are resampled from the original sample. Then by Lemma 2, $\frac{\sqrt{n}(\widehat{\mu_j}-\mu_{M_j}(F_n))}{\sigma_j(F_n)} \xrightarrow{d} \mathcal{N}(0,1)$, $\forall j$. Thus, under the Theorem 1, the bootstrapping statistics for $C_n(1-\widehat{\alpha+\beta,\theta},\widehat{h}_0)$ under R_2 repetitions will be

$$\underset{1\leq j\leq k}{max} \Big\{ \frac{\sqrt{n}(\widehat{\mu_j} - \widehat{\mu_j^*})}{S_j^*} - \frac{\sqrt{n}\lambda_j^*}{S_j^*} \Big\} \tag{29}$$

where $\widehat{\mu_j^*}$ and S_j^* are defined same as above.

3.2 Simulation Study

This simulation example is based on Andrews and Soares (2010b). Consider an interval outcome regression model based on a partial linear model

$$Y = \theta X + h(Z) + U$$

under the assumption of zero conditional mean

$$E(U|X,Z) = 0.$$

Then

$$E[X^2(Y - \theta X - h(Z)] = 0$$

and

$$E[\theta X + h(Z) - Y] = 0.$$

Random variable X and Z are observable, however Y is not. Therefore, if there is no further assumptions or conditions, θ is not identified. For unobservable Y, we could instead observe an interval $Y_l \leq Y \leq Y_u$, where Y_l is the integer part of Y and Y_u is the smallest integer greater or equal to Y. Thus,

$$E(X^2 Y_l) \leq E(X^2 Y) \leq E(X^2 Y_u).$$

Then we could form the moment inequalities

$$\begin{aligned} E[X^2(Y_u - \theta X - h(Z))] &\geq 0 \\ E[\theta X + h(Z) - Y_l] &\geq 0 \end{aligned} \tag{30}$$

where $m_1 \equiv X^2(Y_u - \theta X - h(Z))$ and $m_2 \equiv \theta X + h(Z) - Y_l$. The data available to econometricians are $\{X, Z, Y_l, Y_u\}$.

In this study, we specify that

$$h(Z) = E(X|Z),$$

$(X,Z) \sim \mathcal{N}\Big(\Big(\begin{array}{c} 1 \\ 1 \end{array}\Big), \Big(\begin{array}{cc} 1 & 0.2 \\ 0.2 & 1 \end{array}\Big)\Big)$, $U \sim \mathcal{N}(0,1)$ and $\theta = 1$. Then $h(Z) = 0.2Z + 0.8$. We will apply $T_n^{min} = \min_{1 \le j \le k} \frac{\sqrt{n}(\mu_{M_j}(F) - \widehat{\mu}_j)}{S_j}$ along with $T_n = \max_{1 \le j \le k} - \frac{\sqrt{n}\widehat{\mu_j}}{S_j}$ to find the finite sample coverage probability (CP) of the model above at level $\alpha = 0.05$. And we fix $\beta = 0.005$ for simplicity. The CP is defined as the relative frequency of coverage of the true parameter θ. We will report the CP for sample size $n = 100$, 500 and 2000.

The plan of the simulation includes three parts: (a) find the identified set of θ using an independent sample of size $N = 1,000,000$; (b) generate a new sample and follow the step (1)-(5) of the algorithm described in the last section to determine the binary choice $B \equiv \mathbb{1}\{T_n \le C_n(1 - \widehat{\alpha + \beta}, \theta, \widehat{h}_0)\}$ while $R_1 = R_2 = 1000$; (c) repeat (b) $R' = 20,000$ times to get a collection of value of binary choices $\{B_i\}_{i=1}^{R'}$ and calculate the $CP = \frac{\Sigma_{i=1}^{R'} B_i}{R'}$.

What follows describes the concrete procedure.

(1) Simulate a sample $\{X_i, Z_i, U_i\}_{i=1,\ldots,N}$ from the DGP specified and calculate a set of value of $h(Z_i) = 0.2Z_i + 0.8$;

(2) Find a set of value of $Y_i = X_i + 0.2Z_i + 0.8 + U_i$;

(3) Get a set of value of (Y_{l_i}, Y_{u_i});

(4) Use the set of value of $(X_i, Z_i, Y_{l_i}, Y_{u_i})$ already had along with the moment inequalities (30) to get the identified set of θ, whose element is denoted as θ', by numerical calculation formed by sample moment inequalities

$$\frac{1}{N}\Sigma_{i=1}^{N}[X_i^2 Y_{u_i} - \theta X_i^3 - X_i^2(0.2Z_i + 0.8)] \ge 0$$

and

$$\frac{1}{N}\Sigma_{i=1}^{N}(\theta X_i + 0.2Z_i + 0.8 - Y_{l_i}) \ge 0;$$

(5) Generate another sample of size n following steps (1)-(3) with N replaced by n so that we could have another set of value of $(X_i, Z_i, Y_{l_i}, Y_{u_i})$ and find the series estimator $\widehat{h(z)}$ with intercept using cross-validation criteria GCV or Mallow's C_p to select the series term;

(6) Consider $\theta = \theta'$, then find the sample moments

$$\widehat{\mu_1} = \frac{1}{n}\Sigma_{i=1}^{n}[X_i^2 Y_{u_i} - \theta X_i^3 - X_i^2 \widehat{h(Z_i)}]$$

and

$$\widehat{\mu_2} = \frac{1}{n}\Sigma_{i=1}^{n}(\theta X_i + \widehat{h(Z_i)} - Y_{l_i});$$

(7) Find the estimator of asymptotic standard deviation by equation (18)

$$S_1 = \frac{1}{n}\Sigma_{i=1}^{n}[X_i^2 Y_{u_i} - \theta X_i^3 - X_i^2 \widehat{h(Z_i)} - \widehat{\mu_1} - (\frac{1}{n}\Sigma_{i=1}^{n} X_i^2)(X_i - \widehat{h(Z_i)})]^2$$

and

$$S_2 = \frac{1}{n}\Sigma_{i=1}^{n}[(\theta + 1)X_i - Y_{l_i} - \widehat{\mu_2}]^2;$$

(8) Calculate the T_n;

(9) Use the same sample in step (5) to bootstrap R_1 T_n^{min} each with sample size n by formula (28) with their corresponding estimator $\widehat{h(Z)}$;

(10) Find the 0.005 quantile $K_n^{-1}(\beta, \hat{F}_n)$ as defined according to equation (13);

(11) Find the value of

$$\lambda_j^* = max(\widehat{\mu_j} + \frac{S_j K_n^{-1}(\beta, \widehat{F_n})}{\sqrt{n}}, 0)$$

for $j = 1, 2$;

(12) Use the same sample in step (5) again to bootstrap R_2 T_n^{TS} each with sample size n by formula (29) with λ_j^* just found with their corresponding estimator $\widehat{h(Z)}$;

(13) Find the $1 - \alpha + \beta = 0.955$ quantile $C_n(\widehat{1 - \alpha} + \beta)$ of T_n^{TS} ;

(14) Find the value of B;

(15) Repeat (5)-(14) R' times to get the CP.

The identified set found in this case is $[0.4998, 1.2486]$. We will then simulate for θ' taking values of 0.4498, 0.4998, 1, 1.2486 and 1.3735, which include the cases of the value of θ outside the identified set, on the boundary of the identified set and the interior of the identified set. We should anticipate that the CP will be lower and lower for θ' outside the identified set and far from the boundary of the identified set while all those CPs will be less than 0.95. For all the CPs with respect to the θ' inside the identified set, it will be bigger or equal to 0.95. Particularly speaking, for all θ' taking interior values of the identified set, the CPs will tend to 1. And following the UASC property we derive, the CPs for θ' on the boundary of the identified set will tend to 0.95 as sample size increases. The following table displays the simulation results.

Table 1: Finite Sample Coverage Probabilities of Level 0.05 Confidence Set for θ with $\beta = 0.005$

	Coverage Probabilities for θ Values				
θ	0.4498	0.4998	1.0000	1.2486	1.3735
Position	Out	On	In	On	Out
$n = 100$	0.9199	0.9698	1	0.9325	0.6675
$n = 500$	0.7575	0.9539	1	0.9517	0.1952
$n = 1000$	0.6261	0.9585	1	0.9578	0.0274

According to the results, all the anticipations have been verified, which is consistent with the UASC result we have shown in the Theorem 1. We also conduct the simulation for $\theta = 0.4498$ when $n = 6000$, the coverage probability is 0.0438. Thus, although the finite sample power might not be big enough, the asymptotic power will tend to 1.

In order to motivate the idea of choosing β close to zero, we implement another simulation for $\beta = 0.0001$ while keeping the same random number generator. In addition to all the results discussed above hold, when β is chosen a smaller one, the performance is even better. In particular, the CPs for θ' on the boundary of the identified set become closer to 0.95 and the CPs for θ' outside the identified set will be closer to zero. The following table displays the simulation results. Based on the results, for small β, the test would not be overly conservative.

