AN INTEGRATION-BASED APPROACH TO MOMENT INEQUALITY MODELS

YUANYUAN WAN[†‡]

DEPARTMENT OF ECONOMICS, UNIVERSITY OF TORONTO

ABSTRACT. In this paper we develop new computationally attractive estimation and inference methods for individual parameters in a class of moment inequality models. We propose root-n-consistent estimators for any given projections of the identified set, the computation of which requires little more than making random draws from a known distribution. We construct pointwise nonconservative confidence intervals for individual parameters. Our inference method does not require either projecting a high dimensional confidence set or a resampling procedure. The finite sample performance of the proposed methods is examined in Monte Carlo simulations.

JEL classification: C10, C12, C13, and C63

Keywords: Moment inequality model, Integration-based inference, Markov Chain Monte Carlo

---

Date: Saturday 31st August, 2013.

[†]yuanyuan.wan@utoronto.ca. Max Gluskin House, 150 St. George Street, Toronto, Ontario, M5S 3G7, Canada.

[‡] I am deeply grateful to Joris Pinkse and Sung Jae Jun for their constant guidance and encouragement. I appreciate the valuable suggestions from Haiqing Xu. I also benefited from the discussions with Arie Beresteanu, Herman Bierens, Federico Bugni, Ivan Canay, Paul Grieco, Patrik Guggenberger, Vadim Marmer, Marc Henry, Isabelle Perrigne, Jack Porter, Xiaoxia Shi, Youngki Shin, Quang Vuong, Neil Wallace, Hal White, Daniel Xu and seminar participants at UBC, University of Toronto, University of New South Wales, Monash University, Boston University, Ohio State University, Vanderbilt University, University of Western Ontario, University of Montreal, and Northwestern University.

1. Introduction

In empirical studies it is not uncommon that researchers are primarily interested in making inferences about one particular element of the parameter vector, while treating others as "nuisance parameters". For example, in empirical industrial organization literature, the parameter measuring the strategy interaction among firms often draws more attention than others. When the parameters are partially identified, a traditional way to construct confidence intervals for individual parameters is to first obtain a joint confidence set for the whole vector, then report its projections. However, this approach is not always the most desirable choice as such confidence intervals are likely to be conservative (see discussions in Chernozhukov, Hansen, and Jansson, 2009; Hahn and Ridder, 2011). Moreover, in partially identified models, we often have to invert hypothesis tests at every possible parameter value, which is computationally challenging in the first place.

In this paper we propose easy-to-implement estimation and inference methods for individual parameters in a class of moment inequality models. We compute the $\sqrt{n}$-consistent interval estimators by extending the computationally attractive Laplace-type estimation techniques (Chernozhukov and Hong, 2003) to partially identified models. Numerically, the estimation procedure requires no more than getting random draws from a known distribution, which can be implemented using Markov Chain Monte Carlo (MCMC) algorithms.

To obtain confidence intervals for individual parameters, we do not construct a high–dimensional confidence set to start with, instead we integrate out the rest of the parameters and make an inference directly on the parameter of interest. The proposed inference procedure is therefore easier to implement than the traditional projection method because there is no need to invert hypothesizes at every point in a high-dimensional parameter space. Furthermore, since the confidence intervals are not obtained through projection, they are not conservative.

The key quantity in our methods is the quasi-posterior

$$f_n(\theta) = \frac{\exp(nL_n(\theta))}{\int_\Theta \exp(nL_n(\theta)d\theta},$$

where Θ is a compact parameter space and $L_n(\cdot)$ is the sample analog of a population objective function $L(\cdot)$, which takes the maximum over the identified set Θ_I. f_n is well defined since it is without loss of generality to assume that $L(\theta) \leq 0$. Being integrated to 1, f_n resembles a Bayesian posterior in which the loglikelihood is replaced by the rescaled sample objective function of an extremum estimator. Chernozhukov and Hong (2003, CH2003) first proposed Laplace type estimators (LTEs) as minimizers of quasi-posterior risk functions to simplify the computation when the model is point identified, that is, when Θ_I is a singleton. When the model is partially identified, the "quasi-posterior mass" of f_n over the identified set converges in probability to 1 when the sample size increases. Using this property, we propose to estimate projections of the identified set consistently by choosing two "extreme quantiles" of corresponding marginal quasi-posteriors whose probability levels approach 0 and 1, respectively, as sample size increases. Numerically, it is as simple as taking two rank statistics of a set of random draws from the marginal quasi-posterior.

To construct confidence intervals, we show that two appropriately chosen quantiles of the marginal quasi-posterior form an asymptotically (pointwise) exact confidence interval for the corresponding parameter. We show that the rescaled "tail mass" of the marginal quasi-posterior converges in distribution to a random variable whose distribution can be easily simulated. The distributional information of the limiting random variable provides guidance to choose desirable quantiles. A nice feature is that there are only two distributions to simulate (for the left and right boundaries of the confidence interval), and for each distribution it only requires making random draws from a multivariate normal distribution with a known variance matrix and Monte Carlo integration.

There are numerous papers on inferences in moment inequality models, for example, Andrews and Guggenberger (2009); Andrews and Soares (2010); Beresteanu and Molinari (2008); Bugni (2010); Canay (2010); Chernozhukov, Hong, and Tamer (2007); Imbens and Manski (2004); Kaido (2010); Pakes, Porter, Ho, and Ishii (2006); Romano and Shaikh (2008). Many of the existing methods aim to construct joint confidence sets that have nice asymptotic properties, such as uniform validity. In general, these methods require inverting a hypothesis test, which would sometimes require a resampling procedure at each parameter value that is tested. Our main motivation is to provide empirical researchers an easy-to-implement inference procedure when the objects of interest are confidence intervals on individual parameters; hence this is a useful complement to existing work. Romano and Shaikh (2008) discussed constructing asymptotic uniformly valid confidence sets for a given parameter (and known functions of the parameter vector in general) in which the test statistics are obtained by taking minimums at each parameter value being tested, and critical values are calculated by subsampling. Our method replaces minimization by integration, and under our model setting it only requires simulating two limiting random variables. In addition, the methods developed in this paper do not require subsampling and therefore avoid the issue of choosing subsample sizes.

In the main text of the paper, we focus on the models in which the "extreme points" (defined later) of the identified set is a singleton—a restriction also imposed in the literature (e.g., Pakes, Porter, Ho, and Ishii, 2006, PPHI)[1]. Our methods differ from those of PPHI in both estimation and inference. In estimation, our estimator is implemented using MCMC algorithms. For the same argument as in CH2003, Laplace-type estimators are, in general, computationally attractive compared to extremum estimators, especially when the objective functions are not smooth. In inference, we propose a moment selection mechanism when simulating critical values. This moment selection mechanism picks out the binding moment conditions and delivers exact asymptotic coverage probability. In Section 6.1, we discuss how to construct consistent estimators without making the "singleton extreme points" assumption.

Despite that the quasi-posterior f_n resembles a Bayesian posterior, we aim to construct classical confidence intervals. Our method is therefore different from Bayesian/quasi-Bayesian analysis of moment inequality models. For example, Liao and Jiang (2010) studied large sample properties of a (quasi-) posterior derived from the limited information likelihood (see Kim, 2002) in a similar setup. The credible sets proposed by Liao and Jiang (2010), however, are not valid confidence sets from the classical perspective (too small). See Moon and Schorfheide (2012) for a general comparison between Bayesian approaches and classical approaches in partially identified models.

We illustrate our methods using a simulated example of a 2×2 discrete game of complete information with a two-dimensional parameter vector. In the experiment, our procedure takes about 6% of the time that a projection-based confidence interval would require and has coverage probabilities much closer to the desirable levels. In Section 5 we provide a detailed comparison.

The rest of the paper is organized as follows. We introduce the model and our estimator in Section 2. We discuss the asymptotic properties of the estimator in Section 3. In Section 4, we propose procedures for constructing confidence intervals. Section 5 is the Monte Carlo simulation. In Section 6, we discuss some extensions and conclude this paper.

[1]This scenario could occur in such models that when the value of a particular parameter is given, the rest of parameters are identified. For example, in a two equations Probit model with endogenous dummy regressors and without exclusion restrictions, once the correlation coefficient of the two error terms is fixed, other parameters in the model are point-identified (see Meango and Mourifie, 2013).

2. SETUP

We consider a set of J moment inequalities, $\mathbb{E}[m_i(\theta_0)] \leq 0$, where $m_i(\theta)$ is the J-vector of moment functions evaluated at W_i and θ,

$$m_i(\theta) \equiv m(W_i, \theta) = (m_{(1)}(W_i, \theta), \cdots, m_{(J)}(W_i, \theta))'.$$

$\{W_i\}_{i=1}^n$ are i.i.d. observations. The true parameter $\theta_0 \in \Theta \subset \mathbb{R}^d$. The identified set Θ_I is a collection of parameter values that satisfy the moment inequalities: $\Theta_I = \{\theta \in \Theta : \mathbb{E}[m_i(\theta)] \leq 0\}$. Θ_I is not empty as $\theta_0 \in \Theta_I$ by construction. Throughout this paper, the measurability of $m_{(j)}$ is a maintained assumption.

For any J-vector x, let $\|x\|_+^2 = \sum_j (|x_j|_+)^2$, where $|x_j|_+ = \max\{0, x_j\}$. Following Chernozhukov, Hong, and Tamer (2007, CHT2007), we consider population and sample objective functions of the following forms:

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

where $\bar{m}(\theta) = (1/n)\sum_{i=1}^n m_i(\theta)$.[2] Note that $L(\theta) = 0$ if and only if $\theta \in \Theta_I$.

We define the quasi-posterior density

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

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

We have a few comments on the quasi-posterior f_n. First, f_n is essentially a monotone transformation of the rescaled sample objective function. Under such a transformation, $f_n(\theta) \xrightarrow{p} 0$ at an exponential rate for any $\theta \notin \Theta_I$. Second, the exponential function is not the only possible choice. In the case of point identification (as in CH2003), the exponential transformation is a natural choice because it leads to normal approximation for the quasi-posterior density in large samples. In partially identified models, the exponential transformation has a similar property, which will be illustrated in Example A below. Third, our analysis applies if we define $f_n(\theta) \propto \pi(\theta)\exp(nL_n(\theta))$ for a weighting function π which takes positive values over Θ_I. In this paper, we take $\pi(\theta) \equiv 1$ so that

[2] We could define the objective function as $L(\theta) = -\left\|\mathbb{E}m_i(\theta)\Sigma^{1/2}(\theta)\right\|_+^2$ for some weighting matrix $\Sigma(\theta)$. In this paper, we let $\Sigma(\theta) = I$ for the ease of notation. Our approach can be extended to accommodate such weighting matrices.

the quasi-posterior can be approximated by a uniform distribution over the identified set in large samples.

Let $\theta = (\theta_1, \theta_2')'$ and $\theta_0 = (\theta_{01}, \theta_{02}')'$, where θ_1 is a scalar and θ_2 is potentially vector-valued. From now on we assume that researchers are interested in estimating and making inferences about θ_{01}. Define

$$\Theta_1 = \{\theta_1 : \exists\theta_2 \text{ such that } (\theta_1, \theta_2') \in \Theta\} \text{ and } \Theta_{I1} = \{\theta_1 : \exists\theta_2 \text{ such that } (\theta_1, \theta_2') \in \Theta_I\},$$

that is, Θ_1 and Θ_{I1} are the first dimension of Θ and Θ_I, respectively. Define Θ_2 and Θ_{I2} similarly. Let

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

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

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

be a collection of θ_2s 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 $\tau \in [0, 1]$, we 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

We propose estimators for $\theta_{1\ell}$ and θ_{1u} in this section. Example A below (Example 1 in CHT2007) illustrates the idea behind our estimator.

Example A (interval-observed data)**.** Let $\{(Y_{\ell i}, Y_i, Y_{ui})\}_{i=1}^n$ be a sequence of i.i.d. random vectors. Assume that $Y_{\ell i}$, Y_i, and Y_{ui} have finite first two moments 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. We define the population and sample objective functions as follows:

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

For the purpose of illustration, assume that $\theta_u > \theta_\ell$ and $\bar{Y}_u > \bar{Y}_\ell$. Let Θ be a compact subset of $\mathbb{R}$ whose interior contains $[\theta_\ell, \theta_u]$. Consider first an "infeasible" quasi-posterior:

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

It is easy to observe that $f_\infty(\theta) = 1/(\theta_u - \theta_\ell)$ if $\theta \in \Theta_I$ and that $f_\infty(\theta) = 0$ otherwise. Hence, the support of the "infeasible" quasi-posterior is exactly the identified set. Since the population objective function is unknown, we construct a "feasible" version of f_∞ by replacing L with its sample analog 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 and is flat over the interval $[\bar{Y}_\ell, \bar{Y}_u]$ and declines towards both end points of Θ. Note also that under the exponential transformation, f_n is proportional to normal densities when $\theta \geq \bar{Y}_u$ or $\theta \leq \bar{Y}_\ell$. $\square$

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 sizes, with $Y_{\ell i} \sim U[-1, 0]$ and $Y_{ui} \sim U[1, 2]$. We can see that as n increases, the "quasi-posterior mass" outside of the identified set decreases to 0. When $d > 1$, a similar pattern can be 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 of the marginal quasi-posterior whose probability level converges to 0 (or 1) can be a candidate estimator for $\theta_{1\ell}$ (or θ_{1u}). Specifically, we define the following estimators for the two end points of Θ_{I1}:

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

where $\hat{\tau}_\ell$ and $\hat{\tau}_u$ are chosen by researchers.

