Yale University

EliScholar – A Digital Platform for Scholarly Publishing at Yale

Cowles Foundation Discussion Papers | Cowles Foundation

5-1-2016

Monte Carlo Confidence Sets for Identified Sets

Xiaohong Chen

Timothy M. Christensen

Elie Tamer

Follow this and additional works at: https://elischolar.library.yale.edu/cowles-discussion-paper-series

Part of the Economics Commons

Recommended Citation

Chen, Xiaohong; Christensen, Timothy M.; and Tamer, Elie, "Monte Carlo Confidence Sets for Identified Sets" (2016). *Cowles Foundation Discussion Papers*. 2484.
https://elischolar.library.yale.edu/cowles-discussion-paper-series/2484

This Discussion Paper is brought to you for free and open access by the Cowles Foundation at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Cowles Foundation Discussion Papers by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

MONTE CARLO CONFIDENCE SETS FOR IDENTIFIED SETS

By

Xiaohong Chen, Timothy M. Christensen, and Elie Tamer

May 2016
Revised September 2017

COWLES FOUNDATION DISCUSSION PAPER NO. 2037R2

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

http://cowles.yale.edu/

Monte Carlo Confidence Sets for Identified Sets*

Xiaohong Chen† Timothy M. Christensen‡ Elie Tamer§

First draft: August 2015; Revised September 2017

Abstract

In complicated/nonlinear parametric models, it is generally hard to know whether the model parameters are point identified. We provide computationally attractive procedures to construct confidence sets (CSs) for identified sets of full parameters and of subvectors in models defined through a likelihood or a vector of moment equalities or inequalities. These CSs are based on level sets of optimal sample criterion functions (such as likelihood or optimally-weighted or continuously-updated GMM criterions). The level sets are constructed using cutoffs that are computed via Monte Carlo (MC) simulations directly from the quasi-posterior distributions of the criterions. We establish new Bernstein-von Mises (or Bayesian Wilks) type theorems for the quasi-posterior distributions of the quasi-likelihood ratio (QLR) and profile QLR in partially-identified regular models and some non-regular models. These results imply that our MC CSs have *exact* asymptotic frequentist coverage for identified sets of full parameters and of subvectors in partially-identified regular models, and have valid but potentially conservative coverage in models with reduced-form parameters on the boundary. Our MC CSs for identified sets of subvectors are shown to have *exact* asymptotic coverage in models with singularities. We also provide results on uniform validity of our CSs over classes of DGPs that include point and partially identified models. We demonstrate good finite-sample coverage properties of our procedures in two simulation experiments. Finally, our procedures are applied to two non-trivial empirical examples: an airline entry game and a model of trade flows.

*We are grateful to K. Hirano and three reviewers, S. Bonhomme, and L. P. Hansen for insightful suggestions. We also thank T. Cogley, K. Evdokimov, H. Hong, B. Honoré, M. Keane, K. Menzel, M. Kolesár, U. Müller, J. Montiel Olea, T. Sargent, F. Schorfheide, C. Sims and participants at the 2015 ESWC meetings in Montreal, the September 2015 "Big Data Big Methods" Cambridge-INET conference, and workshops at many institutions for useful comments. We thank Keith O'Hara for his excellent RA work on the MCMC simulations and empirical results reported in the first version.

†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

It is often difficult to verify whether parameters in complicated nonlinear structural models are globally point identified. This is especially the case when conducting a sensitivity analysis to examine the impact of various model assumptions on the estimates of parameters of interest, where relaxing some suspect assumptions may lead to loss of point identification. This difficulty of verifying point identification naturally calls for inference procedures that are valid whether or not the parameters of interest are point identified. Our goal is to contribute to this sensitivity literature by proposing relatively simple inference procedures that allow for partial identification in models defined through a likelihood or a vector of moment equalities or inequalities.

To that extent, we provide computationally attractive and asymptotically valid confidence set (CS) constructions for the identified set Θ_I of the full vector of parameters $\theta \equiv (\mu, \eta) \in \Theta$,[1] and for the identified sets M_I of subvectors μ. As a sensitivity check in an empirical study, a researcher could report conventional CSs based on inverting a t or Wald statistic, which are valid under point identification only, alongside our new CSs that are asymptotically optimal under point identification and robust to failure of point identification.

Our CS constructions are criterion-function based, as in Chernozhukov, Hong, and Tamer (2007) (CHT) and the subsequent literature on CSs for identified sets. That is, contour sets of the sample criterion function are used as CSs for Θ_I and contour sets of the sample profile criterion are used as CSs for M_I. However, our CSs are constructed using critical values that are calculated *differently* from those in the existing literature. In two of our proposed CS constructions, we estimate critical values using quantiles of the sample criterion function (or profile criterion) that are simulated from a quasi-posterior distribution, which is formed by combining the sample criterion function with a prior over the model parameter space Θ.[2]

We propose three procedures for constructing various CSs. To construct a CS for the identified set Θ_I, our Procedure 1 draws a sample $\{\theta^1, ..., \theta^B\}$ from the quasi-posterior, computes the *α-quantile of the sample criterion evaluated at the draws*, and then defines our CS $\widehat{\Theta}_\alpha$ for Θ_I as the contour set at said α-quantile. The computational complexity here is simply as hard as the problem of taking draws from the quasi-posterior, a well-researched and understood area in the literature on Monte Carlo (MC) algorithms in Bayesian computation (see, e.g., Liu (2004), Robert and Casella (2004)). Many MC samplers (including the popular Markov Chain Monte Carlo (MCMC) algorithms) could, in principle, be used for this purpose. In our simulations and empirical applications, we use an

[1] Following the literature, the identified set Θ_I is the argmax of a population criterion over the whole parameter space Θ. A model is point identified if Θ_I is a singleton, say $\{\theta_0\}$, and partially identified if $\{\theta_0\} \subsetneq \Theta_I \subsetneq \Theta$.

[2] In correctly-specified likelihood models the quasi-posterior is a true posterior distribution over Θ. We refer to the distribution as a quasi-posterior because we accommodate non-likelihood based models, such as moment-based models with GMM criterions.

adaptive sequential Monte Carlo (SMC) algorithm that is well-suited to drawing from irregular, multi-modal (quasi-)posteriors and is also easily parallelizable for fast computation (see, e.g., Herbst and Schorfheide (2014), Del Moral, Doucet, and Jasra (2012), Durham and Geweke (2014)). Our Procedure 2 produces a CS $\widehat{M}_\alpha$ for M_I of a general subvector using the same draws from the quasi-posterior as in Procedure 1. Here an added computation step is needed to obtain critical values that guarantee the exact asymptotic coverage for M_I. Finally, our Procedure 3 CS for M_I of a scalar subvector is simply the contour set of the profiled quasi-likelihood ratio (QLR) with its critical value being the α quantile of a chi-square distribution with one degree of freedom. Our Procedure 3 CS is simple to compute but is valid only for scalar subvectors.

Our CS constructions are valid for "optimal" criterions, which include (but are not limited to) correctly-specified likelihood models, GMM models with optimally-weighted or continuously-updated or GEL criterions,[3] or sandwich quasi-likelihoods. For point- or partially-identified regular models, optimal criterions correspond to criterions that satisfy a generalized information equality. But our optimal criterions also allow for some correctly-specified non-regular (or non-standard) models such as models with parameter-dependent support, an important feature of set identified models (see Appendix E). Our Procedure 1 and 2 CSs, $\widehat{\Theta}_\alpha$ and $\widehat{M}_\alpha$, are shown to have *exact* asymptotic coverage for Θ_I and M_I in potentially partially identified regular models, and are valid but possibly conservative in potentially partially identified models with reduced-form parameters on the boundary (in which the local tangent space is a convex cone). Our Procedure 1 and 2 CSs are also shown to be uniformly valid over DGPs that include both point- and partially identified models (see Appendix B). Moreover, our Procedure 2 CS is shown to have *exact* asymptotic coverage for M_I in models with singularities, which are particularly relevant in applications when parameters are close to point-identified or point-identified. Our Procedure 3 CS has *exact* asymptotic coverage in regular models that are point-identified[4]. Although theoretically slightly conservative in partially identified models, our Procedure 3 CS performs well in our simulations and empirical examples.