Table 2: Finite Sample Coverage Probabilities of Level 0.05 Confidence Set for θ with $\beta = 0.0001$

	Coverage Probabilities for θ Values				
θ	0.4498	0.4998	1.0000	1.2486	1.3735
Position	Out	On	In	On	Out
$n = 100$	0.9118	0.9665	1	0.9503	0.7543
$n = 500$	0.7433	0.9485	1	0.9484	0.1853
$n = 1000$	0.6089	0.9537	1	0.9532	0.0254

4 Empirical Study

When there are parameters of interest involved in strategic behaviors of economic agents, one can find the economic equilibrium upon the information structure with a specified solution concept and adopt this economic equilibrium to form a prediction model in terms of Conditional Choice Probability (CCP). And the econometric model in terms of moment conditions can then be established via appropriate transformation of the prediction model so that one can apply certain method to infer the parameters of interest.

A typical general decision making process is that the decision makers first form their updated beliefs about opponents' private signals while there are asymmetric information of private signals among all players and then make the corresponding decisions which leads to certain outcomes. An usual solution concept is Bayesian Nash Equilibrium (BNE). The outside observer econometricians usually have asymmetric information from all the players in terms of the private signals and the parameters in the payoff functions of players, etc. The econometricians could then regard the private signals as random shocks to build up the relationship between the CCP representing the probability of actual choices made by players and the predicted choice probability (PCP) measuring the events of the random shocks that could make any actual choices as one of the equilibrium choices as the prediction model. Then either the conditional or unconditional moment conditions on the basis of the prediction model can be found. Based on the information structure and the reasonable assumptions made, the parameters of interest could then be point identified or partially identified. Under this procedure, the function form of CCP is usually unknown to econometrician and it will be regarded as a nuisance function. In addition, the nuisance functions can also occur to econometricians when the payoff function as a function of covariates is unknown to econometricians even if known to all players. Therefore, addressing with the nuisance functions while conducting inference on the parameters of interest is of a general interest in the empirical work.

4.1 Structure Model

In this section, we illustrate the two-step method with nuisance function (NTS) by a static interaction game with incomplete information, motivated by De Paula and Tang (2012) which addresses with the point identification of the sign of the coordination effect. In our application, we are instead interested in the actual values of the coordination effect, though partly identified.

There are N players each player i with a choice set of two elements $\mathcal{D}_i \equiv \{0, 1\}$. The payoff functions of player i are

$$\begin{aligned} \pi_{1i} &= \beta_i' X + \delta_i' X \sum_{j \neq i} D_j - \epsilon_i, \quad \text{if } D_i = 1 \\ \pi_{0i} &= 0, \quad \text{if } D_i = 0 \end{aligned} \tag{31}$$

where $\epsilon_i | X \sim \mathcal{U}[0, 1]$ and $i \in \mathcal{I} \equiv \{1, \dots, N\}$. The information structure is supposed as follows: $\beta_i' X$ and $\delta_i' X$ is common knowledge among all players, player i privately observe a signal ϵ_i, whose distribution is a common prior among all players, player i will make his own choice D_i but does not know what the opponents $(-i)$ will choose; we outside observer econometrician has access to the covariates X and the strategy profiles taken by the players, i.e. $D \equiv \{D_i\}_{i \in \mathcal{I}}$, besides we also know $\epsilon_i | X \sim \mathcal{U}[0, 1]$ but does not know any realized value of ϵ_i. The dataset thereby consists of G cross-sectional games each $g \in \mathcal{G} \equiv \{1, \dots, G\}$ with their corresponding values of X and D. We assume rationality of all players throughout the games. We are interested in learning the parameters $\theta_0 \equiv \{\beta_i, \delta_i\}_{i \in \mathcal{I}}$. However, like most of the empirical datasets which usually have limited variations in covariates and thereby the standard assumptions for identification such as the rich support proposed in Tamer (2003) and Grieco (2014) is hard to be satisfied in reality, we focus on the partial identification of parameters of interest while the NTS will work.

The solution concept adopted is pure strategy Bayesian Nash Equilibrium (BNE), which is defined by the correspondence mapping a profile of private signals to a set of strategy profiles such that for any player i with signal of player i and any signal-dependent strategy played by the opponents, the expected payoff of player i conditional on his signal is maximized, where the payoff function maps any bundle of a strategy profile and a profile of signals to a real number. We will have multiple equilibriums if the set is non-singleton. The pure-strategy BNE exists for any given X, which can be proved by the Theorem 1 of Athey (2001) applying Kakutani's fixed point theorem where the key is that the payoff functions satisfy the single-crossing condition.

Therefore, conditional on X, the equilibrium condition for player i is

$$D_i(X,\epsilon_i) = \begin{cases} 1, & \text{if } \beta_i' X + \delta_i' X \sum_{j\neq i} E[D_j(X,\epsilon_j)|X,\epsilon_i] - \epsilon_i \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

We assume

$$F_{\epsilon|X}(\cdot|x) = \Pi_{i\leq N} F_{\epsilon_i|X}(\cdot|x) \tag{32}$$

for any $x \in \Omega_X$ and $F_{\epsilon_i|X}(\cdot|x)$ is continuous for all i. Then

$$E[D_j(X,\epsilon_j)|X=x,\epsilon_i] = E[D_j(X,\epsilon_j)|X=x] \equiv p_j(x).$$

This implies a cutoff rule $\kappa(x)$ for best responses to all players and player i will play 1 if

$$\epsilon_i \leq \kappa_i(x) \equiv \beta_i' x + \delta_i' x \sum_{j\neq i} p_j(x)$$

while the function form of $p_j(\cdot)$ is fixed at equilibrium. At the equilibrium threshold $\kappa_i^*(x) \equiv \epsilon_i^*(x)$, player i is indifferent from playing 0 and 1, that is

$$\epsilon_i^*(x) = \beta_i' x + \delta_i' x \sum_{j\neq i} p_j^*(x) = \beta_i' x + \delta_i' x \sum_{j\neq i} \epsilon_j^*(x).$$

Then

$$\epsilon_i^*(x) = \frac{a_i(x) + b_i(x) * \frac{A(x)}{1-B(x)}}{1+b_i(x)}, \tag{33}$$

where $a_i(x) = \beta_i' x$, $b_i(x) = \delta_i' x$, $A(x) = \sum_{i=1}^N \frac{a_i(x)}{1+b_i(x)}$ and $B(x) = \sum_{i=1}^N \frac{b_i(x)}{1+b_i(x)}$, for all i, which defines the unique BNE, i.e. player i will play 1 if observe $\epsilon_i \leq \epsilon_i^*(x)$ and play 0 otherwise.

Note that, in addition to $\kappa(x) \equiv (\kappa_1(x), \ldots, \kappa_N(x))'$, the BNE can also be expressed in terms of conditional choice probability (CCP) , i.e. $p(x) \equiv (p_1(x), \ldots, p_N(x))$' where for any $x \in \Omega_X$, for every i, $p_i(x) = F_{\epsilon_i|X=x}\big(\beta_i' x + \delta_i' x \sum_{j\neq i} p_j(x)\big)$, which implies a conditional independent product of binominals distribution over the strategy profiles. The mapping between $\kappa(x)$ and $p(x)$ is bijective conditional on X when $F_{\epsilon_i|X}(\cdot|x)$ is strictly increasing. Thus, the BNE is equivalently characterized by

$$p_i^*(x) = F_{\epsilon_i|X}(\epsilon_i^*(x)) = E[\mathbf{1}\{\epsilon_i \leq \epsilon_i^*(x)\}] = \epsilon_i^*(x) \tag{34}$$

for any i.

The prediction model applying the BNE is thereby

$$H(\theta, X) \leq p(X) \leq H(\theta, X), \tag{35}$$

where $H(\theta, X) = [H_1(\theta, X), \ldots, H_N(\theta, X)]'$, $H_i(\theta, X) = E[\mathbf{1}\{\epsilon_i \leq \epsilon_i^*(x, \theta)\}]$ for any x almost surely and $p_i(X) = E[D_i|X]$ for any i . The model is apparently coherent and complete due to the uniqueness of BNE.

There is no equilibrium selection mechanism (ESM) involved due to the uniqueness of BNE. One may change the primitive of distribution of $\epsilon|X$ to other continuous distribution which may results in multiple equilibriums and ESM will then naturally enter the model. But that is just an simple conceptually extension with more computation burden and is not the focus of this application. The usual ways to deal with the multiplicity of equilibriums have been discussed in Ciliberto and Tamer (2009), Beresteanu et al. (2011) and Grieco (2014). The first way is extending the prediction model (35) to the inequality of CCP for the counterfactual equilibrium strategy profiles while the lower bound is the probability that the corresponding counterfactual equilibrium is the unique equilibrium and the upper bound is the probability that the corresponding counterfactual equilibrium is one of equilibriums. The ESM is involved but we can grid search θ without find it while there are least restrictions imposed on the ESM. The second way is extending (35) to an equality $p(X) = H(\theta, X, \varrho)$ where ϱ is the ESM. ϱ can be found along with θ and will be unique if the matrix of the predicted outcome probabilities given the equilibriums is invertible. The key in grid search of θ is then to verify whether ϱ exists for all X. Besides, the identification and estimation of ESM has been discussed in Sweeting (2009) and Bajari et al. (2010) under parametrization of ESM.