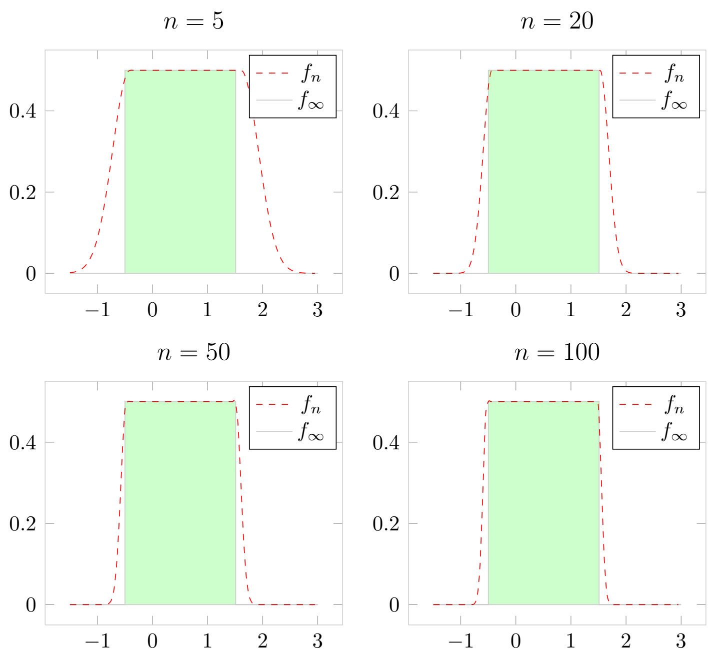

FIGURE 1. Quasi–posterior density (Example A)

3.1. **Consistency.** We develop conditions on $\hat{\tau}_\ell$ and $\hat{\tau}_u$ such that $\hat{\theta}_{1\ell}$ and $\hat{\theta}_{1u}$ are consistent estimators in this subsection. For ease of exposition, we focus on $\hat{\theta}_{1\ell}$; the results for $\hat{\theta}_{1u}$ are similar.

Assumption 3.1. *Θ is compact.*

Assumption 3.1 is standard. It ensures that $D_n = \int_{t\in\Theta} \exp(nL_n(t))dt$ is finite because the integrand is bounded over Θ.

Assumption 3.2. *Θ_I is connected.*

Assumption 3.2 says that Θ_I is not a union of disjoint sets. It ensures that the projections of the identified set onto each axis are single intervals. We impose this assumption to simplify notation. With a simple modification to the computation algorithm, our methods still work without this assumption (see discussions in Section 6.2). Note that we do not require the identified set to be convex.

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 a standard partial identification assumption. It requires that when parameters take values outside the identified set, the expectations of the moments are at least proportional to the smallest distance from the parameter to any points in the identified set.

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

Assumptions 3.1 and 3.4 imply 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 that are uniformly bounded on Θ.

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

Assumption 3.5 is also made in CHT2007. It requires 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.

Assumption 3.6. $\hat{\tau}_\ell$ *and* $\hat{\tau}_u$ *are chosen such that*

$$\hat{\tau}_\ell = \hat{\tau}_u = \frac{\hat{c}}{n^{d/2}D_n},$$

where $\hat{c}$ *is positive and* $\hat{c} \xrightarrow{p} c > 0$*.*[3]

In the expression of $\hat{\tau}_\ell$ and $\hat{\tau}_u$, n is sample size and D_n is defined in Equation (1) and can be computed by Monte Carlo integration. $\hat{c}$ is chosen by researchers. The constant c plays a role similar to the tuning parameter we often seen in the partial identification literature (see discussions in Bugni, 2010). To the best of our knowledge, there is no generic rule of choosing such tuning parameters. Although the asymptotic properties of our estimator do not rely on a specific c, we propose a heuristic rule for choosing c in Section 5. In our Monte Carlo experiment, we find that this rule approximately minimizes the mean squared error of our estimator. Theoretic derivation for an "optimal" choice of c is beyond the scope of this paper and will be left for future research.

[3] In practice, we may want to choose $\hat{\tau}_\ell = \hat{\tau}_u = \min\left\{\frac{\hat{c}}{n^{d/2}D_n}, \frac{1}{2}\right\}$ to ensure $0 \leq \hat{\tau}_\ell \leq 1 - \hat{\tau}_u \leq 1$ in finite samples. This does not affect the asymptotics of the estimators because when the model is not point-identified, for any $c > 0$, the minimum is obtained at $\hat{c}/n^{d/2}D_n$ with probability one.

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. □

We have some comments on Theorem 3.1. First, in a large sample, the "quasi-posterior mass" over Θ_I converges in probability to 1. Therefore, the "marginal quasi-posterior mass" over $[\theta_{1\ell}, \theta_{1u}]$ also coverages in probability to 1. To achieve consistency, we "cut" two properly-sized tails off from the marginal quasi-posterior. When the identified set has a positive Lebesgue measure, D_n converges in probability to a positive constant, implying that $\hat{\tau}_\ell = \hat{c}/(n^{d/2}D_n) \propto n^{-d/2}$. In this case, we essentially cut off two tails whose mass converges to zero at the rate $n^{-d/2}$.

Second, the rate requirement in Assumption 3.6 can be relaxed in specific scenarios. For example, if one knows that the identified set has a positive Lebesgue measure, one can allow that $\hat{c}$ diverges as long as $\hat{c}/n^{d/2} \xrightarrow{p} 0$ at a polynomial rate. Assumption 3.6 is stronger than necessary because it also ensures the consistency of this estimator even when the identified set has an empty interior (i.e., if Θ_I is a singleton or other lower dimension subset of $\mathbb{R}^d$). For example, when $\Theta_I = \{\theta_0\}$, $n^{d/2}D_n = O_p(1)$, our estimator is essentially a random quantile of the marginal quasi-posterior. This coincides with the result in CH2003 that any quantiles are consistent estimators in point-identified models.

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

3.2. **Rate of convergence.** In this subsection, we provide conditions under which the estimators are $\sqrt{n}$-consistent regardless of θ_0 being point- or partially identified. The convergence rate is needed to construct confidence sets for θ_{01}.

Example A continued. We illustrate the idea of obtaining the $\sqrt{n}$-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 also be a $\sqrt{n}$-consistent estimator. 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 the 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. □

When θ_0 is a vector, the appropriate choice of $\hat{\tau}_\ell$ depends on how fast the tail mass of the marginal quasi-posterior decreases to zero, which in turn depends on the shape of the set $\Theta_{I2}(\theta_{1\ell}) = \{\theta_2 : (\theta_{1\ell}, \theta_2) \in \Theta_I\}$, a set of "extreme points". In Pakes, Porter, Ho, and Ishii (2006), this set is assumed to be a singleton. In the main text of this paper, we also assume $\Theta_{I2}(\theta_{1\ell})$ is a singleton. However, our estimation method can be extended and allows $\Theta_{I2}(\theta_{1\ell})$ to have a positive Lebesgue measure, see Section 6.1.

Assumption 3.7. $\Theta_{I2}(\theta_{1\ell}) = \{\theta_\ell\}$.

Assumption 3.8. *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 let $\mathbb{E}m_1^{\mathcal{J}}(\theta)$ be the subvector of expectations of binding moments, that is, $\mathbb{E}m_1^{\mathcal{J}}(\theta) = 0$. 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 based on Assumptions 3.3 and 3.8, there exist positive constants k and K such that the absolute value of every component of $Q_1^{\mathcal{J}}(\theta_\ell)$ 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 3.9. *$Q^{\mathcal{J}}(\theta_\ell)$ has full column rank.*

Assumption 3.9 is crucial to ensure that the quasi-posterior decreases sufficiently quickly within a $\sqrt{n}$-local neighborhood of the corner point θ_ℓ. It implies that $J \geq d$ and that there are no more than $J - d$ moment equations that are proportional to each other. Assumption 3.9 is common in moment inequality models (e.g., Kaido, 2010, and PPHI).

Theorem 3.2. *Suppose that Assumptions 3.1 to 3.3 and 3.5 to 3.9 are satisfied; then, $\sqrt{n}(\hat{\theta}_{1\ell} - \theta_{1\ell}) = O_p(1)$.*

Proof. See Appendix B.2. □

4. Inference

In the previous section, we show that the interval formed by two "extreme quantiles" is a $\sqrt{n}$-consistent estimator for $[\theta_{1\ell}, \theta_{1u}]$. The next question we address is how to choose two quantiles of the marginal quasi-posterior such that the resulting interval covers θ_{01} asymptotically with a prespecified probability. There are two issues. First, θ_{01} may be point–identified or partially identified; our procedure accommodates both cases. Second, we use a weighting method to pick out the binding moments.

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

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

The inference about θ_{01} is based on the following observation. Assume that θ_ℓ is in the interior of Θ and $n^{d/2}D_nF_{1n}(\theta_{1\ell})$ converges in distribution to a continuous random variable ξ_ℓ (will be shown later). Let $c_\ell(\alpha)$ be the α–th quantile of ξ_ℓ; then

$$\begin{aligned}\mathbb{P}\left\{\theta_{1\ell} \geq F_{1n}^{-1}\left(\frac{c_\ell(\alpha)}{n^{d/2}D_n}\right)\right\} &= \mathbb{P}\left\{n^{d/2}D_nF_{1n}(\theta_{1\ell}) \geq c_\ell(\alpha)\right\} \\ &= \mathbb{P}\{\xi_\ell \geq c_\ell(\alpha)\} + o(1) = 1-\alpha+o(1).\end{aligned}$$

Therefore, 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 two-sided confidence intervals for θ_{01}.

Assumption 4.1. $[\theta_{1\ell}, \theta_{1u}]$ *belongs to the interior of* Θ_1.

Lemma 4.1. *Let* $\xi_{\ell n} = n^{d/2}D_nF_{1n}(\theta_{1\ell})$*. Suppose that Assumptions 3.1 to 3.3, 3.5, 3.8, 3.9 and 4.1 are satisfied. Then*

$$\xi_{\ell n} \xrightarrow{d} \xi_\ell = \int_{\{h:h_1\leq 0\}} \exp\left(-\|\Delta^{\mathcal{J}}(\theta_\ell) + Q^{\mathcal{J}}(\theta_\ell)h\|_+^2\right) dh, \tag{2}$$

where Δ *is defined in Assumption 3.5 and* $\mathcal{J}$ *indicates the identity of binding moments.*

Proof. See Appendix C.1. □

Note that the distribution of ξ_ℓ depends on the identities of the binding moments. Let $\gamma^*(\theta)$ be a J-dimensional vector of 1s and 0s, indicating the moment is binding or not binding. Then, we can write,

$$\xi_\ell = \int_{\{h:h_1\leq 0\}} \exp(-\|\gamma^*(\theta_\ell)\otimes(\Delta(\theta_\ell)+Q(\theta_\ell)h)\|_+^2)d\theta_2 dh_1,$$

where $\otimes$ stands for the component-wise product.

To have a confidence set with correct asymptotic coverage probability for θ_{01}, one also needs to take into account of the length of the interval $[\theta_{1\ell}, \theta_{1u}]$. Let $T = \theta_{1u} - \theta_{1\ell}$ and $\hat{T}$ be a $\sqrt{n}$-consistent estimator for T. We construct a confidence interval for θ_{01} as

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

where $\tau^{\mathrm{I}}_{\ell} = c^{\mathrm{I}}_{\ell}/(n^{d/2}D_n)$, $\tau^{\mathrm{I}}_{u} = c^{\mathrm{I}}_{u}/(n^{d/2}D_n)$, and $(c^{\mathrm{I}}_{\ell}, c^{\mathrm{I}}_{u})$ is a solution to the following problem:[4]

$$\begin{aligned}
(c^{\mathrm{I}}_{\ell}, c^{\mathrm{I}}_{u}) = \underset{(c_\ell, c_u)\in\mathbb{R}^+\times\mathbb{R}^+}{\operatorname{argmin}} & \left|F_{1n}^{-1}\left(\frac{c_\ell}{n^{d/2}D_n}\right) - F_{1n}^{-1}\left(1-\frac{c_u}{n^{d/2}D_n}\right)\right| \\
\text{s.t.}\quad & \mathbb{P}\left\{c_\ell \le \xi_\ell, c_u \le \sqrt{n}\nu\left(\frac{\beta_n}{n\hat{T}}\right) + \xi_u\right\} = 1-\alpha, \\
& \mathbb{P}\left\{c_u \le \xi_u, c_\ell \le \sqrt{n}\nu\left(\frac{\beta_n}{n\hat{T}}\right) + \xi_\ell\right\} = 1-\alpha,
\end{aligned} \tag{3}$$

where $\nu(x) = \phi(x)/\phi(0)$ and ϕ is the standard normal density. β_n is a tuning parameter satisfying Assumption 4.2 below.

Assumption 4.2. *$\beta_n/\sqrt{n} \to \infty$ and $\beta_n/n \to 0$.*

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 αth quantiles of ξ_ℓ and ξ_u, 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 ξ_ℓ and ξ_u. $\sqrt{n}\nu\left(\beta_n/n\hat{T}\right)$ plays the same role as the shrinkage parameter of Stoye (2009). One possible choice for β_n is to use the iterated logarithm: $\beta_n = n/(2\ln\ln n)$.