Our Procedure 1 and 2 CSs are Monte Carlo (MC) based. To establish their theoretical validity, we derive new Bernstein-von Mises (or Bayesian Wilks) type theorems for the (quasi-)posterior distributions of the QLR and profile QLR in partially identified models, allowing for regular models and some important non-regular cases (e.g. models in which the local tangent space is a convex cone, models with singularities, and models with parameter-dependent support). These theorems establish that the (quasi-)posterior distributions of the QLR and profile QLR converge to their frequentist counterparts in regular models; see Section 4 and Appendix E for similar results in some non-regular cases. As an illustration we briefly mention some results for Procedure 1 here: Section 4 presents conditions under which the sample QLR statistic and the (quasi-)posterior distribution

[3]Moment inequality-based models are special cases of moment equality-based models as one can add nuisance parameters to transform moment inequalities into moment equalities. Although moment inequality models are allowed, our criterion differs from the popular GMS criterion for moment inequalities in Andrews and Soares (2010) and others; see Subsections 3.1.1, 5.2.1 and 5.3.3.

[4]In fact, all three of our procedures are efficient in point-identified regular models.

of the QLR both converge to a chi-square distribution with unknown degree of freedom in regular models.[5] Appendix E shows that the QLR and the (quasi-)posterior of the QLR both converge to a gamma distribution with scale parameter of 2 and unknown shape parameter in more general partially-identified models. These results ensure that the quantiles of the QLR evaluated at the MC draws from its quasi-posterior consistently estimate the correct critical values needed for Procedure 1 CS to have exact asymptotic coverage for Θ_I. See Section 4 for similar results for the profile QLR and Procedure 2 CSs for M_I for subvectors.

We demonstrate the computational feasibility and good finite-sample coverage of our proposed methods in two simulation experiments: a missing data example and a complete information entry game with correlated payoff shocks. We use the missing data example to illustrate the conceptual difficulties in a transparent way, studying both numerically and theoretically the behaviors of our CSs when this model is partially-identified, close to point-identified, and point-identified. Although the length of a confidence interval for the identified set M_I of a scalar μ is by definition no shorter than that for μ itself, our simulations demonstrate that the differences in length between our Procedures 2 and 3 CSs for M_I and the GMS CSs of Andrews and Soares (2010) for μ are negligible. Finally, our CS constructions are applied to two real data examples: an airline entry game with correlated payoff shocks and an empirical trade flow model. The airline entry game example has 17 partially-identified structural parameters. Our empirical findings using Procedures 2 and 3 CSs show that the data are informative about some equilibrium selection probabilities. The trade example has 46 structural parameters. Here, point-identification may be difficult to verify, especially when conducting a sensitivity analysis of restrictive model assumptions.

Literature Review. Several papers have recently proposed Bayesian (or pseudo Bayesian) methods for constructing CSs for Θ_I that have correct frequentist coverage properties. See section 3.3 in 2009 NBER working paper version of Moon and Schorfheide (2012), Kitagawa (2012), Norets and Tang (2014), Kline and Tamer (2016), Liao and Simoni (2016) and the references therein. All these papers consider *separable* regular models and use various renderings of a similar intuition. First, there exists a finite-dimensional reduced-form parameter, say ϕ, that is (globally) point-identified and $\sqrt{n}$-consistently and asymptotically normal estimable from the data, and is linked to the model structural parameter θ via a *known* global mapping. Second, a prior is placed on the reduced-form parameter ϕ, and third, a classical Bernstein-von Mises theorem stating the asymptotic normality of the posterior distribution for ϕ is assumed to hold. Finally, the known global mapping between the reduced-form and the structural parameters is inverted, which, by step 3, guarantees correct coverage for Θ_I in large samples. In addition to this literature's focus on separable models, it is not clear whether the results there remain valid in various non-regular models we study.

[5] In point-identified models, Wilks-type asymptotics imply the degree of freedom is equal to the dimension of θ for QLR statistics. In partially identified models, the degree of freedom is some d^*, typically less than or equal to $\dim(\theta)$. The correct d^* may not be easy to infer from the context, which is why we refer to it as "unknown".

Our approach is valid regardless of whether the model is separable or not. We show that for general separable or non-separable partially identified likelihood or moment-based models, a *local reduced-form reparameterization* exists (see Section 5). We use this local reparameterization as a proof device to show that the (quasi-)posterior distributions of the QLR and the profile QLR statistics have a frequentist interpretation in large samples. Importantly, since our Procedures 1 and 2 impose priors on the model parameter θ only, there is no need for obtaining a global reduced-form reparameterization or deriving its dimension to implement our procedures. This is in contrast with the above-mentioned existing Bayesian methods for partially identified separable models, for which researchers need to impose priors on global reduced-form parameters ϕ to ensure that its posterior lies on $\{\phi(\theta) : \theta \in \Theta\}$ (i.e. the set of reduced-form parameters consistent with the structural model), which could be difficult even in some empirically relevant separable models; see the airline entry game application in Section 3.2. Moreover, our new Bernstein-von Mises type theorems for the (quasi-)posterior distributions of the QLR and profile QLR allow for several important non-regular cases in which the local reduced-form parameter is typically not $\sqrt{n}$-consistent and asymptotically normally estimable.

When specialized to point- or partially-identified likelihood models, our Procedure 1 CS for Θ_I is equivalent to Bayesian credible set for θ based on inverting a LR statistic. With flat priors, these CSs are also the highest posterior density (HPD) credible sets. Our general theoretical results imply that HPD credible sets give correct frequentist coverage in partially identified regular models and conservative coverage in some non-standard circumstances. These findings complement those of Moon and Schorfheide (2012) who showed that HPD credible sets can under-cover (in a frequentist sense) in separable partially identified regular models under their conditions.[6] In point-identified regular models satisfying a generalized information equality with $\sqrt{n}$-consistent and asymptotically normally estimable parameters θ, Chernozhukov and Hong (2003) (CH hereafter) propose constructing CSs for scalar subvectors μ by taking the upper and lower quantiles of the MCMC draws $\{\mu^1, \ldots, \mu^B\}$ where $(\mu^b, \eta^b) \equiv \theta^b$. Our CS constructions for scalar subvectors are asymptotically equivalent to CH's CSs in such models, but they differ otherwise. Our CS constructions, which are based on quantiles of the *criterion* evaluated at the MC draws $\{\theta^1, \ldots, \theta^B\}$ rather than of the raw parameter draws themselves, are valid irrespective of whether the model is point- or partially-identified. Intuitively, this is because the population criterion is always point-identified irrespective of whether θ is point- or partially-identified.

There are several published works on frequentist CS constructions for Θ_I: see, e.g., 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 Θ_I is strictly convex. For inference on identified sets of subvectors, both the subsampling-based papers

[6] Note that this is not a contradiction since our priors are imposed on structural parameters θ only, violating Assumption 2 in Moon and Schorfheide (2012).

of CHT and Romano and Shaikh (2010) deliver valid tests with a judicious choice of the subsample size for a profiled criterion function. The subsampling-based CS construction allows for general criterion functions, but is computationally demanding and sensitive to choice of subsample size in realistic empirical structural models.[7] Our methods are computationally attractive and typically have asymptotically correct coverage, but require "optimal" criterion functions.

