MCMC Confidence Sets for Identified Sets*

Xiaohong Chen† Timothy M. Christensen‡ Elie Tamer§

First draft: August 2015; Revised May 3, 2016

Abstract

In complicated/nonlinear parametric models, it is hard to determine whether a parameter of interest is formally 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 directly from Markov Chain Monte Carlo (MCMC) simulations of 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 MCMC 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 MCMC 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.

*We are very grateful to S. Bonhomme and L. P. Hansen for insightful suggestions, and to K. O'Hara for outstanding research assistance. We also thank K. Evdokimov, R. Gallant, B. Honoré, M. Keane, Y. Kitamura, U. Müller, C. Sims and participants at the ESWC meetings in Montreal, the September 2015 "Big Data Big Methods" Cambridge-INET conference, and seminar audiences at University of Chicago, Princeton, Yale, Vanderbilt and Cornell 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, Harvard University. E-mail address: elietamer@fas.harvard.edu

1 Introduction

In complicated (nonlinear) structural models, it is typically difficult to rigorously verify that the model parameters are 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. This 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 are 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 Markov Chain Monte Carlo (MCMC) 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 generally rely on the need for choosing 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 MCMC 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 function (or a transformation of it) with a given prior (over the parameter space Θ). We draw

[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.

a MCMC 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 Bayesian MCMC computations (see, e.g., Liu (2004) and the references therein). There are many different MCMC samplers one could use for fast simulation from a (quasi) posterior 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.[3] 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 MCMC 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.[4] 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 C 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 MCMC criterion chains (from the posterior), which are sufficient to construct CSs for the IdS. In addition, we show in Appendix B that our MCMC CSs are uniformly valid over DGPs that include both partially- and point-identified models.

Our MCMC CSs are equivalent to Bayesian credible sets based on inverting a LR statistic in point-identified likelihood models, which are very closely related to Bayesian highest posterior

[3]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.

[4]CHT and Romano and Shaikh (2010) use subsampling based methods to estimate the quantile of the maximal (over the IdS) QLR statistic, we instead estimate it using the quantile of simulated QLR chains from a quasi-posterior and hence our need for the new Bernstein-von Mises type results under partial identification.

density (HPD) credible regions. More generally, for point-identified likelihood or moment-based models our MCMC 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 rate asymptotically normally estimable.[5] Note that our MCMC 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 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 MCMC CS and CH's MCMC 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 MCMC 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 strongly 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) mapping. Second, a prior is placed on the reduced-form parameter ϕ, and third, an existing Bernstein-von Mises theorem stating the asymptotic normality of the posterior distribution for ϕ is assumed to hold. Finally, the known mapping between the reduced-form and the structural parameters is inverted which, by step 3, guarantees

[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 C 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.

correct coverage for the IdS Θ_I in large samples. Broadly, all these papers focus on a class of separable models with a particular structure that allows one to relate a reduced-form parameter to the structural parameters.

Our MCMC 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 MCMC to estimate consistently the relevant quantile of this statistic. Importantly, our local reparametrization is a proof device only, and so one does not need to know this reparametrization or its dimension explicitly for the actual construction of our proposed MCMC CSs for Θ_I. Our more general Bernstein-von Mises theorem for the posterior of QLR in Appendix C 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 at a root-n rate and asymptotically normal). In particular, and in comparison to all the existing other Bayesian works on set inference, 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 profile QLR statistics. Further, our methods are shown to be uniformly valid over a class of DGPs that include both partially-identified and point-identified models (see Appendix B).

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.[9] Our proposed methods are computationally attractive and typically have asymptotically correct coverage, but

[9]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 discrete binary 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.

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 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). Appendix C presents a more general Bernstein-von Mises type theorems which show that the limiting distributions of the posterior QLR is as a gamma distribution with scale parameter 2 but a unknown shape parameter. Appendix D 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.[10] Consider a population objective function $L : \Theta \to \mathbb{R}$ where L can be a log likelihood function 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

[10]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 $\mathscr{B}(\Theta)$.

set (IdS) is defined as follows:

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

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 (asymptotic) frequentist guarantees.

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[11] $p(.;\theta)$, 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,[12] or an optimally-weighted GMM objective 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

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

[12] 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).

the IdS or functions of the IdS with a prespecified probability (in repeated samples) as sample size gets large.

Our main construction is based on Monte Carlo simulation methods using a well defined quasi posterior that is constructed as follows. Given L_n and a prior measure Π on $(\Theta, \mathscr{B}(\Theta))$ (such as a flat prior), define the quasi-posterior Π_n for θ given $\mathbf{X}_n$:

$$\Pi_n(A \,|\, \mathbf{X}_n) = \frac{\int_A e^{nL_n(\theta)} d\Pi(\theta)}{\int_\Theta e^{nL_n(\theta)} d\Pi(\theta)} \tag{4}$$

for $A \in \mathscr{B}(\Theta)$.

We first describe our computational procedure for covering the IdS Θ_I. We then describe procedures for covering a function of Θ_I, such as a subvector. We also describe an extremely simple procedure for covering the identified set for a scalar subvector in certain situations.

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 a 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:

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

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.

Chernozhukov et al. (2007) 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 (and asymptotic approximation) based methods were 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 density Π. This prior density 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 when 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 following lemma presents high-level conditions under which *any* $100\alpha\%$ criterion-based CS for Θ_I is asymptotically valid. Similar results appear in Chernozhukov et al. (2007) and Romano and Shaikh (2010).

Lemma 2.1. *Let (i) $\sup_{\theta\in\Theta_I} Q_n(\theta) \rightsquigarrow W$ where W is a random variable whose probability distribution is tight and 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 are shown to be valid by verifying parts (i) and (ii) with $w_{n,\alpha} = \xi^{mc}_{n,\alpha}$. To verify part (ii), we establish a new Bernstein-von Mises (BvM) result for the posterior distribution of the QLR under loss of identifiability for likelihood and GMM models (see Section 4 for primitive sufficient conditions). Therefore, although our Procedure 1 above appears Bayesian,[13] we show that $\widehat{\Theta}_\alpha$ has correct frequentist coverage.

2.2 Confidence sets for functions of the identified set

In many problems, it may be of interest to provide a confidence set 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:

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

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, $\widehat{M}^{proj}_\alpha$ is typically conservative, and could be very conservative when the dimension of μ is small relative to the dimension of θ. Our simulations below show that $\widehat{M}^{proj}_\alpha$ can be very conservative even in reasonably low-dimensional parametric models.

In the following we propose CSs $\widehat{M}_\alpha$ for M_I that 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., $\mu^{-1}(m) = \{\theta \in \Theta :$

[13] 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.

$\mu(\theta) = m\}$ 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: Exact CSs for functions of the identified set]

1. Draw a MCMC chain $\theta^1, \ldots, \theta^B$ from the quasi-posterior distribution Π_n in (4).
2. Calculate the $(1-\alpha)$ quantile of $\{\inf_{m \in \mu(\Delta(\theta^b))} \sup_{\theta \in \mu^{-1}(m)} L_n(\theta) : b = 1, \ldots, B\}$ 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. For *point-identified root-n estimable parameters* θ, Chernozhukov and Hong (2003) show that an asymptotically valid CS for a scalar subvector μ (of θ) can be obtained by taking a MCMC draw $\theta^1, \ldots, \theta^B$ then computing the upper and lower $100(1-\alpha)/2$ percentiles of $\mu(\theta^1), \ldots, \mu(\theta^B)$. However, this approach is no longer valid under partial identification of θ and has particularly poor coverage, 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 result 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 probability distribution is tight and 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 $w_{n,\alpha} = \xi_{n,\alpha}^{mc,p}$. To verify part (ii), we derive a new BvM result for the posterior of the profile QLR under loss of identifiability for likelihood and GMM objective functions (see Section 4 for sufficient conditions). Therefore, although our Procedure 2 above appears Bayesian,[14] we show that $\widehat{M}_\alpha$ has correct frequentist coverage.

2.3 A simple but slightly conservative CS for scalar subvectors

The CSs $\widehat{M}_\alpha$ described in Procedure 2 and Remark 2 above have asymptotically exact coverage for M_I under sufficient conditions and are valid for general M_I in general partially identified models. 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 is slightly conservative (whereas $\widehat{M}_\alpha$ is asymptotically exact), but it's coverage is typically 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 $L_n(\hat{\theta}) \geq \sup_{\theta \in \Theta} L_n(\theta) + o_{\mathbb{P}}(n^{-1})$.

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

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

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

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 whenever 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. This allows for computationally simple construction using repeated evaluations on a scalar grid. Unlike $\widehat{M}_\alpha$, the CS $\widehat{M}_\alpha^\chi$ has no Bayesian justification, is typically asymptotically conservative and is only valid for scalar functions of Θ_I in a certain class of models (see Section 3.3). Nevertheless it is extremely simple to implement and perform favorably in simulations.

To get an idea of the degree of conservativeness of $\widehat{M}_\alpha^\chi$, consider the class of models for which $\widehat{M}_\alpha^\chi$ is valid (see Section 3.3). Figure 1 plots the asymptotic coverage of $\widehat{M}_\alpha$ and $\widehat{M}_\alpha^\chi$ against nominal coverage for models in this class for which $\widehat{M}_\alpha^\chi$ is most conservative. We refer to as 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 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.

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,

[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 $N(\ell^b, cI)$ with c set so that the acceptance rate is approximately one third.

