AN INTEGRATION–BASED APPROACH TO MOMENT INEQUALITY MODELS

YUANYUAN WAN†
THE PENNSYLVANIA STATE UNIVERSITY

ABSTRACT. I provide new integration–based estimation and inference methods in models with partial identification. Under the assumption that the identified set is connected, I propose a root–n–consistent estimator of any given projection of the identified set. I construct confidence sets for a class of scalar–valued functions of the parameter vector, which are easy to compute, have correct asymptotic coverage probability (pointwisely in the nuisance parameters), and are not conservative as projection–based confidence intervals.

JEL classification: C10, C11, C12, and C13

Keywords: Integration–based approach, moment inequality model, confidence set.

Job Market Paper

Date: Monday 3rd January, 2011.
†yxw162@psu.edu. Address: 605 Kern Graduate Building, Department of Economics, The Pennsylvania State University, University Park, PA 16802.
I am deeply grateful to Joris Pinkse and Sung Jae Jun for their constant guidance, inspiration and encouragement. I appreciate the valuable suggestions from Haiqing Xu. I also benefited from the discussions with Arie Beresteanu, Herman Bierens, Federico Bugni, Paul Grieco, Patrik Guggenberger, Isabelle Perrigne, Jack Porter, Xiaoxia Shi, Quang Vuong, Neil Wallace, Hal White, and Daniel Xu.

1. INTRODUCTION

I provide new integration–based estimation and inference methods in models with partial identification. Under the assumption that the identified set is connected, I propose a root–n–consistent estimator of any given projection of the identified set. I construct confidence sets for a class of scalar–valued functions of the parameter vector, which are easy to compute, have correct asymptotic coverage probability (pointwisely in the nuisance parameters), and are not conservative as projection–based confidence intervals. Embedded in this methodology is an automatic moment selection procedure.

The methods proposed in this paper apply to a class of models in which the parameters of interest may not be uniquely determined by economic restrictions and distributions of observables. Examples include missing data, censored or interval–observed data (e.g. Horowitz and Manski, 2000; Manski, 2003; Manski and Pepper, 2000; Manski and Tamer, 2002), models with level–k rationality (Aradillas-Lopez and Tamer, 2008), models with multiple equilibria (Andrews, Berry, and Jia, 2004; Ciliberto and Tamer, 2009; Grieco, 2010; Pakes, Porter, Ho, and Ishii, 2006), asset pricing models with incomplete markets (Kaido and White, 2009), incomplete auction models (Haile and Tamer, 2003), and instrumental variable regressions in which the rank condition is violated (Phillips, 1989). In this paper, I focus on models characterized by a set of moment inequalities with connected and closed identified set.[1]

My integration–based approach is built on the methods of Laplace estimation, which Chernozhukov and Hong (2003, CH) proposed to simplify computation in models with point identified parameters.[2] Laplace estimators are defined as minimizers of quasi-posterior risk functions, where the quasi-posteriors resemble Bayesian posteriors in which the loglikelihood is replaced by a rescaled sample objective function of a difficult–to–compute extremum estimator. In my paper, I estimate the projections of the identified set and make inferences about scalar–valued functions of the parameter vector in moment inequality models using marginal quasi–posteriors.

[1] An equality can be expressed by two inequalities.

[2] The estimators in CH are called Laplace type estimators (LTEs). LTEs have other nice properties, e.g. estimates can be as efficient as the extremum estimates, inference procedures based on the quantiles of the quasi-posterior distribution can yield asymptotically valid confidence intervals.

As is well–noted in the literature (e.g. Chernozhukov, Hansen, and Jansson, 2009; Hahn and Ridder, 2009; Jun, 2008), existing methods of projecting a high–dimensional confidence set generally leads to a conservative inference on the corresponding component of the parameter vector. By using marginal quasi–posteriors, my approach targets directly on the the parameter of interest and avoids the conservativeness caused by projection. Pakes, Porter, Ho, and Ishii (2006, PPHI) study inferences about one component of the parameter vector in moment inequality models. They assume that the "extreme" of the identified set is a singleton, but I do not.[3]

The asymptotics that I consider in this paper is pointwise in that the nuisance parameters of the model are assumed to be fixed (see discussions in Romano and Shaikh, 2010). I take this approach, because one of my primary goals is to develop an inference method for scalar–valued functions of the parameter vector which is not conservative and is valid (pointwise in nuisance parameters). As Andrews and Soares (2010) pointout, local uniformity over the set of nuisance parameters is a desirable property and I intend to pursue it in a follow up research.

The estimation and inference procedures proposed in this paper are computationally attractive. My method of estimation requires no more than getting random draws from marginal quasi–posteriors, which can be implemented using a Markov Chain Monte Carlo (MCMC) algorithm. To make inferences, I do not invert a hypothesis test, which would sometimes require a resampling procedure for each parameter value that is tested (e.g., Andrews and Guggenberger, 2009; Bugni, 2010; Canay, 2010; Chernozhukov, Hong, and Tamer, 2007; Romano and Shaikh, 2010). Instead, my confidence sets are constructed by using quantiles of the marginal quasi–posterior, which can be obtained by simulation; it requires little more than drawing random numbers from a multivariate normal with an estimated covariance matrix.

I use a "smooth" moment selection mechanism when calculating critical values. I assign different weights to moment equations. A binding moment receives a weight of one with probability approaching one; a non–binding moment receives a weight that converges in

[3] Let Θ_I be the identified set. Let $\theta_{1\ell} = \inf_{\theta_1}\{\exists\theta_2 : (\theta_1, \theta_2) \in \Theta_I\}$ and $\Theta_{I2}(\theta_{1\ell}) = \{\theta_2 : \theta_1 = \theta_{1\ell}, \theta \in \Theta_I\}$. PPHI assume that $\Theta_{I2}(\theta_{1\ell})$ is a singleton (their assumption A1 (d)). My methods allow that $\Theta_{I2}(\theta_{1\ell})$ has positive Lebesgue measure in $\mathbb{R}^{d-1}$.

probability to zero exponentially. In the limit, this moment selection mechanism picks out the binding moment conditions, as does the Chernozhukov, Hong, and Tamer (2007, CHT) recommendation (adopted by Rosen (2008), among others). In finite samples, however, my method always assigns positive weights to all the moments. There are other moment selection methods, e.g., the generalized moment selection (GMS) (Andrews and Soares, 2010) with this property. With GMS, a measure of slackness is added to the demeaned limiting variable, whereas in my setup the slackness of the moment equations is reflected in the weights.[4]

My method relates to the literature in which the parameters are scalar–valued and "interval–identified" (e.g. Imbens and Manski, 2004; Stoye, 2009; Woutersen, 2006). For researchers who are interested in a scalar–valued function of the parameter vector, my paper provides a natural way of transforming a multi–dimensional task into a one dimensional problem with "interval–identified" parameter. My method therefore provide a way to bridge existing results in the models with "interval–identified" parameter and the models with multi–dimensional parameters.

An alternative to my procedure is a Bayesian approach. Moon and Schorfheide (2009) show that classical confidence sets and Bayesian credible sets are generally different (the two types of sets may coincide when the parameter are point identified). Liao and Jiang (2010, LJ) have recently studied large sample properties of the posterior distribution based on the limited information likelihood for moment inequality models. They establish the consistency of their estimator and study moment selection issues. The credible sets proposed by LJ, however, are not valid confidence sets from a frequentist point of view. For example, when the parameter of interest is scalar–valued, LJ construct a 95% credible set by taking 2.5% and 97.5% quantiles of their posterior distribution. The credible set hence will be contained in the identified set with probability approaching one, thus is not a valid confidence set for either the true parameter value or the identified set itself. My confidence sets, on the other hand, are intervals defined by two marginal quasi–posterior quantiles which are chosen based on the asymptotic behavior of the tails of the marginal quasi–posterior. My confidence sets cover the true values with prespecified probability.

[4]GMS provides (uniform) correct asymptotic coverage probability. In this paper, I consider fixed models.

My estimation procedure uses MCMC algorithms, which have been applied by others to partially identified models to reduce computational burden. Chernozhukov, Hansen, and Jansson (2009, CHJ) use an MCMC algorithm to construct CHT type confidence sets for the parameter vector in quantile regression models. My paper exploits the advantage of MCMC algorithms for general moment inequality models and construct less conservative confidence sets for individual components of the the parameter vector.

In addition to what I have discussed before, there are other approaches to estimation and inference in models with partial identification. Beresteanu and Molinari (2008) and Beresteanu, Molchanov, and Molinari (2010) study inference on the identified set based on the theory of random sets; Kaido (2010) considers a duality approach for inference about the identified set, where subsampling procedures are employed; Bontemps, Magnac, and Maurin (2007) study inference for the parameter value in set–identified linear models.

The rest of the paper is organized as follows. I introduce the model and my estimator in section 2. I discuss the asymptotic properties of my estimator in section 3. In section 4, I propose procedures for constructing confidence sets both for the true parameter value and the identified set. I have some further discussions in section 5 before I conclude this paper in section 6.

2. Setup

Suppose that there is a true parameter $\theta_0 \in \Theta \subset \mathbb{R}^d$ that satisfies a set of moment inequalities

$$\mathbb{E}\left[m_{(j)}(W_i, \theta_0)\right] \leq 0, \quad j = 1, \cdots, J, \tag{1}$$

where $\{W_i\}_{i=1}^n$ are i.i.d. observations and Θ is the parameter space.

The identified set Θ_I is a collection of parameter values that satisfy the moment inequalities:

$$\Theta_I = \left\{\theta \in \Theta : \forall j = 1, \cdots, J : \mathbb{E}\left[m_{(j)}(W_i, \theta)\right] \leq 0\right\}.$$

Θ_I is not empty because $\theta_0 \in \Theta_I$ by construction.

In what follows, I assume that there is some $\tilde{\delta} \in (0, \infty)$ such that $m_{(j)}(W_i, \cdot)$ is defined on the $\tilde{\delta}$–expansion of Θ: $\Theta^{\tilde{\delta}} = \{\theta \in \mathbb{R}^d : \inf_{t \in \tilde{\Theta}} \|t - \theta\| \leq \tilde{\delta}\}$. Let

$$m(W_i, \theta) = (m_{(1)}(W_i, \theta), \cdots, m_{(j)}(W_i, \theta))'$$

and $m_i(\theta)$ be the abbreviation for $m(W_i, \theta)$. Throughout this paper, the measurability of $m_{(j)}$ is a maintained assumption.

For any J–vector x, let $\|x\|_+^2 = \sum_j \left(|x_j|_+\right)^2$, where $|x_j|_+ = \max\{0, x_j\}$. Following CHT, I consider the following population and sample objective functions

$$L(\theta) = -\left\|\mathbb{E}m_i(\theta)\right\|_+^2, \qquad L_n(\theta) = -\left\|\bar{m}(\theta)\right\|_+^2,$$

where $\bar{m}(\theta) = (1/n)\sum_{i=1}^n m_i(\theta)$.[5] Note that $L(\theta) = 0$ if and only if $\theta \in \Theta_I$.

I define the quasi–posterior density

$$f_n(\theta) = \frac{1}{D_n} \exp(nL_n(\theta)) \tag{2}$$

over the expanded parameter space Θ, where $D_n = \int_{t \in \Theta} \exp(nL_n(t))dt$ is for normalization. Note that f_n is not a Bayesian posterior because $\exp(nL_n(\theta))$ is not a likelihood.

The quasi–posterior density is defined such that it is always non–negative and declines sufficiently fast as the sample size increases for any θ outside of an ϵ–expansion of the identified set. The exponential function is not the only possible choice . In the case of point identification (CH), the exponential transformation is a natural choice because it leads to normal approximation for the quasi–posterior density in large sample. In partially identified models, the exponential transformation results in attractive properties also, which will be clear later.

From now on I focus on the first element of the parameter vector. Write $\theta = (\theta_1, \theta_2')'$ and $\theta_0 = (\theta_{01}, \theta_{02}')'$, where θ_1 is a scalar and θ_2 is potentially vector–valued. Let Θ_1 be the first dimension of Θ, i.e., the set of all possible θ_1's for which there exists a θ_2 such that $(\theta_1, \theta_2) \in \Theta$. Define Θ_2, Θ_{I1} and Θ_{I2} similarly. The object of interest is a confidence set for

[5]In CHT, the objective function has a weighting matrix $\Sigma(\theta)$: $L(\theta) = -\left\|\mathbb{E}m_i(\theta)\Sigma^{1/2}(\theta)\right\|_+^2$. In this paper, I let $\Sigma(\theta) = I$ for the sake of notational simplicity. My approach can be extended to accommodate weighting matrices.

θ_{01}. Let

$$\theta_{1\ell} = \inf_{\theta_1 \in \Theta_{I1}} \theta_1, \qquad \theta_{1u} = \sup_{\theta_1 \in \Theta_{I1}} \theta_1.$$

So $\theta_{1\ell}$ and θ_{1u} are the end points of Θ_{I1}. Likewise, $\underline{\theta}_1$ and $\overline{\theta}_1$ be the end points of Θ_1. They are assumed to be known to the researcher. For any $\theta_1^* \in \Theta_{I1}$, let

$$\Theta_{I2}(\theta_1^*) = \{\theta_2 : (\theta_1^*, \theta_2) \in \Theta_I\},$$

i.e., $\Theta_{I2}(\theta_1^*)$ is a collection of θ_2's such that (θ_1^*, θ_2) belongs to the identified set. This notation will be used repeatedly throughout this paper.

Let f_{1n} be the marginal quasi–posterior density for θ_1: $f_{1n}(\theta_1) = \int_{\theta_2 \in \Theta_2} f_n(\theta_1, \theta_2) d\theta_2$. Let F_{1n} be the distribution function of the marginal quasi–posterior. For any $\tau \in [0, 1]$, I define the τ–th quantile of the marginal quasi–posterior as

$$F_{1n}^{-1}(\tau) = \inf\{\theta_1 \in [\underline{\theta}_1, \overline{\theta}_1] : F_{1n}(\theta_1) \geq \tau\}.$$

3. ESTIMATION

I discuss the estimation of $\theta_{1\ell}$ and θ_{1u} in this section. Example A below (example 1 in CHT) illustrates the idea behind my estimator. In example A, the parameter is scalar–valued. My estimator is applicable when the parameters are vector–valued.

Example A (interval–observed data)**.** Suppose that there are i.i.d. random variables $\{(Y_{\ell i}, Y_i, Y_{ui})\}_{i=1}^n$. Assume that $Y_{\ell i}$, Y_i, and Y_{ui} have finite expectations and satisfy $Y_{\ell 1} \leq Y_1 \leq Y_{u1}$ a.s.. The parameter of interest is $\theta_0 = \mathbb{E}[Y_1]$. Researchers only observe $\{(Y_{\ell i}, Y_{ui})\}_{i=1}^n$. This model can be characterized by two moment inequalities

$$\mathbb{E}Y_{\ell 1} \leq \theta_0 \leq \mathbb{E}Y_{u1}.$$

In this model, θ_0 is not identified whereas the bounds of $\Theta_I = [\theta_\ell, \theta_u] = [\mathbb{E}Y_{\ell 1}, \mathbb{E}Y_{u1}]$ are identified. For the purpose of illustration, assume that $\theta_u > \theta_\ell$. One choice for the population and sample objective functions is

$$L(\theta) = -|\mathbb{E}Y_{\ell 1} - \theta|_+^2 - |\theta - \mathbb{E}Y_{u1}|_+^2, \quad L_n(\theta) = -|\bar{Y}_\ell - \theta|_+^2 - |\theta - \bar{Y}_u|_+^2.$$

Let Θ be a compact subset of $\mathbb{R}$ of which $[\theta_\ell, \theta_u]$ belongs to the interior. Consider an "infeasible" quasi–posterior[6]

$$f_\infty(\theta) = \lim_{n\to\infty} \frac{\exp(nL(\theta))}{\int_{t\in\Theta} \exp(nL(t))dt}.$$

It can be seen that $f_\infty(\theta) = 1/(\theta_u - \theta_\ell)$ if $\theta \in \Theta_I$ and $f_\infty(\theta) = 0$ otherwise. Hence the support of the "infeasible" quasi–posterior is exactly the identified set. Since the population objective function is unknown, I construct a "feasible" quasi–posterior using the sample objective function L_n:

$$f_n(\theta) = \frac{\exp(-n|\bar{Y}_\ell - \theta|_+^2 - n|\theta - \bar{Y}_u|_+^2)}{\int_{t\in\Theta} \exp(-n|\bar{Y}_\ell - t|_+^2 - n|t - \bar{Y}_u|_+^2)dt}.$$

By construction, the quasi–posterior density is maximized over the interval $[\bar{Y}_\ell, \bar{Y}_u]$ and declines towards both end points of Θ.

Figure 1 shows the shapes of f_∞ and f_n (renormalized such that the maximum of f_n equals $1/(\theta_u - \theta_\ell)$) for different sample size, with $Y_{\ell i} \sim U[-1, 0]$ and $Y_{ui} \sim U[1, 2]$. $\square$

As shown in fig. 1, the quasi–posterior concentrates on the identified set as the sample size increases. When $d > 1$, a similar pattern is expected for the marginal quasi–posterior f_{1n}. Example A, despite its simplicity, suggests a natural method of estimation for $\theta_{1\ell}$ and θ_{1u}. A quantile whose corresponding (quasi–posterior) probability level converges to zero (one) can be a candidate estimator for $\theta_{1\ell}$ (θ_{1u}).

Specifically, I propose the following estimators:

$$\hat{\theta}_{1\ell} = F_{1n}^{-1}(\hat{\tau}_\ell), \quad \hat{\theta}_{1u} = F_{1n}^{-1}(1 - \hat{\tau}_u).$$

Here $\hat{\tau}_\ell$ and $\hat{\tau}_u$ take values from $(0, 1)$; the choice of (quasi–posterior) probability levels $\hat{\tau}_\ell$ and $\hat{\tau}_u$ will be discussed later. The estimators $\hat{\theta}_{1\ell}$ and $\hat{\theta}_{1u}$ are just quantiles of the marginal quasi–posterior.

3.1. **Consistency.** In this subsection I will propose a way of choosing $\hat{\tau}_\ell$ and $\hat{\tau}_u$ such that $\hat{\theta}_{1\ell}$ and $\hat{\theta}_{1u}$ are consistent estimators regardless of the length of $[\theta_{1\ell}, \theta_{1u}]$. From here on I illustrate my approach and conditions needed for the case of $d = 2$ and I focus on estimation

[6]The limit may not exist (e.g. when the identified set has an empty interior), but it does exist in this example.

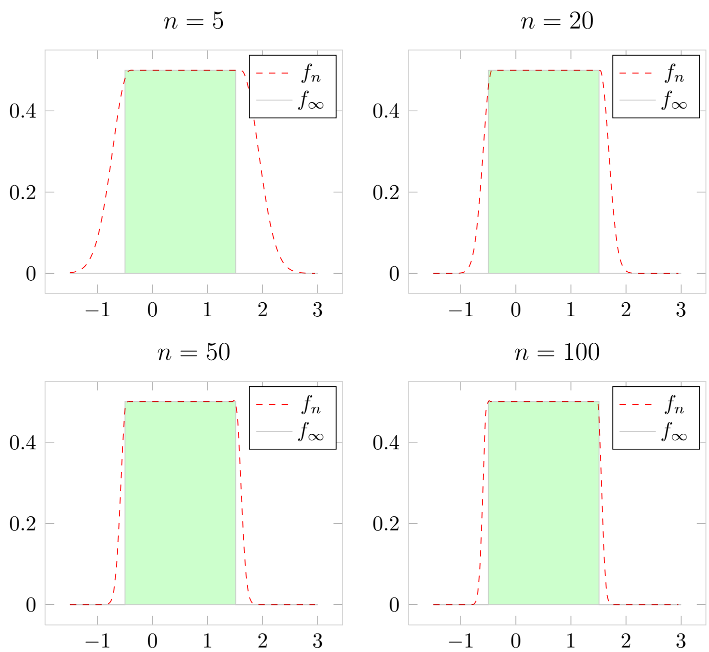

FIGURE 1. Quasi–posterior density (example A)

of $\theta_{1\ell}$. The upper bound θ_{1u} can be dealt with similarly. I provide formulas for generic d in appendix F.

Assumption 3.1. *Θ is compact.*

Assumption 3.1 is standard. It implies that $D_n = \int_{t \in \Theta} \exp(nL_n(t))dt$ is finite because $\exp(nL_n(\theta))$ is bounded uniformly in Θ.

Assumption 3.2. *Θ_I is connected.*

Connectedness says that Θ_I can not be represented by two or more disjoint subsets. It ensures that the projections of the identified set onto each axis are intervals.

For any $\theta \in \Theta$, let $d(\theta, \Theta_I) = \inf_{t \in \Theta_I} \|t - \theta\|$.

Assumption 3.3. *There exist constants $C > 0$ and $\delta > 0$ such that for all $\theta \in \Theta$*

$$\|\mathbb{E}m_1(\theta)\|_+ \geq \min\{Cd(\theta, \Theta_I), \delta\}.$$

Assumption 3.3 is also made in CHT. It requires that when parameters take values outside of the identified set, the expectations of the moments are at least proportional to the distance between the parameters and the identified set.

Assumption 3.4. *$\mathbb{E}m_1$ is Lipschitz continuous on Θ.*

Note that assumptions 3.1 and 3.4 implies that Θ_I is closed.

Let $\Delta_n(\theta) = \sqrt{n}\,(\bar{m}(\theta) - \mathbb{E}m_1(\theta))$. Let $\mathscr{L}^\infty(\Theta)$ be the set of functions which are uniformly bounded on Θ.

Assumption 3.5. *$\Delta_n(\theta)$ weakly converge to a Gaussian process $\Delta(\theta)$ on $\mathscr{L}^\infty(\Theta)$.*

Assumption 3.5 is also made in CHT. It requires that convergence in distribution for every $\theta \in \Theta$ and stochastic tightness of the process Δ_n (see section 2.1, Van der Vaart and Wellner, 1996). In example A, assumption 3.5 is satisfied if $\mathbb{E}Y_{\ell 1}^2$ and $\mathbb{E}Y_{u1}^2$ are finite. In example B below, assumption 3.5 is satisfied if $\mathbb{E}\|W_1\|^2$ is finite.

The (quasi–posterior) probability levels that I propose for my estimators are those using

$$\hat{\tau}_\ell = \hat{\tau}_u = \min\{\hat{c}/(nD_n), 1/2\}, \tag{3}$$

where $\hat{c} > 0$ is chosen by the researcher. Note that $\hat{\tau}_\ell$ is truncated at $1/2$. I introduce this truncation to accommodate the point identification case. I show in the proof of theorem 3.1 that $\hat{\tau}_\ell < 1/2$ with probability approaching one unless $\theta_{1\ell} = \theta_{1u}$. The truncation ensures that $\hat{\theta}_{1\ell} \leq \hat{\theta}_{1u}$.

Assumption 3.6. *$\hat{c} \xrightarrow{p} c$ for some $c \geq 0$. When $c = 0$, $\hat{c}$ converges at a polynomial rate.*

Theorem 3.1. *Let assumptions 3.1 to 3.6 hold. Then $\hat{\theta}_{1\ell} \xrightarrow{p} \theta_{1\ell}$.*

Proof. See appendix B.1. □

I have some comments on theorem 3.1. First, in the limit (when n goes to ∞), the quasi–posterior "will be concentrated on" Θ_I. Therefore, the marginal posterior "will be concentrated on" $[\theta_{1\ell}, \theta_{1u}]$, also. To achieve the consistency of my estimators for the $\theta_{1\ell}$ and θ_{1u}, I just "cut" two properly–sized tails off the marginal posterior. When the identified set has positive Lebesgue measure, D_n converges in probability to a positive constant, implying that $\hat{\tau}_\ell = \hat{c}_\ell/(nD_n) \xrightarrow{p} 0$. In this case, I essentially cut off two tails whose mass converges to zero.

Second, the rate requirement in assumption 3.6 can be relaxed in specific scenarios. For example, if one knows that the identified set has positive Lebesgue measure, one can allow that $\hat{c}$ diverges as long as $\hat{c}/n \xrightarrow{p} 0$ at a polynomial rate. Assumption 3.6 is stronger because it ensures the consistency of this estimator even when the identified set has an empty interior (i.e., Θ_I be a single point or other lower dimension subset of $\mathbb{R}^2$).

Third, for estimation, I do not assume that Θ_I belongs to the interior of Θ. When Θ_I intersects with the boundary of Θ, e.g., $\theta_{1\ell} = \underline{\theta}_1$ (the smallest value for the first dimension of the parameter space), my estimator $\hat{\theta}_\ell$ converges to $\theta_{1\ell}$ from above.

3.2. **Rate of convergence.** In this subsection, I will provide conditions under which my estimators are $\sqrt{n}$–consistent regardless of θ_0 being point or partially identified. The convergence rate is needed for constructing confidence sets for θ_{01}.

Example A continued. I first illustrate the idea of obtaining the rate using example A. If one chooses a probability level $\hat{\tau}_\ell$ in such a way that $F_n^{-1}(\hat{\tau}_\ell) - \bar{Y}_\ell = O_p(1/\sqrt{n})$, then since $\bar{Y}_\ell$ is a $\sqrt{n}$–consistent estimator of θ_ℓ, the quantile $F_n^{-1}(\hat{\tau}_\ell)$ will be a $\sqrt{n}$–consistent estimator also. In example A, it turns out that $F_n(\bar{Y}_\ell)$, which is the mass on the left tail of the quasi–posterior, decreases to zero at rate $1/\sqrt{n}$. Hence a choice of $\hat{\tau}_\ell \propto 1/\sqrt{n}$ ensures that $F_n^{-1}(\hat{\tau}_\ell)$ falls into a $\sqrt{n}$–neighborhood of θ_ℓ. $\square$

When θ_0 is a vector, then the appropriate choice of $\hat{\tau}_\ell$ depends on how fast the tail mass of the marginal posterior decreases to zero, which as I will illustrate in example B, depends crucially on the shape of the set $\Theta_{I2}(\theta_{1\ell}) = \{\theta_2 : (\theta_{1\ell}, \theta_2) \in \Theta_I\}$.

Example B (Linear moments)**.** Consider the following four linear moment inequalities,

$$\begin{aligned}
\mathbb{E}m_{(1)}(W_i,\theta) &= \mathbb{E}X_i\theta_1 + \mathbb{E}Y_i\theta_2 - 1 \le 0,\\
\mathbb{E}m_{(2)}(W_i,\theta) &= \mathbb{E}X_i\theta_1 - \mathbb{E}Y_i\theta_2 + 3 \le 0,\\
\mathbb{E}m_{(3)}(W_i,\theta) &= \theta_1 - \mathbb{E}Z_i \le 0,
\end{aligned}$$

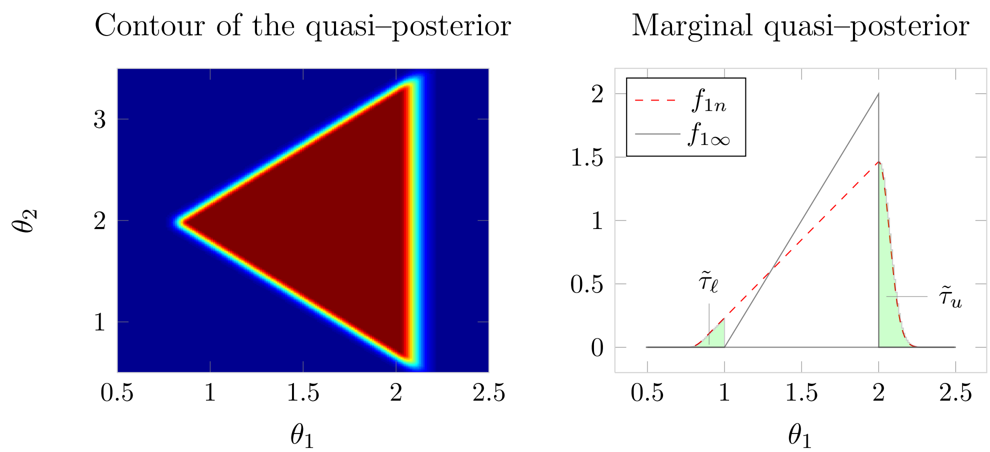

FIGURE 2. Marginal quasi–posterior density

where $\{W_i = (X_i, Y_i, Z_i)\}$ are i.i.d. observations such that $\mathbb{E}X_1 = -1$, $\mathbb{E}Y_1 = 1$, $\mathbb{E}Z_1 = 2$ and $\mathbb{E}\|W_1\|^2 < \infty$. Let $\Theta = [0.5, 2.5] \times [0.5, 3.5]$. I take $\theta_0 = (1, 2)'$. X_i, Y_i, and Z_i are independent. Let $\Delta_{nW} = \sqrt{n}(\bar{W} - \mathbb{E}W)$ and Δ_W be the (distributional) limit of Δ_{nW}.