The rest of the paper is organized as follows. Section 2 describes our new procedures for CSs for identified sets Θ_I and M_I. Section 3 presents simulations and real data applications. Section 4 first establishes new BvM (or Bayesian Wilks) results for the QLR and profile QLR in partially identified models. It then derives the frequentist validity of our CSs. Section 5 provides some sufficient conditions to the key regularity conditions for the general theory in Section 4. Section 6 briefly concludes. Appendix A describes the implementation details for the simulations and real data applications in Section 3. Appendix B shows that our CSs for Θ_I and M_I are valid uniformly over a class of DGPs. Appendix C verifies the main regularity conditions for uniform validity in the missing data and a moment inequality examples. Appendix D presents results on local power. Appendix E establishes a new BvM (or Bayesian Wilks) result which shows that the limiting (quasi-)posterior distribution of the QLR in a partially identified model is a gamma distribution with unknown shape parameter and scale parameter of 2. There, results on models with parameter-dependent support are given. Appendix F contains all the proofs and additional lemmas.

2 Description of our Procedures

In this section we first describe our method for constructing CSs for Θ_I. We then describe methods for constructing CSs for M_I of any subvector. We finally present an extremely simple method for constructing CSs for M_I of a scalar subvector in certain situations.

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. Consider a population objective function $L : \Theta \to \mathbb{R}$, such as a log-likelihood function for correctly specified likelihood models, an optimally-weighted or continuously-updated GMM objective function, or a sandwich quasi-likelihood function. The function L is assumed to be an upper semi-continuous function of θ with $\sup_{\theta\in\Theta} L(\theta) < \infty$. 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:

$$\Theta_I := \{\theta \in \Theta : L(\theta) = \sup_{\vartheta\in\Theta} L(\vartheta)\} \ . \tag{1}$$

[7] There is a large literature on frequentist approach for *inference on the true parameter* $\theta \in \Theta_I$ or $\mu \in M_I$ (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 uses discontinuous-in-parameters asymptotic (repeated sampling) approximations to test statistics. These existing frequentist methods are difficult to implement in realistic empirical models.

The set Θ_I is our first object of interest. In many applications, it may be of interest to provide a CS for a *subvector* of interest. Write $\theta \equiv (\mu, \eta)$ where μ is the subvector of interest and η is a nuisance parameter. Our second object of interest is the identified set for the subvector μ:

$$M_I := \{\mu : (\mu, \eta) \in \Theta_I \text{ for some } \eta\} . \tag{2}$$

Given the data $\mathbf{X}_n$, we seek to construct computationally attractive CSs that cover Θ_I or M_I with a pre-specified probability (in repeated samples) as sample size n gets large.

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 can be a natural sample analogue 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 $p_\theta(.)$ (with respect to some dominating measure), the identified set is $\Theta_I = \{\theta \in \Theta : P_0 = P_\theta\}$ where P_0 is the true data distribution. We take L_n to be the average log-likelihood function:

$$L_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \log p_\theta(X_i) . \tag{3}$$

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

$$L_n(\theta) = -\frac{1}{2} \rho_n(\theta)' W_n(\theta) \rho_n(\theta) \tag{4}$$

where $\rho_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \rho_\theta(X_i)$ and $W_n(\theta) = \left(\frac{1}{n} \sum_{i=1}^{n} \rho_\theta(X_i)\rho_\theta(X_i)' - \rho_n(\theta)\rho_n(\theta)'\right)^{-}$ (the superscript $^{-}$ denotes generalized inverse) for iid data or other suitable choices. Given an optimal weighting matrix $\widehat{W}_n$, we could also use an optimally-weighted GMM objective function:

$$L_n(\theta) = -\frac{1}{2} \rho_n(\theta)' \widehat{W}_n \rho_n(\theta) . \tag{5}$$

Generalized empirical likelihood objective functions could also be used with our procedures.

Our main CS constructions (Procedures 1 and 2 below) are based on Monte Carlo (MC) simulation methods from a quasi-posterior. Given L_n and a prior Π over Θ, the quasi-posterior distribution Π_n for θ given $\mathbf{X}_n$ is defined as

$$\mathrm{d}\Pi_n(\theta|\mathbf{X}_n) = \frac{e^{nL_n(\theta)} \mathrm{d}\Pi(\theta)}{\int_\Theta e^{nL_n(\theta)} \mathrm{d}\Pi(\theta)} . \tag{6}$$

Our procedures 1 and 2 require drawing a sample $\{\theta^1, \ldots, \theta^B\}$ from the quasi-posterior Π_n. In

practice we use an adaptive sequential Monte Carlo (SMC) algorithm which is known to be well suited to drawing from irregular, multi-modal distributions, but any MC sampler could, in principle, be used. The SMC algorithm is described in detail in Appendix A.1.

2.1 Confidence sets for the identified set Θ_I

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

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

[Procedure 1: Confidence sets for the identified set]

1. Draw a sample $\{\theta^1, \ldots, \theta^B\}$ from the quasi-posterior distribution Π_n in (6).
2. Calculate the $(1-\alpha)$ quantile of $\{L_n(\theta^1), \ldots, L_n(\theta^B)\}$; call it $\zeta^{mc}_{n,\alpha}$.
3. Our $100\alpha\%$ confidence set for Θ_I is then:

$$\widehat{\Theta}_\alpha = \{\theta \in \Theta : L_n(\theta) \geq \zeta^{mc}_{n,\alpha}\}\,. \tag{7}$$

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 MC draws $\{\theta^1, \ldots, \theta^B\}$ will concentrate around Θ_I and thereby indicate the regions in Θ over which to search.

CHT considered inference on the set of minimizers of a *nonnegative population criterion function* $Q : \Theta \to \mathbb{R}_+$ using a sample analogue Q_n of Q. Let $\xi_{n,\alpha}$ denote a consistent estimator of the α quantile of $\sup_{\theta\in\Theta_I} Q_n(\theta)$. The $100\alpha\%$ CS for Θ_I at level $\alpha \in (0,1)$ proposed is $\widehat{\Theta}^{CHT}_\alpha = \{\theta \in \Theta : Q_n(\theta) \leq \xi_{n,\alpha}\}$. In the existing literature, subsampling or bootstrap based methods have been used to compute $\xi_{n,\alpha}$ which can be tedious to implement. Instead, our procedure replaces $\xi_{n,\alpha}$ with a cut off based on Monte Carlo simulations. 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.

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:

$$Q_n(\theta) = 2n[L_n(\hat{\theta}) - L_n(\theta)]\,. \tag{8}$$

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 (7) because $L_n(\theta) \geq \zeta_{n,\alpha}^{mc}$ if and only if $Q_n(\theta) \leq \xi_{n,\alpha}^{mc}$.

In Procedure 1 and Remark 1 above, the posterior-like quantity involves the use of a prior distribution Π over Θ. This prior is user chosen and typically would be the uniform prior but other choices are possible. In our simulations, various choices of prior did not matter much, unless they assigned extremely small mass near the true parameter (which is avoided by using a uniform prior whenever Θ is compact).

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

Lemma 2.1. *Let (i) $\sup_{\theta \in \Theta_I} Q_n(\theta) \rightsquigarrow W$ where W is a random variable for which F_W is continuous at 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 MC CSs for Θ_I are shown to be valid by verifying parts (i) and (ii) with $w_{n,\alpha} = \xi_{n,\alpha}^{mc}$. To verify part (ii), we shall establish a new Bernstein-von Mises (BvM) (or a new Bayesian Wilks) type result for the quasi-posterior distribution of the QLR under loss of identifiability.