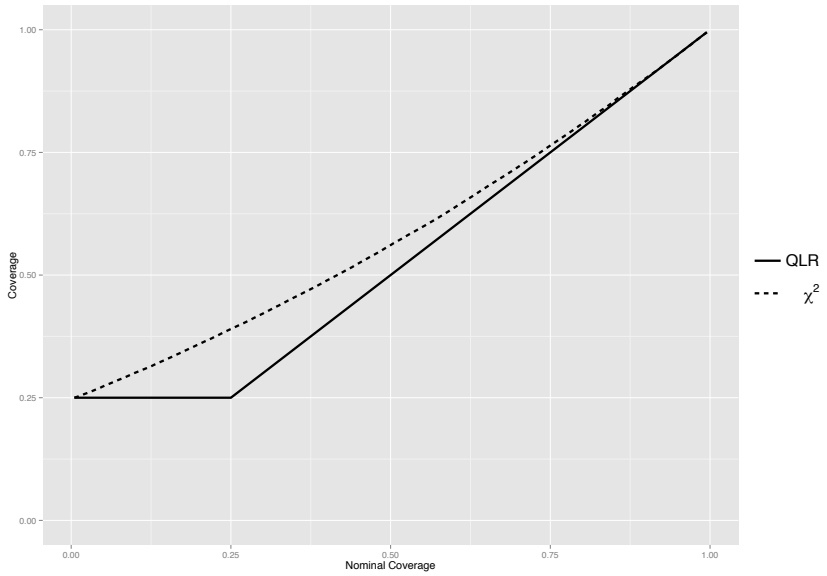

Figure 1: Comparison of asymptotic coverage of $\widehat{M}_\alpha$ (QLR – solid kinked line) of $\widehat{M}^\chi_\alpha$ (χ^2 – dashed curved line) with their nominal coverage for the class of models for which $\widehat{M}^\chi_\alpha$ is valid but most conservative (see Section 3.3).

the parameter chain may not settle down but the criterion chain is stable. We replicate each experiment 5000 times.

2.4.1 Missing data

Here we consider the simplest but most insightful case when we observe $\{(D_i, Y_i D_i)\}_{i=1}^n$ with both the outcome variable Y_i and the selection variable D_i are binary variables. The main parameter of interest is (usually) the mean $\mu = \mathbb{E}[Y_i]$. Without further assumptions, it is clear that μ is not point identified when $\Pr(D_i = 0) > 0$. 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 *reduced form parameters* that can be consistently estimated from the data. The reduced form parameters are functions of the structural parameter θ. The likelihood of the i-th observation $(D_i, Y_i D_i) = (d, yd)$ is

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

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

$$E\Big[\mathbb{1}\big((D_i, Y_i D_i) = (1,1)\big) - \kappa_{11}(\theta)\Big] = 0$$
$$E\Big[\mathbb{1}\big(D_i = 0\big) - \kappa_{00}(\theta)\Big] = 0\,.$$

Consider the model parameterized by $\theta = (\mu, \beta, \rho)$ where $\mu = \mathbb{E}[Y_i]$, $\beta = \Pr(Y_i = 1 | D_i = 0)$, and $\rho = \Pr(D_i = 1)$. The parameter space is

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

The parameter $\theta \in \Theta$ is related to the reduced form parameters via the following equalities:

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

The identified set for θ is:

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

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 $\mu = E[Y] \in M_I$:

$$\mu \leq E[Y|D=1]P(D=1) + P(D=0)$$
$$\mu \geq E[Y|D=1]P(D=1)\,.$$

Generally, all moment inequality models (with finitely many moment inequalities) can be written as moment equality model with added parameters with a known sign (see Subsection 4.2.1). The moment equality approach allows us to obtain a quasi posterior based on an optimal objective function.

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 5).

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 $\widehat{\Theta}_\alpha$ for different parameterizations of the model and different nominal coverage probabilities with a flat prior. Throughout, for set coverage, we use Procedure 1. 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 properties 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 are expected to be conservative (see Theorem 3.2 below for this case), 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). In Table 2 we provide results for the case where we use a curved prior. 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 using Procedure 1 with a continuously-updated GMM objective function (rather than a likelihood) are presented in Table 3. As can be seen, the results look similar to those for 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 2 plots the MCMC chain for a sample with $\rho = .8$. The chain is stable for ρ (which 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 follow Chernozhukov and Hong (2003) and construct a confidence interval for μ as follows: given the MCMC chain $\theta^1, \ldots, \theta^B$ for θ, one picks off the subvector chain $\mu^1, \ldots, \mu^B$ for μ, and then constructs a CS for M_I by taking the upper and lower $100(1-\alpha)/2$ percentiles of $\mu^1, \ldots, \mu^B$. Chernozhukov and Hong (2003) show this approach is valid in likelihood and optimally weighted GMM problems when θ (and hence μ) are

	$\rho = 0.20$	$\rho = 0.80$	$\rho = 0.95$	$\rho = 0.99$	$\rho = 1.00$
			$n = 100$		
$\alpha = 0.90$	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 for $\widehat{\Theta}_\alpha$ using Procedure 1 with 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 for $\widehat{\Theta}_\alpha$ using Procedure 1 with a likelihood for L_n and a curved prior on Θ.

	$\rho = 0.20$	$\rho = 0.80$	$\rho = 0.95$	$\rho = 0.99$	$\rho = 1.00$
			$n = 100$		
$\alpha = 0.90$	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 for $\widehat{\Theta}_\alpha$ using Procedure 1 with 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 for projection confidence sets $\widehat{M}_\alpha^{proj}$ of M_I with a likelihood for L_n and a flat prior on Θ.

point-identified and root-n asymptotically normally estimable. However, this simple percentile approach fails badly under partial identification. Table 5 reports the MC coverage probabilities of the percentile CSs for μ. It is clear that these CSs dramatically undercover, 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 in partially-identified models which takes *quantiles of the parameter chain* severely undercovers 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 2 shows the MCMC chain for $Q_n(\theta)$ is stable. Figure 6 (in Appendix A), which is computed from the draws for the structural parameter presented in Figure 2, shows that the MCMC chain for the reduced-form probabilities is also stable. In Table 6, we provide coverage results $\widehat{M}_\alpha$ with a flat prior using our Procedure 2. Theoretically, we show below (see Theorem 3.3) that the coverage probabilities of $\widehat{M}_\alpha$ 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 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

Figure 2: 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 for CS of μ taking percentiles of parameter chain, flat prior on Θ. Chernozhukov and Hong (2003) show this procedure is valid for point-identified parameters (which corresponds to $\rho = 1$).

	$\rho = 0.20$	$\rho = 0.80$	$\rho = 0.95$	$\rho = 0.99$	$\rho = 1.00$
			$n = 100$		
$\alpha = 0.90$	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 for $\widehat{M}_\alpha$ of M_I using Procedure 2 with 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 for $\widehat{M}_\alpha^\chi$ of M_I using Procedure 3 with 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 for $\widehat{M}_\alpha^\chi$ of M_I using Procedure 3 with a CU-GMM for L_n.

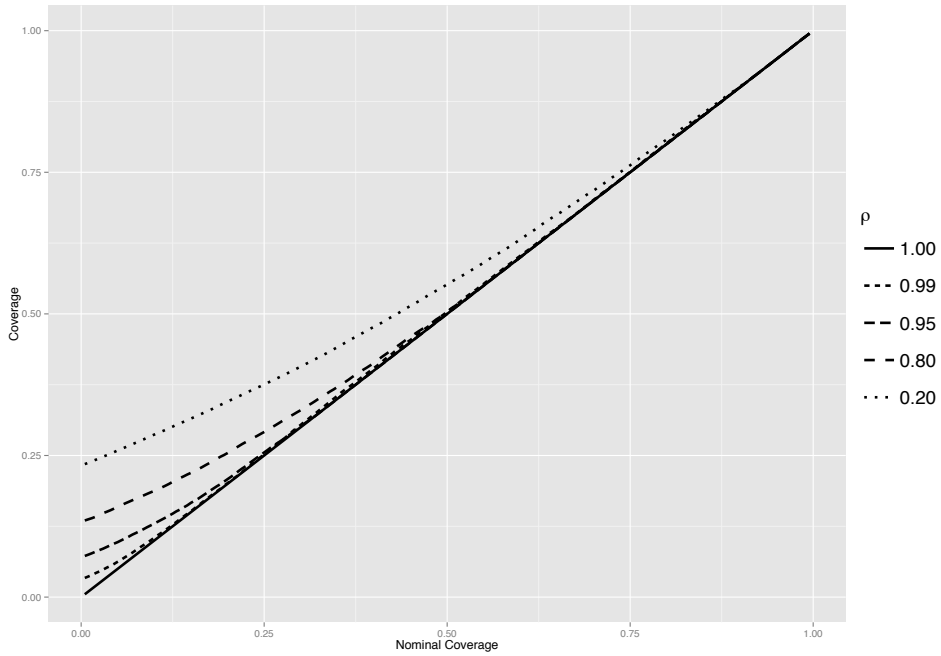

Figure 3: Comparison of asymptotic coverage of $\widehat{M}_\alpha^\chi$ of M_I for different ρ values.

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 $-\beta_i \leq \epsilon_i \leq -\beta_i - \Delta_i$, $i = 1, 2$, the game has two equilibria: for given values of the epsilons in this region, the model predicts $(1, 0)$ *and* $(0, 1)$. Let $D_{a_1 a_2, i}$ 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 data $\{(D_{00,i}, D_{10,i}, D_{01,i}, D_{11,i})\}_{i=1}^n$. So the data provides information of four choice probabilities $(P(0,0), P(1,0), P(0,1), P(1,1))$ (denoted as the true reduced-form parameter values $(\kappa_{00}, \kappa_{10}, \kappa_{01}, \kappa_{11})$), whereas 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 a the equilibrium selection probability.