In this example, the identified set is a triangle. Researchers are interested in the parameter θ_{01} and the projection of the identified set on the axis of θ_1: $\Theta_{I1} = [1, 2]$. $\square$

Figure 2 shows the quasi–posteriors as well as the marginal quasi–posteriors for example B, with sample size $n = 100$. Note that the tail masses $\tilde{\tau}_\ell$ and $\tilde{\tau}_u$ of marginal quasi–posterior outside of the projection of the identified set are very different: $\tilde{\tau}_\ell$ is a lot smaller than $\tilde{\tau}_u$. It turns out that on the left end, because $\Theta_{I2}(\theta_{1\ell})$ only contains a single point, the tail mass decreases at rate n. On the right end, $\Theta_{I2}(\theta_{1u})$ contains an interval with positive length, the tail mass decreases at rate of root–n. Before formally stating this result in lemma 3.1, I make a few more assumptions.

Assumption 3.7. *For all $\theta \in \Theta$, $\mathbb{E}m_1(\theta)$ is continuously differentiable.*

Let $Q(\theta)$ be the $J \times d$ derivative matrix of $\mathbb{E}m_1$ evaluated at θ. Let $\mathcal{J}(\theta) \subseteq \{1, 2, \cdots, J\}$ be the set of indices of binding moments at θ and $\mathbb{E}m_1^{\mathcal{J}}(\theta)$ be the subvector of expectations of binding moments, i.e., $\mathbb{E}m_1^{\mathcal{J}}(\theta) = 0$. Let $Q^{\mathcal{J}'}(\theta) = \partial \mathbb{E}m_1^{\mathcal{J}}(\theta)/\partial\theta$. $\Delta_n^{\mathcal{J}}(\theta)$ and $\Delta^{\mathcal{J}}(\theta)$ are similarly defined. Let $Q_1(\theta)$ be the first column of $Q(\theta)$.

Note that by assumptions 3.3 and 3.7, there exist positive constants k and K such that for any θ_2^* belonging to the interior of $\Theta_{I2}(\theta_{1\ell})$, the absolute value of every component of $Q_1^{\mathcal{J}}(\theta_{1\ell}, \theta_2^*)$ takes values in $[k, K]$. In addition, assumption 3.3 ensures that each row and each column of $Q^{\mathcal{J}}(\theta)$ has at least one non–zero element.

Assumption 3.8. *When $\Theta_{I2}(\theta_{1\ell}) = \{\theta_\ell\}$, $Q^{\mathcal{J}}(\theta_\ell)$ has full column rank.*

Assumption 3.8 is crucial to ensure that the quasi–posterior decline sufficiently fast within a $\sqrt{n}$ local neighborhood of the corner point θ_ℓ. When all the moment equations are linear and $\Theta_{I2}(\theta_{1\ell})$ is a singleton, assumption 3.8 implies that $J \geq d$ and that there are no more than $J - d$ moment equations be proportional to each other. A full column rank condition is also assumed in PPHI.

Let $\text{int}(\Theta_I)$ be the interior of Θ_I and $\partial(\Theta_I)$ be the boundary of Θ_I. For $\delta > 0$ and $\theta \in \Theta$, let $B_\theta^\delta = \{t \in \Theta : \|t - \theta\| \leq \delta\}$.

Assumption 3.9. *Θ_I satisfies either (1) or (2) below.*

(1) *Θ_I is convex.*

(2) *For any $\theta \in \partial(\Theta_I)$ and any $\delta > 0$, $B_\theta^\delta \cap int(\Theta_I)$ is not empty.*

Convexity is required in Bontemps, Magnac, and Maurin (2007), Beresteanu and Molinari (2008), Kaido (2010). One can impose assumptions on the moment equations, e.g. linear moment equations, to ensure the convexity. When convexity does not satisfies, my analysis goes through provided condition 3.9–(2) holds. I will discuss more on this assumption after theorem 3.2.

Assumption 3.10. *$\hat{c} \xrightarrow{p} c^*$ for some $c^* > 0$.*

To show consistency of my estimators, I allow $\hat{c} \xrightarrow{p} 0$ at polynomial rate. Here, I require that $\hat{c}$ converges in probability to a positive constant.

Lemma 3.1 shows that to obtain $\sqrt{n}$–consistency, the choice of $\hat{\tau}_\ell$ depends on the shape of $\Theta_{I2}(\theta_{1\ell})$. Let $\hat{\tau}_\ell^a = \min\{\hat{c}/(nD_n), 1/2\}$ and $\hat{\tau}_\ell^b = \min\{\hat{c}/(\sqrt{n}D_n), 1/2\}$. Note that $\hat{\tau}_\ell^a$ and $\hat{\tau}_\ell^b$ have different rates.

Lemma 3.1. *Suppose that assumptions 3.1 to 3.5 and 3.7 to 3.10 are satisfied.*

(1) *If $\Theta_{I2}(\theta_{1\ell})$ contains an interval of positive length, then*

   (a) *for any $K > 0$, $\lim_{n\to\infty} \mathbb{P}(\sqrt{n}(\theta_{1\ell} - F_{1n}^{-1}(\hat{\tau}_\ell^a)) > K) = 1$ whenever $\theta_{1\ell} > \underline{\theta}_1$.*

   (b) $\sqrt{n}(\theta_{1\ell} - F_{1n}^{-1}(\hat{\tau}_\ell^b)) = O_p(1)$.

(2) *If $\Theta_{I2}(\theta_{1\ell})$ contains only singletons, then $\sqrt{n}(\theta_{1\ell} - F_{1n}^{-1}(\hat{\tau}_\ell^a)) = O_p(1)$.*

Proof. See appendix B.2. □

The implication of lemma 3.1 is that one has to choose different quantiles for different cases to obtain $\sqrt{n}$–consistency. The choice depends on the unknown shape of $\Theta_{I2}(\theta_{1\ell})$. So it is desirable to construct an estimator $\hat{\theta}_{1\ell}^*$ which can automatically adapt to the shape of $\Theta_{I2}(\theta_{1\ell})$. This is feasible because the quasi–posterior provides corresponding information. To see this, consider an infeasible version $\tilde{\theta}_{1\ell}^*$ of the estimator $\hat{\theta}_{1\ell}^*$

$$\tilde{\theta}_{1\ell}^* = F_{1n}^{-1}\left(\hat{\tau}_\ell(\theta_{1\ell})\right), \quad \text{with} \quad \hat{\tau}_\ell(\theta_{1\ell}) = \min\left\{\frac{\hat{c}U_n(\theta_{1\ell})}{nD_n}, \frac{1}{2}\right\},$$

where

$$U_n(\theta_1) = \sqrt{n}\int_{\theta_2} \exp(-n\|\bar{m}(\theta_1,\theta_2)\|_+^2)d\theta_2.$$

By construction, $\sup_{\theta_1} U_n(\theta_1) \le C\sqrt{n}$ for some $C > 0$. It can be shown (in the proof of theorem 3.2) that when $\Theta_{I2}(\theta_{1\ell})$ is a singleton, $U_{1n}(\theta_{1\ell}) = O_p(1)$, in which case I am essentially using a probability level decreasing at rate of n; whereas if $\Theta_{I2}(\theta_{1\ell})$ is an interval, $U_{1n}(\theta_{1\ell}) = O_p(\sqrt{n})$ and I am using a probability level decreasing at rate root–n. The quantity $U_{1n}(\theta_{1\ell})$ hence picks out the correct rate.

$U_{1n}(\theta_{1\ell})$ is unknown because it depends on $\theta_{1\ell}$. I define a feasible version of $\tilde{\theta}_{1\ell}^*$

$$\hat{\theta}_{1\ell}^* = F_{1n}^{-1}\left(\hat{\tau}_\ell(\hat{\theta}_{1\ell}^*)\right), \quad \hat{\tau}_\ell(\hat{\theta}_{1\ell}^*) = \min\left\{\frac{\hat{c}U_n(\hat{\theta}_{1\ell}^*)}{nD_n}, \frac{1}{2}\right\}. \tag{4}$$

If there are multiple solutions to eq. (4), I choose an arbitrary one.

Theorem 3.2. *Suppose that assumptions 3.1 to 3.3, 3.5 and 3.7 to 3.10 are satisfied. Let $\hat{\theta}_{1\ell}^*$ be constructed from eq.* (4)*, then $\sqrt{n}(\hat{\theta}_{1\ell}^* - \theta_{1\ell}) = O_p(1)$.*

Proof. See appendix B.3. □

Under assumption 3.7, the marginal quasi–posterior allows for an expansion within in a $\sqrt{n}$–neighborhood of $\theta_{1\ell}$, hence I can focus on a $\sqrt{n}$–localized marginal quasi–posterior. Assumption 3.8 ensures that when $\Theta_{I2}(\theta_{1\ell})$ is a singleton, the localized quasi–posterior (outside of the identified set) around θ_ℓ is integrable. With assumptions 3.7 and 3.8, my estimator $\hat{\theta}^*_{1\ell}$ can not "underestimate" $\theta_{1\ell}$ more than an order of $1/\sqrt{n}$.

Assumption 3.9, on the other hand, ensures that $\hat{\theta}^*_{1\ell}$ does not "overestimate" $\theta_{1\ell}$ more than an order of $1/\sqrt{n}$. To see this, note that assumption 3.9–(2) implies that there exists a subset of the interior of Θ_I with positive Lebesgue measure, and on which the quasi–posterior density takes positive values in large sample. Because the quasi–posterior probability level associate with $\hat{\theta}^*_{1\ell}$ eventually converges to zero, $\hat{\theta}^*_{1\ell}$ can not exceed $\theta_{1\ell}$ in large sample. Assumption 3.9–(1) plays a similar role.

4. Inference

The next question that I address is how to choose two quantiles of the marginal quasi–posterior such that the resulting interval covers θ_{01} with a prespecified probability asymptotically. In section 4.1, I propose such a procedure which takes care of several issues. First, my confidence set covers θ_{01} with prespecified probability regardless of the shape of $\Theta_{I2}(\theta_{1\ell})$. Second, θ_{01} may be point identified or partially identified; my procedure accommodates both cases. Third, I use a weighting method to pick out the binding moments asymptotically. An algorithm for constructing confidence sets can be found in appendix D.1.

In section 4.2, I show that my confidence set contains any fixed alternative with probability approaching zero and has nontrivial power against local alternatives. My method can be used to construct confidence sets for scalar–valued functions of the true parameter vector, which will be discussed in section 4.3. I construct a confidence set for $[\theta_{1\ell}, \theta_{1u}]$ in section 4.4.

There are other ways of constructing confidence sets for θ_{01}. One way is to construct a confidence set for θ_0, then take the first dimension as the confidence set for θ_{01}. In section 4.5, I provide an example showing that the projection of a high–dimensional confidence set is conservative.

4.1. **Construct confidence sets for θ_{01}.**

4.1.1. *An infeasible confidence set.* It is convenient to introduce an infeasible confidence set $\Theta^{\mathrm{I}}_{\alpha n}$ first; I will propose a feasible confidence set $\widehat{\Theta}^{\mathrm{I}}_{\alpha}$ in section 4.1.2.

The inference about θ_{01} is based on the following observation. Suppose that $\theta_{1\ell}$ is in the interior of $\tilde{\Theta}_1$. Then for some properly chosen $\alpha_{\ell n} \to \infty$, $\alpha_{\ell n} D_n F_{1n}(\theta_{1\ell})$ converges in distribution to a continuous random variable ξ_ℓ (this will be formally shown later). I can then use the distributional information of ξ_ℓ to construct confidence sets for $\theta_{1\ell}$.

To illustrate the idea, suppose that the object of interest is a one–sided confidence set for $\theta_{1\ell}$. Let $c_\ell(\alpha)$ be the α–th quantile of ξ_ℓ, then

$$\begin{aligned}\mathbb{P}\left\{\theta_{1\ell} \geq F_{1n}^{-1}\left(\frac{c_\ell(\alpha)}{\alpha_{\ell n} D_n}\right)\right\} &= \mathbb{P}\left\{\alpha_{\ell n} D_n F_{1n}(\theta_{1\ell}) \geq c_\ell(\alpha)\right\} \\ &= \mathbb{P}\{\xi_\ell \geq c_\ell(\alpha)\} + o(1) = 1 - \alpha + o(1).\end{aligned}$$

Thus a quantile of the marginal quasi–posterior f_{1n} serves as the boundary point of a one–sided confidence set for $\theta_{1\ell}$. This idea can be extended to construct confidence sets for θ_{01}.

There are some difficulties. First, depending on the shape of $\Theta_{I2}(\theta_{1\ell})$, one needs to choose different rates for $\alpha_{\ell n}$ such that the distributional limit of $\alpha_{\ell n} D_n F_{1n}(\theta_{1\ell})$ exists and is nondegenerate. Second, the distributional limits of $\alpha_{\ell n} D_n F_{1n}(\theta_{1\ell})$ may be different for different shapes of $\Theta_{I2}(\theta_{1\ell})$. In principle, $\Theta_{I2}(\theta_{1\ell})$ could be a union of finite number of singletons and intervals. I make a following assumption to simplify the presentation of the theocratic results in this section.

Assumption 4.1. $\Theta_{I2}(\theta_{1\ell})$ *is contains at most one interval and one singleton.*

Assumption 4.1 is satisfied when Θ_I is convex. Theocratic results in this section still holds without this assumption (with complicated notation).

I require that the first dimension of the identified set belongs to the interior of the first dimension of the parameter space.

Assumption 4.2. $[\theta_{1\ell}, \theta_{1u}]$ *belongs to the interior of* Θ_1.

Lemma 4.1. *Suppose that assumptions 3.1 to 3.3, 3.5, 3.7, 3.8 and 4.2 are satisfied. If* $\Theta_{I2}(\theta_{1\ell}) = \{\theta_{2\ell}\}$*, then*

$$nD_nF_{1n}(\theta_{1\ell}) \xrightarrow{d} \xi_\ell^A = \int_{h_1 \leq 0} \int_{h_2 \in \mathbb{R}} \exp\left(-\|\Delta^{\mathcal{J}}(\theta_\ell) + Q^{\mathcal{J}}(\theta_\ell)h\|_+^2\right) dh. \tag{5}$$

If $\Theta_{I2}(\theta_{1\ell}) = [\theta_{2\ell}, \theta_{2u}]$*, then*

$$\sqrt{n}D_nF_{1n}(\theta_{1\ell}) \xrightarrow{d} \xi_\ell^B = \int_{h_1 \leq 0} \int_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \exp\left(-\|\Delta^{\mathcal{J}}(\theta_{1\ell}, \theta_2) + Q_1^{\mathcal{J}}(\theta_{1\ell}, \theta_2)h_1\|_+^2\right) d\theta_2 dh_1. \tag{6}$$

Proof. See appendix C.1. □

The message from lemma 4.1 is similar to that from lemma 3.1. In lemma 3.1, one needs to choose different rates for probability levels according to the shape of $\Theta_{I2}(\theta_{1\ell})$ to achieve $\sqrt{n}$–consistency of the boundary estimators. Here, I need to use different rescaling and different limiting distributions accordingly to construct confidence sets for θ_{01}.

I propose to choose $\alpha_{\ell n}$ as follow:

$$\alpha_{\ell n} = \omega_{\ell n} n + (1 - \omega_{\ell n})\sqrt{n}, \quad \omega_{\ell n} = \nu\left(\frac{M_n(\theta_{1\ell})}{\log n}\right), \tag{7}$$

where $\nu(x) = \phi(x)/\phi(0)$, ϕ is the standard normal density and for some $\beta_n \to \infty$,

$$M_n(\theta_1) = \sqrt{\beta_n} \int \exp(-\beta_n \|\bar{m}(\theta_1, \theta_2)\|_+^2) d\theta_2. \tag{8}$$

So $\alpha_{\ell n}$ is defined as an weighted average of $\sqrt{n}$ and n; the weights depends on the behavior of a quantity $M_n(\theta_\ell)$. Similarly, I consider a random variable $\xi_{\ell n}$ defined as

$$\xi_{\ell n} = \omega_{\ell n}\psi_n\left(\xi_\ell^A\right) + (1 - \omega_{\ell n})\xi_\ell^B, \tag{9}$$

where $\psi_n(x) = \min\{x, \log n\}$. Again, $\xi_{\ell n}$ is a weighted average based on the same weight. The truncation introduced by ψ takes care of the possibility that ξ_ℓ^A explode when $\Theta_{I2}(\theta_{1\ell})$ is an interval.

I make one assumption on the rate of β_n. The rate conditions are sufficient to ensure that the weight $\omega_{\ell n}$ converges to one or zero depending on the shape of $\Theta_{I2}(\theta_{1\ell})$. Hence the weight $\omega_{\ell n}$ automatically select (asymptotically) correct rescaling parameter and limiting distribution (see lemma C.2 for details).

Assumption 4.3. *$\beta_n/(\sqrt{n}\log n) \to \infty$ and $\beta_n/n \to 0$.*

The motivation for the definitions in eqs. (7) and (12) is as follows. If $\Theta_{I2}(\theta_{1\ell})$ is a singleton, I know that $M_n(\theta_{1\ell})/\log n \xrightarrow{p} 0$ and $\omega_{\ell n} \xrightarrow{p} 1$, in which case I rescale $D_n F_{1n}(\theta_{1\ell})$ by $\alpha_{\ell n} = n$ and use ξ_ℓ^A . On the other hand, if $\Theta_{I2}(\theta_{1\ell})$ is an interval, $M_n(\theta_{1\ell})/\log n$ diverges in probability and $\omega_{\ell n} \xrightarrow{p} 0$ exponentially. In this case I rescale $D_n F_{1n}(\theta_{1\ell})$ by $\alpha_{\ell n} = \sqrt{n}$ and use distribution ξ_ℓ^B.

There are other issues I take care of to correctly specify the critical value of the random variable $\xi_{\ell n}$. First, the distribution of ξ_ℓ^A and ξ_ℓ^B depends on the identities of the binding moments. I use a weighting procedure to pick out the binding moments. For each θ, let $\gamma_n(\theta)$ be a sample–size–dependent J–vector of weights with j–th component

$$\gamma_{jn}(\theta) = \frac{\exp(-\beta_n|\frac{1}{n}\sum_{i=1}^n m_{(j)}(W_i,\theta)|^2)}{\exp(-\beta_n|\frac{1}{n}\sum_{i=1}^n m_{(j)}(W_i,\theta)|_+^2)}. \tag{10}$$

By construction, $\gamma_{jn}(\theta) \in (0,1]$. As is shown in lemma C.1, for any $\theta^* \in \Theta_I$, the weight $\gamma_{jn}(\theta^*)$ converges to one if the j–th moment is binding at θ^*; $\gamma_{jn}(\theta^*)$ converges to zero otherwise.[7] Now I consider a weights–adjusted version of $\xi_{\ell n}$,

$$\tilde{\xi}_{\ell n} = \omega_{\ell n}\psi_n\left(\tilde{\xi}_\ell^A\right) + (1-\omega_{\ell n})\tilde{\xi}_\ell^B, \tag{11}$$

where

$$\begin{cases} \tilde{\xi}_\ell^A = \int_{h_1\le 0}\int_{h_2\in\mathbb{R}} \exp(-\gamma_n'(\theta_\ell))\|\Delta(\theta_\ell) + Q(\theta_\ell)h\|_+^2)d\theta_2 dh_1 \\ \tilde{\xi}_\ell^B = \int_{h_1\le 0}\int_{\theta_2\in[\theta_{2\ell},\theta_{2u}]} \exp\left(-\gamma_n'(\theta_{1\ell},\theta_2)\|\Delta(\theta_{1\ell},\theta_2) + Q_1(\theta_{1\ell},\theta_2)h_1\|_+^2\right) d\theta_2 dh_1 \end{cases} \tag{12}$$

Let $\tilde{\xi}_{un}$ and α_{un} be defined in a similar way.

[7] When the j–th moment is nearly binding at θ^*, i.e., when $\mathbb{E}m_{(j)}(\theta^*) = \lambda/\sqrt{n}$ for some $\lambda \in (-\infty, 0)$, $\gamma_{jn}(\theta^*)$ converges in distribution to a random variable takes value from $(0,1)$. In this paper, I do not consider this case.

To make inferences about θ_{01}, one also needs to be careful about the length of the interval $[\theta_{1\ell}, \theta_{1u}]$. Let $T = \theta_{1u} - \theta_{1\ell}$ and $\bar{T}$ be the largest value that T can take. Let $\hat{T}$ be a $\sqrt{n}$–consistent estimator for T. I construct a confidence interval for θ_{01} as

$$\Theta^{\mathrm{I}}_{\alpha n} = \left[F^{-1}_{1n}\left(\tau^{\mathrm{I}}_{\ell}\right), F^{-1}_{1n}\left(1-\tau^{\mathrm{I}}_{u}\right)\right],$$

where $\tau^{\mathrm{I}}_{\ell} = c^{\mathrm{I}}_{\ell}/(\alpha_{\ell n} D_n)$, $\tau^{\mathrm{I}}_{u} = c^{\mathrm{I}}_{u}/(\alpha_{un} D_n)$, and $(c^{\mathrm{I}}_{\ell}, c^{\mathrm{I}}_{u})$ is a solution to the following problem:[8]

$$\begin{aligned}(c^{\mathrm{I}}_{\ell}, c^{\mathrm{I}}_{u}) = &\underset{(c_\ell, c_u)\in\mathbb{R}^+\times\mathbb{R}^+}{\operatorname{argmin}} \left|F^{-1}_{1n}\left(\frac{c_\ell}{\alpha_{\ell n} D_n}\right) - F^{-1}_{1n}\left(1-\frac{c_u}{\alpha_{un} D_n}\right)\right| \\ \text{s.t.}\quad &\mathbb{P}\left\{c_\ell \le \tilde{\xi}_{\ell n}, c_u \le \sqrt{n}\nu\left(\frac{\beta_n}{n\hat{T}}\right) + \tilde{\xi}_{un}\right\} = 1-\alpha, \\ &\mathbb{P}\left\{c_u \le \tilde{\xi}_{un}, c_\ell \le \sqrt{n}\nu\left(\frac{\beta_n}{n\hat{T}}\right) + \tilde{\xi}_{\ell n}\right\} = 1-\alpha,\end{aligned} \tag{13}$$

When $T > 0$, $\sqrt{n}\nu\left(\beta_n/n\hat{T}\right)$ diverges to infinity, in which case c^{I}_{ℓ} and c^{I}_{u} are computed as $1-\alpha$ quantiles of $\tilde{\xi}_{\ell n}$ and $\tilde{\xi}_{un}$ respectively; when T is zero, $\sqrt{n}\nu\left(\beta_n/n\hat{T}\right)$ converges in probability to zero, in which case the confidence set is constructed using the joint distribution of $\tilde{\xi}_{\ell n}$ and $\tilde{\xi}_{un}$.

Theorem 4.1. *Suppose that assumptions 3.1 to 3.3, 3.5, 3.7, 3.8 and 4.1 to 4.3 are satisfied. Then*

$$\lim_{n\to\infty} \inf_{\theta_{01}\in[\theta_{1\ell}, \theta_{1u}]} \mathbb{P}(\theta_{01} \in \Theta^{\mathrm{I}}_{\alpha n}) = 1-\alpha.$$

Proof. See appendix C.2. □

I have several comments on theorem 4.1. First, $\Theta^{\mathrm{I}}_{\alpha n}$ is constructed directly from the marginal quasi–posterior rather than as a projection of a high–dimensional confidence set. Second, my weighting method picks out the binding moments asymptotically. Third, I introduce an additional "shrinkage term" to accommodate the point identification case. Fourth, the critical values c^{I}_{ℓ} and c^{I}_{u} are computed from the joint distribution of $\xi_{\ell n}$ and ξ_{un} to take care of the possible correlation between them.

[8] If there are multiple solutions to eq. (13), I take an arbitrary one.

I use example C to illustrate how my procedure accommodates the point identification case.

Example C. Suppose that θ_0 is the unique value for which $\mathbb{E}m_{(j)}(W_1, \theta_0) = 0$ for $j = 1, 2$, where $\theta_0 \in \Theta \subset \mathbb{R}^2$. Suppose in addition that the moment equations and the parameter space Θ satisfy all the conditions required by theorem 4.1. In this case,

$$\begin{cases} \xi_\ell^A = \int_{h_1 \le 0} \int_{h_2 \in \mathbb{R}} \exp(-\|\Delta(\theta_0) + Q(\theta_0)h\|^2) dh \\ \xi_u^A = \int_{h_1 \ge 0} \int_{h_2 \in \mathbb{R}} \exp(-\|\Delta(\theta_0) + Q(\theta_0)h\|^2) dh. \end{cases} \tag{14}$$

Write Q_0 for $Q(\theta_0)$ and Δ_0 for $\Delta(\theta_0)$. Let $\mathcal{U} = -\frac{1}{2}(Q_0'Q_0)^{-1}Q_0'\Delta_0$ and $\mathcal{U}_1$ be the first component of $\mathcal{U}$. Let V_0 be the variance of Δ_0. Assuming further that $Q_0'V_0Q_0 = 2(Q_0'Q_0)$, it then follows that $\mathcal{U}$ is a mean zero bivariate normal with variance $\Sigma = \frac{1}{2}(Q_0'Q_0)^{-1}$.[9]

With some algebra, it can be shown that c_ℓ^{I} and c_u^{I} need to be chosen such that

$$\mathbb{P}(\xi_\ell^A \ge c_\ell^{\mathrm{I}}, \xi_u^A \ge c_u^{\mathrm{I}}) = \mathbb{P}\left[\Phi_{\Sigma_{11}}(\mathcal{U}_1) \ge c_\ell^{\mathrm{I}}/C, 1 - \Phi_{\Sigma_{11}}(\mathcal{U}_1) \ge c_u^{\mathrm{I}}/C\right] = 1 - \alpha,$$

where $\Phi_{\Sigma_{11}}$ is the distribution function for $\mathcal{N}(0, \Sigma_{11})$ and Σ_{11}, Σ_{11} is the first diagonal element of Σ and C is a constant given by $C = 2\pi|\Sigma|^{1/2}$. It is not difficult to verify that $\Phi_{\Sigma_{11}}(\mathcal{U}_1)$ is uniformly distributed on the unit interval. By observing that $nD_n \to C$, I can conclude that $c_\ell^{\mathrm{I}}/nD_n \xrightarrow{p} \alpha/2$ and $c_u^{\mathrm{I}}/nD_n \xrightarrow{p} \alpha/2$. In other words, the confidence set is essentially an interval introduced by $\alpha/2$ and $1 - \alpha/2$ quantiles of the marginal quasi–posterior. $\square$

4.1.2. *Constructing* $\widehat{\Theta}_\alpha^{\mathrm{I}}$. The confidence set $\Theta_{\alpha n}^{\mathrm{I}}$ is infeasible because the joint distribution of $\tilde{\xi}_{\ell n}$ and $\tilde{\xi}_{un}$ is unknown; as a result, c_ℓ^{I} and c_u^{I} are unknown. In this subsection I propose an algorithm obtaining consistent estimates for c_ℓ^{I} and c_u^{I}. I highlight the big picture of the procedure in the main text ; the detailed algorithm is in appendix D.1 (algorithm 1). Throughout this subsection, I assume that there are $\sqrt{n}$–consistent estimates for $\theta_{1\ell}$ and θ_{1u}, which can be computed from eq. (4).

[9] I.e., the "generalized information equality" in CH is satisfied.

Note that there are several unknown parts in the expression for $\tilde{\xi}_{\ell n}$: $\Delta(\theta_{1\ell}, \theta_2)$, $\gamma_n(\theta)$, $Q_1(\theta_{1\ell}, \theta_2)$, $\alpha_{\ell n}$ and θ_ℓ in the expression of $\tilde{\xi}^A_\ell$. I discuss how to obtain feasible versions of them in turn.