Theorem 4.1. *Suppose that Assumptions 3.1 to 3.3, 3.5, 3.8, 3.9, 4.1 and 4.2 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. □

4.1.2. *Constructing $\widehat{\Theta}^{\mathrm{I}}_{\alpha}$.* The confidence set $\Theta^{\mathrm{I}}_{\alpha n}$ is infeasible because the joint distribution of ξ_ℓ and ξ_u is unknown; as a result, c^{I}_{ℓ} and c^{I}_{u} are unknown. In this subsection, we propose an algorithm to obtain consistent estimates for c^{I}_{ℓ} and c^{I}_{u}. We highlight the big picture in the main text and leave the

[4] If there are multiple solutions to Equation (3), we use an arbitrary one.

detailed algorithm in Appendix E (Algorithm 2). Throughout this subsection, it is assumed that there are $\sqrt{n}$–consistent estimates available. The whole vector θ_ℓ can then be estimated by $\hat{\theta}_\ell = (\hat{\theta}_{1\ell}, \hat{\theta}_2')'$, where

$$\hat{\theta}_2 = \int_{\theta_2 \in \Theta_2} \theta_2 f_n(\hat{\theta}_{1\ell}, \theta_2) d\theta_2. \tag{4}$$

To estimate th critical values c_ℓ^I (c_u^I similar), we consider to approximate the distribution of ξ_ℓ by drawing random numbers from

$$\hat{\xi}_\ell = \int_{\{h: h_1 \le 0\}} \exp(-\|\hat{\gamma}(\hat{\theta}_\ell) \otimes (\hat{\Delta}(\hat{\theta}_\ell) + \widehat{Q}(\hat{\theta}_\ell)h)\|_+^2) dh. \tag{5}$$

We use $\hat{\gamma}$ to estimate the moment selection vector γ^*. In particular, $\hat{\gamma}(\hat{\theta}_\ell)$ is a J–vector of weights with

$$\hat{\gamma}_j(\hat{\theta}_\ell) = \frac{\exp(-\beta_n |\frac{1}{n}\sum_{i=1}^n m_{(j)}(W_i, \hat{\theta}_\ell)|^2)}{\exp(-\beta_n |\frac{1}{n}\sum_{i=1}^n m_{(j)}(W_i, \hat{\theta}_\ell)|_+^2)}. \tag{6}$$

By construction, each element $\hat{\gamma}_j \in (0, 1]$. As is shown in Lemma C.1, for any $\theta \in \Theta_I$, the weight $\hat{\gamma}_j(\theta)$ converges to one if the j–th moment is binding at θ; it converges to zero otherwise.[5]

$\hat{\Delta}(\hat{\theta}_\ell)$ is a J-dimensional normal random vector whose variance $\hat{V}$ is a consistent estimator of $V(\theta_\ell) = \mathbb{E} m_1(\theta_\ell) m_1'(\theta_\ell)$. For example, $V(\theta_\ell)$ can be estimated by a sample analog and θ_ℓ can be replaced by $\hat{\theta}_\ell$. $\widehat{Q} = \partial \bar{m}(\theta)/\partial\theta|_{\theta = \hat{\theta}_\ell}$.

Given the random draws from $\hat{\xi}_\ell$, we construct $\widehat{\Theta}_\alpha^{\mathrm{I}}$ by replacing the unknown parts in Equation (3) with their estimates. We add Assumption 4.3 to ensure that $\widehat{Q}$ converges in probability to its population counterpart. Assumption 4.3 implies that Assumptions 3.4 and 3.8 are satisfied.

Assumption 4.3. *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 $\mathbb{E} d(W_1) < \infty$.*

Theorem 4.2. *Suppose that Assumptions 3.1 to 3.3, 3.5, 3.7, 3.9 and 4.1 to 4.3 are satisfied; then*

$$\lim_{n\to\infty} \inf_{\theta_{01} \in [\theta_{1\ell}, \theta_{1u}]} \mathbb{P}(\theta_{01} \in \widehat{\Theta}_\alpha^{\mathrm{I}}) = 1 - \alpha. \tag{7}$$

Proof. See Appendix C.3. □

[5] When the j-th moment is nearly binding at θ^*, that is, when $\mathbb{E} m_{(j)}(\theta^*) = \lambda/\sqrt{n}$ for some $\lambda \in (-\infty, 0)$, $\hat{\gamma}_j(\theta^*)$ converges in distribution to a random variable that takes value from $(0, 1)$. In this paper, we consider pointwise asymptotics only.

Several points about Theorem 4.2 are worth noting. First, $\widehat{\Theta}_\alpha^{\mathrm{I}}$ is constructed directly from the marginal quasi-posterior rather than as a projection of a high-dimensional confidence set. Constructing $\widehat{\Theta}_\alpha^{\mathrm{I}}$ does not require resampling procedures, instead, we just need to obtain random draws from $\hat{\xi}_\ell$ and $\hat{\xi}_u$. As shall be clear in Algorithm 2, each random draw involves only computing a d-dimensional integral whose integrand is a parametric function of a Gaussian process. Second, our weighting method picks out the binding moments asymptotically. Third, we introduce an additional "shrinkage term" to accommodate the point identification case.

4.2. **Constructing a confidence set for the interval $[\theta_{1\ell}, \theta_{1u}]$.** Our method can be used to construct confidence sets for $[\theta_{1\ell}, \theta_{1u}]$. For any $0 < \alpha < 1/2$, let $(c_\ell^{\mathrm{II}}, c_u^{\mathrm{II}})$ be a solution to the following problem:

$$
\begin{aligned}
(c_\ell^{\mathrm{II}}, c_u^{\mathrm{II}}) = \underset{(c_\ell, c_u) \in \mathbb{R}^+ \times \mathbb{R}^+}{\operatorname{argmin}} \left| F_{1n}^{-1}\left(\frac{c_\ell}{n^{d/2} D_n}\right) - F_{1n}^{-1}\left(1 - \frac{c_u}{n^{d/2} D_n}\right) \right| \\
\text{s.t.} \quad \mathbb{P}\left(c_\ell \le \xi_\ell, c_u \le \xi_u\right) = 1 - \alpha.
\end{aligned}
$$

Let $\Theta_{\alpha n}^{\mathrm{II}} = \left[F_{1n}^{-1}\left(c_\ell^{\mathrm{II}} / n^{d/2} D_n\right), F_{1n}^{-1}\left(1 - c_u^{\mathrm{II}} / n^{d/2} D_n\right)\right]$.

Theorem 4.3. *Suppose that Assumptions 3.1 to 3.3, 3.5, 3.8, 3.9, 4.1 and 4.2 are satisfied. Then*

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

Proof. See Appendix C.4. □

5. Experiment

This section illustrates the integration-based inference approach using Example B, in which partial identification arises as a consequence of multiple equilibria.

Example B (Entry game)**.** We consider a complete information game with two players ($j = 1, 2$) and two actions ($Y_{ij} = 0, 1$) (Bugni, Canay, and Guggenberger, 2010, Example 2.2). An example of this game is that two oligopoly firms decide to enter a local market simultaneously. The index i indicates an observation (a market). The profit for firm j when choosing action $Y_{ij} = 0$ is normalized to be zero regardless of the choice of the other firm ($Y_{i,-j}$). The profit for firm j when choosing action 1 is

$$
\pi_{ij} = u_{ij} - \theta_{0j} \mathbf{1}[Y_{i,-j} = 1],
$$

where u_{ij} is the profit shock, which is uniformly distributed on unit interval and is i.i.d. across i and j. θ_{0j} measures the competition effect. We assume that $\theta_{0j} \in [0, 1]$.

The realizations of profit shocks and the value of $(\theta_{01}, \theta_{02})$ are known to both players. Players play pure strategy Nash equilibria. Researchers observe outcome $\{(Y_{i1}, Y_{i2})\}_{i=1}^n$ and know the distribution of $\{(u_{i1}, u_{i2})\}$. The object of interest is to make inferences about θ_{01} while being agnostic about the equilibrium selection mechanism.

As is illustrated in Figure 2, depending upon the realization of u_{i1} and u_{i2}, there may be multiple equilibria in this game. When (u_{i1}, u_{i2}) belongs to area A, B, or C, the model predicts a unique equilibrium; on the contrary, when (u_{i1}, u_{i2}) belongs to area M, both $(0, 1)$ and $(1, 0)$ could be equilibrium outcomes.

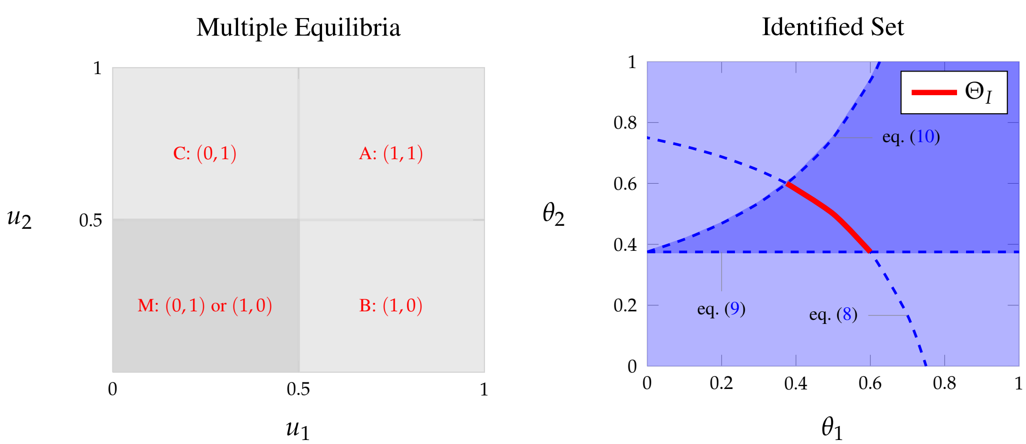

FIGURE 2. Multiple equilibria and the identified set

We can write a set of moment equations based on the model restrictions. The probability of observing outcome $(1, 1)$ is equal to the probability of (u_{i1}, u_{i2}) belongs to area A, which delivers one moment equation (or two moment inequalities):

$$\mathbb{P}\{(1, 1)\} = \mathbb{P}\{A\} = (1 - \theta_{01})(1 - \theta_{02}). \tag{8}$$

The probability statement for outcome $(1, 0)$ is more complicated. Since the equilibrium selection mechanism is unspecified, one can only derive bounds of $\mathbb{P}\{(1, 0)\}$: it is larger than the mass of area B, but smaller than the mass of area $B \cup M$. Hence, we have two more moment inequalities:

$$\mathbb{P}\{(1,0)\} \leq \mathbb{P}\{B \cup M\} = \theta_{02}. \tag{9}$$

$$(1-\theta_{01})\theta_{02} = \mathbb{P}\{B\} \leq \mathbb{P}\{(1,0)\}. \tag{10}$$

Equation (8), inequalities (9) and (10) characterize a set of parameters that is consistent with the model restriction and distribution of observed data. This set is the identified set Θ_I.[6]

To generate random samples from this model, we specify $\theta_{01} = \theta_{02} = 0.5$. When there are multiple equilibria, an equilibrium is selected by tossing a coin (unknown to researchers). Given this design, we can calculate that $\theta_{1\ell} = 0.375$ and $\theta_{1u} = 0.6$. We consider three sample sizes: $n = 200, 400$, and 800.

We use a random walk Metropolis Hasting (MH) algorithm to make random draws from the marginal quasi-posterior, as shown in Algorithm 1.[7] Figure 3 shows the construction of a quasi-posterior. It also plots the last $5,000$ draws from the marginal quasi-posterior.

Algorithm 1. Estimation.

(1) Construct quasi-posterior f_n as in Equation (1).

(2) Draw a sequence of vectors from f_n, $\left\{\theta^1, \theta^2, \cdots, \theta^B\right\}$, as follows:

   (a) $\theta^0 = (\theta_1^0, \theta_2^0) = (0,0)$.

   (b) Update θ_1, holding θ_2^0 fixed.

   - $\theta_1^{temp} = \theta_1^0 + \epsilon$, $\theta_2^{temp} = \theta_2^0$, $\epsilon \sim N(0, 0.04)$.
   - $a \sim U[0,1]$.
   - $\theta_1^1 = \theta_1^{temp}$ if $a \leq f_n(\theta^{temp})/f_n(\theta^0)$; $\theta_1^1 = \theta_1^0$ otherwise.

   (c) Update θ_2, holding θ_1^1 fixed.

   (d) ......

   (e) Reach θ^B.

(3) Let $B_1 = B/2$ be the burn-in period.

(4) Take the first components of the chain after the burn-in period: $\left\{\theta_1^{B_1+1}, \cdots, \theta_1^B\right\}$.

[6] Note that the identified set is not convex.

[7] See Robert and Casella (2004) for a comprehensive summary of MCMC algorithms.

