MCMC CONFIDENCE SETS FOR IDENTIFIED SETS

By

Xiaohong Chen, Timothy M. Christensen, Keith O'Hara, and Elie Tamer

May 2016
Revised July 2016

COWLES FOUNDATION DISCUSSION PAPER NO. 2037R

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

http://cowles.yale.edu/

MCMC Confidence Sets for Identified Sets*

Xiaohong Chen[†] Timothy M. Christensen[‡] Keith O'Hara[§] Elie Tamer[¶]

First draft: August 2015; Revised July 5, 2016

Abstract

In complicated/nonlinear parametric models, it is generally hard to determine whether the model parameters are (globally) point identified. We provide computationally attractive procedures to construct confidence sets (CSs) for identified sets of parameters in econometric models defined through a likelihood or a vector of moments. The CSs for the identified set or for a function of the identified set (such as a subvector) are based on inverting an optimal sample criterion (such as likelihood or continuously updated GMM), where the cutoff values are computed via Monte Carlo simulations directly from a quasi posterior distribution of the criterion. We establish new Bernstein-von Mises type theorems for the posterior distributions of the quasi-likelihood ratio (QLR) and profile QLR statistics in partially identified models, allowing for singularities. These results imply that the Monte Carlo criterion-based CSs have correct frequentist coverage for the identified set as the sample size increases, and that they coincide with Bayesian credible sets based on inverting a LR statistic for point-identified likelihood models. We also show that our Monte Carlo optimal criterion-based CSs are uniformly valid over a class of data generating processes that include both partially- and point-identified models. We demonstrate good finite sample coverage properties of our proposed methods in four non-trivial simulation experiments: missing data, entry game with correlated payoff shocks, Euler equation and finite mixture models. Finally, our proposed procedures are applied in two empirical examples.

*We are grateful to S. Bonhomme and L. P. Hansen for insightful suggestions. We also thank T. Cogley, K. Evdokimov, H. Hong, B. Honoré, M. Keane, K. Menzel, M. Kolesár, U. Müller, J. Montiel Olea, T. Sargent, F. Schorfheide, C. Sims and participants at the ESWC meetings in Montreal, the September 2015 "Big Data Big Methods" Cambridge-INET conference, and seminar audiences at Chicago, Princeton, Yale, Vanderbilt, Kansas, Cornell, Stony Brook, OSU, Wisconsin, NYU and Stanford for useful comments.

†Cowles Foundation for Research in Economics, Yale University. E-mail address: xiaohong.chen@yale.edu

‡Department of Economics, New York University. E-mail address: timothy.christensen@nyu.edu

§Department of Economics, New York University. E-mail address: koh215@nyu.edu

¶Department of Economics, Harvard University. E-mail address: elietamer@fas.harvard.edu

1 Introduction

In complicated (nonlinear) structural models, it is typically difficult to verify that the model parameters are (globally) point identified. This is especially important when one is interested in conducting a sensitivity analysis to examine the impact of various assumptions on parameter estimates where weaker assumptions may lead to loss of point identification. This motivation naturally calls for computationally simple and theoretically attractive inference methods that are valid whether or not the parameter of interest is identified. For example, if we are interested in estimating parameters characterizing the profits of firms using entry data, an important question is whether the estimates obtained from standard methods such as maximum likelihood are sensitive to the functional forms and/or distributional assumptions used to obtain these estimates. Relaxing some of these suspect assumptions (such as replacing the normality assumption on the unobserved fixed costs distribution with a mixture of normals, say) calls into question whether these profit parameters remain (globally) point identified. Our aim is to contribute to this sensitivity literature in parametric models allowing for partial identification.

To that extent, we provide computationally attractive and asymptotically valid confidence set (CS) constructions for the identified set (IdS) or functions of the IdS in models defined through a likelihood or a vector of moments.[1] In particular, we propose Monte Carlo (MC) criterion-based CS for the IdS of the entire structural parameter and for functions of the structural parameter (such as subvectors). The proposed procedures do not require the choice of extra tuning (smoothing) parameters beyond the ability to simulate a draw from the quasi posterior of an optimally weighted sample criterion. As a sensitivity check in an empirical study, a researcher could report a conventional CS based on inverting a t or Wald statistic that is valid under point identification only, and our new MC criterion-based CSs that are robust to failure of point identification.

Following Chernozhukov, Hong, and Tamer (2007) (CHT) and the subsequent literature on the construction of CSs for the IdS, our inference approach is also criterion function based and includes likelihood and generalized method of moment (GMM) models.[2] That is, *contour sets* of the sample criterion function are used as CSs for the IdS. However, unlike CHT and Romano and Shaikh (2010) who use subsampling to estimate critical values, we instead use *the quantile of the simulated sample criterion chain* from a (quasi) posterior to build a CS that has (frequentist) prescribed coverage probability. This posterior combines an optimally weighted sample criterion

[1] Following the literature, the identified set (IdS) Θ_I is the argmax of the population criterion in the parameter space Θ. A model is point identified if the IdS is a singleton $\{\theta_0\}$, and partially identified if the IdS is strictly larger than a singleton but strictly smaller than the whole parameter space.

[2] Unconditional moment inequality based models are a special case of moment (equality) based models in that one can add a nuisance parameter to transform a (unconditional) moment inequality into an equality. See Subsection 4.2.1 for details.

function (or a transformation of it) with a given prior (over the parameter space Θ). We draw a MC sample (chain) $\{\theta^1, ..., \theta^B\}$ from the posterior, compute the quantile of the optimally weighted sample criterion evaluated at these draws at a pre-specified level, and then define our CS for the IdS Θ_I as the contour set at the pre-specified level. The computational complexity of our proposed method for covering the IdS Θ_I of the entire structural parameter is just as hard as the problem of taking draws from a (quasi) posterior. The latter problem is a well researched and understood area in the literature on Monte Carlo (MC) methods in Bayesian posterior computations (see, e.g., Liu (2004), Robert and Casella (2004) and the references therein). There are many different MC samplers one could use for fast simulation from a (quasi) posterior,[3] and no optimization is involved for our CS for the IdS Θ_I. For functions of the IdS (such as a subvector), an added computation step is needed at the simulation draws to obtain level sets that lead to the exact asymptotic coverage of this function of the IdS.[4] We demonstrate the computational feasibility and the good finite sample coverage properties of our proposed methods in four non-trivial simulation experiments: missing data, entry game with correlated shocks, Euler equation and finite mixture models.

Theoretically, the validity of our MC CS construction requires the analysis of the large-sample behavior of the quasi posterior distribution of the likelihood ratio (LR) or optimal GMM criterion under lack of point identification. We establish new Bernstein-von Mises type theorems for quasi-likelihood-ratio (QLR) and profile QLR statistics in partially identified models allowing for singularities. Under regularity conditions, these theorems state that, even for partially identified models, the posterior distributions of the (not-necessarily optimally weighted) QLR and the profile QLR statistics coincide with those of the optimally weighted QLR and the profile QLR statistics as sample size increases to infinity. More precisely, the main text presents some regularity conditions under which the limiting distributions of the posterior QLR and of the maximized (over the IdS Θ_I) sample QLR statistics coincide with a chi-square distribution with an unknown degree of freedom, while Appendix E presents more general regularity conditions under which these limiting distributions coincide with a gamma distribution with an unknown shape parameter and scale parameter of 2. These results allow us to consistently estimate quantiles of the optimally weighted criterion by the quantiles of the MC criterion chains (from the posterior), which are sufficient to construct CSs for the IdS. In addition, we show in Appendix B that our MC CSs are uniformly valid over DGPs that include both partially- and point-identified models. We also present results on local power in Appendix D.

Our MC CSs are equivalent to Bayesian credible sets based on inverting a LR statistic in point-

[3]While many MC samplers could be used, in this paper we often use the terms "Markov Chain Monte Carlo" (MCMC) and "chains" for pedagogical convenience.

[4]We also provide a computationally extremely simple but slightly conservative CS for the identified set of a scalar subvector of a class of partially identified models, which is an optimally weighted profile QLR contour set with its cutoff being the quantile of a chi-square distribution with one degree of freedom.

identified likelihood models, which, with flat priors, are also the Bayesian highest posterior density (HPD) credible regions. More generally, for point-identified likelihood or moment-based models our MC CSs asymptotically coincide with frequentist CSs based on inverting an optimally weighted QLR (or a profile QLR) statistic, even when the true structural parameter may not be root-n consistently, asymptotically normally estimable.[5] Note that our MC CSs are *different* from those of Chernozhukov and Hong (2003) (CH). For point-identified root-n asymptotically normally estimable parameters in likelihood and optimally weighted GMM problems, CH takes the upper and lower $100(1-\alpha)/2$ percentiles of the Markov Chain Monte Carlo (MCMC) parameter chain $\{\theta_j^1,\ldots,\theta_j^B\}$ to construct a CS for a scalar parameter θ_j for $j=1,...,\dim(\theta)$. For such problems, CH's MCMC CS asymptotically coincides with a frequentist CS based on inverting a t statistic. Therefore, our CS and CH's CS are asymptotically first-order equivalent for point-identified scalar parameters that are root-n asymptotically normally estimable, but they differ otherwise. In particular, our methods (which take quantiles of the criterion chain) remain valid for partially-identified models whereas percentile MCMC CSs (which takes quantiles of the parameter chain) undercover. Intuitively this is because the parameter chain fails to stabilize under partial identification while the criterion chain still converges.[6] Indeed, simulation studies demonstrate that our MC CSs have good finite sample coverage properties uniformly over partially-identified or point-identified models.

Several papers have recently proposed Bayesian (or pseudo Bayesian) methods for constructing CSs for IdS Θ_I that have correct frequentist coverage properties. See the 2009 NBER working paper version of Moon and Schorfheide (2012), Kitagawa (2012), Kline and Tamer (2015), Liao and Simoni (2015) and the references therein.[7,8] Theoretically, all these papers consider *separable* models and use various renderings of a similar intuition. First, there exists a finite-dimensional reduced-form parameter, say ϕ, that is (globally) point-identified and root-n consistently and asymptotically normal estimable from the data, and is linked to the structural parameter of interest θ via a *known* (finite-dimensional) global mapping. Second, a prior is placed on the reduced-form parameter ϕ, and third, a classical Bernstein-von Mises theorem stating the asymptotic normality of the posterior distribution for ϕ is assumed to hold. Finally, the known global

[5]In this case an optimally weighted QLR may not be asymptotically chi-square distributed but could still be asymptotically gamma distributed. See Fan, Hung, and Wong (2000) for results on LR statistic in point-identified likelihood models and our Appendix E for an extension to an optimally weighted QLR statistic.

[6]Alternatively, the model structural parameter θ could be point- or partially- identified while the maximal population criterion is always point-identified.

[7]Norets and Tang (2014) propose a method similar to that in the working paper version of Moon and Schorfheide (2012) for constructing CSs for Θ_I in the context of a dynamic binary choice model but do not study formally the frequentist properties of their procedure.

[8]Also, Kitagawa (2012) establishes "bounds" on the posterior for the structural due to a collection of priors. The prior is specified only over the "sufficient parameter." Intuitively, the "sufficient parameter" is a point-identified re-parametrization of the likelihood. He then establishes that this "robust Bayes" approach could deliver a credible set that has correct frequentist coverage under some cases.

mapping between the reduced-form and the structural parameters is inverted, which, by step 3, guarantees correct coverage for the IdS Θ_I in large samples. Broadly, all these papers focus on a class of separable models with known specific structures that map some (globally) identified regular reduced-form parameters to the structural parameters.

Our MC approach to set inference does not require any kind of separability, nor does it require the existence of root-n consistently asymptotically normally estimable reduced-form parameter ϕ of a known finite dimension. Rather, we show that for general (separable or non-separable) partially identified likelihood or GMM models, a *local reduced-form reparameterization* exists under regularity conditions. We then use this reparametrization to show that the posterior distribution of the optimally weighted QLR statistic has a frequentist interpretation when the sample size is large, which enables the use of MC samplers to estimate consistently the relevant quantile of this statistic. Importantly, our local reparametrization is a proof device only, and so a practitioner does not need to know this reparametrization or its dimension explicitly for the actual construction of our proposed MC CSs for Θ_I. Our more general Bernstein-von Mises type theorem for the posterior of QLR in Appendix E even permits the support of the data to depend on the local reduced-form reparametrization (and hence makes it unlikely to estimate the local reduced-form parameter root-n consistently and asymptotically normal). In particular, while most of the existing Bayesian works on set inference place a prior on the reduced-form parameters,[9] we place a prior on the structural parameter $\theta \in \Theta$ only, and characterize the large-sample behaviors of the posterior distributions of the QLR and the profile QLR statistics.

There are several published works on consistent CS constructions for IdSs from the frequentist perspective. See, for example, CHT and Romano and Shaikh (2010) where subsampling based methods are used for general partially identified models, Bugni (2010) and Armstrong (2014) where bootstrap methods are used for moment inequality models, and Beresteanu and Molinari (2008) where random set methods are used when IdS is strictly convex. Also, for inference on functions of the IdS (such as subvectors), both subsampling based papers of CHT and Romano and Shaikh (2010) deliver valid tests with a judicious choice of the subsample size for a profile version of a criterion function. The subsampling based CS construction allows for general criterion functions and general partially identified models, but is computationally demanding and sensitive to choice of subsample size in realistic empirical structural models.[10] Our proposed methods are computationally attractive and typically have asymptotically correct coverage, but

[9] A few Bayesian approaches place a joint prior on both the reduced-form and the structural parameters.

[10] There is a large literature on frequentist approach for *inference on the true parameter* in an IdS (e.g., Imbens and Manski (2004), Rosen (2008), Andrews and Guggenberger (2009), Stoye (2009), Andrews and Soares (2010), Andrews and Barwick (2012), Canay (2010), Romano, Shaikh, and Wolf (2014), Bugni, Canay, and Shi (2016) and Kaido, Molinari, and Stoye (2016) among many others), which generally requires working with discontinuous-in-parameters asymptotic (repeated sampling) approximations to test statistics. These existing frequentist methods based on a guess and verify approach are difficult to implement in realistic empirical models.

require an optimally weighted criterion.

We study two important examples in detail. The first example considers a generic model of missing data. This model is important since its analysis illustrates the conceptual difficulties that arise in a simple and transparent setup. In particular, both numerically and theoretically, we study the behaviors of our CSs when this model is close to point identified, when it is point identified and when it is partially identified. The second model we study is a complete information entry game with correlated payoff shocks. Both these models have been studied in the existing literature as leading examples of partially-identified moment *inequality* models. We instead use them as examples of likelihood and moment equality models. Simulations demonstrate that our proposed CSs have good coverage in small samples. Appendix A contains simulation studies of two additional examples: a weakly identified Euler equation model of Hansen, Heaton, and Yaron (1996) and Stock and Wright (2000), and a mixture of normals example. Finally, our construction is applied to two empirical examples. In the first model based on trade data, we estimate more than 40 parameters using our MC methods, while in the the second example based on airline entry data, we estimate confidence sets for 17 parameters. In both cases, the our procedure show reasonable results.

The rest of the paper is organized as follows. Section 2 describes our new procedures, and demonstrates their good finite sample performance using missing data and entry game examples. Section 3 establishes new Bernstein-von Mises type theorems for QLR and profile QLR statistics in partially-identified models without or with singularities. Section 4 provides some sufficient conditions in several class of models. Section 5 presents an empirical trade application and an airline entry game illustration. Section 6 briefly concludes. Appendix A contains additional simulation evidence using Euler equation and finite mixture models. Appendix B shows that our new CSs for the identified set and its functionals are uniformly valid (over DGPs), and Appendix D presents results on local power. Appendix E establishes a more general Bernstein-von Mises type theorem, showing that the limiting distribution of the posterior QLR in a partially identified parametric model is a gamma distribution with scale parameter 2 but a unknown shape parameter. There, results on models with parameter-dependent support for example are given. Appendix F contains all the proofs and additional lemmas.

2 Description of the Procedures

Let $\mathbf{X}_n = (X_1, \ldots, X_n)$ denote a sample of i.i.d. or strictly stationary and ergodic data of size n.[11] Consider a population objective function $L : \Theta \to \mathbb{R}$ where L can be a log likelihood func-

[11]Throughout we work on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Each X_i takes values in a separable metric space $\mathscr{X}$ equipped with its Borel σ-algebra $\mathscr{B}(\mathscr{X})$. We equip Θ with its Borel σ-algebra $\mathscr{B}(\Theta)$.

tion for correctly specified likelihood models, an optimally-weighted GMM objective function, a continuously-updated GMM objective function, or a sandwich quasi-likelihood function. The function L is assumed to be an upper semicontinuous function of θ with $\sup_{\theta \in \Theta} L(\theta) < \infty$.

The key problem is that the population objective L may not be maximized uniquely over Θ, but rather its maximizers, the *identified set*, may be a nontrivial set of parameters. The identified set (IdS) is defined as follows:

$$\Theta_I := \left\{ \theta \in \Theta : L(\theta) = \sup_{\vartheta \in \Theta} L(\vartheta) \right\} .$$

The set Θ_I is our parameter of interest. We propose methods to construct confidence sets (CSs) for Θ_I that are computationally attractive and have (asymptotically) correct frequentist coverage probabilities.

To describe our approach, let L_n denote an (upper semicontinuous) sample criterion function that is a jointly measurable function of the data $\mathbf{X}_n$ and θ. This objective function $L_n(\cdot)$ can be a natural sample analog of L. We give a few examples of objective functions that we consider.

Parametric likelihood: Given a parametric model: $\{P_\theta : \theta \in \Theta\}$, with a corresponding density[12] $p(.;\theta)$, the identified set can be defined as $\Theta_I = \{\theta \in \Theta : P_0 = P_\theta\}$ where P_0 is the true data distribution. We take L_n to be the average log-likelihood function:

$$L_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \log p(X_i; \theta) . \tag{1}$$

We cover likelihood based models with lack of (point) identification. We could also take L_n to be the average sandwich log-likelihood function in misspecified models (see Remark 3).

GMM models: Consider a set of *moment equalities* $E[\rho(X_i, \theta)] = 0$ such that the solution to this vector of equalities may not be unique. Here, we define the set of interest as $\Theta_I = \{\theta \in \Theta : E[\rho(X_i, \theta)] = 0\}$. The sample objective function L_n can be the continuously-updated GMM objective function:

$$L_n(\theta) = -\frac{1}{2} \left(\frac{1}{n} \sum_{i=1}^{n} \rho(X_i, \theta) \right)' \left(\frac{1}{n} \sum_{i=1}^{n} \rho(X_i, \theta)\rho(X_i, \theta)' \right)^{-} \left(\frac{1}{n} \sum_{i=1}^{n} \rho(X_i, \theta) \right) \tag{2}$$

where A^- denotes a generalized inverse of a matrix A,[13] or an optimally-weighted GMM objective

[12]This density of P_θ is understood to be with respect to a common σ-finite dominating measure.

[13]We could also take the continuously-updated weighting matrix to be $(\frac{1}{n}\sum_{i=1}^{n} \rho(X_i, \theta)\rho(X_i, \theta)' - (\frac{1}{n}\sum_{i=1}^{n} \rho(X_i, \theta))(\frac{1}{n}\sum_{i=1}^{n} \rho(X_i, \theta))')^{-}$ or, for time series data, a form that takes into account any autocorrelations in the residual functions $\rho(X_i, \theta)$. See, e.g., Hansen et al. (1996).

function:

$$L_n(\theta) = -\frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n}\rho(X_i,\theta)\right)' \widehat{W} \left(\frac{1}{n}\sum_{i=1}^{n}\rho(X_i,\theta)\right) \tag{3}$$

for suitable weighting matrix $\widehat{W}$. We could also take L_n to be a generalized empirical likelihood objective function.

The question we pose is given $\mathbf{X}_n$, how to construct computationally attractive CS that covers the IdS Θ_I or functions of the IdS with a prespecified probability (in repeated samples) as sample size gets large.

We first describe our computational method (Procedure 1) for covering the IdS Θ_I. We then describe methods (including Procedure 2) for covering a function of Θ_I, such as a subvector. We also present an extremely simple method (Procedure 3) for covering the identified set for a scalar subvector in certain situations.

Our main CS constructions (Procedures 1 and 2) are based on Monte Carlo (MC) simulation methods using a well defined quasi posterior. Given L_n and a prior measure Π on $(\Theta, \mathscr{B}(\Theta))$ (such as a flat prior), the quasi-posterior distribution Π_n for θ given $\mathbf{X}_n$ is defined as

$$\Pi_n(A \,|\, \mathbf{X}_n) = \frac{\int_A e^{nL_n(\theta)} d\Pi(\theta)}{\int_\Theta e^{nL_n(\theta)} d\Pi(\theta)} \quad \text{for } A \in \mathscr{B}(\Theta)\ . \tag{4}$$

In the following we use MCMC chains for pedagogical convenience, although many MC samplers could be used to draw a sample $\{\theta^1, \ldots, \theta^B\}$ from the quasi-posterior Π_n.

2.1 Confidence sets for the identified set

Given $\mathbf{X}_n$, we seek to construct a $100\alpha\%$ CS $\widehat{\Theta}_\alpha$ for Θ_I using $L_n(\theta)$ that has asymptotically exact coverage, i.e.:

$$\lim_{n\to\infty} \mathbb{P}(\Theta_I \subseteq \widehat{\Theta}_\alpha) = \alpha\,.$$

We propose an MCMC based method to obtain $\widehat{\Theta}_\alpha$ as follows.

[Procedure 1: Confidence sets for the identified set]

1. Draw an MCMC chain $\{\theta^1, \ldots, \theta^B\}$ from the quasi-posterior distribution Π_n in (4).

2. Calculate the $(1-\alpha)$ quantile of $\{L_n(\theta^1), \ldots, L_n(\theta^B)\}$ and call it $\zeta^{mc}_{n,\alpha}$.

3. Our $100\alpha\%$ MCMC confidence set for Θ_I is then:

$$\widehat{\Theta}_\alpha = \{\theta \in \Theta : L_n(\theta) \geq \zeta_{n,\alpha}^{mc}\}\,. \tag{5}$$

Notice that no optimization of L_n itself is required in order to construct $\widehat{\Theta}_\alpha$. Further, an exhaustive grid search over the full parameter space Θ is not required as the MCMC draws $\{\theta^1, \ldots, \theta^B\}$ will concentrate around Θ_I and thereby indicate the regions in Θ over which to search.

CHT considered inference on the set of minimizers of a *nonnegative population criterion function* $Q : \Theta \to \mathbb{R}_+$ using a sample analogue Q_n of Q. Let $\xi_{n,\alpha}$ denote a consistent estimator of the α quantile of $\sup_{\theta \in \Theta_I} Q_n(\theta)$. The $100\alpha\%$ CS for Θ_I at level $\alpha \in (0,1)$ proposed is $\widehat{\Theta}_\alpha^{CHT} = \{\theta \in \Theta : Q_n(\theta) \leq \xi_{n,\alpha}\}$. In the existing literature, subsampling or bootstrap based methods have been used to compute $\xi_{n,\alpha}$. The next remark provides an equivalent approach to Procedure 1 but that is constructed in terms of Q_n, which is the quasi likelihood ratio statistic associated with L_n. So, instead of computationally intensive subsampling and bootstrap, our procedure replaces $\xi_{n,\alpha}$ with a cut off based on Monte Carlo simulations.

Remark 1. *Let $\hat{\theta} \in \Theta$ denote an approximate maximizer of L_n, i.e.:*

$$L_n(\hat{\theta}) = \sup_{\theta \in \Theta} L_n(\theta) + o_{\mathbb{P}}(n^{-1})\,.$$

and define the quasi-likelihood ratio (QLR) (at a point $\theta \in \Theta$) as:

$$Q_n(\theta) = 2n[L_n(\hat{\theta}) - L_n(\theta)]\,. \tag{6}$$

Let $\xi_{n,\alpha}^{mc}$ denote the α quantile of $\{Q_n(\theta_1), \ldots, Q_n(\theta^B)\}$. The confidence set:

$$\widehat{\Theta}'_\alpha = \{\theta \in \Theta : Q_n(\theta) \leq \xi_{n,\alpha}^{mc}\}$$

is equivalent to $\widehat{\Theta}_\alpha$ defined in (5) because $L_n(\theta) \geq \zeta_{n,\alpha}^{mc}$ if and only if $Q_n(\theta) \leq \xi_{n,\alpha}^{mc}$.

In Procedure 1 and Remark 1 above, the posterior like quantity involves the use of a prior distribution Π over Θ. This prior is user defined and typically would be the uniform prior but other choices are possible, and in our simulations, the various choices of this prior did not seem to matter much when the parameter space Θ is compact. Here, the way to obtain the draws $\{\theta^1, \ldots, \theta^B\}$ will rely on a Monte Carlo sampler. We use existing sampling methods to do this. Below we describe how these methods are tuned to our examples. For partially-identified models, the parameter chain $\{\theta^1, \ldots, \theta^B\}$ may not settle down but the criterion chain $\{Q_n(\theta^1), \ldots, Q_n(\theta^B)\}$ still converges. Our MCMC CSs are constructed based on the quantiles of a criterion chain and are intuitively robust to lack of point identification.

The next lemma presents high-level conditions under which *any* $100\alpha\%$ criterion-based CS for Θ_I is asymptotically valid. Similar statements appear in CHT and Romano and Shaikh (2010). Let $F_W(c) := Pr(W \leq c)$ denote the (probability) distribution function of a random variable W and $w_\alpha := \inf\{c \in \mathbb{R} : F_W(c) \geq \alpha\}$ be the α quantile of W.

Lemma 2.1. *Let (i)* $\sup_{\theta \in \Theta_I} Q_n(\theta) \rightsquigarrow W$ *where* W *is a random variable whose distribution function* $F_W()$ *is continuous at its* α *quantile (denoted by* w_α*), and (ii)* $(w_{n,\alpha})_{n\in\mathbb{N}}$ *be a sequence of random variables such that* $w_{n,\alpha} \geq w_\alpha + o_{\mathbb{P}}(1)$*. Define:*

$$\widehat{\Theta}_\alpha = \{\theta \in \Theta : Q_n(\theta) \leq w_{n,\alpha}\} .$$

Then: $\liminf_{n\to\infty} \mathbb{P}(\Theta_I \subseteq \widehat{\Theta}_\alpha) \geq \alpha$*. Moreover, if condition (ii) is replaced by the condition* $w_{n,\alpha} = w_\alpha + o_{\mathbb{P}}(1)$*, then:* $\lim_{n\to\infty} \mathbb{P}(\Theta_I \subseteq \widehat{\Theta}_\alpha) = \alpha$*.*

Our MCMC CSs for Θ_I are shown to be valid by verifying parts (i) and (ii) with $w_{n,\alpha} = \xi^{mc}_{n,\alpha}$. To verify part (ii), we shall establish a new Bernstein-von Mises (BvM) result for the posterior distribution of the QLR under loss of identifiability for likelihood and GMM models. Therefore, although our Procedure 1 above appears Bayesian,[14] we show that $\widehat{\Theta}_\alpha$ has correct frequentist coverage.

2.2 Confidence sets for functions of the identified set

In many applications, it may be of interest to provide a CS for a *subvector* of interest. Suppose that the object of interest is a function of θ, say $\mu(\theta)$, for some continuous function $\mu : \Theta \to \mathbb{R}^k$ for $1 \leq k < \dim(\theta)$. This includes as a special case in which $\mu(\theta)$ is a subvector of θ itself (i.e., $\theta = (\mu, \eta)$ with μ being the subvector of interest and η the nuisance parameter). The identified set for $\mu(\theta)$ is:

$$M_I = \{\mu(\theta) : \theta \in \Theta_I\} .$$

We seek a CS $\widehat{M}_\alpha$ for M_I such that:

$$\lim_{n\to\infty} \mathbb{P}(M_I \subseteq \widehat{M}_\alpha) = \alpha .$$

A well known method to construct a CS for M_I is based on projection, which maps a CS $\widehat{\Theta}_\alpha$ for Θ_I into one that covers a function of Θ_I. In particular, the following MCMC CS:

$$\widehat{M}^{proj}_\alpha = \{\mu(\theta) : \theta \in \widehat{\Theta}_\alpha\} \tag{7}$$

[14] In correctly specified likelihood models with flat priors one may interpret $\widehat{\Theta}_\alpha$ as a highest posterior density $100\alpha\%$ Bayesian credible set (BCS) for Θ_I. Therefore, $\widehat{\Theta}_\alpha$ will have the smallest volume of any BCS for Θ_I.

is a valid $100\alpha\%$ CS for M_I whenever $\widehat{\Theta}_\alpha$ is a valid $100\alpha\%$ CS for Θ_I. As is well known, $\widehat{M}_\alpha^{proj}$ is typically conservative, and could be very conservative when the dimension of μ is small relative to the dimension of θ. Our simulations below indicate that $\widehat{M}_\alpha^{proj}$ is very conservative even in reasonably low-dimensional parametric models.

In the following we propose CSs $\widehat{M}_\alpha$ for M_I that could have asymptotically exact coverage based on a profile criterion for M_I. Let $M = \{\mu(\theta) : \theta \in \Theta\}$ and $\mu^{-1} : M \to \Theta$, i.e., $\mu^{-1}(m) = \{\theta \in \Theta : \mu(\theta) = m\}$ for each $m \in M$. The profile criterion for a point $m \in M$ is

$$\sup_{\theta \in \mu^{-1}(m)} L_n(\theta), \tag{8}$$

and the profile criterion for the identified set M_I is

$$\inf_{m \in M_I} \sup_{\theta \in \mu^{-1}(m)} L_n(\theta). \tag{9}$$

Let $\Delta(\theta^b) = \{\theta \in \Theta : L(\theta) = L(\theta^b)\}$ be an equivalence set for θ^b, $b = 1, ..., B$. For example, in correctly specified likelihood models we have $\Delta(\theta^b) = \{\theta \in \Theta : p(\cdot;\theta) = p(\cdot;\theta^b)\}$ and in GMM models we have $\Delta(\theta^b) = \{\theta \in \Theta : E[\rho(X_i,\theta)] = E[\rho(X_i,\theta^b)]\}$.

[Procedure 2: CSs for functions of the identified set]

1. Draw an MCMC chain $\{\theta^1, \ldots, \theta^B\}$ from the quasi-posterior distribution Π_n in (4).
2. Calculate the $(1-\alpha)$ quantile of $\left\{\inf_{m \in \mu(\Delta(\theta^b))} \sup_{\theta \in \mu^{-1}(m)} L_n(\theta) : b = 1, \ldots, B\right\}$ and call it $\zeta_{n,\alpha}^{mc,p}$.
3. Our $100\alpha\%$ MCMC confidence set for M_I is then:

$$\widehat{M}_\alpha = \left\{ m \in M : \sup_{\theta \in \mu^{-1}(m)} L_n(\theta) \geq \zeta_{n,\alpha}^{mc,p} \right\}. \tag{10}$$

By forming $\widehat{M}_\alpha$ in terms of the profile criterion we avoid having to do an exhaustive grid search over Θ. An additional computational advantage is that the MCMC $\{\mu(\theta^1), \ldots, \mu(\theta^B)\}$ concentrate around M_I, thereby indicating the region in M over which to search.

The following remark describes the numerical equivalence between the CS $\widehat{M}_\alpha$ in (10) and a CS for M_I based on the profile QLR.

Remark 2. *Recall the definition of the QLR Q_n in (6). Let $\xi_{n,\alpha}^{mc,p}$ denote the α quantile of the profile QLR chain:*

$$\left\{ \sup_{m \in \mu(\Delta(\theta^b))} \inf_{\theta \in \mu^{-1}(m)} Q_n(\theta) : b = 1, \ldots, B \right\}.$$

The confidence set:

$$\widehat{M}'_\alpha = \left\{ m \in M : \inf_{\theta \in \mu^{-1}(m)} Q_n(\theta) \leq \xi_{n,\alpha}^{mc,p} \right\}$$

is equivalent to $\widehat{M}_\alpha$ in (10) because $\sup_{\theta \in \mu^{-1}(m)} L_n(\theta) \geq \zeta_{n,\alpha}^{mc,p}$ if and only if $\inf_{\theta \in \mu^{-1}(m)} Q_n(\theta) \leq \xi_{n,\alpha}^{mc,p}$.

Our Procedure 2 and Remark 2 above are *different* from taking quantiles of the MCMC parameter chain. Given the MCMC chain $\{\theta^1, \ldots, \theta^B\}$ for θ, a popular percentile MCMC CS (denoted as $\widehat{M}_\alpha^{perc}$) for a scalar parameter μ is computed by taking the upper and lower $100(1-\alpha)/2$ percentiles of the parameter chain $\{\mu(\theta^1), \ldots, \mu(\theta^B)\}$. For models with *point-identified root-n estimable parameters* θ, this approach is known to be valid for likelihood models in standard Bayesian literature and its validity for optimally weighted GMM models has been established by Chernozhukov and Hong (2003). However, this approach is no longer valid and severely undercovers in partially-identified models, as evidenced in the simulation results below.

The following result presents high-level conditions under which any $100\alpha\%$ criterion-based CS for M_I is asymptotically valid. A similar statement appears in Romano and Shaikh (2010).

Lemma 2.2. *Let (i) $\sup_{m \in M_I} \inf_{\theta \in \mu^{-1}(m)} Q_n(\theta) \rightsquigarrow W$ where W is a random variable whose distribution $F_W()$ is continuous at its α quantile (denoted by w_α) and (ii) $(w_{n,\alpha})_{n \in \mathbb{N}}$ be a sequence of random variables such that $w_{n,\alpha} \geq w_\alpha + o_{\mathbb{P}}(1)$. Define:*

$$\widehat{M}_\alpha = \left\{ m \in M : \inf_{\theta \in \mu^{-1}(m)} Q_n(\theta) \leq w_{n,\alpha} \right\}.$$

Then: $\liminf_{n \to \infty} \mathbb{P}(M_I \subseteq \widehat{M}_\alpha) \geq \alpha$. Moreover, if condition (ii) is replaced by the condition $w_{n,\alpha} = w_\alpha + o_{\mathbb{P}}(1)$, then: $\lim_{n \to \infty} \mathbb{P}(M_I \subseteq \widehat{M}_\alpha) = \alpha$.

Our MCMC CSs for M_I are shown to be valid by verifying parts (i) and (ii) with $w_{n,\alpha} = \xi_{n,\alpha}^{mc,p}$.

2.3 A simple but slightly conservative CS for scalar subvectors

For a class of partially identified models with one-dimensional subvectors $M_I = \{\mu(\theta) \in \mathbb{R} : \theta \in \Theta_I\}$, we now propose another CS $\widehat{M}_\alpha^\chi$ which is extremely simple to construct. This new CS for M_I is slightly conservative (whereas $\widehat{M}_\alpha$ could be asymptotically exact), but it's coverage is much less conservative than that of the projection-based CS $\widehat{M}_\alpha^{proj}$.

[Procedure 3: Simple conservative CSs for scalar subvectors]

1. Calculate a maximizer $\hat{\theta}$ for which $L_n(\hat{\theta}) \geq \sup_{\theta \in \Theta} L_n(\theta) + o_{\mathbb{P}}(n^{-1})$.

2. Our $100\alpha\%$ confidence set for M_I is then:

$$\widehat{M}_\alpha^\chi = \left\{ m \in M \subset \mathbb{R} : \inf_{\theta \in \mu^{-1}(m)} Q_n(\theta) \leq \chi^2_{1,\alpha} \right\} \tag{11}$$

where Q_n is the QLR in (6) and $\chi^2_{1,\alpha}$ denotes the α quantile of the χ^2_1 distribution.

Procedure 3 above is justified when the limit distribution of the profile QLR for $M_I = \{\mu(\theta) \in \mathbb{R} : \theta \in \Theta_I\}$ is stochastically dominated by the χ^2_1 distribution (i.e., $F_W(z) \geq F_{\chi^2_1}(z)$ for all $z \geq 0$ in Lemma 2.2). This allows for computationally simple construction using repeated evaluations on a scalar grid. Unlike $\widehat{M}_\alpha$, the CS $\widehat{M}_\alpha^\chi$ for M_I is typically asymptotically conservative and is only valid for scalar functions of Θ_I (see Section 3.3). Nevertheless, the CS $\widehat{M}_\alpha^\chi$ is asymptotically exact when M_I happens to be a singleton belonging to the interior of M, and, for confidence levels of $\alpha \geq 0.85$, its degree of conservativeness for the set M_I is negligible (see Section 3.3). It is extremely simple to implement and performs very favorably in simulations. As a sensitivity check in empirical estimation of a complicated structural model, one could report the conventional CS based on a t-statistic (that is valid under point identification only) as well as our CS $\widehat{M}_\alpha^\chi$ (that remains valid under partial identification); see Section 5.

2.4 Simulation evidence

In this section we investigate the finite sample behavior of our proposed CSs in the leading missing data and entry game examples. Further simulation evidences for weakly-identified Euler equation models and finite mixture models are presented in Appendix A. We use samples of size n = 100, 250, 500, and 1000. For each sample, we calculate the posterior quantile of the QLR statistic using 10000 draws from a random walk Metropolis-Hastings scheme with a burnin of an additional 10000 draws. The random walk Metropolis-Hastings scheme is tuned so that its acceptance rate is approximately one third.[15] Note that for partially-identified models, the parameter chain may not settle down but the criterion chain is stable. We replicate each experiment 5000 times.

[15] There is a large literature on tuning Metropolis-Hastings algorithms (see, e.g., Besag, Green, Higdon, and Mengersen (1995), Gelman, Roberts, and Gilks (1996) and Roberts, Gelman, and Gilks (1997)). Optimal acceptance ratios for Gaussian models are known to be between 0.23 and 0.44 depending on the dimension of the parameter (Gelman et al., 1996). For concreteness we settle on 0.33, though similar results are achieved with different acceptance rates. To implement the random walk Metropolis-Hastings algorithm we rescale each parameter to have full support $\mathbb{R}$ via a suitably centered and scaled vector logit transform $\ell : \Theta \to \mathbb{R}^d$. We draw each proposal $\ell^{b+1} := \ell(\theta^{b+1})$ from $N(\ell^b, cI)$ where c is chosen so that the acceptance rate is approximately one third.

2.4.1 Missing data

We first consider the simplest but most insightful missing data example. Suppose we observe a random sample $\{(D_i, Y_i D_i)\}_{i=1}^n$ where both the outcome variable Y_i and the selection variable D_i take values in $\{0,1\}$. The main parameter of interest is (usually) the true mean $\mu_0 = \mathbb{E}[Y_i]$. Without further assumptions, μ_0 is not point identified when $\Pr(D_i = 0) > 0$ as we only observe Y_i when $D_i = 1$. We assume that $0 < \Pr(Y_i = 1 | D_i = 1) < 1$. The true probabilities of observing $(D_i, Y_i D_i) = (1,1)$, $(0,0)$ and $(1,0)$ are κ_{11}, κ_{00}, and $\kappa_{10} = 1 - \kappa_{11} - \kappa_{00}$ respectively. We view these as true *reduced-form parameters* that can be consistently estimated from the data. The reduced-form parameters are functions of the structural parameter $\theta = (\mu, \beta, \rho)$ where $\mu = \mathbb{E}[Y_i]$, $\beta = \Pr(Y_i = 1 | D_i = 0)$, and $\rho = \Pr(D_i = 1)$. Using the model and the parametrization above, θ is related to the reduced form parameters via the following equalities:

$$\kappa_{11}(\theta) = \mu - \beta(1-\rho) \qquad \kappa_{10}(\theta) = \rho - \mu + \beta(1-\rho) \qquad \kappa_{00}(\theta) = 1 - \rho\,.$$

and so the parameter space Θ for θ is defined as:

$$\Theta = \{(\mu, \beta, \rho) \in \mathbb{R}^3 : 0 \leq \mu - \beta(1-\rho) \leq \rho, 0 \leq \beta \leq 1, 0 \leq \rho \leq 1\}\,. \tag{12}$$

The likelihood of the i-th observation $(D_i, Y_i D_i) = (d, yd)$ is

$$p(d, yd; \theta) = [\kappa_{11}(\theta)]^{yd}(1 - \kappa_{11}(\theta) - \kappa_{00}(\theta))^{d-yd}[\kappa_{00}(\theta)]^{1-d}\,.$$

In some simulations we also use a continuously-updated GMM objective function based on the moments:

$$\begin{aligned} E\Big[1\!\!1\big((D_i, Y_i D_i) = (1,1)\big) - \kappa_{11}(\theta)\Big] &= 0 \\ E\Big[1\!\!1\big(D_i = 0\big) - \kappa_{00}(\theta)\Big] &= 0\,. \end{aligned}$$

Defining the model via moment equalities, we obtain a quasi posterior based on an optimal objective function.

The identified set for θ is:

$$\Theta_I = \{(\mu, \beta, \rho) \in \Theta : \mu - \beta(1-\rho) = \kappa_{11}, \rho = 1 - \kappa_{00}\}. \tag{13}$$

Here, ρ is always identified but only an affine combination of μ and β are identified. This combination results in the identified set for (μ, β) being a line segment. The identified set for the subvector $\mu = E[Y]$ is

$$M_I = [\kappa_{11}, \kappa_{11} + \kappa_{00}].$$

In the existing literature one typically uses the following *moment inequality model* for inference on $\mu = E[Y] \in M_I$:

$$\begin{aligned} \mu &\leq E[Y|D=1]P(D=1) + P(D=0) \\ \mu &\geq E[Y|D=1]P(D=1)\,. \end{aligned}$$

Generally, all moment inequality models (with finitely many moment inequalities) can be written as moment equality models by adding nuisance parameters with a known sign (see Subsection 4.2.1).

We use two kinds of priors on Θ:

1. A flat prior
2. A curved prior: take $\pi(\mu, \beta, \rho) = \pi_B(\beta)\pi_P(\rho)\pi_{M|B,P}(\mu|\beta,\rho)$ with $\pi_B(\beta) = \text{Beta}(3,8)$, $\pi_P(\rho) = \text{Beta}(8,1)$, and $\pi_{M|B,P}(\mu|\beta,\rho) = U[\beta(1-\rho), \rho + \beta(1-\rho)]$ (see Figure 6).

We set $\mu_0 = 0.5$, $\beta_0 = 0.5$, and vary ρ_0, covering both point- ($\rho_0 = 1$) and partially-identified ($\rho_0 < 1$) cases.

CSs for the identified set Θ_I: Table 1 displays the MC coverage probabilities of $\widehat{\Theta}_\alpha$ (Procedure 1 with a likelihood criterion and a flat prior) for different values of ρ, different sample sizes and different nominal coverage probabilities. The coverage probability should be equal to its nominal value in large samples when $\rho < 1$ (see Theorem 3.1 below). It is perhaps surprising that the nominal and coverage probabilities are this close even in samples as small as $n = 100$; the only exception is the case $\rho = 0.99$ in which the CSs are slightly conservative when $n = 100$. When $\rho = 1$ the CSs (based on the likelihood criterion) for Θ_I are expected to be conservative (see Theorem 3.2 below), which they are. The coverage probabilities are quite insensitive to the size of small to moderate values of ρ. For instance, the coverage probabilities are very similar for $\rho = 0.20$ (corresponding to 80% of data missing) and $\rho = 0.95$ (corresponding to 5% of data missing). Table 2 presents results for the case a curved prior is used. Whether a flat or curved prior is used makes virtually no difference, except for $\widehat{\Theta}_\alpha$ with $\rho = 0.20$ with smaller values of n. In this case the MCMC CS over covers because the prior is of the order of 10^{-4} at $\rho = 0.20$. The posterior distribution assigns very low weight to values of ρ less than one half. The MCMC chain for ρ concentrates relatively far away from $\rho = 0.20$, and, as a consequence, the posterior distribution of the likelihood ratio is larger than it should be. In sum, the performance under both priors is similar and adequate.

Results for CSs $\widehat{\Theta}_\alpha$ using Procedure 1 with a continuously-updated GMM criterion and a flat

	$\rho = 0.20$	$\rho = 0.80$	$\rho = 0.95$	$\rho = 0.99$	$\rho = 1.00$
			$n = 100$		
$\alpha = 0.90$	0.8904	0.8850	0.8856	0.9378	0.9864
$\alpha = 0.95$	0.9458	0.9422	0.9452	0.9702	0.9916
$\alpha = 0.99$	0.9890	0.9868	0.9884	0.9938	0.9982
			$n = 250$		
$\alpha = 0.90$	0.8962	0.8954	0.8980	0.9136	0.9880
$\alpha = 0.95$	0.9454	0.9436	0.9466	0.9578	0.9954
$\alpha = 0.99$	0.9888	0.9890	0.9876	0.9936	0.9986
			$n = 500$		
$\alpha = 0.90$	0.8890	0.8974	0.9024	0.8952	0.9860
$\alpha = 0.95$	0.9494	0.9478	0.9494	0.9534	0.9946
$\alpha = 0.99$	0.9910	0.9900	0.9884	0.9900	0.9994
			$n = 1000$		
$\alpha = 0.90$	0.9018	0.9038	0.8968	0.8994	0.9878
$\alpha = 0.95$	0.9462	0.9520	0.9528	0.9532	0.9956
$\alpha = 0.99$	0.9892	0.9916	0.9908	0.9894	0.9994

Table 1: MC coverage probabilities of $\widehat{\Theta}_\alpha$ (Procedure 1) using a likelihood for L_n and a flat prior on Θ.

	$\rho = 0.20$	$\rho = 0.80$	$\rho = 0.95$	$\rho = 0.99$	$\rho = 1.00$
			$n = 100$		
$\alpha = 0.90$	0.9750	0.8900	0.8722	0.9316	0.9850
$\alpha = 0.95$	0.9906	0.9460	0.9400	0.9642	0.9912
$\alpha = 0.99$	0.9992	0.9870	0.9850	0.9912	0.9984
			$n = 250$		
$\alpha = 0.90$	0.9526	0.8958	0.8932	0.9072	0.9874
$\alpha = 0.95$	0.9794	0.9456	0.9438	0.9560	0.9954
$\alpha = 0.99$	0.9978	0.9896	0.9864	0.9924	0.9986
			$n = 500$		
$\alpha = 0.90$	0.9306	0.8956	0.8996	0.8926	0.9848
$\alpha = 0.95$	0.9710	0.9484	0.9498	0.9518	0.9944
$\alpha = 0.99$	0.9966	0.9900	0.9880	0.9906	0.9994
			$n = 1000$		
$\alpha = 0.90$	0.9222	0.9046	0.8960	0.8988	0.9880
$\alpha = 0.95$	0.9582	0.9536	0.9500	0.9518	0.9958
$\alpha = 0.99$	0.9942	0.9918	0.9902	0.9888	0.9992

Table 2: MC coverage probabilities of $\widehat{\Theta}_\alpha$ (Procedure 1) using a likelihood for L_n and a curved prior on Θ.

prior are presented in Table 3. As can be seen, the results look similar to those based on the likelihood. Even at sample size 100, the coverage is adequate even $\rho = 1$. Theoretical coverage results for the GMM case are provided in Section 4.2 below.

CSs for the identified set of subvectors M_I: We now consider various CSs for the identified set M_I for μ. We first compute the MCMC projection CS $\widehat{M}_\alpha^{proj}$, as defined in (7), for M_I. The coverage results are reported in Table 4. As we can see from the table, for the case when $\alpha = .90$, the lowest coverage probabilities is above .96. Even when $n = 1000$ and for all values of ρ we tried, the coverage is larger than 96%. So the projection CS $\widehat{M}_\alpha^{proj}$ is valid but too conservative.

One may be tempted to use the parameter (θ) chain itself to construct confidence regions. Figure 1 plots the MCMC chain for a sample with $\rho = .8$. The chain is stable for ρ that is point identified, but the chains for μ and β bounce around their respective identified sets $M_I = [\kappa_{11}, \kappa_{11} + \kappa_{00}]$ and $[0, 1]$. One might be tempted to report the simple percentile MCMC CS $\widehat{M}_\alpha^{perc}$ for M_I (of μ) by taking the upper and lower $100(1-\alpha)/2$ percentiles of the parameter chain $\{\mu(\theta^1), \ldots, \mu(\theta^B)\}$. Table 5 reports the MC coverage probabilities of this simple percentile MCMC CS for μ. It has correct coverage when μ is point identified (i.e. when $\rho = 1$). However, it dramatically undercovers as soon as μ is not point identified, even when only a small amount of data is missing. For instance, with a relatively large sample size $n = 1000$, the coverage of a 90% CS is less than 2% when 20% of data is missing ($\rho = .80$), around 42% when only 5% of data is missing ($\rho = .95$), and less than 83% when only 1% of data is missing ($\rho = .99$). This approach to constructing CSs for M_I by taking *quantiles of the parameter chain* severely undercovers in partially-identified models, and is not recommended.