		Player 2: 0		Player 2: 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

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

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

The equalities above naturally suggest a GMM approach via the following moments:

$$
\begin{aligned}
E\Big[\mathbb{1}\big((Y_1, Y_2) = (1,1)\big) - P(\epsilon_1 \geq -\beta_1 - \Delta_1;\, \epsilon_2 \geq -\beta_2 - \Delta_2)\Big] &= 0 \\
E\Big[\mathbb{1}\big((Y_1, Y_2) = (0,0)\big) - P(\epsilon_1 \leq -\beta_1;\, \epsilon_2 \leq -\beta_2)\Big] &= 0 \\
E\Big[\mathbb{1}\big((Y_1, Y_2) = (1,0)\big) - s \times P(-\beta_1 \leq \epsilon_1 \leq -\beta_1 - \Delta_1;\, -\beta_2 \leq \epsilon_2 \leq -\beta_2 - \Delta_2) & \\
- P(\epsilon_1 \geq -\beta_1; \epsilon_2 \leq -\beta_2) - P(\epsilon_1 \geq -\beta_1 - \Delta_1; -\beta_2 \leq \epsilon_2 \leq -\beta_2 - \Delta_2)\Big] &= 0\,.
\end{aligned}
$$

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 7 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 7 and Figure 8 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 smallest sample size. 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] In contrast, the projection CS $\widehat{M}_\alpha^{proj}$ for M_I (of Δ_1) is extremely conservative. And the coverage of percentile-based CSs for Δ_1, which is the Chernozhukov and Hong (2003) procedure for a point-identified parameter, was less than 1% for each sample size (and hence was not tabulated).

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

MC coverage probabilities for $\widehat{\Theta}_\alpha$ (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 for $\widehat{M_\alpha}$ (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 for $\widehat{M_\alpha^\chi}$ (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 for $\widehat{M_\alpha^{proj}}$				
	$n = 100$	$n = 250$	$n = 500$	$n = 1000$
$\alpha = 0.90$	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

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.

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 contribution is the derivation of the large-sample (quasi)-posterior distribution of the QLR statistic for Θ_I and profile QLR statistic for M_I under loss of identifiability.

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

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

Assumption 3.1. *(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 some unknown $d^* \in [1,\infty)$, 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 neighborhood of Θ_I.

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).

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:*

$$\sup_{\theta\in\Theta_{osn}} \left| nL_n(\theta) - \left(\ell_n - \frac{1}{2}\|\sqrt{n}\gamma(\theta)\|^2 + (\sqrt{n}\gamma(\theta))'\mathbb{V}_n \right) \right| = o_{\mathbb{P}}(1) \tag{12}$$

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

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

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

Discussion of Assumptions: Assumption 3.1 is a mild posterior contraction condition. The definition of Θ_{osn} is deliberately general and will typically depend on the particular model under consideration. For example, in likelihood models we could take $\Theta_{osn} = \{\theta \in \Theta : h(P_\theta, P_0) \leq r_n/\sqrt{n}\}$ where h is Hellinger distance and $r_n \to \infty$ slowly as $n \to \infty$. Assumption 3.2(i) is readily verified for likelihood, GMM and generalized empirical likelihood models (see Sections 4.1.1–4.2). 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 $E[v(X_i)] = 0$ and $Var[v(X_i)] = \Sigma$. Parts (ii) is trivially satisfied whenever each K_{osn} contains a ball of radius k_n centered at the origin. More generally, these conditions allow for the origin $\gamma = 0$ to be on the boundary of Γ and are similar to conditions used for identified models when a parameter is on the boundary (see, e.g., Andrews (1999)). The convexity can be weakened (at the cost of more complicated notation) to allow for the cone to be non-convex. Assumption 3.3(i) requires the quasi-posterior to be proper. Part (ii) is a standard prior mass and smoothness condition used to establish Bernstein-von Mises results for identified models (see, e.g., Section 10.2 of van der Vaart (2000)) but applied to Π_Γ. Finally, Assumption 3.4 merely requires that the distribution of the MCMC chain $Q_n(\theta^1), \ldots, Q_n(\theta^B)$ well approximates the posterior distribution of $Q_n(\theta)$.

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 the asymptotic distribution of* $\sup_{\theta\in\Theta_I} Q_n(\theta)$ *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 Bernstein-von Mises type result 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). Note that $\mathbb{V}_n$ is a function of the data. Let $T - \mathbb{V}_n$ denote the cone T translated to have vertex at $-\mathbb{V}_n$. Let $\mathbf{T}$ be the orthogonal projection onto T and $\mathbf{T}^\perp$ denote the orthogonal projection onto the polar cone of T.[17]

[17]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

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

$$\sup_z \left| \Pi_n\big(\{\theta : Q_n(\theta) \leq z\} \big| \mathbf{X}_n\big) - \mathbb{P}_{Z|\mathbf{X}_n}\Big(\|Z\|^2 \leq z + \|\mathbf{T}^\perp \mathbb{V}_n\|^2 \Big| Z \in T - \mathbb{V}_n \Big) \right| = o_{\mathbb{P}}(1) \,. \tag{13}$$

(i) Hence $\Pi_n\big(\{\theta : Q_n(\theta) \leq z\} \big| \mathbf{X}_n\big) \leq \mathbb{P}_{Z|\mathbf{X}_n}(\|\mathbf{T}Z\|^2 \leq z)$ *for all* $z \geq 0$.
(ii) If $T = \mathbb{R}^{d^*}$ *then:*

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

where $F_{\chi^2_{d^*}}$ *denotes the cdf of the* $\chi^2_{d^*}$ *distribution.*

First consider the case in which $T = \mathbb{R}^{d^*}$. Lemma 3.1(ii) shows that the posterior of the QLR is asymptotically $\chi^2_{d^*}$ when $T = \mathbb{R}^{d^*}$. Notice that Lemma 3.1 does not require the generalized information equality $\Sigma = I_{d^*}$ to hold. Theorem 3.1 requires $\Sigma = I_{d^*}$ so that the asymptotic distribution of the QLR is itself $\chi^2_{d^*}$ and therefore coincides with the posterior distribution. Remark 3 discusses this issue in more detail.

Now consider the case in which $T \subsetneq \mathbb{R}^{d^*}$, which will occur when the origin is on the boundary of Γ. When $\Sigma = I_{d^*}$, the cdf of the asymptotic distribution of $\sup_{\theta \in \Theta_I} Q_n(\theta)$ is:

$$F_T(z) = \mathbb{P}_Z(\|\mathbf{T}Z\|^2 \leq z) \tag{14}$$

where $\mathbb{P}_Z$ denotes the distribution of a $N(0, I_{d^*})$ random vector Z. Notice that F_T reduces to the $\chi^2_{d^*}$ distribution when $T = \mathbb{R}^{d^*}$. 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). Lemma 3.1 part (i) shows that the posterior distribution of the QLR asymptotically (first-order) stochastically dominates F_T. It follows that $\widehat{\Theta}_\alpha$ will be asymptotically valid but conservative in this case. The conservativeness of $\widehat{\Theta}_\alpha$ will depend on the shape of T.

Remark 3 (Optimal Weighting)**.** *The Bernstein-von Mises theorem provides conditions under which the posterior distribution of* $\sqrt{n}(\theta - \hat{\theta})$ *(where* $\hat{\theta}$ *is the MLE) in correctly specified identifiable likelihood models converges to the asymptotic distribution of* $\sqrt{n}(\hat{\theta} - \theta_0)$*. It is well known that this equivalence does not hold under misspecification. Instead, the QMLE is asymptotically normal, centered at the pseudo-true parameter with sandwich covariance matrix, whereas the posterior is asymptotically normal, centered at the QMLE, with variance equal to the inverse of the Hessian of* $P_0 \log(p_0/p(\cdot;\theta))$ *(where* P_0 *and* p_0 *denote the true distribution and density) evaluated*

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).

at the pseudo-true parameter (Kleijn and van der Vaart, 2012). Thus, under misspecification the posterior distribution retains the correct centering but has the incorrect scale.

Similarly, we require the quasi-posterior distribution of the QLR to have correct scale in order for our MCMC confidence sets $\widehat{\Theta}$ and $\widehat{M}_\alpha$ to be asymptotically valid $100\alpha\%$ (frequentist) confidence sets for Θ_I and M_I. This means that we require Assumption 3.2 hold with $\Sigma = I_{d^*}$ (a generalized information equality). This is why we confine our attention to correctly-specified likelihood or optimally-weighted or continuously-updated GMM criterion. Our results should extend to empirical-likelihood based criterion functions.

Our results could also be applied to sandwich likelihoods (Müller, 2013) in misspecified, separable likelihood models. In such models we can rewrite the density as $p(\cdot;\theta) = q(\cdot;\gamma(\theta))$ where γ is an identifiable reduced-form parameter (see Section 4.1.1 below). Under misspecification the identified set is $\Theta_I = \{\theta : \gamma(\theta) = \gamma^*\}$ where γ^* is the unique value of γ that minimizes $P_0 \log(p_0(\cdot)/q(\cdot;\gamma))$. Here we could base inference on the sandwich log-likelihood function:

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

where $\hat{\gamma}$ is the QMLE:

$$\frac{1}{n}\sum_{i=1}^{n}\log q(X_i;\hat{\gamma}) \geq \max_{\gamma\in\Gamma}\frac{1}{n}\sum_{i=1}^{n}\log q(X_i;\gamma) + o_{\mathbb{P}}(n^{-1})$$

and $\widehat{\Sigma}_S$ is the sandwich covariance matrix estimator for $\hat{\gamma}$.

3.1.1 Models with singularities

In this section we deal with non-identifiable models with singularities.[18] We show that MCMC CSs $\widehat{\Theta}_\alpha$ (Procedure 1) for Θ_I remain valid but conservative for models with singularities. Importantly and interestingly, Section 3.2 will show that our CSs $\widehat{M}_\alpha$ (Procedure 2) for M_I can have asymptotically correct coverage in this case, even though $\widehat{\Theta}_\alpha$ may be asymptotically conservative.

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_{θ_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 quadratic on a neighborhood of θ_0. By analogy with the identifiable case, we say a non-identifiable model has a singularity if it

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

does not admit a local quadratic approximation like that in Assumption 3.2(i). One prominent example is the missing data model under identification, for which a local quadratic expansion of the form:

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

is obtained for some $\gamma_\perp:\Theta\to\mathbb{R}_+$ (see Section 4.1.1 below). This expansion shows the likelihood is locally quadratic in the reduced-form parameters $\gamma(\theta)$ and locally linear in the reduced-form parameters $\gamma_\perp(\theta)$.

To allow for 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 replace Assumptions 3.2 and 3.3 in the singular case.