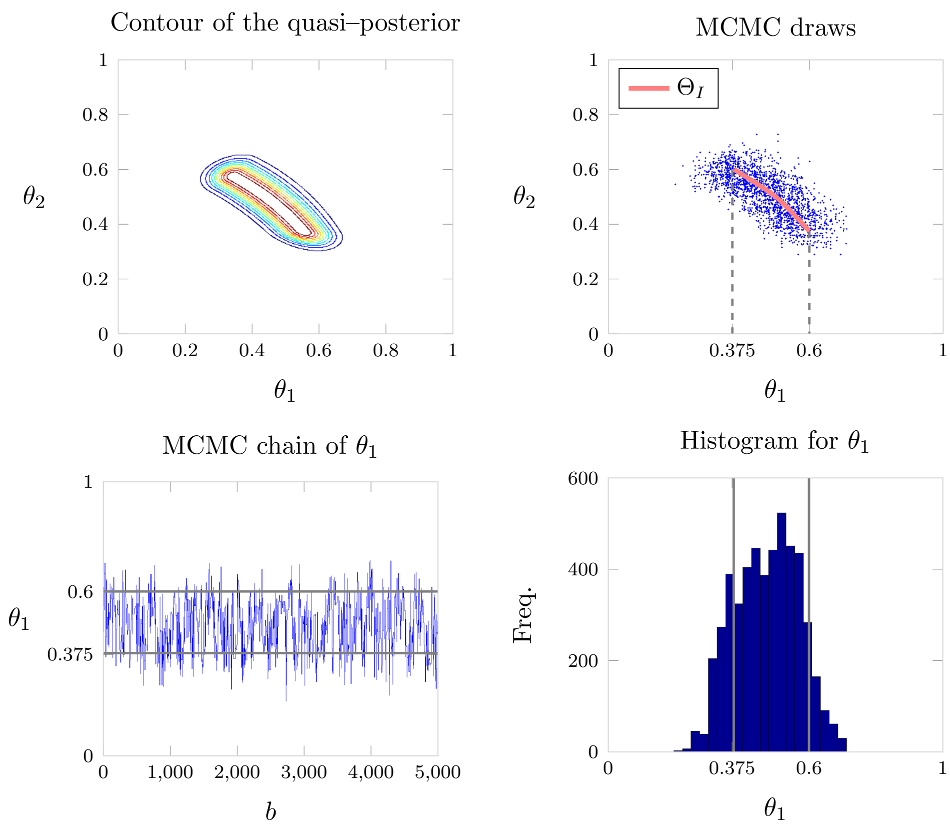

FIGURE 3. MCMC draws

(5) $\hat{\theta}_{1\ell}$ is determined by

$$\hat{\theta}_{1\ell} = F_{1n}^{-1}\left\{\frac{c}{nD_n}\right\}$$

for some constant $c > 0$.

Algorithm 1 does not specify how to choose the tuning parameter c. Although $\hat{\theta}_{1\ell} = F_{1n}^{-1}\{c/(nD_n)\}$ is a $\sqrt{n}$-consistent estimator for any fixed value of c, its finite sample performance will be affected. Figure 4 reports the mean squared error (MSE) and standard deviation of $\hat{\theta}_{1\ell}$ for $c \in \{0.5, 1.0, \cdots, 5.5, 6.0\}$ under different sample sizes (based on $1,000$ replications). The figure for $\hat{\theta}_{1u}$ is similar and is therefore omitted. There are some interesting points worth noting. First, MSEs are small. When the sample size is doubled, MSEs roughly decease by half—support the $\sqrt{n}$-convergence rate. In larger samples, the MSE curve is more flat and the choice of c matters less. Second, it seems that there is a value of c minimizing MSEs. Note that the standard deviations

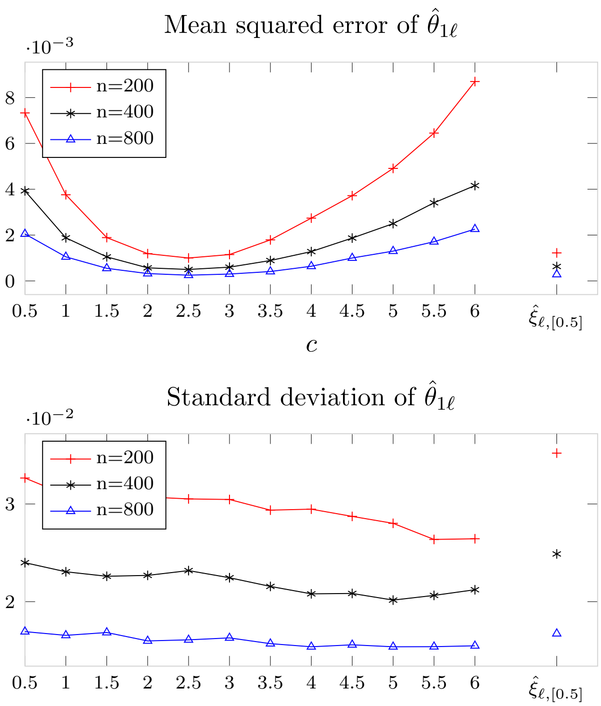

FIGURE 4. Mean squared errors under different choices of c

curve are flat in c; hence the U-shape MSE curve is mainly caused by the finite sample biases. It is intuitive in the sense that at a fixed sample size, $c \uparrow$ implies $c/(nD_n) \uparrow 1$, then $\hat{\theta}_{1\ell}$ will be close to the maximum of the support of the marginal quasi-posterior. This induces positive-valued bias. If $c \downarrow 0$, then $\hat{\theta}_{1\ell}$ is the minimum of the support of the marginal quasi-posterior, which leads to negative-valued bias.

Although deriving an "optimal" choice of c is beyond the scope of this paper, we propose a heuristic rule for choosing c that significantly reduces the biases in our example. Recall that the rescaled tail mass $nD_nF_n(\theta_{1\ell}) \xrightarrow{d} \xi_\ell$. If we choose $c = \xi_{\ell,[0.5]}$, the median of the ξ_ℓ, then asymptotically $\hat{\theta}_{1\ell}$ will be greater than $\theta_{1\ell}$ with a probability of 0.5:

$$\mathbb{P}(\hat{\theta}_{1\ell} > \theta_{1\ell}) = \mathbb{P}(nD_nF_n(\theta_{1\ell}) < \xi_{\ell,[0.5]}) \to 0.5.$$

We can hence achieve a "balance" between the positive-valued bias caused by large c and the negative-valued bias caused by small c. The median $\xi_{\ell,[0.5]}$ is unknown, but can be consistently estimated based on first stage estimates and the simulation procedure described in Algorithm 2. Following this rule, we re-estimate the model by choosing $c = 1$ in the first stage and $c = \hat{\xi}_{\ell,[0.5]}$ in the second stage. For the ease of comparison, we plot the results on the right end of Figure 4 (numbers are in Tables 2 and 3 of Appendix E). We can see that this method of choosing c produces MSE no greater than any other fixed choices of c.

Table 1 compares three inference methods when the object of interest is constructing confidence intervals for $\theta_{1\ell}$. The first two columns report the coverage probability of projected confidence intervals where the joint confidence sets are constructed following the method in CHT2007 (with simulated critical values). Columns 3–7 are based on methods of profiling out nuisance parameters (Romano and Shaikh, 2008), with critical values obtained from subsampling. The last column reports the results of our inference procedure.

TABLE 1. Coverage frequency

	Joint	Proj.-based	Profiling out & subsampling					Int.-based
Subs. size			10	12	15	17	20	
95% level								
$n = 200$	95.6%	99.7%	97.9%	97.7%	98.9%	98.5%	97.4%	93.8%
$n = 400$	94.7%	99.9%	98.6%	97.8%	95.8%	97.6%	97.0%	94.8%
$n = 800$	95.1%	99.8%	97.5%	97.9%	96.0%	95.6%	98.8%	95.2%
90% level								
$n = 200$	90.2%	99.5%	92.9%	94.1%	95.0%	94.8%	94.0%	88.2%
$n = 400$	90.7%	99.4%	92.2%	91.1%	94.4%	95.4%	94.0%	89.7%
$n = 800$	89.9%	99.5%	92.3%	93.2%	90.4%	90.9%	95.8%	89.9%

Based on $1,000$ replications.

We can see from the table that the joint confidence sets performances well across three sample sizes; however, the projections are very conservative. The method of profiling out nuisance parameters and subsample critical values has coverage probabilities close to the desirable levels under suitable choices of subsample sizes. For example, at the 95% level, the choice of 12, 15, and 17 delivers the best results for sample sizes of 200, 400, and 800, respectively. At the 90% level, the choices of 10, 12 and 15 work better than others. The last column shows the performance of our method, which performances reasonably well.

We also documented the computation time of different methods. For the projection method and the integration-based method, we estimate the critical values by making $10,000$ random draws from the (plug-in) limiting distributions. For the method of profiling out nuisance parameters, we generate

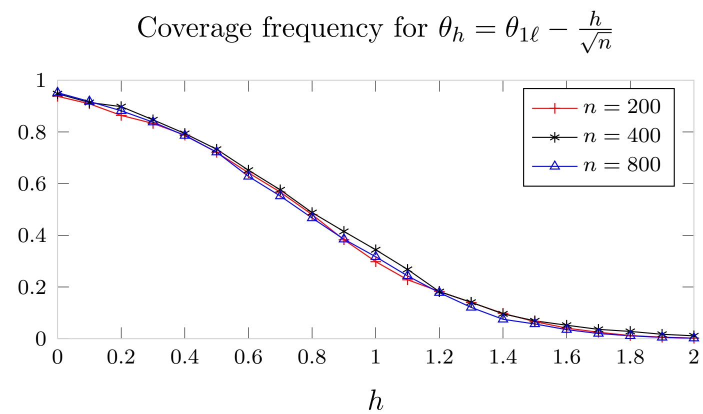

FIGURE 5. Local power

10,000 subsamples. The numerical precision of simulating critical values is therefore comparable across three methods. For our method, we construct an MCMC chain with a length of 500,000. With the acceptance rate at around 32%, the chain explores about 160,000 points in the parameter space $[0,1] \times [0,1]$ and is comparable to 400×400 grids.[8] We approximate the integral in Equation (2) by drawing (h_1, h_2) 10,000 times from a bivariate normal distribution. Under these configuration, with a sample of size 800, it takes the projection method about 367 CPU seconds to finish one replication (inverts 400×400 hypothesis tests). The computation time for the profiling-out method is about 35 seconds (inverts 400 hypothesis tests, and for each test draws 10,000 subsamples). Our method takes about 21 seconds.[9]

There are several points worth noting. First, our method does not require the choice of subsample sizes. Second, our method is expected to have more computational advantage when parameters are high-dimensional. In large samples, "quasi-posterior mass" concentrates on the identified set. Making MCMC draws from the quasi-posterior is numerically similar to making MCMC draws from a muti-dimensional uniform distribution. The running time for a random walk MH algorithm can be bounded by polynomial orders of parameter dimension (e.g., Beskos and Stuart, 2009; Lovasz and Vempala, 2005).[10]

[8] Most of the MCMC draws are located around the identified set, as opposed to the evenly distributed grid points.

[9] We ran programs at SciNet, a Canadian supercomputer center (http://www.scinethpc.ca/).

[10] If th quasi-posterior approaches to a normal density (e.g., when the model is point-identified), the running time of a random walk MH algorithm is bounded in probability by the order of d^2 (see Belloni and Chernozhukov, 2008).

Lastly, Figure 5 shows the coverage probability for alternatives $\theta_h = \theta_{1\ell} - h/\sqrt{n}$ in 1,000 replications as a function of h. It shows that the inference procedure has non-trivial power against local alternatives.

6. Discussions and Conclusion

6.1. **Extreme points are not singletons.** In this subsection, we discuss the $\sqrt{n}$-consistent estimation of $\theta_{1\ell}$ when $\Theta_{I2}(\theta_{1\ell})$ is not necessarily a singleton, that is, without Assumption 3.7. The shape of $\Theta_I(\theta_{1\ell})$ turns out to be important for the $\sqrt{n}$-estimation of $\theta_{1\ell}$. Lemma 6.1 below illustrate in the two dimensional case that to obtain a $\sqrt{n}$-consistent estimator for $\theta_{1\ell}$, the choice of $\hat{\tau}_\ell$ shall depend on the shape of $\Theta_{I2}(\theta_{1\ell})$.

Lemma 6.1. *Suppose that Assumptions 3.1, 3.2, 3.3, 3.5, 3.6, 3.8, 3.9 and 4.1 are satisfied. Suppose that $d = 2$. Let $\hat{\tau}_\ell^a = \hat{c}/(nD_n)$ and $\hat{\tau}_\ell^b = \hat{c}/(\sqrt{n}D_n)$. If $\Theta_{I2}(\theta_{1\ell})$ contains an interval with positive length, then*

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

Proof. See Appendix D.1. □

The implication of Lemma 6.1 is that we have to choose different quantiles according to the shape of $\Theta_I(\theta_{1\ell})$ to obtain $\sqrt{n}$–consistency. Since $\Theta_{I2}(\theta_{1\ell})$ is unknown, it is desirable to construct an estimator $\hat{\theta}_{1\ell}^*$ that 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 $\hat{\theta}_{1\ell}^*$:

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

where

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

By construction, $\sup_{\theta_1} U_n(\theta_1) \le C$ for some $C > 0$. It can be shown (similar to the proof of Theorem 3.2) that when $\Theta_{I2}(\theta_{1\ell})$ is a singleton, $U_{1n}(\theta_{1\ell}) = O_p(1/\sqrt{n})$, in which case we essentially use a probability level decreasing at the rate of n; however if $\Theta_{I2}(\theta_{1\ell})$ is an interval, $U_{1n}(\theta_{1\ell}) = O_p(1)$, and we use a probability level decreasing at the rate $1/\sqrt{n}$. The quantity

$U_{1n}(\theta_{1\ell})$ hence automatically picks out the correct rate. This idea generalizes to any finite dimension d.

In practice, $U_{1n}(\theta_{1\ell})$ is unknown because it depends on $\theta_{1\ell}$; a feasible version of $\tilde{\theta}^*_{1\ell}$ is available:

$$\hat{\theta}^*_{1\ell} = F^{-1}_{1n}\left(\hat{\tau}_\ell(\hat{\theta}^*_{1\ell})\right), \quad \hat{\tau}_\ell(\hat{\theta}^*_{1\ell}) = \frac{\hat{c}U_n(\hat{\theta}^*_{1\ell})}{\sqrt{n}D_n}. \tag{11}$$

Alternatively, it can be written as

$$\frac{f_{1n}(\hat{\theta}^*_{1\ell})}{F_{1n}(\hat{\theta}^*_{1\ell})} = \frac{\sqrt{n}}{\hat{c}}, \tag{12}$$

that is, choosing $\hat{\theta}^*_{1\ell}$ such that the ratio of the density and the distribution function of the marginal quasi-posterior is proportional to $\sqrt{n}$.