The econometric model or the moment inequality model is then

$$E[H(\theta, X)] \leq h \leq E[H(\theta, X)], \tag{36}$$

which we can typically apply the NTS to get the confidence set of θ as defined by (2) to partially identify θ with the nuisance function $h \equiv p(X)$ therein. We take a necessary condition of (35) for a collection of observation of X from different games rather than a particular observation of X , though it could lead to a bigger confidence set of θ if applying (36). One could further consider conditional moment inequality model as discussed by Andrews and Shi (2013) and Chernozhukov et al. (2013) to avoid this problem but it is not the focus of this study. Ciliberto and Tamer (2009) has a similar structure model as the form of (36), however, they use a simple frequency estimator to approximate the nuisance function h without fully consider the approximation error that would have significant impact on the confidence set of interest, through the channel of the estimator of the asymptotic variance of the sample mean of the corresponding moment functions as discussed in section 2.

In the above model, $\beta_i' X$ is regarded as the basic payoff at state X for player i. $\delta_i' X \sum_{j \neq i} D_j$ captures the homogeneous state-dependent interaction effect at state X upon what the opponents do to player i and explains how the market size will affect the payoff of a player. One can extend to the case with heterogeneous state-dependent interaction effect using our method but it is not the focus of this paper. $\epsilon \equiv \{\epsilon_i\}_{i \in \mathcal{I}}$ as private signals are the only random shocks to econometrician. One can extend to the case when public shocks known to all players but unknown to econometrician play a role with our method if the conditional distribution is given. Besides, one can also extend to the case where the signals are correlated or follow a conditional joint distribution known up to finite dimensional parameters. Our method can also be applied to other type of games with different solution concepts, such as entry games under complete information discussed in Ciliberto and Tamer (2009) and voting communication games under incomplete information discussed in Iaryczower et al. (2018).

Regarding Mixed Strategy Nash Equilibriums (MSNE) as discussed in Berry and Tamer (2006) and Beresteanu et al. (2011), one need to further consider the randomization among pure strategies mixed in MSNE, which is again a simple conceptually extension with more computation burden and is not the focus of this study. One could also consider model selection between two different moment inequality models implied by different information structure postulated by a hypothesis testing procedure as discussed in Shi (2015), albeit it is not the focus of this paper. A general framework of forming moment inequality models under certain conditions or assumptions within a microeconomic theory structure has been discussed in Pakes et al. (2015) and a general discussion of applying moment inequality models in different microeconomic context can be found in Molinari (2020).

The key feature of this study compared to Sweeting (2009) and De Paula and Tang (2012) is that the ESM does not get involved at all so that the potential incoherency and incompleteness issue that the sum of all counterfactual CCP is beyond 1 and the conditional prediction is nonunique are avoided and the potential inconsistency issue that the actual observed CCP, a vector of mixture marginal probabilities, is not a BNE is avoided. In addition, the prediction bases on the equilibrium solution rather than equilibrium conditions that commonly preceded in the literature. The rationale is that under our model the unique BNE is achieved and for any observed choice data, players are supposed to have made their best response by directly choosing one pure strategy out of their choice set upon their private signals. And the variation of the dataset solely contribute to the partial identification such that we will have a set-estimator.

4.2 Moment Inequality Model

In this section, we discuss the concrete form of the moment inequality model derived by (36) and the empirical result.

We use the dataset used in De Paula and Tang (2012) to investigate the coordination effect represented by $\{\delta_i\}_{i=1,\ldots,N}$ of three radio stations, i.e. $N = 3$, who make one-shot decisions about whether entering the commercial break markets or not among their programming schedule denoted by D_i. The player i choose to enter the corresponding market if $D_i = 1$. Their decisions are state (X) dependent and private signal (ϵ_i) dependent. The state variable X is the market size of each commercial break markets taking values of whole numbers range from 1 to 3. The total number of cross-sectional games is 26,152. The dataset comes from BIAfn's MediaAccess Pro data base, Media base 24/7 and the 2001 Census. The detailed description of the dataset can be found in Sweeting (2009). We assume the payoff functions π_{1i} and π_{0i} take the form as (31), $\epsilon_i|X \sim \mathcal{U}[0,1]$, conditional independence of ϵ_i for all i as (32), the information structure described above and rationality of all players. We formulate the prediction model based on the BNE (34). Assume all the assumptions in NTS satisfied, the moment inequality model is as followed for $i \in \mathcal{I} \equiv \{1, 2, 3\}$.

$$E[h_i - E(\mathbf{1}\{\epsilon_i \leq \epsilon_i^*(x, \theta)\})] \geq 0$$
$$E[E(\mathbf{1}\{\epsilon_i \leq \epsilon_i^*(x, \theta)\}) - h_i] \geq 0$$

$E(\mathbf{1}\{\epsilon_i \leq \epsilon_i^*(x, \theta)\})$ has a function form (33) as derived. In this model, there are six moment functions while two for each players. The parameters of interest are $\{\beta_i, \delta_i\}_{i \in \mathcal{I}}$ consisting of 6 parameters. In addition, there are three nuisance parameters $h_i = E(D_i|X)$ which can be estimated by series functions with intercept using GCV criteria to select the series term.

For simplicity, we normalize the parameters by $\beta_i = 1$ since only the ratio of β_i and δ_i matters in the case of binary outcome model. Besides, we assume homogeneity in δ_i for this noncooperative game. We then find the confidence set of $\delta_i \equiv \delta$ among $[-100, 100]$ with three decimal places as the parameter space by grid search. And the confidence set of δ is

$$\{[-1.539, -1.412], [0.206, 0.208], [0.412, 0.416]\}.$$

The corresponding state-dependent interaction effect will be $\delta' X$ where X takes positive values only. The result is sharp in terms of the bounds of the subintervals that differ small, which provides sharp and useful values of the interaction effect of the radio stations trying to manage their commercial breaks. The result supports the findings of positive point identified coordination effect provided by Sweeting (2009) and De Paula and Tang (2012). However, it also provides the possibilities of negative interaction effects. The difference from the literature could be the consequence of multiple equilibrium allowed by the literature and or the weakness of certain assumptions in the literature, which are subject to further investigations. Another feature of the confidence set of interest is the nonconvexity, which again motivates our method that applies to nonconvex identified set.

5 Conclusion

In this paper, we provide a general framework to conduct inference on finite-dimensional partially identified parameters of interest in finite number of moment inequalities with point identified nuisance functions, under least restrictive assumptions including uniform integrability and some regularity conditions. We show that our test to form the confidence set of interest achieves the UASC through uniform bootstrapping consistency. Besides, we provide a structure model in game theory to illustrate the importance of our method in terms of CCP that is a nuisance function but has not been fully addressed with the effect of the approximation error of this nuisance function on the confidence set of interest. The theory could be widely applied in industrial organization and along with the microeconomics, especially when we have belief systems or other nuisance functions unknown to econometrician but need to estimate it.

The core way we characterize the effect of the estimator of nuisance functions on the confidence set of interest is by a reparameterized GMM condition that transfers the moment inequality model into a moment equality model that allows us to apply the tools from the semiparametric analysis. The asymptotic variance of the sample mean of the moment functions is thereby derived. Meanwhile, the probability that the estimator of the nuisance parameters violates the moment inequalities when the moment inequalities hold is naturally handled by the two-step procedure.

The main caveat, however, is the CAN property of the nuisance parameters is needed to hold that subject to further primitive conditions for the specific choice of approximation methods, albeit we assume it is point identified. And although our method is computationally feasible to some extent, the further reduction of computation burden will an interesting topic in the future. The general partial identification of conditional moment inequalities is yet also a open question to be discussed.

APPENDIX

Section A provides the Regularity Condition (RC) assumptions, section B provides the auxiliary lemmas and section C delivers the proof of the Theorem 1.

A RC assumptions

Assumption 1. $\mu_M < \infty$.

This technical assumption guarantees the exchange of orders of differentiation and integration.

Assumption 2. *(Pathwise Derivative) (i) the set of scores for regular paths is linear; (ii) for any $\epsilon > 0$ and measurable s(W) with $E[s(W)] = 0$ and $E[s(W)^2) < \infty$, there is a regular path with score $S(W)$ satisfying $E[|s(W) - S(W)|^2] < \epsilon$; (iii) $\widehat{\mu}_M$ is asymptotically linear and regular.*

This assumption guarantees the existence of the pathwise derivative of interest with the crucial requirement for asymptotic equivalence of (20). The definition of regular paths and regular asymptotic linear (RAL) estimators could be found as in Newey (1994a). What follows are the sufficient conditions for the CAN property of the estimator of μ_M or equivalently (20) in this paper.

Assumption 3. *(Linearization) (i) There is a function $\Delta(W, h)$ that is linear in h such that for all h with $\| h - h_0 \| < \epsilon$ for any $\epsilon > 0$, $\| g(W, \mu_M, h) - g(W, \mu_M, h_0) - \Delta(W, h - h_0) \| \leq b(W) \| h - h_0 \|^2$ and $E(b(W))\sqrt{n} \| \widehat{h} - h_0 \|^2 \xrightarrow{p} 0$.*

This assumption guarantees the remainder term from linearization while applying Riesz representation theorem is small and the convergence rate of $\widehat{h}$ is faster than $n^{-\frac{1}{4}}$ when the moment function g is not linear.

Assumption 4. *(Stochastic Equicontinuity) $n^{-\frac{1}{2}} \Sigma_{i=1}^{n} [(\Delta(W_i, \widehat{h} - h_0) - \int \Delta(W, \widehat{h} - h_0) dF_0] \xrightarrow{p} 0$.*

This Assumption in terms of the above empirical process, not necessarily in the form of sample average, is the key assumption to be applied of the generic uniform law of large number as discussed in Newey (1991) and Andrews (1992). One typical sufficiency of stochastic equicontinuity is in terms of Lipschitz condition. The asymptotic normality of semiparametric estimators, or $\widehat{\mu}_M$ in this case, derived by stochastic equicontinuity has been discussed in Andrews (1994). This assumption along with the last one are requirements on second order terms which essentially requires the moment function $g(W, \mu_M, h)$ is sufficiently smooth and the estimator $\widehat{h}$ is well-behaved.