$\Delta(\theta_{1\ell}, \theta_2)$ is a J–dimensional Gaussian process. For each j, j', and θ_2 and θ_2', the covariance between $\Delta_j(\theta_{1\ell}, \theta_2)$ and $\Delta_{j'}(\theta_{1\ell}, \theta_2')$ is

$$H_{j,j',\theta_2,\theta_2'} = \mathbb{E}m_{(j)}(W_1; \theta_{1\ell}, \theta_2)m_{(j')}(W_1; \theta_{1\ell}, \theta_2') - \mathbb{E}m_{(j)}(W_1; \theta_{1\ell}, \theta_2)\mathbb{E}m_{(j')}(W_1; \theta_{1\ell}, \theta_2').$$

It can be estimated by

$$\hat{H}_{j,j',\theta_2,\theta_2'} = \frac{1}{n}\sum_{i=1}^{n} m_{(j)}(W_i; \hat{\theta}_{1\ell}, \theta_2)m_{(j')}(W_i; \hat{\theta}_{1\ell}, \theta_2') - \bar{m}_{(j)}(\hat{\theta}_{1\ell}, \theta_2)\bar{m}_{(j')}(\hat{\theta}_{1\ell}, \theta_2').$$

$Q_1(\theta_{1\ell}, \theta_2)$ can be consistently estimated by $\hat{Q}_1 = \partial\bar{m}(\theta)/\partial\theta|_{\theta=(\hat{\theta}_{1\ell},\theta_2)}$.

I replace $\gamma_n(\theta_{1\ell}, \theta_2)$ with $\hat{\gamma}(\hat{\theta}_{1\ell}, \theta_2)$, where $\theta_{1\ell}$ is replaced by $\hat{\theta}_{1\ell}$ in eq. (10). Similarly, $\omega_{\ell n}$ is replaced by $\hat{\omega}_\ell = \nu_n(M_n(\hat{\theta}_{1\ell})/\log n)$.

When $\Theta_{I2}(\theta_{1\ell})$ is a singleton, θ_ℓ can be estimated by $\hat{\theta}_\ell = (\hat{\theta}_{1\ell}, \hat{\theta}_2)'$, where

$$\hat{\theta}_2 = \int_{\theta_2\in\Theta_2} \theta_2 f_n(\hat{\theta}_{1\ell}, \theta_2)d\theta_2.$$ [10]

Then I construct $\widehat{\Theta}^{\mathrm{I}}_\alpha$ by replacing the unknown parts with their estimates in eq. (13). I add assumptions 4.4 and 4.5 to ensure that $\hat{Q}$ and $\hat{H}$ converge uniformly to their population counterparts. Assumptions 4.4 and 4.5 imply that assumptions 3.4, 3.5 and 3.7 are satisfied.

Assumption 4.4. *For each $w \in \mathscr{W}$, $m(w, \theta)$ is continuously differentiable at each $\theta \in \Theta$. There exists a $d(w)$ such that $\|\partial m(w, \theta)/\partial\theta\| < d(w)$ for all $\theta \in \Theta$ and $\mathbb{E}d(W_1) < \infty$.*

Assumption 4.5. $\mathbb{E}\|m_1(\theta)\|^2 < \infty$ *for all θ.*

Theorem 4.2. *Suppose that assumptions 3.1 to 3.3, 3.8, 3.9 and 4.1 to 4.5 are satisfied, then*

$$\lim_{n\to\infty} \inf_{\theta_{01}\in[\theta_{1\ell},\theta_{1u}]} \mathbb{P}(\theta_{01} \in \widehat{\Theta}^{\mathrm{I}}_\alpha) = 1 - \alpha.$$

Proof. See appendix C.3. □

[10] If $\Theta_{I2}(\theta_{1\ell})$ contains more than one singletons, I need to estimate each of them. In this case, the marginal quasi–posterior $f_n(\theta_{1\ell}, \theta_2)$ of θ_2 have more than mode. Computation will be much harder.

Constructing $\widehat{\Theta}^{\mathrm{I}}_{\alpha}$ does not require resampling procedures. Instead, I just need to obtain random draws from the joint distribution of $\tilde{\xi}_{\ell n}$ and $\tilde{\xi}_{un}$. As shall be clear in algorithm 1 (appendix D.1), each random draw involves only computing a d–dimensional integral whose integrand is a parametric function of a Gaussian process.

4.2. **Local power.** Note that the end points of the confidence set $\Theta^{\mathrm{I}}_{\alpha n}$ are essentially $\sqrt{n}$–consistent estimators for $\theta_{1\ell}$ and θ_{1u}, so it follows that my confidence set contains any fixed alternative with probability approaching zero. In this section, I will analyze the local power property of my inference procedure. I consider local alternatives of the following form: $\theta_{1n} = \theta_{1\ell} - h_1^*/\sqrt{n}$, where $h_1^* \in \mathbb{R}_+$. The case where $\theta_{1n} = \theta_u + h_1^*/\sqrt{n}$ is similar. The object is to derive a lower bound of power function $f_h : \mathbb{R}_+ \to [0,1]$ defined by

$$f_h(h_1^*) = \lim_{n\to\infty} \mathbb{P}(\theta_{1n} \notin \Theta^{\mathrm{I}}_{\alpha n}).$$

Theorem 4.3 shows that my confidence set has non–trivial local power against the $\sqrt{n}$–local alternatives.

Theorem 4.3. *Suppose that assumptions 3.1 to 3.3, 3.5, 3.7, 3.8 and 4.1 to 4.3 are satisfied. Then f_h is non–decreasing and*

$$\lim_{h_1^*\to+\infty} f_h(h_1^*) = 1.$$

Proof. See appendix C.4. □

4.3. **Constructing a confidence set for a scalar–valued function of θ_0.** Researchers may also be interested in $g(\theta_0)$, where $g : \Theta \to \mathbb{R}$ is a scalar–valued function. When $g(\theta_0) = \iota'_{(j)}\theta_0$, where $\iota_{(j)}$ is a vector whose j–th component is one and others are zeros, making inferences about $g(\theta_0)$ is equivalent to making an inference about θ_{0j}, which has been discussed in the previous section.

I take a reparameterization approach to construct a confidence set for $g(\theta_0)$. Let

$$\Theta^g = \{(\theta_1^g, \theta_2^g) : \theta \in \Theta, \theta_2^g = \theta_2, \theta_1^g = g(\theta)\}, \quad \Theta_I^g = \{(\theta_1^g, \theta_2^g) : \theta \in \Theta_I, \theta_2^g = \theta_2, \theta_1^g = g(\theta)\}.$$

Let $\theta_{01}^g = g(\theta_0)$ and $\theta_0^g = (\theta_{01}^g, \theta_{02})$.

Assumption 4.6. *g is bounded and continuously differentiable on Θ. The first order derivative is bounded on Θ.*

Assumption 4.7. *For all $\theta_2 \in \Theta_2$, $g(\cdot, \theta_2)$ is strictly monotone in its first argument. For all $\theta_1 \in \Theta_1$, g is weakly monotone in each elements of θ_2.*

Assumptions 4.6 and 4.7 are satisfied when g is linear.

For any K, let $\mathcal{C}^K = \{\theta \in \Theta : g(\theta) = K\}$ and $\mathcal{D}^K = \partial(\Theta_I) \cap \mathcal{C}^K$. Let $\mathcal{D}_2^K = \{\theta_2 : \exists \theta_1 : (\theta_1, \theta_2) \in \mathcal{D}^K\}$.

Assumption 4.8. *One of the following conditions hold.*

(1) *$\mathcal{D}_2^K$ has positive measure in $\mathbb{R}^{d-1}$.*

(2) *$\mathcal{D}^K$ is a singleton $\{(\theta_{1\ell}^g, \theta_2^{g*})\}$ and $Q^{\mathcal{J}}(g^{-1}(\theta_{1\ell}^g; \theta_2^{g*}), \theta_2^{g*})$ has a full column rank.*

Assumption 4.8 is a technique assumption that ensures that whenever $\Theta_{I2}^g(\theta_{1\ell}^g)$ is a singleton, Q^g has a full column rank. Assumption 4.8 holds if m and g are linear (together with assumption 3.2).

Without loss of generality, assume that the coefficient associate with θ_1 in g is non–zero. For each $\theta^g \in \Theta^g$, let $m_{(j)}^g(\theta^g) = m_{(j)}\left(g^{-1}(\theta_1^g; \theta_2^g), \theta_2^g\right)$, where $g^{-1}(\cdot; \theta_2^g)$ is the inverse of g with respect to its first argument, holding the others fixed.

Theorem 4.4 shows that the tuple $(\Theta^g, \Theta_I^g, m^g)$ satisfies the assumptions of theorem 4.1, provided that (Θ, Θ_I, m) satisfies the corresponding assumptions. Hence under this reparameterization, making inferences about $g(\theta_0)$ is equivalent to making inferences about θ_{01}^g, the first element of the reparameterized vector θ_0^g.

Theorem 4.4. *Suppose that assumptions 3.1 to 3.3, 3.5, 3.7 to 3.9 and 4.1 to 4.8 are satisfied. Then the same inferences can be made for $g(\theta_0)$ as were made for θ_{01} in theorems 4.1 and 4.2.*

Proof. See appendix C.5 □

4.4. **Constructing a confidence set for the interval** $[\theta_{1\ell}, \theta_{1u}]$**.** My method can be used to construct confidence sets for $[\theta_{1\ell}, \theta_{1u}]$. Let $\tilde{\xi}_{\ell n}$ and $\tilde{\xi}_{un}$ be defined as in eq. (12). For any $0 < \alpha < 1/2$, let $(c_\ell^{\mathrm{II}}, c_u^{\mathrm{II}})$ be a solution to the following problem,

$$(c_\ell^{\mathrm{II}}, c_u^{\mathrm{II}}) = \underset{(c_\ell, c_u) \in \mathbb{R}^+ \times \mathbb{R}^+}{\operatorname{argmin}} \left| F_{1n}^{-1}\left(\frac{c_\ell}{\alpha_{\ell n} D_n}\right) - F_{1n}^{-1}\left(1 - \frac{c_u}{\alpha_{\ell n} D_n}\right) \right|$$

$$\text{s.t.} \quad \mathbb{P}\left(c_\ell \leq \tilde{\xi}_{\ell n}, c_u \leq \tilde{\xi}_{un}\right) = 1 - \alpha.$$

Let $\Theta^{\mathrm{II}}_{\alpha n} = \left[F^{-1}_{1n}\left(c^{\mathrm{II}}_\ell / \alpha_{\ell n} D_n\right), F^{-1}_{1n}\left(1 - c^{\mathrm{II}}_u / \alpha_{\ell n} D_n\right)\right]$.

Theorem 4.5. *Suppose that assumptions 3.1 to 3.3, 3.5, 3.7, 3.8 and 4.1 to 4.3 are satisfied. Then*

$$\lim_{n \to \infty} \mathbb{P}([\theta_{1\ell}, \theta_{1u}] \subseteq \Theta^{\mathrm{II}}_{\alpha n}) = 1 - \alpha.$$

Proof. See appendix C.6. □

The constants c^{II}_ℓ and c^{II}_u can be estimated similar to the way I estimate c^{I}_ℓ and c^{I}_u. This reparameterization approach can also be used to construct a confidence set for

$$\Theta^g_{I1} = [\inf_{\theta_0 \in \Theta_I} g(\theta_0), \sup_{\theta_0 \in \Theta_I} g(\theta_0)].$$

Corollary 4.1. *The confidence set of Θ^g_{I1} can be constructed as in theorem 4.5.*

4.5. **Comparison with projection methods: an example.** I compare my confidence set for θ_{01} to a projected confidence set using example D. I show that the projected confidence set covers θ_{01} with probability strictly greater than $1 - \alpha$. I replicate the same exercise for $[\theta_{1\ell}, \theta_{1u}]$ and obtain the same conclusion.

I follow CHT's procedure and construct a confidence set for θ_0 which has $1 - \alpha$ asymptotic coverage probability. The projected confidence set is constructed by taking the first dimension of a confidence set for θ_0.

Example D. Consider the following linear moment inequalities,

$$\mathbb{E}m_{(1)}(W_i, \theta) = \mathbb{E}X_i\theta_1 + \mathbb{E}Y_i\theta_2 \leq 0,$$

$$\mathbb{E}m_{(2)}(W_i, \theta) = \mathbb{E}X_i\theta_1 - \mathbb{E}Z_i\theta_2 + 2 \leq 0,$$

where $\{W_i = (X_i, Y_i, Z_i)\}$ are i.i.d. observations such that $\mathbb{E}X_1 = -1$, $\mathbb{E}Y_1 = \mathbb{E}Z_1 = 1$, $\mathbb{E}\|W_1\|^2 < \infty$. Let the parameter space Θ equal $[0,2] \times [0,2]$ and let $\theta_0 = (1,1)'$. I assume that X_i, Y_i and Z_i are independent. Let $\Delta_{nW} = \sqrt{n}(\bar{W} - \mathbb{E}W)$ and Δ_W be the limit of Δ_{nW}.

In this example, the identified set is a triangle whose three edges are $\theta_2 - \theta_1 \leq 0$, $-\theta_1 - \theta_2 + 2 \leq 0$ and $\theta_1 - 2 \leq 0$. The object of interest is the parameter θ_{01} and the first dimension of the identified set, which equals $[\theta_{1\ell}, \theta_{1u}] = [1, 2]$. For simplicity, I assume that the researcher knows the value of θ_{1u}, but not $\theta_{1\ell}$. $\square$

4.5.1. *The projected confidence set for θ_{01}.* Following CHT, I define a random variable (see Theorem 4.2 in CHT)

$$\mathcal{C} = \lim_{n\to\infty} \sup_{\theta\in\Theta_I} nL_n(\theta) = \sup_{\theta\in\Theta_I} \left\{ |\Delta_X\theta_1 + \Delta_Y\theta_2 + \kappa_1(\theta)|_+^2 + |\Delta_X\theta_1 - \Delta_Z\theta_2 + \kappa_2(\theta)|_+^2 \right\},$$

where $\kappa_j(\theta) = -\infty$ if $\mathbb{E}m_{(j)}(\theta) < 0$; $\kappa_j(\theta) = 0$ if $\mathbb{E}m_{(j)}(\theta) = 0$. Let c_α^{Pro} be the $1-\alpha$ quantile of $\mathcal{C}$: $\mathbb{P}(\mathcal{C} \leq c_\alpha^{Pro}) = 1-\alpha$. For any $\theta \in \Theta_I$, let $\bar{c}(\theta) = \min\{c_\alpha^{Pro}, \hat{c}_\alpha^{Pro}(\theta)\}$ where $\hat{c}_\alpha^{Pro}(\theta)$ is a consistent estimate of the $1-\alpha$ quantile of (see equation (5.5) in CHT)

$$\mathcal{C}(\theta) = \lim_{n\to\infty} nL_n(\theta) = \|\Delta^{\mathcal{J}}(\theta)\|_+^2,$$

where $\Delta(\theta)$ is mean zero normal.

The confidence set $\Theta_\alpha^{Pro,\mathrm{I}}$ for θ_0 is

$$\widehat{\Theta}_\alpha^{Pro,\mathrm{I}} = \{\theta \in \Theta : nL_n(\theta) \leq \bar{c}(\theta)\},$$

I follow the simulation algorithm proposed by CHT to compute $\hat{c}^{Pro}(\theta)$, which by construction is truncated above. Thus one can let $\bar{c}(\theta) = \hat{c}^{Pro}(\theta)$. CHT show that the confidence set constructed this way has $1-\alpha$ asymptotic coverage probability for θ_0.

Note that the projection of $\widehat{\Theta}_\alpha^{Pro,\mathrm{I}}$ onto the axis of θ_1 takes the form of $[\theta_{1\alpha}^{Pro,I}, 2]$.

Lemma 4.2. *In example D,*

$$\lim_{n\to\infty} \inf_{\theta_0\in\Theta_I} \mathbb{P}(\theta_0 \in \widehat{\Theta}_\alpha^{Pro,\mathrm{I}}) = 1-\alpha,$$

and

$$\lim_{n\to\infty} \inf_{\theta_1\in[1,2]} \mathbb{P}(\theta_1 \in [\theta_{1\alpha}^{Pro,I}, 2]) = \mathbb{P}(\theta_{1\alpha}^{Pro,I} \leq 1) > 1-\alpha.$$

Proof. See appendix C.7. $\square$

In this example, although $\widehat{\Theta}_\alpha^{Pro,I}$ has $1-\alpha$ asymptotic coverage probability for θ_0, the projection $[\theta_{1\alpha}^{Pro,I}, 2]$ fails to have this desirable property.

4.5.2. *The projected confidence set for* $[\theta_{1\ell}, \theta_{1u}]$. To construct the confidence set for $[1,2]$ by the projection method, I first construct a confidence set $\Theta_{\alpha}^{Pro,\mathrm{II}}$ for Θ_I, then project it onto the axis of θ_1 to obtain $\Theta_{1\alpha}^{Pro,\mathrm{II}}$, where

$$\Theta_{\alpha}^{Pro,\mathrm{II}} = \{\theta \in \Theta : nL_n(\theta) \leq c_{\alpha}^{Pro}\}.$$

Lemma 4.3. *In example D,* $\lim_{n\to\infty} \mathbb{P}([1,2] \subseteq \Theta_{1\alpha}^{Pro,\mathrm{II}}) > 1 - \alpha$.

Proof. See appendix C.8. □

The intuition behind lemma 4.3 is as follows. Ideally one wants to construct $\Theta_{1\alpha}^{Pro,\mathrm{II}}$ based on the asymptotic behavior of the sample objective function around the point $(1,1)$, because it is the "left–most" corner of the identified set. Nevertheless, the critical value is chosen based on the random variable $\mathcal{C}$, which is the limiting random variable of the sup of the sample objective function over the entire boundary. In this example, the maximizer is different from $(1,1)$ with probability greater than one. Therefore, the critical values are based on a random variable that stochastically dominates the desired one, which results in a conservative confidence set.

5. Further discussion

5.1. **Asymptotic Normality.** In CH, the localized and recentered quasi–posterior density for local parameter h converges in probability to a standard normal density in total variation of moments norm sense (see Theorem 1 in CH). In moment inequality models, the limit of the localized quasi–posterior density is generally not normal. The sample objective function does not allow for a quadratic expansion toward inside of the identified set because of the truncation $\|\cdot\|_+$. In example A, the quasi–posterior density of the local parameter $h_\ell = \sqrt{n}(\theta - \theta_\ell)$ is flat for all $h_\ell > 0$. In addition, towards the outside of the identified set, depending on which moments are binding, the localized quasi–posterior density is generally a mixture of normals. In generally, $\hat{\theta}_{1\ell}$ and $\hat{\theta}_{1u}$ will not have limiting normal distribution either.

There is, however, a way of achieving the asymptotic normality for $\hat{\theta}_{1\ell}$ and $\hat{\theta}_{1u}$ in a class of moment inequality models by "smoothing out" the truncation caused by $\|\cdot\|_+$. Consider a “smoothed” quasi–posterior density

$$f_n^S(\theta) \propto \exp\left(-\|\sqrt{n}\bar{m}(\theta)\|_{\rho_n}^2\right),$$

where $\rho_n \geq 0$. For a vector x, $\|x\|_{\rho_n}^2 = \sum_j^J \left(|x_j|_{\rho_n}\right)^2$, where $|x|_{\rho_n} = x\mathbf{1}[x \geq -\rho_n]$. The new sample objective function essentially penalizes all the sample moments that are not sufficiently smaller than 0. Let F_n^S be the distribution function corresponding to f_n^S. Note that $f_n^S = f_n$ if $\rho_n = 0$.

Assumption 5.1. *$\Theta_{I2}(\theta_{1\ell}) = \{\theta_{2\ell}\}$ is singleton and $\theta_\ell = (\theta_{1\ell}, \theta_{2\ell})'$ is in the interior of parameter space.*

Note that assumption 5.1 is assumed in PPHI.

Assumption 5.2. *The first dimension of Θ_I has non–empty interior.*

Assumption 5.3. *ρ_n diverges to $+\infty$ slower than polynomial rate.*

Let $\hat{\theta}_{1\ell}^S = (F_{1n}^S)^{-1}\left[\hat{c}/(nD_n)\right]$. To achieve asymptotic normality, I requires that $\hat{c}$ satisfies assumption 5.4 below.

Assumption 5.4. $\hat{c} \xrightarrow{p} \frac{\sqrt{2}}{4}(2\pi)^d\sqrt{|(Q_\ell^{\mathcal{J}\prime}Q_\ell^{\mathcal{J}})^{-1}|}$.

Theorem 5.1. *Suppose that assumptions 3.1 to 3.3, 3.5, 3.7, 3.8 and 5.1 to 5.4 are satisfied, then*

$$\sqrt{n}(\hat{\theta}_{1\ell}^S - \theta_{1\ell}) \xrightarrow{d} \mathcal{U}_1,$$

where $\mathcal{U}_1$ is the first component of the vector $\mathcal{U} = (Q_\ell^{\mathcal{J}\prime}Q_\ell^{\mathcal{J}})^{-1}Q_\ell^{\mathcal{J}\prime}\Delta_\ell^{\mathcal{J}}$ and

$$\mathcal{U} \sim \mathcal{N}\left\{0, (Q_\ell^{\mathcal{J}\prime}Q_\ell^{\mathcal{J}})^{-1}(Q_\ell^{\mathcal{J}\prime}V_\ell^{\mathcal{J}}Q_\ell^{\mathcal{J}})(Q_\ell^{\mathcal{J}\prime}Q_\ell^{\mathcal{J}})^{-1}\right\},$$

where $V_\ell^{\mathcal{J}} = \mathbb{E}[\Delta_\ell^{\mathcal{J}}\Delta_\ell^{\mathcal{J}\prime}]$.

Proof. See appendix E.1. □

$Q_\ell^{\mathcal{J}}$ is unknown, but can be consistently estimated using a first step consistent estimator. The variance of $\Delta_\ell^{\mathcal{J}}$ can also be estimated (see discussions in section 3).

6. Conclusion

In this paper I propose integration–based estimation and inference methods for moment inequality models. My confidence sets covers the each component (or a scalar–valued function) of the true parameter vector with prespecified probability and are comparatively easy to compute.

There are issues of potential interest that are not studied in this paper. First, one may be interested in models where the number of moment equations is large (Menzel, 2008), models characterized by conditional moment inequalities (Andrews and Shi, 2010; Shi, 2009), or models are not characterized by moment inequalities (maximum score estimator when the support condition is violated). Second, I focus on a single element (or a scalar–valued function) of the parameter vector. In some applications researchers may be interested in a joint confidence set for a subvector, in which case one needs to study the asymptotic behavior of the marginal quasi–posterior of the corresponding subvector. In this case confidence set can be constructed as a level set of the marginal quasi–posterior (similar to what CHJ do for the entire parameter vector). Third, it is challenging but interesting to extend the current framework and allow for the presence of infinite dimensional nuisance parameters.

References

ANDREWS, D., S. BERRY, AND P. JIA (2004): "Confidence regions for parameters in discrete games with multiple equilibria, with an application to discount chain store location," *manuscript, Yale University*.

ANDREWS, D., AND P. GUGGENBERGER (2009): "Validity of subsampling and plug-in asymptotic inference for parameters defined by moment inequalities," *Econometric Theory*, 25(03), 669–709.

ANDREWS, D., AND X. SHI (2010): "Inference based on conditional moment inequalities," *Working Paper*.

ANDREWS, D., AND G. SOARES (2010): "Inference for parameters defined by moment inequalities using generalized moment selection," *Econometrica*, 78(1), 119–157.

ARADILLAS-LOPEZ, A., AND E. TAMER (2008): "The identification power of equilibrium in simple games," *Journal of Business and Economic Statistics*, 26(3), 261–283.

BERESTEANU, A., I. MOLCHANOV, AND F. MOLINARI (2010): "Partial Identification Using Random Sets Theory," .

BERESTEANU, A., AND F. MOLINARI (2008): "Asymptotic properties for a class of partially identified models," *Econometrica*, 76(4), 763–814.

BILLINGSLEY, P. (1995): *Probability and measure*. Wiley-India.

BONTEMPS, C., T. MAGNAC, AND E. MAURIN (2007): "Set Identified Linear Models," *Manuscript, Toulouse School of Economics*.

BUGNI, F. (2010): "Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set," *Econometrica*, 78(2), 735–753.

CANAY, I. A. (2010): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," *Journal of Econometrics*, 156(2), 408–425.

CHERNOZHUKOV, V., C. HANSEN, AND M. JANSSON (2009): "Finite sample inference for quantile regression models," *Journal of Econometrics*, 152(2), 93–103.

CHERNOZHUKOV, V., AND H. HONG (2003): "An MCMC approach to classical estimation," *Journal of Econometrics*, 115(2), 293–346.

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5), 1243–1284.

CILIBERTO, F., AND E. TAMER (2009): "Market structure and multiple equilibria in airline markets," *Econometrica*, 77(6), 1791–1828.

GRIECO, P. (2010): "Discrete Games with Flexible Information Structures: An Application to Local Grocery Markets," Discussion paper, Working papers.

HAHN, J., AND G. RIDDER (2009): "Partial Identification and Confidence Intervals," *Working Paper*.

HAILE, P., AND E. TAMER (2003): "Inference with an incomplete model of English auctions," *Journal of Political Economy*, 111(1), 1–51.

HOROWITZ, J., AND C. MANSKI (2000): "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data.," *Journal of the American Statistical Association*, 95(449).

IMBENS, G., AND C. MANSKI (2004): "Confidence intervals for partially identified parameters," *Econometrica*, 72(6), 1845–1857.

JUN, S., J. PINKSE, AND Y. WAN (2009): "Cube–root–n and faster convergence, laplace estimator and uniform inference," Discussion paper.

JUN, S. J. (2008): "Weak identification robust tests in an instrumental quantile model," *Journal of Econometrics*, 144(1), 118 – 138.

KAIDO, H. (2010): "A Dual Approach to Inference for Partially Identified Econometric Models," *Working paper*.

KAIDO, H., AND H. WHITE (2009): "Inference on Risk-Neutral Measures for Incomplete Markets," *Journal of Financial Econometrics*.

LIAO, Y., AND W. JIANG (2010): "Bayesian analysis in moment inequality models," *The Annals of Statistics*, 38(1), 275–316.

MANSKI, C. (2003): *Partial identification of probability distributions*. Springer Verlag.

MANSKI, C., AND J. PEPPER (2000): "Monotone instrumental variables: with an application to the returns to schooling," *Econometrica*, 68(4), 997–1010.

MANSKI, C. F., AND E. TAMER (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70(2), 519–546.

MENZEL, K. (2008): "Estimation and Inference with Many Moment Inequalities," *Preprint, Massachussetts Institute of Technology*.

MOON, H., AND F. SCHORFHEIDE (2009): "Bayesian and frequentist inference in partially identified models," *NBER Working Paper*.

NEWEY, W., 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 (2006): "Moment inequalities and their application," *Working Paper*.

PHILLIPS, P. C. B. (1989): "Partially Identified Econometric Models," *Econometric Theory*, 5(2), pp. 181–240.

ROBERT, C., AND G. CASELLA (2004): *Monte Carlo statistical methods*. Springer Verlag.

ROMANO, J., AND A. SHAIKH (2010): "Inference for the identified set in partially identified econometric models," *Econometrica*, 78(1), 169–211.

ROSEN, A. (2008): "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities," *Journal of Econometrics*, 146(1), 107–117.

SHI, X. (2009): "Model Selection Tests for Nonnested Moment Inequality Models," *Working Paper*.

STOYE, J. (2009): "More on confidence intervals for partially identified parameters," *Econometrica*, 77(4), 1299–1315.

VAN DER VAART, A., AND J. WELLNER (1996): *Weak convergence and empirical processes*. Springer Verlag.

WOUTERSEN, T. (2006): "A Simple Way to Calculate Confidence Intervals for Partially Identified Parameters," *Working Paper*.

APPENDIX A. SOME LEMMAS

In this section I present some lemmas which will be used for the proofs in sections 3 and 4. A maintained assumption in appendix A is that Θ_I belongs to the interior of parameter space. The proofs are based on $d = 2$, but can be extended to generic d (see appendix F).

Lemma A.1 says that nD_n is bounded away from 0 with probability approaching one. Lemma A.1 will be used in the proof for consistency.