Proposition 6.1. *Suppose that Assumptions 3.1 to 3.3, 3.5, 3.6, 3.8, 3.9 and 4.1 are satisfied. Let $\hat{\theta}^*_{1\ell}$ satisfies Equation (11) or (12); then $\sqrt{n}(\hat{\theta}^*_{1\ell} - \theta_{1\ell}) = O_p(1)$.*

Proof. See Appendix D.2. □

6.2. **Disconnected identified set.** When the identified is disconnected, the proposed estimators for $\theta_{1\ell}$ and θ_{1u} are consistent and confidence intervals for θ_{10} are asymptotically valid with a modification to Algorithm 1.

To illustrate this, we consider Example B in Section 5 again but expand the parameter space to $[0,2] \times [0,2]$. We can see from the top two panels of Figure 6 that the identified set is the union of two disconnected curves. The projection of the identified set to the θ_1 dimension is $\Theta_{I1} = [0.375, 0.6] \cup [1.25, 2]$. The minimum and maximum values that θ_{01} can possibly take are $\theta_{1\ell} = 0.375$, $\theta_{1u} = 2$. In this case, the quasi-posterior resembles a multi-modal density.

To make random draws from this multi-modal density, we follow Guan, Fleibner, and Joyce (2006) and use an MH algorithm with "small-world proposals". The idea is to add large jumps to the MCMC chain with a small probability, so the chain can explore all the important area of a multi-modal distribution. As illustrated in Figure 6, we can still consistently estimate $\theta_{1\ell} = 0.375$ and $\theta_{1u} = 2$ by taking two extreme quantiles from the marginal quasi-posterior. For inference, confidence intervals can be constructed following the same procedure as described in previous sections. Note that such confidence intervals have correct asymptotic coverage probability for θ_{01}, satisfying Equation (7), although they will cover the redundant interval $(0.6, 1.25)$.

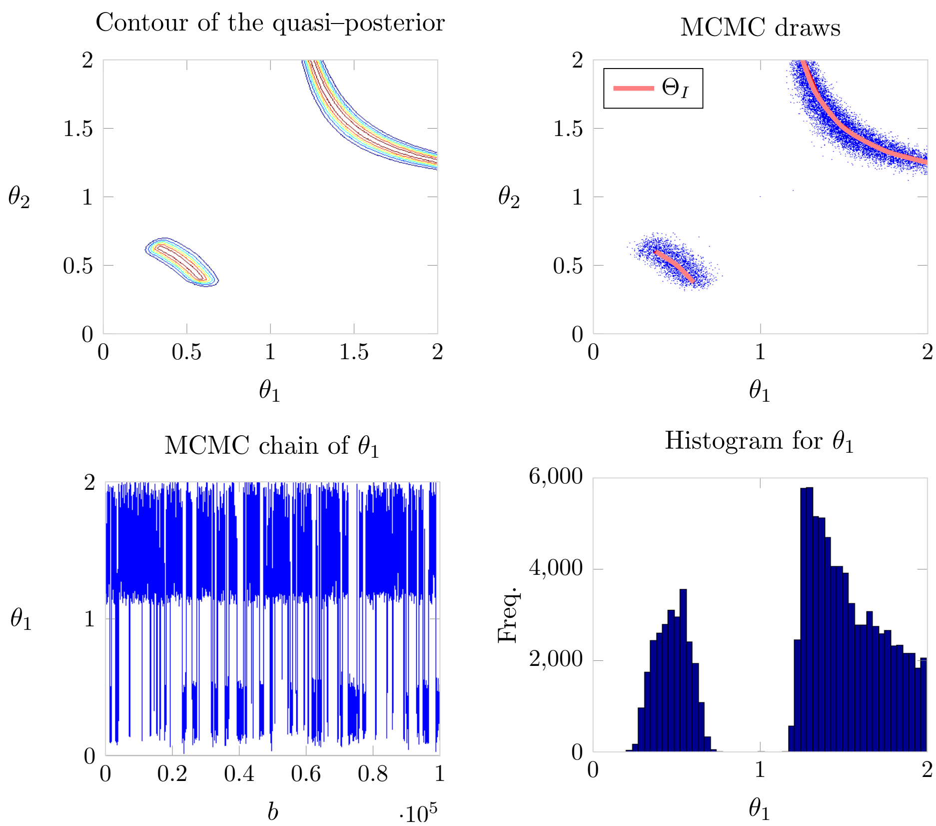

FIGURE 6. MCMC draws with a disconnected identified set

6.3. **Conclusion.** In this paper, we propose integration-based estimation and inference methods for moment inequality models. Our confidence intervals cover each component of the true parameter vector with pre-specified probability and are 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),[11] models characterized by conditional moment inequalities (Andrews and Shi, 2013), or models not characterized by moment inequalities (maximum score estimator when the support condition is violated). Second, we focus on a single element 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

[11] Henry, Meango, and Queyranne (2012) propose a combinatorial inference methods for identified sets that reduces the computational complexity caused by a large number of moments. My paper focuses on inferences on parameter values and aims to reduce the computational burden caused by the high dimensionality of parameters.

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. Third, it is challenging but interesting to extend the current framework to allow for the presence of infinite-dimensional nuisance parameters.

REFERENCES

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 (2013): "Inference Based on Conditional Moment Inequalities," *Econometrica*, 91(4), 609–666.

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

BELLONI, A., AND V. CHERNOZHUKOV (2008): "On the computational complexity of MCMC-based estimators in large samples," *The Annals of Statistics*, 37(4), 2011–2055.

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

BESKOS, A., AND A. STUART (2009): "Computational complexity of Metropolis-Hastings methods in high dimensions," *Proceedings of the 8th International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing (MCQMC 2008)*.

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

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

BUGNI, F., I. CANAY, AND P. GUGGENBERGER (2010): "Distortions of Asymptotic Confidence Size in Locally Misspecified Moment Inequality Models," .

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.

GUAN, Y., R. FLEIBNER, AND P. JOYCE (2006): "Markov Chain Monte Carlo in small worlds," *Stat Comput*, 16, 193–202.

HAHN, J., AND G. RIDDER (2011): "A Dual Approach to Confidence Intervals for Partially Identified Parameters," *Working Paper*.

HENRY, M., R. MEANGO, AND M. QUEYRANNE (2012): "Combinatorial Bootstrap Inference in Partially Identified Incomplete Structural Models," *Working Paper*.

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

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

KIM, J.-Y. (2002): "Limited information likelihood and Bayesian analysis," *Journal of Econometrics*, 107(1-2), 175 – 193.

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

LOVASZ, L., AND S. VEMPALA (2005): "The geometry of logconcave functions and sampling algorithms," *Working Paper*.

MEANGO, R., AND I. MOURIFIE (2013): "Notes on the identification in two equations Probit model with dummy endogenous regressor," *Working Paper*.

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

MOON, H., AND F. SCHORFHEIDE (2012): "Bayesian and frequentist inference in partially identified models," *Econometrica*, 80(2), 755–782.

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*.

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

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

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.

Appendix A. Some Lemmas

We first 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. Lemma A.1 deals with the denominator of the quasi-posterior and lemmas A.2 to A.7 deal with the numerator. All the Lemmas are written for deriving the asymptotics around $\theta_{1\ell}$; they are similar for θ_{1u} and are omitted.

Lemma A.1 says that $n^{d/2}D_n$ 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 $\lim_{n\to\infty} \mathbb{P}(n^{d/2}D_n < C^*) < \epsilon$.*

Proof. Assumption 3.1 ensures that D_n is well defined. For a $C_1 > 0$, define set $\mathscr{A}_n = \{\theta : d(\theta, \Theta_I) \leq \frac{C_1}{\sqrt{n}}\}$. Note that $\mathscr{A}_n$ is compact. Then, by Assumption 3.4, there exist $C_2 > 0$ such that

$$\max_j \sup_{\theta\in\mathscr{A}_n} |\mathbb{E}m_{(j)}(W_1,\theta)|_+ \leq C_2 \sup_{\theta\in\mathscr{A}_n} d(\theta,\Theta_I) \leq C_2C_1/\sqrt{n}.$$

Let $\mu(\mathscr{A}_n)$ be the Lebesgue measure of $\mathscr{A}_n$, then there exist $C_3 > 0$ such that $\mu(\mathscr{A}_n) \geq C_3/n^{d/2}$. Let ι be a J–vector of ones, for any $C > 0$,

$$\mathbb{P}\left\{n^{d/2}D_n < C\right\} \leq \mathbb{P}\left\{n^{d/2}\inf_{\theta\in\mathscr{A}_n}\exp(-n\|\bar{m}(\theta)\|_+^2)\mu(\mathscr{A}_n) < C\right\}$$

$$\leq \mathbb{P}\left\{C_3\inf_{\theta\in\mathscr{A}_n}\exp(-\|\Delta_n(\theta)+\sqrt{n}\mathbb{E}m_1(\theta)\|_+^2) < C\right\} \leq \mathbb{P}\left\{\|\sup_{\theta\in\mathscr{A}_n}\Delta_n(\theta)+C_1C_2\iota\|_+^2 > \log(C_3/C)\right\}.$$

The limit of right hand side is bounded by ϵ by letting C decrease to zero because $\sup_{\theta\in\mathscr{A}_n}\Delta_n(\theta)$ is bounded in probability by Assumption 3.5. □

Lemmas A.2 to A.7 say that when sample size increases to infinity, the integral of the numerator of the quasi-posterior outside of the identified set has the same distribution as a "localized" integral within a $\sqrt{n}$-neighborhood (but outside) of the identified set. The proof follows the same idea as in CH2003 for the point identification case. 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, we can define a integration region for the local parameter $h = \sqrt{n}(\theta - \theta_\ell)$ as $H_n = \{h : -\sqrt{n}(\theta_{1\ell} - \underline{\theta}_1) \le h_1 \le 0, -\sqrt{n}(\theta_{2\ell} - \underline{\theta}_2) \le h_2 \le \sqrt{n}(\overline{\theta}_2 - \theta_{2\ell})\}$. Note that H_n corresponds to the integration region for θ: $\{\theta \in \Theta : \theta_1 \le \theta_{1\ell}\}$. We separate H_n into three parts: $H_{1n} = \{h : \|h\| \le M, h_1 \in H_n\}$, $H_{2n} = \{h : M < \|h\| \le M^*\sqrt{n}, h \in H_n\}$ and $H_{3n} = \{h : \|h\| \ge M^*\sqrt{n}, h \in H_n\}$ for some $M^*, M > 0$. In the rest part of Appendix A, we drop the sup-script $\mathcal{J}$ to simplify the notation. By assumption Assumption 3.2, H_n converges to $\mathbb{R}^- \times \mathbb{R}$ in the Painleve–Kuratowski sense.[12]

Lemma A.2. *Suppose that Assumptions 3.1, 3.2, 3.5 and 3.8 are satisfied, then for any $M > 0$,*

$$\int_{\|h\| \le M} N_n(h) dh \xrightarrow{d} \int_{\|h\| \le M} N_\infty(h) | dh.$$

Proof. Note that $\mathbb{E}m_1(\theta_{1\ell}, \theta_{2\ell}) = 0$. It follows

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

where

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

Note also that

$$\sup_{\|h\| \le M} \|\Delta_n(\theta_\ell) - \Delta_n(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_{2\ell} + \frac{h_2}{\sqrt{n}})\| = o_p(1)$$

since $\Delta_n \xrightarrow{w} \Delta$ by Assumption 3.5. In addition, by Assumption 3.8 (continuous differentiability), $\max_j \sup_{\|h\| \le M} R_{n,j}(h) = o(\|h\|) \le o(M) = o(1)$. The result holds by continuous mapping theorem. □

[12] See discussion in Kaido (2010).

Lemma A.3. *Suppose that Assumptions 3.1, 3.2, 3.5, 3.8 and 3.9 are satisfied, then for any $\epsilon > 0$, there exists an $0 < M < \infty$ such that*

$$\mathbb{P}\left\{\int_{\|h\|\geq M, h_1\leq 0} N_\infty(h)dh \leq 2\epsilon\right\} > 1-2\epsilon.$$

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

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

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

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

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

For every given ω in the sample space, we divide the integration region into H_n^A and H_n^B, where $H_n^A = \{\|h\| \geq M, h_1 \leq 0, Q(\theta_\ell)h + \Delta(\theta_\ell) \geq 0\}$ and H_n^B be the complement. Note

$$\begin{aligned}\mathbb{P}\left\{\int_{H_n^A\cup H_n^B} N_\infty(h)dh \leq 2\epsilon\right\} &\geq \mathbb{P}\left\{\int_{H_n^A} N_\infty(h)dh \leq \epsilon, \int_{H_n^B} N_\infty(h)dh \leq \epsilon\right\}\\ &\geq 1 - \mathbb{P}\left\{\int_{H_n^A} N_\infty(h)dh > \epsilon\right\} - \mathbb{P}\left\{\int_{H_n^B} N_\infty(h)dh > \epsilon\right\}.\end{aligned}$$

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

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

The right hand side term converges to zero because $Q(\theta_\ell)$ has full column rank by Assumption 3.9 and $\Delta(\theta_\ell)$ is bounded in probability by Assumption 3.5.