Assumption 5. *(Mean-square Continuity)(i) There is a function* $\alpha(W)$ *and a measure* $\hat{F}$ *such that* $E(\alpha(W)) = 0$*,* $E[\| \alpha(W) \|^2] < \infty$*, and for all* $\| \widehat{h} - h_0 \| < \epsilon$ *for any* $\epsilon > 0$*,* $\int \Delta(W, \widehat{h} - h_0) dF_0 = \int \alpha(W) d\hat{F}$*; (ii) For the empirical distribution* $\tilde{F}(w) = \frac{1}{n}\Sigma_{i=1}^n \mathbb{1}\{W_i \leq w\}$, $\sqrt{n}[\int \alpha(W) d\hat{F} - \int \alpha(W) d\tilde{F} \xrightarrow{p} 0.$

This smoothness type of assumption, which has be discussed more detailed in Newey (1990), is crucial to find the unique adjustment term $\alpha(W)$ and to derive the asymptotic normality of $n^{-\frac{1}{2}}\Sigma_{i=1}^n g(W_i, \mu_M, \widehat{h})$ given the consistency of $\widehat{h}$, which might converge in a different rate. It regards $\int \alpha(W) dF(W, h)$ that will be differentiable in h if $dF(W, h)^{1/2}$ has mean square a mean square derivative, which leads to the differentiability of $E[g(W, \mu_M, h_\tau)]$ in τ of a path $\{F_\tau\}$ in $\mathbf{F}$. It along with the linearization assumption makes the Riesz representation theorem applicable for the derivation of the influence function $\psi(W)$. The assumption 3-5 together guarantees the proper approximation of $\sqrt{n}\widehat{g}_n(\mu_M)$ by $\Sigma_{i=1}^n [g(W_i, \mu_M, h_0) + \alpha(W_i)]/\sqrt{n} + op(1)$ that leads to the asymptotic normality of $\sqrt{n}\widehat{g}_n(\mu_M)$ given the consistency of $\widehat{h}$ under a proper rate of convergence.

Assumption 6. *There are* $\epsilon, \| h \|, b(W), \tilde{b}(W) > 0$ *such that (i) for all* $\mu \in \mathcal{M}$*,* $g(W, \mu, h_0)$ *is continuous at* μ *with probability one,* $\| g(W, \mu, h_0) \| \leq b(W)$*; (ii)*$\| g(W, \mu, h) - g(W, \mu, h_0) \| \leq \tilde{b}(W)(\| h - h_0 \|)^\epsilon$.

This assumption provides an uniform Lipschitz condition of order ϵ for the uniform convergence of $\| \hat{g}_n(\mu) - E[g(W, \mu, h_0)] \|$ over $\mu \in \mathcal{M}$, i.e. $\underset{\mu \in \mathcal{M}}{sup} \| \hat{g}_n(\mu) - E[g(W, \mu, h_0)] \| \xrightarrow{p} 0$ similar to Andrews (1992).

Assumption 7. *(i) there exists an estimator* $\hat{\Gamma}$ *of a weighting matrix* Γ *such that* $\hat{\Gamma} \xrightarrow{p} \Gamma$*, where* Γ *is positive semi-definite; (ii)* $\Gamma E[g(W, \mu, h_0)] = 0$ *has a unique solution on compact* $\mathcal{M}$ *at* μ_M*.*

This assumption assures the identification of μ_M which then could deliver the consistency of $\widehat{\mu}$ combined with the last assumption.

Assumption 8. *(i) $\mu \in interior(\mathcal{M})$; (ii) there is $\parallel h \parallel, \epsilon > 0$ and a neighborhood $\mathcal{N}$ of μ_M such that for all $\parallel h - h_0 \parallel < \epsilon$, $g(W, \mu, h)$ is differentiable in μ on $\mathcal{N}$; (iii) $G'\Gamma G$ is nonsingular for $G \equiv E\big[\frac{\partial g(W,\mu,h_0)}{\partial \mu}\big|_{\mu=\mu_M}\big]$; (iv) $E[\parallel g(W, \mu_M, h_0) \parallel^2] < \infty$; (v) Assumption 6 is satisfied with $g(W, \mu, h)$ there equal to each row of $\partial g(W, \mu, h)/\partial \mu$.*

This assumption provides an uniform Lipschitz condition of order ϵ as Assumption 6 to deliver the consistency of the Jacobian $\frac{1}{n}\Sigma_{i=1}^n \nabla_\mu g(W_i, \tilde{\mu}, \widehat{h}) \xrightarrow{p} E[\nabla_\mu g(W, \mu_M, h_0)]$, which leads to the asymptotic normality of $\widehat{\mu_M}$ and the consistency of $\widehat{\Sigma}$.

Remark 5. The RC Assumption 3-8 together could deliver the asymptotic normality of $\widehat{\mu_M}$ by mean-value expansion as discussed in Newey and McFadden (1994).

B Auxiliary Results

Lemma 2. *Let W_i, $i = 1, \ldots, n$, be an i.i.d. sequence of random vectors with distribution $F_n \in \mathbf{F}$ on $\mathbb{R}^p$. Suppose $\mathbf{F}$ is such that for all $1 \le j \le k$, $\underset{\lambda\to\infty}{lim}\ \underset{F\in\mathbf{F}}{sup} \mathbb{E}_F\Big[\Big(\frac{\psi_j(W)}{\sigma_j(F)}\Big)^2 \mathbb{1}\Big\{\Big|\frac{\psi_j(W)}{\sigma_j(F)}\Big| > \lambda\Big\}\Big] = 0$. If RC satisfied, then under F_n, $\forall j$, $\frac{\sqrt{n}(\widehat{\mu_j} - \mu_{M_j}(F_n))}{\sigma_j(F_n)} \xrightarrow{d} \mathcal{N}(0, 1)$.*

Proof. By Lemma 1, $\sqrt{n}((\widehat{\mu}_M - \mu_M) = n^{-\frac{1}{2}}\Sigma_{i=1}^n \psi(W_i) + op(1)$ where $\psi(W_i)$ is characterized as (22) and $E(\psi(W_i)) = 0$. Since W_i is i.i.d., $\psi(W_i)$ is i.i.d.. Then by Lemma 11.4.1 of Romano and Lehmann (2005), $\frac{n^{-\frac{1}{2}}\Sigma_{i=1}^n \psi_j(W_i)}{\sigma_j(F_n)} \xrightarrow{d} \mathcal{N}(0, 1)$. The desired result followed then by Slutsky's theorem. □

Remark 6. This lemma relies on that $\sigma_j(F_n)$ cannot tend to zero. There is a DGP of $\psi_j(W)$ on $\mathbb{R}$ uniquely determined by the DGP F of W so that Lemma 11.4.1 of Romano and Lehmann (2005), derived by Lindeberg CLT, is applicable. The asymptotic equivalence expansion is derived upon the true DGP F_0. Given the true DGP $F_0 = F_n$, we could have the above dynamic DGP CLT.

Lemma 3. *Let $Y_{n,i}, \ldots, Y_{n,n}$ be i.i.d. with distribution $G_n \in \tilde{\mathbf{F}}$ with finite mean $\mu(G_n)$ and $\tilde{\mathbf{F}}$ satisfies*

$$\underset{\lambda\to\infty}{lim}\ \underset{F\in\tilde{\mathbf{F}}}{sup} E_F\Big[\frac{|Y - \mu(F)|^2}{\sigma^2(F)} \mathbb{1}\{\{\Big\{\frac{|Y - \mu(F)|}{\sigma(F)} > \lambda\Big\}\Big] = 0 \tag{37}$$

Let $\overline{Y}_n = \Sigma_{i=1}^n \frac{Y_{n,i}}{n}$. *Then under* G_n, $\overline{Y}_n - \mu(G_n) \xrightarrow{p} 0$.

This Lemma is similar to Lemma 11.4.2 of Romano and Lehmann (2005) but under a different assumption.