2.2 Confidence sets for the identified set M_I of subvectors

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 for M_I. The projection CS:

$$\widehat{M}_\alpha^{proj} = \{\mu : (\mu, \eta) \in \widehat{\Theta}_\alpha \text{ for some } \eta\} \tag{9}$$

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 documented, $\widehat{M}_\alpha^{proj}$ is typically conservative, and especially so when the dimension of μ is small relative to the dimension of θ. Indeed, our simulations below indicate that $\widehat{M}_\alpha^{proj}$ is very conservative even in reasonably low-dimensional parametric models.

We propose CSs for M_I based on a profile criterion for M_I. Let $M = \{\mu : (\mu, \eta) \in \Theta \text{ for some } \eta\}$ and $H_\mu = \{\eta : (\mu, \eta) \in \Theta\}$. The profile criterion for a point $\mu \in M$ is $\sup_{\eta \in H_\mu} L_n(\mu, \eta)$, and the profile criterion for M_I is

$$PL_n(M_I) \equiv \inf_{\mu \in M_I} \sup_{\eta \in H_\mu} L_n(\mu, \eta). \tag{10}$$

Let $\Delta(\theta^b)$ be an equivalence set for θ^b. In likelihood models we define $\Delta(\theta^b) = \{\theta \in \Theta : p_\theta = p_{\theta^b}\}$ and in moment-based models we define $\Delta(\theta^b) = \{\theta \in \Theta : E[\rho(X_i, \theta)] = E[\rho(X_i, \theta^b)]\}$. Let $M(\theta^b) = \{\mu : (\mu, \eta) \in \Delta(\theta^b) \text{ for some } \eta\}$, and the profile criterion for $M(\theta^b)$ is

$$PL_n(M(\theta^b)) \equiv \inf_{\mu \in M(\theta^b)} \sup_{\eta \in H_\mu} L_n(\mu, \eta)\,. \tag{11}$$

[Procedure 2: Confidence sets for subvectors]

1. Draw a sample $\{\theta^1, \ldots, \theta^B\}$ from the quasi-posterior distribution Π_n in (6).
2. Calculate the $(1-\alpha)$ quantile of $\left\{PL_n(M(\theta^b)) : b = 1, \ldots, B\right\}$; call it $\zeta_{n,\alpha}^{mc,p}$.
3. Our $100\alpha\%$ confidence set for M_I is then:

$$\widehat{M}_\alpha = \left\{\mu \in M : \sup_{\eta \in H_\mu} L_n(\mu, \eta) \geq \zeta_{n,\alpha}^{mc,p}\right\}. \tag{12}$$

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 subvectors of the draws, say $\{\mu^1, \ldots, \mu^B\}$, concentrate around M_I, thereby indicating the region in M over which to search.

Remark 2. *Recall the definition of the QLR Q_n in (8), we define the profile QLR for the set $M(\theta^b)$ analogously as*

$$PQ_n(M(\theta^b)) \equiv 2n[L_n(\hat{\theta}) - PL_n(M(\theta^b))] \;=\; \sup_{\mu \in M(\theta^b)} \inf_{\eta \in H_\mu} Q_n(\mu, \eta)\,. \tag{13}$$

Let $\xi_{n,\alpha}^{mc,p}$ denote the α quantile of the profile QLR draws $\left\{PQ_n(M(\theta^b)) : b = 1, \ldots, B\right\}$. The confidence set:

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

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

Our Procedure 2 and Remark 2 above are *different* from taking quantiles of the MC parameter draws. A popular percentile CS (denoted as $\widehat{M}_\alpha^{perc}$) for a scalar subvector μ is computed by taking the upper and lower $100(1-\alpha)/2$ percentiles of $\{\mu^1, \ldots, \mu^B\}$. For point-identified regular models with $\sqrt{n}$-consistent and asymptotically normal parameters θ, this approach is known to be valid for correctly-specified likelihood models in the standard Bayesian literature and its validity for criterion-based models satisfying a generalized information equality has been established by Chernozhukov and Hong (2003). However, in partially identified models this approach is no longer valid and under-covers, as evidenced in the simulation results below.

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

Lemma 2.2. *Let (i)* $\sup_{\mu \in M_I} \inf_{\eta \in H_\mu} Q_n(\mu, \eta) \rightsquigarrow W$ *where* W *is a random variable for which* F_W *is continuous at* 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:*

$$\widehat{M}_\alpha = \left\{ \mu \in M : \inf_{\eta \in H_\mu} Q_n(\mu, \eta) \leq w_{n,\alpha} \right\}.$$

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 MC 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 shall derive a new BvM type result for the quasi-posterior of the profile QLR under loss of identifiability.

2.3 A simple but slightly conservative CS for M_I of scalar subvectors

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

[Procedure 3: Simple conservative CSs for scalar subvectors]

1. Calculate a maximizer $\hat{\theta}$ for which $L_n(\hat{\theta}) \geq \sup_{\theta \in \Theta} L_n(\theta) + o_{\mathbb{P}}(n^{-1})$.
2. Our $100\alpha\%$ confidence set for $M_I \subset \mathbb{R}$ is then:

$$\widehat{M}_\alpha^\chi = \left\{ \mu \in M : \inf_{\eta \in H_\mu} Q_n(\mu, \eta) \leq \chi^2_{1,\alpha} \right\} \tag{14}$$

where Q_n is the QLR in (8) and $\chi^2_{1,\alpha}$ denotes the α quantile of the χ^2_1 distribution.

Procedure 3 above is justified when the limit distribution of the profile QLR for M_I is stochastically dominated by the χ^2_1 distribution (i.e., $F_W(z) \geq F_{\chi^2_1}(z)$ for all $z \geq 0$ in Lemma 2.2). This allows for computationally simple construction using repeated evaluations on a scalar grid. Unlike $\widehat{M}_\alpha$, the CS $\widehat{M}^\chi_\alpha$ for M_I is typically asymptotically conservative and is only valid for scalar functions of Θ_I (see Section 4.3). Nevertheless, the CS $\widehat{M}^\chi_\alpha$ is asymptotically exact when M_I happens to be a singleton belonging to the interior of M, and, for confidence levels of $\alpha \geq 0.85$, its degree of conservativeness for the set M_I is negligible (see Section 4.3). It is extremely simple to implement and performs very favorably in simulations. As a sensitivity check in empirical estimation of a complicated structural model, one could report the conventional CS based on a t-statistic (that is valid under point identification only) as well as our CS $\widehat{M}^\chi_\alpha$ (that remains valid under partial identification); see Section 3.2.

3 Simulation Evidence and Empirical Applications

This section presents simulation evidence and empirical applications to demonstrate the good performances of our new procedures for general possibly partially identified models. See Appendix A for implementation details.

3.1 Simulation evidence

In this subsection we investigate the finite-sample behavior of our proposed CSs in two leading examples of partially identified models: missing data and entry game with correlated payoff shocks. Both 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.

We use samples of size $n = 100$, 250, 500, and 1000. For each sample, we calculate the posterior quantile of the QLR or profile QLR statistic using $B = 10000$ draws from an adaptive SMC algorithm (see Appendix A.1 for a description of the algorithm).

3.1.1 Example 1: missing data

We first consider the simple but insightful missing data example. Suppose we observe a random sample $\{(D_i, Y_i D_i)\}_{i=1}^n$ where both the outcome variable Y_i and the selection variable D_i take values in $\{0, 1\}$. The parameter of interest is the true mean $\mu_0 = \mathbb{E}[Y_i]$. Without further assumptions, μ_0 is not point identified when $\Pr(D_i = 0) > 0$ as we only observe Y_i when $D_i = 1$.