Now consider region H_n^B. Let H_{nj}^B be the region of h such that the first element of $\Delta(\theta_\ell) + Q(\theta_\ell)h$ is smaller than zero, that is, $\Delta_j(\theta_\ell) + Q_j(\theta_\ell)h < 0$, where Q_j is the j-th row of Q. So $H_n^B = \cup_{j=1}^J H_{nj}^B$. We show the conclusion is hold for the integral over H_{n1}^B. The others are similar. Treating $\Delta_1(\theta_\ell)$ as the "b", and $Q_1(\theta_\ell)$ as the "a" at the beginning of this lemma, we have

$$\begin{aligned}&\int_{h\in H_{n1}^B} N_\infty(h)dh \leq \int_{h\in H_{n1}^B} \exp(-|Q_1(\theta_\ell)h|_+^2 + |\Delta_1(\theta_\ell)|^2)dh\\ &= \exp(|\Delta_1(\theta_\ell)|^2)\int_{h\in H_{n1}^B} \exp(-|Q_1(\theta_\ell)h|_+^2)dh \leq \exp(|\Delta_1(\theta_\ell)|^2)\int_{\{h:\|h\|\geq M, h_1\leq 0\}} \exp(-|Q_1(\theta_\ell)h|_+^2)dh.\end{aligned}$$

Since $|\Delta_1(\theta_\ell)|$ is bounded in probability by Assumption 3.5, to complete the proof of this lemma, we just need to show

$$\lim_{M\to+\infty}\int_{\{h:\|h\|\geq M,h_1\leq 0\}}\exp(-|Q_1(\theta_\ell)h|_+^2)dh=0.$$

This is true because by Assumption 3.3, there exists at least one element $Q_{1d^*}(\theta_\ell)$ in $Q_1(\theta_\ell)$ such that $Q_{1d^*}(\theta_\ell)<0$.

□

Lemma A.4. *Suppose that Assumptions 3.1, 3.2, 3.5, 3.8 and 3.9 are satisfied, then for any $\epsilon>0$, there exists M^* and M such that*

$$\lim_{n\to\infty}\mathbb{P}\left\{\int_{M<\|h\|\leq M^*\sqrt{n},h_1\leq 0}N_n(h)dh\leq\epsilon\right\}>1-\epsilon.$$

Proof. Let ϵ be arbitrarily given. Note that

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

where

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

For any $M>0$ and $\epsilon^*>0$, by Assumption 3.5, there exists M^* such that:

$$\limsup_{n\to\infty}\mathbb{P}\left\{\sup_{\{h:M\leq\|h\|\leq\sqrt{n}M^*,h_1\leq 0\}}\frac{\|\Delta_n(\theta_\ell+\frac{h}{\sqrt{n}})-\Delta_n(\theta_\ell)\|}{1+\|h\|}>\epsilon^*\right\}<\epsilon^*. \tag{13}$$

Hence with probability at least $1-\epsilon^*$, we have

$$N_n(h)\leq\exp\left(-\sum_j\left\{\left\|\Delta_{nj}(\theta_\ell)+Q_j(\theta_\ell)h\right\|_+^2+\frac{1}{2}\|Q_j(\theta_\ell)h\|\right\}\right),$$

Taking $\epsilon^*=\frac{\epsilon}{2}$, the results holds by following similar arguments 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 $M^*>0$,*

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

Proof. Recall that $H_{3n}=\{h:\|h\|\geq M^*\sqrt{n},h\in H_n\}$. 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}.

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

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

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

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, 3.8 and 3.9 are satisfied, then for any $\epsilon > 0$, there exist an $M > 0$ such that*

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

Proof. Follows from lemmas A.2 to A.5 and triangle inequality. □

Lemma A.7. *Suppose that Assumptions 3.1 to 3.5, 3.8 and 3.9 are satisfied, then*

$$\int_{H_n} N_n(h)dh \xrightarrow{d} \int N_\infty(h)dh.$$

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

Appendix B. Proofs in Section 3

B.1. Proof of Theorem 3.1.

Part 1: $\lim_{n\to\infty} \mathbb{P}(\hat{\theta}_{1\ell} < \theta_{1\ell} - \epsilon) = 0$.

Let $\epsilon > 0$ be arbitrary. We first show that $\sup_{\Theta/\Theta^\epsilon_I} n^r f_n(\theta) = o_p(1)$ for any $r > 0$. By Assumption 3.3, $\inf_{\theta \in \Theta/\Theta^\epsilon_I} \mathbb{E}m(W_1, \theta) > \delta_\epsilon \equiv \min\{C\epsilon, \delta\}$. Then with probability approaching one

$$\sup_{\Theta/\Theta^\epsilon_I} n^r f_n(\theta) = \frac{n^r}{D_n}\exp(-\inf_{\Theta/\Theta^\epsilon_I} n\|\bar{m}(\theta)\|^2_+)$$
$$= \frac{n^r}{D_n}\exp(-\inf_{\Theta/\Theta^\epsilon_I} n\|\bar{m}(\theta) - \mathbb{E}m_1(\theta) + \mathbb{E}m_1(\theta)\|^2_+) \leq \frac{n^{r+d/2}}{n^{d/2}D_n}\exp(-\frac{J\delta^2_\epsilon n}{4}) = o_p(1).$$

The last inequality holds because for every j, $\sup_{\theta\in\Theta}|\bar{m}_{(j)}(\theta)-\mathbb{E}m_{(j)}(W_1,\theta)|<\delta_\epsilon/2$ in probability. The last equality holds because $n^{d/2}D_n$ is bounded away from 0 in probability by Lemma A.1. Therefore,

$$n^r F_{1n}(\theta_{1\ell}-\epsilon)\le n^r\int_{\Theta/\Theta_I^\epsilon} f_n(\theta)d\theta\le\sup_{\Theta/\Theta_I^\epsilon} n^r f_n(\theta)\mu(\Theta/\Theta_I^\epsilon)=o_p(1).$$

On the other hand, by Assumption 3.6, $\hat{\tau}_\ell$ decreases to zero at a polynomial rate. This shows that $\lim_{n\to\infty}\mathbb{P}(\hat{\theta}_{1\ell}<\theta_{1\ell}-\epsilon)=0$.

Part 2: $\lim_{n\to\infty}\mathbb{P}(\hat{\theta}_{1\ell}>\theta_{1\ell}+\epsilon)\to 0$.

There are two cases, $\theta_{1\ell}=\theta_{1u}$ and $\theta_{1\ell}<\theta_{1u}$. Suppose first that $\theta_{1\ell}=\theta_{1u}$, then

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

Note that $F_{1n}(\theta_{1\ell}+\epsilon)\xrightarrow{p}1$ by the argument in part 1. It then follows that the probability on the right hand side coverages to 0. It remains to show same conclusion holds when $\theta_{1\ell}<\theta_{1u}$.

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

where $\mathscr{B}_n=\{\theta:\theta_{1\ell}\le\theta_1\le\theta_{1\ell}+\epsilon,\sqrt{n}\|\mathbb{E}m_1(\theta)\|_+\le\delta^*\}$ for some $\delta^*<\infty$. Note that by Lipschitz Assumption 3.4, there exists some $\delta^{**}>0$ such that $\mu(\mathscr{B}_n)\ge\frac{\delta^{**}\epsilon}{n^{d/2-1}}$.

It thus follows,

$$\begin{aligned}\mathbb{P}\left\{\hat{\tau}_\ell D_n\ge\int_{\mathscr{B}_n}\exp(-n\|\bar{m}(\theta)\|_+^2)d\theta\right\}\le\mathbb{P}\left\{\hat{\tau}_\ell D_n\ge n\mu(\mathscr{B}_n)\inf_{\theta\in\mathscr{B}_n}\exp(-n\|\bar{m}(\theta)\|_+^2\right\}\\ \le\mathbb{P}\left\{\hat{c}\ge\sqrt{n}\delta^{**}\epsilon\inf_{\theta\in\mathscr{B}_n}\exp(-n\|\bar{m}(\theta)\|_+^2\right\}\end{aligned}$$

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

$$\begin{aligned}\inf_{\theta\in\mathscr{B}_n}\exp\left(-n\|\bar{m}(\theta)\|_+^2\right)&=\exp\left(-\sup_{\theta\in\mathscr{B}_n}\|\sqrt{n}\bar{m}(\theta)\|_+^2\right)\\&\ge\exp\left(-\|\sup_{\theta\in\mathscr{B}_n}\Delta_n(\theta)+\sup_{\theta\in\mathscr{B}_n}\sqrt{n}\mathbb{E}m_1(\theta)\|_+^2\right)\\&\ge\exp\left(-\|\sup_{\theta\in\mathscr{B}_n}\Delta_n(\theta)\|_+^2-\|\sup_{\theta\in\mathscr{B}_n}\sqrt{n}\mathbb{E}m_1(\theta)\|_+^2\right).\quad(14)\end{aligned}$$

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 Equation (14) bounded away from zero with probability approaching one. $\square$

B.2. **Proof of Theorem 3.2.**

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

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

Following Lemma A.7,

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

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

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

$$
\begin{aligned}
\mathbb{P}\left\{ \hat{\theta}_{1\ell} \ge \theta_{1\ell} + \frac{K}{\sqrt{n}} \right\} &= \mathbb{P}\left\{ c + o_p(1) \ge n^{d/2} \int_{\{\theta: \theta_1 \le \theta_{1\ell} + \frac{K}{\sqrt{n}}\}} \exp(-n \| \bar{m}(\theta) \|_+^2) \right\} \\
&\le \mathbb{P}\left\{ c + o_p(1) \ge n^{d/2} \int_{\mathscr{B}_n} \exp(-n \| \bar{m}(\theta) \|_+^2) d\theta \right\}, \qquad (15)
\end{aligned}
$$

where

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

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

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

Note that there exists a $C_2 > 0$ such that $\mu(\mathscr{B}_n) = \frac{C_2 K \delta^*}{16 n^{d/2}}$. Therefore,

$$
\begin{aligned}
&\mathbb{P}\left\{ c + o_p(1) \ge n^{d/2} \int_{\mathscr{B}_n} \exp(-n \| \bar{m}(\theta) \|_+^2) d\theta \right\} \\
&\qquad \ge \mathbb{P}\left\{ c + o_p(1) \ge n^{d/2} \mu(\mathscr{B}_n) \inf_{\theta \in \mathscr{B}_n} \exp(-n \| \bar{m}(\theta) \|_+^2 \right\} \\
&\qquad = \mathbb{P}\left\{ c + o_p(1) \ge \frac{C_2 K \delta^*}{16} \inf_{\theta \in \mathscr{B}_n} \exp(-n \| \bar{m}(\theta) \|_+^2) \right\}.
\end{aligned}
$$

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 $\sup_{\theta \in \mathscr{B}_n} \|\sqrt{n}\mathbb{E}m_1(\theta)\|_+^2 \leq \frac{JC_1\delta^*}{4} < +\infty$ and by the same argument as in Equation (14).

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_{01} 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 Theorem holds. □

Appendix C. Proofs in Section 4

Lemmas C.1 to C.4 are needed for the proofs in this section.

Lemma C.1. *Suppose Assumptions 3.5 and 4.2 are satisfied. Define*

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

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

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

The convergence is because $\Delta_{nj} \xrightarrow{w} \Delta_j$ and $\beta_n/n \to 0$ by Assumption 4.2.

Now we consider the case in which there exist a δ such that $\mathbb{E}m_{(j)}(W_i,\theta) < -\delta$, then with probability approaching one the following inequality holds:

$$\gamma_{jn}(\theta) = \exp\left(-\frac{\beta_n}{n}\left|\frac{1}{\sqrt{n}}\sum_i m_{(j)}(W_i,\theta)\right|^2\right)$$

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

The term on the right hand side converges in probability to zero because $\beta_n \to \infty$ by Assumption 4.2. $\square$

Lemma C.2. *Suppose that the assumptions required by Theorem 4.2 are satisfied. Let $\hat{\theta}_\ell = (\hat{\theta}_{1\ell}, \hat{\theta}_2)$ and $\hat{\theta}_2$ be defined in Equation (4), then $\hat{\theta}_\ell \xrightarrow{p} \theta_\ell$.*

Proof. This follows from Theorem 3.1. $\square$

Lemma C.3. *Suppose that the assumptions required by Theorem 4.2 are satisfied, then $\|\hat{\gamma}(\hat{\theta}_\ell) - \gamma_n(\theta_\ell)\| = o_p(1)$.*

Proof. It is sufficient to show the conclusion holds for each element γ_{jn}. Suppose first that $\mathbb{E}m_{(j)}(W_i,(\theta_{1\ell},\theta_2)) = 0$. We 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 we 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 4.2, 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 3.8.