Proof. The proof follows the arguments of the proof of Lemma 11.4.2 of Romano and Lehmann (2005) by define $Z_{n,i} \equiv Y_{n,i}\mathbb{1}\{|Y_{n,i}| \leq n\}$ and $m_n \equiv E(Z_{n,i})$ so that we could prove the result by an intermediate result $\overline{Y_{n,i}} \xrightarrow{p} m_n$, which can be shown by two functions $\tau_n(t) \equiv t[1 - G_n(t) + G_n(-t)]$ and $\kappa_n(t) \equiv \frac{1}{t}\int_{-t}^{t} x^2 dG_n(t)$ that transform it into showing $P(|\overline{Y_n} - m_n| > \epsilon) \leq \epsilon^{-2}\kappa_n(n) + \tau_n(n) \to 0$. WLOG, let $\mu(G_n) = 0$. By observing that: $\underset{n\to\infty}{lim} E_{G_n}\big(|Y_{n,i}|^2\mathbb{1}\{\{|Y_{n,i}| > n\}\big) = \underset{n\to\infty}{lim} E_{G_n}\big(|Y_{n,i}|^2|\{|Y_{n,i}| > n\big) \times P(|Y_{n,i}| > n) \geq \underset{n\to\infty}{lim} E_{G_n}\big(|Y_{n,i}||\{|Y_{n,i}| > n\big) \times P(|Y_{n,i}| > n) = \underset{n\to\infty}{lim} E_{G_n}\big(|Y_{n,i}|\mathbb{1}\{\{|Y_{n,i}| > n\}\big)$,we will still have $\underset{n\to\infty}{lim} \tau_n(n) = 0$ since $0 \leq \tau_n(n) = nP\big(|Y_{n,i}| > n\big) \leq E_{G_n}\big[|Y_{n,i}|\mathbb{1}\{\{|Y_{n,i}| > n\}\big]$ for any $n > 0$ and the uniform integrability condition (37) , which implies $\underset{n\to\infty}{lim} E_{G_n}\big(|Y_{n,i}|^2\mathbb{1}\{\{|Y_{n,i}| > n\}\big) = 0$, and $\kappa_n(n) \to 0$ because $\forall\delta > 0$, $\exists\beta_0$ such that $\underset{n\to\infty}{limsup} E[|Y_{n,i}| \times \mathbb{1}\{|Y_{n,i}| > \beta_0\}] < \frac{\delta}{4}$ while $n = \beta_0$. Then $m_n \xrightarrow{p} 0$ by the same observation. The desired result thereby is achieved by continuous mapping theorem. □

Lemma 4. *Let* W_i, $i = 1, \ldots, n$, *be an i.i.d. sequence of random vectors with distribution* $F_n \in \mathbf{F}$ *on* $\mathbb{R}^p$. *Suppose: (i)* $\mathbf{F}$ *is such that for all* $1 \leq j \leq k$,

$$\underset{\lambda\to\infty}{lim}\ \underset{F\in\mathbf{F}}{sup}\mathbb{E}_F\Big[\Big(\frac{\psi_j(W)}{\sigma_j(F)}\Big)^2\mathbb{1}\Big\{\Big|\frac{\psi_j(W)}{\sigma_j(F)}\Big| > \lambda\Big\}\Big] = 0, \tag{38}$$

(ii) RC satisfied, (iii) $\| \widehat{h} - h_0 \| \xrightarrow{p} 0$, *(iv)* $\frac{1}{n}\Sigma_{i=1}^n \| \widehat{\alpha}(W_i) - \alpha(W_i) \|^2 \xrightarrow{p} 0$ *and (v)* $\| g(W,\mu,h) - g(W,\mu_M,h_0) \| \leq b(W)(\| \mu - \mu_M \| + \| h - h_0 \|)$ *while* $E[b^2(W)] < \infty$, *then under the true DGP* F_n*: (i)*$S_{j,n}^2 \xrightarrow{p} \sigma_j^2(\theta, F_n)$, *(ii)* $\| \widehat{\Sigma} - \Sigma(F_n) \| \xrightarrow{p} 0$ *and (iii)* $\| \widehat{\Omega} - \Omega(F_n) \| \xrightarrow{p} 0$.

Proof. Following the proof of Lemma 1, under assumption (ii), $\psi(W) = g(W,\mu_M,h_0) + \alpha(W)$, where $g(W,\mu_M,h_0) = m(W,\theta,h_0) - \mu_M$, $\alpha(W) = E\big(\frac{\partial m(W,\theta_0,h_0(x))}{\partial h_0(x)}\big|x\big)(y - h_0(x))$ and $E[\psi(W)] = \mathbf{0}$. Under assumption (ii), in particular RC assumption 3-7, and (iii), by Lemma 5.2 of Newey (1994a), $\widehat{\mu} \xrightarrow{p} \mu_M$. Then by assumption (iii) and (v), $\frac{1}{n}\Sigma_{i=1}^n \| g(W,\widehat{\mu},\widehat{h}) - g(W,\mu_M,h_0) \|^2 \xrightarrow{p} 0$, applying Lindeberg-Lévy CLT and continuous mapping theorem. Then by assumption (ii) and (iv), following the proof of Lemma 8.3 of Newey and McFadden (1994),

$$\frac{1}{n}\Sigma_{i=1}^{n}\widehat{\psi(W_i)}\widehat{\psi(W_i)}' \xrightarrow{p} \frac{1}{n}\Sigma_{i=1}^{n}\psi(W_i)\psi(W_i)',$$

where $\widehat{\psi_j(W_i)} = g_j(W_i, \widehat{\mu}, \widehat{h}) + \widehat{\alpha_j}(W_i)$ and $\widehat{\alpha_j}(W_i) = E\big(\frac{\partial m_j(\widehat{W,\theta_0,h_0(x)})}{\partial h_0(x)}|x)(y_i - \widehat{h}_0(x_i))$. Thus, $\frac{1}{n}\Sigma_{i=1}^{n}\widehat{\psi_j(W_i)}\widehat{\psi_j(W_i)}' \xrightarrow{p} \frac{1}{n}\Sigma_{i=1}^{n}\psi_j(W_i)\psi_j(W_i)'$ and $\frac{1}{n}\Sigma_{i=1}^{n}\widehat{\psi_l(W_i)}\widehat{\psi_q(W_i)}' \xrightarrow{p} \frac{1}{n}\Sigma_{i=1}^{n}\psi_l(W_i)\psi_q(W_i)'$, where $l \neq q$. Then by assumption (i) and Lemma 3,

$$\frac{1}{n}\Sigma_{i=1}^{n}\psi_j(W_i)\psi_j(W_i)' \xrightarrow{p} E_{F_n}[\psi_j(W)\psi_j(W)'] = \sigma_j^2(\theta, F_n),$$

which leads to

$$S_{j,n}^2 = \frac{1}{n}\Sigma_{i=1}^{n}\widehat{\psi_j(W_i)}\widehat{\psi_j(W_i)}' \xrightarrow{p} \sigma_j^2(\theta, F_n)$$

by continuous mapping theorem. Regarding (ii), we need further the consistency of the off-diagonal element of $\widehat{\Sigma}$ under a dynamic DGP true F_n when we don't have the corresponding uniform integrability condition as (38). Then also by continuous mapping theorem, it's sufficient to show, under F_n, $\frac{1}{n}\Sigma_{i=1}^{n}\psi_l(W_i)\psi_q(W_i) \xrightarrow{p} E_{F_n}[\psi_l(W)\psi_q(W)']$. Let $Z_{n,i} \equiv \psi_l(W_{n,i})\psi_q(W_{n,i})$, following the proof of Lemma S.7.1 of Romano and Shaikh (2012b), by the inequality

$$|a||b|\mathbb{1}\{\{|a||b| > \lambda\} \leq |a|^2\mathbb{1}\{|a| > \sqrt{\lambda}\} + |b|^2\mathbb{1}\{|b| > \sqrt{\lambda}\},$$

we will have

$$\underset{\lambda\to\infty}{lim}\ \underset{n\to\infty}{limsup} E_{F_n}\big[|Z_{n,i} - E_{F_n}(Z_{n,i})|\mathbb{1}\{|Z_{n,i} - E_{F_n}(Z_{n,i})| > \lambda\}\big] = 0.$$

Therefore, by Lemma 11.4.2 of Romano and Lehmann (2005), the desired result follows. (iii) is a direct consequence of (i) and (ii) by continuous mapping theorem so that $S_{j,n} \xrightarrow{p} \sigma_j(\theta, F_n)$. □

Lemma 5.

$$\underset{\lambda\to\infty}{lim}\ \underset{F\in\mathbf{F}}{sup} E_F\big[\big(\frac{\psi_j(W)}{\sigma_j(F)}\big)^2\{\mathbb{1}\{|\frac{\psi_j(W)}{\sigma_j(F)}| > \lambda\}\big] = 0$$

if and only if $\forall F_n \in \mathbf{F}$,

$$\underset{\lambda\to\infty}{lim}\,\underset{n\to\infty}{limsup} E_{F_n}\big[\big(\frac{\psi_j(W)}{\sigma_j(F_n)}\big)^2\{\mathbb{1}\{|\frac{\psi_j(W)}{\sigma_j(F_n)}| > \lambda\}\big] = 0$$

Proof. (only if): as $\lambda \to \infty$, $\underset{F\in\mathbf{F}}{sup} E_F\big[\big(\frac{\psi_j(W)}{\sigma_j(F)}\big)^2\{\mathbb{1}\{|\frac{\psi_j(W)}{\sigma_j(F)}| > \lambda\}\big] \geq \underset{n\to\infty}{lim} E_{F_n}\big[\big(\frac{\psi_j(W)}{\sigma_j(F_n)}\big)^2\{\mathbb{1}\{|\frac{\psi_j(W)}{\sigma_j(F_n)}| > \lambda\}\big] \geq 0, \forall F_n \in \mathbf{F}$. The desired result followed by squeeze theorem when $\underset{n\to\infty}{lim} E_{F_n}\big[\big(\frac{\psi_j(W)}{\sigma_j(F_n)}\big)^2\{\mathbb{1}\{|\frac{\psi_j(W)}{\sigma_j(F_n)}| > \lambda\}\big] = \underset{n\to\infty}{limsup} E_{F_n}\big[\big(\frac{\psi_j(W)}{\sigma_j(F_n)}\big)^2\{\mathbb{1}\{|\frac{\psi_j(W)}{\sigma_j(F_n)}| > \lambda\}\big]$.