Denote the true probabilities of observing $(D_i, Y_i D_i) = (1, 1)$, $(0, 0)$ and $(1, 0)$ by $\tilde{\gamma}_{11}$, $\tilde{\gamma}_{00}$, and

$\tilde{\gamma}_{10} = 1 - \tilde{\gamma}_{11} - \tilde{\gamma}_{00}$ respectively. We view $\tilde{\gamma}_{00}$ and $\tilde{\gamma}_{11}$ as true *reduced-form parameters* that are consistently estimable. The reduced-form parameters are functions of the structural parameter $\theta = (\mu, \eta_1, \eta_2)$ where $\mu = \mathbb{E}[Y_i]$, $\eta_1 = \Pr(Y_i = 1|D_i = 0)$, and $\eta_2 = \Pr(D_i = 1)$. Under this model parameterization, θ is related to the reduced form parameters via $\tilde{\gamma}_{00}(\theta) = 1 - \eta_2$ and $\tilde{\gamma}_{11}(\theta) = \mu - \eta_1(1 - \eta_2)$. The parameter space Θ for θ is defined as:

$$\Theta = \{(\mu, \eta_1, \eta_2) \in [0, 1]^3 : 0 \leq \mu - \eta_1(1 - \eta_2) \leq \eta_2\}. \tag{15}$$

The identified set for θ is:

$$\Theta_I = \{(\mu, \eta_1, \eta_2) \in \Theta : \tilde{\gamma}_{00} = 1 - \eta_2, \tilde{\gamma}_{11} = \mu - \eta_1(1 - \eta_2)\}. \tag{16}$$

Here, η_2 is point-identified but only an affine combination of μ and η_1 are identified. The identified set for $\mu = E[Y_i]$ is:

$$M_I = [\tilde{\gamma}_{11}, \tilde{\gamma}_{11} + \tilde{\gamma}_{00}]$$

and the identified set for the nuisance parameter η_1 is $[0, 1]$.

We set the true values of the parameters to be $\mu = 0.5$, $\eta_1 = 0.5$, and take $\eta_2 = 1 - c/\sqrt{n}$ for $c = 0, 1, 2$ to cover both partially-identified but "drifting-to-point-identification" ($c = 1, 2$) and point-identified ($c = 0$) cases. We first implement the procedures using a likelihood criterion and a flat prior on Θ. The likelihood function of $(D_i, Y_i D_i) = (d, yd)$ is

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

In Appendix A we present and discuss additional results for a likelihood criterion with a curved prior and a continuously-updated GMM criterion based on the moments $E[1\!\!1\{D_i = 0\} - \tilde{\gamma}_{00}(\theta)] = 0$ and $E[1\!\!1\{(D_i, Y_i D_i) = (1, 1)\} - \tilde{\gamma}_{11}(\theta)] = 0$ with a flat prior (this GMM case may be interpreted as a moment inequality model with $\eta_1(1 - \eta_2)$ playing the role of a slackness parameter).

We implement the SMC algorithm as described in Appendix A.2. To illustrate sampling via the SMC algorithm and the resulting posterior of the QLR, Figure 1 displays histograms of the draws for μ, η_1 and η_2 for one run of the adaptive SMC algorithm for a sample of size 1000 with $\eta_2 = 0.8$. Here μ is partially identified with $M_I = [0.4, 0.6]$. The histograms in Figure 1 show that the draws for μ and η_1 are both approximately flat across their identified sets. In contrast, the draws for η_2, which is point identified, are approximately normally distributed and centered at the MLE. The Q-Q plot in Figure 1 shows that the quantiles of $Q_n(\theta)$ computed from the draws are very close to the quantiles of a χ^2_2 distribution, as predicted by our theoretical results below.

Confidence sets for Θ_I: The top panel of Table 1 displays MC coverage probabilities of $\widehat{\Theta}_\alpha$ for 5000 replications. The MC coverage probability should be equal to its nominal value in large samples when $\eta_2 < 1$ (see Theorem 4.1). It is perhaps surprising that the nominal and MC coverage

	$\eta_2 = 1 - \frac{2}{\sqrt{n}}$						$\eta_2 = 1 - \frac{1}{\sqrt{n}}$						$\eta_2 = 1$ (Point ID)					
	0.90		0.95		0.99		0.90		0.95		0.99		0.90		0.95		0.99	
	$\widehat{\Theta}_\alpha$ (Procedure 1)																	
100	.910	—	.957	—	.994	—	.903	—	.953	—	.993	—	.989	—	.997	—	1.000	—
250	.901	—	.947	—	.991	—	.912	—	.955	—	.992	—	.992	—	.997	—	1.000	—
500	.913	—	.956	—	.991	—	.908	—	.957	—	.991	—	.995	—	.997	—	.999	—
1000	.910	—	.958	—	.992	—	.911	—	.958	—	.994	—	.997	—	.999	—	1.000	—
	$\widehat{M}_\alpha$ (Procedure 2)																	
100	.920	[.32,.68]	.969	[.30,.70]	.997	[.27,.73]	.918	[.37,.63]	.964	[.35,.65]	.994	[.32,.68]	.911	[.42,.59]	.958	[.40,.60]	.990	[.37,.63]
250	.917	[.39,.61]	.961	[.38,.62]	.992	[.36,.64]	.920	[.42,.58]	.963	[.41,.59]	.991	[.39,.61]	.915	[.45,.55]	.959	[.44,.56]	.991	[.42,.58]
500	.914	[.42,.58]	.961	[.41,.59]	.993	[.40,.60]	.914	[.44,.56]	.958	[.43,.57]	.992	[.42,.58]	.916	[.46,.54]	.959	[.46,.54]	.990	[.44,.56]
1000	.917	[.44,.56]	.956	[.44,.56]	.993	[.43,.57]	.914	[.46,.54]	.955	[.45,.55]	.993	[.44,.56]	.916	[.47,.53]	.959	[.47,.53]	.992	[.46,.54]
	$\widehat{M}_\alpha^\chi$ (Procedure 3)																	
100	.920	[.32,.68]	.952	[.31,.69]	.990	[.28,.72]	.916	[.37,.63]	.946	[.36,.64]	.989	[.33,.67]	.902	[.42,.58]	.937	[.41,.60]	.986	[.38,.63]
250	.915	[.39,.61]	.952	[.38,.62]	.990	[.36,.64]	.914	[.42,.58]	.954	[.41,.59]	.990	[.39,.61]	.883	[.45,.55]	.949	[.44,.56]	.991	[.42,.58]
500	.894	[.42,.58]	.954	[.41,.59]	.989	[.40,.60]	.906	[.44,.56]	.949	[.44,.56]	.990	[.42,.58]	.899	[.46,.54]	.945	[.46,.54]	.988	[.44,.56]
1000	.909	[.44,.56]	.950	[.44,.56]	.993	[.43,.57]	.904	[.46,.54]	.954	[.45,.55]	.989	[.45,.55]	.906	[.48,.52]	.946	[.47,.53]	.991	[.46,.54]
	$\widehat{M}_\alpha^{proj}$ (Projection)																	
100	.972	[.30,.70]	.990	[.28,.71]	.999	[.25,.75]	.969	[.35,.65]	.989	[.33,.67]	.998	[.30,.70]	.989	[.37,.63]	.997	[.36,.64]	1.000	[.33,.67]
250	.971	[.37,.63]	.986	[.36,.64]	.998	[.34,.66]	.976	[.40,.60]	.988	[.39,.61]	.998	[.37,.63]	.992	[.42,.58]	.997	[.41,.59]	1.000	[.39,.61]
500	.972	[.41,.59]	.985	[.40,.60]	.999	[.39,.61]	.972	[.43,.57]	.989	[.42,.58]	.999	[.41,.59]	.995	[.44,.56]	.997	[.43,.57]	.999	[.42,.58]
1000	.973	[.44,.56]	.990	[.43,.57]	.999	[.42,.58]	.973	[.45,.55]	.988	[.45,.55]	.999	[.44,.56]	.997	[.45,.55]	.999	[.45,.55]	1.000	[.44,.56]
	$\widehat{M}_\alpha^{perc}$ (Percentile)																	
100	.416	[.38,.62]	.676	[.36,.64]	.945	[.32,.68]	.661	[.40,.59]	.822	[.39,.61]	.963	[.35,.65]	.896	[.42,.58]	.946	[.40,.60]	.989	[.38,.63]
250	.402	[.42,.58]	.669	[.41,.59]	.917	[.38,.62]	.662	[.44,.56]	.822	[.43,.57]	.960	[.41,.59]	.899	[.45,.55]	.950	[.44,.56]	.990	[.42,.58]
500	.400	[.44,.56]	.652	[.43,.57]	.914	[.42,.58]	.652	[.46,.54]	.812	[.45,.55]	.955	[.43,.57]	.903	[.46,.54]	.953	[.46,.54]	.988	[.44,.56]
1000	.405	[.46,.54]	.671	[.45,.55]	.917	[.44,.56]	.662	[.47,.53]	.819	[.46,.54]	.953	[.45,.55]	.905	[.47,.53]	.953	[.47,.53]	.990	[.46,.54]
	Comparison with GMS CSs for μ via moment inequalities																	
100	.815	[.34,.66]	.908	[.32,.68]	.981	[.29,.71]	.803	[.39,.61]	.904	[.37,.63]	.980	[.34,.66]	.889	[.42,.58]	.938	[.40,.60]	.973	[.39,.62]
250	.798	[.40,.60]	.899	[.39,.61]	.979	[.36,.63]	.811	[.43,.57]	.897	[.42,.58]	.980	[.40,.60]	.896	[.45,.55]	.944	[.44,.56]	.981	[.42,.57]
500	.794	[.43,.57]	.898	[.42,.58]	.976	[.40,.60]	.789	[.45,.55]	.892	[.44,.56]	.975	[.43,.57]	.897	[.46,.54]	.948	[.46,.54]	.986	[.45,.55]
1000	.802	[.45,.55]	.900	[.44,.56]	.978	[.43,.57]	.812	[.46,.54]	.900	[.46,.54]	.978	[.45,.55]	.898	[.47,.53]	.949	[.47,.53]	.990	[.46,.54]