In contrast, our MCMC CS procedures are based on the *criterion chain* and remains valid under partial identification. Validity under loss of identifiability is preserved because our procedure effectively samples from the quasi-posterior distribution for an identifiable reduced form parameter. The bottom panel of Figure 1 shows the MCMC chain for $Q_n(\theta)$ is stable. Figure 7 (in Appendix A), which is computed from the draws for the structural parameter presented in Figure 1, shows that the MCMC chain for the reduced-form probabilities is also stable. In Table 6, we provide coverage results for $\widehat{M}_\alpha$ using Procedure 2 with a likelihood criterion and a flat prior. Theoretically, we show below (see Theorem 3.3) that the coverage probabilities of $\widehat{M}_\alpha$ (for M_I) should be equal to their nominal values α when n is large irrespective of whether the model is partially identified with $\rho < 1$ or point identified (with $\rho = 1$). Further, Theorem B.2 shows that our Procedure 2 remains valid *uniformly* over sets of DGPs that include both point- and partially-identified cases. The results in Table 6 show that this is indeed the case, and that the coverage probabilities are close to their nominal level even when $n = 100$. This is remarkable as even in the case when $\rho = .8, .95$, or 1, the coverage is very close to the nominal level even when $n = 100$. The exception is the case in which $\rho = 0.20$, which slightly under-covers in small

	$\rho = 0.20$	$\rho = 0.80$	$\rho = 0.95$	$\rho = 0.99$	$\rho = 1.00$
			$n = 100$		
$\alpha = 0.90$	0.8504	0.8810	0.8242	0.9202	0.9032
$\alpha = 0.95$	0.9048	0.9336	0.9062	0.9604	0.9396
$\alpha = 0.99$	0.9498	0.9820	0.9556	0.9902	0.9870
			$n = 250$		
$\alpha = 0.90$	0.8932	0.8934	0.8788	0.9116	0.8930
$\alpha = 0.95$	0.9338	0.9404	0.9326	0.9570	0.9476
$\alpha = 0.99$	0.9770	0.9874	0.9754	0.9920	0.9896
			$n = 500$		
$\alpha = 0.90$	0.8846	0.8938	0.8978	0.8278	0.8914
$\alpha = 0.95$	0.9416	0.9478	0.9420	0.9120	0.9470
$\alpha = 0.99$	0.9848	0.9888	0.9842	0.9612	0.9884
			$n = 1000$		
$\alpha = 0.90$	0.8970	0.9054	0.8958	0.8698	0.9000
$\alpha = 0.95$	0.9474	0.9516	0.9446	0.9260	0.9494
$\alpha = 0.99$	0.9866	0.9902	0.9882	0.9660	0.9908

Table 3: MC coverage probabilities of $\widehat{\Theta}_\alpha$ (Procedure 1) using a CU-GMM for L_n and a flat prior on Θ.

	$\rho = 0.20$	$\rho = 0.80$	$\rho = 0.95$	$\rho = 0.99$	$\rho = 1.00$
			$n = 100$		
$\alpha = 0.90$	0.9686	0.9658	0.9692	0.9784	0.9864
$\alpha = 0.95$	0.9864	0.9854	0.9856	0.9888	0.9916
$\alpha = 0.99$	0.9978	0.9972	0.9968	0.9986	0.9982
			$n = 250$		
$\alpha = 0.90$	0.9696	0.9676	0.9684	0.9706	0.9880
$\alpha = 0.95$	0.9872	0.9846	0.9866	0.9854	0.9954
$\alpha = 0.99$	0.9976	0.9970	0.9978	0.9986	0.9986
			$n = 500$		
$\alpha = 0.90$	0.9686	0.9674	0.9688	0.9710	0.9860
$\alpha = 0.95$	0.9904	0.9838	0.9864	0.9862	0.9946
$\alpha = 0.99$	0.9988	0.9976	0.9966	0.9970	0.9994
			$n = 1000$		
$\alpha = 0.90$	0.9672	0.9758	0.9706	0.9720	0.9878
$\alpha = 0.95$	0.9854	0.9876	0.9876	0.9886	0.9956
$\alpha = 0.99$	0.9978	0.9980	0.9976	0.9970	0.9994

Table 4: MC coverage probabilities of projection CS $\widehat{M}_\alpha^{proj}$ for M_I using a likelihood for L_n and a flat prior on Θ.

samples. Note however that the identified set in this case is the interval $[0.1, 0.9]$, so the poor performance is likely attributable to the fact that the identified set for μ covers close to the whole parameter space for μ.

In section 4.1.1 below we show that in the missing data case the asymptotic distribution of the profile QLR for M_I is stochastically dominated by the χ^2_1 distribution. Using Procedure 3 above we construct $\widehat{M}^{\chi}_{\alpha}$ as in (11) and present the results in Table 7 for the likelihood and Table 8 for the continuously-updated GMM objective functions. As we can see from these tables, the coverage results look remarkably close to their nominal values even for small sample sizes and for all values of ρ.

2.4.2 Complete information entry game with correlated payoff shocks

We now examine the finite-sample performance of our procedures for CS constructions in a complete information entry game example described in Table 9. In each cell, the first entry is the payoff to player 1, and the second entry is the payoff to player 2. So, if player 2 plays 0, then her payoff is normalized to be zero and if player 1 plays 1, then her payoffs is $\beta_1 + \epsilon_1$. We assume that (ϵ_1, ϵ_2), observed by the players, are jointly normally distributed with variance 1 and correlation ρ, an important parameter of interest. It is also assumed that Δ_1 and Δ_2 are both negative and that players play a pure strategy Nash equilibrium. When $-\beta_j \leq \epsilon_j \leq -\beta_j - \Delta_j$, $j = 1, 2$, the game has two equilibria: for given values of the epsilons in this region, the model predicts $(1, 0)$ *and* $(0, 1)$. Let $D_{a_1 a_2}$ denote a binary random variable taking the value 1 if and only if player 1 takes action a_1 and player 2 takes action a_2. We observe a random sample of $\{(D_{00,i}, D_{10,i}, D_{01,i}, D_{11,i})\}_{i=1}^n$. So the data provides information of four choice probabilities $(P(0,0), P(1,0), P(0,1), P(1,1))$, but there are six parameters that need to be estimated: $\theta = (\beta_1, \beta_2, \Delta_1, \Delta_1, \rho, s)$ where $s \in [0, 1]$ is the equilibrium selection probability. The model parameter is partially identified as we have 4 choice probabilities from which we need to learn about 6 parameters.

To proceed, we can link the choice probabilities (reduced-form parameters) to θ as follows:

$$
\begin{aligned}
\kappa_{11}(\theta) :=& P(\epsilon_1 \geq -\beta_1 - \Delta_1;\ \epsilon_2 \geq -\beta_2 - \Delta_2) \\
\kappa_{00}(\theta) :=& P(\epsilon_1 \leq -\beta_1;\ \epsilon_2 \leq -\beta_2) \\
\kappa_{10}(\theta) :=& s \times P(-\beta_1 \leq \epsilon_1 \leq -\beta_1 - \Delta_1;\ -\beta_2 \leq \epsilon_2 \leq -\beta_2 - \Delta_2) \\
& + P(\epsilon_1 \geq -\beta_1; \epsilon_2 \leq -\beta_2) + P(\epsilon_1 \geq -\beta_1 - \Delta_1; -\beta_2 \leq \epsilon_2 \leq -\beta_2 - \Delta_2)\,.
\end{aligned}
$$

Denote the true choice probabilities $(P(0,0), P(1,0), P(0,1), P(1,1))$ (the true reduced-form parameter values) as $(\kappa_{00}, \kappa_{10}, \kappa_{01}, \kappa_{11})$. Then the equalities above naturally suggest a GMM

Figure 1: MCMC chain for θ and $Q_n(\theta)$ for $n = 1000$ with a flat prior on Θ.

	$\rho = 0.20$	$\rho = 0.80$	$\rho = 0.95$	$\rho = 0.99$	$\rho = 1$ CH
			$n = 100$		
$\alpha = 0.90$	0.0024	0.3546	0.7926	0.8782	0.9072
$\alpha = 0.95$	0.0232	0.6144	0.8846	0.9406	0.9428
$\alpha = 0.99$	0.2488	0.9000	0.9744	0.9862	0.9892
			$n = 250$		
$\alpha = 0.90$	0.0010	0.1340	0.6960	0.8690	0.8978
$\alpha = 0.95$	0.0064	0.3920	0.8306	0.9298	0.9488
$\alpha = 0.99$	0.0798	0.8044	0.9568	0.9842	0.9914
			$n = 500$		
$\alpha = 0.90$	0.0000	0.0474	0.5868	0.8484	0.8916
$\alpha = 0.95$	0.0020	0.1846	0.7660	0.9186	0.9470
$\alpha = 0.99$	0.0202	0.6290	0.9336	0.9832	0.9892
			$n = 1000$		
$\alpha = 0.90$	0.0000	0.0144	0.4162	0.8276	0.9006
$\alpha = 0.95$	0.0002	0.0626	0.6376	0.9086	0.9490
$\alpha = 0.99$	0.0016	0.3178	0.8972	0.9808	0.9908

Table 5: MC coverage probabilities of $\widehat{M}_\alpha^{perc}$ for M_I (of μ) (with a flat prior on Θ). $\widehat{M}_\alpha^{perc}$ becomes CH's percentile CS under point identification (i.e. when $\rho = 1$).

	$\rho = 0.20$	$\rho = 0.80$	$\rho = 0.95$	$\rho = 0.99$	$\rho = 1.00$
			$n = 100$		
$\alpha = 0.90$	0.8674	0.9170	0.9160	0.9166	0.9098
$\alpha = 0.95$	0.9344	0.9522	0.9554	0.9568	0.9558
$\alpha = 0.99$	0.9846	0.9906	0.9908	0.9910	0.9904
			$n = 250$		
$\alpha = 0.90$	0.8778	0.9006	0.9094	0.9118	0.9078
$\alpha = 0.95$	0.9458	0.9506	0.9548	0.9536	0.9532
$\alpha = 0.99$	0.9870	0.9902	0.9922	0.9894	0.9916
			$n = 500$		
$\alpha = 0.90$	0.8878	0.9024	0.9054	0.9042	0.8994
$\alpha = 0.95$	0.9440	0.9510	0.9526	0.9530	0.9510
$\alpha = 0.99$	0.9912	0.9878	0.9918	0.9918	0.9906
			$n = 1000$		
$\alpha = 0.90$	0.8902	0.9064	0.9110	0.9078	0.9060
$\alpha = 0.95$	0.9438	0.9594	0.9532	0.9570	0.9526
$\alpha = 0.99$	0.9882	0.9902	0.9914	0.9910	0.9912

Table 6: MC coverage probabilities of $\widehat{M}_\alpha$ for M_I (Procedure 2) using a likelihood for L_n and a flat prior on Θ.

	$\rho = 0.20$	$\rho = 0.80$	$\rho = 0.95$	$\rho = 0.99$	$\rho = 1.00$
			$n = 100$		
$\alpha = 0.90$	0.9180	0.9118	0.8988	0.8966	0.9156
$\alpha = 0.95$	0.9534	0.9448	0.9586	0.9582	0.9488
$\alpha = 0.99$	0.9894	0.9910	0.9910	0.9908	0.9884
			$n = 250$		
$\alpha = 0.90$	0.9144	0.8946	0.8972	0.8964	0.8914
$\alpha = 0.95$	0.9442	0.9538	0.9552	0.9520	0.9516
$\alpha = 0.99$	0.9922	0.9908	0.9910	0.9912	0.9912
			$n = 500$		
$\alpha = 0.90$	0.9080	0.9120	0.8984	0.8998	0.9060
$\alpha = 0.95$	0.9506	0.9510	0.9554	0.9508	0.9472
$\alpha = 0.99$	0.9936	0.9926	0.9912	0.9896	0.9882
			$n = 1000$		
$\alpha = 0.90$	0.8918	0.8992	0.8890	0.9044	0.9076
$\alpha = 0.95$	0.9540	0.9494	0.9466	0.9484	0.9488
$\alpha = 0.99$	0.9910	0.9928	0.9916	0.9896	0.9906

Table 7: MC coverage probabilities of $\widehat{M}_\alpha^\chi$ for M_I (Procedure 3) using a likelihood for L_n.

	$\rho = 0.20$	$\rho = 0.80$	$\rho = 0.95$	$\rho = 0.99$	$\rho = 1.00$
			$n = 100$		
$\alpha = 0.90$	0.9536	0.9118	0.8988	0.8966	0.9156
$\alpha = 0.95$	0.9786	0.9448	0.9586	0.9582	0.9488
$\alpha = 0.99$	0.9984	0.9910	0.9910	0.9908	0.9884
			$n = 250$		
$\alpha = 0.90$	0.9156	0.8946	0.8972	0.8964	0.8914
$\alpha = 0.95$	0.9656	0.9538	0.9552	0.9520	0.9516
$\alpha = 0.99$	0.9960	0.9908	0.9910	0.9882	0.9912
			$n = 500$		
$\alpha = 0.90$	0.9300	0.9120	0.8984	0.8992	0.9060
$\alpha = 0.95$	0.9666	0.9510	0.9554	0.9508	0.9472
$\alpha = 0.99$	0.9976	0.9926	0.9912	0.9896	0.9882
			$n = 1000$		
$\alpha = 0.90$	0.9088	0.8992	0.9050	0.8908	0.8936
$\alpha = 0.95$	0.9628	0.9494	0.9544	0.9484	0.9488
$\alpha = 0.99$	0.9954	0.9928	0.9916	0.9896	0.9906

Table 8: MC coverage probabilities of $\widehat{M}^\chi_\alpha$ for M_I (Procedure 3) using a CU-GMM for L_n.

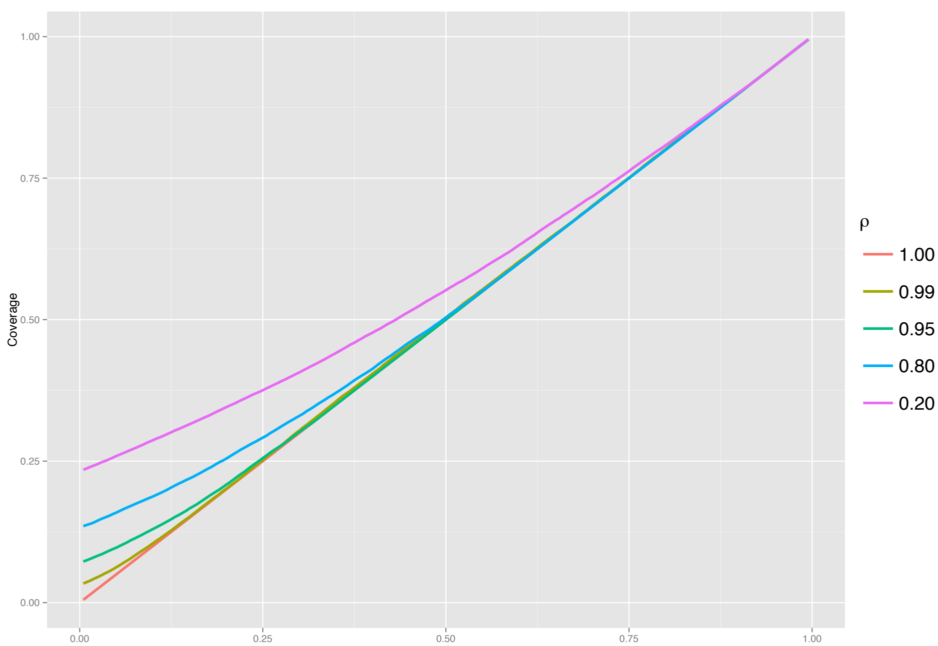

Figure 2: Comparison of asymptotic coverage of $\widehat{M}^\chi_\alpha$ of M_I for different ρ values.

		Player 2			
		0		1	
Player 1	0	0	0	0	$\beta_2 + \epsilon_2$
	1	$\beta_1 + \epsilon_1$	0	$\beta_1 + \Delta_1 + \epsilon_1$	$\beta_2 + \Delta_2 + \epsilon_2$

Table 9: Payoff matrix for the binary entry game

approach via the following moments:

$$\kappa_{11}(\theta) - \kappa_{11} = 0, \quad \kappa_{00}(\theta) - \kappa_{00} = 0, \quad \kappa_{10}(\theta) - \kappa_{10} = 0 \,.$$

In the simulations we use a likelihood approach, where the likelihood of the i-th observation $(D_{00,i}, D_{10,i}, D_{11,i}, D_{01,i}) = (d_{00}, d_{10}, d_{11}, 1 - d_{00} - d_{10} - d_{11})$ is:

$$p(d_{00}, d_{10}, d_{11}; \theta) = [\kappa_{00}(\theta)]^{d_{00}}[\kappa_{10}(\theta)]^{d_{10}}[\kappa_{11}(\theta)]^{d_{11}}[1 - \kappa_{00}(\theta) - \kappa_{10}(\theta) - \kappa_{11}(\theta)]^{1-d_{00}-d_{10}-d_{11}} \,.$$

The parameter space used in the simulations is:

$$\Theta = \{(\beta_1, \beta_2, \Delta_1, \Delta_2, \rho, s) \in \mathbb{R}^6 : -1 \leq \beta_1, \beta_2 \leq 2, -2 \leq \Delta_1, \Delta_2 \leq 0, 0 \leq \rho, s \leq 1\} \,.$$

We simulate the data using $\beta_1 = \beta_2 = 0.2$, $\Delta_1 = \Delta_2 = -0.5$, $\rho = 0.5$ and $s = 0.5$. The identified set for Δ_1 is approximately $M_I = [-1.42, 0]$. Here, it is not as easy to solve for the identified set Θ_I for θ as it needs to be done numerically. We use a flat prior on Θ.

Figure 8 in Appendix A plots the chain for the structural parameters and the chain for the criterion. The chain for ρ bounces between essentially 0 to 1 which indicates that ρ is not identified at all. On the other hand, the data do provide information about (β_1, β_2) as here we see a tighter path. Although the chain for the structural parameters does not converge, Figure 8 and Figure 9 in Appendix A show that the criterion chain and the chain evaluated at the reduced-form probabilities are all stable.

The procedures for computing the CSs for Θ_I and for M_I follow the descriptions given above. In Table 10, we provide the coverage results for the full vector θ and the subvector Δ_1. Coverage of $\widehat{\Theta}_\alpha$ for Θ_I is extremely good, even with the small sample size $n = 100$. Coverages of $\widehat{M}_\alpha$ and $\widehat{M}_\alpha^\chi$ for M_I are slightly conservative for small sample size n but are close to the nominal value for $n = 500$ or larger.[16] The projection CS $\widehat{M}_\alpha^{proj}$ for M_I (of Δ_1) is valid but extremely conservative. The coverage of percentile MCMC CS $\widehat{M}_\alpha^{perc}$ for Δ_1 is less than 1% for each sample size (and hence not valid).

[16]Here we compute Θ_I and $\Delta(\theta^b)$ numerically because ρ is nonzero, so the slight under-coverage of $\widehat{M}_\alpha$ for $n = 1000$ is likely attributable to numerical error.

MC coverage probabilities of $\widehat{\Theta}_\alpha$ (Procedure 1)				
	$n = 100$	$n = 250$	$n = 500$	$n = 1000$
$\alpha = 0.90$	0.9000	0.9000	0.9018	0.9006
$\alpha = 0.95$	0.9476	0.9476	0.9514	0.9506
$\alpha = 0.99$	0.9872	0.9886	0.9902	0.9880

MC coverage probabilities of $\widehat{M}_\alpha$ (Procedure 2)				
	$n = 100$	$n = 250$	$n = 500$	$n = 1000$
$\alpha = 0.90$	0.9683	0.9381	0.9178	0.8865
$\alpha = 0.95$	0.9887	0.9731	0.9584	0.9413
$\alpha = 0.99$	0.9993	0.9954	0.9904	0.9859

MC coverage probabilities of $\widehat{M}_\alpha^\chi$ (Procedure 3)				
	$n = 100$	$n = 250$	$n = 500$	$n = 1000$
$\alpha = 0.90$	0.9404	0.9326	0.9286	0.9110
$\alpha = 0.95$	0.9704	0.9658	0.9618	0.9464
$\alpha = 0.99$	0.9936	0.9928	0.9924	0.9872

MC coverage probabilities of $\widehat{M}_\alpha^{proj}$ (conservative)				
	$n = 100$	$n = 250$	$n = 500$	$n = 1000$
$\alpha = 0.90$	0.9944	0.9920	0.9894	0.9886
$\alpha = 0.95$	0.9972	0.9964	0.9948	0.9968
$\alpha = 0.99$	1.0000	0.9994	0.9990	0.9986

MC coverage probabilities of $\widehat{M}_\alpha^{perc}$ (undercover)				
	$n = 100$	$n = 250$	$n = 500$	$n = 1000$
$\alpha = 0.90$	0.0004	0.0000	0.0000	0.0000
$\alpha = 0.95$	0.0016	0.0000	0.0002	0.0000
$\alpha = 0.99$	0.0058	0.0008	0.0006	0.0000

Table 10: MC coverage probabilities for the complete information game. All CSs are computed with a likelihood for L_n and a flat prior on Θ. CSs $\widehat{M}_\alpha$, $\widehat{M}_\alpha^\chi$ and $\widehat{M}_\alpha^{proj}$ are for M_I of Δ_1, and $\widehat{M}_\alpha^{perc}$ is percentile CS for Δ_1.

3 Large sample properties

This section provides regularity conditions under which $\widehat{\Theta}_\alpha$ (Procedure 1), $\widehat{M}_\alpha$ (Procedure 2) and $\widehat{M}_\alpha^\chi$ (Procedure 3) are asymptotically valid confidence sets for Θ_I and M_I. The main new theoretical contributions are the derivations of the large-sample (quasi)-posterior distributions of the QLR statistic for Θ_I and of the profile QLR statistic for M_I under loss of identifiability.

3.1 Coverage properties of $\widehat{\Theta}_\alpha$ for Θ_I

We first state some high-level regularity conditions. A discussion of these assumptions follows.

Assumption 3.1. *(Consistency, posterior contraction)*
(i) $L_n(\hat\theta) = \sup_{\theta\in\Theta_{osn}} L_n(\theta) + o_{\mathbb{P}}(n^{-1})$, *with* $(\Theta_{osn})_{n\in\mathbb{N}}$ *a sequence of local neighborhoods of* Θ_I*;*
(ii) $\Pi_n(\Theta_{osn}^c|\,\mathbf{X}_n) = o_{\mathbb{P}}(1)$, *where* $\Theta_{osn}^c = \Theta\backslash\Theta_{osn}$.

We presume the existence of a fixed neighborhood Θ_I^N of Θ_I (with $\Theta_{osn} \subset \Theta_I^N$ for all n sufficiently large) upon which there exists a *local* reduced-form reparameterization $\theta \mapsto \gamma(\theta)$ from Θ_I^N into $\Gamma \subseteq \mathbb{R}^{d^*}$ for a possibly unknown dimension $d^* \in [1,\infty)$, with $\gamma(\theta) = 0$ if and only if $\theta \in \Theta_I$. Here γ is merely a proof device and is only required to exist for θ in a fixed neighborhood of Θ_I. Denote $\|\gamma\|^2 := \gamma'\gamma$.

Assumption 3.2. *(Local quadratic approximation)*
(i) There exist sequences of random variables ℓ_n and $\mathbb{R}^{d^*}$-valued random vectors $\mathbb{V}_n$ (both are measurable functions of data $\mathbf{X}_n$) such that:

$$\sup_{\theta\in\Theta_{osn}} \left| nL_n(\theta) - \left(\ell_n - \frac{1}{2}\|\sqrt{n}\gamma(\theta)\|^2 + (\sqrt{n}\gamma(\theta))'\mathbb{V}_n\right)\right| = o_{\mathbb{P}}(1) \tag{14}$$

with $\sup_{\theta\in\Theta_{osn}} \|\gamma(\theta)\| \to 0$ *and* $\mathbb{V}_n \rightsquigarrow N(0,\Sigma)$ *as* $n\to\infty$*;*
(ii) The sets $K_{osn} = \{\sqrt{n}\gamma(\theta) : \theta \in \Theta_{osn}\}$ *cover*[17] *a closed convex cone* $T \subseteq \mathbb{R}^{d^*}$ *as* $n \to \infty$.

Let Π_Γ denote the image measure of the prior Π under the map $\theta \mapsto \gamma(\theta)$ on Θ_I^N, namely $\Pi_\Gamma(A) = \Pi(\{\theta \in \Theta_I^N : \gamma(\theta) \in A\})$. Let $B_\delta \subset \mathbb{R}^{d^*}$ denote a ball of radius δ centered at the origin.

Assumption 3.3. *(Prior)*
(i) $\int_\Theta e^{nL_n(\theta)}\,\mathrm{d}\Pi(\theta) < \infty$ *almost surely;*
(ii) Π_Γ *has a continuous, strictly positive density* π_Γ *on* $B_\delta \cap \Gamma$ *for some* $\delta > 0$.

[17] We say that a sequence of (possibly sample-dependent) sets $A_n \subseteq \mathbb{R}^{d^*}$ *covers* a set $A \subseteq \mathbb{R}^{d^*}$ if (i) $\sup_{b:\|b\|\le M} |\inf_{a\in A_n}\|a-b\|^2 - \inf_{a\in A}\|a-b\|^2| = o_{\mathbb{P}}(1)$ for each M, and (ii) there is a sequence of closed balls B_{k_n} of radius $k_n \to \infty$ centered at the origin with each $C_n := A_n \cap B_{k_n}$ convex, $C_n \subseteq C_{n'}$ for each $n' \ge n$, and $A = \overline{\cup_{n\ge1} C_n}$ (almost surely).

Let $\xi_{n,\alpha}^{post}$ denote the α quantile of $Q_n(\theta)$ under the posterior distribution Π_n, and $\xi_{n,\alpha}^{mc}$ be given in Remark 1.

Assumption 3.4. *(MC convergence)*
$\xi_{n,\alpha}^{mc} = \xi_{n,\alpha}^{post} + o_{\mathbb{P}}(1)$.

Discussion of Assumptions: Assumption 3.1(i) is a standard condition on any approximate extremum estimator, and Assumption 3.1(ii) is a standard posterior contraction condition. The choice of Θ_{osn} is deliberately general and will depend on the particular model under consideration. See Section 4 for verification of Assumption 3.1. Assumption 3.2(i) is a standard local quadratic expansion condition imposed on the local reduced form parameter around $\gamma = 0$. It is readily verified for likelihood and GMM models (see Section 4). For these models with i.i.d. data the vector $\mathbb{V}_n$ is typically of the form: $\mathbb{V}_n = n^{-1/2}\sum_{i=1}^n v(X_i)$ with $\mathbb{E}[v(X_i)] = 0$ and $\mathrm{Var}[v(X_i)] = \Sigma$. Assumption 3.2(ii) is trivially satisfied whenever each K_{osn} contains a ball of radius k_n centered at the origin. This condition allows for the reduced-form true parameter value $\gamma = 0$ to be on the boundary of Γ (see, e.g., Andrews (1999) for similar condition imposed in identified models when a parameter is on the boundary). Assumption 3.3(i) requires the quasi-posterior to be proper. Assumption 3.3(ii) is a standard prior mass and smoothness condition used to establish Bernstein-von Mises theorems for identified parametric models (see, e.g., Section 10.2 of van der Vaart (2000)) but applied to Π_Γ. Under a flat prior on Θ and a continuous local mapping $\gamma : \Theta_I^N \mapsto \Gamma$, this assumption is easily satisfied (see its verification in examples of Section 4). Assumption 3.4 requires that the distribution of the MC chain $\{Q_n(\theta^1), \dots, Q_n(\theta^B)\}$ well approximates the posterior distribution of $Q_n(\theta)$, which is satisfied by many MC samplers.

Let $\mathbf{T}$ be the orthogonal projection onto the tangent space T (at $\gamma = 0$). Assumptions 3.1(i) and 3.2 imply that the QLR statistic for Θ_I satisfies

$$\sup_{\theta \in \Theta_I} Q_n(\theta) = \|\mathbf{T}\mathbb{V}_n\|^2 + o_{\mathbb{P}}(1)$$

(see Lemma F.1). And hence under the generalized information equality $\Sigma = I_{d^*}$, which corresponds to an optimally weighted criterion such as a correctly-specified likelihood, an optimally-weighted or continuously-updated GMM or various (generalized) empirical-likelihood criterion, the asymptotic distribution of $\sup_{\theta \in \Theta_I} Q_n(\theta)$ becomes F_T, which is defined as

$$F_T(z) := \mathbb{P}_Z(\|\mathbf{T}Z\|^2 \le z) \tag{15}$$

where $\mathbb{P}_Z$ denotes the distribution of a $N(0, I_{d^*})$ random vector Z. This recovers the known asymptotic distribution result for optimally weighted QLR statistic under point identification. Note that when $T = \mathbb{R}^{d^*}$, F_T reduces to $F_{\chi^2_{d^*}}$, the cdf of $\chi^2_{d^*}$ (a chi-square random variable with d^* degree of freedom). If T is polyhedral then F_T is the distribution of a chi-bar-squared random variable (i.e. a mixture of chi squares with different degrees of freedom; the mixing weights themselves depending on the shape of T).

Theorem 3.1. *Let Assumptions 3.1, 3.2, 3.3, and 3.4 hold with $\Sigma = I_{d^*}$. Then for any α such that $F_T(\cdot)$ is continuous at its α quantile, we have:*
(i) $\liminf_{n\to\infty} \mathbb{P}(\Theta_I \subseteq \widehat{\Theta}_\alpha) \geq \alpha$*;*
(ii) If $T = \mathbb{R}^{d^*}$ then: $\lim_{n\to\infty} \mathbb{P}(\Theta_I \subseteq \widehat{\Theta}_\alpha) = \alpha$*.*

A key step in the proof of Theorem 3.1 is the following new Bernstein-von Mises type result (Lemma 3.1) for the posterior distribution of the QLR. Let $\mathbb{P}_{Z|\mathbf{X}_n}$ be the distribution of a random vector Z that is $N(0, I_{d^*})$ (conditional on the data). Recall that $\mathbb{V}_n$ is a measurable function of the data. Let $T - \mathbb{V}_n$ denote the cone T translated to have vertex at $-\mathbb{V}_n$. Let $\mathbf{T}^\perp$ denote the orthogonal projection onto the polar cone of T.[18]

Lemma 3.1. *Let Assumptions 3.1, 3.2 and 3.3 hold. Then:*

$$\sup_z \left| \Pi_n\big(\{\theta : Q_n(\theta) \leq z\} \big| \mathbf{X}_n\big) - \mathbb{P}_{Z|\mathbf{X}_n}\Big(\|Z\|^2 \leq z + \|\mathbf{T}^\perp \mathbb{V}_n\|^2 \Big| Z \in T - \mathbb{V}_n\Big) \right| = o_{\mathbb{P}}(1)\,. \quad (16)$$

And hence we have:
(i) If $T \subsetneq \mathbb{R}^{d^*}$ then: $\Pi_n\big(\{\theta : Q_n(\theta) \leq z\} \big| \mathbf{X}_n\big) \leq F_T(z)$ *for all $z \geq 0$.*
(ii) If $T = \mathbb{R}^{d^*}$ then: $\sup_z \left| \Pi_n\big(\{\theta : Q_n(\theta) \leq z\} \,\big|\, \mathbf{X}_n\big) - F_{\chi^2_{d^*}}(z) \right| = o_{\mathbb{P}}(1)$*.*

Note that Lemma 3.1 does not require the generalized information equality $\Sigma = I_{d^*}$ to hold. Therefore, regardless whether a partially-identified model is correctly specified or not, the posterior distribution of the QLR statistic asymptotically (first-order) stochastically dominates F_T when T is a closed convex cone and is asymptotically $\chi^2_{d^*}$ when $T = \mathbb{R}^{d^*}$. This lemma extends the known Bernstein-von Mises theorems for possibly misspecified likelihood models with point-identified root-n asymptotically normally estimable parameters (see, e.g., Kleijn and van der Vaart (2012) and the references therein) to allow for other models with failure of $\Sigma = I_{d^*}$, with partially-identified parameters and/or parameters on a boundary.

Together with Assumption 3.4, Lemma 3.1 implies that our MCMC CS $\widehat{\Theta}_\alpha$ (Procedure 1) is always a well-defined Bayesian credible set for Θ_I regardless whether $\Sigma = I_{d^*}$ holds or not. But, Theorem 3.1 requires $\Sigma = I_{d^*}$ so that our MCMC CS $\widehat{\Theta}_\alpha$ will have a correct frequentist coverage

[18] The orthogonal projection $\mathbf{T}v$ of any vector $v \in \mathbb{R}^{d^*}$ onto a closed convex cone $T \subseteq \mathbb{R}^{d^*}$ is the unique solution to $\inf_{t\in T} \|t - v\|^2$. The polar cone of T is $T^o = \{s \in \mathbb{R}^{d^*} : s't \leq 0 \text{ for all } t \in T\}$ which is also closed and convex. Moreau's decomposition theorem gives $v = \mathbf{T}v + \mathbf{T}^\perp v$ with $\|v\|^2 = \|\mathbf{T}v\|^2 + \|\mathbf{T}^\perp v\|^2$. If $T = \mathbb{R}^{d^*}$ then $\mathbf{T}v = v$, $T^o = \{0\}$ and $\mathbf{T}^\perp v = 0$ for any $v \in \mathbb{R}^{d^*}$. See Chapter A.3.2 of Hiriart-Urruty and Lemaréchal (2001).

probability (for Θ_I).[19] This is because the asymptotic distribution of $\sup_{\theta \in \Theta_I} Q_n(\theta)$ is F_T only under $\Sigma = I_{d^*}$. It follows that, with an optimally weighted criterion, $\widehat{\Theta}_\alpha$ will be asymptotically exact (for Θ_I) when $T = \mathbb{R}^{d^*}$, and asymptotically valid but could be conservative when T is a cone, where the conservativeness of $\widehat{\Theta}_\alpha$ will depend on the shape of T.

Remark 3. *Theorem 3.1 is still applicable to misspecified, separable partially identified likelihood models. For such models we can rewrite the density as $p(\cdot;\theta) = q(\cdot;\tilde{\gamma}(\theta))$ where $\tilde{\gamma}$ is an identifiable reduced-form parameter (see Section 4.1.1 below). Under misspecification the identified set is $\Theta_I = \{\theta : \tilde{\gamma}(\theta) = \tilde{\gamma}^*\}$ where $\tilde{\gamma}^*$ is the unique value of $\tilde{\gamma}$ that minimizes $P_0 \log(p_0(\cdot)/q(\cdot;\tilde{\gamma}))$ (the Kullback-Leibler divergence from the true DGP p_0). Following the insight of Müller (2013), we could base our inference on the sandwich log-likelihood function:*

$$L_n(\theta) = -\frac{1}{2n} \sum_{i=1}^{n} (\tilde{\gamma}(\theta) - \hat{\gamma})'(\widehat{\Sigma}_S)^{-1}(\tilde{\gamma}(\theta) - \hat{\gamma})$$

where $\hat{\gamma}$ approximately maximizes $\frac{1}{n}\sum_{i=1}^{n} \log q(X_i;\tilde{\gamma})$ and $\widehat{\Sigma}_S$ is the sandwich covariance matrix estimator for $\hat{\gamma}$.

3.1.1 Models with singularities

In this subsection we consider (possibly) partially identified models with singularities.[20] In identifiable parametric models $\{P_\theta : \theta \in \Theta\}$, the standard notion of differentiability in quadratic mean requires that the mass of the part of P_θ that is singular with respect to the true distribution $P_0 = P_{\theta_0}$ vanishes faster than $\|\theta - \theta_0\|^2$ as $\theta \to \theta_0$ (Le Cam and Yang, 1990, section 6.2). If this condition fails then the log likelihood will not be locally quadratic at θ_0. By analogy with the identifiable case, we say a non-identifiable model has a singularity if it does not admit a local quadratic approximation (in the reduced-form reparameterization) like that in Assumption 3.2(i). One such an example is the missing data model under identification (see Subsection 4.1.1 below).

To allow for partially identified models with singularities, we first generalize the notion of the local reduced-form reparameterization to be of the form $\theta \mapsto (\gamma(\theta), \gamma_\perp(\theta))$ from Θ_I^N into $\Gamma \times \Gamma_\perp$ where $\Gamma \subseteq \mathbb{R}^{d^*}$ and $\Gamma_\perp \subseteq \mathbb{R}^{\dim(\gamma_\perp)}$ with $(\gamma(\theta), \gamma_\perp(\theta)) = 0$ if and only if $\theta \in \Theta_I$. The following regularity conditions generalize Assumptions 3.2 and 3.3 to allow for singularity.

Assumption 3.2′ *(Local quadratic approximation with singularity)*

[19] This is consistent with the fact that the percentile MCMC CS also needs $\Sigma = I_{d^*}$ in order to have a correct frequentist coverage for a point-identified scalar parameter (see, e.g., Chernozhukov and Hong (2003)), Robert and Casella (2004) and others.

[20] Such models are also referred to as non-regular models or models with non-regular parameters.

(i) There exist sequences of random variables ℓ_n and $\mathbb{R}^{d^*}$-valued random vectors $\mathbb{V}_n$ (both measurable in $\mathbf{X}_n$), and a sequence of measurable functions $f_{n,\perp} : \Gamma_\perp \to \mathbb{R}_+$ with $f_{n,\perp}(0) = 0$, such that:

$$\sup_{\theta \in \Theta_{osn}} \left| nL_n(\theta) - \left(\ell_n - \frac{1}{2} \|\sqrt{n}\gamma(\theta)\|^2 + (\sqrt{n}\gamma(\theta))'\mathbb{V}_n - f_{n,\perp}(\gamma_\perp(\theta)) \right) \right| = o_{\mathbb{P}}(1) \tag{17}$$

with $\sup_{\theta \in \Theta_{osn}} \|(\gamma(\theta), \gamma_\perp(\theta))\| \to 0$, $\mathbb{V}_n \rightsquigarrow N(0, \Sigma)$ as $n \to \infty$;
(ii) $\{(\gamma(\theta), \gamma_\perp(\theta)) : \theta \in \Theta_{osn}\} = \{\gamma(\theta) : \theta \in \Theta_{osn}\} \times \{\gamma_\perp(\theta) : \theta \in \Theta_{osn}\}$;
(iii) The sets $K_{osn} = \{\sqrt{n}\gamma(\theta) : \theta \in \Theta_{osn}\}$ cover a closed convex cone $T \subseteq \mathbb{R}^{d^*}$.

Let Π_{Γ^*} denote the image of the measure Π under the map $\Theta_I^N \ni \theta \mapsto (\gamma(\theta), \gamma_\perp(\theta))$. Let $B_r^* \subset \mathbb{R}^{d^* + \dim(\gamma_\perp)}$ denote a ball of radius r centered at the origin.

Assumption 3.3' *(Prior with singularity)*
(i) $\int_\Theta e^{nL_n(\theta)} \, d\Pi(\theta) < \infty$ almost surely
(ii) Π_{Γ^*} has a continuous, strictly positive density π_{Γ^*} on $B_\delta^* \cap (\Gamma \times \Gamma_\perp)$ for some $\delta > 0$.

Discussion of Assumptions: Assumption 3.2'(i)(iii) is generalization of Assumption 3.2 to the singular case. Assumption 3.2'(i)(ii) implies that the peak of the likelihood does not concentrate on sets of the form $\{\theta : f_{n,\perp}(\gamma_\perp(\theta)) > \epsilon > 0\}$. Recently, Bochkina and Green (2014) established a Bernstein-von Mises result for *identifiable* parametric likelihood models with singularities. They assume that the likelihood is locally quadratic in some parameters and locally linear in others (similar to Assumption 3.2'(i)), and also assume the local parameter space satisfies conditions similar to our Assumption 3.2'(ii)(iii). Finally, Assumption 3.3' generalizes Assumption 3.3 to the singular case. We impose no further restrictions on the set $\{\gamma_\perp(\theta) : \theta \in \Theta_I^N\}$.

Theorem 3.2. *Let Assumptions 3.1, 3.2', 3.3', and 3.4 hold with $\Sigma = I_{d^*}$. Then for any α such that $F_T(\cdot)$ is continuous at its α quantile, we have:*

$$\liminf_{n \to \infty} \mathbb{P}(\Theta_I \subseteq \widehat{\Theta}_\alpha) \geq \alpha \,.$$

For non-singular models Theorem 3.1 establishes that $\widehat{\Theta}_\alpha$ is asymptotically valid for Θ_I, with asymptotically exact coverage when the tangent set T is linear and can be conservative when T is a closed convex cone. For singular models Theorem 3.2 shows that $\widehat{\Theta}_\alpha$ is still asymptotically valid for Θ_I but can be conservative even when T is linear.[21] When applied to the missing data example, Theorems 3.1 and 3.2 imply that $\widehat{\Theta}_\alpha$ for Θ_I is asymptotically exact under partial identification but conservative under point identification; see Section 4.1.1 below for details.

[21]It might be possible to establish asymptotically exact coverage of $\widehat{\Theta}_\alpha$ for Θ_I in singular models where the singular part $f_{n,\perp}(\gamma_\perp(\theta))$ in Assumption 3.2' possesses some extra structures.

3.2 Coverage properties of $\widehat{M}_\alpha$ for M_I

In this section we present conditions under which the CS $\widehat{M}_\alpha$ has correct coverage for the set M_I. Recall that $\mu : \Theta \to M \subset \mathbb{R}^k$ is a known continuous mapping with $1 \leq k < \dim(\theta)$, $M = \{m = \mu(\theta) : \theta \in \Theta\}$, $\mu^{-1}(m) = \{\theta \in \Theta : \mu(\theta) = m\}$, and $\Delta(\theta) = \{\bar{\theta} \in \Theta : L(\bar{\theta}) = L(\theta)\}$. Then $\Theta_I = \Delta(\theta)$ for any $\theta \in \Theta_I$ and $M_I = \{\mu(\theta) : \theta \in \Theta_I\} = \mu(\Delta(\theta))$ for any $\theta \in \Theta_I$.

Define the profile quasi-likelihood for the *set* $\mu(\Delta(\theta)) \subset M$ as:

$$PL_n(\Delta(\theta)) = \inf_{m \in \mu(\Delta(\theta))} \sup_{\bar{\theta} \in \mu^{-1}(m)} L_n(\bar{\theta}) .$$

Since we aim at covering the identified set M_I in a possibly partially identified model, this definition of the profile quasi-likelihood is for a set, and is different from the usual definition (8) of the profile quasi-likelihood for a *point* $m \in M$. Note that $PL_n(\Delta(\theta))$ is defined in the same way as that of the profile quasi-likelihood for the set M_I (see (9)):

$$PL_n(\Delta(\theta)) = PL_n(\Theta_I) = \inf_{m \in M_I} \sup_{\bar{\theta} \in \mu^{-1}(m)} L_n(\bar{\theta}) \quad \text{for all } \theta \in \Theta_I.$$

The profile QLR for the *set* $\mu(\Delta(\theta)) \subset M$ is defined analogously:

$$PQ_n(\Delta(\theta)) = 2n[L_n(\hat{\theta}) - PL_n(\Delta(\theta))] \; = \sup_{m \in \mu(\Delta(\theta))} \inf_{\bar{\theta} \in \mu^{-1}(m)} Q_n(\bar{\theta}) .$$

where $Q_n(\bar{\theta}) = 2n[L_n(\hat{\theta}) - L_n(\bar{\theta})]$ is as defined in (6).

Recall that $\Theta_{osn} \subset \Theta_I^N$ for all n sufficiently large. For $\theta \in \Theta_I^N$, the set $\Delta(\theta)$ can be equivalently expressed as the set $\{\bar{\theta} \in \Theta_I^N : \gamma(\bar{\theta}) = \gamma(\theta)\}$. Also $M_I = \{\mu(\theta) : \gamma(\theta) = 0\}$.

Assumption 3.5. *(Profile QL)*
There exists a measurable function $f : \mathbb{R}^{d^*} \to \mathbb{R}_+$ such that:

$$\sup_{\theta \in \Theta_{osn}} \left| nPL_n(\Delta(\theta)) - \left(\ell_n + \frac{1}{2}\|\mathbb{V}_n\|^2 - \frac{1}{2} f\left(\mathbb{V}_n - \sqrt{n}\gamma(\theta)\right) \right) \right| = o_{\mathbb{P}}(1)$$

with $\mathbb{V}_n$ and γ from Assumption 3.2 or 3.2'.

We also replace Assumption 3.4 by a version appropriate for the profiled case. Let $\xi_{n,\alpha}^{post,p}$ denote the α quantile of the profile QLR $PQ_n(\Delta(\theta))$ under the posterior distribution Π_n, and $\xi_{n,\alpha}^{mc,p}$ be given in Remark 2.

Assumption 3.6. *(MC convergence)*

$\xi^{mc,p}_{n,\alpha} = \xi^{post,p}_{n,\alpha} + o_{\mathbb{P}}(1)$.

Discussion of Assumptions: Assumption 3.5 imposes some structure on the profile QL statistic for M_I over the local neighborhood Θ_{osn}. It effectively deals with models for which the profile QLR for M_I is of the form:

$$PQ_n(\Delta(\theta)) = f(\mathbb{V}_n) - \|\mathbf{T}^{\perp}\mathbb{V}_n\|^2 + o_{\mathbb{P}}(1) \quad \text{for each } \theta \in \Theta_I \tag{18}$$

where $f : \mathbb{R}^{d^*} \to \mathbb{R}_+$ is a measurable function satisfying $f(v) \geq \|\mathbf{T}^{\perp}v\|^2$ for $v \in \mathbb{R}^{d^*}$. The precise functional form of f depends on the local reparameterization γ as well as the mapping μ. When M_I is a singleton then equation (18) is typically satisfied with $f(v) = \inf_{t\in T_1} \|v - t\|^2$ where $T_1 = \mathbb{R}^{d_1^*} \subset T = \mathbb{R}^{d^*}$ (i.e., $d_1^* < d^*$) and the QLR statistic is $\chi^2_{d^*-d_1^*}$ asymptotically. For a non-singleton set M_I, f in equation (18) could be of the form:

$$f(v) = f_0\left(\inf_{t\in T_1} \|v-t\|^2, \dots, \inf_{t\in T_J} \|v-t\|^2\right) + \inf_{t\in T} \|v-t\|^2$$

where $f_0 : \mathbb{R}^J \to \mathbb{R}_+$ and $T_1, \dots, T_J$ are closed cones in $\mathbb{R}^{d^*}$, and the profile QLR statistic could be asymptotically mixtures of χ^2 random variables with different degrees of freedom (i.e. chi-bar-squared random variables) as well as maxima and minima of mixtures of χ^2 random variables. See Section 4 for verification of Assumption 3.5 (or equation (18)) in missing data and moment inequality examples. Note that the existence of such a f is merely a proof device, and one does not need to know its precise expression in the implementation of our MC CS $\widehat{M}_\alpha$ for M_I. Finally, Assumption 3.6 requires that the distribution of the profile QLR statistic computed from the MC chain well approximates the posterior distribution of the profile QLR statistic.

The next theorem is an important consequence of Lemma F.5 (a new BvM type result in Appendix F) for the posterior distribution of the profile QLR for M_I.

Theorem 3.3. *Let Assumptions 3.1, 3.2, 3.3, 3.5, and 3.6 or 3.1, 3.2', 3.3', 3.5, and 3.6 hold with $\Sigma = I_{d^*}$ and $T = \mathbb{R}^{d^*}$, and let the distribution of $f(Z)$ (where $Z \sim N(0, I_{d^*})$) be continuous at its α quantile. Then: $\lim_{n\to\infty} \mathbb{P}(M_I \subseteq \widehat{M}_\alpha) = \alpha$.*

Theorem 3.3 shows that, as long as the tangent set T is linear, our CSs $\widehat{M}_\alpha$ for M_I can have asymptotically exact coverage even when the model is singular. For example, in the missing data example, Theorem 3.3 implies that $\widehat{M}_\alpha$ for M_I is asymptotically exact irrespective of whether the model is point-identified or not; see Subsection 4.1.1 below for details.