Lemma A.1. *Suppose that assumptions 3.1, 3.4 and 3.5 are satisfied, then for any $\epsilon > 0$, there exists a $C^* > 0$ such that $\lim_{n\to\infty} \mathbb{P}(nD_n < C^*) < \epsilon$.*

Proof. Assumption 3.1 ensures that D_n is well defined. For some $C_1 > 0$, define set $A_n = \{\theta : d(\theta, \Theta_I) \leq \frac{C_1}{\sqrt{n}}\}$. Then by Lipschitz assumption 3.4, there exist $C_2 > 0$ such that

$$\max_j \sup_{\theta \in A_n} |\mathbb{E}m_{(j)}(W_1, \theta)|_+ \leq C_2 \sup_{\theta \in A_n} d(\theta, \Theta_I) \leq C_2 C_1/\sqrt{n}.$$

Let $\mu(A_n)$ be the Lebesgue measure of A_n, then exist $C_3 > 0$ and $C_4 > 0$ such that $C_3/n \geq \mu(A_n) \geq C_4/n$. Let ι be a J–vector of ones, then for any $C > 0$,

$$\begin{aligned}
\mathbb{P}(nD_n < C) &\leq \mathbb{P}\left(n \inf_{\theta \in A_n} \exp(-n\|\bar{m}(\theta)\|_+^2)\mu(A_n) < C\right) \\
&\leq \mathbb{P}\left(C_3 \inf_{\theta \in A_n} \exp(-\|\Delta_n(\theta) + \sqrt{n}\mathbb{E}m_1(\theta)\|_+^2) < C\right) \\
&= \mathbb{P}\left(\sup_{\theta \in A_n} \|\Delta(\theta) + \sqrt{n}\mathbb{E}m_1(\theta)\|_+^2 > \log(C_3/C) + o_p(1)\right)
\end{aligned}$$

$$\leq \mathbb{P}\left(\| \sup_{\theta \in A_n} \Delta(\theta) + \sup_{\theta \in A_n} \sqrt{n}\mathbb{E}m_1(\theta)\|_+^2 > \log(C_3/C) + o_p(1)\right)$$
$$\leq \mathbb{P}\left(\| \sup_{\theta \in A_n} \Delta(\theta) + C_1 C_2 \iota\|_+^2 > \log(C_3/C) + o_p(1)\right).$$

The equality holds because $\sup_{\theta \in A_n} \|\Delta_n(\theta) - \Delta(\theta)\| = o_p(1)$ by assumption 3.5 and $\| \cdot \|$ is continuous. The last inequality holds because the monotonicity of $\| \cdot \|_+$. The right hand side probablity converges to zero as C decrease to zero since $\sup_{\theta \in A_n} \Delta(\theta)$ is bounded in probability. □

Lemmas A.2 to A.8 say that the integral of the numerator of the quasi–posterior outside of the identified set can be approximated by a "localized" integral within a $\sqrt{n}$–neighborhood (but outside) of the identified set, with demeaned random variables replaced by their limits. My proof follows the same idea as in Chernozhukov and Hong (2003).

Lemmas A.2 to A.7 deal with the case in which $\Theta_{I2}(\theta_{1\ell})$ is a singleton: $\Theta_{I2}(\theta_{1\ell}) = \{(\theta_{1\ell}, \theta_{2\ell})\} = \{\theta_\ell\}$. Define

$$N_n(h) = D_n f_n(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_{2\ell} + \frac{h_2}{\sqrt{n}}), \quad N_\infty(h) = \exp(-\|\Delta^{\mathcal{J}}(\theta_\ell) + Q^{\mathcal{J}}(\theta_\ell)h\|_+^2).$$

By assumptions 3.1 and 3.2, I can write the integration region for the local parameter $h = \sqrt{n}(\theta - \theta_\ell)$ as $H_n = \{h : -\sqrt{n}(\theta_{1\ell} - \underline{\theta}_1) \leq h_1 \leq 0, -\sqrt{n}(\theta_{2\ell} - \underline{\theta}_2) \leq h_2 \leq \sqrt{n}(\overline{\theta}_2 - \theta_{2\ell})\}$. I separate H_n into three parts: $H_{1n} = \{h : \|h\| < M, h_1 \in H_n\}$, $H_{2n} = \{h : M < \|h\| < M^*\sqrt{n}, h \in H_n\}$ and $H_{3n} = \{h : \|h\| \geq M^*\sqrt{n}, h \in H_n\}$ for some $M^*, M > 0$. In the rest part of appendix A, I drop the sup script $\mathcal{J}$ for the sake of notational simplicity. By assumption assumption 3.2, H_n converges to $\mathbb{R}^- \times \mathbb{R}$ in the Painleve–Kuratowski sense.[11]

Lemma A.2. *Suppose that assumptions 3.1, 3.2, 3.5 and 3.7 are satisfied, then*

$$\int_{h \in H_{1n}} |N_n(h) - N_\infty(h)| dh \xrightarrow{p} 0.$$

Proof. Note that $N_n(h)$ is

$$\begin{aligned}
N_n(h) &= D_n f_n(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_{2\ell} + \frac{h_2}{\sqrt{n}}) \\
&= \exp(-\|\Delta_n(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_{2\ell} + \frac{h_2}{\sqrt{n}}) + \sqrt{n}\mathbb{E}m_1(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_{2\ell} + \frac{h_2}{\sqrt{n}})\|_+^2) \\
&= \exp(-\|\Delta_n(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_{2\ell} + \frac{h_2}{\sqrt{n}}) + Q(\theta_\ell)h + \sqrt{n}\mathbb{E}m_1(\theta_{1\ell}, \theta_{2\ell}) + R_n(h_1, h_2)\|_+^2),
\end{aligned}$$

[11]See discussion in Kaido (2010).

where

$$R_n(h_1,h_2) = \sqrt{n}\mathbb{E}m_1(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_{2\ell} + \frac{h_2}{\sqrt{n}}) - Q(\theta_\ell)h - \sqrt{n}\mathbb{E}m_1(\theta_{1\ell}, \theta_{2\ell}).$$

Note that if $\mathbb{E}m_1(\theta_{1\ell}, \theta_{2\ell}) = 0$ (I already drop the sup script $\mathcal{J}$ for notational simplicity). So it follows that,

$$\int_{h\in H_{1n}} |N_n(h) - N_\infty(h)|dh = \int_{H_{1n}} N_\infty(h)|\exp(T_n(h)) - 1|dh,$$

where $T_n(h) = \|\Delta(\theta_\ell) + Q(\theta_\ell)h\|_+^2 - \|\Delta_n(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_{2\ell} + \frac{h_2}{\sqrt{n}}) + Q(\theta_\ell)h + R_n(h_1,h_2)\|_+^2$, it is sufficient to show that $\sup_{h\in H_{1n}} |T_n(h)| = o_p(1)$ since N_∞ is uniformly bounded in probability over H_{1n}.

$$\begin{aligned}
&\sup_{h\in H_{1n}} \|\Delta(\theta_\ell) - \Delta_n(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_{2\ell} + \frac{h_2}{\sqrt{n}})\| \\
&\quad\le \sup_{h\in H_{1n}} \|\Delta(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_{2\ell} + \frac{h_2}{\sqrt{n}}) - \Delta_n(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_{2\ell} + \frac{h_2}{\sqrt{n}})\| \\
&\qquad + \sup_{h\in H_{1n}} \|\Delta(\theta_\ell) - \Delta(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_{2\ell} + \frac{h_2}{\sqrt{n}})\| = o_p(1).
\end{aligned}$$

The right hand termm is $o_p(1)$ because of assumption 3.5.

By assumption 3.7 (continuous differentiability), I know that $\sup_{h\in H_{1n}} R_{n,j}(h) = o(\|h\|) \le o(M) = o(1)$. Thus $\sup_{h\in H_{1n}} T_n(h) \xrightarrow{p} 0$ and $\int_{h\in H_{1n}} |N_n(h) - N_\infty(h)|dh \xrightarrow{p} 0$. □

Lemma A.3. *Suppose that assumptions 3.1, 3.2, 3.5, 3.7 and 3.8 are satisfied, then for any $\epsilon > 0$, there is a choice of $0 < M < \infty$ such that*

$$\mathbb{P}(\int_{H_{2n}\cup H_{3n}} N_\infty(h)dh \le \epsilon) > 1 - \epsilon.$$

Proof. Let a and b be scalars. Then $|a+b|_+^2 \ge |a|_+^2 - |b|$ unless $a \ge 0$, $b < 0$ and $a + b > 0$. This can be verified as follows.

Case 1, $a \ge 0, b \ge 0$. $|a+b|_+^2 \ge |a|_+^2 - |b|$ holds obviously.

Case 2, $a < 0, b < 0$. $|a+b|_+^2 \ge 0 > |a|_+^2 - |b|$ holds obviously.

Case 3, $a < 0, b \ge 0$. $|a+b|_+^2 \ge 0 \ge -|b| = |a|_+^2 - |b|$.

Case 4, $a \ge 0, b < 0$. In this case, if $a + b < 0$, then $|a+b|_+^2 = 0 \ge |a| - |b| = |a|_+^2 - |b|$.

In this lemma, I will treat $Q(\theta_\ell)h$ term as a and $\Delta(\theta_\ell)$ term as b.

I divide the integration region $H_{2n} \cup H_{3n}$ into H_n^A and H_n^B, where

$$H_n^A = \{h \in H_{2n} \cup H_{3n}, Q(\theta_\ell)h + \Delta(\theta_\ell) \ge 0\},$$

and H_n^B is the complement of H_n^A within $H_{2n} \cup H_{3n}$. Note

$$\mathbb{P}(\int_{H_{2n}\cup H_{3n}} N_\infty(h)dh \leq 2\epsilon) \geq \mathbb{P}(\int_{H_n^A} N_\infty(h)dh \leq \epsilon, \int_{H_n^B} N_\infty(h)dh \leq \epsilon)$$
$$\geq 1 - \mathbb{P}(\int_{H_n^A} N_\infty(h)dh > \epsilon) - \mathbb{P}(\int_{H_n^B} N_\infty(h)dh > \epsilon).$$

For region H_n^A, since $Q(\theta_\ell)h + \Delta(\theta_\ell) \geq 0$, I have

$$\lim_{M\to+\infty} \mathbb{P}(\int_{H_n^A} N_\infty(h)dh > \epsilon) = \lim_{M\to+\infty} \mathbb{P}(\int_{H_n^A} \exp(-\|\Delta(\theta_\ell) + Q(\theta_\ell)h\|_+^2)dh > \epsilon)$$
$$\leq \lim_{M\to+\infty} \mathbb{P}(\int_{\|h\|>M} \exp(-\|\Delta(\theta_\ell) + Q(\theta_\ell)h\|^2)dh > \epsilon) = 0.$$

The right hand side term converges to zero because $Q(\theta_\ell)$ has full column rank by assumption 3.8 and $\Delta(\theta_\ell)$ is bounded in probability by assumption 3.5.

Now consider region H_n^B, observe that $\|\Delta(\theta_\ell) + Q(\theta_\ell)h\|_+^2 \geq \|Q(\theta_\ell)h\|_+^2 - \|\Delta(\theta_\ell)\|$ for all $h \in H_n^B$, it follows

$$\int_{h\in H_n^B} N_\infty(h)dh \leq \int_{h\in H_n^B} \exp(-\|Q(\theta_\ell)h\|_+^2 + \|\Delta(\theta_\ell)\|^2)dh = \exp(\|\Delta(\theta_\ell)\|^2)\int_{h\in H_n^B} \exp(-\|Q(\theta_\ell)h\|_+^2)dh.$$

Since $\|\Delta(\theta_\ell)\|$ is bounded in probability by assumption 3.5, to complete the proof of this lemma, I just need to show

$$\lim_{M\to+\infty} \int_{h\in H_n^B} \exp(-\|Q(\theta_\ell)h\|_+^2)dh = 0.$$

Note that

$$\int_{h\in H_n^B} \exp(-\|Q(\theta_\ell)h\|_+^2)dh \leq \int_{h\in H_n} \exp(-\|Q(\theta_\ell)h\|_+^2)dh$$
$$= \int_{\{h:h_1\leq 0\}} \exp(-\|Q(\theta_\ell)h\|_+^2)dh - \int_{\{h:h_1\leq 0, \|h\|\leq M\}} \exp(-\|Q(\theta_\ell)h\|_+^2)dh.$$

If the first term on the right hand side is finite, then the second term on the right hand side is also finite and moreover is strictly monotonically increasing in M. It covers to the first term as M increases because a bounded monotonically increasing sequence has a limit. It remains to show that the first term is finite.

To show the first term is finite, observe that (i) for all j, Q_{j1}, which is the derivative of the jth binding moment equation with respect to θ_1 evaluating at θ_ℓ, are all negative by definition of $\theta_{1\ell}$, and (ii) by the definition of θ_ℓ, there exists at least one pair of (j, j') such that $Q_{j'2}$ and Q_{j2} is non–zero and take different signs since $\Theta_{I2}(\theta_{1\ell})$ is a singleton. Without loss of generality, assume that $Q_{12} < 0$ and $Q_{22} > 0$.

First note that $h = 0$ is the only possible value from H_n such that all components of the vector $Q(\theta_\ell)h$ be smaller or equal to zero. In other words, for all $h \in H_n$, $h \neq 0$, there exist at least one moment who takes positive value. Now consider the integration region where the first element of $Q(\theta_\ell)h$ is the only element greater than 0. Call this region H_n^{B1}. It implies that $Q_{11}h_1 + Q_{12}h_2 > 0$. Since $Q_{12} < 0$, then $h_2 < -\frac{Q_{11}}{Q_{12}}h_1$. Since $Q_{22} > 0$, then $Q_{21}h_1 + Q_{22}h_2 < 0$ implies that $h_2 < -\frac{Q_{21}}{Q_{22}}h_1$. Since $h_1 < 0$, it turns out $-\frac{Q_{21}}{Q_{22}}h_1 < 0 < -\frac{Q_{11}}{Q_{12}}h_1$, so

$$\int_{h \in H_n^{B1}} \exp(-\|Q_\ell h\|_+^2)dh \leq \int_{h_1 \leq 0} \int_{h_2 < -\frac{Q_{21}}{Q_{22}}h_1, \|h\| \geq M} \exp(-(Q_{11}h_1 + Q_{12}h_2)^2)dh.$$

The right hand side converges to zero as M increases to ∞.

For the integration region where there are more than one elements take positive value, a similar argument applies. In particular, if in the region where all the elements are positive, it is integrable because $\exp(-h'Q(\theta_\ell)'Q(\theta_\ell)h)$ is just a rescaled normal density.

$\square$

Lemma A.4. *Suppose that assumptions 3.1, 3.2, 3.5, 3.7 and 3.8 are satisfied, then for any $\epsilon > 0$, there exists M^* and M such that*

$$\lim_{n\to\infty} \mathbb{P}\left(\int_{H_{2n}} N_n(h)dh \leq \epsilon\right) > 1 - \epsilon.$$

Proof. Let ϵ be arbitrarily given. For any $\epsilon^* > 0$, let M^* be small enough such that the following condition holds:

$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{\{h: \|h\| \leq \sqrt{n}M^*, h_1 \leq 0\}} \frac{\|\Delta_n(\theta_\ell + \frac{h}{\sqrt{n}}) - \Delta_n(\theta_\ell)\|}{1 + \|h\|} > \epsilon^* \right) < \epsilon^*. \tag{15}$$

This is possible by assumption 3.5. Meanwhile,

$$\begin{aligned} N_n(h) &= \exp(-\|\Delta_n(\theta_\ell + \frac{h}{\sqrt{n}}) + \sqrt{n}\mathbb{E}m_1(\theta_\ell + \frac{h}{\sqrt{n}})\|_+^2) \\ &= \exp(-\|\Delta_n(\theta_\ell) + Q(\theta_\ell)h + R_n^a(h_1, h_2) + R_n^b(h_1, h_2)\|_+^2) \\ &= \exp(-\|h\|^2 \|\frac{\Delta_n}{\|h\|} + Q_\ell \frac{h}{\|h\|} + \frac{R_n^a}{\|h\|} + \frac{R_n^b}{\|h\|}\|_+^2). \end{aligned}$$

where

$$R_n^a(h_1, h_2) = \Delta_n(\theta_\ell + \frac{h}{\sqrt{n}}) - \Delta_n(\theta_\ell),$$

$$R_n^b(h_1, h_2) = \sqrt{n}\mathbb{E}m_1(\theta_\ell + \frac{h}{\sqrt{n}}) - Q(\theta_\ell)h = o(\|h\|).$$

Hence we can choose M^* such that for largen n,

$$\begin{aligned}
N_n(h) = \exp\left(-\|h\|^2\left[\left\|\frac{\Delta(\theta_\ell)}{\|h\|} + Q(\theta_\ell)\frac{h}{\|h\|}\right\|_+^2 + o_p(1)\right]\right) \\
\leq \exp\left(-\|h\|^2\left[\left\|\frac{\Delta(\theta_\ell)}{\|h\|} + Q(\theta_\ell)\frac{h}{\|h\|}\right\|_+^2 - \frac{1}{2}\|Q(\theta_\ell)\frac{h}{\|h\|}\|_+^2\right]\right) \\
= \exp\left(-\|\Delta(\theta_\ell) + Q(\theta_\ell)h\|_+^2 + \frac{1}{2}\|Q(\theta_\ell)h\|_+^2\right) \\
\leq \exp\left(-\|Q(\theta_\ell)h\|_+^2 + \|\Delta(\theta_\ell)\|^2 + \frac{1}{2}\|Q(\theta_\ell)h\|_+^2\right) = \exp\left(-\frac{1}{2}\|Q(\theta_\ell)h\|_+^2 + \|\Delta(\theta_\ell)\|^2\right)
\end{aligned}$$

The first inequality is because of eq. (15) and the fact that $\|Q(\theta_\ell)\frac{h}{\|h\|}\|_+^2 > 0$ for all h such that $h \neq 0$ and $h_1 \leq 0$. The right hand side term can be dealt with in the similar way as in lemma A.3. □

Lemma A.5. *Suppose that assumptions 3.1 and 3.3 to 3.5 are satisfied, then for any $\epsilon > 0$, and each $M^* > 0$,*

$$\lim_{n\to\infty} \mathbb{P}(\int_{H_{3n}} N_n(h)dh \leq \epsilon) > 1 - \epsilon.$$

Proof. Let $M^* > 0$ be arbitrary. For any $h \geq M^*\sqrt{n}$, let $\theta_1 = \theta_{1\ell} + \frac{h_1}{\sqrt{n}}$ and $\theta_2 = \theta_{2\ell} + \frac{h_2}{\sqrt{n}}$, let H_{3n}^* be corresponding integration region for θ. Then $\inf_{\theta\in H_{3n}^*} d(\theta, \Theta_I) \geq M^*$. By assumption 3.3, there exists at least one j and some $\delta_m > 0$ such that $\mathbb{E}m_{(j)}(W_1, \theta)_+ \geq \delta_m$ uniformly over H_{3n}^*.

$$\begin{aligned}
\sup_{h\in H_{3n}} N_n(h) = \sup_{h\in H_{3n}} \exp(-\|\Delta_n(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_{2\ell} + \frac{h_2}{\sqrt{n}}) + \sqrt{n}\mathbb{E}m_1(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_{2\ell} + \frac{h_2}{\sqrt{n}})\|_+^2) \\
= \sup_{\theta\in H_{3n}^*} \exp(-\|\Delta_n(\theta_1, \theta_2) + \sqrt{n}\mathbb{E}m_1(\theta_1, \theta_2)\|_+^2) \leq \exp(-|\inf_{\theta\in H_{3n}^*} \Delta_{jn}(\theta_1, \theta_2) + \sqrt{n}\delta_m|_+^2)
\end{aligned}$$

Since $\sup_{\theta\in\Theta} \Delta_n(\theta) = O_P(1)$, for any $\epsilon > 0$

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(\int_{H_{3n}} N_n(h)dh < \epsilon) \geq \lim_{\to\infty} \mathbb{P}(\int_{H_{3n}^*} \sup_{\theta\in H_{3n}^*} \exp(-|\Delta_{jn}(\theta_1, \theta_2) + \sqrt{n}\delta_m|_+^2)d\theta < \epsilon) \\
\geq \lim_{n\to\infty} \mathbb{P}(\mu(H_{3n}^*) \exp(-|\inf_{\theta\in H_{3n}^*} \Delta_{jn}(\theta) + \sqrt{n}\delta_m|_+^2)d\theta < \epsilon) = 1.
\end{aligned}$$

The right hand side converges to one because $\inf_{\theta\in H_{3n}^*} \Delta_n(\theta)$ is bounded in probability. □

Lemma A.6. *Suppose that assumptions 3.1 to 3.5, 3.7 and 3.8 are satisfied, then for any $\epsilon > 0$, there exist an $M > 0$ such that*

$$\lim_{n\to\infty} \mathbb{P}(\int_{H_{2n}\cup H_{3n}} |N_n(h) - N_\infty(h)|dh \le \epsilon) > 1 - \epsilon.$$

Proof. Follows from lemmas A.2 to A.5. □

Lemma A.7. *Suppose that assumptions 3.1 to 3.5, 3.7 and 3.8 are satisfied, then*

$$\int_{H_n} |N_n(h) - N_\infty(h)|dh = o_p(1).$$

Proof. Follows from lemma A.2 and lemma A.6. □

So far I have shown that approximation holds when $\Theta_{I2}(\theta_{1\ell})$ is a singleton (lemma A.7) . Lemma A.8 below shows that a similar approximation holds when $\Theta_{I2}(\theta_{1\ell})$ is an interval.

Lemma A.8. *Suppose that assumptions 3.1 to 3.5, 3.7 and 3.8 are satisfied. Suppose that $\Theta_{I2}(\theta_{1\ell}) = [\theta_{2\ell}, \theta_{2u}]$ with $\theta_{2u} > \theta_{2\ell}$, then*

$$\begin{aligned}\int_{h_1\le 0}\int_{\theta_2\in\Theta_2} & D_n f_n(\theta_{1\ell} + \frac{h}{\sqrt{n}}, \theta_2) d\theta_2 dh_1 \\ &= \int_{h_1\le 0}\int_{\theta_2\in[\theta_{2\ell},\theta_{2u}]} \exp(-\|\Delta^{\mathcal{J}}(\theta_{1\ell},\theta_2) + Q_1^{\mathcal{J}}(\theta_{1\ell},\theta_2)h_1\|_+^2) d\theta_2 dh_1 + o_p(1).\end{aligned}$$

Proof. Without loss of generality, suppose that there is only one binding moment for all $\theta \in \{\theta : \theta_1 = \theta_{1\ell}, \theta_2 \in (\theta_{2\ell}, \theta_{2u})\}$, say the first moment inequality. I can ignore the superscript $\mathcal{J}$ for notational simplicity. If there are more than one binding moments for some $\theta \in \{\theta : \theta_1 = \theta_{1\ell}, \theta_2 \in (\theta_{2\ell}, \theta_{2u})\}$, the same analysis goes through. I first show that,

$$\begin{aligned}\int_{h_1\le 0}\int_{\theta_2\in[\theta_{2\ell},\theta_{2u}]} \Big| & \exp(-\|\Delta_n(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_2) + Q_1(\theta_{1\ell},\theta_2)h_1 + R_n(\theta_{1\ell},\theta_2)\|_+^2 \\ & - \exp(-\|\Delta(\theta_{1\ell},\theta_2) + Q_1(\theta_{1\ell},\theta_2)h_1\|_+^2)\Big| d\theta_2 dh_1 = o_p(1), \quad (16)\end{aligned}$$

where

$$R_n(\theta_{1\ell},\theta_2) = \sqrt{n}\mathbb{E}m_1(\theta_{1\ell} + h_1/\sqrt{n}, \theta_2) - \sqrt{n}\mathbb{E}m_1(\theta_{1\ell},\theta_2) - Q_1(\theta_{1\ell},\theta_2)h_1.$$

By assumption 3.7, R_n is continuous in θ_2, hence we know that $\sup_{\theta_2\in[\theta_{2\ell},\theta_{2u}]} R_n(\theta_{1\ell},\theta_2) = o(\|h\|)$.

Note that for all $\theta_2 \in [\theta_{2\ell}, \theta_{2u}]$, $\mathbb{E}m_1(\theta_{1\ell}, \theta_2) = 0$ (remember I drop the supscript $\mathcal{J}$ for notational simplicity). Now I can divide the integration region for h_1 into three parts. The proof essentially follows from the previous proofs (lemma A.7).

It remains to show the integration outside of $[\theta_{2\ell}, \theta_{2u}$ is $o_p(1)$. Let $\epsilon > 0$ be arbitrary. Then

$$\mathbb{P}\left(\int_{h_1\le 0}\int_{\theta_2\notin[\theta_{2\ell},\theta_{2u}]} D_n f_n(\theta_{1\ell}+\frac{h}{\sqrt{n}},\theta_2)d\theta_2 dh_1 \le 2\epsilon\right)$$
$$\ge 1-\mathbb{P}\left(\int_{h_1\le 0}\int_{\theta_2\in[\theta_{2\ell}-\delta,\theta_{2\ell}]\cup[\theta_{2u},\theta_{2u}+\delta]} D_n f_n(\theta_{1\ell}+\frac{h}{\sqrt{n}},\theta_2)d\theta_2 dh_1 > \epsilon\right)$$
$$-\mathbb{P}\left(\int_{h_1\le 0}\int_{\theta_2\le\theta_{2\ell}-\delta,\theta_2\ge\theta_{2u}+\delta} D_n f_n(\theta_{1\ell}+\frac{h}{\sqrt{n}},\theta_2)d\theta_2 dh_1 > \epsilon\right)$$

The first probability on the right hand side can be made arbitrarily small by taking δ small; the second probability converges to zero as n increases for any $\delta > 0$.

In particular, note that if $Q(\theta_{1\ell},\theta_{2\ell})$ and $Q(\theta_{1\ell},\theta_{2u})$ have full column ranks, then

$$\int_{h_1\le 0}\int_{\theta_2\notin[\theta_{2\ell},\theta_{2u}]} D_n f_n(\theta_{1\ell}+\frac{h}{\sqrt{n}},\theta_2)d\theta_2 dh_1$$
$$=\frac{1}{\sqrt{n}}\int_{h_1\le 0,h_2\ge 0}\exp(-\|\Delta(\theta_{1\ell},\theta_{2u})+Q(\theta_{1\ell},\theta_{2u})h\|_+^2)dh$$
$$+\frac{1}{\sqrt{n}}\int_{h_1\le 0,h_2\le 0}\exp(-\|\Delta(\theta_{1\ell},\theta_{2\ell})+Q(\theta_{1\ell},\theta_{2\ell})h\|_+^2)dh+o_p(\frac{1}{\sqrt{n}})$$
$$=O_p(\frac{1}{\sqrt{n}}).$$

Again, the argument follows from lemma A.7. □

Appendix B. Proofs in section 3

B.1. **Proof of theorem 3.1.** I first show that for any $\epsilon > 0$, and $r > 0$, $\sup_{\Theta/\Theta_I^\epsilon} n^r f_n(\theta) = o_p(1)$, which is sufficient for $\lim_{n\to\infty}\mathbb{P}(\hat\theta_{1\ell} < \theta_{1\ell}-\epsilon)\to 0$. Let $\delta > 0$ be specified in assumption 3.3. Note that with probability approaching one

$$\sup_{\Theta/\Theta_I^\epsilon} n^r f_n(\theta) = \frac{n^r}{D_n}\exp(-\inf_{\Theta/\Theta_I^\epsilon} n\|\bar m(\theta)\|_+^2)$$
$$=\frac{n^r}{D_n}\exp(-\inf_{\Theta/\Theta_I^\epsilon} n\|\bar m(\theta)-\mathbb{E}m_1(\theta)+\mathbb{E}m_1(\theta)\|_+^2)\le\frac{n^{r+1}}{nD_n}\exp(-\frac{\delta_\epsilon^2 n}{4}) = o_p(1).$$

The last inequality is because of assumption 3.3, there exits at least one j such that $\inf_{\theta\in\Theta/\Theta_I^\epsilon}\mathbb{E}m_{(h)}(W_1,\theta) > \delta_\epsilon = \min\{C\epsilon,\delta\}$ and because of $\sup_{\theta\in\Theta}|\bar m(\theta)-\mathbb{E}m_1(\theta)| < \delta_\epsilon/2$ with probability one. The last equality holds because of lemma A.1 that nD_n is bounded away from 0 in probability. Given assumption 3.6, $\hat c$ decreases to zero at polynomial rate. This shows that for any ϵ, $\lim_{n\to\infty}\mathbb{P}(\hat\theta_{1\ell} < \theta_{1\ell}-\epsilon) = 0$.