Table 1: Missing data example: average coverage probabilities for Θ_I and M_I and average lower and upper bounds of CSs for M_I across 5000 MC replications. Procedures 1–3, Projection and Percentile are implemented using a likelihood criterion and flat prior.

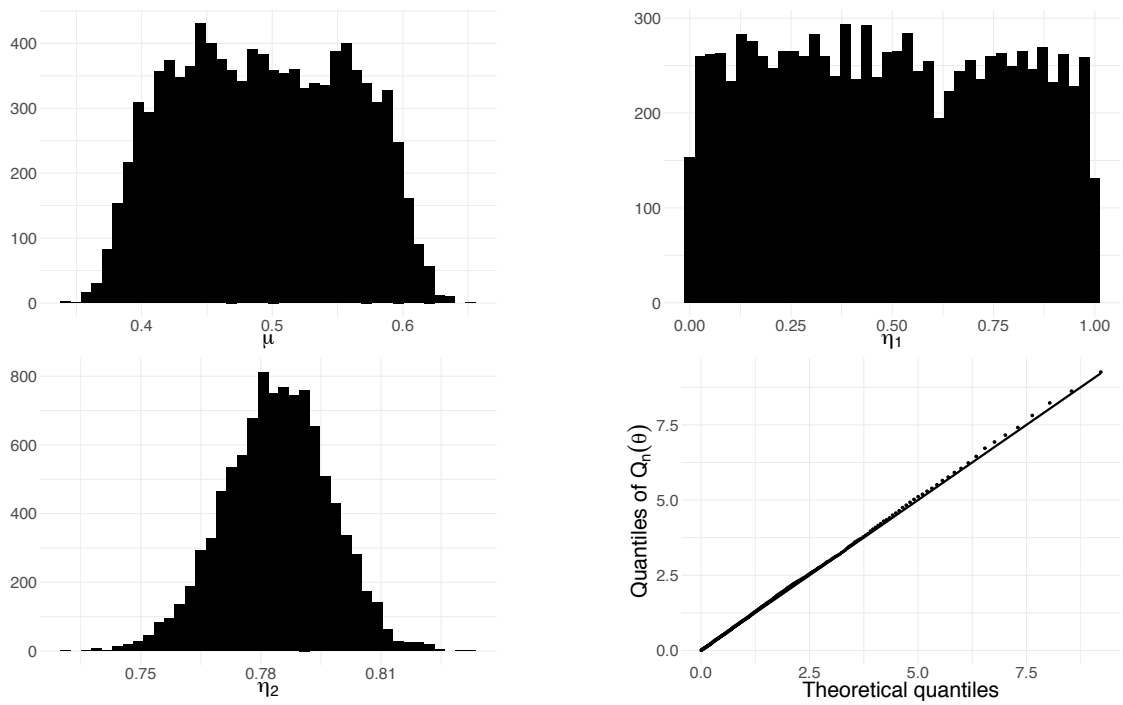

Figure 1: Missing data example: histograms of the SMC draws for μ (top left), η_1 (top right), and η_2 (bottom left) and Q-Q plot of $Q_n(\theta)$ computed from the draws against χ^2_2 quantiles (bottom right) for a sample of size $n = 1000$ with $\eta_2 = 0.8$. The identified sets for μ and η_1 are $[0.4, 0.6]$ and $[0, 1]$, respectively.

probabilities are close even in samples as small as $n = 100$. When $\eta_2 = 1$ the CSs for Θ_I are conservative, as predicted by our theoretical results (see Theorem 4.2).

Confidence sets for M_I: We now consider various CSs for the identified set M_I for μ. We first compute the projection CS $\widehat{M}_\alpha^{proj}$, as defined in (9), for M_I. As we can see from Table 1, this results in conservative CSs for M_I. For example, when $\alpha = 0.90$ the projection CSs cover M_I in around 97% of repeated samples. As the models with $c = 1, 2$ are close to point-identified, one might be tempted to report simple percentile CSs $\widehat{M}_\alpha^{perc}$ for M_I using Chernozhukov and Hong (2003) procedure, which is valid under point identification, and taking the upper and lower $100(1-\alpha)/2$ quantiles from of the draws for μ.[8] The results in Table 1 show that $\widehat{M}_\alpha^{perc}$ has correct coverage when μ is point identified (i.e. $\eta_2 = 1$) but it under-covers when μ is not point identified. For instance, the coverage probabilities of 90% CSs for M_I are about 66% with $c = 1$.

In contrast, our criterion-based procedures 2 and 3 remain valid under partial identification. We show below (see Theorem 4.3) that the coverage probabilities of our procedure 2 CS $\widehat{M}_\alpha$ (for M_I) should be equal to their nominal values α when n is large irrespective of whether the model is

[8]Note that we use exactly the same draws for implementing the percentile CS and procedures 1 and 2. As the SMC algorithm uses a particle approximation to the posterior, in practice we compute posterior quantiles for μ using the particle weights in a manner similar to (28).

partially identified with (i.e. $\eta_2 < 1$) or point identified (i.e. $\eta_2 = 1$). The results in Table 1 show that this is indeed the case, and that the coverage probabilities for procedure 2 are close to their nominal level even for small values of n, irrespective of whether the model is point- or partially-identified. In Section 5.3.1, we show that the asymptotic distribution of the profile QLR for M_I is stochastically dominated by the χ^2_1 distribution. Table 1 also presents results for procedure 3 using $\widehat{M}^{\chi}_{\alpha}$ as in (14). 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 η_2.