(if): Suppose the supremum exists for $E_F\big[\big(\frac{\psi_j(W)}{\sigma_j(F)}\big)^2\{\mathbb{1}\{|\frac{\psi_j(W)}{\sigma_j(F)}| > \lambda\}\big]$, then there exists a subsequence $\{F_n : F_n \in \mathbf{F}\}$ such that $\underset{F\in\mathbf{F}}{sup} E_F\big[\big(\frac{\psi_j(W)}{\sigma_j(F)}\big)^2\{\mathbb{1}\{|\frac{\psi_j(W)}{\sigma_j(F)}| > \lambda\}\big] = \underset{n\to\infty}{lim} E_{F_n}\big[\big(\frac{\psi_j(W)}{\sigma_j(F_n)}\big)^2\{\mathbb{1}\{|\frac{\psi_j(W)}{\sigma_j(F_n)}| > \lambda\}\big]$. Thus $\underset{n\to\infty}{limsup} E_{F_n}\big[\big(\frac{\psi_j(W)}{\sigma_j(F_n)}\big)^2\{\mathbb{1}\{|\frac{\psi_j(W)}{\sigma_j(F_n)}| > \lambda\}\big] = \underset{F\in\mathbf{F}}{sup} E_F\big[\big(\frac{\psi_j(W)}{\sigma_j(F)}\big)^2\{\mathbb{1}\{|\frac{\psi_j(W)}{\sigma_j(F)}| > \lambda\}\big] = 0$ as $\lambda \to \infty$. □

Lemma 6. *Let $J_n(x, F)$ be the distribution of the test statistic T_n^{min} as defined by (10) with $\widehat{\mu} = \frac{1}{n}\Sigma_{i=1}^n m(W_i, \theta, \widehat{h})$. Suppose all the assumptions in Lemma 4 satisfied. Let $\mathbf{F}'$ be the set of all distributions on $\mathbb{R}^p$. Define $\forall (Q, F) \in \mathbf{F}' \times \mathbf{F}$,*

$$\rho(Q, F) = max\{\underset{1\leq j\leq k}{max}\{\int_0^\infty |r_j(\lambda, Q) - r_j(\lambda, F)| exp(-\lambda) d\lambda\}, \parallel \Omega(Q) - \Omega(F) \parallel\}$$

where $\forall 1 \leq j \leq k$,

$$r_j(\lambda, F) = E_F\Big[\Big(\frac{\psi_j(W)}{\sigma_j(F)}\Big)^2 \mathbb{1}\Big\{\Big|\frac{\psi_j(W)}{\sigma_j(F)}\Big| > \lambda\Big\}\Big].$$

Then for all sequences $\{Q_n \in \mathbf{F}' : n \geq 1\}$ and $\{F_n \in \mathbf{F} : n \geq 1\}$ that satisfy $\rho(Q_n, F_n) \to 0$,

$$\underset{n\to\infty}{limsup}\,\underset{x\in\mathbb{R}}{sup} |J_n(x, Q_n) - J_n(x, F_n)| = 0. \tag{39}$$

Proof. For all sequences $\{Q_n \in \mathbf{F}' : n \geq 1\}$ and $\{F_n \in \mathbf{F} : n \geq 1\}$ that satisfy $\rho(Q_n, F_n) \to 0$, following the proof of Lemma S.12.1 of Romano and Shaikh (2012b), by assumption of the uniform integrability (38) and proof by contradiction, we would have

$$\underset{\lambda\to\infty}{lim}\,\underset{n\to\infty}{limsup}\, r_j(\lambda, F_n) = 0$$

$$\underset{\lambda\to\infty}{lim}\,\underset{n\to\infty}{limsup}\, r_j(\lambda, Q_n) = 0$$

by Lemma 5, which is the uniform integrability condition. Then suppose (39) fails, then there exists a subsequence n_v such that $\Omega(F_{n_v}) \to \Omega^*$, $\Omega(Q_{n_v}) \to \Omega^*$ and either

$$\underset{x\in\mathbb{R}}{sup}|J_{n_v}(x, F_{n_v}) - \Phi_{\Omega^*}(x,\ldots,x)| \nrightarrow 0$$

or

$$\underset{x\in\mathbb{R}}{sup}|J_{n_v}(x, Q_{n_v}) - \Phi_{\Omega^*}(x,\ldots,x)| \nrightarrow 0$$

where Φ_Σ is the cdf of $\mathcal{N}(0,\Sigma)$. Consider under F_{n_v},

$$P\big(n^{-\frac{1}{2}}\Sigma_{i=1}^n\psi_1(W_i) \le -\sigma_1(F_{n_v})x, \ldots, n^{-\frac{1}{2}}\Sigma_{i=1}^n\psi_k(W_i) \le -\sigma_k(F_{n_v})x\big)$$

where $E_{F_{n_v}}[\psi_j(W)] = 0\ \forall 1 \le j \le k$. $(-\infty, -\sigma_1(F_{n_v})x] \times \cdots \times (-\infty, -\sigma_k(F_{n_v})x]$ is convex. Then by Lemma 3.1 of Romano and Shaikh (2008) and continuous mapping theorem,

$$|\zeta_{F_{n_v}}(\cdot) - \Phi_{\Omega(F_{n_v})}(\cdot)| \to 0,$$

where $\zeta_{F_{n_v}}(\cdot) \equiv P\big(\frac{n^{-\frac{1}{2}}\Sigma_{i=1}^n\psi_1(W_i)}{\sigma_1(F_{n_v})} \le (\cdot), \ldots, \frac{n^{-\frac{1}{2}}\Sigma_{i=1}^n\psi_k(W_i)}{\sigma_k(F_{n_v})} \le (\cdot)\big)$ and $(\cdot)$ represents any elements in the $(-\infty, -x]^k$. Since $|\Phi_{\Omega(F_{n_v})}(\cdot) - \Phi_{\Omega^*}(\cdot)| \to 0$ by continuity,

$$|\zeta_{F_{n_v}}(\cdot) - \Phi_{\Omega^*}(\cdot)| \to 0$$

by triangular inequality. By Lemma 4 and continuous mapping theorem, $S_{j,n}\xrightarrow{p}\sigma_j(\theta, F_n)$, $\forall 1 \le j \le k$. Then

$$\big|P\big(\frac{n^{-\frac{1}{2}}\Sigma_{i=1}^n\psi_1(W_i)}{S_{1,n}} \le (\cdot), \ldots, \frac{n^{-\frac{1}{2}}\Sigma_{i=1}^n\psi_k(W_i)}{S_{k,n}} \le (\cdot)\big) - \Phi_{\Omega^*}(\cdot)\big| \to 0$$

by Slutsky's Theorem. Following Lemma 1, $\sqrt{n}(\widehat{\mu}_{M_j} - \mu_{M_j}) = n^{-\frac{1}{2}}\Sigma_{i=1}^n\psi_j(W_i) + op(1)$.

$$\big|P\big(\frac{\sqrt{n}(\widehat{\mu}_{M_1} - \mu_{M_1}(F_{n_v}))}{S_{1,n}} \le (\cdot), \ldots, \frac{\sqrt{n}(\widehat{\mu}_{M_k} - \mu_{M_k}(F_{n_v}))}{S_{k,n}} \le (\cdot)\big) - \Phi_{\Omega^*}(\cdot)\big| \to 0$$

by Slutsky's Theorem. Thus,

$$\begin{aligned} J_n(x, F_{n_v}) &= 1 - P\big(\min_{1\le j\le k} \tfrac{\sqrt{n}(\mu_{M_j}(F_{n_v}) - \widehat{\mu}_{M_j})}{S_{j,n}} \ge x\big) \\ &= 1 - P\big(\tfrac{\sqrt{n}(\widehat{\mu}_{M_1} - \mu_{M_1}(F_{n_v}))}{S_{1,n}} \le (\cdot), \dots, \tfrac{\sqrt{n}(\widehat{\mu}_{M_k} - \mu_{M_k}(F_{n_v}))}{S_{k,n}} \le (\cdot)\big). \end{aligned}$$

Therefore,

$$J_n(x, F_{n_v}) \to 1 - \Phi_{\Omega^*}(\cdot) = \Phi_{\Omega^*}(\circ),$$

where $(\circ)$ represents any elements in the $(-\infty, x]^k$. Similarly, one can show that $J_n(x, Q_{n_v}) \to \Phi_{\Omega^*}(\circ)$. Then by Polya's Theorem, $\sup_{x\in\mathbb{R}} |J_{n_v}(x, F_{n_v}) - \Phi_{\Omega^*}(x, \dots, x)| \to 0$ and $\sup_{x\in\mathbb{R}} |J_{n_v}(x, Q_{n_v}) - \Phi_{\Omega^*}(x, \dots, x)| \to 0$, which is a contradiction. □

Lemma 7. *Suppose all the assumptions in Lemma 6 satisfied. Then $M_n(\beta)$ defined by (12)*

$$\liminf_{n\to\infty} \inf_{F\in\mathbf{F}} P\{\mu \in M_n(\beta)\} \ge 1 - \beta.$$