3.3 Coverage properties of $\widehat{M}_\alpha^\chi$ for M_I

This section presents one sufficient condition for validity of the CS $\widehat{M}_\alpha^\chi$ for M_I (Procedure 3).

Assumption 3.7. *(Profile QLR, χ^2 bound)*
$PQ_n(\Delta(\theta)) \rightsquigarrow f(Z) = \inf_{t \in T_1} \|Z - t\|^2 \vee \inf_{t \in T_2} \|Z - t\|^2$ *for all* $\theta \in \Theta_I$, *where* $Z \sim N(0, I_{d^*})$ *for some* $d^* \geq 1$ *and* T_1 *and* T_2 *are closed half-spaces in* $\mathbb{R}^{d^*}$ *with supporting hyperplanes that pass through the origin.*

Note that Assumption 3.7 places additional structure on the function f in Assumption 3.5 or in equation (18).

Theorem 3.4. *Let Assumption 3.7 hold and let the distribution of $f(Z)$ be continuous at its α quantile. Then:* $\liminf_{n\to\infty} \mathbb{P}(M_I \subseteq \widehat{M}_\alpha^\chi) \geq \alpha$.

The exact distribution of $f(Z)$ depends on the geometry of T_1 and T_2. We show in the proof of Theorem 3.4 that the worst-case coverage (i.e., the case in which asymptotic coverage of $\widehat{M}_\alpha^\chi$ will be most conservative) will occur when the polar cones of T_1 and T_2 are orthogonal, in which case $f(Z)$ has the mixture distribution $W^* := \frac{1}{4}\delta_0 + \frac{1}{2}\chi^2_1 + \frac{1}{4}(\chi^2_1 \cdot \chi^2_1)$ where δ_0 is a point mass at zero and $\chi^2_1 \cdot \chi^2_1$ is the distribution of the product of two independent χ^2_1 random variables. The quantiles of the distribution of $f(Z)$ are continuous in α for all $\alpha > \frac{1}{4}$. For all configurations of T_1 and T_2, the distribution of $f(Z)$ (first-order) stochastically dominates F_{W^*} and is (first-order) stochastically dominated by $F_{\chi^2_1}$ (i.e., $F_{W^*}(w) \geq F_{f(Z)}(w) \geq F_{\chi^2_1}(w)$). Notice that this is different from the usual chi-bar-squared case encountered when testing whether a parameter belongs to the identified set on the basis of finitely many moment inequalities (Rosen, 2008).

To get an idea of the degree of conservativeness of $\widehat{M}_\alpha^\chi$, consider the class of models satisfying conditions for Theorem 3.4. Figure 3 plots the asymptotic coverage of $\widehat{M}_\alpha$ and $\widehat{M}_\alpha^\chi$ against nominal coverage for models in this class where $\widehat{M}_\alpha^\chi$ is most conservative (i.e., the worst-case coverage). For each model in this class, the asymptotic coverage of $\widehat{M}_\alpha$ and $\widehat{M}_\alpha^\chi$ is between the nominal coverage and the worst-case coverage. As can be seen, the coverage of $\widehat{M}_\alpha$ is exact at all levels $\alpha \in (0,1)$ for which the distribution of the profile QLR is continuous at its α quantile, as predicted by Lemma 2.2. On the other hand, $\widehat{M}_\alpha^\chi$ is asymptotically conservative, but the level of conservativeness decreases as α increases towards one. Indeed, for levels of α in excess of 0.85 the level of conservativeness is negligible.

The following proposition presents a set of sufficient conditions for Assumption 3.7.

Proposition 3.1. *Let the following hold:*
(i) $\inf_{m \in M_I} \sup_{\theta \in \mu^{-1}(m)} L_n(\theta) = \min_{m \in \{\underline{m}, \overline{m}\}} \sup_{\theta \in \mu^{-1}(m)} L_n(\theta) + o_{\mathbb{P}}(n^{-1})$;

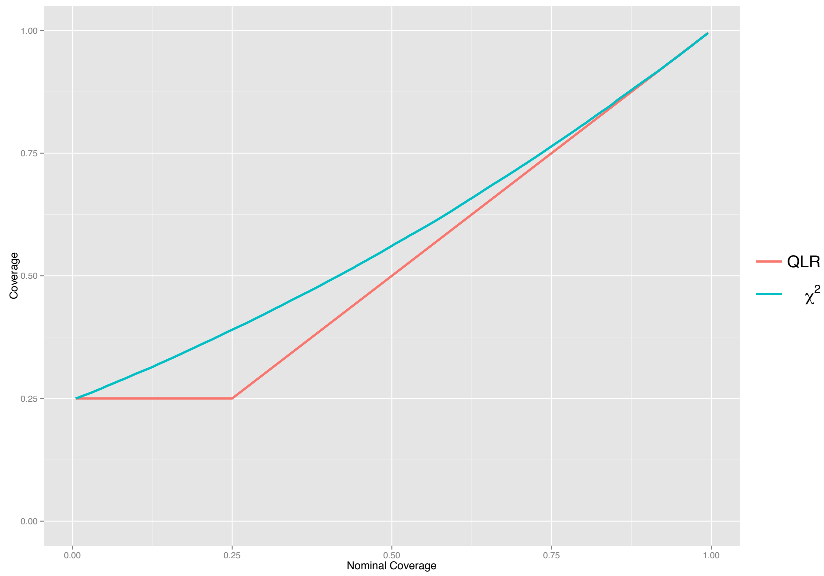

Figure 3: Comparison of asymptotic coverage of $\widehat{M}_\alpha$ (Profile QLR – solid kinked line) and of $\widehat{M}_\alpha^\chi$ (χ^2 – dashed curved line) with their nominal coverage for models where $\widehat{M}_\alpha^\chi$ is valid for M_I but most conservative.

(ii) for each $m \in \{\underline{m}, \overline{m}\}$ there exists a sequence of sets $(\Gamma_{m,osn})_{n\in\mathbb{N}}$ with $\Gamma_{m,osn} \subseteq \Gamma$ for each n and a closed convex cone $T_m \subseteq \mathbb{R}^{d^*}$ with positive volume, such that:

$$\sup_{\theta\in\mu^{-1}(m)} nL_n(\theta) = \sup_{\gamma\in\Gamma_{m,osn}} \left(\ell_n + \frac{1}{2}\|\mathbb{V}_n\|^2 - \frac{1}{2}\|\sqrt{n}\gamma - \mathbb{V}_n\|^2\right) + o_{\mathbb{P}}(1)$$

and $\inf_{\gamma\in\Gamma_{m,osn}} \|\sqrt{n}\gamma - \mathbb{V}_n\|^2 = \inf_{t\in T_m} \|t - \mathbb{V}_n\|^2 + o_{\mathbb{P}}(1)$;
(iii) Assumptions 3.1(i), 3.2 or 3.2' hold with $\Sigma = I_{d^*}$;
(iv) $T = \mathbb{R}^{d^*}$ and both $T_{\underline{m}}$ and $T_{\overline{m}}$ are halfspaces in $\mathbb{R}^{d^*}$.
Then: Assumption 3.7 holds.

Suppose that $M_I = [\underline{m}, \overline{m}]$ with $-\infty < \underline{m} \le \overline{m} < \infty$ (which is the case whenever Θ_I is connected and bounded). If $\sup_{\theta\in\mu^{-1}(m)} L_n(\theta)$ is strictly concave in m then condition (i) of Proposition 3.1 holds. The remaining conditions are then easy to verify.

Since empirical papers typically report a confidence set for scalar parameters, Theorem 3.4 will be very useful in applied work. One could generalize $\widehat{M}_\alpha^\chi$ to allow for quantiles of χ^2_d with higher degrees of freedom $d \in (1, \dim(\theta))$, but it might be difficult to provide sufficient condition to establish result like Theorem 3.4.

4 Sufficient conditions and Examples

This section provides sufficient conditions for Assumption 3.2 in general partially identified likelihood and GMM models with i.i.d. data. We also verify key regularity conditions (Assumptions 3.1(ii), 3.2 (or 3.2'), 3.3, 3.5) in examples. In what follows we use standard empirical process notation (van der Vaart and Wellner, 1996), namely $P_0 g$ denotes the expectation of $g(X_i)$ under the true probability distribution P_0, $\mathbb{P}_n g = n^{-1}\sum_{i=1}^n g(X_i)$ denotes expectation of $g(X_i)$ under the empirical distribution, and $\mathbb{G}_n g = \sqrt{n}(\mathbb{P}_n - P_0)g$ denotes the empirical process.

4.1 Partially identified likelihood models

Consider a parametric likelihood model $\mathcal{P} = \{p(\cdot;\theta) : \theta \in \Theta\}$ where each $p(\cdot;\theta)$ is a probability density with respect to a common σ-finite dominating measure λ. Let $p_0 \in \mathcal{P}$ be the true DGP, $D_{KL}(p(\cdot)||q(\cdot;\theta))$ be the Kullback-Leibler divergence, and $h(p,q)^2 = \int(\sqrt{p} - \sqrt{q})^2 \,\mathrm{d}\lambda$ denote the squared Hellinger distance between two densities p and q. Then the identified set is $\Theta_I = \{\theta \in \Theta : D_{KL}(p_0(\cdot)||p(\cdot;\theta)) = 0\} = \{\theta \in \Theta : h(p_0(\cdot), p(\cdot;\theta)) = 0\}$.

4.1.1 Over-parameterized likelihood models

For a large class of partially identified parametric likelihood models $\mathcal{P} = \{p(\cdot;\theta) : \theta \in \Theta\}$, there exists a measurable function $\tilde{\gamma} : \Theta \to \widetilde{\Gamma} \subset \mathbb{R}^{d^*}$ for some possibly unknown $d^* \in [1, +\infty)$, such that $p(\cdot;\theta) = q(\cdot;\tilde{\gamma}(\theta))$ for each $\theta \in \Theta$ and some densities $\{q(\cdot;\tilde{\gamma}(\theta)) : \tilde{\gamma} \in \widetilde{\Gamma}\}$. In this case we say that the model $\mathcal{P}$ is over-parameterized and admits a (global) reduced-form reparameterization. The reparameterization is assumed to be identifiable, i.e. $D_{KL}(q(\cdot;\tilde{\gamma}_0)||q(\cdot;\tilde{\gamma})) > 0$ for any $\tilde{\gamma} \neq \tilde{\gamma}_0$. Without loss of generality, we may translate the parameter space $\widetilde{\Gamma}$ so that the true density $p_0(\cdot) \equiv q(\cdot;\tilde{\gamma}_0)$ with $\tilde{\gamma}_0 = 0$. The identified set is $\Theta_I = \{\theta \in \Theta : \tilde{\gamma}(\theta) = 0\}$.

In the following we let $\ell_{\tilde{\gamma}}(x) := \log q(x;\tilde{\gamma})$, $\dot{\ell}_{\tilde{\gamma}} = \frac{\partial \ell_{\tilde{\gamma}}}{\partial \tilde{\gamma}}$ and $\ddot{\ell}_{\tilde{\gamma}} = \frac{\partial^2 \ell_{\tilde{\gamma}}}{\partial \tilde{\gamma}\partial \tilde{\gamma}'}$. And let $\mathbb{I}_0 := P_0(\dot{\ell}_{\tilde{\gamma}_0}\dot{\ell}'_{\tilde{\gamma}_0})$ denote the variance of the true score.

Proposition 4.1. *Suppose that $\{q(\cdot;\tilde{\gamma}) : \tilde{\gamma} \in \widetilde{\Gamma}\}$ satisfies the following regularity conditions:*
(a) $X_1, \ldots, X_n$ are i.i.d. drawn from $p_0(\cdot) = q(\cdot;0) \in \{q(\cdot;\tilde{\gamma}) : \tilde{\gamma} \in \widetilde{\Gamma}\}$, where $\widetilde{\Gamma}$ is a compact subset of $\mathbb{R}^{d^*}$;
(b) there exists an open neighborhood $U \subset \widetilde{\Gamma}$ of $\tilde{\gamma}_0 = 0$ upon which $\ell_{\tilde{\gamma}}(x)$ is strictly positive and twice continuously differentiable for each x, with $\sup_{\tilde{\gamma} \in U} \|\ddot{\ell}_{\tilde{\gamma}}(x)\| \leq \bar{\ell}(x)$ for some $\bar{\ell} : \mathscr{X} \to \mathbb{R}$ with $P_0(\bar{\ell}) < \infty$; and $\mathbb{I}_0$ is finite positive definite.
Then: there exists a sequence $(r_n)_{n\in\mathbb{N}}$ with $r_n \to \infty$ and $r_n/\sqrt{n} = o(1)$ as $n \to \infty$ such that

Assumption 3.2 holds for the average log-likelihood (1) over $\Theta_{osn} := \{\theta \in \Theta : \|\gamma(\theta)\| \leq r_n/\sqrt{n}\}$ *with* $\gamma(\theta) = \mathbb{I}_0^{1/2}\tilde{\gamma}(\theta)$, $\mathbb{V}_n = \mathbb{I}_0^{-1/2}\mathbb{G}_n(\dot{\ell}_{\tilde{\gamma}_0}) \rightsquigarrow N(0, I_{d^*})$, *and* $T = \mathbb{R}^{d^*}$.
If, in addition:
(c) π_Γ *is continuous and uniformly bounded away from zero and infinity on* $\Gamma = \{\gamma = \mathbb{I}_0^{1/2}\tilde{\gamma} : \tilde{\gamma} \in \widetilde{\Gamma}\}$*;*
(d) there exists $\alpha > 0$ *such that* $P_0 \log(p_0(\cdot)/q(\cdot;\tilde{\gamma})) \lesssim \|\tilde{\gamma}\|^{2\alpha}$, $P_0[\log(q(\cdot;\tilde{\gamma})/p_0(\cdot))]^2 \lesssim \|\tilde{\gamma}\|^{2\alpha}$, *and* $h(q(\cdot;\tilde{\gamma}_1), q(\cdot;\tilde{\gamma}_2)) \asymp \|\tilde{\gamma}_1 - \tilde{\gamma}_2\|^\alpha$ *all hold on* U*.*
Then: Assumption 3.1(ii) also holds.

Proposition 4.1 shows that Assumption 3.2 holds under conventional smoothness and identification conditions on the reduced-form likelihood. The condition of twice continuous differentiability of the log-likelihood can be weakened by substituting Hellinger differentiability conditions. Sufficient conditions can also be tailored to Markov processes, including DSGE models with latent Markov state variables, and general likelihood-based time series models (see, e.g., Hallin, van den Akker, and Werker (2015)).

Example 1: missing data model in Subsection 2.4.1

We revisit the missing data example in Subsection 2.4.1, where the parameter space for $\theta = (\mu, \beta, \rho)$ is Θ given in (12). The identified set for θ is Θ_I given in (13), and the identified set for $\mu_0 := \mathbb{E}[Y_i]$ is $M_I = [\kappa_{11}, \kappa_{11} + \kappa_{00}]$.

Inference under partial identification: Consider the case in which the model is partially identified (i.e. $0 < \kappa_{00} < 1$). The likelihood of the i-th observation $(D_i, Y_i D_i) = (d, yd)$ is

$$p(d, yd; \theta) = [\kappa_{11}(\theta)]^{yd}[1 - \kappa_{11}(\theta) - \kappa_{00}(\theta)]^{d-yd}[\kappa_{00}(\theta)]^{1-d} = q(d, yd; \tilde{\gamma}(\theta))$$

where the reduced-form reparameterization is:

$$\tilde{\gamma}(\theta) = \begin{pmatrix} \kappa_{11}(\theta) - \kappa_{11} \\ \kappa_{00}(\theta) - \kappa_{00} \end{pmatrix}$$

with $\widetilde{\Gamma} = \{\tilde{\gamma}(\theta) : \theta \in \Theta\} = \{(k_{11} - \kappa_{11}, k_{00} - \kappa_{00}) : (k_{11}, k_{00}) \in [0,1]^2, 0 \leq k_{11} \leq 1 - k_{00}\}$. Conditions (a)-(b) of Proposition 4.1 hold. Hence Assumption 3.2 is satisfied with $\gamma(\theta) = \mathbb{I}_0^{1/2}\tilde{\gamma}(\theta)$ where:

$$\mathbb{I}_0 = \begin{bmatrix} \frac{1}{\kappa_{11}} + \frac{1}{1-\kappa_{11}-\kappa_{00}} & \frac{1}{1-\kappa_{11}-\kappa_{00}} \\ \frac{1}{1-\kappa_{11}-\kappa_{00}} & \frac{1}{\kappa_{00}} + \frac{1}{1-\kappa_{11}-\kappa_{00}} \end{bmatrix}$$

and

$$\mathbb{V}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{I}_0^{-1/2} \begin{pmatrix} \frac{y_i d_i}{\kappa_{11}} - \frac{d_i - y_i d_i}{1-\kappa_{11}-\kappa_{00}} \\ \frac{1-d_i}{\kappa_{00}} - \frac{d_i - y_i d_i}{1-\kappa_{11}-\kappa_{00}} \end{pmatrix} \rightsquigarrow N(0, I_2)$$

and the tangent cone is $T = \mathbb{R}^2$. A flat prior on Θ in (12) induces a flat prior on Γ, which verifies Condition (c) of Proposition 4.1 and Assumption 3.3. Therefore, Theorem 3.1 implies that our MC CSs for Θ_I will have asymptotically exact coverage.

Now consider CSs for $M_I = [\kappa_{11}, \kappa_{11} + \kappa_{00}]$. Note that $\mu^{-1}(m) = \{m\} \times \{(\beta, \rho) \in [0,1]^2 : 0 \leq m - \beta(1-\rho) \leq \rho\}$. By concavity in m, the profile log-likelihood for M_I is:

$$PL_n(\Delta(\theta)) = \min_{m \in \{\kappa_{11}, \kappa_{11}+\kappa_{00}\}} \sup_{\bar\theta \in \mu^{-1}(m)} \mathbb{P}_n \log(p(\cdot; \bar\theta)) \quad \text{for all } \theta \in \Theta_I\,.$$

Rewriting the maximization problem in terms of the reduced-form probabilities:

$$\sup_{\bar\theta \in \mu^{-1}(m)} \mathbb{P}_n \log(p(\cdot;\bar\theta)) = \sup_{\substack{0 \leq k_{11} \leq m \\ m \leq k_{11} + k_{00} \leq 1}} \mathbb{P}_n \Big(yd \log k_{11} + (d - yd)\log(1 - k_{11} - k_{00}) + (1-d)\log k_{00} \Big)\,. \tag{19}$$

at $m = \kappa_{11}$ and $m = \kappa_{11} + \kappa_{00}$. The local parameter spaces for problem (19) at $m = \kappa_{11}$ and $m = \kappa_{11} + \kappa_{00}$ are sketched in Figure 4. Let $\gamma = (\gamma_1, \gamma_2) = (k_{11} - \kappa_{11}, k_{00} - \kappa_{00})$ and let:

$$\begin{aligned}
T_1 &= \bigcup_{n \geq 1} \Big\{ \sqrt{n}\mathbb{I}_0^{1/2}\gamma \,:\, -\kappa_{11} \leq \gamma_1 \leq 0,\ -\kappa_{00} \leq \gamma_1 + \gamma_2 \leq 1 - \kappa_{11} - \kappa_{00},\ \|\gamma\|^2 \leq r_n^2/n \Big\} \\
T_2 &= \bigcup_{n \geq 1} \Big\{ \sqrt{n}\mathbb{I}_0^{1/2}\gamma \,:\, -\kappa_{11} \leq \gamma_1 \leq \kappa_{00},\ 0 \leq \gamma_1 + \gamma_2 \leq 1 - \kappa_{11} - \kappa_{00},\ \ \|\gamma\|^2 \leq r_n^2/n \Big\}
\end{aligned}$$

where r_n is from Proposition 4.1. It follows that for all $\theta \in \Theta_I$:

$$\begin{aligned}
nPL_n(\Delta(\theta)) &= n\mathbb{P}_n \log p_0 + \frac{1}{2}\|\mathbb{V}_n\|^2 - \frac{1}{2}\Big(\inf_{t \in T_1} \|\mathbb{V}_n - t\|^2\Big) \vee \Big(\inf_{t \in T_2} \|\mathbb{V}_n - t\|^2\Big) + o_{\mathbb{P}}(1) \\
PQ_n(\Delta(\theta)) &= \Big(\inf_{t \in T_1} \|\mathbb{V}_n - t\|^2\Big) \vee \Big(\inf_{t \in T_2} \|\mathbb{V}_n - t\|^2\Big) + o_{\mathbb{P}}(1)\,.
\end{aligned}$$

Thus both equation (18) and Assumption 3.7 hold with $f : \mathbb{R}^2 \to \mathbb{R}_+$ given by

$$f(v) = (\inf_{t \in T_1} \|v - t\|^2) \vee (\inf_{t \in T_2} \|v - t\|^2), \tag{20}$$

where T_1 and T_2 are halfspaces in $\mathbb{R}^2$. Theorem 3.4 implies that the CS $\widehat{M}_\alpha^\chi$ is asymptotically valid (but conservative) for M_I.

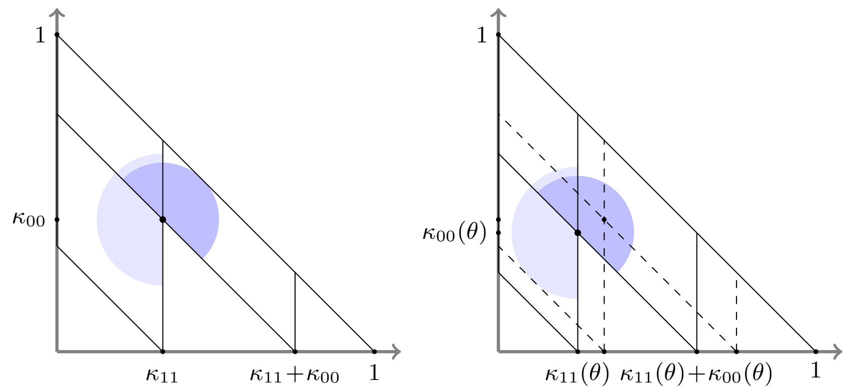

Figure 4: Local parameter spaces for the profile LR statistic for M_I. Left panel: the lightly shaded region is for problem (19) at $m = \kappa_{11}$ and the darker shaded region is for problem (19) at $m = \kappa_{11} + \kappa_{00}$. Right panel: corresponding problems for the profile LR (21) at $\kappa_{11}(\theta)$ and $(\kappa_{11}(\theta), \kappa_{00}(\theta))'$.

To verify Assumption 3.5, take n sufficiently large that $\gamma(\theta) \in \mathrm{int}(\Gamma)$ for all $\theta \in \Theta_{osn}$:

$$PL_n(\Delta(\theta)) = \min_{m \in \{\kappa_{11}(\theta), \kappa_{11}(\theta)+\kappa_{00}(\theta)\}} \sup_{\bar\theta \in \mu^{-1}(m)} \mathbb{P}_n \log p(\cdot, \bar\theta)\,. \tag{21}$$

By analogy with display (19), to calculate $PL_n(\Delta(\theta))$ we need to solve:

$$\sup_{\bar\theta \in \mu^{-1}(m)} \mathbb{P}_n \log(p(\cdot;\bar\theta)) = \sup_{\substack{0 \le k_{11} \le m \\ m \le k_{11}+k_{00} \le 1}} \mathbb{P}_n \Big(yd \log k_{11} + (d - yd)\log(1 - k_{11} - k_{00}) + (1-d)\log k_{00} \Big)$$

at $m = \kappa_{11}(\theta)$ and $m = \kappa_{11}(\theta) + \kappa_{00}(\theta)$.

This problem is geometrically the same as the problem for the profile QLR up to a translation of the local parameter space from $(\kappa_{11}, \kappa_{00})'$ to $(\kappa_{11}(\theta), \kappa_{00}(\theta))'$. The local parameter spaces are approximated by the translated cones $T_1(\theta) = T_1 + \sqrt{n}\gamma(\theta)$ and $T_2(\theta) = T_2 + \sqrt{n}\gamma(\theta)$. It follows that: uniformly in $\theta \in \Theta_{osn}$,

$$nPL_n(\Delta(\theta)) = n\mathbb{P}_n \log p_0 + \frac{1}{2}\|\mathbb{V}_n\|^2 - \frac{1}{2} f\big(\mathbb{V}_n - \sqrt{n}\gamma(\theta)\big) + o_{\mathbb{P}}(1)$$

where f is given in (20), and hence Assumption 3.5 holds. Therefore, Theorem 3.3 implies that our MC CS $\widehat{M}_\alpha$ for M_I will have asymptotically exact coverage.

Inference under identification: Now consider the case in which the model is identified (i.e. true $\kappa_{00} = 0$). In this case each $D_i = 1$ so the likelihood of the i-th observation $(D_i, Y_i D_i) = (1, y)$

is

$$p(1, y; \theta) = [\kappa_{11}(\theta)]^y [1 - \kappa_{11}(\theta) - \kappa_{00}(\theta)]^{1-y} = q(1, y; \tilde{\gamma}(\theta))$$

We again take Θ as in (12) and use a flat prior. Lemma F.6 in Appendix F shows that Π_n concentrates on the local neighborhood Θ_{osn} given by $\Theta_{osn} = \{\theta : |\kappa_{11}(\theta) - \kappa_{11}| \leq r_n/\sqrt{n}, \kappa_{00}(\theta) \leq r_n/n\}$ for any positive sequence $(r_n)_{n\in\mathbb{N}}$ with $r_n \to \infty$, $r_n/\sqrt{n} = o(1)$.

Here the reduced-form parameter $\tilde{\gamma}(\theta)$ is $\tilde{\gamma}(\theta) = \kappa_{11}(\theta) - \kappa_{11}$. Uniformly over Θ_{osn} we obtain:

$$nL_n(\theta) = n\mathbb{P}_n \log p_0 - \frac{1}{2}\frac{(\sqrt{n}\tilde{\gamma}(\theta))^2}{\kappa_{11}(1-\kappa_{11})} + (\sqrt{n}\tilde{\gamma}(\theta))\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{y_i - \kappa_{11}}{\kappa_{11}(1-\kappa_{11})}\right) - n\kappa_{00}(\theta) + o_{\mathbb{P}}(1)$$

which verifies Assumption 3.2'(i) with $\gamma(\theta) = (\kappa_{11}(1-\kappa_{11}))^{-1/2}\tilde{\gamma}(\theta)$, $T = \mathbb{R}$, $f_{n,\perp}(\gamma_\perp(\theta)) = n\gamma_\perp(\theta)$ where $\gamma_\perp(\theta) = \kappa_{00}(\theta) \geq 0$, and

$$\mathbb{V}_n = (\kappa_{11}(1-\kappa_{11}))^{-1/2}\mathbb{G}_n(y) \rightsquigarrow N(0,1)\ .$$

The remaining parts of Assumption 3.2' are easily shown to be satisfied. Therefore, Theorem 3.2 implies that our MC CS $\widehat{\Theta}_\alpha$ for Θ_I will be asymptotically valid but conservative.

For subvector inference on $M_I = \{\mu_0\}$, the profile LR statistic for $M_I = \{\mu_0\}$ is asymptotically χ^2_1, and equation (18) holds with $f : \mathbb{R} \to \mathbb{R}_+$ given by $f(v) = v^2$ and $T = \mathbb{R}$. To verify Assumption 3.5, for each $\theta \in \Theta_{osn}$ we need to solve

$$\sup_{\bar{\theta}\in\mu^{-1}(m)} \mathbb{P}_n \log(p(\cdot;\bar{\theta})) = \sup_{\substack{0\leq k_{11}\leq m \\ m\leq k_{11}+k_{00}\leq 1}} \mathbb{P}_n\Big(y\log k_{11} + (1-y)\log(1-k_{11}-k_{00})\Big)$$

at $m = \kappa_{11}(\theta)$ and $m = \kappa_{11}(\theta) + \kappa_{00}(\theta)$. The maximum is achieved when k_{00} is as small as possible, which occurs along the segment $k_{00} = m - k_{11}$. Substituting in and maximizing with respect to k_{11}:

$$\sup_{\bar{\theta}\in\mu^{-1}(m)} \mathbb{P}_n \log(p(\cdot;\bar{\theta})) = \mathbb{P}_n\big(y\log m + (1-y)\log(1-m)\big)\,.$$

Therefore, we obtain the following expansion uniformly for $\theta \in \Theta_{osn}$:

$$\begin{aligned} nPL_n(\Delta(\theta)) &= n\mathbb{P}_n \log p_0 + \frac{1}{2}(\mathbb{V}_n)^2 \\ &\quad - \frac{1}{2}\Big(\mathbb{V}_n - \sqrt{n}\gamma(\theta)\Big)^2 \vee \frac{1}{2}\Big(\mathbb{V}_n - \sqrt{n}(\gamma(\theta) + \kappa_{00}(\theta))\Big)^2 + o_{\mathbb{P}}(1) \\ &= n\mathbb{P}_n \log p_0 + \frac{1}{2}(\mathbb{V}_n)^2 - \frac{1}{2}\Big(\mathbb{V}_n - \sqrt{n}\gamma(\theta)\Big)^2 + o_{\mathbb{P}}(1) \end{aligned}$$

where the last equality holds because $\sup_{\theta\in\Theta_{osn}} \kappa_{00}(\theta) \le r_n/n = o(n^{-1/2})$. This verifies that Assumption 3.5 holds with $f(v) = v^2$. Thus Theorem 3.3 implies that our MC CS $\widehat{M}_\alpha$ for M_I will have asymptotically exact coverage, even though $\widehat{\Theta}_\alpha$ for Θ_I will be conservative in this case.

Example 2: complete information entry game

Consider the bivariate discrete game with payoffs described in Table 9. Let $D_{a_1a_2}$ denote a binary random variable taking the value 1 if and only if player 1 takes action a_1 and player 2 takes action a_2. We observe a random sample $\{(D_{00,i}, D_{01,i}, D_{10,i}, D_{11,i})\}_{i=1}^n$. The model is parameterized by $\theta = (\beta_1, \beta_2, \Delta_1, \Delta_2, \rho, s)'$, where ρ is the parameter associated with the joint probability distribution (Q_ρ) of (ϵ_1, ϵ_2), and $s \in [0, 1]$ is the selection probability of choosing the $(a_1, a_2) = (0, 1)$ equilibrium when there are multiple equilibria. The reduced-form probabilities of observing D_{00}, D_{01}, D_{11} and D_{10} are $\kappa_{00}(\theta)$, $\kappa_{01}(\theta)$, $\kappa_{11}(\theta)$, and $\kappa_{10}(\theta) = 1 - \kappa_{00}(\theta) - \kappa_{01}(\theta) - \kappa_{11}(\theta)$, given by:

$$
\begin{aligned}
\kappa_{00}(\theta) &= Q_\rho(\epsilon_{1i} \le -\beta_1, \epsilon_{2i} \le -\beta_2) \\
\kappa_{01}(\theta) &= Q_\rho(-\beta_1 \le \epsilon_{1i} \le -\beta_1 - \Delta_1, \epsilon_{2i} \le -\beta_2 - \Delta_2) + Q_\rho(\epsilon_{1i} \le -\beta_1, \epsilon_{2i} \ge -\beta_2) \\
&\quad + sQ_\rho(-\beta_1 \le \epsilon_{1i} \le -\beta_1 - \Delta_1, -\beta_2 \le \epsilon_{2i} \le -\beta_2 - \Delta_2) \\
\kappa_{11}(\theta) &= Q_\rho(\epsilon_{1i} \ge -\beta_1 - \Delta_1, \epsilon_{2i} \ge -\beta_2 - \Delta_2)\,.
\end{aligned}
$$

Let κ_{00}, κ_{01}, and κ_{11} denote the true values of the reduced-form choice probabilities. This model falls into the class of models dealt with in Proposition 4.1 with $\tilde{\gamma}(\theta) = \kappa(\theta) - \kappa_0$ where $\kappa(\theta) = (\kappa_{00}(\theta), \kappa_{01}(\theta), \kappa_{11}(\theta))'$ and $\kappa_0 = (\kappa_{00}, \kappa_{01}, \kappa_{11})'$. The likelihood at the i-th observation is:

$$
\begin{aligned}
p(d_{00}, d_{01}, d_{11}; \theta) &= [\kappa_{00}(\theta)]^{d_{00}}[\kappa_{01}(\theta)]^{d_{01}}[\kappa_{11}(\theta)]^{d_{11}}(1 - \kappa_{00}(\theta) - \kappa_{01}(\theta) - \kappa_{11}(\theta))^{1-d_{00}-d_{01}-d_{11}} \\
&= q(d_{00}, d_{01}, d_{11}; \tilde{\gamma}(\theta))\,.
\end{aligned}
$$

Conditions (a)-(b) and (d) of Proposition 4.1 hold with $\widetilde{\Gamma} = \{\tilde{\gamma}(\theta) : \theta \in \Theta\}$ under very mild conditions on the parameterization $\theta \mapsto \tilde{\gamma}(\theta)$. Hence Assumption 3.2 is satisfied with $\gamma(\theta) = \mathbb{I}_0^{1/2}\tilde{\gamma}(\theta)$ where:

$$
\mathbb{I}_0 = \begin{bmatrix} \frac{1}{\kappa_{11}} & 0 & 0 \\ 0 & \frac{1}{\kappa_{01}} & 0 \\ 0 & 0 & \frac{1}{\kappa_{11}} \end{bmatrix} + \frac{1}{1 - \kappa_{00} - \kappa_{01} - \kappa_{11}} \mathbf{1}_{3\times 3}
$$

where $\mathbf{1}_{3\times 3}$ denotes a 3×3 matrix of ones,

$$\mathbb{V}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbb{I}_0^{-1/2}\begin{pmatrix} \frac{d_{00,i}}{\kappa_{00}} - \frac{1-d_{00,i}-d_{01,i}-d_{11,i}}{1-\kappa_{00}-\kappa_{01}-\kappa_{11}} \\ \frac{d_{01,i}}{\kappa_{01}} - \frac{1-d_{00,i}-d_{01,i}-d_{11,i}}{1-\kappa_{00}-\kappa_{01}-\kappa_{11}} \\ \frac{d_{11,i}}{\kappa_{11}} - \frac{1-d_{00,i}-d_{01,i}-d_{11,i}}{1-\kappa_{00}-\kappa_{01}-\kappa_{11}} \end{pmatrix} \rightsquigarrow N(0, I_3)$$

and $T=\mathbb{R}^3$. Condition (c) of Proposition 4.1 and Assumption 3.3 can be verified under mild conditions on the map $\theta \mapsto \kappa(\theta)$ and the prior Π. For instance, consider the parameterization $\theta = (\beta_1, \beta_2, \Delta_1, \Delta_2, \rho, s)$ where the joint distribution of (ϵ_1, ϵ_2) is a bivariate Normal with means zero, standard deviations one and positive correlation $\rho \in [0,1]$. The parameter space is

$$\Theta = \{(\beta_1, \beta_2, \Delta_1, \Delta_2, \rho, s) \in \mathbb{R}^6 : \underline{\beta} \leq \beta_1, \beta_2 \leq \overline{\beta}, \underline{\Delta} \leq \Delta_1, \Delta_2 \leq \overline{\Delta}, 0 \leq \rho, s \leq 1\}.$$

where $-\infty < \underline{\beta} < \overline{\beta} < \infty$ and $-\infty < \underline{\Delta} < \overline{\Delta} < 0$. The image measure Π_Γ of a flat prior on Θ is positive and continuous on a neighborhood of the origin, which verifies Condition (c) of Proposition 4.1 and Assumption 3.3. Therefore, Theorem 3.1 implies that our MC CSs for Θ_I will have asymptotically exact coverage.

4.1.2 General non-identifiable likelihood models

It is possible to define a local reduced-form reparameterization for non-identifiable likelihood models, even when $\mathcal{P} = \{p(\cdot;\theta) : \theta \in \Theta\}$ does not admit an explicit (global) reduced-form reparameterization. Let $\mathcal{D} \subset L^2(P_0)$ denote the set of all limit points of:

$$\mathcal{D}_\epsilon := \left\{ \frac{\sqrt{p/p_0} - 1}{h(p, p_0)} : p \in \mathcal{P}, 0 < h(p, p_0) \leq \epsilon \right\}$$

as $\epsilon \to 0$. The set $\mathcal{D}$ is the set of generalized Hellinger scores,[22] which consists of functions of X_i with mean zero and unit variance. The cone $\Lambda = \{\tau d : \tau \geq 0, d \in \mathcal{D}\}$ is the tangent cone of the model $\mathcal{P}$ at p_0. We say that $\mathcal{P}$ is differentiable in quadratic mean (DQM) if each $p \in \mathcal{P}$ is absolutely continuous with respect to p_0 and for each $p \in \mathcal{P}$ there are elements $g(p) \in \Lambda$ and remainders $R(p) \in L^2(\lambda)$ such that:

$$\sqrt{p} - \sqrt{p_0} = g(p)\sqrt{p_0} + h(p, p_0)R(p)$$

with $\sup\{\|R(p)\|_{L^2(\lambda)} : h(p, p_0) \leq \varepsilon\} \to 0$ as $\varepsilon \to 0$. If the linear hull $\mathrm{Span}(\Lambda)$ of Λ has finite dimension $d^* \geq 1$, then we can write each $g \in \Lambda$ as $g = c(g)'\psi$ where $c(g) \in \mathbb{R}^{d^*}$ and the

[22] It is possible to define sets of generalized scores via other measures of distance between densities. See Liu and Shao (2003) and Azaïs, Gassiat, and Mercadier (2009). Our results can easily be adapted to these other cases.

elements of $\psi = (\psi_1, \ldots, \psi_{d^*})$ form an orthonormal basis for $\mathrm{Span}(\Lambda)$ in $L^2(P_0)$. Let $\mathbf{\Lambda}$ denote the orthogonal projection onto Λ and let $\gamma(\theta)$ be given by $\mathbf{\Lambda}(2(\sqrt{p(\cdot;\theta)/p_0(\cdot)} - 1)) = \gamma(\theta)'\psi$.[23] Finally, let $\overline{\mathcal{D}}_\varepsilon = \mathcal{D}_\epsilon \cup \mathcal{D}$.

Proposition 4.2. *Suppose that $\mathcal{P}$ satisfies the following regularity conditions:*
(a) $\{\log p : p \in \mathcal{P}\}$ is P_0-Glivenko Cantelli;
(b) $\mathcal{P}$ is DQM, and Λ is convex and $\mathrm{Span}(\Lambda)$ has finite dimension $d^* \geq 1$.
(c) there exists $\varepsilon > 0$ such that $\overline{\mathcal{D}}_\varepsilon$ is Donsker and has envelope $D \in L^2(P_0)$.
Then: there exists a sequence $(r_n)_{n\in\mathbb{N}}$ with $r_n \to \infty$ and $r_n = O(\log n)$ as $n \to \infty$, such that Assumption 3.2(i) holds for the average log-likelihood (1) over $\Theta_{osn} := \{\theta : h(P_\theta, P_0) \leq r_n/\sqrt{n}\}$ with $\mathbb{V}_n = \mathbb{G}_n(\psi)$ and $\gamma(\theta)$ defined by $\mathbf{\Lambda}(2(\sqrt{p(\cdot;\theta)/p_0(\cdot)} - 1)) = \gamma(\theta)'\psi$.

Proposition 4.2 is a set of sufficient conditions in the i.i.d. setting. See Lemma F.7 in Appendix F for a more general result.

4.2 GMM models

Consider the GMM model $\{\rho(X_i, \theta) : \theta \in \Theta\}$ with $\rho : \mathscr{X} \times \Theta \to \mathbb{R}^{\dim(\rho)}$. Let $g(\theta) = E[\rho(X_i, \theta)]$ and the identified set be $\Theta_I = \{\theta \in \Theta : g(\theta) = 0\}$. In models with a moderate or large number of moment conditions, the set $\{g(\theta) : \theta \in \Theta\}$ may not contain a neighborhood of the origin. However, the map $\theta \mapsto g(\theta)$ is typically smooth, in which case $\{g(\theta) : \theta \in \Theta\}$ can be locally approximated at the origin by a closed convex cone $\Lambda \subset \mathbb{R}^{\dim(g)}$ at the origin. For instance, if $\{g(\theta) : \theta \in \Theta\}$ is a differentiable manifold this is trivially true with Λ a linear subspace of $\mathbb{R}^{\dim(g)}$.

Let $\mathbf{\Lambda} : \mathbb{R}^{\dim(g)} \to \Lambda$ denote the orthogonal projection onto Λ. Let $U \in \mathbb{R}^{\dim(g)\times\dim(g)}$ be a unitary matrix (i.e. $U' = U^{-1}$) such that for each $v \in \mathbb{R}^{\dim(g)}$ the first $\dim(\Lambda) = d^*$ (say) elements of Uv are in the linear hull $\mathrm{Span}(\Lambda)$ and the remaining $\dim(g) - d^*$ are orthogonal to $\mathrm{Span}(\Lambda)$. Let $[(U\Omega U')^{-1}]_{11}$ be the $d^* \times d^*$ upper left block of $(U\Omega U')^{-1}$, $[U\mathbf{\Lambda}g(\theta)]_1$ be the first d^* elements of $U\mathbf{\Lambda}g(\theta)$, and $[U\Omega^{-1}\mathbb{G}_n(\rho(\cdot,\theta))]_1$ be the upper d^* subvector of $U\Omega^{-1}\mathbb{G}_n(\rho(\cdot;\theta))$. If $\{g(\theta) : \theta \in \Theta\}$ contains a neighborhood of the origin then we just take $\Lambda = \mathbb{R}^{\dim(g)}$ with $d^* = \dim(g)$, $U = I_{\dim(g)}$, and $\mathbf{\Lambda}g(\theta) = g(\theta)$.

In the following let $\Theta_I^\varepsilon = \{\theta \in \Theta : \|g(\theta)\| \leq \varepsilon\}$ and $\mathcal{R}_\varepsilon = \{\rho(\cdot, \theta) : \theta \in \Theta, \|g(\theta)\| \leq \varepsilon\}$.

Proposition 4.3. *Suppose that data $\{X_i\}_{i=1}^n$ is i.i.d. and the identified set $\Theta_I = \{\theta \in \Theta : E[\rho(X_i, \theta)] = 0\}$ is not empty. Let the following hold:*

[23] If $\Lambda \subseteq L^2(P_0)$ is a closed convex cone, the projection $\mathbf{\Lambda}f$ of any $f \in L^2(P_0)$ is defined as the unique element of Λ such that $\|f - \mathbf{\Lambda}f\|_{L^2(P_0)} = \inf_{t\in\Lambda} \|f - t\|_{L^2(P_0)}$ (see Hiriart-Urruty and Lemaréchal (2001)).

(a) $\sup_{\theta \in \Theta_I^\varepsilon} \|g(\theta) - \mathbf{\Lambda} g(\theta)\| = o(\varepsilon)$ *as* $\varepsilon \to 0$*;*
(b) $E[\rho(X_i,\theta)\rho(X_i,\theta)'] = \Omega$ *for each* $\theta \in \Theta_I$ *and* Ω *is positive definite;*
(c) there exists $\varepsilon_0 > 0$ *such that* $\mathcal{R}_{\varepsilon_0}$ *is Donsker;*
(d) $\sup_{(\theta,\bar\theta):\in \Theta_I^\varepsilon \times \Theta_I} E[\|\rho(X_i,\theta) - \rho(X_i;\bar\theta)\|^2] = o(1)$ *as* $\varepsilon \to 0$*;*
(e) $\sup_{\theta \in \Theta_I^\varepsilon} \|E[(\rho(X_i,\theta)\rho(X_i,\theta)')] - \Omega\| = o(1)$ *as* $\varepsilon \to 0$.
Then: there exists a sequence $(r_n)_{n\in\mathbb{N}}$ *with* $r_n \to \infty$ *and* $r_n = o(n^{1/4})$ *as* $n \to \infty$ *such that Assumption 3.2(i) holds for the continuously-updated GMM criterion (2) over* $\Theta_{osn} = \{\theta \in \Theta : \|g(\theta)\| \leq r_n/\sqrt{n}\}$*, where* $\gamma(\theta) = [(U\Omega U')^{-1}]_{11}^{1/2}[U\mathbf{\Lambda} g(\theta)]_1$*,* $\mathbb{V}_n = -[(U\Omega U')^{-1}]_{11}^{-1/2}[U\Omega^{-1}\mathbb{G}_n(\rho(\cdot,\theta))]_1$ *for any fixed* $\theta \in \Theta_I$*, and* T *equals to the image of* Λ *under the map* $v \mapsto [(U\Omega U')^{-1}]_{11}[Uv]_1$*.*
If $\{g(\theta) : \theta \in \Theta\}$ *contains a neighborhood of the origin then* $\gamma(\theta) = \Omega^{-1/2}g(\theta)$*,* $\mathbb{V}_n = -\Omega^{-1/2}\mathbb{G}_n(\rho(\cdot,\theta))$ *for any fixed* $\theta \in \Theta_I$*, and* $T = \mathbb{R}^{\dim(g)}$*.*

Proposition 4.4. *Let all the conditions of Proposition 4.3 hold, except that its condition (e) is replaced by: (e)* $\|\widehat{W} - \Omega^{-1}\| = o_{\mathbb{P}}(1)$*.*
Then: the conclusions of Proposition 4.3 hold for the optimally-weighted GMM criterion (3).

Andrews and Mikusheva (2016) consider weak identification-robust inference when the null hypothesis is described by a regular C^2 manifold in the parameter space. Let $\{g(\theta) : \theta \in \Theta\}$ be a C^2 manifold in $\mathbb{R}^{\dim(g)}$ that is regular at the origin.[24] Then Condition (a) of Propositions 4.3 and 4.4 hold with Λ equal to the tangent space of $\{g(\theta) : \theta \in \Theta\}$ at the origin, which is a linear subspace of $\mathbb{R}^{\dim(g)}$ (Federer, 1996, p. 234). It is straightforward to verify that K_{osn} is convex and contains a ball B_{k_n} where we may choose $k_n \to \infty$ as $n \to \infty$, hence Assumption 3.2(ii) also hold with $T = \mathbb{R}^{\dim(\Lambda)}$.

4.2.1 Moment inequalities

Consider the moment inequality model $\{\tilde\rho(X_i,\beta) : \beta \in B\}$ with $\tilde\rho : \mathscr{X} \times B \to \mathbb{R}^{\dim(\rho)}$ where the parameter space is $B \subseteq \mathbb{R}^{\dim(\beta)}$. The identified set is $B_I = \{\beta \in B : E[\tilde\rho(X_i,\beta)] \leq 0\}$ (the inequality is understood to hold element-wise). We may reformulate the moment inequality model as a GMM-type moment equality model by augmenting the parameter vector with a vector of slackness parameters $\lambda \in \Lambda \subseteq \mathbb{R}_+^{\dim(\rho)}$. Thus we re-parameterize the model by $\theta = (\beta,\lambda) \in \Theta = B \times \Lambda$ and write the inequality model as a GMM equality model

$$E[\rho(X_i,\theta)] = 0 \; for \; \theta \in \Theta_I, \quad \rho(X_i,\theta) = \tilde\rho(X_i,\beta) + \lambda \,, \tag{22}$$

[24]That is, there exists a neighborhood N of the origin in $\mathbb{R}^{\dim(g)}$, a C^2 homeomorphism $\varphi : N \to \mathbb{R}^{\dim(g)}$, and a linear subspace Φ of $\mathbb{R}^{\dim(g)}$ of dimension $\dim(\Phi)$ such that $\varphi(N \cap \{g(\theta) : \theta \in \Theta\}) = \Phi \cap \mathrm{im}(\varphi)$ where $\mathrm{im}(\varphi)$ is the image of φ. Such manifolds are also called $\dim(\Phi)$-dimensional submanifolds of class 2 of $\mathbb{R}^{\dim(g)}$; see Federer (1996), Chapers 3.1.19-20.

where the identified set for θ is $\Theta_I = \{\theta \in \Theta : E[\rho(X_i, \theta)] = 0\}$ and B_I is the projection of Θ_I onto B. We may then apply Propositions 4.3 or 4.4 to the reparameterized GMM model (22).