Finally, we compare our the length of CSs for M_I using procedures 2 and 3 with the length of CSs for the parameter μ constructed using the generalized moment selection (GMS) procedure of Andrews and Soares (2010). We implement their procedure using the inequalities

$$E[\mu - Y_i D_i] \geq 0 \qquad\qquad E[Y_i D_i + (1 - D_i) - \mu] \geq 0 \tag{17}$$

with their smoothing parameter $\kappa_n = (\log n)^{1/2}$, their GMS function $\varphi_j^{(1)}$, and with critical values computed via a multiplier bootstrap. Of course, GMS CSs are for the parameter μ rather than the set M_I, which is why the coverage for M_I reported in Table 1 appears lower than nominal under partial identification (GMS CSs are known to be asymptotically valid CSs for μ). Importantly, the average lower and upper bounds of our CSs for M_I constructed using Procedures 2 and 3 are very close to those using GMS, whereas projection-based CSs are, in turn, larger. On the other hand, CSs computed using percentiles of the draws for μ are narrower.

3.1.2 Example 2: entry game with correlated payoff shocks

We now consider the complete information entry game example described in Table 2. 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. We also assume that Δ_1 and Δ_2 are both negative and that players play a pure strategy Nash equilibrium. When $-\beta_j \leq \epsilon_j \leq -\beta_j - \Delta_j$, $j = 1, 2$, the game has two equilibria: for given values of the epsilons in this region, the model predicts $(1, 0)$ *and* $(0, 1)$. Let $D_{a_1 a_2}$ denote a binary random variable taking the value 1 if and only if player 1 takes action a_1 and player 2 takes action a_2. We observe a random sample of $\{(D_{00,i}, D_{10,i}, D_{01,i}, D_{11,i})\}_{i=1}^n$. So the data provides information of four choice probabilities $(P(0,0), P(1,0), P(0,1), P(1,1))$, but there are six parameters that need to be estimated: $\theta = (\beta_1, \beta_2, \Delta_1, \Delta_1, \rho, s)$ where $s \in [0, 1]$ is the equilibrium selection probability. The model parameter is partially identified as we have 3 non-redundant choice probabilities from which we need to learn about 6 parameters.

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

Table 2: Payoff matrix for the binary entry game. The first entry in each cell is the payoff to player 1 and the second entry is the payoff to player 2.

We can link the choice probabilities (reduced-form parameters) to θ via:

$$\begin{aligned}
\tilde{\gamma}_{00}(\theta) :=& Q_\rho(\epsilon_1 \leq -\beta_1;\, \epsilon_2 \leq -\beta_2) \\
\tilde{\gamma}_{11}(\theta) :=& Q_\rho(\epsilon_1 \geq -\beta_1 - \Delta_1;\, \epsilon_2 \geq -\beta_2 - \Delta_2) \\
\tilde{\gamma}_{10}(\theta) :=& s \times Q_\rho(-\beta_1 \leq \epsilon_1 \leq -\beta_1 - \Delta_1;\, -\beta_2 \leq \epsilon_2 \leq -\beta_2 - \Delta_2) \\
&+ Q_\rho(\epsilon_1 \geq -\beta_1; \epsilon_2 \leq -\beta_2) + Q_\rho(\epsilon_1 \geq -\beta_1 - \Delta_1; -\beta_2 \leq \epsilon_2 \leq -\beta_2 - \Delta_2)
\end{aligned}$$

and $\tilde{\gamma}_{01}(\theta) = 1 - \tilde{\gamma}_{00}(\theta) - \tilde{\gamma}_{11}(\theta) - \tilde{\gamma}_{10}(\theta)$, where Q_ρ denotes the joint probability distribution of (ϵ_1, ϵ_2) indexed by the correlation parameter ρ. Let $(\tilde{\gamma}_{00}, \tilde{\gamma}_{10}, \tilde{\gamma}_{01}, \tilde{\gamma}_{11})$ denote the true choice probabilities $(P(0,0), P(1,0), P(0,1), P(1,1))$. This naturally suggests a likelihood approach, where the likelihood of $(D_{00,i}, D_{10,i}, D_{11,i}, D_{01,i}) = (d_{00}, d_{10}, d_{11}, 1 - d_{00} - d_{10} - d_{11})$ is:

$$p_\theta(d_{00}, d_{10}, d_{11}) = [\tilde{\gamma}_{00}(\theta)]^{d_{00}} [\tilde{\gamma}_{10}(\theta)]^{d_{10}} [\tilde{\gamma}_{11}(\theta)]^{d_{11}} [1 - \tilde{\gamma}_{00}(\theta) - \tilde{\gamma}_{10}(\theta) - \tilde{\gamma}_{11}(\theta)]^{1-d_{00}-d_{10}-d_{11}} .$$

In the simulations, we use a likelihood criterion with parameter space:

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

We simulate the data using $\beta_1 = \beta_2 = 0.2$, $\Delta_1 = \Delta_2 = -0.5$, $\rho = 0.5$ and $s = 0.5$.

We put a flat prior on Θ and implement the SMC algorithm as described in Appendix A.3. Figure 2 displays histograms of the marginal draws for Δ_1, β_1, s and ρ for one run of the SMC algorithm with a sample of size $n = 1000$. The plots for Δ_2 and β_2 are very similar to those for Δ_1 and β_1 (which is to be expected as the parameters are symmetric) and are therefore omitted. The draws for Δ_1 and β_1 are supported on and around their respective identified sets, which are approximately $[-1.42, 0]$ and $[-0.05, 0.66]$ (the identified sets for ρ and s are $[0, 1]$). Note that here the draws for Δ_1, β_1 and ρ are not flat over their identified sets, in contrast with the draws for μ and η_1 in Figure 1. To see why, consider the marginal identified set for (Δ_1, ρ), plotted as the shaded region in Figure 2. This plot shows that when Δ_1 is close to the upper bound of its identified set, (Δ_1, ρ) is in the shaded region for any $\rho \in [0, 1]$. However, when Δ_1 is close to the lower bound of its identified set, (Δ_1, ρ) is only in the shaded region for very large values of ρ. This structure of Θ_I, together with the flat prior on Θ, means that the marginal posterior for Δ_1 assigns relatively more mass