Proof. Consider any sequence $\{F_n \in \mathbf{F} : n \ge 1\}$ as the true DGP of $W_{n,i}$ and denote $\widehat{F_n}$ as the empirical distribution of $W_{n,i}$. Then trivially $P_{F_n}\{\widehat{F_n} \in \mathbf{F}'\} \to 1$. By Lemma S.12.2 of Romano and Shaikh (2012b), $\int_0^\infty |r_j(\lambda, \widehat{F_n}) - r_j(\lambda, F_n)| exp(-\lambda) d\lambda \xrightarrow{P_{F_n}} 0$, $\forall 1 \le j \le k$. And by Lemma 4, $\| \widehat{\Omega} - \Omega(F_n) \| \xrightarrow{p} 0$. Hence, by continuous mapping theorem, $\rho(\widehat{F_n}, F_n) \xrightarrow{P_{F_n}} 0$. Then by Lemma 6 and Theorem 2.4 of Romano and Shaikh (2012a),

$$\liminf_{n\to\infty} \inf_{F\in\mathbf{F}} P\{J_n^{-1}(\alpha_1, \widehat{F_n}) \le T_n^{min} \le J_n^{-1}(1 - \alpha_2, \widehat{F_n})\} \ge 1 - \alpha_1 - \alpha_2$$

holds for any $\alpha_1 \ge 0$ and $\alpha_2 \ge 0$ satisfying $0 \le \alpha_1 + \alpha_2 \le 1$. The desired result follows immediately when $\alpha_1 = \beta$ and $\alpha_2 = 0$. □

Lemma 8. *Suppose all the assumptions in Lemma 6 satisfied. If $\frac{\sqrt{n}\mu_j(F_n)}{\sigma_j(F_n)} \to \infty$ $\forall 1 \le j \le k$, then $P_{F_n}(M_n(\beta) \subseteq \mathbb{R}^k_+) \to 1$.*

Proof. By definition of $M_n(\beta)$, $\mu \in M_n(\beta)$ is such that $\frac{\sqrt{n}(\mu_j(F_n) - \widehat{\mu_j})}{S_j} \ge K_n^{-1}(\beta, \widehat{F_n})$ $\forall 1 \le j \le k$, where $K_n^{-1}(\beta, \widehat{F_n})$ is defined as (13), that is,

$$\mu_j \geq \frac{K_n^{-1}(\beta, \widehat{F_n}) \times S_j}{\sqrt{n}} + \widehat{\mu_j} = \frac{\sigma_j(F_n)}{\sqrt{n}} \Big[\frac{\sqrt{n}(\widehat{\mu_j} - \mu_j(F_n))}{\sigma_j(F_n)} + \frac{\sqrt{n}\mu_j(F_n)}{\sigma_j(F_n)} + \frac{K_n^{-1}(\beta, \widehat{F_n}) \times S_j}{\sigma_j(F_n)} \Big]$$

$\forall 1 \leq j \leq k$. By Lemma 2, $\frac{\sqrt{n}(\widehat{\mu_j} - \mu_j(F_n))}{\sigma_j(F_n)} = Op_{Fn}(1)$. By Lemma 4 and continuous mapping theorem, $\frac{S_j}{\sigma_j(F_n)} \overset{p_{F_n}}{\to} 1$. By Lemma 6, $K_n(x, \widehat{F_n}) \to \Phi_{\Omega^*}(\circ)$, where $(\circ)$ represents any elements in the $(-\infty, x]^k$. Thus, $K_n^{-1}(\beta, \widehat{F_n}) = Op_{Fn}(1)$. Therefore, when $\frac{\sqrt{n}\mu_j(F_n)}{\sigma_j(F_n)} \to \infty \; \forall 1 \leq j \leq k$, the desired result follows. $\square$

Lemma 9. *Let $J_n(x, F)$ be the distribution of the test statistic T_n given by (4), (6) and (7) with $\widehat{\mu} = \frac{1}{n}\Sigma_{i=1}^n m(W_i, \theta, \widehat{h})$. After replacing $\sqrt{n}\mu$ by $\sqrt{n}\lambda_n$ in T_n, denote $J_n(x, \lambda_n, F)$ as the distribution of it, i.e.*

$$J_n(x, \lambda, F) = P\Big(S\big(S_n^{-1}(\sqrt{n}(\widehat{\mu} - \mu(F))) + S_n^{-1}\sqrt{n}\lambda, \widehat{\Omega_n}\big) \leq x\Big)$$

Suppose all the assumptions in Lemma 4 satisfied. Define $\rho(Q, F)$ as in Lemma 6. Suppose for some $\emptyset \neq I \subseteq \{1, \ldots, k\}$,

$$\frac{\sqrt{n}\mu_j(F_n)}{\sigma_j(F_n)} \to \delta_j$$

for all $j \in I$ and some $\delta_j \geq 0$ and

$$\frac{\sqrt{n}\mu_j(F_n)}{\sigma_j(F_n)} \to \infty$$

for all $j \notin I$. Then

$$P_{F_n}\big(T_n > J_n^{-1}(1 - \alpha + \beta, \mu(F_n), \widehat{F_n})\big) \to \alpha - \beta.$$

Proof. Consider $T_n = \underset{1 \leq j \leq k}{max} - \frac{\sqrt{n}\widehat{\mu_j}}{S_j}$. Under F_n, $\frac{\sqrt{n}\mu_j(F_n)}{S_j} = \frac{\sigma_j(F_n)}{S_j} \frac{\sqrt{n}\mu_j(F_n)}{\sigma_j(F_n)}$. By Lemma 4 and continuous mapping theorem, $\frac{S_j}{\sigma_j(F_n)} \overset{p_{F_n}}{\to} 1$. Thus, by assumption,

$$\frac{\sqrt{n}\mu_j(F_n)}{S_j} \overset{p_{F_n}}{\to} \delta_j$$

for all $j \in I$ and some $\delta_j \geq 0$ and

$$\frac{\sqrt{n}\mu_j(F_n)}{S_j} \overset{p_{F_n}}{\to} \infty$$

for all $j \notin I$. Hence,

$$\underset{1\leq j\leq k}{max} - \frac{\sqrt{n}\widehat{\mu_j}}{S_j} = \underset{j\in I}{max}\{-\frac{\sqrt{n}(\widehat{\mu_j} - \mu_j(F_n))}{S_j} - \frac{\sqrt{n}\mu_j(F_n)}{S_j}\} + op_{F_n}(1). \tag{40}$$

Under Q_n, $\frac{\sqrt{n}\mu_j(Q_n)}{S_j} = \frac{\sigma_j(Q_n)}{S_j}\frac{\sigma_j(F_n)}{\sigma_j(Q_n)}\frac{\sqrt{n}\mu_j(F_n)}{\sigma_j(F_n)}$. For all sequences $\{Q_n \in \mathbf{F}' : n \geq 1\}$ and $\{F_n \in \mathbf{F} : n \geq 1\}$ that satisfy $\rho(Q_n, F_n) \to 0$, following the proof in Lemma 6 and by Lemma 5,

$$\underset{\lambda\to\infty}{lim}\underset{Q_n\in\mathbf{F}'}{sup} E_{Q_n}\big[\big(\frac{\psi_j(W)}{\sigma_j(Q_n)}\big)^2\{\mathbb{1}\{|\frac{\psi_j(W)}{\sigma_j(Q_n)}| > \lambda\}\big] = 0.$$

Thus, similarly, $\frac{S_j}{\sigma_j(Q_n)}\overset{p_{Q_n}}{\to}1$. At the same time, proof by contradiction leads to $\parallel \Omega(Q_n) - \Omega(F_n) \parallel \to 0$. Therefore, $\frac{\sigma_j(Q_n)}{\sigma_j(F_n)} \to 1$. Thus,

$$\frac{\sqrt{n}\mu_j(Q_n)}{S_j}\overset{p_{Q_n}}{\to}\delta_j$$

for all $j \in I$ and some $\delta_j \geq 0$ and

$$\frac{\sqrt{n}\mu_j(Q_n)}{S_j}\overset{p_{Q_n}}{\to}\infty$$

for all $j \notin I$. Hence,

$$\underset{1\leq j\leq k}{max} - \frac{\sqrt{n}\widehat{\mu_j}}{S_j} = \underset{j\in I}{max}\{-\frac{\sqrt{n}(\widehat{\mu_j} - \mu_j(Q_n))}{S_j} - \frac{\sqrt{n}\mu_j(Q_n)}{S_j}\} + op_{Q_n}(1). \tag{41}$$

Following the arguments in the proof of Lemma 6 and further by Slutsky's Theorem, $J_n(x, F_n) \to \Phi_{\Omega^*}(\circ)$, where $(\circ)$ represents any elements in the$(-\infty, x + \delta_j]^{|I|}$ for each $j \in I$ for some $\delta_j \geq 0$. Similarly, $J_n(x, Q_n) \to \Phi_{\Omega^*}(\circ)$. Therefore, applying proof by contradiction as in Lemma 6, for all sequences $\{Q_n \in \mathbf{F}' : n \geq 1\}$ and $\{F_n \in \mathbf{F} : n \geq 1\}$, if $\rho(Q_n, F_n) \to 0$,

$$\underset{n\to\infty}{lim sup}\underset{x\in\mathbb{R}}{sup}|J_n(x, Q_n) - J_n(x, F_n)| = 0.$$

Besides trivially, $P_{F_n}\{\widehat{F_n} \in \mathbf{F}'\} \to 1$. And by the same arguments in the proof of Lemma 7, $\rho(\widehat{F_n}, F_n)\overset{P_{F_n}}{\to}0$. Then by Theorem 2.4 of Romano and Shaikh (2012a), the desired result follows. The case of T_n given by (6) and (7) can be proved similarly. □