Example 3. As a simple illustration, consider the model in which $X_1, \ldots, X_n$ are i.i.d. with unknown mean $\mu \in [0,1] = B$ and unit variance. Suppose that $\beta \in B$ is identified by the moment inequality $\mathbb{E}[\beta - X_i] \leq 0$. The identified set for β is $B_I = [0, \mu]$, which is the argmax of the population criterion function

$$L(\beta) = -\frac{1}{2}((\beta - \mu) \vee 0)^2$$

(see Figure 5(a)). The sample analogue criterion $-\frac{1}{2}((\beta - \bar{X}_n) \vee 0)^2$ is typically used in the moment inequality literature, but does not satisfy our Assumption 3.2. However, we can rewrite the inequality model in terms of the moment equality model: $\mathbb{E}[\beta + \lambda - X_i] = 0$ where $\lambda \in [0, 1-\beta]$ is a slackness parameter. The parameter space for $\theta = (\beta, \lambda)$ is $\Theta = \{(\beta, \lambda) \in B^2 : \beta + \lambda \leq 1\}$. The identified set for θ is $\Theta_I = \{(\beta, \lambda) \in \Theta : \beta + \lambda = \mu\}$ and the identified set for the subvector β is B_I (see Figure 5(b)). The GMM objective function for $\mathbb{E}[\beta + \lambda - X_i] = 0$ is:

$$L_n(\beta, \lambda) = -\frac{1}{2}(\beta + \lambda - \bar{X}_n)^2 .$$

Suppose that $\mu \in (0,1)$. Then wpa1 we can choose $(\hat{\beta}, \hat{\lambda}) \in \Theta$ such that $nL_n(\hat{\beta}, \hat{\lambda}) = 0 + o_{\mathbb{P}}(1)$. Then:

$$\sup_{(\beta,\lambda)\in\Theta} |Q_n(\beta, \lambda) - (\mathbb{V}_n - \sqrt{n}(\beta + \lambda - \mu))^2| = o_{\mathbb{P}}(1)$$

where $\mathbb{V}_n = \sqrt{n}(\bar{X}_n - \mu) \rightsquigarrow N(0,1)$. The profile QLR for B_I is $\sup_{\beta \in B_I} \inf_{\lambda \in B} Q_n(\beta, \lambda)$ where:

$$\inf_{\lambda \in B} Q_n(\beta, \lambda) = \left[\begin{array}{ll} (\mathbb{V}_n - \sqrt{n}(\beta - \mu))^2 & \text{if } \ \mathbb{V}_n/\sqrt{n} - (\beta - \mu) < 0 \\ 0 & \text{if } \ 0 \leq \mathbb{V}_n/\sqrt{n} - (\beta - \mu) \leq 1 \\ (\mathbb{V}_n - \sqrt{n}(\beta + 1 - \mu))^2 & \text{if } \ \mathbb{V}_n/\sqrt{n} - (\beta - \mu) > 1 \, . \end{array} \right.$$

The maximum over B_I is attained at $\beta = \mu$, hence $PQ_n(\Delta(\theta)) = f(\mathbb{V}_n) + o_{\mathbb{P}}(1)$ for all $\theta \in \Theta_I$ where $f(v) = v^2 1\!\!1\{v < 0\}$. Therefore, the profile QLR for B_I is asymptotically a mixture between point mass at zero and a χ^2_1 random variable.

For the posterior distribution of the profile QLR, first observe that this maps into our framework with the local reduced-form parameter $\gamma(\theta) = ((\beta + \lambda) - \mu)$. A flat prior on Θ induces a prior Π_Γ whose density $\pi_\Gamma(\gamma) = 2(\gamma + \mu)$ is positive and continuous at the origin (see Figure 5(c)). The set $\Gamma = \{\gamma(\theta) : \theta \in \Theta\}$ contains a ball of positive radius at the origin when $\mu \in (0,1)$ hence

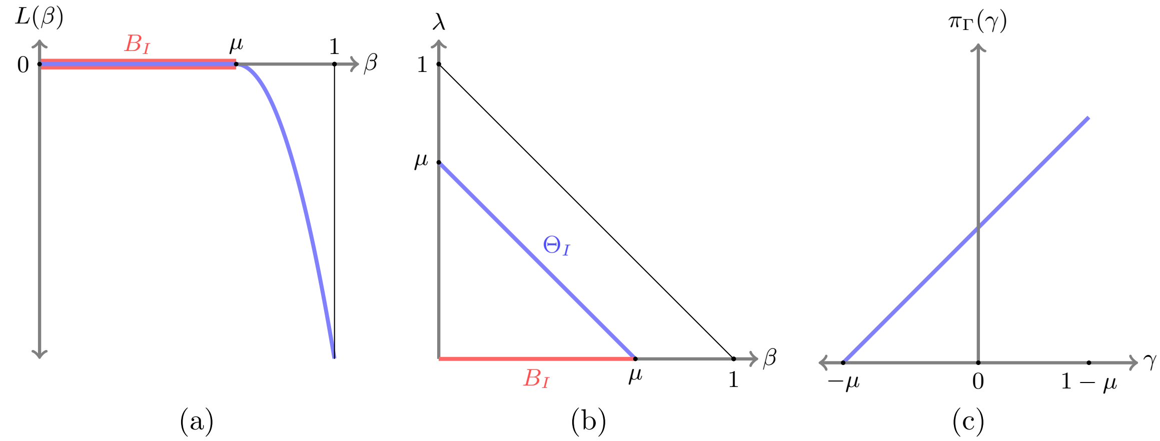

Figure 5: Panel (a): identified set B_I for β with population (moment inequality) criterion $L(\beta) = -\frac{1}{2}((\beta-\mu)\vee 0)^2$. Panel (b): identified set Θ_I for $\theta = (\beta, \lambda)$ with moment equality model $\mathbb{E}[\beta + \lambda - X] = 0$, and identified set B_I for β. Panel (c): induced prior π_Γ for $\gamma(\theta) = (\beta + \lambda - \mu)$ from a flat prior on Θ.

$T = \mathbb{R}$ (otherwise $T = \mathbb{R}_+$ or $\mathbb{R}_-$ when μ is at the boundary of B). Moreover:

$$\Delta(\theta^b) = \{(\beta, \lambda) \in \Theta : \beta + \lambda = \beta^b + \lambda^b\}$$

and so $\mu(\Delta(\theta^b)) = [0, \beta^b + \lambda^b]$. Similar arguments then yield:

$$PQ_n(\Delta(\theta)) = f(\mathbb{V}_n - \sqrt{n}\gamma(\theta)) + o_{\mathbb{P}}(1) \quad \text{uniformly in} \quad \theta \in \Theta_I$$

with $f(v) = v^2 1\!\!1\{v < 0\}$. So all the regularity conditions of Theorem 3.3 hold, and hence our MC CS $\widehat{M}_\alpha$ has asymptotically exact coverage for B_I.

In Appendix C we show that, under a drifting sequence of DGPs towards the boundary $B_I = \{0\}$, our MC CS $\widehat{M}_\alpha$ has asymptotically correct but possibly conservative coverage for B_I while the nonparametric bootstrap based CS for B_I undercovers. This illustrates that our MC CSs are not equivalent to bootstrap CSs.

5 Applications

This section implements our procedures in two empirical illustrations. The first estimates a model of trade flows initially examined in Helpman, Melitz, and Rubinstein (2008) (HMR henceforth). This application uses a version of the empirical model in HMR with more than 40 parameters to be estimated. The second empirical example estimates a simple stylized version of a bivariate binary entry game with data from the US airline industry with 17 parameters to be estimated via a likelihood. Both of these applications illustrate our robust approach to inference: the model is nonlinear and it might be hard to determine whether it point identifies the parameters; and more importantly, examining the robustness of the estimates to various adhoc modelling assumptions can be done in a theoretically valid and computationally feasible way.

5.1 An Empirical Model of Trade Flows

In an influential paper, Helpman et al. (2008) examine the extensive margin of trade using a structural model estimated with current trade data. The following is a brief description of their empirical framework. Let M_{ij} denote the value of country i's imports from country j. This is only observed if country j exports to country i. If a random draw for productivity from country j to i is sufficiently high then j will export to i. To model this, Helpman et al. (2008) introduce a latent variable z^*_{ij} which measures trade volume between i and j. Here z^*_{ij} takes the value zero if j does not export to i and strictly positive otherwise. We adapt slightly their empirical model to obtain a selection model of the form:

$$\log M_{ij} = \begin{cases} \beta_0 + \lambda_j + \chi_i - \nu' d_{ij} + \delta z^*_{ij} + u_{ij} & \text{if } z_{ij^*} > 0 \\ \text{not observed} & \text{if } z_{ij^*} \leq 0 \end{cases}$$

$$z^*_{ij} = \beta^*_0 + \lambda^*_j + \chi^*_i - \nu^{*\prime} d_{ij} + \eta^*_{ij}$$

in which λ_j, χ_i, λ^*_j and χ^*_i are exporting and importing continent fixed effects, d_{ij} is a vector of observable trade frictions between i and j, and u_{ij} and η^*_{ij} are error terms described below. Notice that the model is different from the usual Heckman selection model due to the presence of z^*_{ij} in the outcome equation. Exclusion restrictions can be imposed by setting one or several of the elements of ν equal to zero.

There are three differences between our empirical model and that of Helpman et al. (2008). First, we let z^*_{ij} enter the outcome equation linearly instead of nonlinearly[25]. Second, we use continent fixed effect instead of country fixed effects. This reduces the number of parameters from over 400 to around 40. Third, we allow for heteroskedasticity in the selection equation, which is known to be a problem in trade data. Also, this is one way to illustrate the robustness approach we advocate which relaxes parametric assumptions on part of the model that is suspect (homoskedasticity) without worrying about loss of point identification.

To allow for heteroskedasticity, we suppose that the distribution of (u_{ij}, η^*_{ij}) conditional on

[25] Their nonlinear specification is known to be problematic (see, e.g., Santos Silva and Tenreyro (2015)).

observables is Normal with mean zero and covariance:

$$\Sigma(X_{ij}) = \begin{pmatrix} \sigma_m^2 & \rho\sigma_m\sigma_z(X_{ij}) \\ \rho\sigma_m\sigma_z(X_{ij}) & \sigma_z^2(X_{ij}) \end{pmatrix}$$

where:

$$\sigma_z(X_{ij}) = \exp(\log(\text{distance}_{ij}) + \varpi_1 \log(\text{distance}_{ij})^2)\,.$$

For estimation we estimate the model from data on 24,649 country pairs in the selection equation and 11,156 country pairs in the outcome equation using the same data from 1986 as in Helpman et al. (2008). We also impose the exclusion restriction that the coefficient in ν corresponding to religion is equal to zero, else there is an exact linear relationship between the coefficients in the outcome and selection equation. This leaves a total of 43 parameters to be estimated. We only report estimates for the trade friction coefficients ν in the outcome equation as these are the most important. We estimate the model first by maximum likelihood under homoskedasticity and report conventional ML estimates for ν together with 95% confidence sets based on inverting t-statistics. We then re-estimate the model under heteroskedasticity and report conventional ML estimates together with confidence sets based on inverting t-statistics, the Chernozhukov and Hong (2003) procedure, and our procedures 2 and 3. We use a random walk Metropolis Hastings sampler with chain length of 10000, burnin of 10000 and acceptance rate tuned to be one third.

The results are presented in Table 11. Overall, the results for the heteroskedastic specification show that the confidence sets seem reasonably insensitive to the type of procedure used, which suggests that partial identification may not be an issue even allowing for heteroskedasticity. We also notice some difference in results relative to Helpman et al. (2008). For instance, we find that sharing the same legal system does not significantly impact trade flows whereas they document a strong positive effect. On the other hand, we find that sharing a common language and not being an island has a positive effect on trade flows whereas they document no such effects. Under heteroskedasticity, the magnitudes of coefficients of the trade friction variables are generally smaller than under homoskedasticity but of the same sign. The exception is the legal variable, whose coefficient is negative under homoskedasticity but positive under heteroskedasticity. However, this variable is insignificant for both specifications. A question that one can shed light on is whether the estimates are also sensitive to the normality assumption on the errors. This question can be examined within the context of our results by for example using a flexible form for the joint distribution of the errors.

5.2 Bivariate Entry Game with US Airline Data

This section estimates a version of the entry game that we study in Subsection 2.4.2 above. We use data from the second quarter of 2010's Airline Origin and Destination Survey (DB1B) to estimate a binary game where the payoff for firm i from entering market m is

$$\beta_i + \beta_i^x x_{im} + \Delta_i y_{3-i} + \epsilon_{im} \quad i = 1, 2$$

where the Δ_i's are assumed to be negative (as usually the case in entry models). The data contain 7882 markets which are formally defined as trips between two airports irrespective of stopping and we examine the entry behavior of two kinds of firms: *LC* (low cost) firms,[26] and *OA* (other airlines) which includes all the other firms. The unconditional choice probabilities are $(.16, .61, .07, .15)$ which are respectively the probabilities that OA and LC serve a market, that OA and not LC serve a market, that LC and not OA serve a market, and finally whether no airline serve the market. The regressors we have are *market presence* and *market size*. Market presence is a market and airline specific variable and is defined as follows. From a given airport, we compute the ratio of markets a given carrier (we take the maximum within the category OA or the category LC) serves divided by the total number of markets served from that given airport. The market presence variable (or *MP*) is the average of the ratios from the two endpoints and it provides some proxy for an airline's presence in a given airport (See Berry (1992) for more on this variable). For our purposes here, this variable is important since it acts as an *excluded* regressor: the market presence for OA only enters OA's payoffs - so *MP* is both market and airline specific. The second regressor we use is *market size* (or *MS*) which is defined as the population at the endpoints - so this variable is market specific. We discretize both market size and market presence into binary variables that take the value of one if the variable is higher than its median (in the data) value and zero otherwise. So, the reduced form parameters (or the $\kappa(.)$'s in Subsection 2.4.2) here are conditional on a three dimensional vector. That is, the choice probabilities are $P(y_{OA}, y_{LC}|MS,\ MP_{OA},\ MP_{LC})$ which gives us a set of 4 choice probabilities for every value of the conditioning variables (and there are 8 values for these[27]). To use notation similar to that in Subsection 2.4.2, we call firm *OA* as player 1 and firm *LC* as player 2. Denote $\beta_1(x_{mOA}) := \beta^0_{OA} + \beta'_{OA} x_{mOA}$ and $\beta_2(x_{mLC}) := \beta^0_{LC} + \beta'_{LC} x_{mLC}$. Then the likelihood of market m observation depends on the following choice probabilities:

[26] The low cost carriers are: JetBLue, Frontier, Air Tran, Allegiant Air, Spirit, Sun Country, USA3000, Virgin America, Midwest Air, and Southwest.

[27] With binary values, the conditioning set takes the following eight values: (1,1,1), (1,1,0), (1,0,1), (1,0,0), (0,1,1), (0,1,0), (0,0,1), (0,0,0).

	Homoskedastic		Heteroskedastic				
Variable	MLE	Asy	MLE	Asy	CH	Proc 2	χ^2
Distance	-1.849	(-2.876,-0.823)	-0.694	(-0.838,-0.549)	(-0.819,-0.539)	(-0.844,-0.551)	(-0.839,-0.557)
Border	4.551	(2.536, 6.565)	2.264	(1.786, 2.743)	(1.789, 2.719)	(1.797, 2.730)	(1.797, 2.730)
Island	1.184	(0.278, 2.090)	0.853	(0.504, 1.202)	(0.505, 1.167)	(0.524, 1.186)	(0.513, 1.197)
Landlock	5.794	(2.708, 8.880)	2.366	(1.744, 2.989)	(1.741, 3.104)	(1.702, 3.068)	(1.770, 2.991)
Legal	-0.220	(-0.654, 0.214)	0.058	(-0.096, 0.211)	(-0.086, 0.210)	(-0.092, 0.205)	(-0.092, 0.206)
Language	3.111	(1.140, 5.082)	0.696	(0.470, 0.921)	(0.420, 0.885)	(0.450, 0.954)	(0.480, 0.918)
Colonial ties	13.852	(6.033,21.671)	5.526	(4.389, 6.662)	(4.411, 6.552)	(4.531, 6.737)	(4.515, 6.756)
Currency union	15.066	(6.097,24.035)	5.119	(3.779, 6.459)	(3.767, 6.491)	(3.859, 6.672)	(3.933, 6.565)

Table 11: Estimated coefficients ν of the trade friction variables in the outcome equation together with 95% confidence sets. "Asy" are conventional 95% asymptotic CSs based on inverting t-statistics, CH are CSs based on the upper and lower 2.5% percentiles of the MCMC scalar parameter chain. Proc 2 and χ^2 are 95% CSs calculated using our Procedures 2 and 3, respectively.

$$
\begin{aligned}
\kappa_{11}(\theta; x_m) :=& P(\epsilon_{1m} \geq -\beta_1(x_{mOA}) - \Delta_{OA};\, \epsilon_{2m} \geq -\beta_2(x_{mLC}) - \Delta_{LC}) \\
\kappa_{00}(\theta; x_m) :=& P(\epsilon_{1m} \leq -\beta_1(x_{mOA}); \epsilon_{2m} \leq -\beta_2(x_{mLC})) \\
\kappa_{10}(\theta; x_m) :=& s(x_m) \times P(-\beta_1(x_{mOA}) \leq \epsilon_{1m} \leq -\beta_1(x_{mOA}) - \Delta_{OA}; -\beta_2(x_{mLC}) \leq \epsilon_{2m} \leq -\beta_2(x_{mLC}) - \Delta_{LC}) \\
& + P(\epsilon_{1m} \geq -\beta_1(x_{mOA}); \epsilon_{2m} \leq -\beta_2(x_{mLC})) \\
& + P(\epsilon_{1m} \geq -\beta_1(x_{mOA}) - \Delta_{OA}; -\beta_2(x_{mLC}) \leq \epsilon_{2m} \leq -\beta_2(x_{mLC}) - \Delta_{LC})\,.
\end{aligned}
$$

Here, $x_m = (MS_m, MP_{mOA}, MP_{mLC})'$ and $s(x_m)$ is a nuisance parameter which corresponds to the various *aggregate* equilibrium selection probabilities. This function $s(.)$ is defined on the support of x_m and so in the model this function takes $2^3 = 8$ values each belonging to $[0, 1]$. These selection probabilities are usually considered nuisance parameters. We call this the *full model* where no assumptions are made on equilibrium selection and use the likelihood function to build the confidence regions through the LR statistic as described above. So, the *full model* contains 4 parameters per profit function, the correlation across the ϵ's and the 8 parameters in the aggregate equilibrium choice probabilities (the s's) for a total of 17 parameters. We also estimate another version of the model called the *fixed s*, where we restrict the aggregate selection probabilities to be the same across markets. Note that the above is one version of the econometric model for a game and a more parsimonious version would allow for example the parameters to change with regressor values, or allow for the regressors' support to be richer (and not just binary). Here, we analyze this case precisely to highlight the fact that our CSs provide coverage guarantees regardless of whether the parameter vector is point identified. The empirical findings are presented in the Table 12 below.

The columns labeled *Proc 1* contain *projections* of the identified sets at the prespecified confidence level. In this model with 17 parameters, we expect these projections to be especially conservative. On the other hand, in the columns labeled χ^2, we provide one-dimensional confidence *intervals* for single dimensional (subvector) identified sets that are shown to be slightly conservative. The construction of these intervals follows Procedure 3 above where we profile out the corresponding nuisance parameters for every case and compute the likelihood on a one-dimensional grid. Generally, the χ^2 intervals should be tighter than the projection intervals and that is evident in Table 12.

Starting with the *full model* results, and considering first on the 95% χ^2 results, we see that the estimates are meaningful economically and are inline with recent estimates obtained in the literature. For example, fixed costs (the intercepts) are positive and significant for the large airlines (or *OA*) but are negative for the *LC* carriers. Typically the presence of higher fixed costs can signal various barriers to entry that are usually there to prevent *LC*s from entering. So, the higher these fixed costs the less likely it is for *LC*s to enter. On the other hand, higher fixed costs of large airlines are associated with a bigger presence (such as a hub) and so more likely to enter. As expected, both market presence and market size are associated with a positive probability of entry for both OA and LC regardless of market structure. Note also the very high correlation in the errors obtained here which could indicate missing profitability variables whereby firms enter a market regardless of competition in those markets that are particularly profitable. One interesting observation is the estimates for s_{001} and s_{101}. These are the aggregate selection probabilities and according to the results, they are not identified. This is likely to be due to the rather small number of markets with small size, large presence for OA but small presence for LC (in the case of s_{001}) and also small number of markets with large market size but small presence for LCs but large presence for OAs. The strength of our approach is its *adaptivity* to lack of identification in a particular data set: for example, the identified set for s_{001} is contained in $[0,1]$ with at least 95% probability which indicates that the model (and data) has no information about this parameter while the identified set for s_{111} is contained in $[.97,1]$ with at least 95% probability! Also, in the *fixed* s model, the results for both the Proc 1 and χ^2 procedures are in agreement with the corresponding ones for the full model χ^2 and the results across both Proc 1 (90% and 95%) and χ^2 (or Proc 3) for both full and fixed s models are remarkably similar and tell a consistent story.

6 Conclusion

We propose new methods for constructing frequentist CSs for IdSs in possibly partially-identified econometric models. Our CSs are simple to compute and have asymptotically correct frequentist coverage uniformly over a class of DGPs, including partially- and point- identified parametric likelihood and moment based models. We show that under a set of sufficient conditions, and in some broad classes of models, our set coverage is asymptotically exact. We also show that in models with singularities (such as the missing data example), our MCMC CSs for the IdS Θ_I of the whole parameter vector may be slightly conservative, but our MCMC CSs for M_I (functions of the IdS) could still be asymptotically exact. Monte Carlo experiments showcase the good finite-sample coverage properties of our proposed CS constructions in standard difficult situations. We also illustrate our proposed CSs in two relevant empirical examples.

There are numerous extensions we plan to address in the future. The first natural extension is to allow for semiparametric likelihood or moment based models involving unknown and possibly partially-identified nuisance functions. We think this paper's MCMC approach could be extended to the partially-identified sieve MLE based inference in Chen, Tamer, and Torgovitsky (2011). A second extension is to allow for structural models with latent state variables. Finally, another extension is to study the case with possibly misspecified likelihoods.

	full model				fixed s			
	Proc 1		χ^2		Proc 1		χ^2	
	90%	95%	90%	95%	90%	95%	90%	95%
β^0_{OA}	[0.391 0.643]	[0.381 0.655]	[0.481 0.541]	[0.473 0.548]	[0.386 0.591]	[0.380 0.598]	[0.464 0.516]	[0.457 0.524]
β^{MP}_{OA}	[0.345 0.663]	[0.345 0.663]	[0.459 0.541]	[0.451 0.548]	[0.432 0.693]	[0.429 0.702]	[0.526 0.601]	[0.518 0.608]
β^{MS}_{OA}	[0.301 0.608]	[0.301 0.624]	[0.419 0.493]	[0.405 0.509]	[0.298 0.529]	[0.291 0.532]	[0.379 0.445]	[0.372 0.453]
Δ_{OA}	[-1.637 -0.997]	[-1.637 -0.963]	[-1.482 -1.363]	[-1.490 -1.349]	[-1.618 -1.247]	[-1.626 -1.230]	[-1.501 -1.404]	[-1.516 -1.397]
β^0_{LC}	[-0.999 -0.564]	[-0.999 -0.564]	[-0.942 -0.778]	[-0.982 -0.750]	[-0.980 -0.550]	[-0.999 -0.538]	[-0.839 -0.733]	[-0.856 -0.719]
β^{MP}_{LC}	[1.521 1.917]	[1.508 1.919]	[1.683 1.818]	[1.683 1.832]	[1.521 1.842]	[1.509 1.858]	[1.642 1.742]	[1.642 1.755]
β^{MS}_{LC}	[0.135 0.502]	[0.131 0.512]	[0.270 0.373]	[0.257 0.381]	[0.244 0.499]	[0.236 0.507]	[0.338 0.404]	[0.332 0.412]
Δ_{LC}	[-1.627 -1.103]	[-1.645 -1.100]	[-1.431 -1.298]	[-1.447 -1.278]	[-1.629 -1.268]	[-1.637 -1.249]	[-1.488 -1.391]	[-1.496 -1.377]
ρ	[0.807 0.995]	[0.803 0.996]	[0.918 0.969]	[0.904 0.969]	[0.909 0.996]	[0.901 0.997]	[0.952 0.985]	[0.947 0.987]
s					[0.901 0.989]	[0.895 1.000]	[0.937 0.973]	[0.933 0.975]
s_{111}	[0.897 1.000]	[0.895 1.000]	[0.975 1.000]	[0.970 1.000]				
s_{110}	[0.291 0.941]	[0.276 0.947]	[0.633 0.798]	[0.619 0.831]				
s_{101}	[0.000 0.999]	[0.000 0.999]	[0.000 1.000]	[0.000 1.000]				
s_{100}	[0.045 0.999]	[0.022 0.999]	[0.858 0.995]	[0.824 0.997]				
s_{011}	[0.851 1.000]	[0.843 1.000]	[0.965 1.000]	[0.955 1.000]				
s_{010}	[0.703 1.000]	[0.682 1.000]	[0.895 1.000]	[0.870 1.000]				
s_{001}	[0.000 0.991]	[0.000 0.994]	[0.000 0.856]	[0.000 0.901]				
s_{000}	[0.186 0.986]	[0.184 0.986]	[0.752 0.934]	[0.722 0.943]				

Table 12: Results for our Procedure 1 (Proc 1) and Procedure 3 (χ^2) at 90% and 95% confidence level. The *full model* contains a general specification for equilibrium selection while the *fixed* s restricts s to be the same across markets with different regressor values.

References

Andrews, D. (1999). Estimation when a parameter is on a boundary. *Econometrica 67*(6), 1341–1383.

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

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

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

Andrews, I. and A. Mikusheva (2016). A geometric approach to nonlinear econometric models. forthcoming in econometrica, Harvard and M.I.T.

Armstrong, T. (2014). Weighted ks statistics for inference on conditional moment inequalities. *Journal of Econometrics 181*(2), 92–116.

Azaïs, J.-M., E. Gassiat, and C. Mercadier (2009). The likelihood ratio test for general mixture models with or without structural parameter. *ESAIM: Probability and Statistics 13*, 301–327.

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

Berry, S. T. (1992). Estimation of a model of entry in the airline industry. *Econometrica: Journal of the Econometric Society*, 889–917.

Besag, J., P. Green, D. Higdon, and K. Mengersen (1995). Bayesian computation and stochastic systems. *Statistical Science 10*(1), 3–4 1.

Bochkina, N. A. and P. J. Green (2014). The Bernstein-von Mises theorem and nonregular models. *The Annals of Statistics 42*(5), 1850–1878.

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 X. Shi (2016). Inference for subvectors and other functions of partially identified parameters in moment inequality models. *Quantitative Economics forthcoming.*

Burnside, C. (1998). Solving asset pricing models with gaussian shocks. *Journal of Economic Dynamics and Control 22*(3), 329 – 340.

Canay, I. A. (2010). El inference for partially identified models: Large deviations optimality and bootstrap validity. *Journal of Econometrics 156*(2), 408–425.

Chen, X., E. Tamer, and A. Torgovitsky (2011). Sensitivity analysis in semiparametric likelihood models. Cowles foundation discussion paper no 1836, Yale and Northwestern.

Chernozhukov, V. and H. Hong (2003). An mcmc approach to classical estimation. *Journal of Econometrics 115*(2), 293 – 346.

Chernozhukov, V. and H. Hong (2004). Likelihood estimation and inference in a class of non-regular econometric models. *Econometrica 72*(5), 1445–1480.

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

Fan, J., H.-N. Hung, and W.-H. Wong (2000). Geometric understanding of likelihood ratio statistics. *Journal of the American Statistical Association 95*(451), 836–841.

Federer, H. (1996). *Geometric Measure Theory.* Springer: New York.

Flinn, C. and J. Heckman (1982). New methods for analyzing structural models of labor force dynamics. *Journal of Econometrics 18*(1), 115–168.

Gao, F. (2016). A reverse gaussian correlation inequality by adding cones. Mimeo.

Gassiat, E. (2002). Likelihood ratio inequalities with application to various mixtures. *Ann. I. H. Poincaré – PR 38*(6), 897–906.

Gelman, A., G. O. Roberts, and W. R. Gilks (1996). Efficient metropolis jumping rules. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and F. M. Smith (Eds.), *Bayesian Statistics*. Oxford: Oxford University Press.

Ghosal, S., J. K. Ghosh, and A. W. van der Vaart (2000). Convergence rates of posterior distributions. *The Annals of Statistics 28*(2), 500–531.

Götze, F. (1991). On the rate of convergence in the multivariate clt. *The Annals of Probability 19*(2), 724–739.

Hallin, M., R. van den Akker, and B. Werker (2015). On quadratic expansions of log-likelihoods and a general asymptotic linearity result. In M. Hallin, D. M. Mason, D. Pfeifer, and J. G. Steinebach (Eds.), *Mathematical Statistics and Limit Theorems*, pp. 147–165. Springer International Publishing.

Hansen, L. P., J. Heaton, and A. Yaron (1996). Finite-sample properties of some alternative gmm estimators. *Journal of Business & Economic Statistics 14*(3), 262–280.

Helpman, E., M. Melitz, and Y. Rubinstein (2008). Estimating trade flows: Trading partners and trading volumes. *The Quarterly Journal of Economics 123*(2), 441–487.

Hirano, K. and J. Porter (2003). Asymptotic efficiency in parametric structural models with parameter-dependent support. *Econometrica 71*(5), 1307–1338.

Hiriart-Urruty, J.-B. and C. Lemaréchal (2001). *Fundamentals of Convex Analysis*. Springer.

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

Kaido, H., F. Molinari, and J. Stoye (2016). Confidence intervals for projections of partially identified parameters. Working paper, Boston University and Cornell.

Kitagawa, T. (2012). Estimation and inference for set-identified parameters using posterior lower probability. Working paper, University College London.

Kleijn, B. and A. van der Vaart (2012). The Bernstein-Von-Mises theorem under misspecification. *Electronic Journal of Statistics 6*, 354–381.

Kline, B. and E. Tamer (2015). Bayesian inference in a class of partially identified models. Working paper, UT Austin and Harvard.

Kocherlakota, N. R. (1990). On tests of representative consumer asset pricing models. *Journal of Monetary Economics 26*(2), 285 – 304.

Le Cam, L. and G. L. Yang (1990). *Asymptotics in Statistics: Some Basic Concepts*. Springer.

Liao, Y. and A. Simoni (2015). Posterior properties of the support function for set inference. Working paper, University of Maryland and CREST.

Liu, J. S. (2004). *Monte Carlo Strategies in Scientific Computing*. Springer.

Liu, X. and Y. Shao (2003). Asymptotics for likelihood ratio tests under loss of identifiability. *The Annals of Statistics 31*(3), 807–832.

Moon, H. R. and F. Schorfheide (2012). Bayesian and frequentist inference in partially identified models. *Econometrica 80*(2), pp. 755–782.

Müller, U. K. (2013). Risk of Bayesian inference in misspecified models, and the sandwich covariance matrix. *Econometrica 81*(5), 1805–1849.

Newey, W. K. and D. McFadden (1994). Chapter 36 large sample estimation and hypothesis testing. Volume 4 of *Handbook of Econometrics*, pp. 2111–2245. Elsevier.

Norets, A. and X. Tang (2014). Semiparametric inference in dynamic binary choice models. *The Review of Economic Studies 81*(3), 1229–1262.

Robert, C. P. and G. Casella (2004). *Monte Carlo Statistical Methods*. Springer: New York.

Roberts, G. O., A. Gelman, and W. R. Gilks (1997). Weak convergence and optimal scaling of random walk metropolis algorithms. *The Annals of Applied Probability 7*(1), 110–120.

Romano, J., A. Shaikh, and M. Wolf (2014). Inference for the identified set in partially identified econometric models. *Econometrica*.

Romano, J. P. and A. M. Shaikh (2010). Inference for the identified set in partially identified econometric models. *Econometrica 78*(1), 169–211.

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

Santos Silva, J. M. C. and S. Tenreyro (2015). Trading partners and trading volumes: Implementing the helpmanmelitzrubinstein model empirically. *Oxford Bulletin of Economics and Statistics 77*(1), 93–105.

Stock, J. H. and J. H. Wright (2000). Gmm with weak identification. *Econometrica 68*(5), 1055–1096.

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

Tauchen, G. and R. Hussey (1991). Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models. *Econometrica 59*(2), pp. 371–396.

van der Vaart, A. W. (2000). *Asymptotic statistics*. Cambridge University Press.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag.

A Additional Monte Carlo evidence

A.1 Missing data example

Figure 6 plots the marginal "curved" priors for β and ρ. Figure 7 plots the reduced-form parameters evaluated at the MCMC chain for the structural parameters presented in Figure 1. Although the partially-identified structural parameters μ and β bounce around their respective identified sets, the reduced-form chains in Figure 7 are stable.

A.2 Complete information game

Figure 8 presents the MCMC chain for the structural parameters computed from one simulated data set with $n = 1000$ using a likelihood objective function and a flat prior on Θ. Figure 9 presents the reduced-form probabilities calculated from the chain in Figure 8.

A.3 Euler equations

We simulate data using the design in Hansen et al. (1996) (also used by Kocherlakota (1990) and Stock and Wright (2000)).[28] The simulation design has a representative agent with CRRA preferences indexed by δ (discount rate) and γ (risk-aversion parameter) and a representative dividend-paying asset. The design has log consumption growth c_{t+1} and log dividend growth on a representative asset d_{t+1} evolving as a bivariate VAR(1), with:

$$\begin{pmatrix} d_{t+1} \\ c_{t+1} \end{pmatrix} = \begin{pmatrix} 0.004 \\ 0.021 \end{pmatrix} + \begin{pmatrix} 0.117 & 0.414 \\ 0.017 & 0.161 \end{pmatrix} \begin{pmatrix} d_t \\ c_c \end{pmatrix} + \varepsilon_{t+1}$$

where the ε_{t+1} are i.i.d normal with mean zero and covariance matrix:

$$\begin{pmatrix} 0.01400 & 0.00177 \\ 0.00177 & 0.00120 \end{pmatrix} .$$

Previous studies use the Tauchen and Hussey (1991) method to simulate the data based on a discretized system. Unlike the previous studies, we simulate the VAR directly and use Burnside (1998)'s formula for the price dividend ratio to calculate the return. Therefore we do not incur any numerical approximation error due to discretization.

The only return used in the Euler equation is the gross stock return R_{t+1}, with a constant, lagged consumption growth, and lagged returns used as instruments. Thus the GMM model is:

$$E\left[\left(\delta G_{t+1}^{-\gamma} R_{t+1} - 1\right) \otimes z_t\right] = 0$$

[28] We are grateful to Lars Peter Hansen for suggesting this simulation exercise.

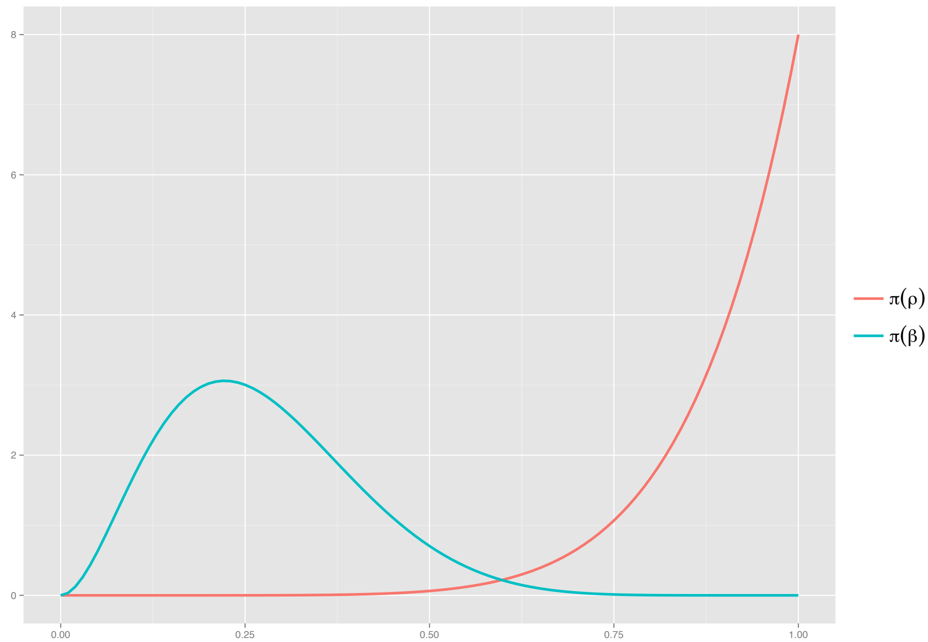

Figure 6: Marginal curved priors for β and ρ for the missing data example.

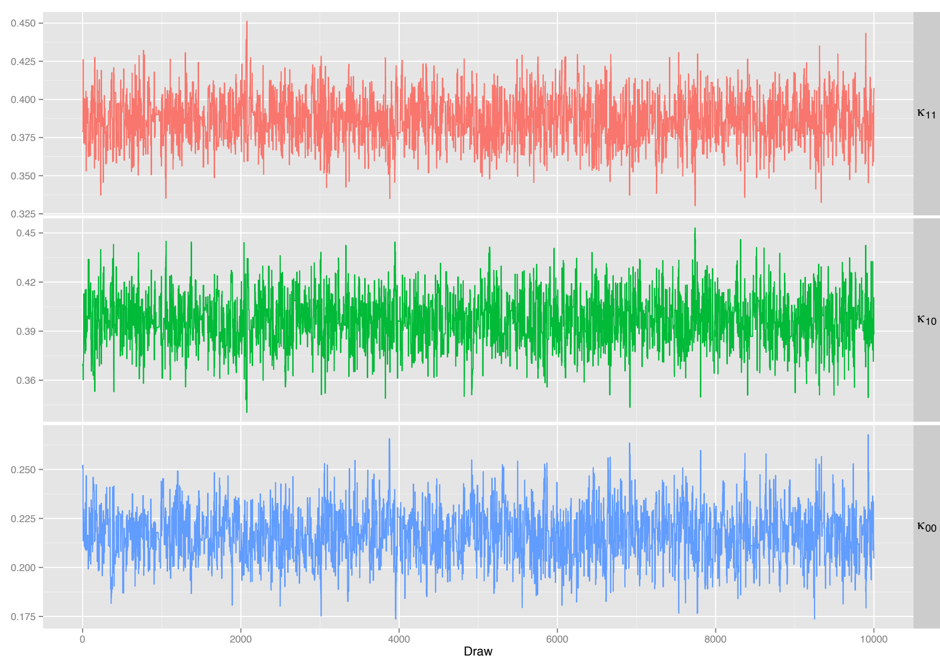

Figure 7: MCMC chain for the reduced-form probabilities $(\kappa_{11}(\theta), \kappa_{10}(\theta), \kappa_{00}(\theta))'$ calculated from the chain in Figure 1. It is clear the chain for the reduced-form probabilities has converged even though the chain for the structural parameters has not.

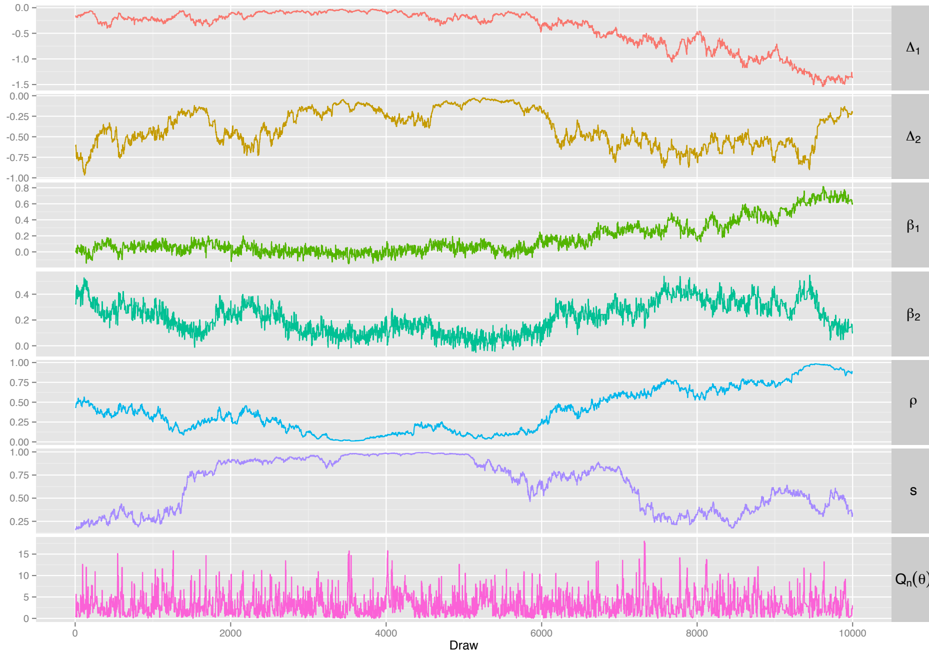

Figure 8: MCMC chain for all structural parameters (top 6 panels) and QLR (bottom panel) with $n = 1000$ using a likelihood for L_n and a flat prior on Θ.

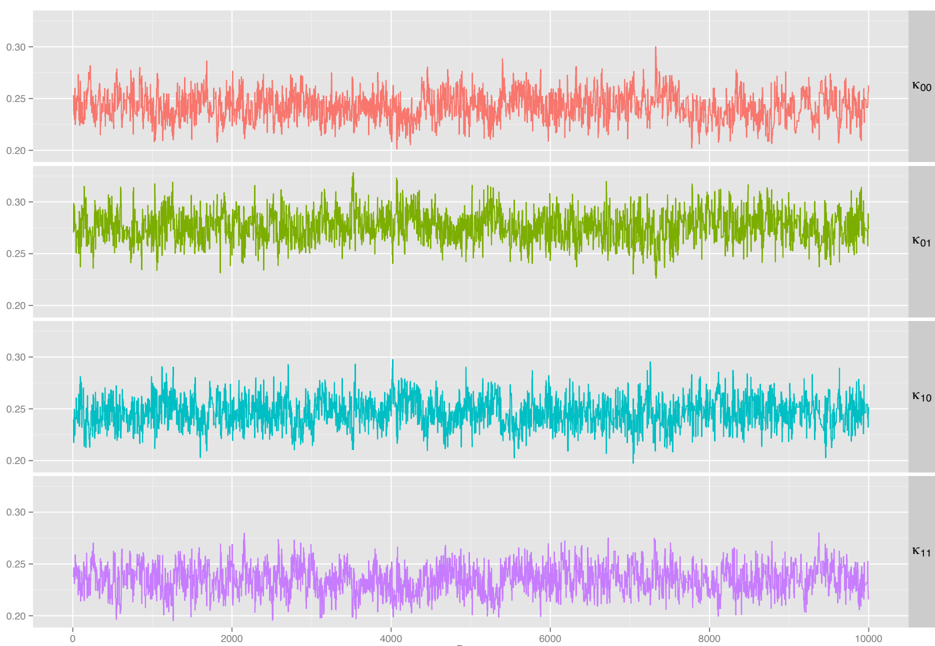

Figure 9: MCMC chain for the reduced-form probabilities calculated from the chain in Figure 8. It is clear that the chain for the reduced-form probabilities has converged, even though the chain for the structural parameters from which they are calculated has not.

with $G_{t+1} = \exp(c_{t+1})$ and $z_t = (1, G_t, R_t)'$. We use a continuously-updated GMM objective function. We again use samples of size $n = 100$, 250, 500, and 1000 with (δ, γ) sampled from the quasi-posterior using a random walk Metropolis Hastings sampler with acceptance rate tuned to be approximately one third. We take a flat prior and vary (δ, γ) in the DGP and the support of the prior.

The model is (weakly) point identified. However, Figure 10 shows that the criterion contains very little information about the true parameters even with $n = 500$. The chain for γ bounces around the region $[10, 40]$ and the chain for δ bounces around $[0.8, 1.05]$. The chain is drawn from the quasi-posterior with a flat prior on $[0, 6, 1.1] \times [0, 40]$. This suggests that conventional percentile-based confidence intervals for δ and γ following Chernozhukov and Hong (2003) may be highly sensitive to the prior. Figure 11 shows a scatter plot of the (δ, γ) chain which illustrates further the sensitivity of the draws to the prior.

Tables 13 and 14 present coverage properties of our Procedure 1 for the full set $\widehat{\Theta}_\alpha$ (CCOT θ in the tables) together with our Procedure 2 for the identified set for δ and γ (CCOT δ and CCOT γ in the tables). Here our Procedure 3 coincides with confidence sets based on inverting the "constrained-minimized" QLR statistic suggested in Hansen et al. (1996) (HHY δ and HHY γ in the tables). We also present the coverage properties of confidence sets formed from the upper and lower $100(1-\alpha)/2$ quantiles of the MCMC chains for γ and δ (i.e. the Chernozhukov and Hong (2003) procedure; CH in the tables) and conventional confidence intervals based on inverting t-statistics (Asy in the tables).

Overall the results are somewhat sensitive to the support for the parameters, even for the full identified set. Results that construct the confidence sets using the quantiles of the actual chain of parameters (CH in the Tables) do not perform well, but whether it over/under covers seems to depend on the support of the prior. For instance, CH is conservative in Table 13 but undercovers badly for γ even with $n = 500$ in Table 14. Confidence sets based on the profiled QLR statistic from the MCMC chain appear to perform better, but can over or under cover by a few percentage points in samples of $n = 100$ and $n = 250$.

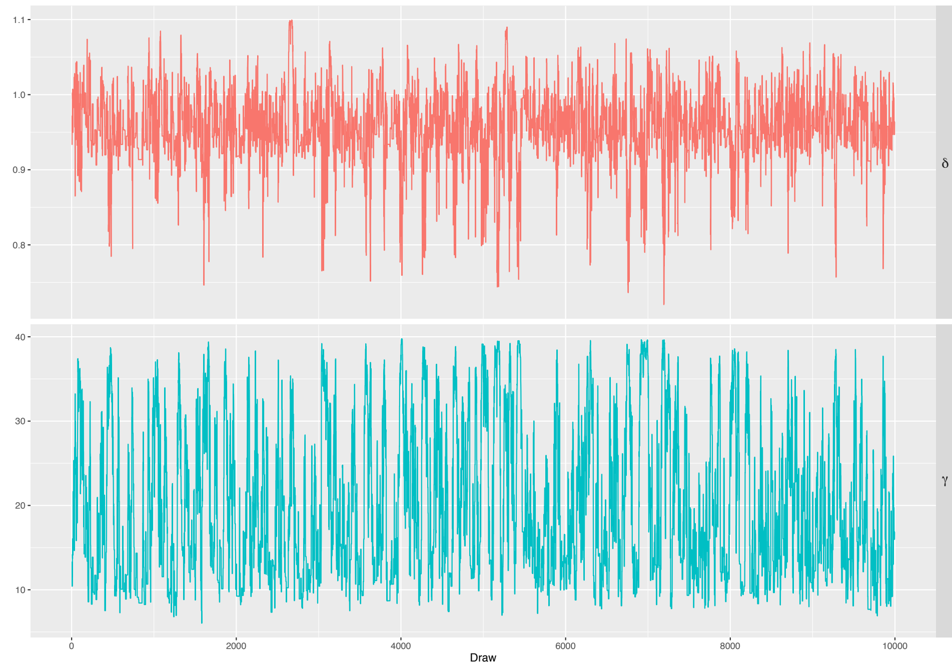

Figure 10: Plots of the MCMC chain for the structural parameter $\theta = (\delta, \gamma)$ with $n = 250$, $\theta_0 = (0.97, 10)$ and a flat prior on $\Theta = [0.6, 1.1] \times [0, 40]$.