Now I show that $\lim_{n\to\infty} \mathbb{P}(\hat{\theta}_{1\ell} > \theta_{1\ell} + \epsilon) \to 0$. There are two cases, $\theta_{1\ell} = \theta_{1u}$ and $\theta_{1\ell} < \theta_{1u}$. When $\theta_{1\ell} = \theta_{1u}$,

$$\mathbb{P}(\hat{\theta}_{1\ell} > \theta_{1\ell} + \epsilon) = \mathbb{P}(\hat{\tau}_\ell \geq F_{1n}(\theta_{1\ell} + \epsilon)) \leq \mathbb{P}\left(\frac{1}{2} \geq F_{1n}(\theta_{1\ell} + \epsilon)\right).$$

Since $F_{1n}(\theta_{1\ell} + \epsilon) \xrightarrow{p} 1$ by the previous argument, it follows the probability on the right hand side is arbitrarily small as n is large. It remains to show same conclusion holds when $\theta_{1\ell} < \theta_{1u}$.

$$\begin{aligned}\mathbb{P}(\hat{\theta}_{1\ell} > \theta_{1\ell} + \epsilon) = \mathbb{P}\left(\hat{c} \geq n \int_{\{\theta:\theta_1 \leq \theta_{1\ell}+\epsilon\}} \exp(-n\|\bar{m}(\theta)\|_+^2) d\theta\right) \\ \leq \mathbb{P}\left(\hat{c} \geq n \int_{\mathscr{B}_n} \exp(-n\|\bar{m}(\theta)\|_+^2) d\theta\right),\end{aligned}$$

where $\mathscr{B}_n = \{\theta : \theta_{1\ell} \leq \theta_1 \leq \theta_{1\ell} + \epsilon, \sqrt{n}\|\mathbb{E}m_1(\theta)\|_+ \leq \delta^*\}$ for some $\delta^* < \infty$. Note that by Lipschitz assumption 3.4, there exists some $\delta^{**} > 0$ such that $\mu(\mathscr{B}_n) \geq \frac{\delta^{**}\epsilon}{\sqrt{n}}$.

It thus follows,

$$\begin{aligned}\mathbb{P}\left(\hat{c} \geq n \int_{\mathscr{B}_n} \exp(-n\|\bar{m}(\theta)\|_+^2) d\theta\right) \leq \mathbb{P}\left(\hat{c} \geq n \int_{\mathscr{B}_n} \exp(-n\|\bar{m}(\theta)\|_+^2) d\theta\right) \\ \leq \mathbb{P}\left(\hat{c} \geq n\mu(\mathscr{B}_n) \inf_{\theta\in\mathscr{B}_n} \exp(-n\|\bar{m}(\theta)\|_+^2\right).\end{aligned}$$

To show that right hand side probability converges to zero as n increases, it sufficient to show that $\inf_{\theta\in\mathscr{B}_n} \exp(-n\|\bar{m}(\theta)\|_+^2)$ is bounded away from zero with probability approaching one. This is true because

$$\begin{aligned}1 \geq \inf_{\theta\in\mathscr{B}_n} \exp\left(-n\|\bar{m}(\theta)\|_+^2\right) = \exp\left(-\sup_{\theta\in\mathscr{B}_n} \|\bar{m}(\theta)\|_+^2\right) \\ \geq \exp\left(-\|\sup_{\theta\in\mathscr{B}_n} \Delta_n(\theta) + \sup_{\theta\in\mathscr{B}_n} \sqrt{n}\mathbb{E}m_1(\theta)\|_+^2\right) \\ \geq \exp\left(-\|\sup_{\theta\in\mathscr{B}_n} \Delta_n(\theta)\|_+^2 - \|\sup_{\theta\in\mathscr{B}_n} \sqrt{n}\mathbb{E}m_1(\theta)\|_+^2\right) \quad (17)\end{aligned}$$

Note that by assumption 3.5, $\|\sup_{\theta\in\mathscr{B}_n} \Delta_n(\theta)\|_+^2$ is $O_p(1)$. Also, $\|\sup_{\theta\in\mathscr{B}_n} \sqrt{n}\mathbb{E}m_1(\theta)\|_+^2$ is finite. Hence the right hand side of eq. (17) bounded away from zero with probability approaching one.

B.2. **Proof of lemma 3.1.** The proof to lemma 3.1 is summarized by the following three lemmas.

B.2.1. $\Theta_{I2}(\theta_{1\ell})$ *is a singleton.*

Lemma B.1. *Suppose that assumptions 3.1, 3.3 to 3.5 and 3.7 to 3.10 are satisfied. Suppose that $\Theta_{I2}(\theta_{1\ell})$ is a singleton. Let $\hat{\tau} = \min\{\frac{\hat{c}_\ell}{nD_n}, \frac{1}{2}\}$, where $\hat{c}_\ell \xrightarrow{p} c_\ell > 0$, let $\hat{\theta}_{1\ell} = F_{1n}^{-1}(\hat{\tau})$, then*

$$\sqrt{n}(\hat{\theta}_{1\ell} - \theta_{1\ell}) = O_p(1).$$

Proof. It is enough to show that for any $K \in \mathbb{R}$, $\lim_{n\to\infty} \mathbb{P}(\sqrt{n}(|\hat{\theta}_{1\ell} - \theta_{1\ell})| > K)$ converges to zero as K increases to infinity.

Part 1. I show that for $\lim_{n\to\infty} \mathbb{P}(\sqrt{n}(\hat{\theta}_{1\ell} - \theta_{1\ell}) \leq -K)$ converges to zero as K increases to infinity first.

$$\begin{aligned}
\mathbb{P}(\sqrt{n}(\theta_{1\ell} - \hat{\theta}_{1\ell}) \geq K) &= \mathbb{P}(\hat{c}_\ell \leq nD_n \int_{\underline{\theta}_1}^{\theta_{1\ell} - \frac{K}{\sqrt{n}}} \int_{\theta_2 \in \Theta_2} f_n(\theta_1, \theta_2) d\theta_2 d\theta_1) \\
&= \mathbb{P}(\hat{c}_\ell \leq D_n \int_{-\sqrt{n}(\theta_{1\ell} - \underline{\theta}_1)}^{-K} \int_{-\sqrt{n}(\theta_{2\ell} - \underline{\theta}_2)}^{\sqrt{n}(\overline{\theta}_2) - \theta_{2\ell}} f_n(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_{2\ell} + \frac{h_2}{\sqrt{n}}) dh_2 dh_1).
\end{aligned}$$

Following lemma A.7,

$$\begin{aligned}
&\mathbb{P}(\hat{c}_\ell \leq D_n \int_{-\infty}^{-K} \int_{-\sqrt{n}(\theta_{2\ell} - \underline{\theta}_2)}^{\sqrt{n}(\overline{\theta}_2) - \theta_{2\ell}} f_n(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_{2\ell} + \frac{h_2}{\sqrt{n}}) dh_2 dh_1) \\
&\qquad = \mathbb{P}(\hat{c}_\ell \leq \int_{\{h: h_1 \leq -K\}} N_n(h) dh) = \mathbb{P}(c_\ell \leq \int_{\{h: h_1 \leq -K\}} N_\infty(h) dh) + o_p(1).
\end{aligned}$$

The right hand side converges to zero as K increases to infinity, as already shown in lemma A.3.

Part 2. Now I show that $\lim_{n\to\infty} \mathbb{P}(\hat{\theta}_{1\ell} > \theta_{1\ell} + \frac{K}{\sqrt{n}}) \to 0$ as K increase. Suppose $\theta_{1u} > \theta_{1\ell}$ for now.

$$\begin{aligned}
\mathbb{P}(\hat{\theta}_{1\ell} \geq \theta_{1\ell} + \frac{K}{\sqrt{n}}) &= \mathbb{P}(c_\ell + o_p(1) \geq n \int_{\{\theta: \theta_1 \leq \theta_{1\ell} + \frac{K}{\sqrt{n}}\}} \exp(-n\|\bar{m}(\theta)\|_+^2)) \\
&\leq \mathbb{P}(c_\ell + o_p(1) \geq n \int_{\mathscr{B}_n} \exp(-n\|\bar{m}(\theta)\|_+^2) d\theta), \quad (18)
\end{aligned}$$

where

$$\mathscr{B}_n = \{\theta : \theta_{1\ell} + \frac{K}{4\sqrt{n}} \leq \theta_1 \leq \theta_{1\ell} + \frac{3K}{4\sqrt{n}}, d(\theta, \Theta_I) \leq \frac{\delta^*}{\sqrt{n}}\}.$$

By assumption 3.4, there exist $C_1 > 0$ such that

$$\max_j \sup_{\theta \in \mathscr{B}_n} |\mathbb{E}m_{(j)}(W_1, \theta)|_+ \leq \sup_{\theta \in \mathscr{B}_n} C_1 d(\theta, \Theta_I) = \frac{C_1 \delta^*}{4\sqrt{n}}.$$

It is not difficult to verify that $\mu(\mathscr{B}_n) = \frac{C_2 K \delta^*}{16n}$ for some $C_2 > 0$. It thus follows,

$$\mathbb{P}(c_\ell + o_p(1) \geq n \int_{\mathscr{B}_n} \exp(-n\|\bar{m}(\theta)\|_+^2) d\theta)$$

$$\geq \mathbb{P}(c_\ell + o_p(1) \geq n\mu(\mathscr{B}_n) \inf_{\theta\in\mathscr{B}_n} \exp(-n\|\bar{m}(\theta)\|_+^2)$$
$$= \mathbb{P}(c_\ell + o_p(1) \geq \frac{C_2 K \delta^*}{16} \inf_{\theta\in\mathscr{B}_n} \exp(-n\|\bar{m}(\theta)\|_+^2).$$

It remains to show that $\inf_{\theta\in\mathscr{B}_n} \exp(-n\|\bar{m}(\theta)\|_+^2)$ is bounded away from zero with probability approaching one. This is true since $\sup_{\theta\in\mathscr{B}_n} \|\sqrt{n}\mathbb{E}m_1(\theta)\|_+^2 \leq \frac{C_1\delta^*}{4} < +\infty$ and by the same argument as in eq. (17).

To complete part 2, it remains to show $\lim_{n\to\infty} \mathbb{P}(\hat{\theta}_{1\ell} > \theta_{1\ell} + \frac{K}{\sqrt{n}}) \to 0$ when $\theta_{1\ell} = \theta_{1u}$. In this case, θ_0 is point identified.

$$\mathbb{P}\left(\hat{\theta}_{1\ell} > \theta_{1\ell} + \frac{K}{\sqrt{n}}\right) = \mathbb{P}\left(\hat{\tau}_\ell \geq F_{1n}(\theta_{1\ell} + \frac{K}{\sqrt{n}})\right) \leq \mathbb{P}\left(\frac{1}{2} \geq F_{1n}(\theta_{1\ell} + \frac{K}{\sqrt{n}})\right).$$

The probability limit of $F_{1n}(\theta_{1\ell} + \frac{K}{\sqrt{n}})$ can be made arbitrarily close to one as K increases. So the conclusion follows.

Combine part 1 and 2, the statement of the Lemma follows. $\square$

B.2.2. $\Theta_{I2}(\theta_{1\ell})$ *is an interval.*

Lemma B.2. *Suppose that assumptions 3.1, 3.3 to 3.5 and 3.7 to 3.10 are satisfied. Suppose that* $\Theta_{I2}(\theta_{1\ell}) = [\theta_{2\ell}, \theta_{2u}]$. *Let* $\hat{\tau} = \min\{\frac{\hat{c}_\ell}{nD_n}, \frac{1}{2}\}$, *where* $\hat{c}_\ell \xrightarrow{p} c_\ell > 0$, *let* $\hat{\theta}_{1\ell} = F_{1n}^{-1}(\hat{\tau})$, *then with probability approaching 1,*

$$\sqrt{n}(\hat{\theta}_{1\ell} - \theta_{1\ell}) \to -\infty$$

Proof. I want to show that for any $K > 0$,

$$\lim_{n\to\infty} \mathbb{P}(\sqrt{n}(\theta_{1\ell} - \hat{\theta}_{1\ell}) > K) = 1.$$

Note that

$$\mathbb{P}(\sqrt{n}(\theta_{1\ell} - \hat{\theta}_{1\ell}) > K) = \mathbb{P}\left(\hat{c}_\ell \leq nD_n \int_{\underline{\theta}_1}^{\theta_{1\ell}-\frac{K}{\sqrt{n}}} \int_{\theta_2\in\Theta_2} f_n(\theta_1,\theta_2) d\theta_2 d\theta_1\right)$$
$$= \mathbb{P}\left(\hat{c}_\ell \leq \sqrt{n} D_n \int_{-\infty}^{-K} \int_{\theta_2\in\Theta_2} f_n(\theta_{1\ell} + h_1/\sqrt{n}, \theta_2) d\theta_2 dh_1\right) = \mathbb{P}(\hat{c}_\ell \leq A_n + B_n) \geq \mathbb{P}(\hat{c} \leq A_n),$$

where

$$A_n = \sqrt{n}\int_{-\infty}^{-K} \int_{\theta_2\in[\theta_{2\ell},\theta_{2u}]} D_n f_n(\theta_{1\ell} + h_1/\sqrt{n}, \theta_2) d\theta_2 dh_1$$
$$B_n = \sqrt{n}\int_{-\infty}^{-K} \int_{\theta_2\notin[\theta_{2\ell},\theta_{2u}]} D_n f_n(\theta_{1\ell} + h_1/\sqrt{n}, \theta_2) d\theta_2 dh_1$$

I will show that A_n diverges to $+\infty$ with probability approaching one.

By assumptions 3.4 and 3.7, there exists some $C_1 < 0$ such that (note that $h_1 \le 0$ here)

$$0 \le \max_j \sup_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \sqrt{n}\mathbb{E}m_{(j)}(W_1; \theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_2) \le C_1 h_1.$$

Hence

$$\begin{aligned}
A_n = \sqrt{n} \int_{-\infty}^{-K} \int_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \exp(-\|\Delta_n(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_2) + \sqrt{n}\mathbb{E}m_1(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_2)\|_+^2) d\theta_2 dh_1 \\
\ge \sqrt{n} \int_{-\infty}^{-K} \int_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \exp(-\|\Delta_n(\theta_{1\ell}, \theta_2) + C_1 h_1 + o_p(1)\|_+^2) d\theta_2 dh_1.
\end{aligned}$$

For every given K, A_n diverges to $+\infty$ since the integral is bounded away from zero with probability approaching one. □

Lemma B.3. *Suppose that assumptions 3.1, 3.3 to 3.5 and 3.7 to 3.10 are satisfied. Suppose that $\Theta_{I2}(\theta_{1\ell}) = [\theta_{2\ell}, \theta_{2u}]$. Let $\hat{\tau} = \min\{\frac{\hat{c}_\ell}{\sqrt{n}D_n}, \frac{1}{2}\}$, where $\hat{c}_\ell \xrightarrow{p} c_\ell > 0$, let $\hat{\theta}_{1\ell} = F_{1n}^{-1}(\hat{\tau})$, then with probability approaching 1,*

$$\sqrt{n}(\hat{\theta}_{1\ell} - \theta_{1\ell}) = O_p(1)$$

Proof. **Part 1**. Let $K > 0$ be arbitrary, I want to show that

$$\lim_{K \to +\infty} \lim_{n \to \infty} \mathbb{P}(\sqrt{n}(\theta_{1\ell} - \hat{\theta}_{1\ell}) > K) = 0.$$

Note that

$$\begin{aligned}
\mathbb{P}(\sqrt{n}(\theta_{1\ell} - \hat{\theta}_{1\ell}) > K) = \mathbb{P}(\hat{c}_\ell \le \sqrt{n} D_n \int_{\underline{\theta}_1}^{\theta_{1\ell} - \frac{K}{\sqrt{n}}} \int_{\theta_2 \in \Theta_2} f_n(\theta_1, \theta_2) d\theta_2 d\theta_1) \\
= \mathbb{P}(\hat{c}_\ell \le D_n \int_{-\infty}^{-K} \int_{\theta_2 \in \Theta_2} f_n(\theta_{1\ell} + h_1/\sqrt{n}, \theta_2) d\theta_2 dh_1) = \mathbb{P}(\hat{c}_\ell \le A_n + B_n),
\end{aligned}$$

where

$$\begin{aligned}
A_n &= \int_{-\infty}^{-K} \int_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} D_n f_n(\theta_{1\ell} + h_1/\sqrt{n}, \theta_2) d\theta_2 dh_1 \\
B_n &= \int_{-\infty}^{-K} \int_{\theta_2 \notin [\theta_{2\ell}, \theta_{2u}]} D_n f_n(\theta_{1\ell} + h_1/\sqrt{n}, \theta_2) d\theta_2 dh_1
\end{aligned}$$

B_n is $O_p(1/\sqrt{n})$ by the same argument in lemma B.1. It remains to show that $\lim_{n\to\infty} \mathbb{P}(\hat{c} \le A_n)$ decrease to zero as K increases to infinity.

By assumptions 3.4 and 3.7, there exists at least one j and some $C_2 < 0$ such that

$$\inf_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \sqrt{n}\mathbb{E}m_{(j)}(W_1, (\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_2)) \ge C_2 h_1 \ge 0.$$

Then with probability approaching one,

$$A_n = \int_{-\infty}^{-K} \int_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \exp(-\|\Delta_n(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_2) + \sqrt{n}\mathbb{E}m_1(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_2)\|_+^2) d\theta_2 dh_1$$
$$\leq (\theta_{2u} - \theta_{2\ell}) \int_{-\infty}^{-K} \exp(-| \inf_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \Delta_{jn}(\theta_{1\ell}, \theta_2) + C_2 h_1 + o_p(1)|_+^2) dh_1$$
$$= (\theta_{2u} - \theta_{2\ell}) \int_{-\infty}^{-K} \exp(-| \inf_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \Delta_j(\theta_{1\ell}, \theta_2) + C_2 h_1|_+^2) dh_1 + o_p(1).$$

The conclusion follows because $\hat{c} \xrightarrow{p} c > 0$ and $\inf_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \Delta(\theta_{1\ell}, \theta_2)$ is $O_p(1)$.

Part 2. I want to show that

$$\lim_{K \to +\infty} \lim_{n \to \infty} \mathbb{P}(\sqrt{n}(\hat{\theta}_{1\ell} - \theta_{1\ell}) > K) = 0.$$

The case in which $\theta_{1u} = \theta_{1\ell}$ is similar to part 1. Suppose $\theta_{1u} > \theta_{1\ell}$ for now.

$$\mathbb{P}\left(\hat{\theta}_{1\ell} \geq \theta_{1\ell} + \frac{K}{\sqrt{n}}\right) = \mathbb{P}\left(\hat{c}_\ell \geq \sqrt{n} \int_{\{\theta : \theta_1 \leq \theta_{1\ell} + \frac{K}{\sqrt{n}}\}} \exp(-n\|\bar{m}(\theta)\|_+^2)\right)$$
$$= \mathbb{P}\left(c_\ell + o_p(1) \geq \sqrt{n} \int_{\{\theta : \theta_1 \leq \theta_{1\ell} + \frac{K}{\sqrt{n}}\}} \exp(-n\|\bar{m}(\theta)\|_+^2)\right)$$
$$\leq \mathbb{P}\left(c_\ell + o_p(1) \geq \sqrt{n} \int_{\mathscr{B}_n} \exp(-n\|\bar{m}(\theta)\|_+^2) d\theta\right),$$

where $\mathscr{B}_n = \{\theta : \theta_{1\ell} \leq \theta_1 \leq \theta_{1\ell} + \frac{K}{\sqrt{n}}, d(\theta, \Theta_I) \leq \delta^* / \sqrt{n}\}$. By assumption 3.4, $\sup_{\theta \in \mathscr{B}_n} \|\mathbb{E}m_1(\theta)\|_+ \leq \sup_{\theta \in \mathscr{B}_n} C_1 d(\theta, \Theta_I) \leq \frac{\delta^*}{\sqrt{n}}$ for some $\delta^* > 0$.

It is not difficult to verify that $\mu(\mathscr{B}_n) \geq \frac{K\delta^{**}}{\sqrt{n}}$ for some δ^{**} under assumption 3.9. If assumption 3.9–(2) holds, i.e., Θ_I has non–empty interior around $(\theta_{1\ell}, \theta_2)$, this conclusion holds immediately. When Θ_I is convex (under assumption 3.9–1), just note that $\mathscr{B}_n$ always contains a triangle with three corner points: $(\theta_{1\ell}, \theta_{2\ell})$, $(\theta_{1\ell}, \theta_{2u})$ and $(\theta_{1\ell} + \frac{K}{\sqrt{n}}, (\theta_{2u} + \theta_{2\ell})/2)$. I can take $\delta^{**} = (\theta_{2u} - \theta_{2\ell})/2$.

It thus follows,

$$\mathbb{P}(c_\ell + o_p(1) \geq n \int_{\mathscr{B}_n} \exp(-n\|\bar{m}(\theta)\|_+^2) d\theta)$$
$$\geq \mathbb{P}(c_\ell + o_p(1) \geq n\mu(\mathscr{B}_n) \inf_{\theta \in \mathscr{B}_n} \exp(-n\|\bar{m}(\theta)\|_+^2)$$
$$= \mathbb{P}(c_\ell + o_p(1) \geq \frac{K\delta^*}{16} \inf_{\theta \in \mathscr{B}_n} \exp(-n\|\bar{m}(\theta)\|_+^2).$$

The limit probability converges to zero as K increases. □

B.3. **Proof of theorem 3.2.** Consistency is ensured by construction and theorem 3.1. To see this, I just need to show that for any $\epsilon > 0$, the following equation holds with probability approaching zero.

$$\frac{n\int_{\underline{\theta}_1}^{\theta_{1\ell}-\epsilon}\int_{\theta_2\in\Theta_2}\exp(-n\|\bar{m}(\theta_1,\theta_2)\|_+^2)d\theta_2 d\theta_1}{\sqrt{n}\int_{\theta_2\in\Theta_2}\exp(-n\|\bar{m}(\theta_{1\ell}-\epsilon,\theta_2)\|_+^2)d\theta_2} = \hat{c}.$$

Suppose that there exist one, called ϵ^* such that it holds with probability greater than zero; then by mean value theorem, with probability greater than zero, there exists an $\epsilon^{**} \leq \epsilon^*$ such that

$$\frac{\sqrt{n}(\theta_{1\ell}-\epsilon^*-\underline{\theta}_1)\int_{\theta_2\in\Theta_2}\exp(-n\|\bar{m}(\theta_1-\epsilon^{**},\theta_2)\|_+^2)d\theta_2 d\theta_1}{\int_{\theta_2\in\Theta_2}\exp(-n\|\bar{m}(\theta_{1\ell}-\epsilon^*,\theta_2)\|_+^2)d\theta_2} = \hat{c}.$$

This cannot happen since the numerator and the denominator have different rates and $\hat{c} \xrightarrow{p} c > 0$.

Now I establish $\sqrt{n}$–consistency. I show that in the $\sqrt{n}$–neighborhood of $\theta_{1\ell}$, with probability approaching one, there exists at least one solution to eq. (4).

Equation (4) can be written as

$$\frac{n\int_{\underline{\theta}_1}^{\hat{\theta}_{1\ell}^*}\int_{\theta_2\in\Theta_2}\exp(-n\|\bar{m}(\theta_1,\theta_2)\|_+^2)d\theta_2 d\theta_1}{\sqrt{n}\int_{\theta_2\in\Theta_2}\exp(-n\|\bar{m}(\hat{\theta}_{1\ell}^*,\theta_2)\|_+^2)d\theta_2} = \hat{c}_\ell.$$

Let $h_1 = \sqrt{n}(\theta_{1\ell}-\theta_{1\ell})$, consider a random variable q_n indexed by $h_1^* \in \mathbb{R}$.

$$q_n(h_1^*) = \frac{\int_{\sqrt{n}(\underline{\theta}_1-\theta_{1\ell})}^{h_1^*}\int_{\theta_2\in\Theta_2}\exp(-n\|\bar{m}(\theta_{1\ell}+\frac{h_1}{\sqrt{n}},\theta_2)\|_+^2)d\theta_2 dh_1}{\int_{\theta_2}\exp(-n\|\bar{m}(\theta_{1\ell}+\frac{h_1^*}{\sqrt{n}},\theta_2)\|_+^2)d\theta_2}. \tag{19}$$

I will show that for any $\epsilon > 0$, $\lim_{n\to\infty}\mathbb{P}(q_n(h_1^*) > \epsilon)$ can be made arbitrarily small if I let $h_1^* \to -\infty$; and for any $C > 0$, $\lim_{n\to\infty}\mathbb{P}(q_n(h_1^*) < C)$ can be made arbitrarily small if I let $h_1^* \to +\infty$. If those two statements are true, then by continuity of $q_n(\cdot)$, the probability of exists one $\tilde{h}_1$ such that $q_n(\tilde{h}_1) = c$ holds can be made arbitrarily close to one. By defining $\hat{\theta}_{1\ell}^* = \theta_{1\ell} + \tilde{h}/\sqrt{n}$, I show that the probability of eq. (4) having a solution approaches 1.

B.3.1. $h_1^* \to -\infty$. There are two cases: $\Theta_{I2}(\theta_{1\ell}) = [\theta_{2\ell},\theta_{2u}]$ and $\Theta_{I2}(\theta_{1\ell}) = \{\theta_{2\ell}\}$.

case 1. $\Theta_{I2}(\theta_{1\ell}) = [\theta_{2\ell},\theta_{2u}]$. Consider the denominator first. For given h_1^*, by lemma A.8

$$\int_{\theta_2\in\Theta_2}\exp(-\|\bar{m}(\theta_{1\ell}+\frac{h_1^*}{\sqrt{n}},\theta_2)\|_+^2)d\theta_2 = \int_{\theta_2\in\Theta_2}\exp(-\|\Delta_n(\theta_{1\ell}+\frac{h_1^*}{\sqrt{n}},\theta_2)+\sqrt{n}\mathbb{E}(\theta_{1\ell}+\frac{h_1^*}{\sqrt{n}},\theta_2)\|_+^2)d\theta_2$$

$$= \int_{\theta_2\in[\theta_{2\ell},\theta_{2u}]}\exp(-\|\Delta(\theta_{1\ell},\theta_2)+Q_1(\theta_{1\ell},\theta_2)h_1^*+o_p(1)\|_+^2)d\theta_2 + O_p(1/\sqrt{n})$$

$$= \int_{\theta_2 \in [\theta_{2\ell}\theta_{2u}]} \exp(-\|\psi(h_1^*, \theta_2) + o_p(1)\|_+^2) d\theta_2 + O_p(1/\sqrt{n}),$$

where $\psi(h_1, \theta_2) = \Delta(\theta_{1\ell}, \theta_2) + Q_1(\theta_{1\ell}, \theta_2) h_1$. Again, without loss of generality, I omit the sup–script $\mathcal{J}$ for notation simplicity.

On the other hand, the numerator can be written as

$$\begin{aligned}
&\int_{\sqrt{n}(\underline{\theta}_1 - \theta_{1\ell})}^{h_1^*} \int_{\theta_2 \in \Theta_2} \exp(-\|\bar{m}(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_2)\|_+^2) d\theta_2 dh_1 \\
&\quad = \int_{\sqrt{n}(\underline{\theta}_1 - \theta_{1\ell})}^{h_1^*} \int_{\theta_2 \in \Theta_2} \exp(-\|\Delta_n(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_2) + \sqrt{n}\mathbb{E}m_1(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_2)\|_+^2) d\theta_2 dh_1 \\
&\quad = \int_{\sqrt{n}(\underline{\theta}_1 - \theta_{1\ell})}^{h_1^*} \int_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \exp(-\|\Delta(\theta_{1\ell}, \theta_2) + Q_1(\theta_{1\ell}, \theta_2) h_1 + o_p(1)\|_+^2) d\theta_2 dh_1 + O_p(1/\sqrt{n}) \\
&\quad = \int_{\sqrt{n}(\underline{\theta}_1 - \theta_{1\ell})}^{h_1^*} \int_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \exp(-\|\psi(h_1, \theta_2) + o_p(1\|_+^2) d\theta_2 dh_1 + O_p(1/\sqrt{n}),
\end{aligned}$$

Hence

$$\begin{aligned}
&\mathbb{P}(q_n(h_1^*) \geq \epsilon) = \mathbb{P}\left(\frac{\int_{\sqrt{n}(\underline{\theta}_1 - \theta_{1\ell})}^{h_1^*} \int_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \exp(-\|\psi(h_1, \theta_2) + o_p(1\|_+^2) d\theta_2 dh_1 + O_p(1/\sqrt{n})}{\int_{\theta_2 \in [\theta_{2\ell}\theta_{2u}]} \exp(-\|\psi(h_1^*, \theta_2) + o_p(1\|_+^2) d\theta_2 + O_p(1/\sqrt{n})} \geq \epsilon \right) \\
&\leq \mathbb{P}\left(\frac{\int_{\sqrt{n}(\underline{\theta}_1 - \theta_{1\ell})}^{h_1^*} \int_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \exp(-\|\psi(h_1, \theta_2)\|^2) d\theta_2 dh_1 + O_p(1/\sqrt{n})}{\int_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \exp(-\|\psi(h_1^*, \theta_2)\|^2) d\theta_2 + O_p(1/\sqrt{n})} \geq \epsilon | \inf_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \psi(h_1^*, \theta_2) \geq 0 \right) \\
&\qquad \times \mathbb{P}(\inf_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \psi(h_1^*, \theta_2) \geq 0) + \mathbb{P}(\inf_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \psi(h_1^*, \theta_2) < 0) + o_p(1).
\end{aligned}$$

$\lim_{n \to \infty} \mathbb{P}(\inf_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \psi(h_1^*, \theta_2) < 0)$ can be made arbitrarily small by letting $h_1^* \to -\infty$. It remains to verify that the limit of the first probability on the right hand side goes to zero as $h_1^* \to -\infty$. Note that

$$\begin{aligned}
&\mathbb{P}\left(\frac{\int_{\sqrt{n}(\underline{\theta}_1 - \theta_{1\ell})}^{h_1^*} \int_{\theta_2 \in [\theta_{1\ell}, \theta_{2u}]} \exp(-\|\psi(h_1, \theta_2)\|^2) d\theta_2 dh_1}{\int_{\theta_2 \in [\theta_{1\ell}, \theta_{2u}]} \exp(-\|\psi(h_1^*, \theta_2)\|^2) d\theta_2} \geq \epsilon \right) \\
&\quad = \mathbb{P}\left(\int_{\theta_2 \in [\theta_{1\ell}, \theta_{2u}]} \left[\int_{-\infty}^{h_1^*} \exp(-\|\psi(h_1, \theta_2)\|^2) dh_1 - \epsilon \exp(-\|\psi(h_1^*, \theta_2)\|^2) \right] d\theta_2 d \geq 0 \right) \\
&\quad \leq \mathbb{P}\left(\sup_{\theta_2 \in [\theta_{1\ell}, \theta_{2u}]} \left\{ \int_{-\infty}^{h_1^*} \exp(-\|\psi(h_1, \theta_2)\|^2) dh_1 - \epsilon \exp(-\|\psi(h_1^*, \theta_2)\|^2) \right\} \geq 0 \right) \\
&\quad = \mathbb{P}\left(\sup_{\theta_2 \in [\theta_{1\ell}, \theta_{2u}]} \frac{\int_{-\infty}^{h_1^*} \exp(-\|\psi(h_1, \theta_2)\|^2) dh_1}{\exp(-\|\psi(h_1^*, \theta_2)\|^2} > \epsilon \right)
\end{aligned}$$

So It is sufficient to show that

$$\lim_{h_1^* \to -\infty} \mathbb{P}\left(\sup_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \frac{\int_{-\infty}^{h_1^*} \exp(-\|\psi(h_1, \theta_2)\|^2) dh_1}{\exp(-\|\psi(h_1^*, \theta_2)\|^2)} > \epsilon \right) = 0.$$

This is true because $\sup_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \|\Delta(\theta_{1\ell}, \theta_2)\|$ is bounded in probability.