C Proof of Theorem 1

Under H_0 and the assumption that $h_0 = E[Y|X]$, its sufficient to show $\limsup_{n\to\infty} \underset{F\in\mathbf{F}}{sup} E_F[\phi_n] \leq \alpha$. The proof then follows that of Theorem 2.1 in Romano et al. (2014). Suppose (9) fails, then there exists a subsequence n_t and $\eta > \alpha$ such that $E_{F_{n_t}}(\phi_{n_t}) \to \eta$. Observe that

$$E_{F_{n_t}}(\phi_{n_t}) = 1-P_{F_{n_t}}\Big(\{M_{n_t}(\beta) \subseteq \mathbb{R}_+^k\} \cup \{T_{n_t}(\theta) \leq C_{n_t}(1\widehat{-\alpha+}\beta,\theta)\}\Big) \leq 1-P_{F_{n_t}}\Big(\{M_{n_t}(\beta) \subseteq \mathbb{R}_+^k\}\Big)$$

and that

$$E_{F_{n_t}}(\phi_{n_t}) \leq 1 - P_{F_{n_t}}\Big(\{T_{n_t}(\theta) \leq C_{n_t}(1\widehat{-\alpha+}\beta,\theta)\}\Big).$$

By Lemma 8, when $\frac{\sqrt{n_t}\mu_j(F_{n_t})}{\sigma_j(F_{n_t})} \to \infty\ \forall 1 \leq j \leq k$, then $P_{F_{n_t}}(M_{n_t}(\beta) \subseteq \mathbb{R}_+^k) \to 1$. Hence, $E_{F_{n_t}}(\phi_{n_t}) \to 0$, which is a contradiction.

Then consider the case when $\frac{\sqrt{n_t}\mu_j(F_{n_t})}{\sigma_j(F_{n_t})} \to \delta_j$ for all $j \in I$ and some $\delta_j \geq 0$ and $\frac{\sqrt{n_t}\mu_j(F_{n_t})}{\sigma_j(F_{n_t})} \to \infty$ for all $j \notin I$. By observing that $C_n(1\widehat{-\alpha+}\beta,\theta)$ is the critical value of $J_n^{-1}(1-\alpha+\beta,\lambda,\widehat{F_n})$ at level $1-\alpha+\beta$ when $\lambda_{min} = \underset{\lambda\in M_n(\beta)\bigcap\mathbb{R}_+^k}{arg\,sup} J_n^{-1}(1-\alpha+\beta,\lambda,\widehat{F_n})$, which is weakly greater than $J_n^{-1}(1-\alpha+\beta,\mu,\widehat{F_n}) = J_n^{-1}(1-\alpha+\beta,\widehat{F_n})$. Hence,

$$E_{F_{n_t}}(\phi_{n_t}) \leq P_{F_{n_t}}\Big(\{T_{n_t} \geq J_{n_t}^{-1}(1-\alpha+\beta,\widehat{F_{n_t}})\}\Big) + \beta,$$

where $0 < \beta < \alpha$. By Lemma 9, $P_{F_{n_t}}\big(T_{n_t} > J_{n_t}^{-1}(1-\alpha+\beta,\widehat{F_{n_t}})\big) \to \alpha-\beta$. Then $E_{F_{n_t}}(\phi_{n_t}) \to \alpha$, which is a contradiction. Therefore, the desired result follows.

Q.E.D.

References

Ackerberg, D., X. Chen, J. Hahn, and Z. Liao (2014). Asymptotic efficiency of semiparametric two-step gmm. *Review of Economic Studies 81*(3), 919–943.

Andrews, D. W. (1992). Generic uniform convergence. *Econometric theory 8*(2), 241–257.

Andrews, D. W. (1994). Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica: Journal of the Econometric Society*, 43–72.

Andrews, D. W. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica*, 399–405.

Andrews, D. W. and P. J. Barwick (2012). Inference for parameters defined by moment inequalities: A recommended moment selection procedure. *Econometrica 80*(6), 2805–2826.

Andrews, D. W. and P. Guggenberger (2009). Validity of subsampling and 'plug-in asymptotic' inference for parameters defined by moment inequalities. *Econometric Theory 25*(3), 669–709.

Andrews, D. W. and X. Shi (2013). Inference based on conditional moment inequalities. *Econometrica 81*(2), 609–666.

Andrews, D. W. and G. Soares (2010a). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica 78*(1), 119–157.

Andrews, D. W. and G. Soares (2010b). Supplement to "inference for parameters defined by moment inequalities using generalized moment selection". *Econometrica 78*(1), 119–157.

Andrews, I., J. Roth, and A. Pakes (2023). Inference for linear conditional moment inequalities. *Review of Economic Studies 90*(6), 2763–2791.

Athey, S. (2001). Single crossing properties and the existence of pure strategy equilibria in games of incomplete information. *Econometrica 69*(4), 861–889.

Bajari, P., H. Hong, and S. P. Ryan (2010). Identification and estimation of a discrete game of complete information. *Econometrica 78*(5), 1529–1568.

Beresteanu, A., I. Molchanov, and F. Molinari (2011). Sharp identification regions in models with convex moment predictions. *Econometrica 79*(6), 1785–1821.

Beresteanu, A. and F. Molinari (2008). Asymptotic properties for a class of partially identified models. *Econometrica 76*(4), 763–814.

Berry, S. and E. Tamer (2006). Identification in models of oligopoly entry. *Econometric Society Monographs 42*, 46.

Chernozhukov, V., J. C. Escanciano, H. Ichimura, W. K. Newey, and J. M. Robins (2022). Locally robust semiparametric estimation. *Econometrica 90*(4), 1501–1535.

Chernozhukov, V., H. Hong, and E. Tamer (2007). Estimation and confidence regions for parameter sets in econometric models 1. *Econometrica 75*(5), 1243–1284.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013). Intersection bounds: Estimation and inference. *Econometrica 81*(2), 667–737.

Ciliberto, F. and E. Tamer (2009). Market structure and multiple equilibria in airline markets. *Econometrica 77*(6), 1791–1828.

Cox, G. and X. Shi (2023). Simple adaptive size-exact testing for full-vector and subvector inference in moment inequality models. *The Review of Economic Studies 90*(1), 201–228.

De Paula, A. and X. Tang (2012). Inference of signs of interaction effects in simultaneous games with incomplete information. *Econometrica 80*(1), 143–172.

Grieco, P. L. (2014). Discrete games with flexible information structures: An application to local grocery markets. *The RAND Journal of Economics 45*(2), 303–340.

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*, 315–331.

Hirano, K., G. W. Imbens, and G. Ridder (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica 71*(4), 1161–1189.

Iaryczower, M., X. Shi, and M. Shum (2018). Can words get in the way? the effect of deliberation in collective decision making. *Journal of Political Economy 126*(2), 688–734.

Imbens, G. W. and C. F. Manski (2004). Confidence intervals for partially identified parameters. *Econometrica 72*(6), 1845–1857.

Kaido, H., F. Molinari, and J. Stoye (2019). Confidence intervals for projections of partially identified parameters. *Econometrica 87*(4), 1397–1432.

Manski, C. F. and E. Tamer (2002). Inference on regressions with interval data on a regressor or outcome. *Econometrica 70*(2), 519–546.

McCloskey, A. (2017). Bonferroni-based size-correction for nonstandard testing problems. *Journal of Econometrics 200*(1), 17–35.

Molinari, F. (2020). Microeconometrics with partial identification. *Handbook of econometrics 7*, 355–486.

Newey, W. K. (1990). Semiparametric efficiency bounds. *Journal of applied econometrics 5*(2), 99–135.

Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica: Journal of the Econometric Society*, 1161–1167.

Newey, W. K. (1994a). The asymptotic variance of semiparametric estimators. *Econometrica: Journal of the Econometric Society*, 1349–1382.

Newey, W. K. (1994b). Kernel estimation of partial means and a general variance estimator. *Econometric Theory 10*(2), 1–21.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics 4*, 2111–2245.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2015). Moment inequalities and their application. *Econometrica 83*(1), 315–334.

Robinson, P. M. (1987). Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. *Econometrica: Journal of the Econometric Society*, 875–891.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, 931–954.

Romano, J. P. and E. Lehmann (2005). Testing statistical hypotheses.

Romano, J. P. and A. M. Shaikh (2008). Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference 138*(9), 2786–2807.

Romano, J. P. and A. M. Shaikh (2012a). On the uniform asymptotic validity of subsampling and the bootstrap. *The Annals of Statistics*, 2798–2822.

Romano, J. P. and A. M. Shaikh (2012b). Supplement to ”on the uniform asymptotic validity of subsampling and the bootstrap”.

Romano, J. P., A. M. Shaikh, and M. Wolf (2014). A practical two-step method for testing moment inequalities. *Econometrica 82*(5), 1979–2002.

Shi, X. (2015). Model selection tests for moment inequality models. *Journal of Econometrics 187*(1), 1–17.

Shi, X. and M. Shum (2015). Simple two-stage inference for a class of partially identified models. *Econometric Theory 31*(3), 493–520.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557–586.

Sweeting, A. (2009). The strategic timing incentives of commercial radio stations: An empirical analysis using multiple equilibria. *The RAND Journal of Economics 40*(4), 710–742.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies 70*(1), 147–165.