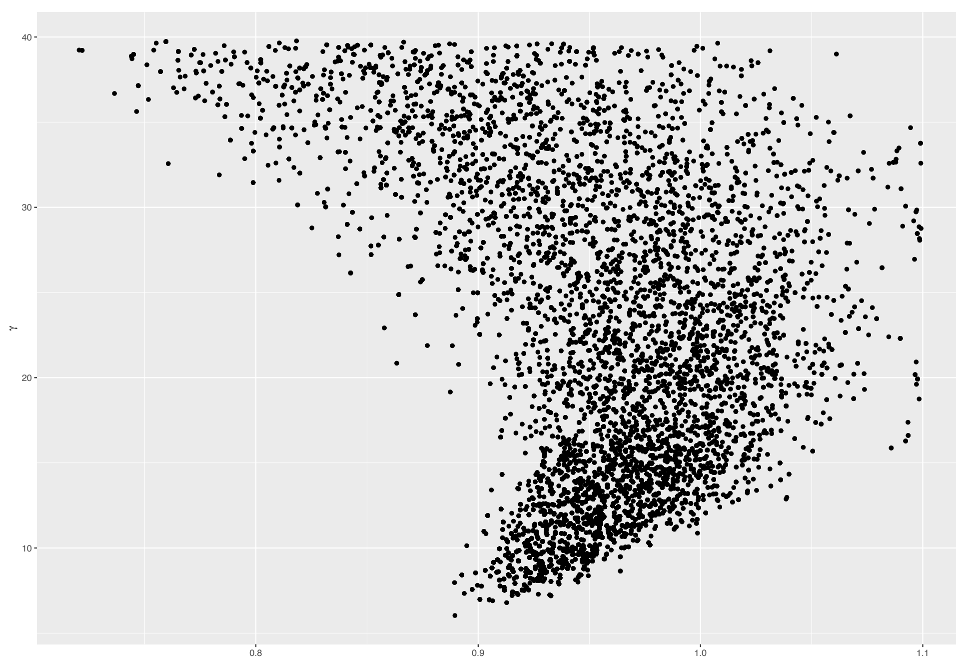

Figure 11: Scatter plot of the chain depicted in 10.

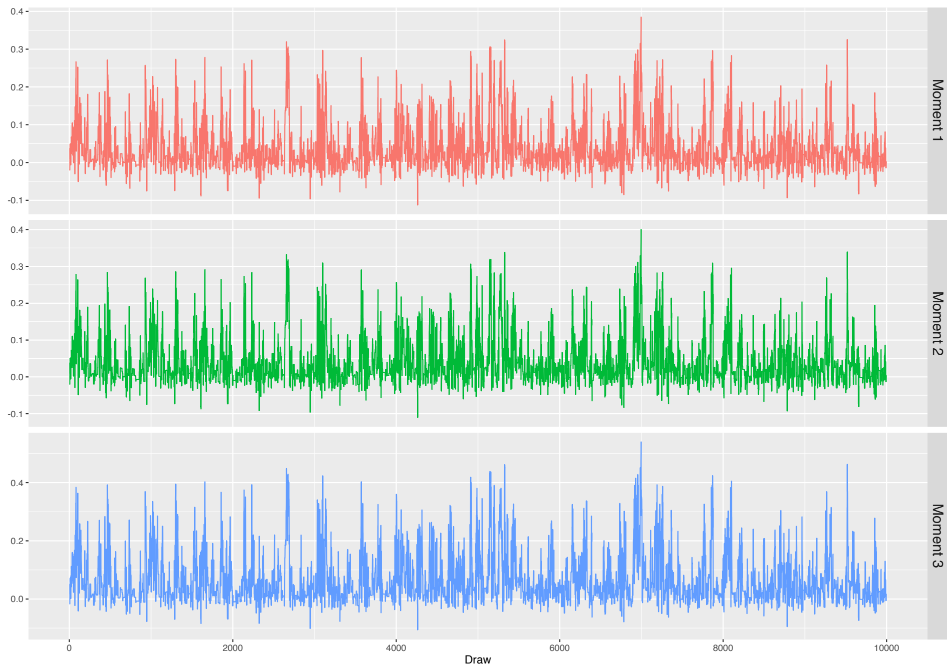

Figure 12: Plots of the moments calculated from the chain in Figure 10.

	CCOT θ	CCOT δ	CCOT γ	HHY δ	HHY γ	CH δ	CH γ
				$n = 100$			
$\alpha = 0.90$	0.8796	0.9478	0.9554	0.9344	0.8584	0.9900	0.9886
$\alpha = 0.95$	0.9388	0.9858	0.9870	0.9728	0.8954	0.9974	0.9950
$\alpha = 0.99$	0.9860	0.9996	0.9982	0.9940	0.9364	1.0000	0.9998
				$n = 250$			
$\alpha = 0.90$	0.8828	0.9492	0.9542	0.9184	0.8716	0.9860	0.9874
$\alpha = 0.95$	0.9360	0.9844	0.9846	0.9596	0.9076	0.9958	0.9940
$\alpha = 0.99$	0.9836	0.9990	0.9976	0.9908	0.9330	0.9996	0.9990
				$n = 500$			
$\alpha = 0.90$	0.8848	0.9286	0.9230	0.9038	0.8850	0.9764	0.9708
$\alpha = 0.95$	0.9404	0.9756	0.9720	0.9548	0.9312	0.9900	0.9894
$\alpha = 0.99$	0.9888	0.9974	0.9972	0.9856	0.9594	0.9986	0.9988
				$n = 1000$			
$\alpha = 0.90$	0.8840	0.8842	0.8774	0.9056	0.8984	0.9514	0.9518
$\alpha = 0.95$	0.9440	0.9540	0.9548	0.9532	0.9516	0.9812	0.9796
$\alpha = 0.99$	0.9866	0.9954	0.9938	0.9898	0.9852	0.9968	0.9972

Table 13: MC coverage probabilities for $\delta = 0.97 \in [0.8, 1]$, $\gamma = 1.3 \in [0, 10]$.

	CCOT θ	CCOT δ	CCOT γ	HHY δ	HHY γ	CH δ	CH γ
				$n = 100$			
$\alpha = 0.90$	0.8212	0.9098	0.7830	0.8940	0.8764	0.9658	0.3434
$\alpha = 0.95$	0.8820	0.9564	0.8218	0.9394	0.9288	0.9886	0.4954
$\alpha = 0.99$	0.9614	0.9934	0.8780	0.9846	0.9732	0.9984	0.8098
				$n = 250$			
$\alpha = 0.90$	0.8774	0.9538	0.8560	0.8758	0.8914	0.9768	0.4068
$\alpha = 0.95$	0.9244	0.9784	0.8908	0.9260	0.9468	0.9920	0.5402
$\alpha = 0.99$	0.9756	0.9982	0.9392	0.9780	0.9856	0.9990	0.7552
				$n = 500$			
$\alpha = 0.90$	0.9116	0.9600	0.9060	0.8668	0.8952	0.9704	0.5504
$\alpha = 0.95$	0.9494	0.9866	0.9412	0.9136	0.9504	0.9892	0.6130
$\alpha = 0.99$	0.9880	0.9978	0.9758	0.9640	0.9890	0.9986	0.7070
				$n = 1000$			
$\alpha = 0.90$	0.9046	0.9134	0.8952	0.8838	0.8988	0.9198	0.8864
$\alpha = 0.95$	0.9582	0.9614	0.9556	0.9216	0.9528	0.9586	0.9284
$\alpha = 0.99$	0.9882	0.9930	0.9922	0.9594	0.9914	0.9884	0.9600

Table 14: MC coverage probabilities for $\delta = 0.97 \in [0.6, 1.1]$, $\gamma = 1.3 \in [0, 40]$.

A.4 Gaussian mixtures

Consider the bivariate normal mixture where each X_i is iid with density f given by:

$$f(x_i) = \eta\phi(x_i - \mu) + (1 - \eta)\phi(x_i)$$

where $\eta \in [0, 1]$ is the mixing weight and $\mu \in [-M, M]$ is the location parameter and ϕ is the standard normal pdf. We restrict μ to have compact support because of Hartigan (1985). If $\mu = 0$ or $\eta = 0$ then the model is partially identified and the identified set for $\theta = (\mu, \eta)'$ is $[-M, M] \times \{0\} \cup \{0\} \times [0, 1]$. However, if $\mu \neq 0$ and $\eta > 0$ then the model is point identified.

We are interested in doing inference on the identified set M_I for μ and H_I for η. For each simulation, we simulate a chain $\theta^1, \ldots, \theta^B$ using Gibbs sampling.[29] We calculate the profile QLR ratio for μ, which is:

$$\begin{cases} L_n(\hat{\theta}) - \sup_{\eta \in [0,1]} L_n(\mu^b, \eta) & \text{if both } \mu^b \neq 0 \text{ and } \eta^b > 0 \\ L_n(\hat{\theta}) - \min_{\mu \in [-M,M]} \sup_{\eta \in [0,1]} L_n(\mu, \eta) & \text{else} \end{cases}$$

and the profile QLR ratio for η, which is:

$$\begin{cases} L_n(\hat{\theta}) - \sup_{\mu \in [-M,M]} L_n(\mu, \eta^b) & \text{if both } \mu^b \neq 0 \text{ and } \eta^b > 0 \\ L_n(\hat{\theta}) - \min_{\eta \in [0,1]} \sup_{\mu \in [-M,M]} L_n(\mu, \eta) & \text{else}\,. \end{cases}$$

We take the 100α percentile of the QLRs and call them ξ^{μ}_{α} and ξ^{η}_{α}. Confidence sets for M_I and

[29] Unlike the previous examples, here we use hierarchical Gibbs sampling instead of a random walk Metropolis-Hastings algorithm as it allows us to draw exactly from the posterior.

H_I (using Procedure 2) are:

$$\widehat{M}_\alpha = \left\{ \mu \in [-M, M] : L_n(\hat{\theta}) - \sup_{\eta \in [0,1]} L_n(\mu, \eta) \leq \xi_\alpha^\mu \right\}$$
$$\widehat{H}_\alpha = \left\{ \eta \in [0, 1] : L_n(\hat{\theta}) - \sup_{\mu \in [-M,M]} L_n(\mu, \eta) \leq \xi_\alpha^\eta \right\}.$$

Unlike the missing data and game models, here the set of parameters θ under which the model is partially identified is a set of measure zero in the full parameter space. So naïve MCMC sampling won't going to give us the correct critical values when the model is partially identified unless we choose a prior that puts positive probability on the partially identified region.

Therefore, we use a truncated normal prior for μ:

$$\pi(\mu) = \frac{1}{\Phi(\frac{M-a}{b}) - \Phi(\frac{-M-a}{b})} \frac{1}{b\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\mu-a}{b}\right)^2} 1\!\mathrm{l}\{\mu \in [-M, M]\}$$

with hyperparameters (a, b). Conjugate beta priors for η are most commonly used. However, they do not assign positive probability to $\eta = 0$. Instead we take the following empirical Bayes approach. Let:

$$\pi(\eta) = q\delta_0 + (1 - q) f_{B(\alpha,\beta)}(\eta)$$

where $q \in [0, 1]$, δ_0 is point mass at the origin, and $B(\alpha, \beta)$ is the Beta distribution pdf. We'll treat the hyperparameters α, β, a, b as fixed but estimate the mixing proportion q from the data. The posterior distribution for $\theta = (\mu, \eta)$ is:

$$\Pi((\mu, \eta)|\mathbf{X}_n; q) = \frac{e^{L_n(\theta)} \pi(\mu)\pi(\eta|q)}{\int_0^1 \int_{-M}^M e^{L_n(\theta)} \pi(\mu)\pi(\eta|q) d\mu d\eta}.$$

The denominator is proportional to the marginal distribution for $\mathbf{X}_n$ given q. For the "empirical Bayes" bit we choose q to maximize this expression. Therefore, we choose:

$$\hat{q} = \begin{cases} 1 & \text{if } \prod_{i=1}^n \phi(X_i) \geq \int_{-M}^M \int_0^1 \prod_{i=1}^n (\eta\phi(X_i - \mu) + (1 - \eta)\phi(X_i)) f_{B(\alpha,\beta)}(\eta)\pi(\mu) d\eta d\mu \\ 0 & \text{else}. \end{cases}$$

We then plug $\hat{q}$ back in to the prior for η. The posterior distribution we use for the MCMC chain is:

$$\Pi((\mu, \eta)|\mathbf{X}_n; \hat{q}) = \frac{e^{L_n(\theta)} \pi(\mu)\pi(\eta|\hat{q})}{\int \int e^{L_n(\theta)} \pi(\mu)\pi(\eta|\hat{q}) d\mu d\eta}.$$

where $\pi(\mu)$ is as above and

$$\pi(\eta|\hat{q}) = \begin{cases} \delta_0 & \text{if } q = 1 \\ f_{B(\alpha,\beta)} & \text{if } q = 0. \end{cases}$$

When $\hat{q} = 1$ we have $\eta = 0$ for every draw, and when $q = 0$ we can use the hierarchical Gibbs method to draw μ and η.

For the simulations we take $M = 3$ with $\mu_0 = 1$. The prior for μ is a $N(0, 1)$ truncated to $[-M, M]$. We take $\alpha = 1.5$ and $\beta = 3$ in the prior for η. We vary η_0, taking $\eta_0 = 0.5, 0.2, 0.1$

(point identified) and $\eta_0 = 0$ (partially identified; see Figure 13). We use 5,000 replications with chain length 10,000 and a burnin of 1,000. For confidence sets for Θ_I we use Procedure 1 with the prior $\pi(\eta) = f_{B(\alpha,\beta)}(\eta)$ with $\alpha = 1.5$ and $\beta = 3$ and $\pi(\mu)$ is a $N(0,1)$ truncated to $[-M, M]$. We again use a hierarchical Gibbs sampler with chain length 10,000 and burnin of 1,000.

The first two Tables 15 and 16 present coverage probabilities of $\widehat{M}_\alpha$ and $\widehat{H}_\alpha$ using Procedure 2. Our procedure is valid but conservative in the partially identified case (here the identified set for the subvectors μ and η is the full parameter space which is why the procedure is conservative). However the method under-covers for moderate sample sizes when the mixing weight is small but nonzero. Tables 17 and 18 present results using our Procedure 3. This works well as expected under point identification (since the QLR is exactly χ^2_1 in this case). Under partial identification this method performs poorly for M_I. The final Table 19 presents coverage probabilities of $\widehat{\Theta}_\alpha$ using Procedure 1 which shows that its coverage is good in both the point and partially-identified cases, though again it can under-cover slightly in small to moderate sample sizes when the mixing weight is close to zero.

	$\eta_0 = 0.50$	$\eta_0 = 0.20$	$\eta_0 = 0.10$	$\eta_0 = 0.00$
		$n = 100$		
$\alpha = 0.90$	0.9368	0.9760	0.9872	0.9712
$\alpha = 0.95$	0.9782	0.9980	0.9980	0.9712
$\alpha = 0.99$	0.9968	0.9996	0.9994	0.9712
avg $\hat{q}$	0.0052	0.5634	0.8604	0.9712
		$n = 250$		
$\alpha = 0.90$	0.8884	0.8646	0.9322	0.9838
$\alpha = 0.95$	0.9514	0.9522	0.9794	0.9838
$\alpha = 0.99$	0.9938	0.9978	0.9998	0.9838
avg $\hat{q}$	0.0000	0.2278	0.7706	0.9838
		$n = 500$		
$\alpha = 0.90$	0.8826	0.8434	0.8846	0.9886
$\alpha = 0.95$	0.9396	0.9090	0.9346	0.9886
$\alpha = 0.99$	0.9880	0.9892	0.9944	0.9886
avg $\hat{q}$	0.0000	0.0324	0.6062	0.9886
		$n = 1000$		
$\alpha = 0.90$	0.8900	0.8844	0.8546	0.9888
$\alpha = 0.95$	0.9390	0.9208	0.8906	0.9888
$\alpha = 0.99$	0.9882	0.9776	0.9798	0.9888
avg $\hat{q}$	0.0000	0.0002	0.3150	0.9888
		$n = 2500$		
$\alpha = 0.90$	0.8932	0.9010	0.8970	0.9942
$\alpha = 0.95$	0.9454	0.9456	0.9236	0.9942
$\alpha = 0.99$	0.9902	0.9842	0.9654	0.9942
avg $\hat{q}$	0.0000	0.0000	0.0166	0.9942

Table 15: MC coverage probabilities for $\widehat{M}_\alpha$ (Procedure 2).

	$\eta_0 = 0.50$	$\eta_0 = 0.20$	$\eta_0 = 0.10$	$\eta_0 = 0.00$
		$n = 100$		
$\alpha = 0.90$	0.9470	0.9252	0.8964	0.9742
$\alpha = 0.95$	0.9820	0.9718	0.9438	0.9752
$\alpha = 0.99$	0.9986	0.9970	0.9902	0.9768
avg $\hat{q}$	0.0052	0.5634	0.8604	0.9712
		$n = 250$		
$\alpha = 0.90$	0.9008	0.8886	0.8744	0.9864
$\alpha = 0.95$	0.9594	0.9520	0.9288	0.9872
$\alpha = 0.99$	0.9956	0.9926	0.9898	0.9882
avg $\hat{q}$	0.0000	0.2278	0.7706	0.9838
		$n = 500$		
$\alpha = 0.90$	0.8826	0.8798	0.8508	0.9900
$\alpha = 0.95$	0.9432	0.9356	0.9118	0.9902
$\alpha = 0.99$	0.9918	0.9890	0.9764	0.9908
avg $\hat{q}$	0.0000	0.0324	0.6062	0.9886
		$n = 1000$		
$\alpha = 0.90$	0.8892	0.8900	0.8582	0.9922
$\alpha = 0.95$	0.9440	0.9314	0.9076	0.9922
$\alpha = 0.99$	0.9886	0.9842	0.9722	0.9928
avg $\hat{q}$	0.0000	0.0002	0.3150	0.9888
		$n = 2500$		
$\alpha = 0.90$	0.8938	0.8956	0.9022	0.9954
$\alpha = 0.95$	0.9460	0.9460	0.9342	0.9956
$\alpha = 0.99$	0.9870	0.9866	0.9730	0.9962
avg $\hat{q}$	0.0000	0.0000	0.0166	0.9942

Table 16: MC coverage probabilities for $\widehat{H}_\alpha$ (Procedure 2).

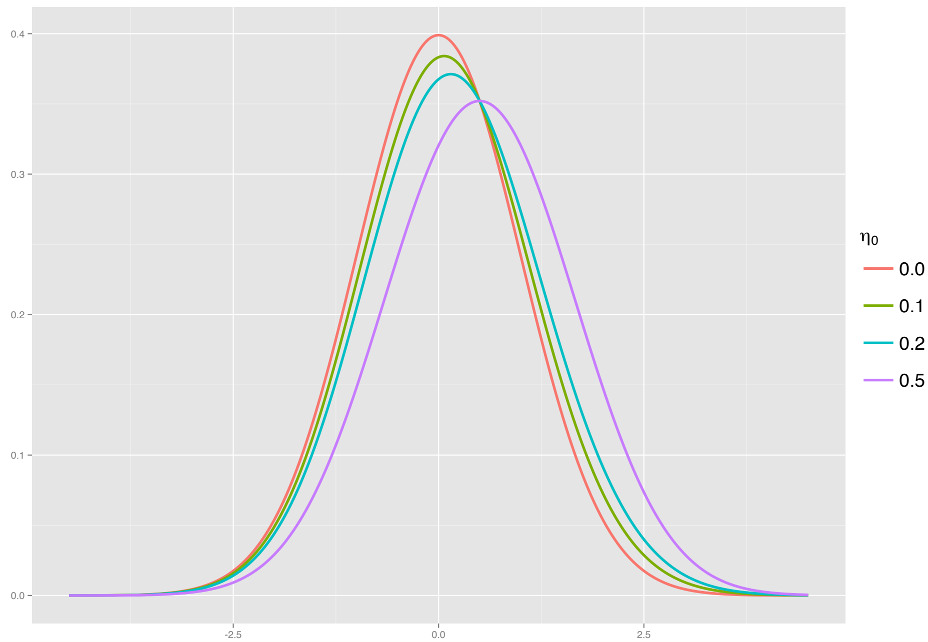

Figure 13: PDFs for the normal mixture MC design for different values of η_0.

	$\eta_0 = 0.50$	$\eta_0 = 0.20$	$\eta_0 = 0.10$	$\eta_0 = 0.00$
	$n = 100$			
$\alpha = 0.90$	0.8978	0.9190	0.9372	0.8208
$\alpha = 0.95$	0.9516	0.9684	0.9718	0.9020
$\alpha = 0.99$	0.9938	0.9958	0.9954	0.9796
	$n = 250$			
$\alpha = 0.90$	0.8996	0.8960	0.9180	0.8248
$\alpha = 0.95$	0.9514	0.9486	0.9602	0.9042
$\alpha = 0.99$	0.9882	0.9926	0.9944	0.9752
	$n = 500$			
$\alpha = 0.90$	0.8998	0.8916	0.9030	0.8240
$\alpha = 0.95$	0.9474	0.9434	0.9500	0.9042
$\alpha = 0.99$	0.9898	0.9874	0.9904	0.9756
	$n = 1000$			
$\alpha = 0.90$	0.9028	0.9026	0.8984	0.8214
$\alpha = 0.95$	0.9514	0.9538	0.9502	0.8986
$\alpha = 0.99$	0.9902	0.9912	0.9930	0.9788
	$n = 2500$			
$\alpha = 0.90$	0.8998	0.8966	0.8968	0.8098
$\alpha = 0.95$	0.9520	0.9489	0.9442	0.8916
$\alpha = 0.99$	0.9912	0.9902	0.9882	0.9720

Table 17: MC coverage probabilities for $\widehat{M}_\alpha^\chi$ (Procedure 3).

	$\eta_0 = 0.50$	$\eta_0 = 0.20$	$\eta_0 = 0.10$	$\eta_0 = 0.00$
		$n = 100$		
$\alpha = 0.90$	0.9024	0.9182	0.9426	0.8920
$\alpha = 0.95$	0.9528	0.9622	0.9738	0.9434
$\alpha = 0.99$	0.9916	0.9946	0.9950	0.9890
		$n = 250$		
$\alpha = 0.90$	0.8974	0.8970	0.9216	0.8948
$\alpha = 0.95$	0.9432	0.9466	0.9600	0.9444
$\alpha = 0.99$	0.9908	0.9894	0.9928	0.9880
		$n = 500$		
$\alpha = 0.90$	0.9026	0.8948	0.9080	0.8954
$\alpha = 0.95$	0.9472	0.9454	0.9550	0.9476
$\alpha = 0.99$	0.9886	0.9886	0.9914	0.9898
		$n = 1000$		
$\alpha = 0.90$	0.8960	0.9006	0.8964	0.8972
$\alpha = 0.95$	0.9442	0.9524	0.9476	0.9522
$\alpha = 0.99$	0.9878	0.9884	0.9892	0.9914
		$n = 2500$		
$\alpha = 0.90$	0.9052	0.9038	0.9036	0.8954
$\alpha = 0.95$	0.9504	0.9490	0.9502	0.9480
$\alpha = 0.99$	0.9906	0.9892	0.9900	0.9922

Table 18: MC coverage probabilities for $\widehat{H}^{\chi}_{\alpha}$ (Procedure 3).

	$\eta_0 = 0.50$	$\eta_0 = 0.20$	$\eta_0 = 0.10$	$\eta_0 = 0.00$
		$n = 100$		
$\alpha = 0.90$	0.9170	0.8696	0.8654	0.9294
$\alpha = 0.95$	0.9610	0.9250	0.9342	0.9724
$\alpha = 0.99$	0.9926	0.9824	0.9880	0.9960
		$n = 250$		
$\alpha = 0.90$	0.8962	0.8932	0.8682	0.9192
$\alpha = 0.95$	0.9498	0.9468	0.9358	0.9654
$\alpha = 0.99$	0.9918	0.9876	0.9872	0.9938
		$n = 500$		
$\alpha = 0.90$	0.8922	0.8842	0.8706	0.9034
$\alpha = 0.95$	0.9464	0.9464	0.9310	0.9536
$\alpha = 0.99$	0.9898	0.9902	0.9846	0.9926
		$n = 1000$		
$\alpha = 0.90$	0.8980	0.8964	0.8832	0.9134
$\alpha = 0.95$	0.9456	0.9478	0.9376	0.9594
$\alpha = 0.99$	0.9872	0.9888	0.9882	0.9932
		$n = 2500$		
$\alpha = 0.90$	0.8986	0.8960	0.9036	0.9026
$\alpha = 0.95$	0.9522	0.9466	0.9468	0.9520
$\alpha = 0.99$	0.9918	0.9886	0.9896	0.9916

Table 19: MC coverage probabilities for $\widehat{\Theta}_{\alpha}$ (Procedure 1).

B Uniformity

Let $\mathbf{P}$ denote the class of distributions over which we want the confidence sets to be uniformly valid. Let $L(\theta;\mathbb{P})$ denote the population objective function. We again assume that $L(\cdot;\mathbb{P})$ and L_n are upper semicontinuous and that $\sup_{\theta\in\Theta} L(\theta;\mathbb{P}) < \infty$ holds for each $\mathbb{P} \in \mathbf{P}$. The identified set is $\Theta_I(\mathbb{P}) = \{\theta \in \Theta : L(\theta;\mathbb{P}) = \sup_{\vartheta\in\Theta} L(\vartheta;\mathbb{P})\}$ and the identified set for a function μ of $\Theta_I(\mathbb{P})$ is $M_I(\mathbb{P}) = \{\mu(\theta) : \theta \in \Theta_I(\mathbb{P})\}$. We now show that, under slight strengthening of our regularity conditions, $\widehat{\Theta}_\alpha$ and $\widehat{M}_\alpha$ are uniformly valid, i.e.:

$$\liminf_{n\to\infty} \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\Theta_I(\mathbb{P}) \subseteq \widehat{\Theta}_\alpha) \geq \alpha \tag{23}$$

$$\liminf_{n\to\infty} \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(M_I(\mathbb{P}) \subseteq \widehat{M}_\alpha) \geq \alpha \tag{24}$$

both hold.

The following results are modest extensions of Lemmas 2.1 and 2.2. Let $(\upsilon_n)_{n\in\mathbb{N}}$ be a sequence of random variables. We say that $\upsilon_n = o_{\mathbb{P}}(1)$ uniformly for $\mathbb{P} \in \mathbf{P}$ if $\lim_{n\to\infty} \sup_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(|\upsilon_n| > \epsilon) = 0$ for each $\epsilon > 0$. We say that $\upsilon_n \leq o_{\mathbb{P}}(1)$ uniformly for $\mathbb{P} \in \mathbf{P}$ if $\lim_{n\to\infty} \sup_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\upsilon_n > \epsilon) = 0$ for each $\epsilon > 0$

Lemma B.1. *Let (i) $\sup_{\theta\in\Theta_I(\mathbb{P})} Q_n(\theta) \overset{\mathbb{P}}{\rightsquigarrow} W_{\mathbb{P}}$ where $W_{\mathbb{P}}$ is a random variable whose probability distribution is continuous at its α quantile (denoted by $w_{\alpha,\mathbb{P}}$) for each $\mathbb{P} \in \mathbf{P}$, and:*

$$\lim_{n\to\infty} \sup_{\mathbb{P}\in\mathbf{P}} \left| \mathbb{P}\Big(\sup_{\theta\in\Theta_I(\mathbb{P})} Q_n(\theta) \leq w_{\alpha,\mathbb{P}} - \eta_n \Big) - \alpha \right| = 0$$

for any sequence $(\eta_n)_{n\in\mathbb{N}}$ with $\eta_n = o(1)$; and (ii) $(w_{n,\alpha})_{n\in\mathbb{N}}$ be a sequence of random variables such that $w_{n,\alpha} \geq w_{\alpha,\mathbb{P}} + o_{\mathbb{P}}(1)$ uniformly for $\mathbb{P} \in \mathbf{P}$.
Then: (23) holds for $\widehat{\Theta}_\alpha = \{\theta \in \Theta : Q_n(\theta) \leq w_{n,\alpha}\}$. Moreover, if $w_{n,\alpha} = w_{\alpha,\mathbb{P}} + o_{\mathbb{P}}(1)$ uniformly for $\mathbb{P} \in \mathbf{P}$ then (23) holds with equality.

Lemma B.2. *Let (i) $\sup_{m\in M_I(\mathbb{P})} \inf_{\theta\in\mu^{-1}(m)} Q_n(\theta) \overset{\mathbb{P}}{\rightsquigarrow} W_{\mathbb{P}}$ where $W_{\mathbb{P}}$ is a random variable whose probability distribution is continuous at its α quantile (denoted by $w_{\alpha,\mathbb{P}}$) for each $\mathbb{P} \in \mathbf{P}$ and:*

$$\lim_{n\to\infty} \sup_{\mathbb{P}\in\mathbf{P}} \left| \mathbb{P}\Big(\sup_{m\in M_I(\mathbb{P})} \inf_{\theta\in\mu^{-1}(m)} Q_n(\theta) \leq w_{\alpha,\mathbb{P}} - \eta_n \Big) - \alpha \right| = 0$$

for any sequence $(\eta_n)_{n\in\mathbb{N}}$ with $\eta_n = o(1)$; and (ii) $(w_{n,\alpha})_{n\in\mathbb{N}}$ be a sequence of random variables such that $w_{n,\alpha} \geq w_{\alpha,\mathbb{P}} + o_{\mathbb{P}}(1)$ uniformly for $\mathbb{P} \in \mathbf{P}$.
Then: (24) holds for $\widehat{M}_\alpha = \{\mu(\theta) : \theta \in \Theta, Q_n(\theta) \leq w_{n,\alpha}\}$. Moreover, if $w_{n,\alpha} = w_{\alpha,\mathbb{P}} + o_{\mathbb{P}}(1)$ uniformly for $\mathbb{P} \in \mathbf{P}$ then (24) holds with equality.

The following regularity conditions ensure that $\widehat{\Theta}_\alpha$ and $\widehat{M}_\alpha$ are uniformly valid over $\mathbf{P}$. Let $(\Theta_{osn}(\mathbb{P}))_{n\in\mathbb{N}}$ denote a sequence of local neighborhoods of $\Theta_I(\mathbb{P})$ such that $\Theta_{osn}(\mathbb{P}) \in \mathscr{B}(\Theta)$ and $\Theta_I(\mathbb{P}) \subseteq \Theta_{osn}(\mathbb{P})$ for each n and for each $\mathbb{P} \in \mathbf{P}$. In what follows we omit the dependence of $\Theta_{osn}(\mathbb{P})$ on $\mathbb{P}$ to simplify notation.

Assumption B.1. *(Consistency, posterior contraction)*
(i) $L_n(\hat{\theta}) = \sup_{\theta \in \Theta_{osn}} L_n(\theta) + o_{\mathbb{P}}(n^{-1})$ *uniformly for* $\mathbb{P} \in \mathbf{P}$.
(ii) $\Pi_n(\Theta_{osn}^c | \mathbf{X}_n) = o_{\mathbb{P}}(1)$ *uniformly for* $\mathbb{P} \in \mathbf{P}$.

We restate our conditions on local quadratic approximation of the criterion allowing for singularity. Recall that a local reduced-form reparameterization is defined on a neighborhood Θ_I^N of Θ_I. We require that $\Theta_{osn}(\mathbb{P}) \subseteq \Theta_I^N(\mathbb{P})$ for all $\mathbb{P} \in \mathbf{P}$, for all n sufficiently large. For nonsingular $\mathbb{P} \in \mathbf{P}$ the reparameterization is of the form $\theta \mapsto \gamma(\theta; \mathbb{P})$ from $\Theta_I^N(\mathbb{P})$ into $\Gamma(\mathbb{P})$ where $\gamma(\theta) = 0$ if and only if $\theta \in \Theta_I(\mathbb{P})$. For singular $\mathbb{P} \in \mathbf{P}$ the reparameterization is of the form $\theta \mapsto (\gamma(\theta; \mathbb{P}), \gamma_\perp(\theta; \mathbb{P}))$ from $\Theta_I^N(\mathbb{P})$ into $\Gamma(\mathbb{P}) \times \Gamma_\perp(\mathbb{P})$ where $(\gamma(\theta; \mathbb{P}), \gamma_\perp(\theta; \mathbb{P})) = 0$ if and only if $\theta \in \Theta_I(\mathbb{P})$. We require the dimension of $\gamma(\cdot; \mathbb{P})$ to be between 1 and $\overline{d}$ for each $\mathbb{P} \in \mathbf{P}$, with $\overline{d} < \infty$ independent of $\mathbb{P}$.

To simply notation, in what follows we omit dependence of d^*, Θ_I^N, T, γ, $\gamma_\perp$, Γ, $\Gamma_\perp$, ℓ_n, $\mathbb{V}_n$, Σ, and $f_{n,\perp}$ on $\mathbb{P}$. We present results for the case in which each $T = \mathbb{R}^{d^*}$; extension to the case where some T are cones are straightforward.

Assumption B.2. *(Local quadratic approximation)*
(i) There exist sequences of random variables ℓ_n, $\mathbb{R}^{d^*}$*-valued random vectors* $\mathbb{V}_n$ *and, for singular* $\mathbb{P} \in \mathbf{P}$, *a sequence of non-negative measurable functions* $f_{n,\perp} : \Gamma_\perp \to \mathbb{R}$ *with* $f_{n,\perp}(0) = 0$ *(we take* $\gamma_\perp \equiv 0$ *and* $f_{n,\perp} \equiv 0$ *for nonsingular* $\mathbb{P} \in \mathbf{P}$*), such that* $\sup_{\mathbb{P} \in \mathbf{P}} \sup_{\theta \in \Theta_{osn}} \|(\gamma(\theta), \gamma_\perp(\theta))\| \to 0$ *and*

$$\sup_{\theta \in \Theta_{osn}} \left| nL_n(\theta) - \left(\ell_n - \frac{1}{2} \|\sqrt{n}\gamma(\theta)\|^2 + (\sqrt{n}\gamma(\theta))'\mathbb{V}_n - f_{n,\perp}(\gamma_\perp(\theta)) \right) \right| = o_{\mathbb{P}}(1) \qquad (25)$$

uniformly for $\mathbb{P} \in \mathbf{P}$, *with* $\mathbb{V}_n \overset{\mathbb{P}}{\rightsquigarrow} N(0, \Sigma)$ *as* $n \to \infty$ *for each* $\mathbb{P} \in \mathbf{P}$*;*
(ii) for each singular $\mathbb{P} \in \mathbf{P}$*:* $\{(\gamma(\theta), \gamma_\perp(\theta)) : \theta \in \Theta_{osn}\} = \{\gamma(\theta) : \theta \in \Theta_{osn}\} \times \{\gamma_\perp(\theta) : \theta \in \Theta_{osn}\}$*;*
(iii) $K_{osn} := \{\sqrt{n}\gamma(\theta) : \theta \in \Theta_{osn}\} \supseteq B_{k_n}$ *for each* $\mathbb{P} \in \mathbf{P}$ *and* $\inf_{\mathbb{P} \in \mathbf{P}} k_n \to \infty$ *as* $n \to \infty$*;*
(iv) $\sup_{\mathbb{P} \in \mathbf{P}} \sup_z |\mathbb{P}(\|\Sigma^{-1/2}\mathbb{V}_n\|^2 \le z) - F_{\chi^2_{d^*}}(z)| = o(1)$.

Notice that k_n in Part (iii) may depend on $\mathbb{P}$. Part (iv) can be verified via Berry-Esseen type results provided higher moments of $\Sigma^{-1/2}\mathbb{V}_n$ are bounded uniformly in $\mathbb{P}$ (see, e.g., Götze (1991)).

Let Π_{Γ^*} denote the image measure of Π on Γ under the map $\Theta_I^N \ni \theta \mapsto \gamma(\theta)$ if $\mathbb{P}$ is nonsingular and $\Theta_I^N \ni \theta \mapsto (\gamma(\theta), \gamma_\perp(\theta))$ if $\mathbb{P}$ is singular. Also let B_δ^* denote a ball of radius δ centered at the origin in $\mathbb{R}^{d^*}$ if $\mathbb{P}$ is nonsingular and in $\mathbb{R}^{d^* + \dim(\gamma_\perp)}$ if $\mathbb{P}$ is singular. In what follows we omit dependence of Π_{Γ^*}, B_δ^*, and π_{Γ^*} on $\mathbb{P}$.

Assumption B.3. *(Prior)*
(i) $\int_\theta e^{nL_n(\theta)} \, \mathrm{d}\Pi(\theta) < \infty$ $\mathbb{P}$*-almost surely for each* $\mathbb{P} \in \mathbf{P}$*;*
(ii) Each Π_{Γ^*} *has a density* π_{Γ^*} *on* $B_\delta^* \cap (\Gamma \times \Gamma_\perp)$ *(or* $B_\delta^* \cap \Gamma$ *if* $\mathbb{P}$ *is nonsingular) for some* $\delta > 0$ *which are uniformly (in* $\mathbb{P}$*) positive and continuous at the origin.*

As before, we let $\xi_{n,\alpha}^{post}$ denote the α quantile of $Q_n(\theta)$ under the posterior distribution Π_n.

Assumption B.4. *(MC convergence)*
$\xi_{n,\alpha}^{mc} = \xi_{n,\alpha}^{post} + o_{\mathbb{P}}(1)$ *uniformly for* $\mathbb{P} \in \mathbf{P}$.

The following result is uniform (in $\mathbb{P} \in \mathbf{P}$) extension of Theorems 3.1 and 3.2.

Theorem B.1. *Let Assumptions B.1, B.2, B.3, and B.4 hold with $\Sigma(\mathbb{P}) = I_{d^*}$ for each $\mathbb{P} \in \mathbf{P}$.*
(i) If there is at least a singular $\mathbb{P} \in \mathbf{P}$, then: (23) holds.
(ii) If no $\mathbb{P} \in \mathbf{P}$ is singular, then: (23) holds with equality.

To establish (24) we require a uniform version of Assumptions 3.5 and 3.6. Let $\mathbb{P}_Z$ denote the distribution of a $N(0, I_{d^*})$ random vector. In what follows, we omit dependence of f on $\mathbb{P}$ to simplify notation. Let $\xi_{\alpha,\mathbb{P}}$ denote the α quantile of $f(Z)$.

Assumption B.5. *(Profile QL)*
(i) For each $\mathbb{P} \in \mathbf{P}$ there exists a measurable function $f : \mathbb{R}^{d^*} \to \mathbb{R}$ such that:

$$\sup_{\theta \in \Theta_{osn}} \left| nPL_n(\Delta(\theta)) - \left(\ell_n + \frac{1}{2}\|\mathbb{V}_n\|^2 - \frac{1}{2} f\left(\mathbb{V}_n - \sqrt{n}\gamma(\theta)\right)\right)\right| = o_{\mathbb{P}}(1)$$

uniformly for $\mathbb{P} \in \mathbf{P}$, with $\mathbb{V}_n$, ℓ_n, and γ from Assumption B.2;
(ii) There exist $\underline{z}, \overline{z} \in \mathbb{R}$ with $\underline{z} < \inf_{\mathbb{P}\in\mathbf{P}} \xi_{\alpha,\mathbb{P}} \leq \sup_{\mathbb{P}\in\mathbf{P}} \xi_{\alpha,\mathbb{P}} < \overline{z}$ such that the functions $[\underline{z}, \overline{z}] \ni z \mapsto \mathbb{P}_Z(f(Z) \leq z)$ are uniformly equicontinuous and invertible with uniformly equicontinuous inverse;
(iii) $\sup_{\mathbb{P}\in\mathbf{P}} \sup_{z\in[\underline{z},\overline{z}]} |\mathbb{P}(f(\Sigma^{-1/2}\mathbb{V}_n) \leq z) - \mathbb{P}_Z(f(Z) \leq z)| = o(1)$.

Let $\xi^{post,p}_{n,\alpha}$ denote the α quantile of $PQ_n(\Delta(\theta))$ under the posterior distribution Π_n.

Assumption B.6. *(MC convergence)*
$\xi^{mc,p}_{n,\alpha} = \xi^{post,p}_{n,\alpha} + o_{\mathbb{P}}(1)$ uniformly for $\mathbb{P} \in \mathbf{P}$.

The following result is uniform (in $\mathbb{P} \in \mathbf{P}$) extension of Theorem 3.3.

Theorem B.2. *Let Assumptions B.1, B.2, B.3, B.5, and B.6 hold with $\Sigma(\mathbb{P}) = I_{d^*}$ for each $\mathbb{P} \in \mathbf{P}$. Then: (24) holds with equality.*

C Example 3: parameters drifting to boundary and point-identification

We return to Example 3 considered in Section 4.2.1 and examine the coverage properties of $\widehat{M}_\alpha$ for the identified set $B_I = [0, \mu]$ along certain drifting sequences of distributions. As will be seen, our MC CSs (based on the posterior distribution of the profile QLR) remain valid in certain situations while bootstrap-based CSs (based on the bootstrap distribution of the profile QLR) can undercover.

Recall that $X_1, \ldots, X_n$ are i.i.d. with unknown mean $\mu \in [0, 1]$ and unit variance. Here we consider coverage of the CS for $B_I = [0, \mu]$ as the mean $\mu \in [0, 1]$ drifts to the lower bound $\mu = 0$ of the parameter space. Suppose that $\beta \in B$ is identified by the moment inequality $\mathbb{E}[\beta - X_i] \leq 0$. The identified set for β is $B_I = [0, \mu]$, which is the argmax of the population criterion $L(\beta) = -\frac{1}{2}((\beta - \mu) \vee 0)^2$.
We write this as a moment equality model $\mathbb{E}[\beta + \lambda - X_i] = 0$ where $\lambda \in [0, 1 - \beta]$ is a slackness parameter. The parameter space for $\theta = (\beta, \lambda)$ is $\Theta = \{(\beta, \lambda) \in [0, 1]^2 : \beta + \lambda \leq 1\}$. The

identified set for θ is $\Theta_I = \{(\beta, \lambda) \in \Theta : \beta + \lambda = \mu\}$ and the identified set for the subvector β is $B_I = [0, \mu]$. The CU-GMM objective function is:

$$L_n(\beta, \lambda) = -\frac{1}{2}(\beta + \lambda - \bar{X}_n)^2 .$$

Drifting to point identification. We take $\mu = c/\sqrt{n}$ with $c > 0$. Then:

$$nL_n(\beta, \lambda) = -\frac{1}{2}(\sqrt{n}(\beta + \lambda) - c - \mathbb{V}_n)^2$$

where $\mathbb{V}_n = \sqrt{n}(\bar{X}_n - \mu) \rightsquigarrow N(0, 1)$. It is straightforward to show that:

$$2nL_n(\hat{\beta}, \hat{\lambda}) = -(c + \mathbb{V}_n)^2 1\!\mathrm{l}\{c + \mathbb{V}_n \leq 0\} + o_{\mathbb{P}}(1) .$$

Similarly:

$$\sup_{\lambda \in [0,1-\beta]} 2nL_n(\beta, \lambda) = \left[\begin{array}{ll} -(c + \mathbb{V}_n - \sqrt{n}\beta)^2 & c + \mathbb{V}_n - \sqrt{n}\beta \leq 0 \\ 0 & 0 \leq c + \mathbb{V}_n - \sqrt{n}\beta \leq \sqrt{n} - \sqrt{n}\beta \\ -(c + \mathbb{V}_n - \sqrt{n})^2 & c + \mathbb{V}_n \geq \sqrt{n} \end{array} \right.$$

hence:

$$\inf_{\beta \in \mu(\Delta(\theta^b))} \sup_{\lambda \in [0,1]} 2nL_n(\beta, \lambda) = -(c + \mathbb{V}_n - \sqrt{n}(\beta^b + \lambda^b))^2 1\!\mathrm{l}\{c + \mathbb{V}_n - \sqrt{n}(\beta^b + \lambda^b) \leq 0\} + o_{\mathbb{P}}(1)$$

and:

$$PQ_n(\Delta(\theta^b)) = (c + \mathbb{V}_n - \sqrt{n}(\beta^b + \lambda^b))^2 1\!\mathrm{l}\{c + \mathbb{V}_n - \sqrt{n}(\beta^b + \lambda^b) \leq 0\} - (c + \mathbb{V}_n)^2 1\!\mathrm{l}\{c + \mathbb{V}_n \leq 0\} + o_{\mathbb{P}}(1) .$$

In particular, we have:

$$\sup_{\theta \in \Theta_I} PQ_n(\Delta(\theta)) = (\mathbb{V}_n)^2 1\!\mathrm{l}\{\mathbb{V}_n \leq 0\} - (c + \mathbb{V}_n)^2 1\!\mathrm{l}\{c + \mathbb{V}_n \leq 0\} + o_{\mathbb{P}}(1) .$$

Suppose we choose a prior on (β, λ) that induces a flat prior on $\gamma = \beta + \lambda$. Also let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(\kappa) = \kappa^2 1\!\mathrm{l}\{\kappa \leq 0\}$ and let $z_n^* = z + (c + \mathbb{V}_n)^2 1\!\mathrm{l}\{c + \mathbb{V}_n \leq 0\}$. Ignoring asymptotically negligible terms, we have:

$$\begin{aligned}
\Pi_n(\{\theta : PQ_n(\Delta(\theta)) \le z\}|\mathbf{X}_n) &= \Pi_n\big(\{\theta : f(c + \mathbb{V}_n - \sqrt{n}\gamma(\theta)) \le z_n^*\}\big|\mathbf{X}_n\big) \\
&= \frac{\int_0^1 1\!\!1\{f(c + \mathbb{V}_n - \sqrt{n}\gamma) \le z_n^*\} e^{-\frac{1}{2}(c+\mathbb{V}_n - \sqrt{n}\gamma)^2} d\gamma}{\int_0^1 e^{-\frac{1}{2}(c+\mathbb{V}_n-\sqrt{n}\gamma)^2} d\gamma} \\
&= \frac{\int_{c+\mathbb{V}_n-\sqrt{n}}^{c+\mathbb{V}_n} 1\!\!1\{f(\kappa) \le z_n^*\} e^{-\frac{1}{2}\kappa^2} d\kappa}{\int_{c+\mathbb{V}_n-\sqrt{n}}^{c+\mathbb{V}_n} e^{-\frac{1}{2}\kappa^2} d\kappa} \\
&= \frac{\int_{-\infty}^{c+\mathbb{V}_n} 1\!\!1\{f(\kappa) \le z_n^*\} e^{-\frac{1}{2}\kappa^2} d\kappa}{\int_{-\infty}^{c+\mathbb{V}_n} e^{-\frac{1}{2}\kappa^2} d\kappa} + o_{\mathbb{P}}(1)\,.
\end{aligned}$$

Since $f(\kappa) \le z_n^*$ holds if and only if $\kappa \ge -\sqrt{z_n^*}$, we have:

$$\Pi_n(\{\theta : PQ_n(\Delta(\theta)) \le z\}|\mathbf{X}_n) = \frac{\mathbb{P}_{Z|\mathbf{X}_n}(-\sqrt{z_n^*} \le Z \le c + \mathbb{V}_n)}{\mathbb{P}_{Z|\mathbf{X}_n}(Z \le c + \mathbb{V}_n)} + o_{\mathbb{P}}(1)\,.$$

We choose $z_{n,\alpha} = z_n^* - (c + \mathbb{V}_n)^2 1\!\!1\{c + \mathbb{V}_n < 0\} \ge 0$ so that the right-hand side is equal to α (notice that in some cases we will choose $z_{n,\alpha} = 0$ with the right-hand side $\ge \alpha$). Ignoring asymptotically negligible terms, this gives:

$$-\sqrt{z_n^*} = 0 \wedge \Phi^{-1}\big((1-\alpha)\Phi(c + \mathbb{V}_n)\big)$$

hence:

$$z_{n,\alpha} = \Big(0 \wedge \Phi^{-1}\big((1-\alpha)\Phi(c+\mathbb{V}_n)\big)\Big)^2 - (c+\mathbb{V}_n)^2 1\!\!1\{c + \mathbb{V}_n < 0\} + o_{\mathbb{P}}(1)\,.$$

Therefore, the asymptotic coverage of the CS $\widehat{M}_\alpha$ for B_I is:

$$\lim_{n\to\infty} \mathbb{P}(B_I \subseteq \widehat{M}_\alpha) = \mathbb{P}_Z\bigg((Z)^2 1\!\!1\{Z \le 0\} \le \Big(0 \wedge \Phi^{-1}\big((1-\alpha)\Phi(c+Z)\big)\Big)^2\bigg)$$

where $Z \sim N(0,1)$. One can verify numerically that $\lim_{n\to\infty} \mathbb{P}(B_I \subseteq \widehat{M}_\alpha) \ge \alpha$ for all $\alpha \in (0,1)$ and $c \ge 0$ (see Figure 14 below).