Case 2. $\Theta_{I2}(\theta_{1\ell}) = \{\theta_{2\ell}\}$.

In this case, following the above argument, and ignoring the supscript $\mathcal{J}$ for notational simplicity, it can be shown that

$$q_n(h_1^*) = \frac{\int_{\sqrt{n}(\underline{\theta}_1 - \theta_{1\ell})}^{h_1^*} \int_{h_2 \in \mathbb{R}} \exp(-\|\Delta(\theta_{1\ell}, \theta_{2\ell}) + Q(\theta_{1\ell}, \theta_{2\ell})h)\|_+^2) dh + o_p(1)}{\int_{h_2 \in \mathbb{R}} \exp(-\|\Delta(\theta_{1\ell}, \theta_{2\ell}) + Q(\theta_{1\ell}, \theta_{2\ell})(h_1^*, h_2)')\|_+^2) dh_2 + o_p(1)}.$$

The probability limit of right hand side can be made arbitrarily small as h_1^* decreases to $-\infty$ by similar argument as in case 1.

B.3.2. $h_1^* \to +\infty$. Now I show that when $h_1^* \to +\infty$, $q_n(h_1^*)$ diverges with probability 1. Note that if $\theta_{1\ell} = \theta_{1u}$, the conclusion holds immediately (this is because of the denominator converges to zero and the numerator is bounded away from zero). So I focus on $\theta_{2u} > \theta_{1\ell}$.

$$q_n(h_1^*) = \frac{\int_{\sqrt{n}(\underline{\theta}_1 - \theta_{1\ell})}^{h_1^*} \int_{\theta_2} \exp(-n\|\bar{m}(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_2)\|_+^2) d\theta_2 dh_1}{\int_{\theta_2} \exp(-n\|\bar{m}(\theta_{1\ell} + \frac{h_1^*}{\sqrt{n}}, \theta_2)\|_+^2) d\theta_2}.$$

$$\geq \frac{\int_0^{h_1^*} \int_{\theta_2} \exp(-n\|\bar{m}(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_2)\|_+^2) d\theta_2 dh_1}{\int_{\theta_2} \exp(-n\|\bar{m}(\theta_{1\ell} + \frac{h_1^*}{\sqrt{n}}, \theta_2)\|_+^2) d\theta_2}.$$

If $\Theta_{I2}(\theta_{1\ell})$ is singleton, the the denominator is of order $1/\sqrt{n}$, it follows from the similar argument as in section 1 in the proof of lemma B.1 (with h_1^* taking the place of K) that $q_n(h_1^*)$ diverges in probability. If $\Theta_{I2}(\theta_{1\ell})$ is an interval, then the denominator is of order $O_p(1)$, it follows from the similar argument as in the proof of lemma B.3 (with h_1^* taking the place of K) that $q_n(h_1^*)$ diverges in probability.

Appendix C. Proofs in section 4

C.1. **proof to lemma 4.1.** The results follows from lemmas A.7 and A.8.The continuity of random variable ξ_ℓ^A and ξ_ℓ^B holds because the integration, $\exp(\cdot)$ and $\|\cdot\|_+$ are all continuous operations and Δ is continuous random process.

C.2. **Proof of theorem 4.1 .** Lemmas C.1 to C.3 which will be used in the rest of the proof.

Lemma C.1. *Suppose assumptions 3.5 and 4.3 are satisfied. Then for any $\theta \in \Theta_I$, $\gamma_{jn}(\theta) \xrightarrow{p} 1$ if $\mathbb{E}m_{(j)}(W_i, \theta) = 0$; $\gamma_{jn}(\theta) \xrightarrow{p} 0$ if there exist a δ such that $\mathbb{E}m_{(j)}(W_i, \theta) < -\delta$.*

Proof. Suppose j is a binding moment. If $\frac{1}{n}\sum_i m_{(j)}(W_i, \theta) \geq 0$, then $\gamma_{jn}(\theta_{1\ell}, \theta_2) = 1$; so I only consider the case $\frac{1}{n}\sum_i m_{(j)}(W_i, \theta) < 0$.

$$\gamma_{jn}(\theta) = \exp\left(-\frac{\beta_n}{n}\left|\frac{1}{\sqrt{n}}\sum_i m_{(j)}(W_i, \theta)\right|^2\right)$$
$$= \exp\left(-\frac{\beta_n}{n}\left|\Delta_{nj}(\theta) + \sqrt{n}\mathbb{E}m_{(j)}(W_i, \theta)\right|^2\right) \xrightarrow{p} 1.$$

The convergence is because $\Delta_{nj}(\theta_{1\ell}, \theta_2) \xrightarrow{d} \Delta_j(\theta_{1\ell}, \theta_2)$ and $\beta_n/n \to 0$ by assumption 4.3.

Now I consider the case in which there exist a δ such that $\mathbb{E}m_{(j)}(W_i, \theta) < -\delta$, then for large n

$$\gamma_{jn}(\theta) = \exp\left(-\frac{\beta_n}{n}\left|\frac{1}{\sqrt{n}}\sum_i m_{(j)}(W_i, \theta)\right|^2\right)$$
$$= \exp\left(-\frac{\beta_n}{n}\left|\Delta_{nj}(\theta) + \sqrt{n}\mathbb{E}m_{(j)}(W_i, \theta)\right|^2\right) \leq \exp\left(-\frac{\beta_n}{n}\left|\Delta_{nj}(\theta) - \sqrt{n}\delta/2\right|^2\right) \xrightarrow{p} 0.$$

The term on the right hand side converges in probability to zero because $\beta_n \to \infty$ by assumption 4.3. □

Lemma C.2. *Suppose that assumptions 3.1 to 3.3, 3.5, 3.7, 3.8, 4.1 and 4.3 are satisfied. Then $\omega_{\ell n} \xrightarrow{p} 0$ and $\alpha_{\ell n}/n = 1 + o_p(1)$ if $\Theta_{I2}(\theta_{1\ell})$ only contains countable disjoint singletons; $\omega_{\ell n} \xrightarrow{p} 1$ and $\alpha_{\ell n}/\sqrt{n} = 1 + o_p(1)$ if $\Theta_{I2}(\theta_{1\ell})$ contains at least one interval with positive length.*

Proof. Suppose first that $\Theta_{I2}(\theta_{1\ell})$ is an interval with positive length: $\theta_{2u} > \theta_{2\ell}$, then

$$M_n(\theta_{1\ell}) = \sqrt{\beta_n}\int \exp(-\|\sqrt{\beta_n n^{-1}}\Delta_n(\theta_{1\ell}, \theta_2) + \sqrt{\beta_n}\mathbb{E}m_1(\theta_{1\ell}, \theta_2)\|_+^2)d\theta_2$$
$$= \sqrt{\beta_n}\int_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \exp(-\|\sqrt{\beta_n n^{-1}}\Delta_n(\theta_{1\ell}, \theta_2) + \sqrt{\beta_n}\mathbb{E}m_1(\theta_{1\ell}, \theta_2)\|_+^2)d\theta_2$$
$$+ \sqrt{\beta_n}\int_{\theta_2 \notin [\theta_{2\ell}, \theta_{2u}]} \exp(-\|\sqrt{\beta_n n^{-1}}\Delta_n(\theta_{1\ell}, \theta_2) + \sqrt{\beta_n}\mathbb{E}m_1(\theta_{1\ell}, \theta_2)\|_+^2)d\theta_2$$
$$= A_n + B_n.$$

I show $A_n/\sqrt{\beta_n} \xrightarrow{p} (\theta_{2u} - \theta_{1\ell})$ first. Note that the unbinding moments automatically drop out because $\beta_n \to \infty$. So for notational simplicity and without loss of generality, assume that

$\mathbb{E}m_1(\theta_{1\ell}, \theta_2) = 0$ for all $\theta_2 \in [\theta_{2\ell}, \theta_{2u}]$.

$$A_n/\sqrt{\beta_n} = \int_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \exp(-\|\sqrt{\beta_n n^{-1}}\Delta_n(\theta_{1\ell}, \theta_2)\|_+^2) d\theta_2 \xrightarrow{p} \int_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} d\theta_2 = \theta_{2u} - \theta_{2\ell}.$$

Now I show there exist some $C > 0$ such that $B_n \leq C$ with probability approaching one. By assumptions 3.3 and 3.7, for $h_2 > 0$, there exists some $C_3 > 0$ such that

$$\|\sqrt{\beta_n}\mathbb{E}m_1(\theta_{1\ell}, \theta_2^* + \frac{h_2}{\sqrt{\beta_n}}\|_+ \geq C_3 h_2.$$

Hence I just need to show that for large n

$$\begin{aligned}
&\int_{\theta_2 > \theta_{2u}} \sqrt{\beta_n} \exp(-\|\sqrt{\beta_n n^{-1}}\Delta_n(\theta_{1\ell}, \theta_2) + \sqrt{\beta_n}\mathbb{E}m_1(\theta_{1\ell}, \theta_2)\|_+^2) d\theta_2 \\
&\quad = \int_{h_2 \in \mathbb{R}^+} \exp(-\|\sqrt{\beta_n n^{-1}}\Delta_n(\theta_{1\ell}, \theta_{2u} + \frac{h_2}{\sqrt{\beta_n}}) + \sqrt{\beta_n}\mathbb{E}m_1(\theta_{1\ell}, \theta_{2u} + \frac{h_2}{\sqrt{\beta_n}})\|_+^2) dh_2 \\
&\quad \leq \int_{h_2 \in \mathbb{R}} \exp(\|\sqrt{\beta_n n^{-1}}\Delta_n(\theta_{1\ell}, \theta_{2u} + \frac{h_2}{\sqrt{\beta_n}})\|^2 - \|\sqrt{\beta_n}\mathbb{E}m_1(\theta_{1\ell}, \theta_{2u} + \frac{h_2}{\sqrt{\beta_n}}\|_+^2) dh_2 \\
&\qquad\qquad \leq \int_{h_2 \in \mathbb{R}} \exp(\|\sqrt{\beta_n n^{-1}}\Delta_n(\theta_{1\ell}, \theta_{2u} + \frac{h_2}{\sqrt{\beta_n}})\|^2 - \|C_3 h_2\|_+^2) dh_2 \xrightarrow{p} C^* > 0.
\end{aligned}$$

The convergence holds because of assumption 3.5. Similar argument can be applied to the region $\theta_2 < \theta_{2\ell}$.

Remember that $\nu(x) \to 1$ when $x \to 0$; $\nu(x) \to 0$ exponentially when $x \to \infty$. Since $M_n(\theta_{1\ell}) = O_p(\sqrt{\beta_n})$ and $\sqrt{\beta_n}/\log n \to \infty$ at polynomial rate (by assumption 4.3), it follows that

$$\sqrt{n}\omega_{\ell n} = \sqrt{n}\nu\left(\frac{M_n(\theta_{1\ell})}{\log n}\right) \xrightarrow{p} 0,$$

and

$$\frac{\alpha_{\ell n}}{\sqrt{n}} = (1 - \omega_{\ell n}) + \sqrt{n}\omega_{\ell n} \xrightarrow{p} 1.$$

When $\Theta_{I2}(\theta_{1\ell})$ contains only singletons, $M_n(\theta_{1\ell})/\log n = o_p(1)$, it follows that $\omega_{\ell n} \xrightarrow{p} 1$ and

$$\frac{\alpha_{\ell n}}{n} = \frac{1}{\sqrt{n}}(1 - \omega_{\ell n}) + \omega_{\ell n} \xrightarrow{p} 1.$$

□

Lemma C.3. *Suppose that assumptions 3.1 to 3.3, 3.5, 3.7, 3.8 and 4.1 to 4.3 are satisfied. Then* $\tilde{\xi}_{\ell n} \xrightarrow{d} \xi_\ell^A$ *if* $\Theta_{I2}(\theta_{1\ell})$ *is a singleton.* $\tilde{\xi}_{\ell n} \xrightarrow{d} \xi_\ell^B$ *if* $\Theta_{I2}(\theta_{1\ell})$ *is an interval.*

Proof. The convergence follows from lemmas C.1 and C.2. □

Now I start to show the proposed confidence set has correct coverage probability.

Case 1. $T > 0$. In this case, $\nu\left(\beta_n/n\hat{T}\right) \xrightarrow{p} 1$ and $\sqrt{n}\nu\left(\beta_n/n\hat{T}\right)$ diverges. The values c_k^{I} are computed as,

$$\mathbb{P}\left(c_\ell^{\mathrm{I}} \leq \tilde{\xi}_{\ell n}\right) = 1-\alpha, \quad \mathbb{P}\left(c_u^{\mathrm{I}} \leq \tilde{\xi}_{un}\right) = 1-\alpha. \tag{20}$$

Let $\theta_\lambda = \lambda\theta_{1\ell} + (1-\lambda)\theta_{1u}$, $\lambda \in [0,1]$. Let $\mathcal{P}_n(\lambda)$ be the probability of θ_λ belongs to the confidence set.

$$\begin{aligned}\mathcal{P}_n(\lambda) &= \mathbb{P}\left(\theta_\lambda \in \Theta_{\alpha n}^{\mathrm{I}}\right) = \mathbb{P}\left(\int_{\underline{\theta}_1}^{\theta_\lambda} f_{1n}(\theta_1)d\theta_1 \geq \frac{c_\ell^{\mathrm{I}}}{\alpha_{un}D_n}, \int_{\theta_\lambda}^{\bar{\theta}_1} f_{1n}(\theta_1)d\theta_1 \geq \frac{c_u^{\mathrm{I}}}{\alpha_{\ell n}D_n}\right)\\ &= \mathbb{P}\left(\tilde{\xi}_{\ell n} + o_p(1) + \alpha_{\ell n}D_n\int_{\theta_{1\ell}}^{\theta_\lambda} f_{1n}(\theta_1)d\theta_1 \geq c_\ell^{\mathrm{I}}, \alpha_{un}D_n\int_{\theta_\lambda}^{\theta_{1u}} f_{1n}(\theta_1)d\theta_1 + \tilde{\xi}_{un} + o_p(1) \geq c_u^{\mathrm{I}}\right).\end{aligned}$$

Note that if $T > 0$, then for any $\lambda \in (0,1)$, $\alpha_{\ell n}D_n \int_{\theta_\lambda}^{\theta_{1u}} f_{1n}(\theta_1)d\theta_1$ or $\alpha_{un}D_n \int_{\theta_{1\ell}}^{\theta_\lambda} f_{1n}(\theta_1)d\theta_1$ (or both) diverges in probability. Hence $\mathcal{P}_n(\lambda)$ is minimized at $\lambda^* = 0$ or $\lambda^* = 1$ for large n. In both cases $\lim_{n\to\infty}\mathcal{P}_n(\lambda^*) = 1-\alpha$ because of eq. (20). This shows that

$$\lim_{n\to\infty} \inf_{\lambda\in[0,1]} \inf_{\{T_n:\sqrt{n}T_n \geq \log n\}} \mathcal{P}_n(\lambda) = 1-\alpha.$$

Case 2. $T = 0$. In this case, $\sqrt{n}\nu\left(\beta_n/n\hat{T}\right) \xrightarrow{p} 0$. The values c_k^{I} are computed as,

$$\mathbb{P}\left(c_\ell^{\mathrm{I}} \leq \tilde{\xi}_{\ell n}, c_u^{\mathrm{I}} \leq \tilde{\xi}_{un}\right) = 1-\alpha. \tag{21}$$

Again,

$$\begin{aligned}\mathcal{P}_n(\lambda) &= \mathbb{P}(\theta_\lambda \in \Theta_{\alpha n}^{\mathrm{I}}) = \mathbb{P}(\int_{\underline{\theta}_1}^{\theta_\lambda} f_{1n}(\theta_1)d\theta_1 \geq \frac{c_\ell^{\mathrm{I}}}{\alpha_{\ell n}D_n}, \int_{\theta_\lambda}^{\bar{\theta}_1} f_{1n}(\theta_1)d\theta_1 \geq \frac{c_u^{\mathrm{I}}}{\alpha_{un}D_n})\\ &= \mathbb{P}(\tilde{\xi}_{\ell n} + o_p(1) + \alpha_{\ell n}D_n\int_{\theta_{1\ell}}^{\theta_\lambda} f_{1n}(\theta_1)d\theta_1 \geq c_\ell^{\mathrm{I}}, \alpha_{un}D_n\int_{\theta_\lambda}^{\theta_{1u}} f_{1n}(\theta_1)d\theta_1 + \tilde{\xi}_{un} + o_p(1) \geq c_u^{\mathrm{I}}).\end{aligned}$$

The validity is ensured by eq. (21) since $\theta_{1\ell} = \theta_\lambda = \theta_{1u}$.

C.3. **Proof of theorem 4.2.** Let

$$\hat{\xi}_\ell = \hat{\omega}_\ell\psi_n(\hat{\xi}_\ell^A) + (1-\hat{\omega}_\ell)\hat{\xi}_\ell^B$$

be computed from algorithm 1. The target is to show that,

$$\hat{\xi}_\ell \xrightarrow{d} \tilde{\xi}_{\ell n}.$$

The proof needs Lemmas C.4 to C.9.

Lemma C.4. *Suppose that the assumptions required by theorem 4.2 are satisfied. Then $\hat{\theta}_\ell \xrightarrow{p} \theta_\ell$ when $\Theta_{I2}(\theta_{1\ell})$ is a singleton.*

Proof. This follows automatically from theorem 3.1. □

Lemma C.5. *Suppose that the assumptions required by theorem 4.2 are satisfied. Then $\hat{\alpha}_\ell/\alpha_{\ell n} \xrightarrow{p} 1$.*

Proof. Suppose $\theta_{2u} > \theta_{2\ell}$, then

$$\begin{aligned}
\hat{M}_\ell = \sqrt{\beta_n} \int & \exp(-\| \sqrt{\beta_n n^{-1}} \Delta_n(\hat{\theta}_{1\ell}, \theta_2) + \sqrt{\beta_n} \mathbb{E} m_1(\hat{\theta}_{1\ell}, \theta_2) \|_+^2) d\theta_2 \\
&= \sqrt{\beta_n} \int_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \exp(-\| \sqrt{\beta_n n^{-1}} \Delta_n(\hat{\theta}_{1\ell}, \theta_2) + \sqrt{\beta_n} \mathbb{E} m_1(\hat{\theta}_{1\ell}, \theta_2) \|_+^2) d\theta_2 \\
&+ \sqrt{\beta_n} \int_{\theta_2 \notin [\theta_{2\ell}, \theta_{2u}]} \exp(-\| \sqrt{\beta_n n^{-1}} \Delta_n(\hat{\theta}_{1\ell}, \theta_2) + \sqrt{\beta_n} \mathbb{E} m_1(\hat{\theta}_{1\ell}, \theta_2) \|_+^2) d\theta_2 \\
&= \hat{A} + \hat{B}.
\end{aligned}$$

I show $\hat{A}/\sqrt{\beta_n} \xrightarrow{p} \theta_{2u} - \theta_{2\ell}$ and for some $C > 0$, $\hat{B} \le C$ with probability 1. For some $\theta_{1\ell}^*$ between $\theta_{1\ell}$ and $\hat{\theta}_{1\ell}$

$$\begin{aligned}
\hat{A}/\sqrt{\beta_n} &= \int_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \exp(-\| \sqrt{\beta_n n^{-1}} \Delta_n(\hat{\theta}_{1\ell}, \theta_2) + \sqrt{\beta_n} Q_1(\theta_{1\ell}^*, \theta_2)(\hat{\theta}_{1\ell} - \theta_{1\ell}) + o_p(1) \|_+^2) d\theta_2 \\
&= \int_{\theta_2 \in [\theta_{2\ell}, \theta_{2u}]} \exp(-\| \sqrt{\beta_n n^{-1}} \Delta_n(\hat{\theta}_{1\ell}, \theta_2) + \sqrt{\beta_n n^{-1}} Q_1(\theta_{1\ell}^*, \theta_2) \sqrt{n}(\hat{\theta}_{1\ell} - \theta_{1\ell}) + o_p(1) \|_+^2) d\theta_2 \\
&\xrightarrow{p} \theta_{2u} - \theta_{2\ell}.
\end{aligned}$$

The convergence in probability is because of assumption 4.3 and that $\hat{\theta}_{1\ell}$ is $\sqrt{n}$–consistent.

Now I show $\hat{B}/\sqrt{\beta_n} \to 0$. For some constant C_3,

$$\begin{aligned}
\int_{\theta_2 > \theta_{2u}} & \sqrt{\beta_n} \exp(-\| \sqrt{\beta_n n^{-1}} \Delta_n((\hat{\theta}_{1\ell}, \theta_2) + \sqrt{\beta_n} \mathbb{E} m_1((\hat{\theta}_{1\ell}, \theta_2) \|_+^2) d\theta_2 \\
&= \int_{h_2 \in \mathbb{R}^+} \exp(-\| \sqrt{\beta_n n^{-1}} \Delta_n(\hat{\theta}_{1\ell}, \theta_{2u} + \frac{h_2}{\sqrt{\beta_n}}) + \sqrt{\beta_n} \mathbb{E} m_1(\hat{\theta}_{1\ell}, \theta_{2u} + \frac{h_2}{\sqrt{\beta_n}}) \|_+^2) dh_2 \\
&\le \int_{h_2 \in \mathbb{R}} \exp(\| \sqrt{\beta_n n^{-1}} \Delta_n(\hat{\theta}_{1\ell}, \theta_{2u} + \frac{h_2}{\sqrt{\beta_n}}) \|^2 - \| \sqrt{\beta_n} \mathbb{E} m_1(\hat{\theta}_{1\ell}, \theta_{2u} + \frac{h_2}{\sqrt{\beta_n}} \|_+^2) dh_2 \\
&\le \int_{h_2 \in \mathbb{R}} \exp(\| \sqrt{\beta_n n^{-1}} \Delta_n(\hat{\theta}_{1\ell}, \theta_{2u} + \frac{h_2}{\sqrt{\beta_n}}) \|^2 - \| C_3 h_2 \|_+^2) dh_2 \xrightarrow{p} C^* > 0.
\end{aligned}$$

The last inequality holds because $(\hat{\theta}_{1\ell}, \theta_{2u} + h_2/\sqrt{\beta_n}) \notin \Theta_I$ for large sample as a result of $\hat{\theta}_{1\ell}$ is roon–n–consistent and $\beta_n/n \to 0$. Hence,

$$\sqrt{\beta_n}\mathbb{E}m_1(\hat{\theta}_{1\ell}, \theta_{2u} + \frac{h_2}{\sqrt{\beta_n}}) \geq C_3 h_2.$$

Now consider the case where $\Theta_{I2}(\theta_{1\ell})$ is a singleton. The proof will be similar to the proof of part B_n above.

The result that $\hat{\alpha}_\ell/\alpha_{\ell n} \xrightarrow{p} 1$ follows by a similar argument as in lemma C.2. □

Lemma C.6. *Suppose that the assumptions required by theorem 4.2 are satisfied, then* $\sup_{\theta_2 \in \Theta_2} \|\hat{\gamma}(\hat{\theta}_{1\ell}, \theta_2) - \gamma_n(\theta_{1\ell}, \theta_2)\| = o_p(1)$.

Proof. It is sufficient to show the conclusion holds for each element γ_{jn}. Suppose first that $\mathbb{E}m_{(j)}(W_i, (\theta_{1\ell}, \theta_2)) = 0$ for some $K > 0$. I know that in this case $\gamma_{jn} \xrightarrow{p} 1$. It remains to show that $\hat{\gamma}_{jn}(\hat{\theta}_{1\ell}, \theta_2) \xrightarrow{p} 1$ too. If $\frac{1}{\sqrt{n}} \sum_i m_{(j)}(W_i, (\hat{\theta}_{1\ell}, \theta_2)) \geq 0$, $\hat{\gamma}_{jn}(\hat{\theta}_{1\ell}, \theta_2) = 1$; so I only consider the case in which $\frac{1}{\sqrt{n}} \sum_i m_{(j)}(W_i, (\hat{\theta}_{1\ell}, \theta_2)) < 0$.

$$\begin{aligned}
\hat{\gamma}_{jn}(\hat{\theta}_{1\ell}, \theta_2) &= \exp\left(-\frac{\beta_n}{n}\left|\frac{1}{\sqrt{n}}\sum_i m_{(j)}(W_i, (\hat{\theta}_{1\ell}, \theta_2))\right|^2\right) \\
&= \exp\left(-\frac{\beta_n}{n}\left|\Delta_{nj}(\theta) + \mathbb{E}m_{(j)}(W_i, (\theta_{1\ell}, \theta_2)) + Q_1(\theta_{1\ell}, \theta_2)\sqrt{n}(\hat{\theta}_{1\ell} - \theta_{1\ell})\right|^2\right) \xrightarrow{p} 1.
\end{aligned}$$

The convergence is because $\Delta_{nj}(\theta_{1\ell}, \theta_2) \xrightarrow{d} \Delta_j(\theta_{1\ell}, \theta_2)$, $\frac{\beta_n}{n} \to 0$ by assumption 4.3, as well as $\sqrt{n}(\hat{\theta}_{1\ell} - \theta_{1\ell}) = O_p(1)$. Note that the convergence holds uniformly over θ_2 by assumptions 3.1 and 3.7.

Now I consider the case in which there exist a δ such that $\mathbb{E}m_{(j)}(W_i, (\theta_{1\ell}, \theta_2)) < -\delta$, then for large n,

$$\begin{aligned}
\hat{\gamma}_{jn}(\hat{\theta}_{1\ell}, \theta_2) &= \exp\left(-\frac{\beta_n}{n}\left|\frac{1}{\sqrt{n}}\sum_i m_{(j)}(W_i, (\hat{\theta}_{1\ell}, \theta_2))\right|^2\right) \\
&= \exp\left(-\frac{\beta_n}{n}\left|\Delta_{nj}(\theta) + \sqrt{n}\mathbb{E}m_{(j)}(W_i, (\theta_{1\ell}, \theta_2)) + Q_1(\theta_{1\ell}, \theta_2)\sqrt{n}(\hat{\theta}_{1\ell} - \theta_{1\ell})\right|^2\right) \\
&\leq \exp\left(-\frac{\beta_n}{n}\left|\Delta_{nj}(\theta) - \sqrt{n}\delta/2 + Q_1(\theta_{1\ell}, \theta_2)\sqrt{n}(\hat{\theta}_{1\ell} - \theta_{1\ell})\right|^2\right).
\end{aligned}$$

The term on the right hand side converges in probability to zero. □

Lemma C.7. *Suppose that the assumptions required by theorem 4.2 are satisfied, then* $\sup_{\theta_2} \|\hat{Q}(\hat{\theta}_{1\ell}, \theta_2) - Q(\theta_{1\ell}, \theta_2)\| = o_p(1)$.