Assumption 3.2′ *(Local quadratic approximation with singularity)*
(i) There exist sequences of random variables ℓ_n and $\mathbb{R}^{d^*}$-valued random vectors $\mathbb{V}_n$ (both measurable of data $\mathbf{X}_n$), and a sequence of functions $f_{n,\perp}:\Theta\to\mathbb{R}_+$ that is measurable jointly in $\mathbf{X}_n$ and θ, such that:

$$\sup_{\theta\in\Theta_{osn}}\left|nL_n(\theta)-\left(\ell_n-\frac{1}{2}\|\sqrt{n}\gamma(\theta)\|^2+(\sqrt{n}\gamma(\theta))'\mathbb{V}_n-f_{n,\perp}(\gamma_\perp(\theta))\right)\right|=o_{\mathbb{P}}(1)\qquad(15)$$

with $\sup_{\theta\in\Theta_{osn}}\|(\gamma(\theta),\gamma_\perp(\theta))\|\to 0$ and $\mathbb{V}_n\rightsquigarrow N(0,\Sigma)$ 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. Part (ii) requires that the peak of the likelihood does not concentrate on sets of the form $\{\theta:f_{n,\perp}(\gamma_\perp(\theta))>\epsilon\}$, and may be weakened but at the cost of more complicated notation. Recently, Bochkina and Green (2014) established a Bernstein-von Mises result for *identifiable* singular likelihood models. They assume the likelihood is locally quadratic in some parameters and locally linear in others (similar to Assumption 3.2′(i)). They also assume the local parameter space satisfies conditions similar to parts (ii) and (iii). Finally, Assumption 3.3′

generalizes Assumption 3.3 to the singular case. Note that we impose no further restrictions on the set $\{\gamma_\perp(\theta) : \theta \in \Theta\}$.

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 the asymptotic distribution of $\sup_{\theta\in\Theta_I} Q_n(\theta)$ is continuous at its α quantile, we have:*

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

Theorem 3.2 shows that $\widehat{\Theta}_\alpha$ is asymptotically valid for Θ_I but conservative in singular models whereas Theorem 3.1 shows $\widehat{\Theta}_\alpha$ is valid with asymptotically correct coverage in non-singular models when the tangent cone T is linear and conservative in non-singular models when T is a cone. Importantly, in Section 3.2 below we show that our CSs for the identified set for functions of Θ_I (including subvectors) can have asymptotically correct coverage irrespective of whether the model is singular or not. Consider the missing data example. as we show in Section 4.1.1, $\widehat{\Theta}_\alpha$ will be conservative under point identification but asymptotically correct under partial identification, whereas $\widehat{M}_\alpha$ for the identified set M_I of the mean parameter is asymptotically exact irrespective of whether the model is point-identified or not.

The key step in the proof of Theorem 3.2 is to show that the posterior distribution of the QLR asymptotically (first-order) stochastically dominates the asymptotic distribution of the QLR, namely F_T defined in (14).[19]

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

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

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:

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

[19] In particular, this implies that the posterior distribution of the QLR asymptotically dominates the $\chi^2_{d^*}$ distribution when $T = \mathbb{R}^{d^*}$.

which is different form the definition (8) of the profile quasi-likelihood for *a point* $m \in M$. But, $PL_n(\Delta(\theta))$ is the same as the definition (9) of the profile quasi-likelihood for the set M_I:

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

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 $Q_n(\bar{\theta}) = 2n(L_n(\hat{\theta}) - L_n(\bar{\theta}))$ as in (6). We use these non-standard definitions of PL_n and PQ_n as we are concerned with inference on the whole identified set M_I rather than testing whether a particular point $m = \mu(\theta)$ belongs to M_I. In particular the profile QLR for the set M_I is

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

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

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

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

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 $\{\bar{\theta} \in \Theta_I^N : \gamma(\bar{\theta}) = \gamma(\theta)\}$.

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

Discussion of Assumptions: Assumption 3.5 imposes mild structure on the posterior distribution of the QLR statistic for M_I on the local neighborhood Θ_{osn}. In section 4.1.1 we show that Assumption 3.5 holds for the missing data model under partial identification with $f : \mathbb{R}^2 \to \mathbb{R}_+$ given by $f(v) = (\inf_{t \in T_1} \|v - t\|^2) \vee (\inf_{t \in T_2} \|v - t\|^2)$ where T_1 and T_2 are halfspaces in $\mathbb{R}^2$, and under identification with $f : \mathbb{R} \to \mathbb{R}_+$ given by $f(v) = v^2$. In general, we deal with models for which the profile QLR for M_I is of the form:

$$PQ_n(\Delta(\theta)) = f(\mathbb{V}_n) - \|\mathbf{T}^{\perp}\mathbb{V}_n\|^2 + o_{\mathbb{P}}(1) \quad \text{for each } \theta \in \Theta_I \tag{16}$$

where $f : \mathbb{R}^{d^*} \to \mathbb{R}_+$ is a measurable function which satisfies $f(v) \geq \|\mathbf{T}^\perp v\|^2$ for $v \in \mathbb{R}^{d^*}$. We do not need to know the expression of f in the implementation of our MCMC CS construction, it is merely a proof device. Examples of cases in which the profile QLR for M_I is of the form (16) includes the special case in which M_I is a singleton, with $f(v) = \inf_{t \in T_1} \|v - t\|^2$ where $T_1 = \mathbb{R}^{(\dim(T_1)} \subset T = \mathbb{R}^{d^*}$ and the QLR statistic is $\chi^2_{d^* - \dim(T_1)}$. More generally, we allow for the profile QLR statistic to be 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. One such example is when f is of the form:

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

where $f_0 : \mathbb{R}^J \to \mathbb{R}_+$ and $T_1, \ldots, T_J$ are closed cones in $\mathbb{R}^{d^*}$.
Assumption 3.6 requires that the distribution of the profile QLR statistic computed from the MCMC chain well approximates the posterior distribution of the profile QLR statistic.

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 $\widehat{M}_\alpha$ has asymptotically correct coverage irrespective of whether the model is singular or not.

A key step in the proof of Theorem 3.3 is the following new BvM type result for the posterior distribution of the profile QLR for $M_I = \mu(\Delta(\theta))$ for $\theta \in \Theta_I$. For any $S \subset \mathbb{R}_+$ and $\epsilon > 0$, let $S^{-\epsilon}$ denote the ϵ-contraction of S and let $S_n^{-\epsilon} = \{s - \|\mathbf{T}^\perp \mathbb{V}_n\|^2 : s \in S^{-\epsilon}\}$.[20]

Lemma 3.3. *Let Assumptions 3.1, 3.2, 3.3, and 3.5 or 3.1, 3.2', 3.3', and 3.5 hold, and let $z \mapsto \mathbb{P}_Z(f(Z) \leq z)$ 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:*

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

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

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

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

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

The following result presents just one set of sufficient conditions for validity of the CS $\widehat{M}_\alpha^\chi$ for M_I. This condition places additional structure on the function f in Assumption 3.5. There may exist other sufficient conditions. One can also generalize $\widehat{M}_\alpha^\chi$ to allow for quantiles of χ^2 with higher degrees of freedom.

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 $\mathbb{R}^{d^*}$ whose supporting hyperplanes pass through the origin.

Sufficient conditions for Assumption 3.7 are in Proposition 3.1 below.

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. 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 $\frac{1}{4}\delta_0 + \frac{1}{2}\chi_1^2 + \frac{1}{4}(\chi_1^2 \cdot \chi_1^2)$ where δ_0 denotes point mass at zero and $\chi_1^2 \cdot \chi_1^2$ denotes the distribution of the product of two independent χ_1^2 random variables. The quantiles of the distribution of $f(Z)$ are continuous in α for all $\alpha > \frac{1}{4}$. In other configurations of T_1 and T_2, the distribution of $f(Z)$ will (first-order) stochastically dominate $\frac{1}{4}\delta_0 + \frac{1}{2}\chi_1^2 + \frac{1}{4}(\chi_1^2 \cdot \chi_1^2)$ and will itself be (first-order) stochastically dominated by χ_1^2. 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).

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

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

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

and $\inf_{\gamma\in\Gamma_{m,osn}} \|\sqrt{n}\gamma - \mathbb{V}_n\| = \inf_{t\in T_m} \|t - \mathbb{V}_n\|^2 + o_{\mathbb{P}}(1)$;
(iii) Assumptions 3.1(i), 3.2(i)(ii) or 3.2'(i)(ii)(iii) hold;

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

4 Sufficient conditions

This section provides sufficient conditions for the main results derived in Section 3. We start with likelihood models and then consider GMM models.

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, $D_{KL}(p_0(\cdot)||p(\cdot;\theta))$ 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 $\Theta_I = \{\theta \in \Theta : D_{KL}(p_0(\cdot)||p(\cdot;\theta)) = 0\} = \{\theta \in \Theta : h(p_0(\cdot), p(\cdot;\theta)) = 0\}$. 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.

4.1.1 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. with density* $p_0(\cdot) \in \{q(\cdot;\tilde{\gamma}) : \tilde{\gamma} \in \widetilde{\Gamma}\}$*, where* $\widetilde{\Gamma}$ *is a compact subset in* $\mathbb{R}^{d^*}$*;*
(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 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)).

Missing data example

Let D_i be a binary selection variable and Y_i be a binary outcome variable. We observe (D_i, Y_iD_i). The (reduced-form) probabilities of observing $(D_i, Y_iD_i) = (1,1)$, $(1,0)$, and $(0,0)$ are $\kappa_{11}(\theta)$, $\kappa_{10}(\theta)$, and $\kappa_{00}(\theta)$, where θ is a structural parameter. Let κ_{11}, κ_{10}, and κ_{00} denote the true probabilities. The parameter of interest is usually $\mu_0 := \mathbb{E}[Y_i]$ which is partially identified when $\kappa_{00} > 0$ with $M_I = [\kappa_{11}, \kappa_{11} + \kappa_{00}]$. We assume that $0 < \Pr(Y_i = 1|D_i = 1) < 1$.