Comparison with the nonparametric bootstrap. Let $\mathbf{X}_n^*$ denote the bootstrap sample of size n. The bootstrap criterion function is

$$\begin{aligned}
2nL_n^*(\beta,\lambda) &= -(\sqrt{n}(\beta+\lambda) - (\sqrt{n}\bar{X}_n) - \sqrt{n}(\bar{X}_n^* - \bar{X}_n))^2 \\
&= -(\sqrt{n}(\beta+\lambda) - c_n - \mathbb{V}_n^*)^2
\end{aligned}$$

where $c_n = \sqrt{n}\bar{X}_n = c + \mathbb{V}_n$ and $\mathbb{V}_n^* = \sqrt{n}(\bar{X}_n^* - \bar{X}_n) \rightsquigarrow N(0,1)$. By similar arguments, we have:

$$2nL_n^*(\hat{\beta}^*, \hat{\lambda}^*) = -(c_n + \mathbb{V}_n^*)^2 1\!\!1\{c_n + \mathbb{V}_n^* \leq 0\} + o_{\mathbb{P}}(1)$$

$$\inf_{\beta \in [0,(\bar{X}_n \vee 0)]} \sup_{\lambda \in [0,1]} 2nL_n^*(\beta, \lambda) = -(\mathbb{V}_n^*)^2 1\!\!1\{\mathbb{V}_n^* \leq 0\} \wedge -(c_n + \mathbb{V}_n^*)^2 1\!\!1\{c_n + \mathbb{V}_n^* \leq 0\} + o_{\mathbb{P}}(1)$$

so the bootstrap profile QLR statistic for B_I is:

$$\Big((\mathbb{V}_n^*)^2 1\!\!1\{\mathbb{V}_n^* \leq 0\} - (c_n + \mathbb{V}_n^*)^2 1\!\!1\{c_n + \mathbb{V}_n^* \leq 0\}\Big) \vee 0 + o_{\mathbb{P}}(1)\,.$$

Figure 14 presents the asymptotic coverage of our MCMC CS $\widehat{M}_\alpha$ for B_I and a CS based on bootstrapping the QLR statistic for B_I for the case in which $\mu = c/\sqrt{n}$ with $c = 2.0$. It is clear that our MCMC CS remains valid whereas the bootstrap CS undercovers. Similar results are obtained with other values of $c \geq 0$. This example clearly shows that the posterior distribution of the profile QLR statistic and the bootstrap distribution of the profile QLR statistic can indeed behave differently. Thus, our MCMC CSs do not necessarily run into coverage problems in certain situations in which bootstrap-based CSs undercover.

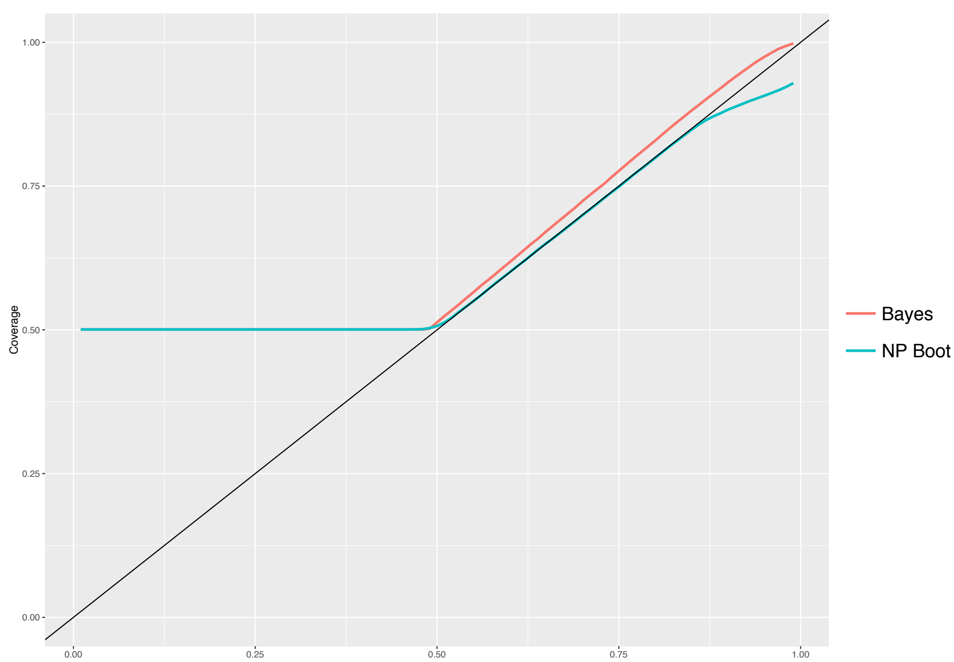

Figure 14: Comparison of the asymptotic coverage probabilities of our MCMC CS for B_I (Bayes) and a CS based on bootstrapping the profile QLR statistic for B_I (NP Boot).

D Local power

In this appendix we study the behavior of the CSs $\widehat{\Theta}_\alpha$ and $\widehat{M}_\alpha$ under $n^{-1/2}$-local (contiguous) alternatives. We maintain the same setup as in Section 3.

Assumption D.1. *There exist sequences of distributions $(\mathrm{P}_{n,a})_{n\in\mathbb{N}}$ for fixed $a \in \mathbb{R}^{d^*}$ that satisfy:*
(i) $L_n(\hat\theta) = \sup_{\theta\in\Theta_{osn}} L_n(\theta) + o_{\mathrm{P}_{n,a}}(n^{-1})$*;*
(ii) $\Pi_n(\Theta_{osn}^c|\mathbf{X}_n) = o_{\mathrm{P}_{n,a}}(1)$*;*
(iii) There exist sequences of random variables ℓ_n and $\mathbb{R}^{d^*}$-valued random vectors $\mathbb{V}_n$ (both of which are measurable functions of data $\mathbf{X}_n$) such that:

$$\sup_{\theta\in\Theta_{osn}} \left| nL_n(\theta) - \left(\ell_n - \frac{1}{2}\|\sqrt{n}\gamma(\theta)\|^2 + (\sqrt{n}\gamma(\theta))'\mathbb{V}_n\right)\right| = o_{\mathrm{P}_{n,a}}(1) \tag{26}$$

with $\sup_{\theta\in\Theta_{osn}} \|\gamma(\theta)\| \to 0$ *and* $\mathbb{V}_n \overset{\mathrm{P}_{n,a}}{\rightsquigarrow} N(a, I_{d^*})$ *as* $n \to \infty$*;*
(iv) The sets $K_{osn} = \{\sqrt{n}\gamma(\theta) : \theta \in \Theta_{osn}\}$ *cover* $\mathbb{R}^{d^*}$*;*
(v) $\int_\Theta e^{nL_n(\theta)}\,\mathrm{d}\Pi(\theta) < \infty$ $\mathrm{P}_{n,a}$*-almost surely;*
(vi) Π_Γ *has a continuous, strictly positive density* π_Γ *on* $B_\delta \cap \Gamma$ *for some* $\delta > 0$*;*
(vii) $\xi^{mc}_{n,\alpha} = \xi^{post}_{n,\alpha} + o_{\mathrm{P}_{n,a}}(1)$.

Assumption D.1 is essentially a restatement of Assumptions 3.1 to 3.4 with a modified quadratic expansion. Notice that with $a = 0$ we obtain $\mathrm{P}_{n,a} = \mathbb{P}$ and Assumption D.1 corresponds to Assumptions 3.1 to 3.4 with optimal weighting $\Sigma = I_{d^*}$.

In the following result, let $\chi^2_{d^*}(a'a)$ denote the noncentral chi-square distribution with d^* degrees of freedom and noncentrality parameter $a'a$ and let $F_{\chi^2_{d^*}(a'a)}$ denote its cdf. Also let $\chi^2_{d^*,\alpha}$ denote the α quantile of the (standard) $\chi^2_{d^*}$ distribution $F_{\chi^2_{d^*}}$.

Theorem D.1. *Let Assumption D.1(i)(iii)(iv) hold. Then:*

$$\sup_{\theta\in\Theta_I} Q_n(\theta) \overset{\mathrm{P}_{n,a}}{\rightsquigarrow} \chi^2_{d^*}(a'a);$$

if further Assumption D.1(ii)(v)(vi) hold, then:

$$\sup_z \left|\Pi_n\big(\{\theta : Q_n(\theta) \le z\}\,\big|\,\mathbf{X}_n\big) - F_{\chi^2_{d^*}}(z)\right| = o_{\mathrm{P}_{n,a}}(1);$$

and if further Assumption D.1(vii) holds, then:

$$\lim_{n\to\infty} \mathrm{P}_{n,a}(\Theta_I \subseteq \widehat{\Theta}_\alpha) = F_{\chi^2_{d^*}(a'a)}(\chi^2_{d^*,\alpha}) < \alpha$$

whenever $a \neq 0$.

We now present a similar result for $\widehat{M}_\alpha$. In order to do so, we extend slightly the conditions in Assumption D.1.

Assumption D.1. *Let the following also hold under the local alternatives:*
(viii) There exists a measurable $f : \mathbb{R}^{d^*} \to \mathbb{R}_+$ *such that:*

$$\sup_{\theta \in \Theta_{osn}} \left| nPL_n(\Delta(\theta)) - \left(\ell_n + \frac{1}{2}\|\mathbb{V}_n\|^2 - \frac{1}{2} f\left(\mathbb{V}_n - \sqrt{n}\gamma(\theta)\right) \right) \right| = o_{\mathrm{P}_{n,a}}(1)\,.$$

(vii′) $\xi^{mc,p}_{n,\alpha} = \xi^{post,p}_{n,\alpha} + o_{\mathrm{P}_{n,a}}(1)$.

Assumption D.1(viii) and (vii′) are essentially Assumptions 3.5 and 3.6.

Let $Z \sim N(0, I_{d^*})$ and $\mathbb{P}_Z$ denote the distribution of Z. Let the distribution of $f(Z)$ be continuous at its α-quantile, which we denote by z_α.

Theorem D.2. *Let Assumption D.1(i)(iii)(iv)(viii) hold. Then:*

$$\sup_{\theta \in \Theta_I} PQ_n(\Delta(\theta)) \overset{\mathrm{P}_{n,a}}{\rightsquigarrow} f(Z + a)\ ;$$

if further Assumption D.1(ii)(v)(vi) hold, then:

$$\sup_{z \in S_n^{-\epsilon}} \left| \Pi_n\big(\{\theta : PQ_n(\Delta(\theta)) \leq z\} \,\big|\, \mathbf{X}_n\big) - \mathbb{P}_{Z|\mathbf{X}_n}\big(f(Z) \leq z\big) \right| = o_{\mathrm{P}_{n,a}}(1)$$

and if further Assumption D.1(vii′) holds, then:

$$\lim_{n \to \infty} \mathrm{P}_{n,a}(M_I \subseteq \widehat{M}_\alpha) = \mathbb{P}_Z(f(Z + a) \leq z_\alpha)\ .$$

We can thus deduce from Anderson's lemma (van der Vaart, 2000, Lemma 8.5) that the coverage $\lim_{n\to\infty} \mathrm{P}_{n,a}(M_I \subseteq \widehat{M}_\alpha) \leq \alpha$ whenever f is subconvex. In particular, this includes the case in which M_I is a singleton.

E Parameter-dependent support

In this appendix we briefly describe how our procedure may be applied to models with parameter dependent support under loss of identifiability. Parameter-dependent support is a feature of certain auction models (e.g., Hirano and Porter (2003), Chernozhukov and Hong (2004)) and some structural models in labor economics (e.g., Flinn and Heckman (1982)). For simplicity we just deal with inference on the full vector, though the following results could be extended to subvector inference in this context.

We again presume the existence of a local reduced-form parameter γ such that $\gamma(\theta) = 0$ if and only if $\theta \in \Theta_I$. In what follows we assume without loss of generality that $L_n(\hat\theta) = \sup_{\theta \in \Theta_{osn}} L_n(\theta)$ since $\hat\theta$ is not required in order to compute the confidence set. We replace Assumption 3.2 (local quadratic approximation) with the following assumption, which permits the support of the data to depend on certain components of the local reduced-form parameter γ.

Assumption E.2. *(i) There exist functions* $\gamma : \Theta_I^N \to \Gamma \subseteq \mathbb{R}^{d^*}$ *and* $h : \Gamma \to \mathbb{R}_+$*, a sequence of* $\mathbb{R}^{d^*}$*-valued random vectors* $\hat{\gamma}_n$*, and a positive sequence* $(a_n)_{n\in\mathbb{N}}$ *with* $a_n \to 0$ *such that:*

$$\sup_{\theta\in\Theta_{osn}} \left| \frac{\frac{a_n}{2} Q_n(\theta) - h(\gamma(\theta) - \hat{\gamma}_n)}{h(\gamma(\theta) - \hat{\gamma}_n)} \right| = o_{\mathbb{P}}(1)$$

with $\sup_{\theta\in\Theta_{osn}} \|\gamma(\theta)\| \to 0$ *and* $\inf\{h(\gamma) : \|\gamma\| = 1\} > 0$*;*
(ii) there exist $r_1, \ldots, r_{d^*} > 0$ *such that* $th(\gamma) = h(t^{r_1}\gamma_1, t^{r_2}\gamma_2, \ldots, t^{r_{d^*}}\gamma_{d^*})$ *for each* $t > 0$*;*
(iii) the sets $K_{osn} = \{(b_n^{-r_1}(\gamma_1(\theta) - \hat{\gamma}_{n,1}), \ldots, b_n^{-r_{d^*}}(\gamma_{d^*}(\theta) - \hat{\gamma}_{n,d^*}))' : \theta \in \Theta_{osn}\}$ *cover* $\mathbb{R}_+^{d^*}$ *for any positive sequence* $(b_n)_{n\in\mathbb{N}}$ *with* $b_n \to 0$ *and* $a_n/b_n \to 1$*.*

This assumption is similar to Assumptions 2-3 in Fan et al. (2000) but has been modified to allow for non-identifiable parameters θ. Let F_Γ denote a Gamma distribution with shape parameter $r^* = \sum_{i=1}^{d^*} r_i$ and scale parameter 2. The following lemma shows that the posterior distribution of the QLR converges to F_Γ.

Lemma E.1. *Let Assumptions 3.1, E.2, and 3.3 hold. Then:*

$$\sup_z |\Pi_n(\{\theta : Q_n(\theta) \le z\}|\mathbf{X}_n) - F_\Gamma(z)| = o_p(1)\,.$$

By modifying appropriately the arguments in Fan et al. (2000) one can show that, under Assumption E.2, $\sup_{\theta\in\Theta_I} Q_n(\theta) \rightsquigarrow F_\Gamma$. The following theorem states that one still obtains asymptotically correct frequentist coverage of $\widehat{\Theta}_\alpha$ for the IdS Θ_I.

Theorem E.1. *Let Assumptions 3.1, E.2, 3.3, and 3.4 hold and* $\sup_{\theta\in\Theta_I} Q_n(\theta) \rightsquigarrow F_\Gamma$*. Then:*

$$\lim_{n\to\infty} \mathbb{P}(\Theta_I \subseteq \widehat{\Theta}_\alpha) = \alpha\,.$$

We finish this section with a simple example. Consider a model in which $X_1, \ldots, X_n$ are i.i.d. $U[0, (\theta_1 \vee \theta_2)]$ where $(\theta_1, \theta_2) \in \Theta = \mathbb{R}_+^2$. Let the true distribution of the data be $U[0, \tilde{\gamma}]$. The identified set is $\Theta_I = \{\theta \in \Theta : \theta_1 \vee \theta_2 = \tilde{\gamma}\}$.

Then we use the reduced-form parameter $\gamma(\theta) = (\theta_1 \vee \theta_2) - \tilde{\gamma}$. Let $\hat{\gamma}_n = \max_{1\le i\le n} X_i - \tilde{\gamma}$. Here we take $\Theta_{osn} = \{\theta : (1+\varepsilon_n)\hat{\gamma}_n \ge \gamma(\theta) \ge \hat{\gamma}_n\}$ where $\varepsilon_n \to 0$ slower than n^{-1} (e.g. $\varepsilon_n = (\log n)/n$). It is straightforward to show that:

$$\sup_{\theta\in\Theta_I} Q_n(\theta) = 2n \log\left(\frac{\tilde{\gamma}}{\hat{\gamma}_n + \tilde{\gamma}}\right) \rightsquigarrow F_\Gamma$$

where F_Γ denotes the Gamma distribution with shape parameter $r^* = 1$ and scale parameter 2. Furthermore, taking $a_n = n^{-1}$ and $h(\gamma(\theta) - \hat{\gamma}_n) = \tilde{\gamma}^{-1}(\gamma(\theta) - \hat{\gamma}_n)$ we may deduce that:

$$\sup_{\theta\in\Theta_{osn}} \left| \frac{\frac{1}{2n} Q_n(\theta) - h(\gamma(\theta) - \hat{\gamma}_n)}{h(\gamma(\theta) - \hat{\gamma}_n)} \right| = o_{\mathbb{P}}(1)\,.$$

Notice that $r^* = 1$ and that the sets $K_{osn} = \{n(\gamma(\theta) - \hat{\gamma}_n) : \theta \in \Theta_{osn}\} = \{n(\gamma - \hat{\gamma}_n) : (1+\varepsilon_n)\hat{\gamma} \ge$

$\gamma \geq \hat{\gamma}_n\}$ cover $\mathbb{R}^+$. A smooth prior on Θ will induce a smooth prior on $\gamma(\theta)$, and the result follows from Theorem E.1.

F Proofs and Additional Results

F.1 Proofs and Additional Lemmas for Sections 2 and 3

Proof of Lemma 2.1. By (ii), there is a positive sequence $(\eta_n)_{n\in\mathbb{N}}$ with $\eta_n = o(1)$ such that $w_{n,\alpha} \geq w_\alpha - \eta_n$ holds wpa1. Therefore:

$$\begin{aligned}\mathbb{P}(\Theta_I \subseteq \widehat{\Theta}_\alpha) &= \mathbb{P}(\sup_{\theta\in\Theta_I} Q_n(\theta) \leq w_{n,\alpha}) \\ &\geq \mathbb{P}(\sup_{\theta\in\Theta_I} Q_n(\theta) \leq w_\alpha - \eta_n) + o(1)\end{aligned}$$

and the result follows by part (i). If $w_{n,\alpha} = w_\alpha + o_{\mathbb{P}}(1)$ then we may replace the preceding inequality by an equality. ■

Proof of Lemma 2.2. Follows by similar arguments to the proof of Lemma 2.1. ■

In this appendix we often use the following expression (27) that is equivalent to equation (14) of Assumption 3.2(i):

$$\sup_{\theta\in\Theta_{osn}} \left| nL_n(\theta) - \ell_n - \frac{1}{2}\|\mathbb{V}_n\|^2 + \frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 \right| = o_{\mathbb{P}}(1)\,. \tag{27}$$

Lemma F.1. *Let Assumptions 3.1(i) and 3.2 hold. Then:*

$$\sup_{\theta\in\Theta_{osn}} \left| Q_n(\theta) - \left(-\inf_{t\in T} \|t - \mathbb{V}_n\|^2 + \|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 \right) \right| = o_{\mathbb{P}}(1)\,. \tag{28}$$

And hence

$$\sup_{\theta\in\Theta_I} Q_n(\theta) = \|\mathbf{T}\mathbb{V}_n\|^2 + o_{\mathbb{P}}(1)\ .$$

Proof of Lemma F.1. Applying successively Assumption 3.2(i) or expression (27), then using Assumptions 3.1(i) and 3.2(ii), we obtain:

$$\begin{aligned}nL_n(\hat{\theta}) &= \sup_{\theta\in\Theta_{osn}} nL_n(\theta) + o_{\mathbb{P}}(1) \\ &= \ell_n + \frac{1}{2}\|\mathbb{V}_n\|^2 - \inf_{\theta\in\Theta_{osn}} \frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 + o_{\mathbb{P}}(1) \\ &= \ell_n + \frac{1}{2}\|\mathbb{V}_n\|^2 - \inf_{t\in T} \frac{1}{2}\|t - \mathbb{V}_n\|^2 + o_{\mathbb{P}}(1)\,. \end{aligned} \tag{29}$$

Then using Assumption 3.2(i) or expression (27), we have:

$$\begin{aligned} Q_n(\theta) &= 2\left(\ell_n + \frac{1}{2}\|\mathbb{V}_n\|^2 - \inf_{t\in T}\frac{1}{2}\|t-\mathbb{V}_n\|^2 + o_{\mathbb{P}}(1)\right) - 2\left(\ell_n + \frac{1}{2}\|\mathbb{V}_n\|^2 - \frac{1}{2}\|\sqrt{n}\gamma(\theta)-\mathbb{V}_n\|^2 + o_{\mathbb{P}}(1)\right) \\ &= \|\sqrt{n}\gamma(\theta)-\mathbb{V}_n\|^2 - \inf_{t\in T}\|t-\mathbb{V}_n\|^2 + o_{\mathbb{P}}(1) \end{aligned}$$

where the $o_{\mathbb{P}}(1)$ term holds uniformly over Θ_{osn}. This proves expression (28).

Next, since $\gamma(\theta) = 0$ for $\theta \in \Theta_I$, we have:

$$\begin{aligned} \sup_{\theta\in\Theta_I} Q_n(\theta) &= \|\mathbb{V}_n\|^2 - \inf_{t\in T}\|t-\mathbb{V}_n\|^2 + o_{\mathbb{P}}(1) \\ &= \|\mathbb{V}_n\|^2 - \|\mathbb{V}_n - \mathbf{T}\mathbb{V}_n\|^2 + o_{\mathbb{P}}(1) \\ &= \|\mathbf{T}\mathbb{V}_n\|^2 + o_{\mathbb{P}}(1) \end{aligned}$$

where the second equality is by definition of the projection onto the closed convex cone T, and the third equality is by Moreau's decomposition Theorem (Hiriart-Urruty and Lemaréchal, 2001, Theorem 3.2.5, p.51). ∎

Proof of Theorem 3.1. We verify the conditions of Lemma 2.1. We may assume without loss of generality that $L_n(\hat\theta) = \sup_{\theta\in\Theta_{osn}} L_n(\theta) + o_{\mathbb{P}}(n^{-1})$ because $\widehat\Theta_\alpha$ does not depend on the precise $\hat\theta$ used (cf. Remark 1). By Lemma F.1 we have:

$$\sup_{\theta\in\Theta_I} Q_n(\theta) = \|\mathbf{T}\mathbb{V}_n\|^2 + o_{\mathbb{P}}(1) \rightsquigarrow \|\mathbf{T}Z\|^2$$

with $Z \sim N(0, I_{d^*})$ when $\Sigma = I_{d^*}$, where the final result is by the continuous mapping theorem. In the following let z_α denote the α quantile of the distribution of $\|\mathbf{T}Z\|^2$.

For part (i), Lemma 3.1(i) shows that the posterior distribution of the QLR asymptotically (first-order) stochastically dominates the distribution of $\|\mathbf{T}Z\|^2$ which implies that $\xi^{post}_{n,\alpha} \geq z_\alpha + o_{\mathbb{P}}(1)$. Therefore:

$$\xi^{mc}_{n,\alpha} = z_\alpha + (\xi^{post}_{n,\alpha} - z_\alpha) + (\xi^{mc}_{n,\alpha} - \xi^{post}_{n,\alpha}) \geq z_\alpha + (\xi^{mc}_{n,\alpha} - \xi^{post}_{n,\alpha}) + o_{\mathbb{P}}(1) = z_\alpha + o_{\mathbb{P}}(1)$$

where the final equality is by Assumption 3.4.

For part (ii), when $T = \mathbb{R}^{d^*}$ and $\Sigma = I_{d^*}$, we have:

$$\sup_{\theta\in\Theta_I} Q_n(\theta) = \|\mathbb{V}_n\|^2 + o_{\mathbb{P}}(1) \rightsquigarrow \chi^2_{d^*}\,, \quad \text{and hence} \quad z_\alpha = \chi^2_{d^*,\alpha}\,.$$

Further:

$$\xi^{mc}_{n,\alpha} = \chi^2_{d^*,\alpha} + (\xi^{post}_{n,\alpha} - \chi^2_{d^*,\alpha}) + (\xi^{mc}_{n,\alpha} - \xi^{post}_{n,\alpha}) = \chi^2_{d^*,\alpha} + o_{\mathbb{P}}(1)$$

by Lemma 3.1(ii) and Assumption 3.4. ∎

Proof of Lemma 3.1. We first prove equation (16). Since $|\Pr(A) - \Pr(A \cap B)| \le \Pr(B^c)$, we have:

$$\sup_z \left|\Pi_n(\{\theta : Q_n(\theta) \le z\}|\mathbf{X}_n) - \Pi_n(\{\theta : Q_n(\theta) \le z\} \cap \Theta_{osn}|\mathbf{X}_n)\right| \le \Pi_n(\Theta_{osn}^c|\mathbf{X}_n) = o_{\mathbb{P}}(1) \quad (30)$$

by Assumption 3.1(ii). Moreover by Assumptions 3.1(ii) and 3.3(i),

$$\left|\frac{\int_{\Theta_{osn}} e^{nL_n(\theta)} \mathrm{d}\Pi(\theta)}{\int_{\Theta} e^{nL_n(\theta)} \mathrm{d}\Pi(\theta)} - 1\right| = \Pi_n(\Theta_{osn}^c|\mathbf{X}_n) = o_{\mathbb{P}}(1)$$

and hence:

$$\sup_z \left|\Pi_n(\{\theta : Q_n(\theta) \le z\} \cap \Theta_{osn} \,|\, \mathbf{X}_n) - \frac{\int_{\{\theta: Q_n(\theta) \le z\}} e^{nL_n(\theta)} \mathrm{d}\Pi(\theta)}{\int_{\Theta_{osn}} e^{nL_n(\theta)} \mathrm{d}\Pi(\theta)}\right| = o_{\mathbb{P}}(1)\,. \quad (31)$$

In view of (30) and (31), it suffices to characterize the large-sample behavior of:

$$R_n(z) := \frac{\int_{\{\theta: Q_n(\theta) \le z\} \cap \Theta_{osn}} e^{nL_n(\theta) - \ell_n - \frac{1}{2}\|\mathbb{V}_n\|^2} \mathrm{d}\Pi(\theta)}{\int_{\Theta_{osn}} e^{nL_n(\theta) - \ell_n - \frac{1}{2}\|\mathbb{V}_n\|^2} \mathrm{d}\Pi(\theta)}\,. \quad (32)$$

Lemma F.1 and expression (27) imply that there exists a positive sequence $(\eta_n)_{n\in\mathbb{N}}$ independent of z with $\eta_n = o(1)$ such that the inequalities:

$$\sup_{\theta \in \Theta_{osn}} \left| Q_n(\theta) - \left(- \inf_{t \in T} \|t - \mathbb{V}_n\|^2 + \|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 \right) \right| \le \eta_n$$

$$\sup_{\theta \in \Theta_{osn}} \left| nL_n(\theta) - \ell_n - \frac{1}{2}\|\mathbb{V}_n\|^2 + \frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 \right| \le \frac{\eta_n}{2}$$

both hold wpa1. Therefore, wpa1 we have:

$$\begin{aligned} & e^{-\eta_n} \frac{\int_{\{\theta: \|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 \le z + \inf_{t\in T}\|t - \mathbb{V}_n\|^2 - \eta_n\} \cap \Theta_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2} \mathrm{d}\Pi(\theta)}{\int_{\Theta_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2} \mathrm{d}\Pi(\theta)} \\ & \le R_n(z) \le e^{\eta_n} \frac{\int_{\{\theta: \|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 \le z + \inf_{t\in T}\|t - \mathbb{V}_n\|^2 + \eta_n\} \cap \Theta_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2} \mathrm{d}\Pi(\theta)}{\int_{\Theta_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2} \mathrm{d}\Pi(\theta)} \end{aligned}$$

uniformly in z. Let $\Gamma_{osn} = \{\gamma(\theta) : \theta \in \Theta_{osn}\}$. A change of variables yields:

$$\begin{aligned} & e^{-\eta_n} \frac{\int_{\{\gamma: \|\sqrt{n}\gamma - \mathbb{V}_n\|^2 \le z + \inf_{t\in T}\|t - \mathbb{V}_n\|^2 - \eta_n\} \cap \Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma - \mathbb{V}_n\|^2} \mathrm{d}\Pi_\Gamma(\gamma)}{\int_{\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma - \mathbb{V}_n\|^2} \mathrm{d}\Pi_\Gamma(\gamma)} \\ & \le R_n(z) \le e^{\eta_n} \frac{\int_{\{\gamma: \|\sqrt{n}\gamma - \mathbb{V}_n\|^2 \le z + \inf_{t\in T}\|t - \mathbb{V}_n\|^2 + \eta_n\} \cap \Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma - \mathbb{V}_n\|^2} \mathrm{d}\Pi_\Gamma(\gamma)}{\int_{\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma - \mathbb{V}_n\|^2} \mathrm{d}\Pi_\Gamma(\gamma)} \end{aligned} \quad (33)$$

uniformly in z.

Recall B_δ from Assumption 3.3(ii). The inclusion $\Gamma_{osn} \subset B_\delta \cap \Gamma$ holds for all n sufficiently large by Assumption 3.2(ii). Taking n sufficiently large and using Assumption 3.3(ii), we may deduce that there exists a positive sequence $(\bar{\eta}_n)_{n\in\mathbb{N}}$ with $\bar{\eta}_n = o(1)$ such that:

$$\left| \frac{\sup_{\gamma\in\Gamma_{osn}} \pi_\Gamma(\gamma)}{\inf_{\gamma\in\Gamma_{osn}} \pi_\Gamma(\gamma)} - 1 \right| \leq \bar{\eta}_n$$

for each n. Substituting into (33) yields:

$$\begin{aligned}
&(1-\bar{\eta}_n)e^{-\eta_n} \frac{\int_{\{\gamma:\|\sqrt{n}\gamma-\mathbb{V}_n\|^2\leq z+\inf_{t\in T}\|t-\mathbb{V}_n\|^2-\eta_n\}\cap\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma-\mathbb{V}_n\|^2}\mathrm{d}\gamma}{\int_{\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma-\mathbb{V}_n\|^2}\mathrm{d}\gamma} \\
&\leq R_n(z) \leq (1+\bar{\eta}_n)e^{\eta_n} \frac{\int_{\{\gamma:\|\sqrt{n}\gamma-\mathbb{V}_n\|^2\leq z+\inf_{t\in T}\|t-\mathbb{V}_n\|^2+\eta_n\}\cap\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma-\mathbb{V}_n\|^2}\mathrm{d}\gamma}{\int_{\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma-\mathbb{V}_n\|^2}\mathrm{d}\gamma}
\end{aligned}$$

where integration "$\mathrm{d}\gamma$" should be interpreted as integration with respect to Lebesgue measure on $\mathbb{R}^{d^*}$.

Let $K_{osn} = \{\sqrt{n}\gamma : \gamma \in \Gamma_{osn}\}$ and $\Delta_n(z) = \{\kappa \in \mathbb{R}^{d^*} : \|\kappa\|^2 \leq z + \inf_{t\in T}\|t - \mathbb{V}_n\|^2\}$. Using the change of variables $\sqrt{n}\gamma - \mathbb{V}_n \mapsto \kappa$, we can rewrite the preceding inequalities as:

$$(1-\bar{\eta}_n)e^{-\eta_n} \frac{\int_{\Delta_n(z-\eta_n)\cap(K_{osn}-\mathbb{V}_n)} e^{-\frac{1}{2}\|\kappa\|^2}\mathrm{d}\kappa}{\int_{(K_{osn}-\mathbb{V}_n)} e^{-\frac{1}{2}\|\kappa\|^2}\mathrm{d}\kappa} \leq R_n(z) \leq (1+\bar{\eta}_n)e^{\eta_n} \frac{\int_{\Delta_n(z+\eta_n)\cap(K_{osn}-\mathbb{V}_n)} e^{-\frac{1}{2}\|\kappa\|^2}\mathrm{d}\kappa}{\int_{(K_{osn}-\mathbb{V}_n)} e^{-\frac{1}{2}\|\kappa\|^2}\mathrm{d}\kappa}.$$

Let $\nu_{d^*}(A) = (2\pi)^{-d^*/2}\int_A e^{-\frac{1}{2}\|\kappa\|^2}\,\mathrm{d}\kappa$ denote the Gaussian measure of a set $A \in \mathscr{B}(\mathbb{R}^{d^*})$. To complete the proof, it is enough to show that:

$$\sup_z \left| \frac{\nu_{d^*}(\Delta_n(z\pm\eta_n)\cap(K_{osn}-\mathbb{V}_n))}{\nu_{d^*}(K_{osn}-\mathbb{V}_n)} - \frac{\nu_{d^*}(\Delta_n(z\pm\eta_n)\cap(K_{osn}-\mathbb{V}_n))}{\nu_{d^*}(T-\mathbb{V}_n)} \right| = o_{\mathbb{P}}(1) \tag{34}$$

$$\sup_z \left| \frac{\nu_{d^*}(\Delta_n(z\pm\eta_n)\cap(K_{osn}-\mathbb{V}_n))}{\nu_{d^*}(T-\mathbb{V}_n)} - \frac{\nu_{d^*}(\Delta_n(z)\cap(T-\mathbb{V}_n))}{\nu_{d^*}(T-\mathbb{V}_n)} \right| = o_{\mathbb{P}}(1)\,. \tag{35}$$

Consider (34). Simple algebra yields:

$$\begin{aligned}
&\sup_z \left| \frac{\nu_{d^*}(\Delta_n(z\pm\eta_n)\cap(K_{osn}-\mathbb{V}_n))}{\nu_{d^*}(K_{osn}-\mathbb{V}_n)} - \frac{\nu_{d^*}(\Delta_n(z\pm\eta_n)\cap(K_{osn}-\mathbb{V}_n))}{\nu_{d^*}(T-\mathbb{V}_n)} \right| \\
&\leq \frac{\nu_{d^*}((T\setminus K_{osn})-\mathbb{V}_n)}{\nu_{d^*}(T-\mathbb{V}_n)} \\
&\leq \frac{\nu_{d^*}((T\setminus K_{osn})\cap B_{k_n}-\mathbb{V}_n)}{\nu_{d^*}(T-\mathbb{V}_n)} + \frac{\nu_{d^*}(B^c_{k_n}-\mathbb{V}_n)}{\nu_{d^*}(T-\mathbb{V}_n)}\,.
\end{aligned} \tag{36}$$

Since $\mathbb{V}_n$ is tight and the cone T has positive volume, for any $\epsilon > 0$ there exists $L_\epsilon > 0$ such that $\limsup_{n\to\infty} \mathbb{P}(\nu_{d^*}(T - \mathbb{V}_n) < L_\epsilon) \le \epsilon$ whence

$$1/\nu_{d^*}(T - \mathbb{V}_n) = O_{\mathbb{P}}(1)\,. \tag{37}$$

By tightness of $\mathbb{V}_n$ and the fact that $k_n \to \infty$ as $n \to \infty$ we may deduce $\nu_{d^*}(B_{k_n}^c - \mathbb{V}_n) = o_{\mathbb{P}}(1)$. Therefore, the second term on the right-hand side of (36) is $o_{\mathbb{P}}(1)$. The first term on the right-hand side of (36) may also be shown to be $o_{\mathbb{P}}(1)$ by tightness of $\mathbb{V}_n$ and Assumption 3.2(ii).

Now consider (35). Simple algebra yields:

$$\begin{aligned}
&\sup_z \left| \frac{\nu_{d^*}(\Delta_n(z \pm \eta_n) \cap (K_{osn} - \mathbb{V}_n))}{\nu_{d^*}(T - \mathbb{V}_n)} - \frac{\nu_{d^*}(\Delta_n(z \pm \eta_n) \cap (T - \mathbb{V}_n))}{\nu_{d^*}(T - \mathbb{V}_n)} \right| \\
&= \sup_z \frac{\nu_{d^*}(\Delta_n(z \pm \eta_n) \cap ((T \setminus K_{osn}) - \mathbb{V}_n))}{\nu_{d^*}(T - \mathbb{V}_n)} \\
&\le \frac{\nu_{d^*}((T \setminus K_{osn}) - \mathbb{V}_n)}{\nu_{d^*}(T - \mathbb{V}_n)}
\end{aligned}$$

which is $o_{\mathbb{P}}(1)$ by the preceding argument. Finally:

$$\begin{aligned}
&\sup_z |\nu_{d^*}(\Delta_n(z) \cap (T - \mathbb{V}_n)) - \nu_{d^*}(\Delta_n(z - \eta_n) \cap (T - \mathbb{V}_n))| \\
&\le \sup_z (\nu_{d^*}(\Delta_n(z)) - \nu_{d^*}(\Delta_n(z - \eta_n))) \\
&= \sup_z \left(F_{\chi^2_{d^*}}(z) - F_{\chi^2_{d^*}}(z - \eta_n) \right) = o(1)
\end{aligned} \tag{38}$$

where $F_{\chi^2_d}$ is the cdf of the χ^2_d distribution (which is uniformly continuous on $\mathbb{R}$). The $+\eta_n$ case handled similarly. Result (35) follows by combining (37) and (38).

Part (i) follows by combining (16) and the following inequality (39) due to Gao (2016):

$$\sup_z \left(\mathbb{P}_Z \Big(\|Z\|^2 \le z + \|\mathbf{T}^\perp v\|^2 \Big| Z \in T - v \Big) - \mathbb{P}_Z(\|\mathbf{T}Z\|^2 \le z) \right) \le 0 \tag{39}$$

holds for every $v \in \mathbb{R}^{d^*}$, where $\mathbb{P}_Z$ denotes the distribution of a $N(0, I_{d^*})$ random vector.

Part (ii) follows from (16) by observing that if $T = \mathbb{R}^{d^*}$ then $T - \mathbb{V}_n = \mathbb{R}^{d^*}$ and $\|\mathbf{T}^\perp V_n\| = 0$. ∎

In this appendix we often use the following expression (40) that is equivalent to equation (17) of Assumption 3.2'(i):

$$\sup_{\theta \in \Theta_{osn}} \left| nL_n(\theta) - \ell_n - \frac{1}{2}\|\mathbb{V}_n\|^2 + \frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 + f_{n,\perp}(\gamma_\perp(\theta)) \right| = o_{\mathbb{P}}(1)\,. \tag{40}$$

Lemma F.2. *Let Assumptions 3.1(i) and 3.2' hold. Then:*

$$\sup_{\theta\in\Theta_{osn}}\left|Q_n(\theta)-\left(-\inf_{t\in T}\|t-\mathbb{V}_n\|^2+\|\sqrt{n}\gamma(\theta)-\mathbb{V}_n\|^2+2f_{n,\perp}(\gamma_\perp(\theta))\right)\right|=o_{\mathbb{P}}(1)\,. \tag{41}$$

And hence

$$\sup_{\theta\in\Theta_I}Q_n(\theta)=\|\mathbf{T}\mathbb{V}_n\|^2+o_{\mathbb{P}}(1)\ .$$

Proof of Lemma F.2. Using successively Assumptions 3.1(i) and 3.2'(i) or (40), (ii) and (iii), we obtain:

$$\begin{aligned}
nL_n(\hat\theta)&=\sup_{\theta\in\Theta_{osn}}\left(\ell_n+\frac{1}{2}\|\mathbb{V}_n\|^2-\frac{1}{2}\|\sqrt{n}\gamma(\theta)-\mathbb{V}_n\|^2-f_{n,\perp}(\gamma_\perp(\theta))\right)+o_{\mathbb{P}}(1)\\
&=\sup_{\theta\in\Theta_{osn}}\left(\ell_n+\frac{1}{2}\|\mathbb{V}_n\|^2-\frac{1}{2}\|\sqrt{n}\gamma(\theta)-\mathbb{V}_n\|^2\right)-\inf_{\theta\in\Theta_{osn}}f_{n,\perp}(\gamma_\perp(\theta))+o_{\mathbb{P}}(1)\\
&=\ell_n+\frac{1}{2}\|\mathbb{V}_n\|^2-\inf_{t\in T}\frac{1}{2}\|t-\mathbb{V}_n\|^2+o_{\mathbb{P}}(1)\,,
\end{aligned}\tag{42}$$

where the last equality is due to the fact that $f_{n,\perp}(\cdot)\geq 0$ with $f_{n,\perp}(0)=0$, $\gamma_\perp(\theta)=0$ for all $\theta\in\Theta_I$,

$$0\leq\inf_{\theta\in\Theta_{osn}}f_{n,\perp}(\gamma_\perp(\theta))\leq f_{n,\perp}(\gamma_\perp(\bar\theta))=0\ \text{ for any }\bar\theta\in\Theta_I\ .$$

Then by Assumption 3.2'(i) or (40), and definition of Q_n, we obtain:

$$\begin{aligned}
Q_n(\theta)&=2\left(\ell_n+\frac{1}{2}\|\mathbb{V}_n\|^2-\inf_{t\in T}\frac{1}{2}\|t-\mathbb{V}_n\|^2+o_{\mathbb{P}}(1)\right)\\
&\quad-2\left(\ell_n+\frac{1}{2}\|\mathbb{V}_n\|^2-\frac{1}{2}\|\sqrt{n}\gamma(\theta)-\mathbb{V}_n\|^2-f_{n,\perp}(\gamma_\perp(\theta))+o_{\mathbb{P}}(1)\right)\\
&=-\inf_{t\in T}\|t-\mathbb{V}_n\|^2+\|\sqrt{n}\gamma(\theta)-\mathbb{V}_n\|^2+2f_{n,\perp}(\gamma_\perp(\theta))+o_{\mathbb{P}}(1)
\end{aligned}$$

where the $o_{\mathbb{P}}(1)$ term holds uniformly over Θ_{osn}. This proves expression (41).

Next, since $\gamma(\theta)=0$ and $\gamma_\perp(\theta)=0$ for $\theta\in\Theta_I$, and $f_{n,\perp}(0)=0$, we have:

$$\begin{aligned}
\sup_{\theta\in\Theta_I}Q_n(\theta)&=\|\mathbb{V}_n\|^2-\inf_{t\in T}\|t-\mathbb{V}_n\|^2+2\sup_{\theta\in\Theta_I}f_{n,\perp}(\gamma_\perp(\theta))+o_{\mathbb{P}}(1)\\
&=\|\mathbb{V}_n\|^2-\|\mathbb{V}_n-\mathbf{T}\mathbb{V}_n\|^2+o_{\mathbb{P}}(1)\\
&=\|\mathbf{T}\mathbb{V}_n\|^2+o_{\mathbb{P}}(1)
\end{aligned}$$

where the second equality is by definition of the projection onto the closed convex cone T, and the third inequality is by Moreau's decomposition Theorem (Hiriart-Urruty and Lemaréchal, 2001, Theorem 3.2.5, p.51). ∎

The key step in the proof of Theorem 3.2 is to establish the following lemma, which states that the posterior distribution of the QLR asymptotically (first-order) stochastically dominates the asymptotic distribution of the QLR, namely F_T defined in (15).

Lemma F.3. *Let Assumptions 3.1, 3.2' and 3.3' hold. Then:*

$$\sup_z \left(\Pi_n\big(\{\theta : Q_n(\theta) \le z\}\big|\,\mathbf{X}_n\big) - F_T(z)\right) \le o_{\mathbb{P}}(1)\,.$$

Proof of Lemma F.3. We first show that:

$$\sup_z \left(\Pi_n\big(\{\theta : Q_n(\theta) \le z\}\big|\,\mathbf{X}_n\big) - \mathbb{P}_{Z|\mathbf{X}_n}\Big(\|Z\|^2 \le z + \|\mathbf{T}^\perp \mathbb{V}_n\|^2 \Big| Z \in T - \mathbb{V}_n\Big)\right) \le o_{\mathbb{P}}(1) \tag{43}$$

holds. By identical arguments to the proof of Lemma 3.1, it is enough to characterize the large-sample behavior of $R_n(z)$ defined in (32). By Lemma F.2 and expression (40), there exists a positive sequence $(\eta_n)_{n\in\mathbb{N}}$ independent of z with $\eta_n = o(1)$ such that:

$$\sup_{\theta\in\Theta_{osn}} \left| Q_n(\theta) - \left(-\inf_{t\in T} \|t - \mathbb{V}_n\|^2 + \|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 + 2 f_{n,\perp}(\gamma_\perp(\theta)) \right) \right| \le \eta_n$$

$$\sup_{\theta\in\Theta_{osn}} \left| nL_n(\theta) - \ell_n - \frac{1}{2}\|\mathbb{V}_n\|^2 + \frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 + f_{n,\perp}(\gamma_\perp(\theta)) \right| \le \eta_n$$

both hold wpa1. Also note that for any z, we have

$$\begin{aligned}
&\left\{\theta \in \Theta_{osn} : -\inf_{t\in T} \|t - \mathbb{V}_n\|^2 + \|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 + 2 f_{n,\perp}(\gamma_\perp(\theta)) + \eta_n \le z\right\} \\
&\subseteq \left\{\theta \in \Theta_{osn} : -\inf_{t\in T} \|t - \mathbb{V}_n\|^2 + \|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 + \eta_n \le z\right\}
\end{aligned}$$

because $f_{n,\perp}(\cdot) \ge 0$. Therefore, wpa1 we have:

$$R_n(z) \le e^{\eta_n} \frac{\int_{\{\theta : \|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 \le z + \inf_{t\in T}\|t - \mathbb{V}_n\|^2 + \eta_n\} \cap \Theta_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 - f_{n,\perp}(\gamma_\perp(\theta))} \mathrm{d}\Pi(\theta)}{\int_{\Theta_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 - f_{n,\perp}(\gamma_\perp(\theta))} \mathrm{d}\Pi(\theta)}$$

uniformly in z.