Now we 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. $\square$

Lemma C.4. *Suppose that the assumptions required by Theorem 4.2 are satisfied, then $\|\widehat{Q}(\hat{\theta}_\ell) - Q(\theta_\ell)\| = o_p(1)$.*

Proof. The uniform consistency of $\widehat{Q}$ follows from the Assumption 4.3 and the compactness of Θ, (see lemma 2.1, Newey and McFadden, 1994). $\square$

C.1. **proof to Lemma 4.1.** The results follows from Lemma A.7. The continuity of random variable ξ_ℓ and ξ_ℓ 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 .**

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,

$$\mathbb{P}\left\{c_\ell^{\mathrm{I}} \leq \xi_\ell\right\} = 1-\alpha, \quad \mathbb{P}\left\{c_u^{\mathrm{I}} \leq \xi_u\right\} = 1-\alpha. \tag{16}$$

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.

$$\begin{aligned}\mathcal{P}_n(\lambda) = \mathbb{P}\left\{\theta_\lambda \in \Theta_{\alpha n}^{\mathrm{I}}\right\} &= \mathbb{P}\left\{\int_{\underline{\theta}_1}^{\theta_\lambda} f_{1n}(\theta_1)d\theta_1 \geq \frac{c_\ell^{\mathrm{I}}}{n^{d/2}D_n}, \int_{\theta_\lambda}^{\bar{\theta}_1} f_{1n}(\theta_1)d\theta_1 \geq \frac{c_u^{\mathrm{I}}}{n^{d/2}D_n}\right\}\\ &= \mathbb{P}\left\{\xi_{\ell n} + n^{d/2}D_n\int_{\theta_{1\ell}}^{\theta_\lambda} f_{1n}(\theta_1)d\theta_1 \geq c_\ell^{\mathrm{I}}, n^{d/2}D_n\int_{\theta_\lambda}^{\theta_{1u}} f_{1n}(\theta_1)d\theta_1 + \xi_{un} \geq c_u^{\mathrm{I}}\right\}.\end{aligned}$$

Note that if $T > 0$, then for any $\lambda \in (0,1)$, $n^{d/2}D_n\int_{\theta_\lambda}^{\theta_{1u}} f_{1n}(\theta_1)d\theta_1$ or $n^{d/2}D_n\int_{\theta_{1\ell}}^{\theta_\lambda} f_{1n}(\theta_1)d\theta_1$ (or both) diverges to infinite as sample increases. 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 Equation (16). This shows that

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

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,

$$\mathbb{P}\left\{c_\ell^{\mathrm{I}} \leq \xi_\ell, c_u^{\mathrm{I}} \leq \xi_u\right\} = 1-\alpha. \tag{17}$$

Again,

$$\begin{aligned}\mathcal{P}_n(\lambda) = \mathbb{P}\left\{\theta_\lambda \in \Theta_{\alpha n}^{\mathrm{I}}\right\} &= \mathbb{P}\left\{\int_{\underline{\theta}_1}^{\theta_\lambda} f_{1n}(\theta_1)d\theta_1 \geq \frac{c_\ell^{\mathrm{I}}}{n^{d/2}D_n}, \int_{\theta_\lambda}^{\bar{\theta}_1} f_{1n}(\theta_1)d\theta_1 \geq \frac{c_u^{\mathrm{I}}}{n^{d/2}D_n}\right\}\\ &= \mathbb{P}\left\{\xi_{\ell n} + n^{d/2}D_n\int_{\theta_{1\ell}}^{\theta_\lambda} f_{1n}(\theta_1)d\theta_1 \geq c_\ell^{\mathrm{I}}, n^{d/2}D_n\int_{\theta_\lambda}^{\theta_{1u}} f_{1n}(\theta_1)d\theta_1 + \xi_{un} \geq c_u^{\mathrm{I}}\right\}.\end{aligned}$$

The validity is ensured by Equation (17) since $\theta_{1\ell} = \theta_\lambda = \theta_{1u}$. □

C.3. **Proof of Theorem 4.2.** Let $\hat{\xi}_\ell$ be simulated from Algorithm 2. $\hat{\xi}_\ell$ and ξ_ℓ have the same distribution in the limit is ensured by Lemmas C.3 and C.4 and by dominated convergence theorem (see theorem 16.4, Billingsley, 1995) and the fact that $\exp(-\|\cdot\|_+^2)$ is integrable (since $Q_1(\theta_\ell) < 0$), the convergence result follows. □

C.4. **Proof of Theorem 4.3.** When $\theta_{1\ell} < \theta_{1u}$, the conclusion follows immediately. Now consider the case $\theta_{1\ell} = \theta_{1u}$. Let θ_u^α and θ_u^α be the two end points of the confidence set.

$$
\begin{aligned}
\mathbb{P}(\{\theta_{1\ell}\} \subseteq \Theta_{\alpha n}^{\mathrm{II}}) = \mathbb{P}\left\{\theta_\ell^\alpha \leq \theta_{1\ell} \leq \theta_u^\alpha\right\} \\
&= \mathbb{P}\left\{\int_{\underline{\theta}_1}^{\theta_{1\ell}} f_{1n}(\theta_1)d\theta_1 \leq 1 - \frac{c^{\mathrm{II}}}{n^{d/2}D_n}, \int_{\underline{\theta}_1}^{\theta_{1\ell}} f_{1n}(\theta_1)d\theta_1 \geq \frac{c^{\mathrm{II}}}{n^{d/2}D_n}\right\} \\
&= \mathbb{P}\left\{\int_{\theta_{1\ell}}^{\bar{\theta}_1} f_{1n}(\theta_1)d\theta_1 \geq \frac{c^{\mathrm{II}}}{n^{d/2}D_n}, \int_{\underline{\theta}_1}^{\theta_{1\ell}} f_{1n}(\theta_1)d\theta_1 \geq \frac{c^{\mathrm{II}}}{n^{d/2}D_n}\right\} \\
&= \mathbb{P}\left\{\xi_\ell \geq c_\ell^{\mathrm{II}}, \xi_u \geq c_u^{\mathrm{II}}\right\} + o(1) = 1 - \alpha + o(1).
\end{aligned}
$$

□

APPENDIX D. PROOFS IN SECTION 6

D.1. **Proof of Lemma 6.1.**

Part 1. For the first statement in Lemma 6.1,

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

where

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

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

By Assumption 3.8, there exists some $C_1 > 0$ such that (note that $h_1 \leq 0$ here)

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

Hence

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

For every given K, A_n diverges in probability since the integrand is bounded away from zero with probability approaching one.

Part 2. Now we show the second statement of Lemma 6.1. I show first that

$$\lim_{K\to+\infty}\lim_{n\to\infty}\mathbb{P}\left\{\sqrt{n}(\theta_{1\ell}-F_{1n}^{-1}(\hat{\tau}_\ell^b))>K\right\}=0.$$

Note that

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

where

$$\begin{aligned}A_n&=\int_{-\infty}^{-K}\int_{\theta_2\in[\theta_{2\ell},\theta_{2u}]}D_nf_n(\theta_{1\ell}+h_1/\sqrt{n},\theta_2)d\theta_2dh_1,\\B_n&=\int_{-\infty}^{-K}\int_{\theta_2\ge\theta_{2u}}D_nf_n(\theta_{1\ell}+h_1/\sqrt{n},\theta_2)d\theta_2dh_1,\\C_n&=\int_{-\infty}^{-K}\int_{\theta_2\le\theta_{2\ell}}D_nf_n(\theta_{1\ell}+h_1/\sqrt{n},\theta_2)d\theta_2dh_1.\end{aligned}$$

B_n and C_n is $O_p(1/\sqrt{n})$ by the same argument in Appendix A. 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 Assumption 3.3, there exists at least one j^* and some $C_2>0$ such that for $h_1<0$

$$\inf_{\theta_2\in[\theta_{2\ell},\theta_{2u}]}\sqrt{n}\mathbb{E}m_{(j^*)}(W_1,(\theta_{1\ell}+\frac{h_1}{\sqrt{n}},\theta_2))\ge|C_2h_1|>0.$$

Then we know that,

$$\begin{aligned}A_n&=\int_{-\infty}^{-K}\int_{\theta_2\in[\theta_{2\ell},\theta_{2u}]}\exp(-\|\Delta_n(\theta_{1\ell}+\frac{h_1}{\sqrt{n}},\theta_2)+\sqrt{n}\mathbb{E}m_1(\theta_{1\ell}+\frac{h_1}{\sqrt{n}},\theta_2)\|_+^2)d\theta_2dh_1\\&\le(\theta_{2u}-\theta_{2\ell})\int_{-\infty}^{-K}\exp(-|\inf_{\theta_2\in[\theta_{2\ell},\theta_{2u}]}\Delta_{j^*n}(\theta_{1\ell},\theta_2)+|C_2h_1|+o_p(1)|_+^2)dh_1\equiv\overline{A}_n.\end{aligned}$$

So we have

$$\lim_{n\to\infty}\mathbb{P}(\hat{c}\le A_n)\le\lim_{n\to\infty}\mathbb{P}(\hat{c}\le\overline{A}_n)$$

The right hand side converges to zero as $K\to+\infty$ because $\hat{c}\xrightarrow{p}c>0$ and $\inf_{\theta_2\in[\theta_{2\ell},\theta_{2u}]}\Delta_n(\theta_{1\ell},\theta_2)$ is $O_p(1)$.

It remains to show

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

The case in which $\theta_{1u} = \theta_{1\ell}$ can be shown with similar argument as above. Suppose $\theta_{1u} > \theta_{1\ell}$ for now.

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

where $\mathscr{B}_n = \{\theta : \theta_{1\ell} \leq \theta_1 \leq \theta_{1\ell} + \frac{K}{\sqrt{n}}, d(\theta, \Theta_I) \leq \delta^*/\sqrt{n}\}$. It is not difficult to verify that $\mu(\mathscr{B}_n) \geq \frac{K\delta^{**}}{\sqrt{n}}$ for some $\delta^{**} > 0$. It thus follows,

$$\begin{aligned}
&\mathbb{P}\left\{c_\ell + o_p(1) \geq \sqrt{n}\int_{\mathscr{B}_n} \exp(-n\|\bar{m}(\theta)\|_+^2) d\theta\right\} \\
&\leq \mathbb{P}\left\{c_\ell + o_p(1) \geq \sqrt{n}\mu(\mathscr{B}_n) \inf_{\theta\in\mathscr{B}_n} \exp(-n\|\bar{m}(\theta)\|_+^2\right\} \\
&\leq \mathbb{P}\left\{c_\ell + o_p(1) \geq K\delta^{**} \inf_{\theta\in\mathscr{B}_n} \exp(-n\|\bar{m}(\theta)\|_+^2\right\}.
\end{aligned}$$

The limit probability converges to zero as K increases because $\inf_{\theta\in\mathscr{B}_n} \exp(-n\|\bar{m}(\theta)\|_+^2$ is $O_p(1)$ by similar reason as in Equation (14). □

D.2. **Sketch of Proof of Proposition 6.1.** Equation (11) or (12) can be written as

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

where $\hat{c}_\ell \xrightarrow{p} c > 0$.

It is sufficient to show that with probability approaching one, there exists an h_1^* such that

$$\frac{\mathbb{N}_n}{\mathbb{D}_n} \equiv \frac{\int_{-\infty}^{h_1^*}\int_{\theta_2\in\Theta_2} \exp(-n\|\bar{m}(\theta_{1\ell}+\frac{h_1}{\sqrt{n}},\theta_2)\|_+^2) d\theta_2 dh_1}{\int_{\theta_2\in\Theta_2} \exp(-n\|\bar{m}(\theta_{1\ell}+\frac{h_1^*}{\sqrt{n}},\theta_2)\|_+^2) d\theta_2} = c.$$

Note that for every $\omega \in \Omega$, the left hand side is a continuous function of h_1^*. We just need to show that for every ω, the left hand side converges to zero as h_1^* diverges to $-\infty$; and diverges when h_1^* goes to $+\infty$.

Part 1: $h_1^* \to \infty$**.** We first show that for any $c > 0$, there exists h_1^* such that $\mathbb{P}(\mathbb{N}_n/\mathbb{D}_n > c) \to 1$. For some $\kappa > 0$, let $\kappa_n = \kappa/\sqrt{n}$. Let $\Theta_I^{\kappa_n}$ be the κ_n–expansion of Θ_I, and define

$$\Theta_2^{\kappa_n} = \left\{\theta_2 : \exists h_1 \text{ such that } (\theta_{1\ell} + h_1/\sqrt{n}, \theta_2) \in \Theta_I^{\kappa_n}\right\}.$$

Then it follows in a similar argument as in Lemma A.5 that

$$\frac{\mathbb{N}_n}{\mathbb{D}_n} + o_p(1) = \frac{\mathbb{N}_n^*}{\mathbb{D}_n^*} \equiv \frac{\int_{-\infty}^{h_1^*} \int_{\theta_2 \in \Theta_2^{\kappa_n}} \exp(-n\|\bar{m}(\theta_{1\ell} + \frac{h_1}{\sqrt{n}}, \theta_2)\|_+^2) d\theta_2 dh_1}{\int_{\theta_2 \in \Theta_2^{\kappa_n}} \exp(-n\|\bar{m}(\theta_{1\ell} + \frac{h_1^*}{\sqrt{n}}, \theta_2)\|_+^2) d\theta_2}.$$

In the meantime, let

$$\mathscr{B}_n = \left\{(h_1, \theta_2) : 0 \le h_1 \le h_1^*, \text{ and } (\theta_{1\ell} + h_1/\sqrt{n}, \theta_2) \in \Theta_I^{\kappa_n}\right\}.$$

Then the Lebesgue measure $\mu(\mathscr{B}_n) \ge k h_1^* \mu(\Theta_2^{\kappa_n})$ for some $k > 0$.

$$\begin{aligned}
&\mathbb{P}(\mathbb{N}_n/\mathbb{D}_n > c) = \mathbb{P}(\mathbb{N}_n^*/\mathbb{D}_n^* > c) + o(1) \\
&\ge \mathbb{P}\left(\frac{\mu(\mathscr{B}_n) \inf_{\theta \in \Theta_I^{\kappa_n}} \exp\left(-n\|\bar{m}(\theta)\|_+^2\right)}{\mu(\Theta_2^{\kappa_n}) \sup_{\theta \in \Theta_I^{\kappa_n}} \exp\left(-n\|\bar{m}(\theta)\|_+^2\right)} > c\right) \ge \mathbb{P}\left(\frac{k \inf_{\theta \in \Theta_I^{\kappa_n}} \exp\left(-n\|\bar{m}(\theta)\|_+^2\right)}{\sup_{\theta \in \Theta_I^{\kappa_n}} \exp\left(-n\|\bar{m}(\theta\|_+^2\right)} > \frac{c}{h_1^*}\right).
\end{aligned}$$

Following similar arguments in Equation (14), $\sup_{\theta \in \Theta_I^{\kappa_n}} \exp\left(-n\|\bar{m}(\theta)\|_+^2\right)$ and $\inf_{\theta \in \Theta_I^{\kappa_n}} \exp\left(-n\|\bar{m}(\theta)\|_+^2\right)$ are $O_p(1)$. Then the results follows since the probability can be made arbitrarily small by taking h_1^* large.