Proof. The uniform consistency of $\hat{Q}$ follows from the assumption 4.4 and the compactness of Θ, (see lemma 2.1, Newey and McFadden, 1994). □

Lemma C.8. *Suppose that the assumptions required by theorem 4.2 are satisfied, then* $\sup_{\theta_2,\theta_2',j,j'} |\hat{H}_{\theta_2,\theta_2',j,j'} - H_{\theta_2,\theta_2',j,j'}| = o_p(1)$.

Proof. This follows from assumptions 4.4 and 4.5 and the compactness of Θ. □

Let $\mathscr{T}_1, \mathscr{T}_2 \cdots$ be a sequence of compact subsets of $\mathbb{R}^-$ such that $\cup_{t=1}^{\infty} \mathscr{T}_t = \mathbb{R}^-$.

Lemma C.9. *Suppose that the assumptions required by theorem 4.2 are satisfied, then* $\hat{\Delta} \xrightarrow{w} \Delta$ *on* $\mathscr{L}^{\infty}(\mathscr{T}_t \times \Theta_2)$.

Proof. The weak convergence of $\hat{\Delta}$ follows from the consistency of $\hat{H}_{\theta_2,\theta_2',j,j'}$ to $H_{\theta_2,\theta_2',j,j'}$. □

Now I can show that $\hat{\xi}_\ell^A \xrightarrow{d} \xi_\ell^A$ and $\hat{\xi}_\ell^B \xrightarrow{d} \xi_\ell^B$. Given lemmas C.5 to C.9, the convergences of $\hat{\xi}_\ell^A$ and $\hat{\xi}_\ell^B$ can be shown in a similar way as in Jun, Pinkse, and Wan (2009), Appendix F. Then the convergence of $\hat{\xi}_\ell$ follows from continuous mapping theorem.

Note first by Skorokhod representation theorem (see theorem 25.6, Billingsley, 1995), there exists a copy $(\Delta^*, \hat{\Delta}^*)$ of $(\Delta, \hat{\Delta})$ such that they have the same properties and for each ω in the sample space, $\hat{\Delta}^*(\omega) \to \Delta^*(\omega)$. Then by dominated convergence theorem (see theorem 16.4, Billingsley, 1995) and the fact that $\exp(-\|\cdot\|_+^2)$ is integrable (since $Q_1(\theta_{1\ell}, \theta_2) < 0$ for all $\theta_2 \in [\theta_{2\ell}, \theta_{2u}]$), the convergence result follows.

C.4. **Proof of theorem 4.3. Case 1.** $\Theta_{I2}(\theta_{1\ell})$ **is a singleton.**
Let $\theta_{\alpha\ell}^{\mathrm{I}}$ be the left end point of $\Theta_{\alpha n}^{\mathrm{I}}$. Then

$$\begin{aligned}
\mathbb{P}\left(\theta_{1n} \in \Theta_{\alpha n}^{\mathrm{I}}\right) &\leq \mathbb{P}\left(\theta_{1n} \leq \theta_{\alpha\ell}^{\mathrm{I}}\right) = \mathbb{P}\left(\int_{\underline{\theta}_1}^{\theta_{1n}} f_{1n}(\theta_1) d\theta_1 \geq \frac{c_\ell^{\mathrm{I}}}{\alpha_{\ell n} D_n}\right) \\
&= \mathbb{P}\left(\xi_\ell - \int_{-h_1^*}^{0} \int \exp(-\|\Delta(\theta_\ell) + \kappa(\theta_\ell) + Q(\theta_\ell) h\|_+^2) dh \geq c_\ell^{\mathrm{I}}, \int_{\theta_{1n}}^{\bar{\theta}_1} f_{1n}(\theta_1) d\theta_1 \geq \frac{c_u^{\mathrm{I}}}{\alpha_{\ell n} D_n}\right) \\
&= \mathbb{P}\left(\xi_\ell - \int_{-h_1^*}^{0} \int \exp(-\|\Delta(\theta_\ell) + \kappa(\theta_\ell) + Q(\theta_\ell) h\|_+^2) dh \geq c_\ell^{\mathrm{I}}\right) + o_p(1).
\end{aligned}$$

Hence as $h_1^* \to \infty$,

$$f_h(h_1^*) = \lim_{n\to\infty} \mathbb{P}\left(\theta_{1n} \notin \Theta_{\alpha n}^{\mathrm{I}}\right)$$
$$\geq \lim_{n\to\infty} \mathbb{P}\left(\int_{-\infty}^{-h_1^*}\int \exp(-\|\Delta(\theta_\ell)+\kappa(\theta_\ell)+Q(\theta_\ell)h\|_+^2)dh > c_\ell^{\mathrm{I}}\right) \to 1.$$

Case 2. $\Theta_{I2}(\theta_{1\ell})$ **is an interval.** Let $\theta_{\alpha\ell}^{\mathrm{I}}$ be the left end point of $\Theta_{\alpha n}^{\mathrm{I}}$. Then

$$\mathbb{P}\left(\theta_{1n} \in \Theta_{\alpha n}^{\mathrm{I}}\right) \leq \mathbb{P}\left(\theta_{1n} \geq \theta_{\alpha\ell}^{\mathrm{I}}\right) = \mathbb{P}\left(\int_{\underline{\theta}_1}^{\theta_{1n}} f_{1n}(\theta_1)d\theta_1 \geq \frac{c_\ell^{\mathrm{I}}}{\alpha_{\ell n}D_n}\right)$$
$$= \mathbb{P}\left(\xi_\ell - \int_{-h_1^*}^{0}\int_{\theta_2\in[\theta_{2\ell},\theta_{2u}]} \exp(-\|\Delta(\theta_{1\ell},\theta_2)+\kappa(\theta_{1\ell},\theta_2)+Q_1(\theta_{1\ell},\theta_2)h_1\|_+^2)d\theta_2 dh_1 \geq c_\ell^{\mathrm{I}}\right) + o_p(1).$$

It follows that as $h_1^* \to \infty$,

$$f_h(h_1^*) = \lim_{n\to\infty} \mathbb{P}\left(\theta_{1n} \notin \Theta_{\alpha n}^{\mathrm{I}}\right)$$
$$\geq \lim_{n\to\infty} \mathbb{P}\left(-\int_{-\infty}^{-h_1^*}\int_{\theta_2\in[\theta_{2\ell},\theta_{2u}]} \exp(-\|\Delta(\theta_{1\ell},\theta_2)+\kappa(\theta_{1\ell},\theta_2)+Q_1(\theta_{1\ell},\theta_2)h_1\|_+^2)d\theta_2 dh_1 > c_\ell^{\mathrm{I}}\right)$$
$$\to 1.$$

Note that the above argument does not depend on the length of the interval $[\theta_{1\ell}, \theta_{1u}]$.

C.5. **Proof of theorem 4.4.** lemmas C.10 and C.11 show that the tuple $(m^g, \Theta^g, \Theta_I^g)$ satisfies assumptions 3.1 to 3.3, 3.5, 3.7, 3.8 and 4.1 provided the tuple (m, Θ, Θ_I) satisfy corresponding assumptions. Then I can just apply theorems 4.1 and 4.2. Given assumption 4.7, without loss of generality, I assume that g is strictly increasing in its first argument.

Lemma C.10. *Suppose that assumptions 3.1, 3.2, 4.1, 4.2 and 4.6 to 4.8 are satisfied, then Θ_I^g satisfies assumptions 3.1, 3.2, 4.1 and 4.2.*

Proof. Remember that Θ^g and Θ_I^g are defined as,

$$\Theta^g = \{(\theta_1^g, \theta_2^g) : \theta \in \Theta, \theta_2^g = \theta_2, \theta_1^g = g(\theta)\}, \quad \Theta_I^g = \{(\theta_1^g, \theta_2^g) : \theta \in \Theta_I, \theta_2^g = \theta_2, \theta_1^g = g(\theta)\}.$$

Assumptions 3.1 and 3.2 holds straightforwardly by assumption 4.6. Assumption 4.1 is satisfied because g is weakly monotone in θ_2. Assumption 4.2 because g is strictly monotone in its first argument.

□

Lemma C.11. $\mathbb{E}m^g_{(j)}(\theta^g) \leq 0 \quad j = 1, 2, \cdots, J$ *if and only if* $\theta^g \in \Theta^g_I$. *Meanwhile, if assumptions 3.3 to 3.5, 3.7 and 3.8 are satisfied, then* $(m^g, \Theta^g_I, \Theta^g)$ *also satisfies assumptions 3.3 to 3.5, 3.7 and 3.8.*

Proof. Let $\theta^g \in \Theta^g_I$ be arbitrary. Then there exists a $\theta_1 \in \Theta_{I1}$ such that $\theta_1 = g^{-1}(\theta^g_1; \theta^g_2)$. It is also known that $\theta^g_2 \in \Theta_{I2}$, then it follows that $(g^{-1}(\theta^g_1; \theta^g_2), \theta^g_2) \in \Theta_I$, hence $\mathbb{E}m_{(j)}(g^{-1}(\theta^g_1; \theta^g_2), \theta^g_2) \leq 0$.

To see the inverse, suppose that $\mathbb{E}m^g_{(j)}(\theta^g) \leq 0$ for some θ^g. If $\theta^g \notin \Theta^g_{I1}$, then either $\theta^g_2 \notin \Theta_{I2}$ or $\theta^g_1 \notin \Theta^g_{I1}$, or both. If $\theta^g_2 \notin \Theta_{I2}$, It is a contradiction already. If $\theta^g_1 \notin \Theta^g_{I1}$, then $g^{-1}(\theta^g_1; \theta^g_2) \notin \Theta_{I1}$, contradiction again. So I can conclude that $\theta^g \in \Theta^g_I$.

Assumptions 3.3, 3.5 and 3.7 holds for $(m^g, \Theta^g, \Theta^g_I)$ because of the strict monotonicity of g in θ_1 and the boundedness of the first derivative over Θ. I show that assumption 3.8 holds too. Without loss of generality, let g is strictly increasing in its first argument and weakly increasing in θ_2.

Let $\theta^g_{1\ell}$ is lower bound of Θ^g_{I1}. Let $\Theta^g_{I2}(\theta^g_{1\ell}) = \{\theta^{g^*}_2\}$. Let $\mathcal{J}_g$ be the set of moment equations which are binding at $(\theta^g_{1\ell}, \theta^{g*}_2)$, i.e., $\mathbb{E}m^g_{(j)}(\theta^g_{1\ell}, \theta^{g*}_2) = 0$ for all $j \in \mathcal{J}_g$. But $\mathbb{E}m_{(j)}(g^{-1}(\theta^g_{1\ell}, \theta^{g^*}_2), \theta^{g*}_2) = \mathbb{E}m^g_{(j)}(\theta^g_{1\ell}, \theta^{g*}_2)$, so the j–th moment $m_{(j)}$ is also binding at $(g^{-1}(\theta^g_{1\ell}, \theta^{g^*}_2), \theta^{g*}_2)$. It follows that,

$$Q^{g\mathcal{J}_g}(\theta^g_{1\ell}, \theta^{g*}_2) = \frac{\partial \mathbb{E}m^g}{\partial \theta^g}\Big|_{\theta^g = (\theta^g_{1\ell}, \theta^{g*}_2)} = Q^{\mathcal{J}}(g^{-1}(\theta^g_{1\ell}; \theta^{g*}_2), \theta^{g*}_2) \times \mathcal{G}$$

where $Q^{\mathcal{J}}(g^{-1}(\theta^g_{1\ell}; \theta^{g*}_2), \theta^{g*}_2)$ has maximum rank by assumption 4.8 and $\mathcal{G}$ is a d by d matrix takes following form

$$\begin{pmatrix} \frac{\partial g^{-1}}{\partial \theta_1} & \frac{\partial g^{-1}}{\partial \theta_2} & \cdots & \frac{\partial g^{-1}}{\partial \theta_d} \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Note that $\frac{\partial g^{-1}}{\partial \theta_1} \neq 0$. It is easy to verify that $(Q^{g\mathcal{J}})'(Q^{g\mathcal{J}}) = \mathcal{G}' Q^{\mathcal{J}'} Q^{\mathcal{J}} \mathcal{G}$ is positive definite. □

C.6. **Proof of theorem 4.5.** When $\theta_{1\ell} < \theta_{1u}$, the conclusion follows immediately. Now consider the case $\theta_{1\ell} = \theta_{1u}$. In this case, since $\xi_\ell = \xi_u = \xi$ and $c^{\mathrm{II}}_\ell = c^{\mathrm{II}}_u = c^{\mathrm{II}}$. Let θ^α_u and θ^α_u be the two end points of the confidence set.

$$\begin{aligned} \mathbb{P}([\theta_{1\ell}, \theta_{1u}] \subseteq \Theta^{\mathrm{II}}_{\alpha n}) &= \mathbb{P}(\theta^\alpha_\ell \leq \theta_{1\ell} \leq \theta^\alpha_u) \\ &= \mathbb{P}\Big(\int_{\underline{\theta}_1}^{\theta_{1\ell}} f_{1n}(\theta_1) d\theta_1 \leq 1 - \frac{c^{\mathrm{II}}}{\alpha_n D_n}, \int_{\underline{\theta}_1}^{\theta_{1\ell}} f_{1n}(\theta_1) d\theta_1 \geq \frac{c^{\mathrm{II}}}{\alpha_n D_n}\Big) \\ &= \mathbb{P}\Big(\int_{\theta_{1\ell}}^{\bar{\theta}_1} f_{1n}(\theta_1) d\theta_1 \geq \frac{c^{\mathrm{II}}}{\alpha_n D_n}, \int_{\underline{\theta}_1}^{\theta_{1\ell}} f_{1n}(\theta_1) d\theta_1 \geq \frac{c^{\mathrm{II}}}{\alpha_n D_n}\Big) \\ &= \mathbb{P}(\xi \geq c^{\mathrm{II}}) = 1 - \alpha. \end{aligned}$$

C.7. **Proof of lemma 4.2.**

Lemma C.12. *Let* $0 < M < +\infty$. *Let* $\mathscr{A}$ *be the event that*

$$\min_{h_2 \in [-M,M]} \{\|\Delta_X + \Delta_Y + h_2\|_+^2 + \|\Delta_X - \Delta_Z - h_2\|_+^2\} < \|\Delta_X + \Delta_Y\|_+^2 + \|\Delta_X - \Delta_Z\|_+^2,$$

then $\mathbb{P}(\mathscr{A}) > 0$.

Proof. Consider the even $\mathscr{B} = \{0 < \Delta_X - \Delta_Z < -\Delta_X - \Delta_Y < M\}$. Since Δ_W are all independent mean zero normal random variables, it is straightforward to see that $\mathbb{P}(\mathscr{B}) > 0$. Given $\mathscr{B}$, for any $h_2 \in (\Delta_X - \Delta_Z, -\Delta_X - \Delta_Y)$,

$$\|\Delta_X + \Delta_Y + h_2\|_+^2 + \|\Delta_X - \Delta_Z - h_2\|_+^2 = 0 < \|\Delta_X + \Delta_Y\|_+^2 + \|\Delta_X - \Delta_Z\|_+^2.$$

So the conclusion follows directly. □

Lemma C.13. *Let* $0 < M < +\infty$. *Then* $\sup_{h_2 \in [-M,M]} |\hat{c}^{Pro}(1,1) - \hat{c}^{Pro}(1, 1 + \frac{h_2}{\sqrt{n}})| = o_p(1)$.

Proof. Note that $\hat{c}^{Pro}(1,1)$ is the $1-\alpha$ quantile of

$$\|\Delta_{X,n} + \Delta_{Y,n} + \hat{\kappa}_1(1,1)\|_+^2 + \|\Delta_{X,n} - \Delta_{Z,n} + \hat{\kappa}_2(1,1)\|_+^2,$$

and $\hat{c}^{Pro}(1, 1 + h_2/\sqrt{n})$ is the $1-\alpha$ quantile of

$$\|\Delta_{X,n} + \Delta_{Y,n}(1 + h_2/\sqrt{n}) + \hat{\kappa}_1(1, 1 + h_2/\sqrt{n})\|_+^2 + \|\Delta_{X,n} - \Delta_{Z,n} + \hat{\kappa}_2(1, 1 + h_2/\sqrt{n})\|_+^2.$$

For $\hat{\kappa}$, I use the $\varphi_j^{(1)}$ function in Andrews and Soares (2010). Note that $\frac{1}{n}\sum_{i=1}^n m_{1i}(W_i, (1,1)) = \bar{X} + \bar{Y} > -\sqrt{\log n / n}$ with probability approaching 1, so $\hat{\kappa}_1(1,1) \xrightarrow{p} 0$. The same conclusion holds for $\hat{\kappa}_2(1,1)$. Likewise, $\frac{1}{n}\sum_{i=1}^n m_{1i}(W_i, (1, 1 + h_2/\sqrt{n})) = \bar{X} + \bar{Y}(1 + h_2/\sqrt{n}) = \hat{\kappa}_1(1,1) + \bar{Y}h_2/\sqrt{n} > -\sqrt{\log n / n}$ with probability approaching 1 since $|h_2| \le M$. So $\hat{\kappa}_1(1, 1 + h_2/\sqrt{n}) \xrightarrow{p} 0$.

Then conclusion of the lemma follows because $\sup_{h \in [-M,M]} \Delta_{W,n} \frac{h_2}{\sqrt{n}} = o_p(1)$ and the continuity of both random variables at their $1-\alpha$ quantile. □

Now I show that the conclusion of the lemma 4.2 holds. For large n and some large constant $M > 0$,

$$\begin{aligned}
\mathbb{P}(\theta_{1\alpha}^{Pro,\mathrm{I}} \le 1) &= \mathbb{P}(\exists \theta_2 : (1, \theta_2) \in \Theta_\alpha^{Pro,\mathrm{I}}) \\
&= 1 - \mathbb{P}(\forall \theta_2 : (1, \theta_2) \notin \Theta_\alpha^{Pro,\mathrm{I}}) = 1 - \mathbb{P}(\forall \theta_2 : nL_n(1, \theta_2) > \hat{c}^{Pro}(1, \theta_2)) \\
&= 1 - \mathbb{P}(\forall \theta_2 : \|\Delta_{X,n} + \Delta_{Y,n}\theta_2 + \sqrt{n}(\theta_2 - 1)\|_+^2 + \|\Delta_{X,n} - \Delta_{Z,n}\theta_2 - \sqrt{n}(\theta_2 - 1)\|_+^2 > \hat{c}^{Pro}(1, \theta_2))
\end{aligned}$$

$$\geq 1-\mathbb{P}(\forall h_2 \in [-M,M] : \|\Delta_{X,n}+\Delta_{Y,n}(1+\frac{h_2}{\sqrt{n}})+h_2\|_+^2+\|\Delta_{X,n}-\Delta_{Z,n}(1+\frac{h_2}{\sqrt{n}})-h_2\|_+^2 > \hat{c}^{Pro}(1,1+\frac{h_2}{\sqrt{n}}))$$
$$= 1-\mathbb{P}(\forall h_2 \in [-M,M] : \|\Delta_X+\Delta_Y+h_2\|_+^2+\|\Delta_X-\Delta_Z-h_2\|_+^2 > c^{Pro}(1,1)+o_p(1))$$
$$= 1-\mathbb{P}(\min_{h_2\in[-M,M]}\{\|\Delta_X+\Delta_Y+h_2\|_+^2+\|\Delta_X-\Delta_Z-h_2\|_+^2\} > c^{Pro}(1,1))+o(1).$$

The second last equality comes from the fact that $\Delta_{W,n} \xrightarrow{d} \Delta_W$ and $\hat{c}^{Pro}(1,1+\frac{h_2}{\sqrt{n}}) \xrightarrow{p} \hat{c}^{Pro}(1,1)$ uniformly over $[-M,M]$.

Note that $c^{Pro}(1,1)$ is the $(1-\alpha)$ quantile of $\|\Delta_X+\Delta_Y\|_+^2+\|\Delta_X-\Delta_Z\|_+^2$, and because $\min_{h_2\in[-M,M]}\{\|\Delta_X+\Delta_Y+h_2\|_+^2+\|\Delta_X-\Delta_Z-h_2\|_+^2\}$ is continuous random variable,

$$\mathbb{P}(\min_{h_2\in[-M,M]}\{\|\Delta_X+\Delta_Y+h_2\|_+^2+\|\Delta_X-\Delta_Z-h_2\|_+^2\} > c^{Pro}(1,1)|\mathscr{A})$$
$$< \mathbb{P}(\|\Delta_X+\Delta_Y\|_+^2+\|\Delta_X-\Delta_Z\|_+^2 > c^{Pro}(1,1)|\mathscr{A}) = \alpha.$$

It follows,

$$\mathbb{P}(\theta_{1\alpha}^{Pro,\mathrm{I}} \leq 1)+o_p(1) \geq \mathbb{P}(\min_{h_2\in[-M,M]}\{\|\Delta_X+\Delta_Y+h_2\|_+^2+\|\Delta_X-\Delta_Z-h_2\|_+^2\} > c^{Pro}(1,1))$$
$$= \mathbb{P}(\min_{h_2\in[-M,M]}\{\|\Delta_X+\Delta_Y+h_2\|_+^2+\|\Delta_X-\Delta_Z-h_2\|_+^2\} > c^{Pro}(1,1)|\mathscr{A})\mathbb{P}(\mathscr{A})$$
$$+\mathbb{P}(\min_{h_2\in[-M,M]}\{\|\Delta_X+\Delta_Y+h_2\|_+^2+\|\Delta_X-\Delta_Z-h_2\|_+^2\} > c^{Pro}(1,1)|\mathscr{A}^C)\mathbb{P}(\mathscr{A}^C)$$
$$\geq \mathbb{P}(\min_{h_2\in[-M,M]}\{\|\Delta_X+\Delta_Y+h_2\|_+^2+\|\Delta_X-\Delta_Z-h_2\|_+^2\} > c^{Pro}(1,1)|\mathscr{A})\mathbb{P}(\mathscr{A})$$
$$+\alpha\mathbb{P}(\mathscr{A}^C) > \alpha\mathbb{P}(\mathscr{A})+\alpha\mathbb{P}(\mathscr{A}^C) = 1-\alpha.$$

So I can conclude that $\mathbb{P}(\theta_{1\alpha}^{Pro,\mathrm{I}} \leq 1) > 1-\alpha+o(1)$.

C.8. **Proof of lemma 4.3.** It can be shown that,

$$\Theta_{1\alpha}^{Pro,\mathrm{II}} = \left[\frac{-2\bar{Y}}{\bar{X}(\bar{Y}+\bar{Z})}+q(\alpha,n),2\right].$$

where $q(\alpha,n)=\sqrt{\frac{c_\alpha^{Pro}}{n}}(-1+o_p(1))$. Let $\theta_{1\alpha}^{Pro,\mathrm{II}}=\frac{-2\bar{Y}}{\bar{X}(\bar{Y}+\bar{Z})}+q(\alpha,n)$. I want to show that $\mathbb{P}(\theta_{1\alpha}^{Pro,\mathrm{II}} \leq 1) > 1-\alpha$.

$$\mathbb{P}([1,2]\subseteq[\theta_{1\alpha}^{Pro,\mathrm{II}},2]) = \mathbb{P}(\theta_\alpha^{Pro}\leq 1) = \mathbb{P}(\frac{-2\bar{Y}}{\bar{X}(\bar{Y}+\bar{Z})}-1 \leq \sqrt{\frac{c_\alpha^{Pro}}{n}}+o_p(\frac{1}{n}))$$
$$= \mathbb{P}(\frac{1}{2}\|2\Delta_{n,X}+\Delta_{n,Y}-\Delta_{n,Z}\|_+^2 \leq c_\alpha^{Pro}+o(1)) = \mathbb{P}(\frac{1}{2}\|2\Delta_X+\Delta_Y-\Delta_Z\|_+^2 \leq c_\alpha^{Pro})+o(1)$$

$$\geq \mathbb{P}(\|\Delta_X + \Delta_Y\|_+^2 + \|\Delta_X - \Delta_Z\|_+^2 \leq c_\alpha^{Pro}) + o(1)$$
$$> \mathbb{P}(\mathcal{C} \leq c_\alpha^{Pro}) + o(1) = 1 - \alpha + o(1).$$

The strict inequality holds because $\|\Delta_X + \Delta_Y\|_+^2 + \|\Delta_X - \Delta_Z\|_+^2$ are continuously distributed on $(0, +\infty)$ and $\mathbb{P}(\mathcal{C} > \|\Delta_X + \Delta_Y\|_+^2 + \|\Delta_X - \Delta_Z\|_+^2) > 0$, as shown in the following Lemma.

Lemma C.14. *In example D,* $\mathbb{P}(\mathcal{C} \geq \|\Delta_X + \Delta_Y\|_+^2 + \|\Delta_X - \Delta_Z\|_+^2) = 1$ *and the strict inequality holds with positive probability.*

Proof. Since when θ belongs to the interior of Θ_I, $\kappa(\theta) = -\infty$ and $\mathcal{C} = 0$, the sup is necessarily achieved on the boundary of Θ_I: $\partial\Theta_I$. $\mathbb{P}(\mathcal{C} \geq \|\Delta_X + \Delta_Y\|_+^2 + \|\Delta_X - \Delta_Z\|_+^2) = 1$ because $(1,1)$ belongs to the boundary.

In this example, $\mathcal{C}$ takes the form of

$$\mathcal{C} = \max\{\sup_{\theta_1 \in [1,2]} \theta_1 \|\Delta_X + \Delta_Y\|_+^2, \sup_{\theta_1 \in [1,2]} \|(\Delta_X + \Delta_Z)\theta_1 - 2\Delta_Z\|_+^2\}.$$

Let A be the event such that $\Delta_X + \Delta_Y > 0$, $\Delta_X - \Delta_Z < 0$ and $\Delta_X + \Delta_Z > 0$, then under event A,

$$\mathcal{C} = \max\{2(\Delta_X + \Delta_Y)^2, 4\Delta_X^2\} \geq 2(\Delta_X + \Delta_Y)^2 > \|\Delta_X + \Delta_Y\|_+^2 + \|\Delta_X - \Delta_Z\|_+^2.$$

Event A occurs with probability strictly greater than 0 under this assumption that X, Y and Z are independent. So I can conclude that

$$\mathbb{P}(\mathcal{C} = \|\Delta_X + \Delta_Y\|_+^2 + \|\Delta_X - \Delta_Z\|_+^2) \leq 1 - \mathbb{P}(A) < 1.$$

$\square$

Appendix D. Algorithm

D.1. **Algorithm for theorem 4.2.** Now I summarize the procedure of computing the confidence set.

Algorithm 1. Construct confidence set by simulation. Let S, R and B be a some positive integers.

(1) Estimation.
   (a) Choose one initial value $\theta_{(0)} \in \Theta$. One can choose $\theta_{(0)}$ such that $\bar{m}_{(j)}(\theta_{(0)}) = 0$ for some j.
   (b) Construct an MCMC chain $\{\theta_{(b)}\}_{b=0}^B$ based on f_n using Metropolis–Hastings Algorithm (See Robert and Casella, 2004, Chapter 7). Obtain a stationary chain $\{\theta_{(b)}\}_{b=B'}^B$.

(c) Obtain the first component $\theta_{1(b)}$ of $\theta_{(b)}$ for all $B' \leq b \leq B$. Then $\{\theta_{1(b)}\}_{b=B'}^{B}$ are random draws from the marginal quasi–posterior f_{1n}.

(d) Sort $\{\theta_{1(b)}\}_{b=B'}^{B}$ and compute $\hat{\theta}_{1\ell}$ and $\hat{\theta}_{1u}$ using eq. (4).

(2) Draw another stationary MCMC chain $\{\theta_{2(b)}\}$ for θ_2 from $f_n(\hat{\theta}_{1\ell}, \theta_2)$ using Metropolis–Hastings Algorithm.

(3) Take S random draws $\{\theta_{2,s}\}_{s=1}^{S}$ from the chain $\{\theta_{2(b)}\}$ with replacement.