Inference under partial identification: We first discuss the case in which the model is partially identified (i.e. $0 < \kappa_{00} < 1$). The likelihood 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 if $\theta \mapsto \tilde{\gamma}(\theta)$ is twice continuously differentiable. The local quadratic expansion in Assumption 3.2 is obtained 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 $T = \mathbb{R}^2$.

We use the parameterization $\theta = (\mu, \beta, \rho)$ where $\mu = \mathbb{E}[Y_i]$, $\beta = \Pr(Y_i = 1 | D_i = 0)$, and $\rho = \Pr(D_i = 1)$. The parameter space is

$$\Theta = \{(\mu, \beta, \rho) \in \mathbb{R}^3 : 0 \leq \mu - \beta(1-\rho) \leq \rho, 0 \leq \beta \leq 1, 0 \leq \rho \leq 1\}. \tag{17}$$

The reduced-form probabilities are $\kappa_{11}(\theta) = \mu - \beta(1-\rho)$, $\kappa_{10}(\theta) = \rho - \mu + \beta(1-\rho)$, and $\kappa_{00}(\theta) = 1 - \rho$. The identified set is:

$$\Theta_I = \{(\mu, \beta, \rho) \in \Theta : \mu - \beta(1-\rho) = \kappa_{11}, \rho = \rho_0\}$$

so ρ is always identified but only an affine combination of μ and β are identified. A flat prior on Θ in (17) induces a flat prior on Γ, which verifies Condition (c) and Assumption 3.3. Therefore, MCMC confidence sets for Θ_I will have asymptotically correct coverage.

Now consider subvector inference on μ. The identified set is $M_I = [\kappa_{11}, \kappa_{11} + \kappa_{00}]$. We have $\mu^{-1}(m) = \{m\} \times \{(\beta, \rho) \in [0,1]^2 : 0 \leq m - \beta(1-\rho) \leq \rho\}$. By concavity in m, the profile likelihood for M_I is:

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

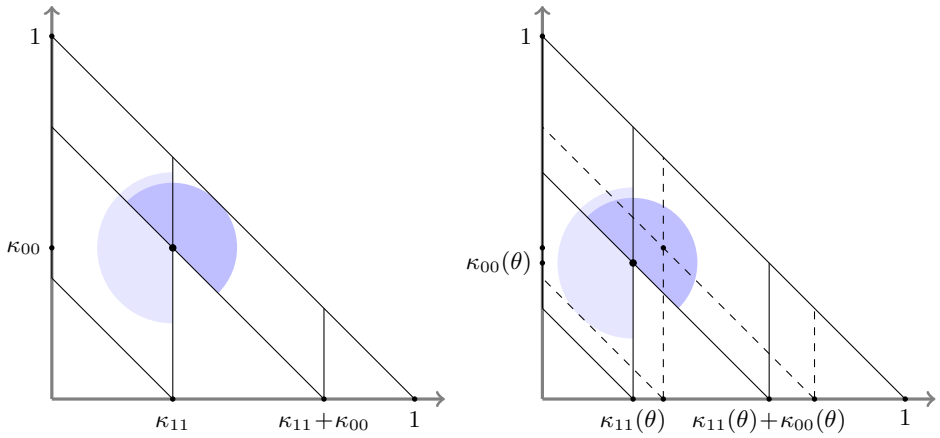

Figure 4: Local parameter spaces for the profile LR statistic for M_I. Left panel: the lightly shaded region is for the maximization problem (18) at $m = \kappa_{11}$ and the darker shaded region is for the maximization problem at $m = \kappa_{11} + \kappa_{00}$. Right panel: corresponding problems for the profile LR (19) at $\kappa_{11}(\theta)$ and $(\kappa_{11}(\theta), \kappa_{00}(\theta))'$.

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

$$\begin{aligned}
&\sup_{\bar{\theta}\in\mu^{-1}(m)} n\mathbb{P}_n \log(p(\cdot;\bar{\theta})) \\
&= \sup_{\substack{0\le k_{11}\le m \\ m\le k_{11}+k_{00}\le 1}} n\mathbb{P}_n\Big(yd\log k_{11} + (d-yd)\log(1-k_{11}-k_{00}) + (1-d)\log k_{00}\Big)\,. \qquad (18)
\end{aligned}$$

at $m = \kappa_{11}$ and $m = \kappa_{11} + \kappa_{00}$. The local parameter spaces for problem (18) 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:

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

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

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

This is of the form (16) with $f(v) = (\inf_{t\in T_1}\|v-t\|^2)\vee(\inf_{t\in T_2}\|v-t\|^2)$.

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

$$nPL_n(\Delta(\theta)) = \min_{m \in \{\kappa_{11}(\theta), \kappa_{11}(\theta)+\kappa_{00}(\theta)\}} \sup_{\bar{\theta} \in \mu^{-1}(m)} n\mathbb{P}_n \log p(\cdot, \bar{\theta}) \,. \tag{19}$$

By analogy with display (18), to calculate $PL_n(\Delta(\theta))$ we need to solve:

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

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:

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

uniformly for $\theta \in \Theta_{osn}$, verifying Assumption 3.5. Therefore, MCMC confidence sets for M_I will have asymptotically correct 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 log-likelihood reduces to:

$$L_n(\theta) = n\mathbb{P}_n \big(y \log(\kappa_{11}(\theta)) + (1 - y) \log(1 - \kappa_{11}(\theta) - \kappa_{00}(\theta)) \big) \,.$$

We again take Θ as in (17) and use a flat prior. Π_n concentrates on the local neighborhood Θ_{osn} given by $\Theta_{osn} = \{\theta : |\kappa_{11}(\theta) - \kappa_{11}| \le r_n/\sqrt{n}, \kappa_{00}(\theta) \le r_n/n\}$ 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:

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

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 $\gamma_\perp(\theta) = \kappa_{00}(\theta) \ge 0$, 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, $\widehat{\Theta}_\alpha$ will be valid but conservative.

For subvector inference on μ, the profile LR statistic for $M_I = \{\mu_0\}$ is asymptotically χ^2_1 with

$f(\mathbb{V}_n) = \mathbb{V}_n^2$. To verify Assumption 3.5, for each $\theta \in \Theta_{osn}$ we need to solve

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

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}:

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

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 final line is 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 $\widehat{M}_\alpha$ for M_I will have asymptotically exact coverage, even though $\widehat{\Theta}_\alpha$ for Θ_I will be conservative.

Complete information entry game example

Consider the bivariate discrete game with payoffs described in Table 9. Let $D_{a_1a_2,i}$ 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 $(D_{00,i}, D_{01,i}, D_{10,i}, D_{11,i})$. 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 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 $\theta \mapsto \kappa(\theta)$. The local quadratic expansion in Assumption 3.2 is obtained 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) and Assumption 3.3 can be verified under mild conditions on the map $\theta \mapsto \kappa(\theta)$ and the prior Π.

As a concrete example, 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, verifying Condition (c) and Assumption 3.3. Therefore, MCMC CSs for Θ_I will have asymptotically correct 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 reduced-form reparameteri-

zation. 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,[21] 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 $\text{Span}(\Lambda)$ 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 elements of $\psi = (\psi_1, \ldots, \psi_{d^*})$ form an orthonormal basis for $\text{Span}(\Lambda)$ in $L^2(P_0)$. Let $\mathbf{\Lambda}$ denote the orthogonal projection onto Λ and let $\gamma(\theta)$ be given by $\mathbf{\Lambda}(2(\sqrt{p(\cdot;\theta)/p_0(\cdot)} - 1)) = \gamma(\theta)'\psi$.[22] 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 $\text{Span}(\Lambda)$ 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 $\mathbf{\Lambda}(2(\sqrt{p(\cdot;\theta)/p_0(\cdot)} - 1)) = \gamma(\theta)'\psi$.

4.2 GMM models

Consider the GMM model $\{\rho(X_i, \theta) : \theta \in \Theta\}$ with $\rho : \mathscr{X} \times \Theta \to \mathbb{R}^{\dim(\rho)}$. The identified set is $\Theta_I = \{\theta \in \Theta : E[\rho(X_i, \theta)] = 0\}$. Let $g(\theta) = E[\rho(X_i, \theta)]$ and $\Omega(\theta) = E[\rho(X_i, \theta)\rho(X_i, \theta)']$. An equivalent definition of Θ_I is $\Theta_I = \{\theta \in \Theta : g(\theta) = 0\}$. 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.

[21] 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.

[22] 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)).

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 $\text{Span}(\Lambda)$ and the remaining $\dim(g) - d^*$ are orthogonal to $\text{Span}(\Lambda)$. If $\{g(\theta) : \theta \in \Theta\}$ contains a neighborhood of the origin then we just take $\Lambda = \mathbb{R}^{\dim(g)}$, $U = I_{\dim(g)}$, and $\mathbf{\Lambda}g(\theta) = g(\theta)$. Also define $\mathcal{R}_\varepsilon = \{\rho(\cdot,\theta) : \theta \in \Theta, \|g(\theta)\| \leq \varepsilon\}$ and $\Theta_I^\varepsilon = \{\theta \in \Theta : \|g(\theta)\| \leq \varepsilon\}$.

Proposition 4.3. *Let the following hold:*
(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 objective function (2) over* $\Theta_{osn} = \{\theta \in \Theta : \|g(\theta)\| \leq r_n/\sqrt{n}\}$ *with* $\gamma(\theta) = [(U\Omega U')^{-1}]_{11}[U\mathbf{\Lambda}g(\theta)]_1$ *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 first* d^* *elements of* $U\mathbf{\Lambda}g(\theta)$, $\mathbb{V}_n = -[(U\Omega U')^{-1}]_{11}^{-1/2}[U\Omega^{-1}\mathbb{G}_n(\rho(\cdot,\theta))]_1$ *where* $[U\Omega^{-1}\mathbb{G}_n(\rho(\cdot,\theta))]_1$ *is the upper* d^* *subvector of* $U\Omega^{-1}\mathbb{G}_n(\rho(\cdot;\theta))$ *for any fixed* $\theta \in \Theta_I$*, and with* T *equal 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 we simply have* $\gamma(\theta) = \Omega^{-1/2}g(\theta)$, $\mathbb{V}_n = \Omega^{-1/2}\mathbb{G}_n(\rho(\cdot,\theta))$ *for any* $\theta \in \Theta_I$*, and* $T = \mathbb{R}^{\dim(g)}$.