Part 2: $h_1^* \to -\infty$. Similarly, we only need to show that for any $c > 0$, there exists a $h_1^* < 0$ such that $\mathbb{P}(\mathbb{N}_n^*/\mathbb{D}_n^* < c) \to 1$.

Notice that we can separate the integration region for θ_2 into Θ_{I2} and $\Theta_2^{\kappa_n}/\Theta_{I2}$. It then follows

$$\begin{aligned}
\frac{\mathbb{N}_n^*}{\mathbb{D}_n^*} &= \frac{\int_{h_1} \int_{\Theta_{I2}} \cdots d\theta_2 dh_1 + \int_{h_1} \int_{\Theta_2^{\kappa_n}/\Theta_{I2}} \cdots d\theta_2 dh_1}{\int_{\Theta_{I2}} \cdots d\theta_2 + \int_{\Theta_2^{\kappa_n}/\Theta_{I2}} \cdots d\theta_2} \\
&\le \frac{\int_{h_1} \int_{\Theta_{I2}} \cdots d\theta_2 dh_1}{\int_{\Theta_{I2}} \cdots d\theta_2} + \frac{\int_{h_1} \int_{\Theta_2^{\kappa_n}/\Theta_{I2}} \cdots d\theta_2 dh_1}{\int_{\Theta_2^{\kappa_n}/\Theta_{I2}} \cdots d\theta_2} \equiv \frac{\mathbb{N}_{1n}^*}{\mathbb{D}_{1n}^*} + \frac{\mathbb{N}_{2n}^*}{\mathbb{D}_{2n}^*},
\end{aligned}$$

Consider $\mathbb{N}_{1n}^*/\mathbb{D}_{1n}^*$ first. For a given $\theta_2 \in \Theta_{I2}(\theta_{1\ell})$,

$$\sqrt{n}\bar{m}(\theta_1 + \frac{h_1}{\sqrt{n}}, \theta_2) = \sqrt{n}\bar{m}(\theta_{1\ell}, \theta_2) + \bar{Q}_1(\theta_{1\ell}, \theta_2) h_1 + R_n(h_1),$$

where $\bar{Q}_1$ is the first order derivative of $\bar{m}$ with respect to first element and R_n is the remainder term satisfying for any $\delta > 0$

$$\sup_{|h_1| < \delta\sqrt{n}} \frac{R_n(h_1)}{|h_1|} \xrightarrow{p} 0.$$

Let $\Delta_n(\theta_{1\ell}, \theta_2) = \sqrt{n}(\bar{m}(\theta_{1\ell}, \theta_2) - \mathbb{E}m(\theta_{1\ell}, \theta_2))$, and $Q_1 = \mathbb{E}[\bar{Q}_1]$, then

$$\frac{\mathbb{N}_{1n}^*}{\mathbb{D}_{1n}^*} = \frac{\int_{-\infty}^{h_1^*} \int_{\Theta_{I2}} \exp(-\|\Delta_n(\theta_{1\ell}, \theta_2) + Q_1 h_1\|_+^2) d\theta_2 dh_1}{\int_{\Theta_{I2}} \exp(-\|\Delta_n(\theta_{1\ell}, \theta_2) + Q_1 h_1^*\|_+^2) d\theta_2} + o_p(1)$$

$$= \frac{\int_{\Theta_{I2}} \int_{\Delta_n + Q_1 h_1^*}^{+\infty} \exp(-\|t_1\|_+^2) dt_1 d\theta_2}{\int_{\Theta_{I2}} \exp(-\|\Delta_n(\theta_{1\ell}, \theta_2) + Q_1 h_1^*\|_+^2) d\theta_2} + o_p(1)$$

With probability approaching one, the right hand side can be made arbitrarily small by choosing $h_1^* \to -\infty$ because $Q_1 h_1^* \to +\infty$ as $h_1^* \to -\infty$.

Now we consider $\mathbb{N}_{2n}^* / \mathbb{D}_{2n}^*$. Notice that the integrand is a continuous function of θ_2, hence by mean value theorem, we have

$$\frac{\mathbb{N}_{2n}^*}{\mathbb{D}_{2n}^*} = \frac{\int_{-\infty}^{h_1^*} \exp(-n\|\bar{m}(\theta_{1\ell} + h_1/\sqrt{n}, \tilde{\theta}_2(h_1))\|_+^2) dh_1}{\exp(-n\|\bar{m}(\theta_{1\ell} + h_1^*/\sqrt{n}, \tilde{\theta}_2(h_1^*))\|_+^2)},$$

where the mean value $\tilde{\theta}_2(\cdot)$ takes value in $\Theta_2^{\kappa_n}/\Theta_{I2}$. Taylor expanding $\bar{m}$ on both the direction of θ_1 and θ_2 and by a similar argument as above, with probability approaching one, $\mathbb{N}_{2n}^*/\mathbb{D}_{2n}^*$ can be made arbitrarily small as $h_1^* \to -\infty$. □

Appendix E. Algorithm and Tables

We summarize the procedure of estimation and inference in Algorithm 2. As before, we focus on the $\theta_{1\ell}$; the procedure for θ_{1u} is similar.

Algorithm 2. Construct confidence set by simulation. Let $S = 10,000$, $Q = 1,000$ and $B = 500,000$.

(1) Estimation.

(a) Compute D_n by Monte Carlo integration:

$$D_n = \frac{1}{S} \sum_{s=1}^{S} \exp\left(nL_n(\theta_s)\right),$$

where $\{\theta_s\}_s^S$ are i.i.d. draws from uniform distribution over $[0,1] \times [0,1]$.

(b) Choose one initial value $\theta_{(0)} \in \Theta$. One can choose $\theta_{(0)}$ such that $\bar{m}_{(j)}(\theta_{(0)}) = 0$ for some j.

(c) Construct an MCMC chain $\{\theta_{(b)}\}_{b=0}^B$ based on f_n using Algorithm 1 (For more options, see Robert and Casella, 2004, Chapter 7). Discard the first half as burn-in period.

(d) Obtain the first component $\theta_{1(b)}$ of $\theta_{(b)}$ for all $B/2 + 1 \le b \le B$. $\{\theta_{1(b)}\}_{b=B/2+1}^B$ are used as $B/2$ random draws from the marginal quasi-posterior f_{1n}.

(e) Sort $\{\theta_{1(b)}\}_{b=B/2}^B$ and compute $\hat{\theta}_{1\ell}$ and $\hat{\theta}_{1u}$ by taking two empirical quantiles of the chain, that is,

$$\hat{\theta}_{1\ell} = F_{1n}^{-1}\left\{\frac{1}{nD_n}\right\}, \quad \hat{\theta}_{1u} = F_{1n}^{-1}\left\{1 - \frac{1}{nD_n}\right\}.$$

(2) Draw another MCMC chain $\{\theta_{2(b)}\}_{b=1}^B$ for θ_2 from the conditional density $f_n(\theta_2|\theta_1 = \hat{\theta}_{1\ell})$ following the same procedure as in Algorithm 1.

(3) Compute $\hat{\theta}_2 = (2/B)\sum_{b=B/2+1}^{B} \theta_{2(b)}$.

(4) Compute $\hat{\gamma}(\hat{\theta}_{1\ell})$.

(5) Estimate $\widehat{Q}_1(\hat{\theta}_{1\ell}) = \frac{\partial \bar{m}(\theta)}{\partial \theta_1}|_{\theta=\hat{\theta}_\ell}$.

(6) Let $\hat{V}$ be a J by J matrix

$$\frac{1}{n}\sum_{i=1}^{n}\left[m_i(\hat{\theta}_{1\ell}) - \bar{m}(\hat{\theta}_{1\ell})\right]' \times \left[m_i(\hat{\theta}_{1\ell}) - \bar{m}(\hat{\theta}_{1\ell})\right].$$

(7) Independently draw $\{(h_{1,s}, h_{2,s})\}_{s=1}^{S}$ from bivariate normal distribution with identity covariance matrix.[13]

(8) For $q = 1, \cdots, Q$, independently across q,

   (a) draw a J-vector mean zero normal random variable $\mathbf{w}_q$ with (j, j') element in the variance matrix equals to

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

   (b) compute

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

(9) $\{\hat{\xi}_{\ell,q}\}_{q=1}^{Q}$ are used as Q independent random draws to approximate the distribution of ξ_ℓ.

(10) Obtain the simulated distribution for $\hat{\xi}_u$ in a similar way (step 2 to step 9).

(11) Let $\hat{c}_\ell^{\mathrm{I}}$ and $\hat{c}_u^{\mathrm{I}}$ be computed using the maximization problem in Equation (3). If there are more than one solutions, pick an arbitrary one.

(12) Construct the confidence interval $\widehat{\Theta}_\alpha^{\mathrm{I}} = [F_{1n}^{-1}(\hat{c}_\ell^{\mathrm{I}}/(n^{d/2}D_n)), F_{1n}^{-1}(1 - \hat{c}_u^{\mathrm{I}}/(n^{d/2}D_n))]$.

(13) Compute the median $\hat{\xi}_{\ell,[0.5]}$ of $\hat{\xi}_\ell$, and the median $\hat{\xi}_{u,[0.5]}$ of $\hat{\xi}_u$, respectively; then obtain the updated estimators:

$$\tilde{\theta}_{1\ell} = F_{1n}^{-1}\left\{\frac{\hat{\xi}_{\ell,[0.5]}}{nD_n}\right\}, \quad \tilde{\theta}_{1u} = F_{1n}^{-1}\left\{1 - \frac{\hat{\xi}_{u,[0.5]}}{nD_n}\right\}.$$

□

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

TABLE 2. MSE under different choices of c (for Figure 4)

Sample Size	$\hat{\theta}_{1\ell}$ 200	400	800	$\hat{\theta}_{1u}$ 200	400	800
$c = 0.5$	0.00734	0.00393	0.00205	0.01771	0.00794	0.00393
$c = 1.0$	0.00377	0.00188	0.00105	0.01081	0.00484	0.00227
$c = 1.5$	0.00189	0.00105	0.00055	0.00697	0.00288	0.00154
$c = 2.0$	0.00119	0.00057	0.00032	0.00458	0.00219	0.00106
$c = 2.5$	0.00100	0.00050	0.00025	0.00374	0.00183	0.00071
$c = 3.0$	0.00115	0.00060	0.00030	0.00274	0.00134	0.00061
$c = 3.5$	0.00179	0.00089	0.00041	0.00261	0.00107	0.00052
$c = 4.0$	0.00274	0.00128	0.00064	0.00267	0.00113	0.00054
$c = 4.5$	0.00372	0.00187	0.00100	0.00285	0.00134	0.00060
$c = 5.0$	0.00492	0.00250	0.00130	0.00323	0.00151	0.00077
$c = 5.5$	0.00645	0.00341	0.00171	0.00418	0.00184	0.00092
$c = 6.0$	0.00870	0.00416	0.00226	0.00482	0.00251	0.00115
$\hat{\xi}_{\ell,[0.5]}$ or $\hat{\xi}_{u,[0.5]}$	0.00124	0.00063	0.00028	0.00228	0.00099	0.00050

Based on $1,000$ replications.

TABLE 3. Standard deviation under different choices of c (for Figure 4)

Sample Size	$\hat{\theta}_{1\ell}$ 200	400	800	$\hat{\theta}_{1u}$ 200	400	800
$c = 0.5$	0.03266	0.02399	0.01693	0.04839	0.03226	0.02225
$c = 1.0$	0.03093	0.02306	0.01655	0.04686	0.03357	0.02264
$c = 1.5$	0.03117	0.02260	0.01684	0.04675	0.03230	0.02329
$c = 2.0$	0.03072	0.02269	0.01598	0.04939	0.03361	0.02224
$c = 2.5$	0.03052	0.02318	0.01609	0.04997	0.03316	0.02291
$c = 3.0$	0.03046	0.02245	0.01629	0.04903	0.03441	0.02288
$c = 3.5$	0.02937	0.02156	0.01570	0.04886	0.03313	0.02258
$c = 4.0$	0.02947	0.02080	0.01538	0.05311	0.03387	0.02287
$c = 4.5$	0.02873	0.02084	0.01558	0.05055	0.03377	0.02374
$c = 5.0$	0.02802	0.02016	0.01538	0.05304	0.03375	0.02395
$c = 5.5$	0.02637	0.02064	0.01539	0.05316	0.03496	0.02443
$c = 6.0$	0.02644	0.02122	0.01548	0.05521	0.03602	0.02386
$\hat{\xi}_{\ell,[0.5]}$ or $\hat{\xi}_{u,[0.5]}$	0.03520	0.02488	0.01672	0.04601	0.03080	0.02217

Based on $1,000$ replications.