(4) Replace $M_n(\theta_{1\ell})$ by

$$\widehat{M}_\ell = \frac{\sqrt{\beta_n}}{S} \sum_{s=1}^{S} \frac{\exp(-\beta_n \|\bar{m}(\hat{\theta}_{1\ell}, \theta_{2,s})\|_+^2)}{f_n(\hat{\theta}_{1\ell}, \theta_{2,s})}.$$

(5) Compute $\hat{\alpha}_\ell$ and $\hat{\omega}_\ell$ using $\widehat{M}_\ell$ in place of $M_n(\theta_{1\ell})$.

(6) Compute $\hat{\theta}_2 = (1/S) \sum_{s=1}^{S} \theta_{2,s}$.

(7) For each $(h_{1,s}, \theta_{2,s})$, compute $\hat{\gamma}(\hat{\theta}_{1\ell}, \theta_{2,s})$.

(8) Independently draw $\{(h_{1,s}, h_{2,s})\}_{s=1}^{S}$ from bivariate normal distribution with identity covariance matrix.[12]

(9) For each $(h_{1,s}, \theta_{2,s})$, estimate $\hat{Q}_1(\hat{\theta}_{1\ell}, \theta_{2,s}) = \frac{\partial \bar{m}(\theta)}{\partial \theta_1}|_{\theta_1 = \hat{\theta}_{1\ell}, \theta_2 = \theta_{2,s}}$.

(10) Let $\hat{H}$ be a JS by JS matrix whose $[(s-1) \times J + j]$ by $[(s'-1) \times J + j']$–th element equals

$$\frac{1}{n} \sum_{i=1}^{n} \left[m_{(j)}(W_i, (\hat{\theta}_{1\ell}, \theta_{2,s})) - \bar{m}_{(j)}(\hat{\theta}_{1\ell}, \theta_{2,s}) \right] \times \left[m_{(j')}(W_i, (\hat{\theta}_{1\ell}, \theta_{2,s'})) - \bar{m}_{(j')}(\hat{\theta}_{1\ell}, \theta_{2,s'}) \right].$$

(11) For $r = 1, \cdots, R$, independently across r,

(a) Draw JS by 1 random vector $\mathbf{z}_r$, each element is drew independently from standard normal distribution.

(b) Let $\tilde{\mathbf{z}}_r = \hat{H}^{1/2} \mathbf{z}_r$. Let $\Delta_{j,s,r}$ be the $((s-1) \times J + j)$–th element of $\tilde{\mathbf{z}}_r$.

(c) Compute

$$\hat{\xi}_{\ell,r}^{B} = \frac{1}{S} \sum_{s=1}^{S} \frac{\exp(-\sum_{j=1}^{J} \hat{\gamma}_j(\hat{\theta}_{1\ell}, \theta_{2,s}) |\Delta_{j,s,r} + \hat{Q}_{1j} h_{1,s}|_+^2) \mathbf{1}(h_{1,s} < 0)}{\phi(h_{1,s}) f_n(\hat{\theta}_{1\ell}, \theta_{2,s})}.$$

(d) Draw a J–vector mean zero normal random variable $\mathbf{w}_r$ with (j, j') element in the variance matrix equals to

$$\frac{1}{n} \sum_{i=1}^{n} \left[m_{(j)}(W_i; \hat{\theta}_{1\ell}, \hat{\theta}_2) - \bar{m}_{(j)}(\hat{\theta}_{1\ell}, \hat{\theta}_2) \right] \times \left[m_{(j')}(W_i; \hat{\theta}_{1\ell}, \hat{\theta}_2) - \bar{m}_{(j')}(\hat{\theta}_{1\ell}, \hat{\theta}_2) \right].$$

[12] One can choose covariance matrix be $2(\hat{Q}'\hat{Q})^{-1}$ to improve the performance.

(e) Compute

$$\hat{\xi}^A_{\ell,r} = \frac{1}{S}\sum_{s=1}^{S} \frac{\exp(-\sum_{j=1}^{J}\hat{\gamma}_j(\hat{\theta}_{1\ell},\hat{\theta}_2)|w_{j,r}+\hat{Q}'(\hat{\theta}_{1\ell},\hat{\theta}_2)h_s|_+^2)\mathbf{1}(h_{1,s}<0)}{\phi(h_{1,s})\phi(h_{2,s})}.$$

(f) Compute $\hat{\xi}_{\ell,r} = \hat{\omega}\psi_n(\hat{\xi}^A_{\ell,r}) + (1-\hat{\omega})\hat{\xi}^B_{\ell,r}$.

(12) Obtain the simulated distribution for $\hat{\xi}_u$ in a similar way (step 4 to step 11).

(13) Let $\hat{c}^{\mathrm{I}}_\ell$ and $\hat{c}^{\mathrm{I}}_u$ be computed using the maximization problem in eq. (13). If there are more than one solutions, pick an arbitrary one.

(14) Construct the confidence interval $\widehat{\Theta}^{\mathrm{I}}_\alpha = [F^{-1}_{1n}(\hat{c}^{\mathrm{I}}_\ell/(\hat{\alpha}_\ell D_n)), F^{-1}_{1n}(1-\hat{c}^{\mathrm{I}}_u/(\hat{\alpha}_u D_n))]$.

Appendix E. Proofs in section 5

E.1. **Sketch of proof of theorem 5.1.** For notation simplicity, I ignore the sup–script $\mathcal{J}$.

What I need to show is that $\lim_{n\to\infty}\mathbb{P}(\sqrt{n}(\hat{\theta}_{1\ell}-\theta_{1\ell})<K) = \mathbb{P}(\mathcal{U}_1<K)$ for any K.

$$\begin{aligned}
\mathbb{P}(\sqrt{n}(\hat{\theta}_{1\ell}-\theta_{1\ell})<K) &= \mathbb{P}\left(\hat{c} < n\int_{\underline{\theta}}^{\theta_{1\ell}+K/\sqrt{n}}\int \exp(-\|\sqrt{n}\bar{m}(\theta)\|^2_{\rho_n})d\theta\right)\\
&= \mathbb{P}\left(\hat{c} < \int_{-\infty}^{K}\int \exp(-\|\Delta_n(\theta_\ell+h/\sqrt{n})+Q_\ell h+o(\|h\|)\|^2_{\rho_n})dh\right)\\
&= \mathbb{P}\left(\hat{c} < \int_{-\infty}^{\mathcal{U}_1}\int \exp(-\|\Delta_n(\theta_\ell+h/\sqrt{n})+Q_\ell h+o(\|h\|)\|^2_{\rho_n})dh\right.\\
&\qquad\left. + \int_{\mathcal{U}_1}^{K}\int \exp(-\|\Delta_n(\theta_\ell+h/\sqrt{n})+Q_\ell h+o(\|h\|)\|^2_{\rho_n})dh\right)
\end{aligned}$$

Now I show that $\hat{c}-\int_{-\infty}^{\mathcal{U}_1}\int\exp(-\|\Delta_n(\theta_\ell+h/\sqrt{n})+Q_\ell h+o(\|h\|)\|^2_{\rho_n})dh \xrightarrow{p} 0$, then the conclusion follows. Let $\mu = (Q'Q)^{-1}Q'\Delta$.

$$\begin{aligned}
&\int_{-\infty}^{\mathcal{U}_1}\int \exp(-\|\Delta_n(\theta_\ell+h/\sqrt{n})+Q_\ell h+o(\|h\|)\|^2_{\rho_n})dh\\
&\quad= \int_{-\infty}^{\mathcal{U}_1}\int \exp(-\|\Delta_\ell+Q_\ell h\|^2_{\rho_n})dh+o_p(1) = \int_{-\infty}^{\mathcal{U}_1}\int \exp(-\|\Delta_\ell+Q_\ell h\|^2)dh+o_p(1)\\
&= \int_{-\infty}^{\mathcal{U}_1}\int \exp(-\frac{1}{2}(h+\mu)'(2Q'_\ell Q_\ell)(h+\mu))dh+o_p(1) = \int_{-\infty}^{\mathcal{U}_1+\mu_1}\int \exp(-\frac{1}{2}\hat{h}'(2Q'_\ell Q_\ell)\hat{h})d\hat{h}+o_p(1)\\
&\quad= \int_{-\infty}^{0}\int \exp(-\frac{1}{2}\hat{h}'(2Q'_\ell Q_\ell)\hat{h})d\hat{h}+o_p(1) = \frac{1}{2}\sqrt{\pi|(Q'_\ell Q_\ell)^{-1}|}+o_p(1)\xrightarrow{p}\pi\sqrt{|(2Q'_\ell Q_\ell)^{-1}|}.
\end{aligned}$$

the integrability of the left hand side can is ensured by lemma A.3. By assumption 5.4, I know that $\hat{c}-\int_{-\infty}^{\mathcal{U}_1}\int\exp(-\|\Delta_n(\theta_\ell+h/\sqrt{n})+Q_\ell h+o(\|h\|)\|^2_{\rho_n})dh \xrightarrow{p} 0$. Then it follows that

$$\mathbb{P}(\sqrt{n}(\hat{\theta}_{1\ell}-\theta_{1\ell})<K)=\mathbb{P}\left(\int_{\mathcal{U}_1}^{K}\int \exp(-\|\Delta_n(\theta_\ell+h/\sqrt{n})+Q_\ell h+o(\|h\|)\|_{\rho_n}^2)dh\geq o_p(1)\right)$$
$$=\mathbb{P}\left(\mathcal{U}_1\leq K+o_p(1)\right)\xrightarrow{p}\mathbb{P}\left(\mathcal{U}_1\leq K\right).$$

APPENDIX F. GENERIC DIMENSION

In the main text, this results are stated for $d=2$. In this section, I give a general results for any $d>2$. I state Theorems only. In appendix F.1, I outline the results for $\sqrt{n}$–consistent estimation. In appendix F.2, I show a parallel result to theorem 4.1 holds, also.

F.1. **Estimation.**

For generic d, I define the quantile

$$\hat{\tau}=\min\left\{\frac{\hat{c}}{n^{d/2}D_n},\frac{1}{2}\right\},\tag{22}$$

where $\hat{c}\xrightarrow{p}c\geq 0$.

Theorem F.1 (consistency). *Suppose that the conditions required by theorem 3.1 are satisfied. Suppose in addition that $\hat{\tau}$ is defined in eq.* (22), *then $\hat{\theta}_{1\ell}\xrightarrow{p}\theta_{1\ell}$.*

Sketch of proof. By a similar argument in lemma A.1, it follows that for any $\eta>0$, there exists a $C^*>0$ such that $\lim_{n\to\infty}\mathbb{P}(n^{d/2}D_n<C^*)<\eta$. Following the proof to theorem 3.1, $\lim_{n\to\infty}\mathbb{P}(\hat{\theta}_{1\ell}<\theta_{1\ell}-\epsilon)\to 0$ for any $\epsilon>0$.

It remains to show that $\lim_{n\to\infty}\mathbb{P}(\hat{\theta}_{1\ell}>\theta_{1\ell}+\epsilon)\to 0$ for any $\epsilon>0$. If $\theta_{1\ell}=\theta_{1u}$, then the proof is same as before. When $\theta_{1\ell}<\theta_{1u}$, there exists a set

$$\mathscr{B}_n=\{\theta:\theta_1\in[\theta_{1\ell},\theta_{1\ell}+\epsilon],\sqrt{n}\|\mathbb{E}m_1(\theta)\|_+\leq\delta^*\},$$

for some $\delta^*>0$. By the Lipschitz 3.4, there exists $\delta^{**}>0$ such that $\mu(\mathscr{B}_n)\geq\delta^{**}n^{-(d-1)/2}$, which implies that $D_n\int_{\theta\in\mathscr{B}_n}f_n(\theta)d\theta\geq O_p(\delta^{**}n^{-(d-1)/2})$. On the other hand, $D_n\hat{\tau}=\hat{c}/n^{d/2}\leq O_p(n^{-d/2})$. So we can conclude that $\lim_{n\to\infty}\mathbb{P}(\hat{\theta}_{1\ell}>\theta_{1\ell}+\epsilon)\to 0$. □

To derive the convergence rate, I require that $\hat{c}\xrightarrow{p}c>0$. As in the main text, I define a estimator $\hat{\theta}^*_{1\ell}$

$$\hat{\theta}^*_{1\ell}=\max\{F_{1n}^{-1}\left(\hat{\tau}_\ell(\hat{\theta}^*_{1\ell})\right),\tilde{\underline{\theta}}_1\},\quad \hat{\tau}_\ell(\hat{\theta}^*_{1\ell})=\min\left\{\frac{\hat{c}U_n(\hat{\theta}^*_{1\ell})}{\sqrt{n}D_n},\frac{1}{2}\right\},\tag{23}$$

where

$$U_n(\theta_1)=\int\exp(-n\|\bar{m}(\theta_1,\theta_2)\|_+^2)d\theta_2.$$

Let $\Theta_{I2n}(\theta_{1\ell})$ be an $\frac{\delta}{\sqrt{n}}$ expansion of $\Theta_{I2}(\theta_{1\ell})$ in Θ_2:

$$\Theta_{I2n}(\theta_{1\ell}) = \left\{\theta_2 : \delta > 0 : d(\theta_2, \Theta_{I2}(\theta_{1\ell})) \leq \frac{\delta}{\sqrt{n}}\right\}.$$

For any set $\mathcal{A} \subset \mathbb{R}^{d-1}$, let $\mu_{d-1}(\mathcal{A})$ be the Lebesgue measure of $\mathcal{A}$ in $\mathbb{R}^{d-1}$. For example, if $d = 3$ and if $\mathcal{A}$ is a single point or a line in $\mathbb{R}^2$, then $\mu_{d-1}(\mathcal{A}) = 0$; if $\mathcal{A}$ is a rectangle in $\mathbb{R}^2$, then $\mu_{d-1}(\mathcal{A}) > 0$. Write $\mu_n = \mu_{d-1}(\Theta_{I2n}(\theta_{1\ell}))$.

Lemma F.1. *There exists a $d^* \in \{0, 1, \cdots, d-1\}$ such that $\mu_n = O(1/\sqrt{n^{d^*}})$.*

Proof. Since $\Theta_{I2n}(\theta_{1\ell})$ is a $(d-1)$–dimensional bounded set, the measure of which can at most be $O(1)$; if $\Theta_{I2}(\theta_{1\ell})$ is a singleton, then $\mu_n = O(1/\sqrt{n^{d-1}})$. □

For example, when $d = 2$, $\mu_n = 2\delta/\sqrt{n}$ when $\Theta_{I2}(\theta_{1\ell})$ is singleton; $\mu_n = \theta_{2u} - \theta_{2\ell} + 2\delta/\sqrt{n}$ when $\Theta_{I2}(\theta_{1\ell})$ is an interval. It is d^* that determines the rate at which the tail mass of the marginal quasi–posterior decreases to zero.

Lemma F.2. $\sqrt{n^{d^*}}U_n(\theta_{1\ell}) = O_p(1)$. $\lim_{C\to 0}\lim_{n\to\infty}\mathbb{P}(\sqrt{n^{d^*}}U_n(\theta_{1\ell}) < C) = 0$.

Sketch of proof. Note that with probability approaching one, we have

$$\mu_n \inf_{\theta_2 \in \Theta_{I2n}(\theta_{1\ell})} \exp(-\|\bar{m}(\theta_{1\ell}, \theta_2)\|_+^2) \leq U_n(\theta_{1\ell}) \leq \mu_n \sup_{\theta_2 \in \Theta_{I2n}(\theta_{1\ell})} \exp(-\|\bar{m}(\theta_{1\ell}, \theta_2)\|_+^2).$$

I show that $\inf_{\theta_2 \in \Theta_{I2n}(\theta_{1\ell})} \exp(-\|\bar{m}(\theta_{1\ell}, \theta_2)\|_+^2) = O_p(1)$. The sup is bounded by one already. Note that for all $\theta_2 \in \Theta_{I2n}(\theta_{1\ell})$,

$$\begin{aligned}
-\|\bar{m}(\theta_{1\ell}, \theta_2)\|_+ &= -\|\Delta_n(\theta_{1\ell}, \theta_2) + \sqrt{n}\mathbb{E}m_1(\theta_{1\ell}, \theta_2)\|_+ \\
&\geq -\|\Delta_n(\theta_{1\ell}, \theta_2)\|_+ - \|\sqrt{n}\mathbb{E}m_1(\theta_{1\ell}, \theta_2)\|_+ \geq -\|\Delta_n(\theta_{1\ell}, \theta_2)\|_+ - \delta.
\end{aligned}$$

The last inequality holds because by assumption 3.4, for all $\theta_2 \in \Theta_{I2n}(\theta_{1\ell})$, $\|\sqrt{n}\mathbb{E}m_1(\theta_{1\ell}, \theta_2)\|_+ \leq \sqrt{n}d(\theta_2, \Theta_{I2}(\theta_{1\ell})) \leq \delta$. $\Delta_n(\theta_{1\ell}, \theta_2)$ is stochastic bounded by assumption 3.5. □

I focus my attention to the models in which $\Theta_{I2}(\theta_{1\ell})$ and $\Theta_{I2}(\theta_{1u})$ satisfies assumption F.1.

Assumption F.1. *The $d^* = 0$ or $d^* = d - 1$.*

Assumption F.1 is more restrictive than assumption 4.1, which I made for the models with $d = 2$. There, assumption 4.1 is made for the sake of notational simplicity; here assumption F.1 excludes the cases in which $\Theta_{I2}(\theta_{1\ell})$ is not a singleton but has zero Lebesgue measure in $\mathbb{R}^{d-1}$.

Theorem F.2 ($\sqrt{n}$–consistency)**.** *Suppose that assumptions 3.1 to 3.5, 3.7 to 3.10 and F.1 Suppose in addition that $\hat{\theta}^*_{1\ell}$ is defined in eq.* (23), *then $\sqrt{n}(\hat{\theta}^*_{1\ell} - \theta_{1\ell}) = O_p(1)$.*

Sketch of proof. As before, let

$$q_n(h_1^*) = \frac{\int_{\sqrt{n}(\underline{\theta}_1 - \theta_{1\ell})}^{h_1^*} \int_{\theta_2 \in \Theta_2} \exp(-n\|\bar{m}(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_2)\|_+^2) d\theta_2 dh_1}{\int_{\theta_2 \in \Theta_2} \exp(-n\|\bar{m}(\theta_{1\ell} + \frac{h_1^*}{\sqrt{n}}, \theta_2)\|_+^2) d\theta_2}. \tag{24}$$

I will show that for any $\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}(q_n(h_1^*) \geq \epsilon)$ can be made arbitrarily small if I let $h_1^* \to -\infty$; and for any $C > 0$, $\lim_{n\to\infty} \mathbb{P}(q_n(h_1^*) < C)$ can be made arbitrarily small if I let $h_1^* \to +\infty$.

When $d^* = d - 1$, $\Theta_{I2}(\theta_{1\ell})$ has positive measure in $\mathbb{R}^{d-1}$. Consider $h_1^* \to -\infty$ first. It can be shown that

$$\begin{aligned}
\mathbb{P}(q_n(h_1^*) \geq \epsilon) &= \mathbb{P}\left(\frac{\int_{\sqrt{n}(\underline{\theta}_1 - \theta_{1\ell})}^{h_1^*} \int_{\theta_2 \in \Theta_{I2}(\theta_{1\ell})} \exp(-\|\psi(h_1, \theta_2) + o_p(1)\|_+^2) d\theta_2 + o_p(1)}{\int_{\theta_2 \in \Theta_{I2}(\theta_{1\ell})} \exp(-\|\psi(h_1^*, \theta_2) + o_p(1)\|_+^2) d\theta_2 + o_p(1)} \geq \epsilon \right) \\
&\leq \mathbb{P}\left(\frac{\int_{\sqrt{n}(\underline{\theta}_1 - \theta_{1\ell})}^{h_1^*} \int_{\theta_2 \in \Theta_{I2}(\theta_{1\ell})} \exp(-\|\psi(h_1, \theta_2)\|^2) dh_1 + o_p(1)}{\int_{\theta_2 \in \Theta_{I2}(\theta_{1\ell})} \exp(-\|\psi(h_1^*, \theta_2)\|^2) + o_p(1)} \geq \epsilon | \inf_{\theta_2 \in \Theta_{I2}(\theta_{1\ell})} \psi(h_1^*, \theta_2) \geq 0 \right) \\
&\qquad \times \mathbb{P}(\inf_{\theta_2 \in \Theta_{I2}(\theta_{1\ell})} \psi(h_1^*, \theta_2) \geq 0) + \mathbb{P}(\inf_{\theta_2 \in \Theta_{I2}(\theta_{1\ell})} \psi(h_1^*, \theta_2) < 0) + o_p(1),
\end{aligned}$$

where $\psi(h_1, \theta_2) = \Delta^{\mathcal{J}}(\theta_{1\ell}, \theta_2) + Q_1^{\mathcal{J}}(\theta_{1\ell}, \theta_2) h_1 + \sqrt{n}\mathbb{E}m^{\mathcal{J}}(\theta_{1\ell}, \theta_2)$. Note that $\|\sqrt{n}\mathbb{E}m^{\mathcal{J}}(\theta_{1\ell}, \theta_2)\|_+$ is bounded over $\Theta_{I2}(\theta_{1\ell})$. Hence $\lim_{n\to\infty} \mathbb{P}(\inf_{\theta_2 \in \Theta_{I2}(\theta_{1\ell})} \psi(h_1^*, \theta_2) < 0)$ can be made arbitrarily small by letting $h_1^* \to -\infty$. It remains to verify that the limit of the first probability on the right hand side goes to zero as $h_1^* \to -\infty$.

$$\begin{aligned}
&\mathbb{P}\left(\frac{\int_{\sqrt{n}(\underline{\theta}_1 - \theta_{1\ell})}^{h_1^*} \int_{\theta_2 \in \Theta_{I2}(\theta_{1\ell})} \exp(-\|\psi(h_1, \theta_2)\|^2) dh_1}{\int_{\theta_2 \in \Theta_{I2}(\theta_{1\ell})} \exp(-\|\psi(h_1^*, \theta_2)\|^2)} \geq \epsilon \right) \\
&\quad = \mathbb{P}\left(\int_{\theta_2 \in \Theta_{I2}(\theta_{1\ell})} \left[\int_{-\infty}^{h_1^*} \exp(-\|\psi(h_1, \theta_2)\|^2) dh_1 - \epsilon \exp(-\|\psi(h_1^*, \theta_2)\|^2) \right] d\theta_2 \geq 0 \right) \\
&\quad \leq \mathbb{P}\left(\sup_{\theta_2 \in \Theta_{I2}(\theta_{1\ell})} \left\{ \int_{-\infty}^{h_1^*} \exp(-\|\psi(h_1, \theta_2)\|^2) dh_1 - \epsilon \exp(-\|\psi(h_1^*, \theta_2)\|^2) \right\} \geq 0 \right)
\end{aligned}$$

$$= \mathbb{P}\left(\sup_{\theta_2 \in \Theta_{I2}(\theta_{1\ell})} \frac{\int_{-\infty}^{h_1^*} \exp(-\|\psi(h_1, \theta_2)\|^2) dh_1}{\exp(-\|\psi(h_1^*, \theta_2)\|^2} > \epsilon \right)$$

The probability converges to zero as $h_1^* \to -\infty$ because $\sup_{\theta_2 \in \Theta_{I2}(\theta_{1\ell})} \|\Delta(\theta_{1\ell}, \theta_2)\|$ is bounded in probability.

Now consider $h_1^* \to +\infty$. First note that the $\sqrt{n^{d^*}}$ times the denominator in eq. (24) is $O_p(1)$, following a similar argument in lemma F.2 . It remains to show that $\sqrt{n^{d^*}}$ times the numerator in eq. (24) diverges in probability as $h_1^* \to \infty$.

$$\sqrt{n^{d^*}} \int_{\sqrt{n}(\theta_1 - \theta_{1\ell})}^{h_1^*} \int_{\theta_2} \exp(-n\|\bar{m}(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_2)\|_+^2) d\theta_2 dh_1 \geq \sqrt{n^{d^*+1}} \int_{\theta \in \mathscr{C}_n^*} \exp(-n\|\bar{m}(\theta)\|_+^2) d\theta$$
$$\geq \sqrt{n^{d^*+1}} \mu(\mathscr{C}_n^*) \inf_{\theta \in \mathscr{C}_n^*} \exp(-n\|\bar{m}(\theta)\|_+^2),$$

where

$$\mathscr{C}_n^* = \{\theta : \theta_{1\ell} \leq \theta_{1\ell} \leq \theta_1 + h_1^*/\sqrt{n}, \theta_2 \in \Theta_{I2n}(\theta_{1\ell})\}.$$

Note that there exists some C^* such that $\sqrt{n^{d^*+1}} \mu(\mathscr{C}_n^*) \geq C^* h_1^* \to \infty$ as $h_1^* \to \infty$. Note that $\inf_{\theta \in \mathscr{C}_n^*} \exp(-n\|\bar{m}(\theta)\|_+^2)$ is $O_p(1)$ by assumption 3.5, we can conclude that numerator in eq. (24) diverges in probability.

The case where $d^* = 0$, i.e., $\Theta_{I2}(\theta_{1\ell})$ is a singleton, can be shown in a similar way as in lemma A.8. □

F.2. **Inference.** In this section I discuss the inference about θ_{01}. Lemma F.3 is analog to lemma 4.1.

Lemma F.3. *Suppose that assumptions 3.1 to 3.5, 3.7 to 3.9, 4.2, 4.3 and F.1 are satisfied. If $\Theta_{I2} = \{\theta_{2\ell}\}$ is a singleton, then*

$$\sqrt{n^d} D_n F_{1n}(\theta_{1\ell}) \xrightarrow{d} \xi_\ell^A = \int_{h_1 \leq 0} \int_{h_2 \in \mathbb{R}} \exp\left(-\|\Delta^{\mathcal{J}}(\theta_\ell) + Q^{\mathcal{J}}(\theta_\ell) h\|_+^2\right) dh. \tag{25}$$

If $\mu_{d-1}\{\Theta_{I2}(\theta_{1\ell})\} > 0$, then

$$\sqrt{n} D_n F_{1n}(\theta_{1\ell}) \xrightarrow{d} \xi_\ell^B = \int_{h_1 \leq 0} \int_{\theta_2 \in \Theta_{I2}(\theta_{1\ell})} \exp\left(-\|\Delta^{\mathcal{J}}(\theta_{1\ell}, \theta_2) + Q_1^{\mathcal{J}}(\theta_{1\ell}, \theta_2) h_1\|_+^2\right) d\theta_2 dh_1. \tag{26}$$

Proof. The proof is similar to the proof of lemma 4.1. □

For inference, let $\tau_\ell^{\mathrm{III}} = c_\ell^{\mathrm{III}}/(\alpha_{\ell n} D_n)$ and $\tau_u^{\mathrm{III}} = c_u^{\mathrm{III}}/(\alpha_{un} D_n)$, where c_ℓ^{III} and c_u^{III} are solution to the following problem

$$(c_\ell^{\mathrm{III}}, c_u^{\mathrm{III}}) = \underset{(c_\ell,c_u)\in\mathbb{R}^+\times\mathbb{R}^+}{\operatorname{argmin}} \left| F_{1n}^{-1}\left(\frac{c_\ell}{\alpha_{\ell n} D_n}\right) - F_{1n}^{-1}\left(1 - \frac{c_u}{\alpha_{un} D_n}\right) \right| \tag{27}$$

$$\text{s.t.} \quad \mathbb{P}\left(c_\ell \le \tilde{\tilde{\xi}}_{\ell n}, c_u \le \sqrt{n}\nu\left(\frac{\beta_n}{n\hat{T}}\right) + \tilde{\tilde{\xi}}_{un}\right) = 1 - \alpha,$$

$$\mathbb{P}\left(c_u \le \tilde{\tilde{\xi}}_{un}, c_l \le \sqrt{n}\nu\left(\frac{\beta_n}{n\hat{T}}\right) + \tilde{\tilde{\xi}}_{\ell n}\right) = 1 - \alpha.$$

If there are more than one solutions, I take an arbitrary one. I thus construct a confidence interval for θ_{01} as $\Theta_{\alpha n}^{\mathrm{III}} = [F_{1n}^{-1}\left(\tau_\ell^{\mathrm{III}}\right), F_{1n}^{-1}\left(1 - \tau_u^{\mathrm{III}}\right)]$.

Theorem F.3 (Inference)**.** *Suppose that assumptions 3.1 to 3.5, 3.7 to 3.9, 4.1 to 4.3 and F.1 are satisfied. Then*

$$\lim_{n\to\infty} \inf_{\theta_{01}\in[\theta_{1\ell},\theta_{1u}]} \mathbb{P}(\theta_{01} \in \Theta_{\alpha n}^{\mathrm{III}}) = 1 - \alpha.$$

Proof. The proof is similar to the proof of theorem 4.1. □

$\widehat{\Theta}_\alpha^{\mathrm{III}}$ can be constructed using algorithm 1.