A similar result holds for the optimally-weighted GMM objective function.

Proposition 4.4. *Let conditions (a)-(d) of Proposition 4.3 hold, and its condition (e) be replaced by:*
(e) $\|\widehat{W} - \Omega^{-1}\| = o_{\mathbb{P}}(1)$.
Then: the conclusions of Proposition 4.3 remain valid for the optimally-weighted GMM objective function (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.[23] Then Condition (a) of Propositions 4.3

[23]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 $\varphi(N \cap \{g(\theta) : \theta \in \Theta\}) = \Phi \cap \text{im}(\varphi)$ where $\text{im}(\varphi)$ 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.

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 $\theta = (\beta, \lambda) \in B \times \Lambda$ and write the inequality model as a GMM equality model

$$E[\rho(X_i, \theta)] = 0 \text{ for } \theta \in \Theta_I, \quad \rho(X_i, \theta) = \tilde{\rho}(X_i, \beta) + \lambda \ , \tag{20}$$

where the identified set for θ is $\Theta_I = \{\theta \in B \times \Lambda : E[\rho(X_i, \theta)] = 0\}$ 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 (20).

An example.[24] As a simple illustration, consider the model in which $X_1, \ldots, X_n$ are i.i.d. with unknown mean $\mu \in [0, \bar{b}] = B$ and variance $\sigma^2 < \infty$. Suppose that $\beta \in B$ is identified by the moment inequality $E[\tilde{\rho}(X_i, \beta)] \leq 0$ where $\tilde{\rho}(X_i, \beta) = \beta - X_i$ and so $B_I = [0, \mu]$. We rewrite this as a moment equality model by introducing the slackness parameter $\lambda \in B$ and writing the residual function as $\rho(X_i, \theta) = \lambda + \beta - X_i$ for $\theta = (\beta, \lambda) \in B^2 = \Theta$. The CU-GMM objective function is:

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

where $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2$. Suppose that $\mu \in (0, \bar{b})$. Then wpa1 we can choose $(\beta, \lambda) \in \Theta$ such that $\bar{X}_n = \lambda + \beta$. Let $\Theta_{osn} = \{(\beta, \lambda) \in \Theta : |\beta + \lambda - \mu| \leq k_n/\sqrt{n}\}$ where $k_n \to \infty$ slowly enough that $k_n^2(\frac{\sigma^2}{\hat{\sigma}^2} - 1) = o_{\mathbb{P}}(1)$. Then:

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

[24] We thank Kirill Evdokimov for suggesting this example, which clearly highlights the fact that our approach uses a *different* criterion function from that is typically used in moment inequality literature.

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

$$\inf_{\lambda \in B} Q_n(\beta, \lambda) = \begin{cases} (\mathbb{V}_n - \sqrt{n}(\beta - \mu)/\sigma)^2 & \text{if } \sigma\mathbb{V}_n/\sqrt{n} - (\beta - \mu) < 0 \\ 0 & \text{if } 0 \leq \sigma\mathbb{V}_n/\sqrt{n} - (\beta - \mu) \leq \bar{b} \\ (\mathbb{V}_n - \sqrt{n}(\beta + \bar{b} - \mu)/\sigma)^2 & \text{if } \sigma\mathbb{V}_n/\sqrt{n} - (\beta - \mu) > \bar{b}\,. \end{cases}$$

As the maximum of the above for $\beta \in B_I$ is attained at $\beta = \mu$, we have $\sup_{\beta \in B_I} \inf_{\lambda \in B} Q_n(\beta, \lambda) = f(\mathbb{V}_n) + o_{\mathbb{P}}(1)$ where $f(v) = v^2 \mathbb{1}\{v < 0\}$. 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 $\gamma(\theta) = ((\beta + \lambda) - \mu)/\sigma$ and $\Sigma = 1$. The set $\Gamma = \{\gamma(\theta) : \theta \in \Theta\}$ contains a ball of positive radius at the origin when $\mu \in (0, \bar{b})$ hence $T = \mathbb{R}$ (otherwise $T = \mathbb{R}_+$ or $\mathbb{R}_-$ when μ is at the boundary of B). Moreover:

$$\Delta(\theta^b) = \{(\beta, \lambda) \in \Theta : E[\rho(X_i; (\beta, \lambda))] = E[\rho(X_i; \theta^b)]\} = \{(\beta, \lambda) \in \Theta : \beta + \lambda = \beta^b + \lambda^b\}$$

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

$$\sup_{\beta \in \mu(\Delta(\theta))} \inf_{\lambda \in B} Q_n(\beta, \lambda) = f(\mathbb{V}_n - \sqrt{n}\gamma(\theta)) + o_{\mathbb{P}}(1) \quad \text{uniformly in} \quad \theta \in \Theta_{osn}$$

with f as defined above. A flat prior on Θ induces a smooth prior on γ. It is also straightforward to show directly that Assumption 3.1 holds. So all the regularity conditions of Theorem 3.3 hold and we will have asymptotically correct coverage for B_I.

5 Conclusion

We propose new methods for constructing CSs for IdSs in possibly partially-identified structural models. Our MCMC 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) can still be asymptotically exact. Further, our CSs are shown to be asymptotically conservative in models where the tangent space of the reduced-form parameter is a closed convex cone, but asymptotically exact in models where the support of the data could depend on a reduced-form parameter (in Appendix C).

Monte Carlo experiments showcase the good finite-sample coverage properties of our proposed CS constructions in standard difficult situations. This is highlighted in the missing data model with a range of designs that span partial to point identification, the entry game model with correlated shocks, the weakly-identified Euler equation model, and also the finite mixture models.

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). The second extension is to allow for structural models with latent state variables. The third extension is to allow for possibly misspecified likelihoods.

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.

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 nonregular 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.

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

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.

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.

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.

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

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.

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 5 plots the marginal "curved" priors for β and ρ. Figure 6 plots the reduced-form parameters evaluated at the MCMC chain for the structural parameters presented in Figure 2. Although the partially-identified structural parameters μ and β bounce around their respective identified sets, the reduced-form chains in Figure 6 are stable.

A.2 Complete information game

Figure 7 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 8 presents the reduced-form probabilities calculated from the chain in Figure 7.

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)).[25] 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:

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

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

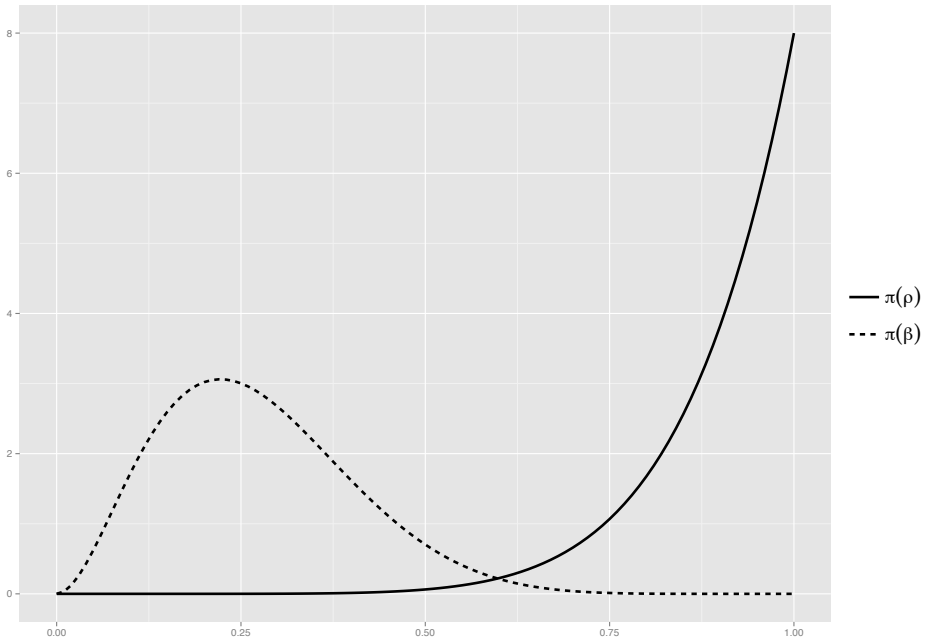

Figure 5: Marginal curved priors for β and ρ for the missing data example.

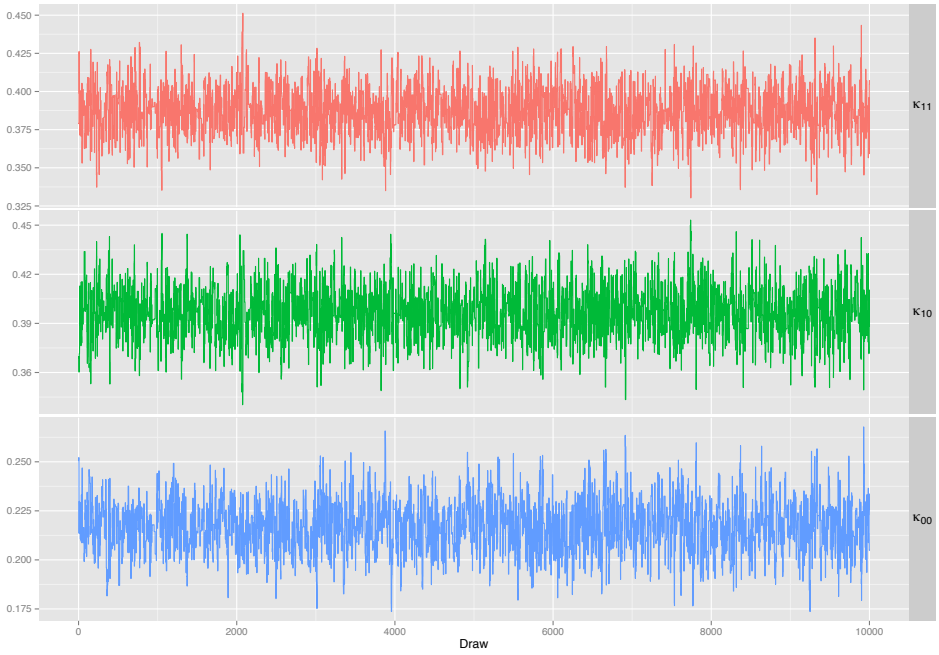

Figure 6: MCMC chain for the reduced-form probabilities $(\kappa_{11}(\theta), \kappa_{10}(\theta), \kappa_{00}(\theta))'$ calculated from the chain in Figure 2. It is clear the chain for the reduced-form probabilities has converged even though the chain for the structural parameters has not.