Define $\Gamma_{osn} = \{\gamma(\theta) : \theta \in \Theta_{osn}\}$ and $\Gamma_{\perp,osn} = \{\gamma_\perp(\theta) : \theta \in \Theta_{osn}\}$. By similar arguments to the proof of Lemma 3.1, we use Assumption 3.3'(ii) and a change of variables to obtain:

$$\begin{aligned}
R_n(z) &\le e^{\eta_n}(1 + \bar{\eta}_n) \\
&\quad \times \frac{\int_{((\{\gamma : \|\sqrt{n}\gamma - \mathbb{V}_n\|^2 \le z + \inf_{t\in T}\|t - \mathbb{V}_n\|^2 + \eta_n) \cap \Gamma_{osn}) \times \Gamma_{\perp,osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma - \mathbb{V}_n\|^2 - f_{n,\perp}(\gamma_\perp)} \mathrm{d}(\gamma, \gamma_\perp)}{\int_{\Gamma_{osn} \times \Gamma_{\perp,osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma - \mathbb{V}_n\|^2 - f_{n,\perp}(\gamma_\perp)} \mathrm{d}(\gamma, \gamma_\perp)}
\end{aligned} \tag{44}$$

which holds uniformly in z (wpa1) for some $\bar{\eta}_n = o(1)$. Then by Tonelli's theorem and Assumption 3.2'(ii) we obtain:

$$R_n(z) \le e^{\eta_n}(1+\bar{\eta}_n)\frac{\int_{(\{\gamma:\|\sqrt{n}\gamma-\mathbb{V}_n\|^2\le z+\inf_{t\in T}\|t-\mathbb{V}_n\|^2+\eta_n)\cap\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma-\mathbb{V}_n\|^2}\mathrm{d}\gamma}{\int_{\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma-\mathbb{V}_n\|^2}\mathrm{d}\gamma} \tag{45}$$

uniformly in z. The inequality (43) then follows by similar arguments to the proof of Lemma 3.1. The result follows by combining inequality (43) with Gao (2016)'s inequality in (39). ■

Proof of Theorem 3.2. We verify the conditions of Lemma 2.1. Again, we assume wlog that $L_n(\hat{\theta}) = \sup_{\theta\in\Theta_{osn}} L_n(\theta) + o_{\mathbb{P}}(n^{-1})$. By Lemma F.2, when $\Sigma = I_{d^*}$, we have:

$$\sup_{\theta\in\Theta_I} Q_n(\theta) = \|\mathbf{T}\mathbb{V}_n\|^2 + o_{\mathbb{P}}(1) \rightsquigarrow \|\mathbf{T}Z\|^2 \tag{46}$$

where $Z \sim N(0, I_{d^*})$. Lemma F.3 shows that the posterior distribution of the QLR asymptotically (first-order) stochastically dominates the F_T distribution. The result follows by the same arguments as the proof of Theorem 3.1(i). ■

Lemma F.4. *Let Assumptions 3.1(i) and 3.2 or 3.2' and 3.5 hold. Then:*

$$\sup_{\theta\in\Theta_{osn}}\left|PQ_n(\Delta(\theta)) - f\left(\mathbb{V}_n - \sqrt{n}\gamma(\theta)\right) + \inf_{t\in T}\|\mathbb{V}_n - t\|^2\right| = o_{\mathbb{P}}(1)\,.$$

Proof of Lemma F.4. By display (29) in the proof of Lemma F.1 or display (42) in the proof of Lemma F.2 and Assumption 3.5, we obtain:

$$\begin{aligned} PQ_n(\Delta(\theta)) &= 2nL_n(\hat{\theta}) - 2nPL_n(\Delta(\theta)) \\ &= \left(2\ell_n + \|\mathbb{V}_n\|^2 - \inf_{t\in T}\|t-\mathbb{V}_n\|^2\right) - \left(2\ell_n + \|\mathbb{V}_n\|^2 - f\left(\mathbb{V}_n - \sqrt{n}\gamma(\theta)\right)\right) + o_{\mathbb{P}}(1) \end{aligned}$$

where the $o_{\mathbb{P}}(1)$ term holds uniformly over Θ_{osn}. The result is immediate. ■

For any open set $S \subset \mathbb{R}_+$ and any small $\epsilon > 0$, let $S^{-\epsilon}$ denote the ϵ-contraction of S and let $S_n^{-\epsilon} = \{s - \|\mathbf{T}^\perp\mathbb{V}_n\|^2 : s \in S^{-\epsilon}\}$.[30]

Lemma F.5. *Let Assumptions 3.1, 3.2, 3.3, and 3.5 or 3.1, 3.2', 3.3', and 3.5 hold, and let $z \mapsto \mathbb{P}_Z(f(Z) \le z)$ be uniformly continuous on $S \subset \mathbb{R}_+$ (where $Z \sim N(0, I_{d^*})$). Then for any $\epsilon > 0$ such that $S^{-\epsilon}$ is not empty:*

$$\sup_{z\in S_n^{-\epsilon}}\left|\Pi_n\big(\{\theta : PQ_n(\Delta(\theta)) \le z\}\,\big|\,\mathbf{X}_n\big) - \mathbb{P}_{Z|\mathbf{X}_n}\Big(f(Z) \le z + \|\mathbf{T}^\perp\mathbb{V}_n\|^2\Big| Z \in \mathbb{V}_n - T\Big)\right| = o_{\mathbb{P}}(1)\,.$$

[30]The ϵ-contraction of S is defined as $S^{-\epsilon} = \{z \in \mathbb{R} : \inf_{z'\in(\mathbb{R}\setminus S)}|z - z'| \ge \epsilon\}$. For instance, if $S = (0,\infty)$ then $S^{-\epsilon} = [\epsilon,\infty)$ and $S_n^{-\epsilon} = [\epsilon - \|\mathbf{T}^\perp\mathbb{V}_n\|^2, \infty)$.

If, in addition, $T = \mathbb{R}^{d^*}$, then:

$$\sup_{z \in S^{-\epsilon}} \left| \Pi_n\big(\{\theta : PQ_n(\Delta(\theta)) \le z\} \,\big|\, \mathbf{X}_n\big) - \mathbb{P}_{Z|\mathbf{X}_n}\big(f(Z) \le z\big) \right| = o_{\mathbb{P}}(1)\,.$$

Proof of Lemma F.5. We prove the result under Assumptions 3.1, 3.2', 3.3', and 3.5. The proof under Assumptions 3.1, 3.2, 3.3, and 3.5 follows similarly. By the same arguments as the proof of Lemma 3.1, it suffices to characterize the large-sample behavior of:

$$R_n(z) := \frac{\int_{\{\theta : PQ_n(\Delta(\theta)) \le z\} \cap \Theta_{osn}} e^{L_n(\theta) - \ell_n - \frac{1}{2}\|\mathbb{V}_n\|^2} \, \mathrm{d}\Pi(\theta)}{\int_{\Theta_{osn}} e^{L_n(\theta) - \ell_n - \frac{1}{2}\|\mathbb{V}_n\|^2} \, \mathrm{d}\Pi(\theta)}\,. \tag{47}$$

By Lemma F.4 and expression (40), there exists a positive sequence $(\eta_n)_{n \in \mathbb{N}}$ independent of z with $\eta_n = o(1)$ such that the inequalities:

$$\sup_{\theta \in \Theta_{osn}} \left| PQ_n(\Delta(\theta)) - h_n(\mathbb{V}_n - \sqrt{n}\gamma(\theta)) \right| \le \eta_n$$

$$\sup_{\theta \in \Theta_{osn}} \left| nL_n(\theta) - \ell_n - \frac{1}{2}\|\mathbb{V}_n\|^2 - \left(-\frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 - f_{n,\perp}(\gamma_\perp(\theta)) \right) \right| \le \frac{\eta_n}{2}$$

both hold wpa1, where $h_n(\mathbb{V}_n - \sqrt{n}\gamma(\theta)) = f(\mathbb{V}_n - \sqrt{n}\gamma(\theta)) - \inf_{t \in T} \|\mathbb{V}_n - t\|^2$. Therefore, wpa1 we have:

$$\begin{aligned} & e^{-\eta_n} \frac{\int_{\{\theta : h_n(\mathbb{V}_n - \sqrt{n}\gamma(\theta)) \le z - \eta_n\} \cap \Theta_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 - f_{n,\perp}(\gamma_\perp(\theta))} \mathrm{d}\Pi(\theta)}{\int_{\Theta_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 - f_{n,\perp}(\gamma_\perp(\theta))} \mathrm{d}\Pi(\theta)} \\ & \le R_n(z) \le e^{\eta_n} \frac{\int_{\{\theta : h_n(\mathbb{V}_n - \sqrt{n}\gamma(\theta)) \le z + \eta_n\} \cap \Theta_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 - f_{n,\perp}(\gamma_\perp(\theta))} \mathrm{d}\Pi(\theta)}{\int_{\Theta_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 - f_{n,\perp}(\gamma_\perp(\theta))} \mathrm{d}\Pi(\theta)} \end{aligned}$$

uniformly in z. By similar arguments to the proof of Lemma F.3, we may use the change of variables $\theta \mapsto (\gamma(\theta), \gamma_\perp(\theta))$, continuity of π_{Γ^*} (Assumption 3.3'(ii)), and Tonelli's theorem to rewrite the above system of inequalities as:

$$\begin{aligned} & (1 - \bar{\eta}_n) e^{-\eta_n} \frac{\int_{\{\gamma : h_n(\mathbb{V}_n - \sqrt{n}\gamma) \le z - \eta_n\} \cap \Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma - \mathbb{V}_n\|^2} \mathrm{d}\gamma}{\int_{\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma - \mathbb{V}_n\|^2} \mathrm{d}\gamma} \\ & \le R_n(z) \le (1 + \bar{\eta}_n) e^{\eta_n} \frac{\int_{\{\gamma : h_n(\mathbb{V}_n - \sqrt{n}\gamma) \le z + \eta_n\} \cap \Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma - \mathbb{V}_n\|^2} \mathrm{d}\gamma}{\int_{\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma - \mathbb{V}_n\|^2} \mathrm{d}\gamma} \end{aligned}$$

which holds uniformly in z (wpa1) for some $\bar{\eta}_n = o(1)$. Let $K_{osn} = \{\sqrt{n}\gamma : \gamma \in \Gamma_{osn}\}$ and $H_n(z) = \{\kappa \in \mathbb{R}^{d^*} : h_n(\kappa) \le z\} = \{\kappa \in \mathbb{R}^{d^*} : f(\kappa) \le z + \inf_{t \in T} \|t - \mathbb{V}_n\|^2\}$. A second change of variables $\mathbb{V}_n - \sqrt{n}\gamma \mapsto \kappa$ yields:

$$(1 - \bar{\eta}_n) e^{-\eta_n} \frac{\int_{H_n(z - \eta_n) \cap (\mathbb{V}_n - K_{osn})} e^{-\frac{1}{2}\|\kappa\|^2} \mathrm{d}\kappa}{\int_{(\mathbb{V}_n - K_{osn})} e^{-\frac{1}{2}\|\kappa\|^2} \mathrm{d}\kappa} \le R_n(z) \le (1 + \bar{\eta}_n) e^{\eta_n} \frac{\int_{H_n(z + \eta_n) \cap (\mathbb{V}_n - K_{osn})} e^{-\frac{1}{2}\|\kappa\|^2} \mathrm{d}\kappa}{\int_{(\mathbb{V}_n - K_{osn})} e^{-\frac{1}{2}\|\kappa\|^2} \mathrm{d}\kappa}$$

which holds uniformly in z (wpa1).

To complete the proof, it remains to show that:

$$\sup_{z \in S_n^{-\epsilon}} \left| \frac{\nu_{d^*}(H_n(z \pm \eta_n) \cap (\mathbb{V}_n - K_{osn})}{\nu_{d^*}((\mathbb{V}_n - K_{osn}))} - \frac{\nu_{d^*}(H_n(z) \cap (\mathbb{V}_n - T)}{\nu_{d^*}((\mathbb{V}_n - T))} \right| = o_{\mathbb{P}}(1) \,. \tag{48}$$

By similar arguments to the proof of Lemma 3.1, it is enough to show that:

$$\sup_{z \in S_n^{-\epsilon}} |\nu_{d^*}(H_n(z \pm \eta_n) \cap (\mathbb{V}_n - T)) - \nu_{d^*}(H_n(z) \cap (\mathbb{V}_n - T))| = o_{\mathbb{P}}(1)$$

Notice that:

$$\begin{aligned}
&\sup_{z \in S_n^{-\epsilon}} |\nu_{d^*}(H_n(z - \eta_n) \cap (\mathbb{V}_n - T)) - \nu_{d^*}(H_n(z) \cap (\mathbb{V}_n - T)| \\
&\leq \sup_{z \in S_n^{-\epsilon}} |\nu_{d^*}(H_n(z - \eta_n)) - \nu_{d^*}(H_n(z))| \\
&= \sup_{z \in S^{-\epsilon}} |\nu_{d^*}(\{\kappa : f(\kappa) \leq z - \eta_n\}) - \nu_{d^*}(\{\kappa : f(\kappa) \leq z\})| = o(1)
\end{aligned}$$

by uniform continuity of $z \mapsto \nu_{d^*}(\{\kappa : f(\kappa) \leq z\})$ on S. The $+\eta$ case is handled similarly. ∎

Proof of Theorem 3.3. We verify the conditions of Lemma 2.2. Again, we assume wlog that $L_n(\hat{\theta}) = \sup_{\theta \in \Theta_{osn}} L_n(\theta) + o_{\mathbb{P}}(n^{-1})$. It follows from Lemma F.4 (taking $\gamma(\theta) = 0$ for any $\theta \in \Theta_I$) that when $T = \mathbb{R}^{d^*}$ and $\Sigma = I_{d^*}$ that:

$$PQ_n(\Delta(\theta)) = f(\mathbb{V}_n) + o_{\mathbb{P}}(1) \rightsquigarrow f(Z) \quad \text{for all } \theta \in \Theta_I \,.$$

where $Z \sim N(0, I_{d^*})$. Let ξ_α denote the α quantile of $f(Z)$. Then:

$$\xi_{n,\alpha}^{mc} = \xi_\alpha + (\xi_{n,\alpha}^{post} - \xi_\alpha) + (\xi_{n,\alpha}^{mc} - \xi_{n,\alpha}^{post}) = \xi_\alpha + o_{\mathbb{P}}(1)$$

by Lemma F.5 and Assumption 3.6. ∎

Proof of Theorem 3.4. It is enough to verify the conditions of Lemma 2.2. Let T_1^c and T_2^c denote the complement of T_1 and T_2 in $\mathbb{R}^{d^*}$ and let $\mathbb{P}_Z$ denote the distribution of $Z \sim N(0, I_{d^*})$.

Now suppose that $T_1^c \cap T_2^c$ is not empty. For any $w \geq 0$:

$$\begin{aligned}
\mathbb{P}(f(Z) \leq w) &= \mathbb{P}_Z(f(Z) \leq w | Z \in T_1 \cap T_2)\mathbb{P}_Z(Z \in T_1 \cap T_2) \\
&\quad + \mathbb{P}_Z(f(Z) \leq w | Z \in T_1^c \cap T_2^c)\mathbb{P}_Z(Z \in T_1^c \cap T_2^c) \\
&\quad + \mathbb{P}_Z(f(Z) \leq w | Z \in T_1^c \cap T_2)\mathbb{P}_Z(Z \in T_1^c \cap T_2) \\
&\quad + \mathbb{P}_Z(f(Z) \leq w | Z \in T_1 \cap T_2^c)\mathbb{P}_Z(Z \in T_1 \cap T_2^c) \\
&= p\big(\mathbb{P}_Z(f(Z) \leq w | Z \in T_1 \cap T_2) + \mathbb{P}_Z(f(Z) \leq w | Z \in T_1^c \cap T_2^c)\big) \\
&\quad + (1-2p)\mathbb{P}_Z(f(Z) \leq w | Z \in T_1^c \cap T_2) \qquad (49)
\end{aligned}$$

by symmetry of Gaussian measure, where $p = \mathbb{P}_Z(Z \in T_1 \cap T_2)$. We omit the term conditional on $T_1^c \cap T_2^c$ whenever $T_1^c \cap T_2^c = \emptyset$.

If $Z \in T_1 \cap T_2$ then $f(Z) = 0$ and hence

$$\mathbb{P}_Z(f(Z) \leq w | Z \in T_1 \cap T_2) = 1 \,. \qquad (50)$$

If $Z \in T_1^c \cap T_2$ then we have $\inf_{t \in T_2} \|Z - t\|^2 = 0$ and hence, by Moreau's decomposition theorem (Hiriart-Urruty and Lemaréchal, 2001, Theorem 3.2.5, p.51), we obtain:

$$f(Z) = \inf_{t \in T_1} \|Z - t\|^2 = \|\mathbf{T}_1^{\perp} Z\|^2$$

where $\mathbf{T}_1^{\perp}$ denotes the projection onto the polar cone T_1^o of T_1. Here $T_1^o \subset T_1^c$ is a ray extending from the origin that is orthogonal to the supporting hyperplane for T_1. Since the orthogonal projection of the standard normal random vector Z onto a line passing through the origin is distributed as χ_1^2 and the length $\|Z\|$ and direction $Z/\|Z\|$ of Z are independently distributed, we may deduce:

$$\mathbb{P}_Z(f(Z) \leq w | Z \in T_1^c \cap T_2) = \mathbb{P}_Z(\|\mathbf{T}_1^{\perp} Z\|^2 \leq w | Z \in T_1^c \cap T_2) = F_{\chi_1^2}(w) \,. \qquad (51)$$

By similar arguments, if $Z \in T_1^c \cap T_2^c$ we have $f(Z) = \|\mathbf{T}_1^{\perp} Z\|^2 \vee \|\mathbf{T}_2^{\perp} Z\|^2$. Each $\|\mathbf{T}_i^{\perp} Z\|^2$ is distributed as χ_1^2 (conditionally upon $Z \in T_1^c \cap T_2^c$). Therefore:

$$\mathbb{P}_Z(\|\mathbf{T}_1^{\perp} Z\|^2 \vee \|\mathbf{T}_2^{\perp} Z\|^2 \leq w | Z \in T_1^c \cap T_2^c)$$

is minimized when T_1^o and T_2^o are orthogonal, in which case $\|\mathbf{T}_1^{\perp} Z\|^2$ and $\|\mathbf{T}_2^{\perp} Z\|^2$ are independent χ_1^2 random variables (conditional upon $Z \in T_1^c \cap T_2^c$). It follows that:

$$\mathbb{P}_Z(f(Z) \leq w | Z \in T_1^c \cap T_2^c) \geq F_{\chi_1^2}(w)^2 \,. \qquad (52)$$

Now, substituting (50), (51), and (52) into (49) yields:

$$\mathbb{P}(f(Z) \leq w) \geq p(1 + F_{\chi_1^2}(w)^2) + (1-2p)F_{\chi_1^2}(w) \,.$$

and hence:

$$\mathbb{P}(f(Z) \leq w) - F_{\chi_1^2}(w) \geq p(1 - F_{\chi_1^2}(w))^2 \geq 0 \,.$$

Let w_α be the α quantile of $W = f(Z)$ in Lemma 2.2. It follows that $\chi^2_{1,\alpha} \geq w_\alpha$. ■

Proof of Proposition 3.1. It follows from part (iii) and display (29) or display (42) that:

$$2nL_n(\hat\theta) = 2\ell_n + \|\mathbb{V}_n\|^2 - \inf_{t \in T} \|\mathbb{V}_n - t\|^2 + o_{\mathbb{P}}(1)\,.$$

Moreover, applying parts (i) and (ii), we obtain:

$$\begin{aligned}\inf_{m \in M_I} \sup_{\theta \in \mu^{-1}(m)} 2nL_n(\theta) &= \min_{m \in \{\underline{m}, \overline{m}\}} \sup_{\theta \in \mu^{-1}(m)} 2nL_n(\theta) + o_{\mathbb{P}}(1) \\ &= \min_{m \in \{\underline{m}, \overline{m}\}} \left(2\ell_n + \|\mathbb{V}_n\|^2 - \inf_{t \in T_m} \|\mathbb{V}_n - t\|^2\right) + o_{\mathbb{P}}(1)\,.\end{aligned}$$

Therefore:

$$\sup_{m \in M_I} \inf_{\theta \in \mu^{-1}(m)} Q_n(\theta) = \left(\inf_{t \in T_{\underline{m}}} \|\mathbb{V}_n - t\|^2 \vee \inf_{t \in T_{\overline{m}}} \|\mathbb{V}_n - t\|^2\right) - \inf_{t \in T} \|\mathbb{V}_n - t\|^2 + o_{\mathbb{P}}(1)\,.$$

The result follows by part (iv) and $\Sigma = I_{d^*}$. ■

F.2 Proofs and Additional Lemmas for Section 4

Proof of Proposition 4.1. Wlog we can take $\tilde\gamma_0 = 0$. By condition (b), for any $\tilde\gamma \in U$ we have:

$$nL_n(\tilde\gamma) = nL_n(\tilde\gamma_0) + (\sqrt{n}\tilde\gamma)'(\sqrt{n}\mathbb{P}_n\dot\ell_{\tilde\gamma_0}) + \frac{1}{2}(\sqrt{n}\tilde\gamma)'(\mathbb{P}_n\ddot\ell_{\tilde\gamma^*})(\sqrt{n}\tilde\gamma)$$

where $\tilde\gamma^*$ is in the segment between $\tilde\gamma$ and $\tilde\gamma_0$ for each element of $\mathbb{P}_n\ddot\ell_{\tilde\gamma^*}$. We may deduce from Lemma 2.4 of Newey and McFadden (1994) that $\sup_{\tilde\gamma:\|\tilde\gamma\| \leq n^{1/4}} \|(\mathbb{P}_n\ddot\ell_{\tilde\gamma^*}) - P_0(\ddot\ell_{\tilde\gamma_0})\| = o_{\mathbb{P}}(1)$ holds under conditions (a) and (b). Since this term is $o_{\mathbb{P}}(1)$, we can choose a positive sequence $(r_n)_{n \in \mathbb{N}}$ with $r_n \to \infty$, $r_n = o(n^{1/4})$ such that $r_n^2 \sup_{\tilde\gamma:\|\tilde\gamma\| \leq n^{1/4}} \|(\mathbb{P}_n\ddot\ell_{\tilde\gamma^*}) - P_0(\ddot\ell_{\tilde\gamma_0})\| = o_{\mathbb{P}}(1)$ holds. Take $\Theta_{osn} = \{\theta \in \Theta : \|\tilde\gamma(\theta)\| \leq r_n/\sqrt{n}\}$. Assumption 3.2(i) then holds with $\gamma(\theta) = \mathbb{I}_{\tilde\gamma_0}^{1/2}\tilde\gamma(\theta)$. Assumption 3.2(ii) is also trivially satisfied with $T = \mathbb{R}^{d^*}$ because $\tilde\gamma_0 = 0 \in \mathrm{int}(\widetilde\Gamma)$ by Condition (b). Assumption 3.1(ii) follows under conditions (c) and (d) by Theorem 5.1 of Ghosal, Ghosh, and van der Vaart (2000). ■

Lemma F.6. *Consider the missing data model with Θ as in (12) and a flat prior on Θ. Then Assumption 3.1(ii) is satisfied with $\Theta_{osn} := \{\theta : |\kappa_{11}(\theta) - \kappa_{11}| \leq k_n/\sqrt{n}, \kappa_{00}(\theta) \leq k_n/n\}$ for any positive sequence $(k_n)_{n \in \mathbb{N}}$ with $k_n \to \infty$, $k_n/\sqrt{n} = o(1)$.*

Proof of Lemma F.6. Let $S_n = \sum_{i=1}^n Y_i$. The flat prior under the map $\theta \mapsto (\kappa_{11}(\theta), \kappa_{00}(\theta))'$ induces a flat prior on $\{(a,b) \in [0,1] : 0 \leq a \leq 1 - b\}$. Take n sufficiently large that $[\kappa_{11} -$

$k_n/\sqrt{n}, \kappa_{11} + k_n/\sqrt{n}] \subseteq [0,1]$ and $k_n/n < 1 - \kappa_{11}$. Then:

$$
\begin{aligned}
&\Pi_n(\Theta_{osn}^c|\mathbf{X}_n) \\
&= \frac{\int_0^{\kappa_{11}-k_n/\sqrt{n}} \int_0^{1-a} (a)^{S_n}(1-a-b)^{n-S_n}\,\mathrm{d}b\,\mathrm{d}a}{\int_0^1 \int_0^{1-a} (a)^{S_n}(1-a-b)^{n-S_n}\,\mathrm{d}b\,\mathrm{d}a} + \frac{\int_{\kappa_{11}+k_n/\sqrt{n}}^1 \int_0^{1-a} (a)^{S_n}(1-a-b)^{n-S_n}\,\mathrm{d}b\,\mathrm{d}a}{\int_0^1 \int_0^{1-a} (a)^{S_n}(1-a-b)^{n-S_n}\,\mathrm{d}b\,\mathrm{d}a} \\
&\quad + \frac{\int_{\kappa_{11}-k_n/\sqrt{n}}^{\kappa_{11}+k_n/\sqrt{n}} \int_{k_n/n}^{1-a} (a)^{S_n}(1-a-b)^{n-S_n}\,\mathrm{d}b\,\mathrm{d}a}{\int_0^1 \int_0^{1-a} (a)^{S_n}(1-a-b)^{n-S_n}\,\mathrm{d}b\,\mathrm{d}a} \quad =: \quad I_1 + I_2 + I_3\,.
\end{aligned}
$$

Integrating first with respect to b yields:

$$
\begin{aligned}
I_1 + I_2 &= \frac{\int_0^{\kappa_{11}-k_n/\sqrt{n}} (a)^{S_n}(1-a)^{n-S_n+1}\,\mathrm{d}a}{\int_0^1 (a)^{S_n}(1-a)^{n-S_n+1}\,\mathrm{d}a} + \frac{\int_{\kappa_{11}+k_n/\sqrt{n}}^1 (a)^{S_n}(1-a)^{n-S_n+1}\,\mathrm{d}a}{\int_0^1 (a)^{S_n}(1-a)^{n-S_n+1}\,\mathrm{d}a} \\
&= \mathbb{P}_{U|S_n}(|U-\kappa_{11}| > k_n/\sqrt{n})
\end{aligned}
$$

where $U|S_n \sim \mathrm{Beta}(S_n+1, n-S_n+2)$. By properties of the Beta distribution:

$$
\begin{aligned}
\mathbb{E}[U|S_n] &= \frac{S_n+1}{n+3} \\
\mathrm{Var}[U|S_n] &= \frac{(S_n+1)(n-S_n+2)}{(n+3)^2(n+4)}\,.
\end{aligned}
$$

By the triangle inequality, the fact that $\mathbb{E}[U|S_n] = \kappa_{11} + O_{\mathbb{P}}(n^{-1/2})$, and Chebyshev's inequality:

$$
\begin{aligned}
I_1 + I_2 &\le \mathbb{P}_{U|S_n}\Big(|U - \mathbb{E}[U|S_n]| > k_n/(2\sqrt{n})\Big) + \mathbb{1}\Big\{|\mathbb{E}[U|S_n] - \kappa_{11}| > k_n/(2\sqrt{n})\Big\} \\
&= \mathbb{P}_{U|S_n}\Big(|U - \mathbb{E}[U|S_n]| > k_n/(2\sqrt{n})\Big) + o_{\mathbb{P}}(1) \\
&\le \frac{4n}{k_n^2}\frac{(S_n+1)(n-S_n+2)}{(n+3)^2(n+4)} + o_{\mathbb{P}}(1) \\
&\le \frac{4}{k_n^2}\frac{(\frac{S_n}{n}+\frac{1}{n})(1-\frac{S_n}{n}+\frac{2}{n})}{(1+\frac{3}{n})^2(1+\frac{4}{n})} + o_{\mathbb{P}}(1)
\end{aligned}
$$

which is $o_{\mathbb{P}}(1)$ (because $k_n \to \infty$).

Similarly, for I_3 we have:

$$
\begin{aligned}
I_3 &= \frac{\int_{\kappa_{11}-k_n/\sqrt{n}}^{\kappa_{11}+k_n/\sqrt{n}} (a)^{S_n}(1-a-(k_n/n))^{n-S_n+1}\,\mathrm{d}a}{\int_0^1 (a)^{S_n}(1-a)^{n-S_n+1}\,\mathrm{d}a} \\
&\le \frac{\int_0^{1-(k_n/n)} (a)^{S_n}(1-a-(k_n/n))^{n-S_n+1}\,\mathrm{d}a}{\int_0^1 (a)^{S_n}(1-a)^{n-S_n+1}\,\mathrm{d}a}\,.
\end{aligned}
$$

Using the change of variables $a \mapsto c(a) := \frac{1-a-k_n/n}{1-k_n/n}$ in the numerator yields:

$$I_3 \leq (1-(k_n/n))^{n+2} \frac{\int_0^1 (1-c)^{S_n}(c)^{n-S_n+1}\,\mathrm{d}c}{\int_0^1 (a)^{S_n}(1-a)^{n-S_n+1}\,\mathrm{d}a} = (1-(k_n/n))^{n+2}$$

and so $I_3 \to 0$ as $n \to \infty$ (because $k_n \to \infty$). ■

F.2.1 Additional results for the quadratic expansion of the log-likelihood in general non-identifiable models

For the following lemma, we let $(r_n)_{n\in\mathbb{N}}$ be a positive sequence with $r_n \to \infty$ and $r_n = o(n^{1/2})$, and let $\mathcal{P}_{osn} = \{p \in \mathcal{P} : h(p,p_0) \leq r_n/\sqrt{n}\}$ and $\Theta_{osn} = \{\theta \in \Theta : h(p_\theta,p_0) \leq r_n/\sqrt{n}\}$. For each $p \in \mathcal{P}$ with $h(p,p_0) > 0$ we also define $S_p = \sqrt{p/p_0} - 1$ and $s_p = S_p/h(p,p_0)$. Recall the definitions of $\overline{\mathcal{D}}_\varepsilon$, the tagent cone Λ and the projection $\mathbf{\Lambda}$ from Section 4.1.2. Finally, we say $\mathcal{P}$ is r_n-DQM (with respect to p_0) if each p is absolutely continuous with respect to p_0 and for each $p \in \mathcal{P}$ there are elements $g(p) \in \Lambda$ and remainders $R(p) \in L^2(\lambda)$ such that:

$$\sqrt{p} - \sqrt{p_0} = g(p)\sqrt{p_0} + h(p,p_0)R(p)$$

with $\sup\{r_n\|R(p)\|_{L^2(\lambda)} : h(p,p_0) \leq r_n/\sqrt{n}\} \to 0$ as $n \to \infty$.

Lemma F.7. *Let the following conditions hold.*
(i) $\mathcal{P}$ is r_n-DQM
(ii) There exists $\varepsilon > 0$ such that $\{s_p^2 : s_p \in \overline{\mathcal{D}}_\varepsilon\}$ is P_0-Glivenko Cantelli
(iii) $\overline{\mathcal{D}}_\varepsilon$ has a measurable envelope $D : \mathscr{X} \to \mathbb{R}$ with $\max_{i\leq i\leq n} D(X_i) = o_{\mathbb{P}}(\sqrt{n}/r_n^3)$
(iv) $\sup_{p\in\mathcal{P}_{osn}} |\mathbb{G}_n(S_p - \mathbf{\Lambda}S_p)| = o_{\mathbb{P}}(n^{-1/2})$
(v) $\sup_{p\in\mathcal{P}_{osn}} |(\mathbb{P}_n - P_0)S_p^2| = o_{\mathbb{P}}(n^{-1})$.
Then:

$$\sup_{\theta\in\Theta_{osn}} \left| nL_n(\theta) - \left(n\mathbb{P}_n \log p_0 + n\mathbb{P}_n(2\mathbf{\Lambda}S_{p_\theta}) - \frac{1}{2}nP_0((2\mathbf{\Lambda}S_{p_\theta})^2) \right) \right| = o_{\mathbb{P}}(1)\,.$$

When the linear hull $\mathrm{Span}(\Lambda)$ *has finite dimension $d^* \geq 1$ we may restate this result as:*

$$\sup_{\theta\in\Theta_{osn}} \left| nL_n(\theta) - \left(n\mathbb{P}_n \log p_0 + (\sqrt{n}\gamma(\theta))'V_n - \frac{1}{2}\|\sqrt{n}\gamma(\theta)\|^2 \right) \right| = o_{\mathbb{P}}(1)\,.$$

where $V_n = \mathbb{G}_n(\psi)$, $\psi = (\psi_1,\ldots,\psi_{d^*})'$, $\psi_1,\ldots,\psi_{d^*}$ is an orthonormal basis for the linear hull $\mathrm{Span}(\Lambda)$, *and $\gamma(\theta)$ is defined by $\mathbf{\Lambda}(2S_{p_\theta}) = \gamma(\theta)'\psi$.*

Proof of Lemma F.7. We first show that:

$$\sup_{p\in\mathcal{P}_{osn}} \left| n\mathbb{P}_n \log(p/p_0) - 2n\mathbb{P}_n(S_p - P_0(S_p)) + n(\mathbb{P}_n S_p^2 + h^2(p,p_0)) \right| = o_{\mathbb{P}}(1) \qquad (53)$$

holds. To do so, we adapt arguments used in Theorem 1 of Azaïs et al. (2009), Theorem 3.1 in Gassiat (2002), and Theorem 2.1 in Liu and Shao (2003). Take n sufficiently large that

$r_n/\sqrt{n} \leq \varepsilon$ (whence $\mathcal{P}_{osn} \subseteq \overline{\mathcal{D}}_\varepsilon$). For each $p \in \mathcal{P}_{osn} \setminus \{p_0\}$ we have:

$$n\mathbb{P}_n \log(p/p_0) = 2n\mathbb{P}_n S_p - n\mathbb{P}_n S_p^2 + 2h(p,p_0)^2 n\mathbb{P}_n s_p^2 r(h(p,p_0)s_p) \tag{54}$$

where $r(u) = (\log(1+u) - u - \frac{1}{2}u^2)/u^2$ and $\lim_{u\to 0} |r(u)/(\frac{1}{3}u) - 1| = 0$. Condition (iii) implies:

$$\sup_{p\in\mathcal{P}_{osn}} \max_{1\leq i\leq n} h(p,p_0)|s_p| \leq \frac{r_n}{\sqrt{n}} \max_{1\leq i\leq n} D(X_i) = o_{\mathbb{P}}(r_n^{-2})$$

and hence:

$$\sup_{p\in\mathcal{P}_{osn}} \max_{1\leq i\leq n} |r(h(p,p_0)s_p)| = o_{\mathbb{P}}(r_n^{-2})$$

Therefore:

$$\begin{aligned}\sup_{p\in\mathcal{P}_{osn}} |2h(p,p_0)^2 n\mathbb{P}_n s_p^2 r(h(p,p_0)s_p)| &\leq 2r_n^2 \times o_{\mathbb{P}}(r_n^{-2}) \times \sup_{p\in\mathcal{P}_{osn}} \mathbb{P}_n s_p^2 \\ &\leq 2r_n^2 \times o_{\mathbb{P}}(r_n^{-2}) \times (1 + o_{\mathbb{P}}(1)) = o_{\mathbb{P}}(1)\end{aligned}$$

where the second inequality is by Condition (ii). Expression (53) follows by adding and subtracting $2nP_0(S_p) = -nh^2(p,p_0)$ to the right-hand side of (54).

To complete the proof, it remains to show:

$$\sup_{p\in\mathcal{P}_{osn}} |\mathbb{P}_n(S_p - P_0(S_p) - \mathbf{\Lambda} S_p)| = o_{\mathbb{P}}(n^{-1}) \tag{55}$$

$$\sup_{p\in\mathcal{P}_{osn}} \left|\mathbb{P}_n(S_p^2) + h^2(p,p_0) - 2P_0((\mathbf{\Lambda} S_p)^2)\right| = o_{\mathbb{P}}(n^{-1})\,. \tag{56}$$

Each element of Λ has mean zero, hence $P_0(\mathbf{\Lambda} S_p) = 0$ for each $p \in \mathcal{P}$, we can deduce:

$$\sup_{p\in\mathcal{P}_{osn}} |\mathbb{P}_n(S_p - P_0(S_p) - \mathbf{\Lambda} S_p)| = n^{-1/2} \times \sup_{p\in\mathcal{P}_{osn}} |\mathbb{G}_n(S_p - \mathbf{\Lambda} S_p)|$$

from which (55) follows by Condition (iv).

As $P_0(S_p^2) = h^2(p,p_0)$, in order to prove (56) it suffices to prove:

$$\sup_{p\in\mathcal{P}_{osn}} |(\mathbb{P}_n - P_0)(S_p^2)| = o_{\mathbb{P}}(n^{-1}) \tag{56a}$$

$$\sup_{p\in\mathcal{P}_{osn}} |P_0(S_p^2) - P_0((\mathbf{\Lambda} S_p)^2)| = o_{\mathbb{P}}(n^{-1})\,. \tag{56b}$$

Result (56a) holds by condition (v). It remains to prove (56b). Under Condition (i), for each $p \in \mathcal{P}$ there is a $g(p) \in \Lambda$ and remainder $R^*(p)$ such that:

$$S_p = g(p) + h(p,p_0)R^*(p)$$

with $\sup\{r_n \|R^*(p)\|_{L^2(P_0)} : h(p,p_0) \le r_n/\sqrt{n}\} \to 0$ as $n \to \infty$. It follows by definition of $\mathbf{\Lambda}$ that:

$$\|S_p - \mathbf{\Lambda} S_p\|_{L^2(P_0)} \le \|S_p - g(p)\|_{L^2(P_0)} = h(p,p_0)\|R^*(p)\|_{L^2(P_0)} \tag{57}$$

for each $p \in \mathcal{P}$. By Moreau's decomposition theorem (Hiriart-Urruty and Lemaréchal, 2001, Theorem 3.2.5, p.51) and inequality (57) we may deduce:

$$\sup_{p\in\mathcal{P}_{osn}} |P_0(S_p^2) - P_0((\mathbf{\Lambda} S_p)^2)| = \sup_{p\in\mathcal{P}_{osn}} \|S_p - \mathbf{\Lambda} S_p\|^2_{L^2(P_0)} \le \sup_{p\in\mathcal{P}_{osn}} h(p,p_0)^2 \|R^*(p)\|^2_{L^2(P_0)}\,.$$

Result (56b) then follows from definition of $\mathcal{P}_{osn}$ and Condition (i). This proves the first result.

The second result is immediate by defining $V_n = \mathbb{G}_n(\psi)$, $\psi = (\psi_1,\ldots,\psi_{d^*})'$, $\psi_1,\ldots,\psi_{d^*}$ is an orthonormal basis for the linear hull Span(Λ), and $\gamma(\theta)$ by $\mathbf{\Lambda}(2S_{p_\theta}) = \gamma(\theta)'\psi$, then noting that $P_0((\mathbf{\Lambda}(2S_{p_\theta}))^2) = \gamma(\theta)' P_0(\psi\psi')\gamma(\theta) = \|\gamma(\theta)\|^2$. ∎

Proof of Proposition 4.2. To verify Assumption 3.2(i) it suffices to verify the conditions of Lemma F.7. By DQM (condition (b)) we have $\sup\{\|R(p)\|_{L^2(\lambda)} : h(p,p_0) \le n^{-1/4}\} \to 0$ as $n \to \infty$. Therefore, we may choose a slowly diverging sequence $(a_n)_{n\in\mathbb{N}}$ with $a_n \le n^{1/4}$ such that

$$\sup\{a_n\|R(p)\|_{L^2(\lambda)} : h(p,p_0) \le a_n/\sqrt{n}\} \to 0 \quad \text{as } n \to \infty$$

and hence

$$\sup\{r_n\|R(p)\|_{L^2(\lambda)} : h(p,p_0) \le r_n/\sqrt{n}\} \to 0 \quad \text{as } n \to \infty$$

for any positive sequence $(r_n)_{n\in\mathbb{N}}$ with $r_n \le a_n$. This verifies Condition (i) of Lemma F.7.

Condition (c) implies $\overline{\mathcal{D}}_\varepsilon$ is Donsker and so $\{s_p^2 : s_p \in \overline{\mathcal{D}}_\varepsilon\}$ is Glivenko-Cantelli (van der Vaart and Wellner, 1996, Lemma 2.10.14), which verifies Condition (ii) of Lemma F.7. Choose a positive sequence $(b_n)_{n\in\mathbb{N}}$ with $b_n \to \infty$ such that $b_n^2 \sup_{s_p\in\overline{\mathcal{D}}_\varepsilon} |(\mathbb{P}_n - P_0)s_p^2| = o_{\mathbb{P}}(1)$ and so:

$$\sup_{p:h(p,p_0)\le r_n/\sqrt{n}} |(\mathbb{P}_n - P_0)S_p^2| \le \sup_{p:h(p,p_0)\le r_n/\sqrt{n}} r_n^2 |(\mathbb{P}_n - P_0)s_p^2|/n = o_{\mathbb{P}}(n^{-1})$$

for any sequence $(r_n)_{n\in\mathbb{N}}$ with $r_n \le b_n$. This verifies Condition (v) of Lemma F.7. Moreover, it follows from the envelope condition (in Condition (c)) that $\max_{1\le i\le n} D(X_i) = o_{\mathbb{P}}(n^{1/2})$. Therefore, we can choose a positive sequence $(c_n)_{n\in\mathbb{N}}$ with $c_n \to \infty$ such that $c_n^3 \max_{1\le i\le n} D(X_i) = o_{\mathbb{P}}(n^{1/2})$ or equivalently $\max_{1\le i\le n} D(X_i) = o_{\mathbb{P}}(n^{1/2}/r_n^3)$ for any diverging sequence $(r_n)_{n\in\mathbb{N}}$ with $r_n \le c_n$. This verifies Condition (iii) of Lemma F.7.

$\overline{\mathcal{D}}_\varepsilon$ is Donsker by Condition (c). $\overline{\mathcal{D}}_{\varepsilon,\Lambda} := \{\mathbf{\Lambda} s_p : s_p \in \overline{\mathcal{D}}_\varepsilon\} \subseteq \{f \in \Lambda : \|f\|_{L^2(P_0)} \le 1\}$ is Donsker because the linear hull Span(Λ) is finite dimensional. Therefore, $\Delta\overline{\mathcal{D}}_\varepsilon := \{s_p - \mathbf{\Lambda} s_p : s_p \in \overline{\mathcal{D}}_\varepsilon\}$ is also Donsker and hence $\mathbb{G}_n \rightsquigarrow W$ in $\ell^\infty(\Delta\overline{\mathcal{D}}_\varepsilon)$ where W is the isonormal Gaussian process. In view of the Skorohod-Dudley-Wichura theorem (van der Vaart and Wellner, 1996, Theorem 1.10.3) we represent $\mathbb{G}_n \rightsquigarrow W$ in $\ell^\infty(\Delta\overline{\mathcal{D}}_\varepsilon)$ by $\mathbb{G}_n \to_{a.s.} W$ in a suitable probability space. Therefore:

$$\sup_{d\in\Delta\overline{\mathcal{D}}_\varepsilon} |\mathbb{G}_n(d)| = \sup_{d\in\Delta\overline{\mathcal{D}}_\varepsilon} |W(d)| + \eta_n$$

where $\eta_n = o_{\mathbb{P}}(1)$. Again, we may choose a positive sequence $(d_n)_{n\in\mathbb{N}}$ with $d_n \to \infty$ sufficiently slowly that $d_n\eta_n = o_{\mathbb{P}}(1)$ and hence $r_n\eta_n = o_{\mathbb{P}}(1)$ for any positive sequence $(r_n)_{n\in\mathbb{N}}$ with $r_n \leq d_n$. The singleton $\{0\}$ is the only limit point of $\Delta\overline{\mathcal{D}}_\varepsilon$ as $\varepsilon \searrow 0$ because:

$$\begin{aligned}\sup\{\|d\|_{L^2(P_0)} : d \in \Delta\overline{\mathcal{D}}_\varepsilon\} &= \sup\{\|s_p - \mathbf{\Lambda} s_p\|_{L^2(P_0)} : h(p,p_0) \leq \epsilon\} \\ &\leq \sup\{\|R(p)\|_{L^2(\lambda)} : h(p,p_0) \leq \varepsilon\} \\ &\to 0 \quad (\text{as } \varepsilon \to 0)\end{aligned}$$

by DQM (condition (b)). Define $D(\varepsilon') = \sup\{\|d\|_{L^2(P_0)} : d \in \Delta\overline{\mathcal{D}}_{\varepsilon'}\}$ and $H(v) = \int_0^v (N(\Delta\overline{\mathcal{D}}_\varepsilon, u))^{1/2} du$ where $N(\Delta\overline{\mathcal{D}}_\varepsilon, u)$ is the covering number of $\Delta\overline{\mathcal{D}}_\varepsilon$ with respect to the intrinsic semimetric. Clearly $D(\varepsilon) \to 0$ as $\varepsilon \to 0$ and $H(v) \to 0$ as $v \to 0$. By Corollary 2.2.8 of van der Vaart and Wellner (1996):

$$\sup_{d\in\Delta\overline{\mathcal{D}}_\varepsilon} |W(d)| = O_{\mathbb{P}}\big(H(D(\varepsilon))\big) \to_{\mathbb{P}} 0 \quad (\text{as } \varepsilon \to 0)\,.$$

Taking $\varepsilon_n = n^{-1/4}$, we can choose a positive sequence $(e_n)_{n\in\mathbb{N}}$ with $e_n \to \infty$ as $n \to \infty$ sufficiently slowly that $e_n H(D(\varepsilon_n)) = o(1)$. For any sequence $(r_n)_{n\in\mathbb{N}}$ with $r_n \leq (d_n \vee e_n \vee n^{1/4})$ we then have:

$$\begin{aligned}\sup_{p:h(p,p_0)\leq r_n/\sqrt{n}} \mathbb{G}_n(S_p - \mathbf{\Lambda} S_p) &\leq \frac{r_n}{\sqrt{n}} \sup_{p:h(p,p_0)\leq r_n/\sqrt{n}} \mathbb{G}_n(s_p - \mathbf{\Lambda} s_p) \\ &\leq \frac{1}{\sqrt{n}} \times r_n \sup_{d\in\Delta\overline{\mathcal{D}}_{\varepsilon_n}} |\mathbb{G}_n(d)| = \frac{1}{\sqrt{n}} \times o_{\mathbb{P}}(1)\,.\end{aligned}$$

This verifies Condition (iv) of Lemma F.7. Take $r_n = (a_n \wedge b_n \wedge c_n \wedge d_n \wedge e_n \wedge \log n)$. ■

Proof of Proposition 4.3. We first show that there exists a positive sequence $(r_n)_{n\in\mathbb{N}}$ with $r_n \to \infty$ such that:

$$\sup_{\theta:\|g(\theta)\|\leq r_n/\sqrt{n}} \left| nL_n(\theta) - \left(-\frac{1}{2}(\mathbf{\Lambda}(\sqrt{n}g(\theta)) + Z_n)'\Omega^{-1}(\mathbf{\Lambda}(\sqrt{n}g(\theta)) + Z_n)\right)\right| = o_{\mathbb{P}}(1) \qquad (58)$$

for some sequence of random vectors $(Z_n)_{n\in\mathbb{N}}$ with $Z_n \rightsquigarrow N(0,\Omega)$.

In this proof we often abuse notation and use ρ_θ to denote $\rho(\cdot,\theta)$. Take n large enough that $n^{-1/4} \leq \varepsilon_0$. Let $\varepsilon_n = n^{1/4}$. I.i.d. data and Conditions (c)(e) imply that

$$\sup_{\theta:\|g(\theta)\|\leq\varepsilon_n} \|\mathbb{P}_n(\rho_\theta\rho_\theta') - \Omega\| = o_{\mathbb{P}}(1)$$

(van der Vaart and Wellner, 1996, Lemma 2.10.14). Therefore, we may choose a positive sequence $(a_n)_{n\in\mathbb{N}}$ with $a_n \to \infty$, $a_n = o(n^{1/4})$ such that $\sup_{\theta:\|g(\theta)\|\leq\varepsilon_n} a_n^2\|\mathbb{P}_n(\rho_\theta\rho_\theta') - \Omega\| = o_{\mathbb{P}}(1)$ and hence:

$$\sup_{\theta:\|g(\theta)\|\leq r_n/\sqrt{n}} \|\mathbb{P}_n(\rho_\theta\rho_\theta') - \Omega\| = o_{\mathbb{P}}(r_n^{-2}) \qquad (59)$$

holds for any sequence $(r_n)_{n\in\mathbb{N}}$ with $r_n \leq a_n$.

For any $\varepsilon \in (0, \varepsilon_0]$, under i.i.d. data and Condition (c) there exists a Gaussian process W defined on $\mathcal{R}_\varepsilon$ with $E[W(\rho)W(\bar\rho)'] = E[(\rho - E[\rho])(\bar\rho - E[\bar\rho])']$ for any $\rho, \bar\rho \in \mathcal{R}_\varepsilon$ such that $\mathbb{G}_n \rightsquigarrow W$ in $\ell^\infty(\mathcal{R}_\varepsilon)$. Fix any $\theta^* \in \Theta_I$ and set $Z_n = \mathbb{G}_n(\rho_{\theta^*})$ where $Z_n \rightsquigarrow N(0, \Omega)$ by condition (b). Representing weak convergence of $\mathbb{G}_n$ to W as almost sure convergence in a suitable probability (van der Vaart and Wellner, 1996, Theorem 1.10.3), we have:

$$\sup_{\rho \in \mathcal{R}_\varepsilon} |\mathbb{G}_n(\rho) - Z_n| = \sup_{\rho \in \mathcal{R}_\varepsilon} |W(\rho) - W(\rho_{\theta^*})| + \eta_n$$

where $\eta_n = o_{\mathbb{P}}(1)$. Choose a positive sequence $(b_n)_{n \in \mathbb{N}}$ with $b_n \to \infty$ slowly such that $b_n \eta_n = o_{\mathbb{P}}(1)$ and so $r_n \eta_n = o_{\mathbb{P}}(1)$ holds for any sequence $(r_n)_{n \in \mathbb{N}}$ with $r_n \leq b_n$. The intrinsic semimetric $d_I(\rho, \bar\rho) := (E[\|\rho - \bar\rho - E[\rho - \bar\rho]\|^2])^{1/2}$ satisfies:

$$\sup_{\theta \in \Theta_I^\delta} d_I(\rho_\theta, \rho_{\theta^*})^2 = \sup_{\theta \in \Theta_I^\delta} \big(E[\|\rho_\theta - \rho_{\theta^*}\|^2] - \|g(\theta)\|^2 \big) \to 0 \quad (\text{as } \delta \to 0)$$

by Condition (d). Let $D(\delta) = \sup_{\theta \in \Theta_I^\delta} d_I(\rho_\theta, \rho_{\theta^*})$ and let $H(v) = \int_0^v (N(\mathcal{R}_\varepsilon, u))^{1/2} du$ where $N(\mathcal{R}_\varepsilon, u)$ is the covering number of $\mathcal{R}_\varepsilon$ with respect to the intrinsic semimetric. Clearly $H(v) \to 0$ as $v \to 0$. Using Corollary 2.2.8 of van der Vaart and Wellner (1996) we can deduce:

$$\sup_{\rho \in \mathcal{R}_\varepsilon} |W(\rho) - W(\rho_{\theta^*})| = O_{\mathbb{P}}\big(H(D(\varepsilon))\big) \to_{\mathbb{P}} 0 \quad (\text{as } \varepsilon \to 0)\,.$$

Taking $\varepsilon_n = n^{-1/4}$, we can choose a positive sequence $(c_n)_{n \in \mathbb{N}}$ with $c_n \to \infty$ as $n \to \infty$ sufficiently slowly that $c_n H(D(\varepsilon_n)) = o(1)$. It follows that:

$$\sup_{\theta : \|g(\theta)\| \leq r_n/\sqrt{n}} |\sqrt{n}\mathbb{P}_n \rho_\theta - (\sqrt{n} g(\theta) + Z_n)| = o_{\mathbb{P}}(r_n^{-1})\,. \tag{60}$$

for any sequence $(r_n)_{n \in \mathbb{N}}$ with $r_n \leq (b_n \vee c_n \vee n^{1/4})$.