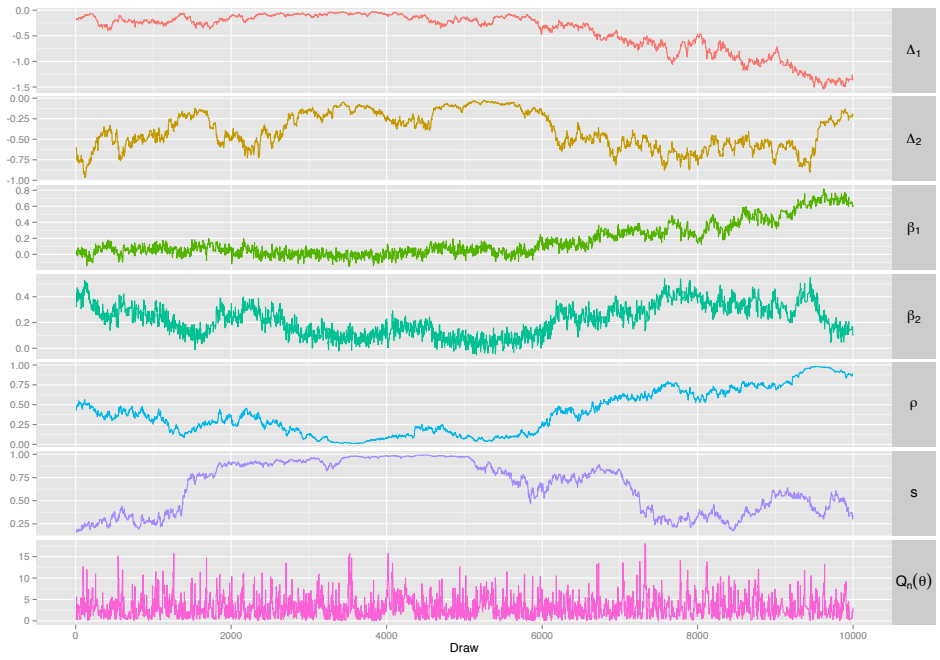

Figure 7: MCMC chain for all structural parameters (top 6 panels) and QLR (bottom panel) with $n = 1000$ using a likelihood for L_n and a flat prior on Θ.

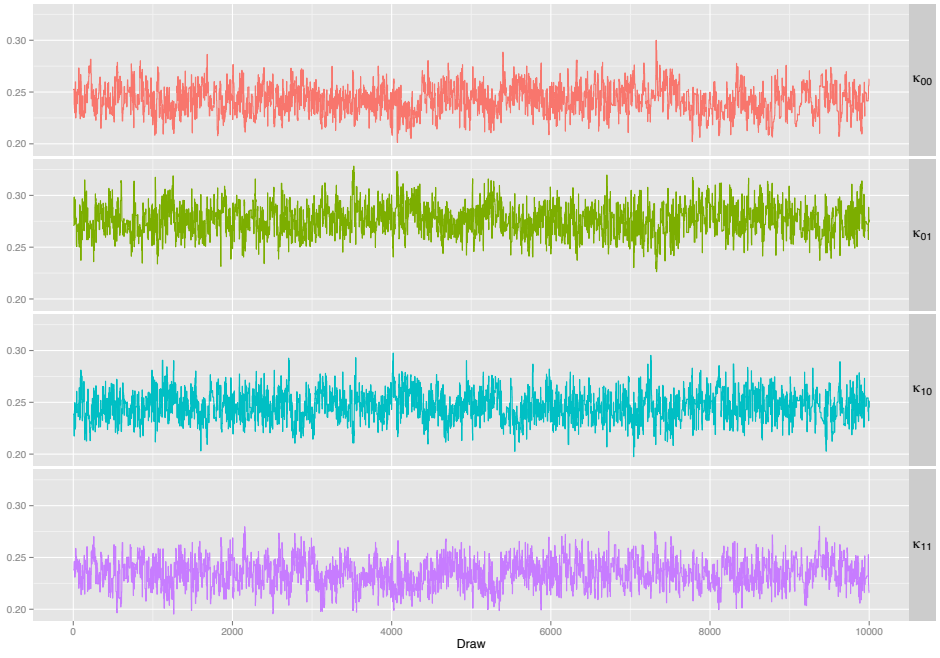

Figure 8: MCMC chain for the reduced-form probabilities calculated from the chain in Figure 7. It is clear that the chain for the reduced-form probabilities has converged, even though the chain for the structural parameters from which they are calculated has not.

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 9 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 10 shows a scatter plot of the (δ, γ) chain which illustrates further the sensitivity of the draws to the prior.

Tables 11 and 12 present coverage properties of our Procedure 1 for the full set $\widehat{\Theta}_\alpha$ (CCT θ in the tables) together with our Procedure 2 for the identified set for δ and γ (CCT δ and CCT 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 11 but undercovers badly for γ even with $n = 500$ in Table 12. 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 $n = 250$.

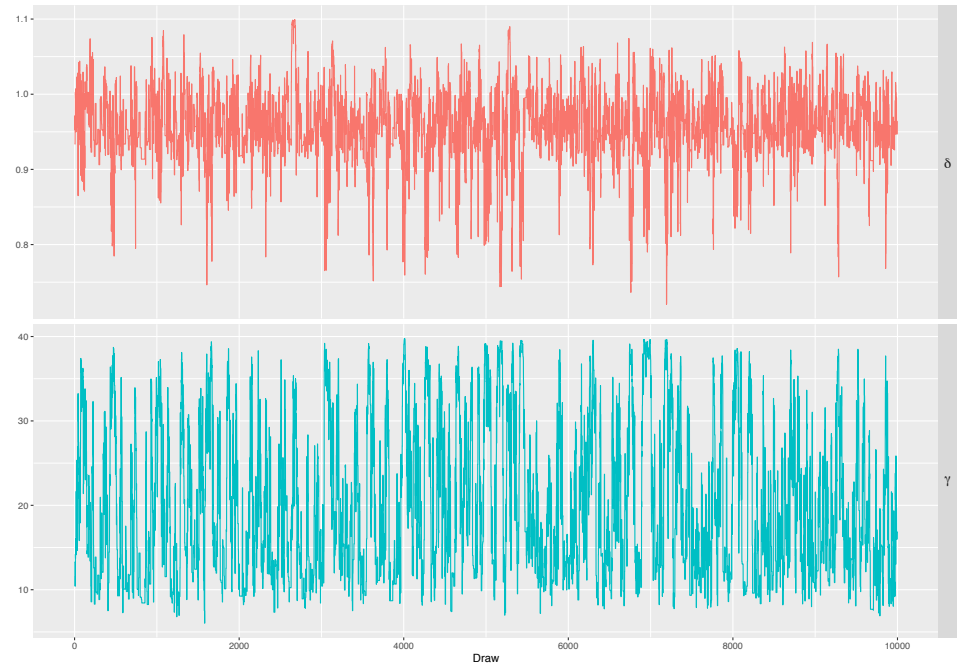

Figure 9: Plots of the MCMC chain for the structural parameter $\theta = (\delta, \gamma)$ with $n = 250$, $\theta_0 = (0.97, 10)$ and a flat prior on $\Theta = [0.6, 1.1] \times [0, 40]$.

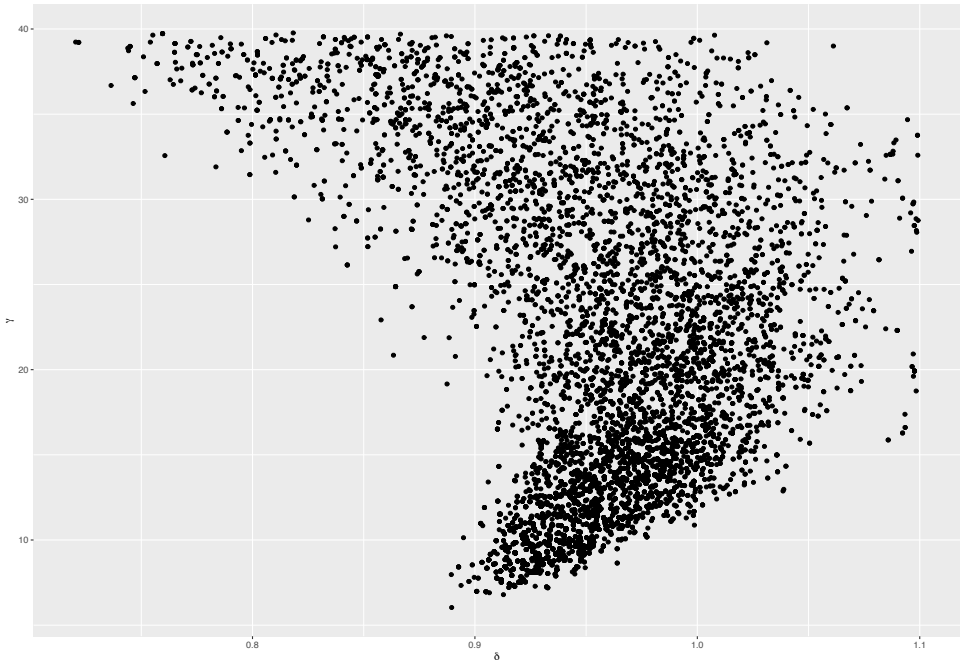

Figure 10: Scatter plot of the chain depicted in 9.

Figure 11: Plots of the moments calculated from the chain in Figure 9.

	CCT θ	CCT δ	CCT 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 11: MC coverage probabilities for $\delta = 0.97 \in [0.8, 1]$, $\gamma = 1.3 \in [0, 10]$.

	CCT θ	CCT δ	CCT 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 12: 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.[26] 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:

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

[26] 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.

We take the 100α percentile of the QLRs and call them ξ_α^μ and ξ_α^η. Confidence sets for M_I and 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:

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

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

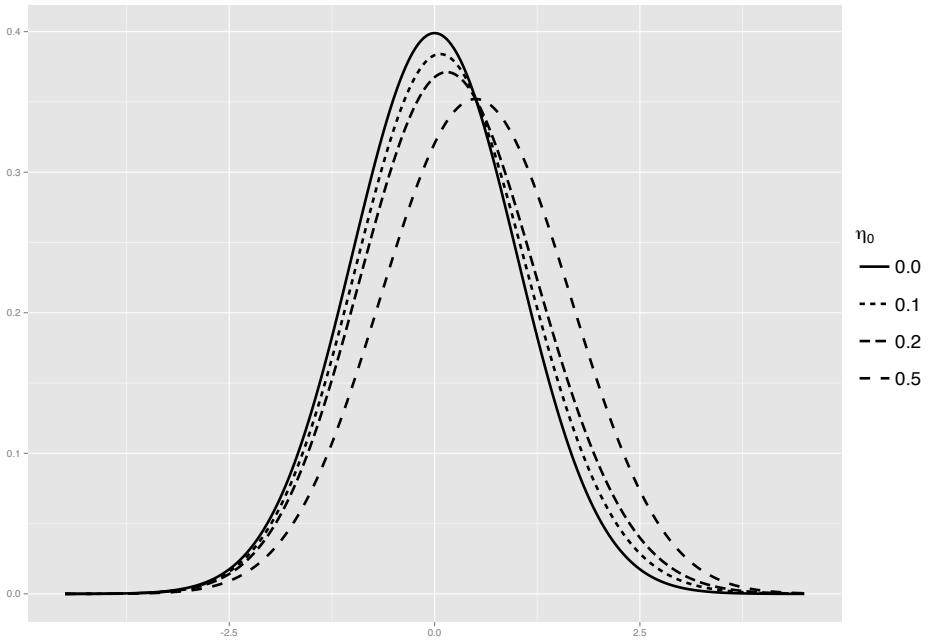

Figure 12: PDFs for the normal mixture MC design for different values of mixing weight η_0.

$[-M, M]$. 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 12). 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 13 and 14 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 15 and 16 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 17 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 13: MC coverage probabilities for $\widehat{M}_\alpha$ (Procedure 2) for different values of η_0 with $\mu_0 = 1$

	$\eta_0 = 0.50$	$\eta_0 = 0.20$	$\eta_0 = 0.10$	$\eta_0 = 0.00$
		$n = 100$		
$\alpha = 0.90$	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 14: MC coverage probabilities for $\widehat{H}_\alpha$ (Procedure 2) for different values of η_0 with $\mu_0 = 1$

	$\eta_0 = 0.50$	$\eta_0 = 0.20$	$\eta_0 = 0.10$	$\eta_0 = 0.00$
		$n = 100$		
$\alpha = 0.90$	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 15: MC coverage probabilities for $\widehat{M}_\alpha^\chi$ (Procedure 3) for different values of η_0 with $\mu_0 = 1$

	$\eta_0 = 0.50$	$\eta_0 = 0.20$	$\eta_0 = 0.10$	$\eta_0 = 0.00$
		$n = 100$		
$\alpha = 0.90$	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 16: MC coverage probabilities for $\widehat{H}_\alpha^\chi$ (Procedure 3) for different values of η_0 with $\mu_0 = 1$.

	$\eta_0 = 0.50$	$\eta_0 = 0.20$	$\eta_0 = 0.10$	$\eta_0 = 0.00$
		$n = 100$		
$\alpha = 0.90$	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 17: MC coverage probabilities for $\widehat{\Theta}_\alpha$ (Procedure 1) for different values of η_0 with $\mu_0 = 1$.

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.:

$$\liminf_{n\to\infty} \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\Theta_I(\mathbb{P}) \subseteq \widehat{\Theta}_\alpha) \geq \alpha \tag{21}$$

$$\liminf_{n\to\infty} \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(M_I(\mathbb{P}) \subseteq \widehat{M}_\alpha) \geq \alpha \tag{22}$$

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 tight and 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: (21) holds for $\widehat{\Theta}_\alpha = \{\theta\in\Theta : Q_n(\theta) \leq w_{n,\alpha}\}$.

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 tight and 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: (22) holds for $\widehat{M}_\alpha = \{\mu(\theta) : \theta\in\Theta, Q_n(\theta) \leq w_{n,\alpha}\}$.

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. *(Posterior contraction)*
$\Pi_n(\Theta_{osn}^c|\,\mathbf{X}_n) = o_{\mathbb{P}}(1)$ *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 $\Theta_I^N(\mathbb{P}) \subseteq \Theta_{osn}(\mathbb{P})$ 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 functions $f_{n,\perp} : \Theta \to \mathbb{R}$ where $f_{n,\perp}$ is jointly measurable in $\mathbf{X}_n$ and θ (we take $\gamma_\perp \equiv 0$ and $f_{n,\perp} \equiv 0$ for nonsingular $\mathbb{P} \in \mathbf{P}$), such that:

$$\sup_{\theta \in \Theta_{osn}} \left| nL_n(\theta) - \left(\ell_n - \frac{1}{2}\|\sqrt{n}\gamma(\theta)\|^2 + (\sqrt{n}\gamma(\theta))'\mathbb{V}_n - f_{n,\perp}(\gamma_\perp(\theta)) \right) \right| = o_{\mathbb{P}}(1) \tag{23}$$

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 (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_r^* denote a ball of radius r 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_r^*, and π_{γ^*} on $\mathbb{P}$.

Assumption B.3. *(Prior)*
(i) $\int_\theta e^{nL_n(\theta)}\, d\Pi(\theta) < \infty$ $\mathbb{P}$-almost surely for each $\mathbb{P} \in \mathbf{P}$;
(ii) Each Π_{Γ^*} has a continuous and strictly positive density π_{Γ^*} on $B_\delta^* \cap (\Gamma \times \Gamma_\perp)$ (or $B_\delta^* \cap \Gamma$ if $\mathbb{P}$ is nonsingular) for some $\delta > 0$ and $\{(\gamma(\theta), \gamma_\perp(\theta)) : \theta \in \Theta_{osn}\} \subseteq B_\delta^*$ (or $\{\gamma(\theta) : \theta \in \Theta_{osn}\} \subseteq B_\delta^*$ if $\mathbb{P}$ is nonsingular) holds uniformly in $\mathbb{P}$ for all n sufficiently large.

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

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

The following results are uniform extensions of Theorem 3.2 and Lemma 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 $\mathbb{P} \in \mathbf{P}$. Then: (21) holds.*

Lemma B.3. *Let Assumptions B.1, B.2 and B.3 hold and let* $L_n(\hat{\theta}) = \sup_{\theta\in\Theta_{osn}} L_n(\theta)+o_{\mathbb{P}}(n^{-1})$ *uniformly for* $\mathbb{P}\in\mathbf{P}$*. 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 $\mathbb{P}\in\mathbf{P}$*.*

To establish (22) 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 QLR statistic)*
(i) For each $\mathbb{P}\in\mathbf{P}$ *there exists a measurable* $f:\mathbb{R}^{d^*}\to\mathbb{R}$ *such that:*

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

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

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

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

The following results are uniform extensions of Theorems 3.3 and Lemma 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: (22) holds.*

Lemma B.4. *Let Assumptions B.1, B.2, B.3, and B.5 hold and let* $L_n(\hat{\theta}) = \sup_{\theta\in\Theta_{osn}} L_n(\theta) + o_{\mathbb{P}}(n^{-1})$ *uniformly for* $\mathbb{P}\in\mathbf{P}$*. Let* $\mathbb{P}_Z$ *denote the distribution of a* $N(0, I_{d^*})$ *random vector. 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 $\mathbb{P}\in\mathbf{P}$*.*

C 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.

Here we replace Assumption 3.2 with the following assumption, which permits the support of the data to depend on certain components of the local reduced-form parameter γ. 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. The following assumption is similar to Assumptions 2-3 in Fan et al. (2000) but has been modified to allow for non-identifiable parameters.

Assumption C.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 $a_n/b_n \to 1$.

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

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

The asymptotic distribution of the QLR under Assumption C.2 may be derived by modifying appropriately the arguments in Fan et al. (2000). The following theorem shows that one still obtains asymptotically correct frequentist coverage of $\widehat{\Theta}_\alpha$ for the IdS Θ_I, even though the posterior distribution of the QLR is asymptotically a gamma F_Γ.

Theorem C.1. *(i) Let Assumptions 3.1, C.2, 3.3, and 3.4 hold and let $\sup_{\theta \in \Theta_I} Q_n(\theta) \rightsquigarrow F_\Gamma$. 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 $\hat{\gamma}_n = \max_{1\leq i\leq n} X_i - \tilde{\gamma}$. Here we take $\Theta_{osn} = \{\theta : (1+\varepsilon_n)\hat{\gamma}_n \geq \gamma(\theta) \geq \hat{\gamma}_n\}$ 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 also that $r=1$ and that the sets $K_{osn} = \{n(\gamma(\theta)-\hat{\gamma}_n) : \theta\in\Theta_{osn}\} = \{n(\gamma-\hat{\gamma}_n) : (1+\varepsilon_n)\hat{\gamma}\geq\gamma\geq\hat{\gamma}_n\}$ cover $\mathbb{R}^+$. A smooth prior on Θ will induce a smooth prior on $\gamma(\theta)$, and the result follows.