Condition (a) implies $\sup_{\theta : \|g(\theta)\| \leq d_n/\sqrt{n}} \|g(\theta) - \mathbf{\Lambda} g(\theta)\| = o(d_n/\sqrt{n})$ for any sequence (d_n) with $d_n \to \infty$, $d_n = o(\sqrt{n})$. Choose $d_n \to \infty$ slowly so that: $\sup_{\theta : \|g(\theta)\| \leq d_n/\sqrt{n}} \|g(\theta) - \mathbf{\Lambda} g(\theta)\| = o(1/\sqrt{n})$. Then choose another positive sequence $(e_n)_{n \in \mathbb{N}}$ with $e_n \to \infty$ such that:

$$\sup_{\theta : \|g(\theta)\| \leq d_n/\sqrt{n}} e_n \|g(\theta) - \mathbf{\Lambda} g(\theta)\| = o(1/\sqrt{n})$$

and hence, by the fact that $\mathbf{\Lambda}(\sqrt{n} g(\theta)) = \sqrt{n} \mathbf{\Lambda} g(\theta)$ (Hiriart-Urruty and Lemaréchal, 2001, p. 51), we obtain:

$$\sup_{\theta : \|g(\theta)\| \leq r_n/\sqrt{n}} \|\sqrt{n} g(\theta) - \mathbf{\Lambda}(\sqrt{n} g(\theta))\| = o(r_n^{-1}) \tag{61}$$

for any sequence $(r_n)_{n \in \mathbb{N}}$ with $r_n \to \infty$ slowly such that $r_n \leq (d_n \wedge e_n)$.

Taking $r_n = (a_n \wedge b_n \wedge c_n \wedge d_n \wedge e_n \wedge n^{1/4})$ and using (59), (60) and (61) yields:

$$nL_n(\theta) = -\frac{1}{2}(\sqrt{n}\mathbf{\Lambda} g(\theta) + Z_n + o_{\mathbb{P}}(r_n^{-1}))(\Omega^{-1} + o_{\mathbb{P}}(r_n^{-2}))(\sqrt{n}\mathbf{\Lambda} g(\theta) + Z_n + o_{\mathbb{P}}(r_n^{-1}))$$

where the $o_{\mathbb{P}}(r_n^{-1})$ and $o_{\mathbb{P}}(r_n^{-2})$ terms hold uniformly over $\Theta_{osn} := \{\theta : \|g(\theta)\| \leq r_n/\sqrt{n}\}$. This proves (58).

Let U be a unitary matrix as described and recall that $U^{-1} = U'$. The result follows from (58), by expanding the quadratic and using:

$$
\begin{aligned}
(\sqrt{n}\mathbf{\Lambda}g(\theta))'\Omega^{-1}(\sqrt{n}\mathbf{\Lambda}g(\theta)) &= (\sqrt{n}U\mathbf{\Lambda}(g(\theta)))'(U\Omega U')^{-1}(\sqrt{n}U\mathbf{\Lambda}(g(\theta))) \\
&= (\sqrt{n}[U\mathbf{\Lambda}g(\theta)]_1)'[(U\Omega U')^{-1}]_{11}(\sqrt{n}[U\mathbf{\Lambda}g(\theta)]_1) \\
&\equiv (\sqrt{n}\gamma(\theta))'(\sqrt{n}\gamma(\theta))
\end{aligned}
$$

where $[(U\Omega U')^{-1}]_{11}$ is the $d^* \times d^*$ upper-left block of $(U\Omega U')^{-1}$ and $[U\mathbf{\Lambda}g(\theta)]_1$ is the upper d^* subvector of $U\mathbf{\Lambda}g(\theta)$, and:

$$
\begin{aligned}
(\sqrt{n}\mathbf{\Lambda}g(\theta))'\Omega^{-1}Z_n &= (\sqrt{n}U\mathbf{\Lambda}g(\theta))'(U\Omega^{-1}Z_n) \\
&= (\sqrt{n}\gamma(\theta))'[(U\Omega U')^{-1}]_{11}^{-1/2}[U\Omega^{-1}Z_n]_1 \\
&\equiv -(\sqrt{n}\gamma(\theta))'\mathbb{V}_n
\end{aligned}
$$

where $[U\Omega^{-1}Z_n]_1$ is the upper d^* subvector of $U\Omega^{-1}Z_n$. ∎

Proof of Proposition 4.4. Follows by similar arguments to the proof of Proposition 4.3, noting that by condition (e) we may choose a positive sequence $(a_n)_{n\in\mathbb{N}}$ with $a_n \to \infty$ slowly such that $a_n^2\|\widehat{W} - \Omega^{-1}\| = o_{\mathbb{P}}(1)$. Therefore $\|\widehat{W} - \Omega^{-1}\| = o_{\mathbb{P}}(r_n^{-2})$ holds for any sequence $(r_n)_{n\in\mathbb{N}}$ with $r_n \to \infty$ such that $r_n = O(a_n)$. ∎

F.3 Proofs and Additional Lemmas for Appendix B

Proof of Lemma B.1. By (ii), there exists a positive sequence $(\eta_n)_{n\in\mathbb{N}}$ with $\eta_n = o(1)$ such that $\sup_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(w_{n,\alpha} < w_{\alpha,\mathbb{P}} - \eta_n) = o(1)$. Therefore:

$$
\begin{aligned}
\inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\Theta_I(\mathbb{P}) \subseteq \widehat{\Theta}_\alpha) &\geq \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\{\Theta_I(\mathbb{P}) \subseteq \widehat{\Theta}_\alpha\} \cap \{w_{n,\alpha} \geq w_{\alpha,\mathbb{P}} - \eta_n\}) \\
&= \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\{\sup\nolimits_{\theta\in\Theta_I(\mathbb{P})} Q_n(\theta) \leq w_{n,\alpha}\} \cap \{w_{n,\alpha} \geq w_{\alpha,\mathbb{P}} - \eta_n\}) \\
&\geq \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\{\sup\nolimits_{\theta\in\Theta_I(\mathbb{P})} Q_n(\theta) \leq w_{\alpha,\mathbb{P}} - \eta_n\} \cap \{w_{n,\alpha} \geq w_{\alpha,\mathbb{P}} - \eta_n\})\,.
\end{aligned}
$$

Since $\mathbb{P}(A \cap B) \geq 1 - \mathbb{P}(A^c) - \mathbb{P}(B^c)$, we have:

$$
\begin{aligned}
&\inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\{\sup_{\theta\in\Theta_I(\mathbb{P})} Q_n(\theta) \leq w_{\alpha,\mathbb{P}} - \eta_n\} \cap \{w_{n,\alpha} \geq w_{\alpha,\mathbb{P}} - \eta_n\}) \\
&\geq 1 - \sup_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\sup_{\theta\in\Theta_I(\mathbb{P})} Q_n(\theta) > w_{\alpha,\mathbb{P}} - \eta_n) - \sup_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(w_{n,\alpha} < w_{\alpha,\mathbb{P}} - \eta_n) \\
&= 1 - \sup_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\sup_{\theta\in\Theta_I(\mathbb{P})} Q_n(\theta) > w_{\alpha,\mathbb{P}} - \eta_n) - o(1) \\
&= 1 - (1 - \alpha + o(1)) - o(1)\,.
\end{aligned}
$$

where the second last line is by definition of η_n and the final line is by part (i).

To prove the case with equality, it suffices to show that $\inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\Theta_I(\mathbb{P}) \subseteq \widehat{\Theta}_\alpha) \leq \alpha + o(1)$. Since $w_{n,\alpha} = w_{\alpha,\mathbb{P}} + o_{\mathbb{P}}(1)$ uniformly for $\mathbb{P} \in \mathbf{P}$, there exists a positive sequence $(\eta_n)_{n\in\mathbb{N}}$ with $\eta_n = o(1)$ such that $\sup_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(|w_{n,\alpha} - w_{\alpha,\mathbb{P}}| > \eta_n) = o(1)$. Therefore:

$$
\begin{aligned}
\inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\Theta_I(\mathbb{P}) \subseteq \widehat{\Theta}_\alpha) &= \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\{\Theta_I(\mathbb{P}) \subseteq \widehat{\Theta}_\alpha\} \cap \{|w_{n,\alpha} - w_{\alpha,\mathbb{P}}| \leq \eta_n\}) + o(1) \\
&= \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\{\sup_{\theta\in\Theta_I(\mathbb{P})} Q_n(\theta) \leq w_{n,\alpha}\} \cap \{|w_{n,\alpha} - w_{\alpha,\mathbb{P}}| \leq \eta_n\}) + o(1) \\
&\leq \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\{\sup_{\theta\in\Theta_I(\mathbb{P})} Q_n(\theta) \leq w_{\alpha,\mathbb{P}} + \eta_n\} \cap \{|w_{n,\alpha} - w_{\alpha,\mathbb{P}}| \leq \eta_n\}) + o(1) \\
&\leq \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\sup_{\theta\in\Theta_I(\mathbb{P})} Q_n(\theta) \leq w_{\alpha,\mathbb{P}} + \eta_n) + o(1) \\
&= \alpha + o(1)
\end{aligned}
$$

where the final line is by part (i). ∎

Proof of Lemma B.2. Follows by similar arguments to the proof of Lemma B.1. ∎

In the following we often use the following expression (62) that is equivalent to equation (25) of Assumption B.2(i):

$$
\sup_{\theta\in\Theta_{osn}} \left| nL_n(\theta) - \ell_n - \frac{1}{2}\|\mathbb{V}_n\|^2 + \left(\frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 + f_{n,\perp}(\gamma_\perp(\theta))\right)\right| = o_{\mathbb{P}}(1) \tag{62}
$$

uniformly for $\mathbb{P} \in \mathbf{P}$.

Lemma F.8. *Let Assumptions B.1(i) and B.2 hold. Then:*

$$
\sup_{\theta\in\Theta_{osn}} \left|Q_n(\theta) - \left(\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 + 2f_{n,\perp}(\gamma_\perp(\theta))\right)\right| = o_{\mathbb{P}}(1) \tag{63}
$$

uniformly for $\mathbb{P} \in \mathbf{P}$. *If, in addition, Assumption B.5(i) holds, then:*

$$
\sup_{\theta\in\Theta_{osn}} \left|PQ_n(\Delta(\theta)) - f\left(\mathbb{V}_n - \sqrt{n}\gamma(\theta)\right)\right| = o_{\mathbb{P}}(1) \tag{64}
$$

uniformly for $\mathbb{P} \in \mathbf{P}$.

Proof of Lemma F.8. To show (63), using Assumptions B.1(i), B.2(i) (or expression (62)) and (ii) and completing the square, we obtain:

$$
\begin{aligned}
nL_n(\hat\theta) &= \sup_{\theta\in\Theta_{osn}} \left(\ell_n - \frac{1}{2}\|\sqrt{n}\gamma(\theta)\|^2 + (\sqrt{n}\gamma(\theta))'\mathbb{V}_n - f_{n,\perp}(\gamma_\perp(\theta))\right) + o_{\mathbb{P}}(1) \\
&= \sup_{\theta\in\Theta_{osn}} \left(\ell_n - \frac{1}{2}\|\sqrt{n}\gamma(\theta)\|^2 + (\sqrt{n}\gamma(\theta))'\mathbb{V}_n \right) + o_{\mathbb{P}}(1) \\
&= \ell_n + \frac{1}{2}\|\mathbb{V}_n\|^2 - \inf_{\theta\in\Theta_{osn}} \frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 + o_{\mathbb{P}}(1) \qquad (65)
\end{aligned}
$$

uniformly for $\mathbb{P}\in\mathbf{P}$. But observe that for any $\epsilon > 0$:

$$
\begin{aligned}
&\sup_{\mathbb{P}\in\mathbf{P}} \mathbb{P}\left(\inf_{\theta\in\Theta_{osn}} \|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 > \epsilon\right) \\
&\le \sup_{\mathbb{P}\in\mathbf{P}} \mathbb{P}\left(\left\{\inf_{\theta\in\Theta_{osn}} \|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 > \epsilon\right\} \cap \{\|\mathbb{V}_n\| < k_n\}\right) + \sup_{\mathbb{P}\in\mathbf{P}} \mathbb{P}\left(\|\mathbb{V}_n\| \ge k_n\right) \\
&= \sup_{\mathbb{P}\in\mathbf{P}} \mathbb{P}\left(\|\mathbb{V}_n\| \ge k_n\right) = o(1)
\end{aligned}
$$

by Assumption B.2(iii)(iv). This proves (63). Result (64) follows by Assumption B.5(i). ∎

Lemma F.9. *Let Assumptions B.1, B.2 and B.3 hold. Then:*

$$
\sup_z \left(\Pi_n\big(\{\theta : Q_n(\theta) \le z\} \,\big|\, \mathbf{X}_n\big) - F_{\chi^2_{d^*}}(z)\right) \le o_{\mathbb{P}}(1)
$$

uniformly for $\mathbb{P}\in\mathbf{P}$. If no $\mathbb{P}\in\mathbf{P}$ is singular, then:

$$
\sup_z \left|\Pi_n\big(\{\theta : Q_n(\theta) \le z\} \,\big|\, \mathbf{X}_n\big) - F_{\chi^2_{d^*}}(z)\right| = o_{\mathbb{P}}(1)\,.
$$

uniformly for $\mathbb{P}\in\mathbf{P}$

Proof of Lemma F.9. We only prove the case with singularity. The (simpler) case without singularity follows similarly.

By identical arguments to the proof of Lemma F.3, it is enough to characterize the large-sample behavior of $R_n(z)$ defined in equation (32) uniformly for $\mathbb{P}\in\mathbf{P}$. By Lemma F.8 and expression (62), there exist a positive sequence $(\eta_n)_{n\in\mathbb{N}}$ independent of z with $\eta_n = o(1)$ and a sequence of events $(\mathcal{A}_n)_{n\in\mathbb{N}} \subset \mathcal{F}$ with $\inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\mathcal{A}_n) = 1 - o(1)$ such that:

$$
\sup_{\theta\in\Theta_{osn}} \left|Q_n(\theta) - \left(\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 + 2f_{n,\perp}(\gamma_\perp(\theta))\right)\right| \le \eta_n
$$

$$
\sup_{\theta\in\Theta_{osn}} \left|nL_n(\theta) - \ell_n - \frac{1}{2}\|\mathbb{V}_n\|^2 + \left(\frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 + f_{n,\perp}(\gamma_\perp(\theta))\right)\right| \le \frac{\eta_n}{2}
$$

both hold on $\mathcal{A}_n$ for all $\mathbb{P} \in \mathbf{P}$. Also note that for any $z \in \mathbb{R}$ and any singular $\mathbb{P} \in \mathbf{P}$, we have

$$\left\{\theta \in \Theta_{osn} : \|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 + 2f_{n,\perp}(\gamma_\perp(\theta)) + \eta_n \le z\right\} \subseteq \left\{\theta \in \Theta_{osn} : \|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 + \eta_n \le z\right\}$$

because $f_{n,\perp} \ge 0$. Therefore, on $\mathcal{A}_n$ we have:

$$R_n(z) \le e^{\eta_n} \frac{\int_{\{\theta : \|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 \le z + \eta_n\} \cap \Theta_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 - f_{n,\perp}(\gamma_\perp(\theta))} \mathrm{d}\Pi(\theta)}{\int_{\Theta_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 - f_{n,\perp}(\gamma_\perp(\theta))} \mathrm{d}\Pi(\theta)}$$

uniformly in z for all $\mathbb{P} \in \mathbf{P}$.

Define $\Gamma_{osn} = \{\gamma(\theta) : \theta \in \Theta_{osn}\}$ and $\Gamma_{\perp,osn} = \{\gamma_\perp(\theta) : \theta \in \Theta_{osn}\}$ (if $\mathbb{P}$ is singular). The condition $\sup_{\mathbb{P} \in \mathbf{P}} \sup_{\theta \in \Theta_{osn}} \|(\gamma(\theta), \gamma_\perp(\theta))\| \to 0$ in Assumption B.2(i) implies that for all n sufficiently large we have $\Gamma_{osn} \times \Gamma_{\perp,osn} \subset B_\delta^*$ for all $\mathbb{P} \in \mathbf{P}$. By similar arguments to the proof of Lemma F.3, we use Assumption B.3(ii), a change of variables and Tonelli's theorem to obtain:

$$R_n(z) \le e^{\eta_n}(1 + \bar{\eta}_n) \frac{\int_{(\{\gamma : \|\sqrt{n}\gamma - \mathbb{V}_n\|^2 \le z + \eta_n) \cap \Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma - \mathbb{V}_n\|^2} \mathrm{d}\gamma}{\int_{\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma - \mathbb{V}_n\|^2} \mathrm{d}\gamma} \tag{66}$$

which holds uniformly in z for all $\mathbb{P} \in \mathbf{P}$ (on $\mathcal{A}_n$ with n sufficiently large). A second change of variables with $\sqrt{n}\gamma - \mathbb{V}_n \mapsto \kappa$ allows us to rewrite (66) as:

$$R_n(z) \le e^{\eta_n}(1 + \bar{\eta}_n) \frac{\nu_{d^*}(\{\gamma : \|\kappa\|^2 \le z + \eta_n\} \cap (K_{osn} - \mathbb{V}_n))}{\nu_{d^*}(K_{osn} - \mathbb{V}_n)}.$$

To complete the proof, it is enough to show that:

$$\sup_z \left| \frac{\nu_{d^*}(\{\kappa : \|\kappa\|^2 \le z + \eta_n\} \cap (K_{osn} - \mathbb{V}_n))}{\nu_{d^*}(K_{osn} - \mathbb{V}_n)} - \nu_{d^*}(\{\kappa : \|\kappa\|^2 \le z + \eta_n\} \cap (K_{osn} - \mathbb{V}_n)) \right| = o_{\mathbb{P}}(1) \tag{67}$$

$$\sup_z \left| \nu_{d^*}(\{\kappa : \|\kappa\|^2 \le z + \eta_n\} \cap (K_{osn} - \mathbb{V}_n)) - \nu_{d^*}(\{\kappa : \|\kappa\|^2 \le z\}) \right| = o_{\mathbb{P}}(1) \tag{68}$$

uniformly for $\mathbb{P} \in \mathbf{P}$.

Simple algebra shows that the left-hand side of (67) is bounded by $\nu_{d^*}((\mathbb{R}^{d^*} \setminus K_{osn}) - \mathbb{V}_n)$ which, in turn, is bounded by $\nu_{d^*}(B_{k_n}^c - \mathbb{V}_n)$ (cf. Assumption B.2(iii)). Now fix any $\epsilon > 0$ and notice that Assumption B.2(iii)(iv) and the fact that $d^* \le \bar{d}$ for all $\mathbb{P} \in \mathbf{P}$ implies $\sup_{\mathbb{P} \in \mathbf{P}} \mathbb{P}(\|\mathbb{V}_n\|^2 \le k_n) = o(1)$. Therefore:

$$\begin{aligned}
&\sup_{\mathbb{P} \in \mathbf{P}} \mathbb{P}(\nu_{d^*}(B_{k_n}^c - \mathbb{V}_n) > \epsilon) \\
&\le \sup_{\mathbb{P} \in \mathbf{P}} \mathbb{P}(\{\nu_{d^*}(B_{k_n}^c - \mathbb{V}_n) > \epsilon\} \cap \{\|\mathbb{V}_n\| \le k_n/2\}) + \sup_{\mathbb{P} \in \mathbf{P}} \mathbb{P}(\|\mathbb{V}_n\| > k_n/2) \\
&\le \sup_{\mathbb{P} \in \mathbf{P}} 1\!\!1\{\nu_{d^*}(B_{k_n/2}^c) > \epsilon\} + o(1) = o(1)
\end{aligned}$$

by Assumption B.2(iii).

Now consider (68). Simple algebra yields:

$$\begin{aligned}&\sup_z \left|\nu_{d^*}(\{\kappa : \|\kappa\|^2 \le z + \eta_n\} \cap (K_{osn} - \mathbb{V}_n)) - \nu_{d^*}(\{\kappa : \|\kappa\|^2 \le z + \eta_n\})\right| \\ &\le \nu_{d^*}((\mathbb{R}^{d^*} \setminus K_{osn}) - \mathbb{V}_n) = o_{\mathbb{P}}(1)\end{aligned}$$

uniformly for $\mathbb{P} \in \mathbf{P}$ by the preceding argument. Finally:

$$\begin{aligned}&\sup_z \left|\nu_{d^*}(\{\kappa : \|\kappa\|^2 \le z + \eta_n\}) - \nu_{d^*}(\{\kappa : \|\kappa\|^2 \le z\})\right| \\ &= \sup_z \Big(F_{\chi^2_{d^*}}(z + \eta_n) - F_{\chi^2_{d^*}}(z)\Big) = o(1)\end{aligned}$$

uniformly for $\mathbb{P} \in \mathbf{P}$ (the χ^2 cdfs are uniformly equicontinuous on $\mathbb{R}$). ■

Proof of Theorem B.1. We first prove part (i). To do so, we verify the conditions of Lemma B.1. As in the proof of Theorem 3.1, we may assume without loss of generality that $L_n(\hat\theta) = \sup_{\theta \in \Theta_{osn}} L_n(\theta) + o_{\mathbb{P}}(n^{-1})$ uniformly for $\mathbb{P} \in \mathbf{P}$. To verify condition (i) of Lemma B.1, by display (63) in Lemma F.8 we have $\sup_{\theta \in \Theta_I(\mathbb{P})} Q_n(\theta) = \|\mathbb{V}_n\|^2 + o_{\mathbb{P}}(1)$ uniformly for $\mathbb{P} \in \mathbf{P}$, where $\|\mathbb{V}_n\|^2 \overset{\mathbb{P}}{\rightsquigarrow} \chi^2_{d^*}$ for each $\mathbb{P} \in \mathbf{P}$ (since $\Sigma = I_{d^*}$). Condition (i) then follows by Assumption B.2(iv) and uniform equicontinuity of the χ^2 distribution functions.

To verify condition (ii) of Lemma B.1, by Lemma F.9 there exists a sequence of positive constants $(\eta_n)_{n\in\mathbb{N}}$ with $\eta_n = o(1)$ and a sequence of events $(\mathcal{A}_n)_{n\in\mathbb{N}} \subset \mathcal{F}$ with $\inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\mathcal{A}_n) = 1 - o(1)$ such that:

$$\sup_z \Big(\Pi_n\big(\{\theta : Q_n(\theta) \le z\} \,\big|\, \mathbf{X}_n\big) - F_{\chi^2_{d^*}}(z)\Big) \le \eta_n$$

holds on $\mathcal{A}_n$ for all $\mathbb{P} \in \mathbf{P}$. Substituting in $z = \xi^{post}_{n,\alpha}$:

$$\Pi_n\big(\{\theta : Q_n(\theta) \le \xi^{post}_{n,\alpha}\} \,\big|\, \mathbf{X}_n\big) - F_{\chi^2_{d^*}}(\xi^{post}_{n,\alpha}) = \alpha - F_{\chi^2_{d^*}}(\xi^{post}_{n,\alpha}) \le \eta_n$$

and hence:

$$F_{\chi^2_{d^*}}(\chi^2_{d^*,\alpha}) - F_{\chi^2_{d^*}}(\xi^{post}_{n,\alpha}) \le \eta_n \tag{69}$$

holds for all $\mathbb{P} \in \mathbf{P}$ on $\mathcal{A}_n$. The $\chi^2_{d^*}$ cdfs with $1 \le d^* \le \bar d < \infty$ are strictly monotone and their inverses are all uniformly continuous on a fixed neighborhood of $\chi^2_{d^*,\alpha}$. Hence by (69) there exists a positive sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ with $\varepsilon_n = o(1)$ such that:

$$\xi^{post}_{n,\alpha} \ge \chi^2_{d^*,\alpha} - \varepsilon_n$$

holds for all $\mathbb{P} \in \mathbf{P}$ on $\mathcal{A}_n$. Therefore, $\xi^{post}_{n,\alpha} \ge \chi^2_{d^*,\alpha} + o_{\mathbb{P}}(1)$ uniformly for $\mathbb{P} \in \mathbf{P}$. Combining this with Assumption B.4 we obtain $\xi^{mc}_{n,\alpha} \ge \chi^2_{d^*,\alpha} + o_{\mathbb{P}}(1)$, as required.

We now prove part (ii) of the theorem, again by verifying the conditions of Lemma B.1. Condition (i) is verified above. For condition (ii) of Lemma B.1, by Lemma F.9 if no $\mathbb{P} \in \mathbf{P}$ is singular, then we have:

$$\sup_z \left|\Pi_n\big(\{\theta : Q_n(\theta) \le z\} \,\big|\, \mathbf{X}_n\big) - F_{\chi^2_{d^*}}(z)\right| \le \eta_n$$

holds on a sequence of events $(\mathcal{A}_n)_{n\in\mathbb{N}} \subset \mathcal{F}$ with $\inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\mathcal{A}_n) = 1 - o(1)$ for some sequence of positive constants $(\eta_n)_{n\in\mathbb{N}}$ with $\eta_n = o(1)$, hence:

$$\sup_{\mathbb{P}\in\mathbf{P}} |F_{\chi^2_{d^*}}(\chi^2_{d^*,\alpha}) - F_{\chi^2_{d^*}}(\xi^{post}_{n,\alpha})| \le \eta_n$$

holds on $\mathcal{A}_n$. Arguing as above, we have that $\sup_{\mathbb{P}\in\mathbf{P}} |\xi^{post}_{n,\alpha} - \chi^2_{d^*,\alpha}| \le \epsilon_n$ on $\mathcal{A}_n$, for some positive sequence $(\epsilon_n)_{n\in\mathbb{N}}$ with $\epsilon_n = o(1)$. ∎

Lemma F.10. *Let Assumptions B.1, B.2, B.3, and B.5 hold. Then for any* $0 < \epsilon < (\overline{z} - \underline{z})/2$*:*

$$\sup_{z\in[\underline{z}+\epsilon,\overline{z}-\epsilon]} \left|\Pi_n\big(\{\theta : PQ_n(\Delta(\theta)) \le z\} \,\big|\, \mathbf{X}_n\big) - \mathbb{P}_Z(f(Z) \le z)\right| = o_{\mathbb{P}}(1)\,.$$

uniformly for $\mathbb{P} \in \mathbf{P}$*.*

Proof of Lemma F.10. By the same arguments as the proof of Lemma F.5, it suffices to characterize the large-sample behavior of $R_n(z)$ defined in (47) uniformly for $\mathbb{P} \in \mathbf{P}$. By Lemma F.8 and expression (62), there exist a positive sequence $(\eta_n)_{n\in\mathbb{N}}$ independent of z with $\eta_n = o(1)$ and a sequence of events $(\mathcal{A}_n)_{n\in\mathbb{N}} \subset \mathcal{F}$ with $\inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\mathcal{A}_n) = 1 - o(1)$ such that:

$$\sup_{\theta\in\Theta_{osn}} \left|PQ_n(\Delta(\theta)) - f(\mathbb{V}_n - \sqrt{n}\gamma(\theta))\right| \le \eta_n$$

$$\sup_{\theta\in\Theta_{osn}} \left|nL_n(\theta) - \ell_n - \frac{1}{2}\|\mathbb{V}_n\|^2 + \left(\frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 + f_{n,\perp}(\gamma_\perp(\theta))\right)\right| \le \frac{\eta_n}{2}$$

both hold on $\mathcal{A}_n$ for all $\mathbb{P} \in \mathbf{P}$. So on $\mathcal{A}_n$ we obtain: Therefore, wpa1 we have:

$$\begin{aligned}
&e^{-\eta_n} \frac{\int_{\{\theta: f(\mathbb{V}_n - \sqrt{n}\gamma(\theta)) \le z - \eta_n\}\cap\Theta_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 - f_{n,\perp}(\gamma_\perp(\theta))} \mathrm{d}\Pi(\theta)}{\int_{\Theta_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 - f_{n,\perp}(\gamma_\perp(\theta))} \mathrm{d}\Pi(\theta)} \\
&\le R_n(z) \le e^{\eta_n} \frac{\int_{\{\theta: f(\mathbb{V}_n - \sqrt{n}\gamma(\theta)) \le z + \eta_n\}\cap\Theta_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 - f_{n,\perp}(\gamma_\perp(\theta))} \mathrm{d}\Pi(\theta)}{\int_{\Theta_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma(\theta) - \mathbb{V}_n\|^2 - f_{n,\perp}(\gamma_\perp(\theta))} \mathrm{d}\Pi(\theta)}
\end{aligned}$$

uniformly in z for all $\mathbb{P} \in \mathbf{P}$. By similar arguments to the proof of Lemma F.9, we may use the change of variables $\theta \mapsto (\gamma(\theta), \gamma_\perp(\theta))$, smoothness of π_{Γ^*}, and Tonelli's theorem to rewrite the above system of inequalities as:

$$\begin{aligned}
&(1 - \bar{\eta}_n) e^{-\eta_n} \frac{\int_{\{\gamma: f(\mathbb{V}_n - \sqrt{n}\gamma) \le z - \eta_n\}\cap\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma - \mathbb{V}_n\|^2} \mathrm{d}\gamma}{\int_{\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma - \mathbb{V}_n\|^2} \mathrm{d}\gamma} \\
&\le R_n(z) \le (1 + \bar{\eta}_n) e^{\eta_n} \frac{\int_{\{\gamma: f(\mathbb{V}_n - \sqrt{n}\gamma) \le z + \eta_n\}\cap\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma - \mathbb{V}_n\|^2} \mathrm{d}\gamma}{\int_{\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma - \mathbb{V}_n\|^2} \mathrm{d}\gamma}
\end{aligned}$$

which holds uniformly in z for all $\mathbb{P} \in \mathbf{P}$ (on $\mathcal{A}_n$ with n sufficiently large), for some positive sequence $(\bar{\eta}_n)_{n\in\mathbb{N}}$ with $\bar{\eta}_n = o(1)$. Let $K_{osn} = \{\sqrt{n}\gamma : \gamma \in \Gamma_{osn}\}$. A second change of variables $\sqrt{n}\gamma - \mathbb{V}_n \mapsto \kappa$ yields:

$$\begin{aligned}(1-\bar{\eta}_n)e^{-\eta_n}&\frac{\int_{\{\kappa:f(\kappa)\le z-\eta_n\}\cap(\mathbb{V}_n-K_{osn})} e^{-\frac{1}{2}\|\kappa\|^2}\mathrm{d}\kappa}{\int_{(\mathbb{V}_n-K_{osn})} e^{-\frac{1}{2}\|\kappa\|^2}\mathrm{d}\kappa}\\ &\le R_n(z) \le (1+\bar{\eta}_n)e^{\eta_n}\frac{\int_{\{\kappa:f(\kappa)\le z+\eta_n\}\cap(\mathbb{V}_n-K_{osn})} e^{-\frac{1}{2}\|\kappa\|^2}\mathrm{d}\kappa}{\int_{(\mathbb{V}_n-K_{osn})} e^{-\frac{1}{2}\|\kappa\|^2}\mathrm{d}\kappa}\end{aligned}$$

which holds uniformly in z for all $\mathbb{P} \in \mathbf{P}$ (on $\mathcal{A}_n$ with n sufficiently large).

To complete the proof, it remains to show that:

$$\sup_{z\in[\underline{z}+\epsilon,\overline{z}-\epsilon]}\left|\frac{\nu_{d^*}(\{\kappa : f(\kappa)\le z\pm\eta_n\}\cap(\mathbb{V}_n-K_{osn}))}{\nu_{d^*}((\mathbb{V}_n-K_{osn}))} - \nu_{d^*}(\{\kappa : f(\kappa)\le z\})\right| = o_{\mathbb{P}}(1)$$

uniformly for $\mathbb{P} \in \mathbf{P}$. By the proof of Lemma F.9, it is enough to show that:

$$\sup_{\mathbb{P}\in\mathbf{P}}\sup_{z\in[\underline{z}+\epsilon,\overline{z}-\epsilon]}|\nu_{d^*}(\{\kappa : f(\kappa)\le z\pm\eta_n\}) - \nu_{d^*}(\{\kappa : f(\kappa)\le z\})| = o(1)\,.$$

This follows directly from Assumption B.5(ii). ∎

Proof of Theorem B.2. We verify the conditions of Lemma B.2. As in the proof of Theorem 3.3, we may assume without loss of generality that $L_n(\hat{\theta}) = \sup_{\theta\in\Theta_{osn}} L_n(\theta) + o_{\mathbb{P}}(n^{-1})$ uniformly for $\mathbb{P} \in \mathbf{P}$. To verify condition (i) of Lemma B.2, by display (64) in Lemma F.8 we have:

$$\sup_{\mu\in M_I(\mathbb{P})}\inf_{\theta\in\mu^{-1}(m)} Q_n(\theta) = f(\mathbb{V}_n) + o_{\mathbb{P}}(1)$$

uniformly for $\mathbb{P} \in \mathbf{P}$, where $\mathbb{V}_n \overset{\mathbb{P}}{\rightsquigarrow} N(0, I_{d^*})$ for each $\mathbb{P} \in \mathbf{P}$ (since $\Sigma = I_{d^*}$). Part (i) follows by Assumption B.5(ii)(iii).

To verify condition (ii) of Lemma B.2 with equality, take $\epsilon > 0$ such that $\epsilon < \inf_{\mathbb{P}\in\mathbf{P}} \xi_{\alpha,\mathbb{P}} - \underline{z}$ and $\epsilon < \overline{z} - \sup_{\mathbb{P}\in\mathbf{P}} \xi_{\alpha,\mathbb{P}}$. By Lemma F.10 there exists a sequence of positive constants $(\eta_n)_{n\in\mathbb{N}}$ with $\eta_n = o(1)$ and a sequence of events $(\mathcal{A}_n)_{n\in\mathbb{N}} \subset \mathcal{F}$ with $\inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\mathcal{A}_n) = 1 - o(1)$ such that:

$$\sup_{z\in[\underline{z}+\epsilon,\overline{z}-\epsilon]}\left|\Pi_n\big(\{\theta : PQ_n(\Delta(\theta))\le z\}\,\big|\,\mathbf{X}_n\big) - \mathbb{P}_Z(f(Z)\le z)\right| \le \eta_n$$

holds on $\mathcal{A}_n$ for all $\mathbb{P} \in \mathbf{P}$. Substituting in $z = \xi^{post}_{n,\alpha}$ (which is in $[\underline{z}+\epsilon, \overline{z}-\epsilon]$ for all $\mathbb{P} \in \mathbf{P}$, for all n sufficiently large by Assumption B.5(ii)), we can deduce that:

$$\left|\mathbb{P}_Z(f(Z)\le \xi_{\alpha,\mathbb{P}}) - \mathbb{P}_Z(f(Z)\le \xi^{post}_{n,\alpha})\right| \le \eta_n$$

holds for all $\mathbb{P} \in \mathbf{P}$ on $\mathcal{A}_n$, for all n sufficiently large. Uniform equicontinuity of the inverse of $z \mapsto \mathbb{P}_Z(f(Z) \le z)$ (Assumption B.5(ii)) implies that there exists a positive sequence $(\varepsilon_n)_{n\in\mathbb{N}}$

with $\varepsilon_n = o(1)$ such that

$$|\xi^{post}_{n,\alpha} - \xi_{\alpha,\mathbb{P}}| \leq \varepsilon_n$$

holds for all $\mathbb{P} \in \mathbf{P}$ on $\mathcal{A}_n$, for all n sufficiently large. Therefore, $\xi^{post}_{n,\alpha} - \xi_{\alpha,\mathbb{P}} = o_{\mathbb{P}}(1)$ uniformly for $\mathbb{P} \in \mathbf{P}$. The result follows by Assumption B.6. ■

F.4 Proofs for Appendix D

Proof of Theorem D.1. We first derive the asymptotic distribution of $\sup_{\theta\in\Theta_I} Q_n(\theta)$ under $\mathrm{P}_{n,a}$. By similar arguments to the proof of Theorem 3.1, we have:

$$\sup_{\theta\in\Theta_I} Q_n(\theta) = \|\mathbb{V}_n\|^2 + o_{\mathrm{P}_{n,a}}(1) \overset{\mathrm{P}_{n,a}}{\rightsquigarrow} \chi^2_{d^*}(a'a)\,.$$

Identical arguments to the proof of Lemma 3.1 yield:

$$\sup_z |\Pi_n(\{\theta : Q_n(\theta) \leq z\}|\mathbf{X}_n) - F_{\chi^2_{d^*}}(z)| = o_{\mathrm{P}_{n,a}}(1)\,.$$

Therefore, $\xi^{mc}_{n,\alpha} = \chi^2_{d^*,\alpha} + o_{\mathrm{P}_{n,a}}(1)$ and we obtain:

$$\mathrm{P}_{n,a}(\Theta_I \subseteq \widehat{\Theta}_\alpha) = \Pr(\chi^2_{d^*}(a'a) \leq \chi^2_{d^*,\alpha}) + o(1)$$

as required. ■

Proof of Theorem D.2. Similar arguments to the proof of Theorem 3.3, uniformly for $\theta \in \Theta_I$ we have:

$$PQ_n(\Delta(\theta)) = 2nL_n(\hat\theta) - 2nPL_n(\theta) = f(\mathbb{V}_n) + o_{\mathrm{P}_{n,a}}(1)\,.$$

hence:

$$\sup_{\theta\in\Theta_I} PQ_n(\Delta(\theta)) \overset{\mathrm{P}_{n,a}}{\rightsquigarrow} f(Z + a)$$

where $Z \sim N(0, I_{d^*})$. Identical arguments to the proof of Lemma F.5 yield:

$$\sup_{z\in S^{-\epsilon}} \left|\Pi_n\big(\{\theta : PQ_n(\Delta(\theta)) \leq z\} \,\big|\, \mathbf{X}_n\big) - \mathbb{P}_{Z|\mathbf{X}_n}\big(f(Z) \leq z\big)\right| = o_{\mathbb{P}}(1)$$

for an open set $S^{-\epsilon}$ containing z_α. Therefore, $\xi^{mc,p}_{n,\alpha} = z_\alpha + o_{\mathrm{P}_{n,a}}(1)$ and we obtain:

$$\mathrm{P}_{n,a}(M_I \subseteq \widehat{M}_\alpha) = \mathbb{P}_Z(f(Z + a) \leq z_\alpha) + o(1)$$

as required. ■

F.5 Proofs for Appendix E

Proof of Lemma E.1. By equations (30) and (31) in the proof of Lemma 3.1, it suffices to characterize the large-sample behavior of:

$$R_n(z) := \frac{\int_{\{\theta:Q_n(\theta)\le z\}\cap\Theta_{osn}} e^{-\frac{1}{2}Q_n(\theta)}\mathrm{d}\Pi(\theta)}{\int_{\Theta_{osn}} e^{-\frac{1}{2}Q_n(\theta)}\mathrm{d}\Pi(\theta)} .$$

By Assumption E.2(i), there exists a positive sequence $(\eta_n)_{n\in\mathbb{N}}$ with $\eta_n = o(1)$ such that: $(1-\eta_n)h(\gamma(\theta)-\hat\gamma_n) \le \frac{a_n}{2}Q_n(\theta) \le (1+\eta_n)h(\gamma(\theta)-\hat\gamma_n)$ holds uniformly over Θ_{osn}. Therefore:

$$\begin{aligned}
&\frac{\int_{\{\theta:2a_n^{-1}(1+\eta_n)h(\gamma(\theta)-\hat\gamma_n)\le z\}\cap\Theta_{osn}} e^{-a_n^{-1}(1+\eta_n)h(\gamma(\theta)-\hat\gamma_n)}\mathrm{d}\Pi(\theta)}{\int_{\Theta_{osn}} e^{-a_n^{-1}(1-\eta_n)h(\gamma(\theta)-\hat\gamma_n)}\mathrm{d}\Pi(\theta)} \\
&\le R_n(z) \le \frac{\int_{\{\theta:2a_n^{-1}(1-\eta_n)h(\gamma(\theta)-\hat\gamma_n)\le z\}\cap\Theta_{osn}} e^{-a_n^{-1}(1-\eta_n)h(\gamma(\theta)-\hat\gamma_n)}\mathrm{d}\Pi(\theta)}{\int_{\Theta_{osn}} e^{-a_n^{-1}(1+\eta_n)h(\gamma(\theta)-\hat\gamma_n)}\mathrm{d}\Pi(\theta)} .
\end{aligned}$$

By similar arguments to the proof of Lemma 3.1, under Assumption 3.3 there exists a positive sequence $(\bar\eta_n)_{n\in\mathbb{N}}$ with $\bar\eta_n = o(1)$ such that for all n sufficiently large we have:

$$\begin{aligned}
&(1-\bar\eta_n)\frac{\int_{\{\gamma:2a_n^{-1}(1+\eta_n)h(\gamma-\hat\gamma_n)\le z\}\cap\Gamma_{osn}} e^{-a_n^{-1}(1+\eta_n)h(\gamma-\hat\gamma_n)}\mathrm{d}\gamma}{\int_{\Gamma_{osn}} e^{-a_n^{-1}(1-\eta_n)h(\gamma-\hat\gamma_n)}\mathrm{d}\gamma} \\
&\le R_n(z) \le (1+\bar\eta_n)\frac{\int_{\{\gamma:2a_n^{-1}(1-\eta_n)h(\gamma-\hat\gamma_n)\le z\}\cap\Gamma_{osn}} e^{-a_n^{-1}(1-\eta_n)h(\gamma-\hat\gamma_n)}\mathrm{d}\gamma}{\int_{\Gamma_{osn}} e^{-a_n^{-1}(1+\eta_n)h(\gamma-\hat\gamma_n)}\mathrm{d}\gamma} .
\end{aligned}$$

under the change of variables $\theta \mapsto \gamma(\theta)$, where $\Gamma_{osn} = \{\gamma(\theta) : \theta \in \Theta_{osn}\}$.

Assumption E.2(ii) implies that:

$$a_n^{-1}(1\pm\eta_n)h(\gamma-\hat\gamma_n) = h\Big(a_n^{-r_1}(1\pm\eta_n)^{r_1}(\gamma_1-\hat\gamma_{n,1}),\ldots,a_n^{-r_{d^*}}(1\pm\eta_n)^{r_{d^*}}(\gamma_{d^*}-\hat\gamma_{n,d^*})\Big) .$$

Using a change of variables:

$$\gamma \mapsto \kappa_\pm(\gamma) = \big(a_n^{-r_1}(1\pm\eta_n)^{r_1}(\gamma_1-\hat\gamma_{n,1}),\ldots,a_n^{-r_{d^*}}(1\pm\eta_n)^{r_{d^*}}(\gamma_{d^*}-\hat\gamma_{n,d^*})\big)$$

(with choice of sign as appropriate) and setting $r^* = r_1 + \ldots + r_{d^*}$, we obtain:

$$\begin{aligned}
&(1-\bar\eta_n)\frac{(1-\eta_n)^{r^*}}{(1+\eta_n)^{r^*}}\frac{\int_{\{\kappa:2h(\kappa)\le z\}\cap K_{osn}^+} e^{-h(\kappa)}\mathrm{d}\kappa}{\int e^{-h(\kappa)}\mathrm{d}\kappa} \\
&\le R_n(z) \le (1+\bar\eta_n)\frac{(1+\eta_n)^{r^*}}{(1-\eta_n)^{r^*}}\frac{\int_{\{\kappa:2h(\kappa)\le z\}} e^{-h(\kappa)}\mathrm{d}\kappa}{\int_{K_{osn}^+} e^{-h(\kappa)}\mathrm{d}\kappa}
\end{aligned} \tag{70}$$

uniformly in z, where $K_{osn}^+ = \{\kappa_+(\gamma) : \gamma \in \Gamma_{osn}\}$.

We can use a change of variables for $\kappa \mapsto t = 2h(\kappa)$ to obtain:

$$\int_{\{\kappa:h(\kappa)\leq z/2\}} e^{-h(\kappa)}\mathrm{d}\kappa = 2^{-r^*}\mathrm{V}(S)\int_0^z e^{-t/2}t^{r^*-1}dt \quad \int e^{-h(\kappa)}\mathrm{d}\kappa = 2^{-r^*}\mathrm{V}(S)\int_0^\infty e^{-t/2}t^{r^*-1}dt \tag{71}$$

where $\mathrm{V}(S)$ denotes the volume of the set $S = \{\kappa : h(\kappa) = 1\}$.

For the remaining integrals over K_{osn}^+ we first fix any $\omega \in \Omega$ so that $K_{osn}^+(\omega)$ becomes a deterministic sequence of sets. Let $C_n(\omega) = K_{osn}^+(\omega) \cap B_{k_n}$. Assumption E.2(iii) gives $\mathbb{R}^{d^*} = \overline{\cup_{n\geq 1}C_n(\omega)}$ for almost every ω. Now clearly:

$$\int e^{-h(\kappa)}\mathrm{d}\kappa \geq \int_{K_{osn}^+(\omega)} e^{-h(\kappa)}\mathrm{d}\kappa \geq \int 1\!\mathrm{l}\{\kappa \in C_n(\omega)\}e^{-h(\kappa)}\mathrm{d}\kappa \to \int e^{-h(\kappa)}\,\mathrm{d}\kappa$$

(by dominated convergence) for almost every ω. Therefore:

$$\int_{K_{osn}^+} e^{-h(\kappa)}\mathrm{d}\kappa \to_p 2^{-r^*}\mathrm{V}(S)\int_0^\infty e^{-t/2}t^{r^*-1}dt\,. \tag{72}$$

We may similarly deduce that:

$$\sup_z \left|\int_{\{\kappa:h(\kappa)\leq 2z\}\cap K_{osn}^+} e^{-h(\kappa)}\mathrm{d}\kappa - 2^{-r^*}\mathrm{V}(S)\int_0^z e^{-t/2}t^{r^*-1}dt\right| \to_p 0\,. \tag{73}$$

The result follows by substituting (71), (72), and (73) into (70). ∎

Proof of Theorem E.1. We verify the conditions of Lemma 2.1.

Lemma E.1 shows that the posterior distribution of the QLR is asymptotically $F_\Gamma = \Gamma(r^*, 1/2)$, and hence $\xi_{n,\alpha}^{post} = z_\alpha + o_{\mathbb{P}}(1)$, where z_α denotes the α quantile of the F_Γ. By Assumption $\sup_{\theta\in\Theta_I} Q_n(\theta) \rightsquigarrow F_\Gamma$. Then:

$$\xi_{n,\alpha}^{mc} = z_\alpha + (\xi_{n,\alpha}^{post} - z_\alpha) + (\xi_{n,\alpha}^{mc} - \xi_{n,\alpha}^{post}) = z_\alpha + o_{\mathbb{P}}(1)$$

where the final equality is by Assumption 3.4. ∎