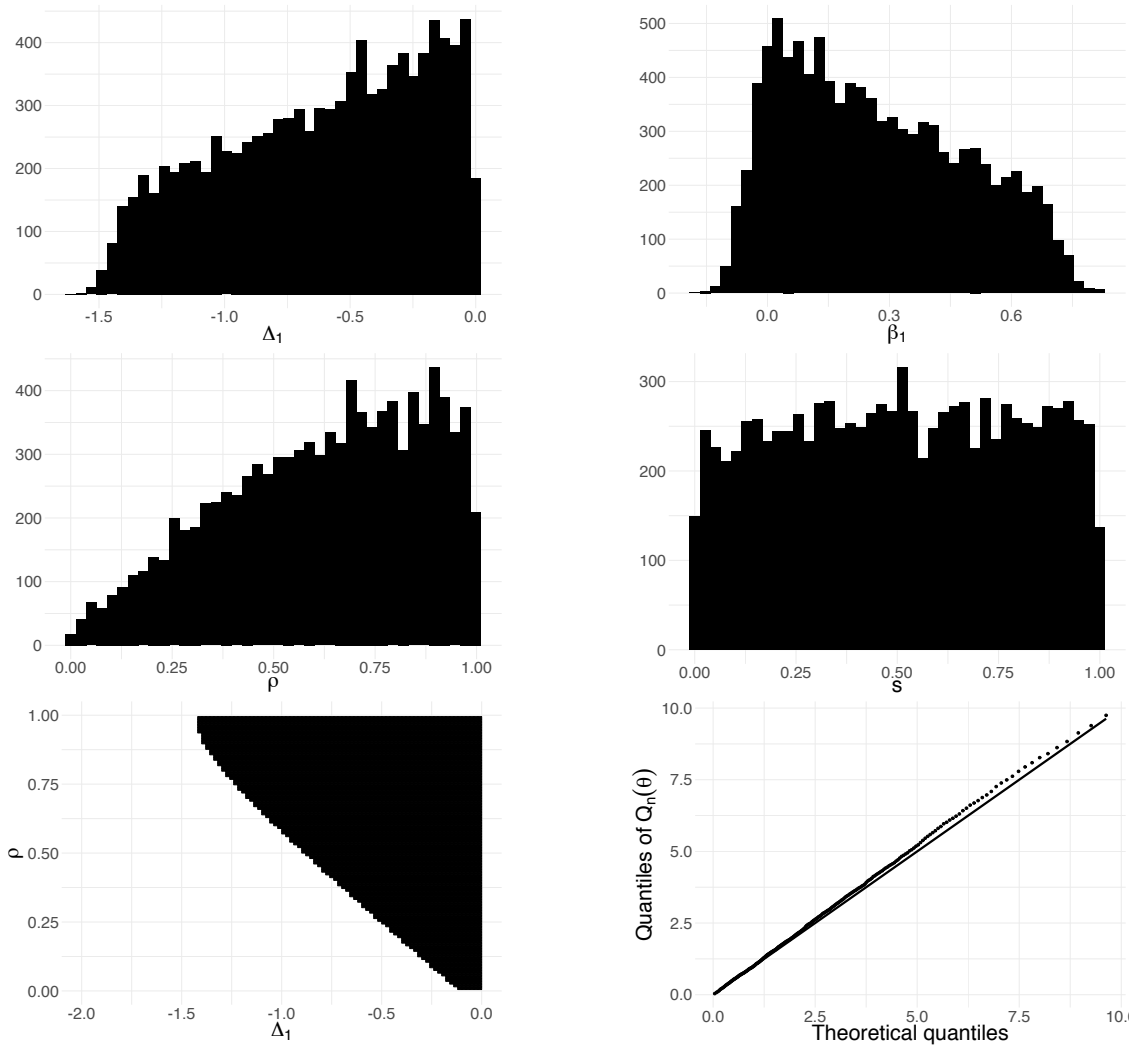

Figure 2: Entry game example: histograms of the SMC draws for Δ_1 (top left), β_1 (top right) ρ (mid left), and s (mid right), and Q-Q plot of $Q_n(\theta)$ computed from the draws against χ^2_3 quantiles (bottom right) for a sample of size $n = 1000$. The identified set for (Δ_1, ρ) is also shown (bottom left).

towards the upper limit of the identified set for Δ_1. Similar logic applies for β_1 and ρ. Figure 2 also shows that the quantiles of $Q_n(\theta)$ computed from the draws are very close to the χ^2_3 quantiles, as predicted by our theoretical results below.

Table 3 reports average coverage probabilities and CS limits for the various procedures across 1000 replications. We form CSs for Θ_I using procedure 1, as well as CSs for the identified sets of scalar subvectors Δ_1 and β_1 using procedures 2 and 3.[9] We also compare our CS for identified sets for Δ_1 and β_1 with projection-based and percentile-based CSs. Appendix A.3 provides additional details on computation of $M(\theta)$ for implementation of procedure 2. We do not use the reduced-

[9] As the parameterization is symmetric, the identified sets for Δ_2 and β_2 are the same as for Δ_1 and β_1 so we omit them. We also omit CSs for ρ and s, whose identified sets are both $[0, 1]$.

form reparameterization in terms of choice probabilities to compute $M(\theta)$. Coverage of $\widehat{\Theta}_\alpha$ for Θ_I is extremely good, even with the small sample size $n = 100$. Coverage of procedures 2 and 3 for the identified sets for Δ_1 and β_1 is slightly conservative for the small sample size n, but close to nominal for $n = 1000$. As expected, projection CSs are valid but very conservative (the coverage probabilities of 90% CSs are all at least 98%) whereas percentile-based CSs undercover.

3.2 Empirical applications

This subsection implements our procedures in two non-trivial empirical applications. The first application estimates an entry game with correlated payoff shocks using data from the US airline industry. Here there are 17 model parameters to be estimated. The second application estimates a model of trade flows initially examined in Helpman, Melitz, and Rubinstein (2008) (HMR henceforth). We use a version of the empirical model in HMR with 46 parameters to be estimated.

Although the entry game model is separable, we do not make use of separability in implementing our procedures. In fact, the existing Bayesian approaches that impose priors on the globally-identified reduced-form parameters will be problematic in this example. This separable model has 24 non-redundant choice probabilities (global reduced-form parameters, i.e., $\dim(\phi) = 24$) and 17 model structural parameters (i.e., $\dim(\theta) = 17$), and there is no explicit closed form expression for the identified set. Both Moon and Schorfheide (2012) and Kline and Tamer (2016) would sample from the posterior for the reduced-form parameter ϕ. But, unless the posterior for ϕ is constrained to lie on $\{\phi(\theta) : \theta \in \Theta\}$ (i.e. the set of reduced-form probabilities consistent with the model, rather than the full 24-dimensional space), certain values of ϕ drawn from their posteriors for ϕ will not be consistent with the model.

The empirical trade example is a *nonseparable* likelihood model that cannot be handled by either (a) existing Bayesian approaches that rely on a point-identified, $\sqrt{n}$-estimable and asymptotically normal reduced-form parameter, or (b) inference procedures based on moment inequalities.

In both applications, our approach only puts a prior on the model structural parameter θ so it does not matter whether the model is separable or not. Both applications illustrate how our procedures may be used to examine the robustness of estimates to various ad hoc modeling assumptions in a theoretically valid and computationally feasible way.

3.2.1 Bivariate Entry Game with US Airline Data

This section estimates a version of the entry game that we study in Subsection 3.1.2 above. We use data from the second quarter of 2010's Airline Origin and Destination Survey (DB1B) to estimate

	0.90		0.95		0.99	
	CSs for the identified set Θ_I					
	$\widehat{\Theta}_\alpha$ (Procedure 1)					
100	0.924	—	0.965	—	0.993	—
250	0.901	—	0.952	—	0.996	—
500	0.913	—	0.958	—	0.991	—
1000	0.913	—	0.964	—	0.992	—
	CSs for the identified set for Δ_1					
	$\widehat{M}_\alpha$ (Procedure 2)					
100	0.958	[−1.70,0.00]	0.986	[−1.77,0.00]	0.997	[−1.87,0.00]
250	0.930	[−1.58,0.00]	0.960	[−1.62,0.00]	0.997	[−1.70,0.00]
500	0.923	[−1.52,0.00]	0.961	[−1.55,0.00]	0.996	[−1.60,0.00]
1000	0.886	[−1.48,0.00]	0.952	[−1.50,0.00]	0.989	[−1.54,0.00]
	$\widehat{M}^{\chi}_\alpha$ (Procedure 3)					
100	0.944	[−1.70,0.00]	0.973	[−1.75,0.00]	0.993	[−1.84,0.00]
250	0.939	[−1.59,0.00]	0.957	[−1.62,0.00]	0.997	[−1.69,0.00]
500	0.937	[−1.53,0.00]	0.971	[−1.55,0.00]	0.996	[−1.60,0.00]
1000	0.924	[−1.49,0.00]	0.966	[−1.51,0.00]	0.992	[−1.54,0.00]
	$\widehat{M}^{proj}_\alpha$ (Projection)					
100	0.993	[−1.84,0.00]	0.997	[−1.88,0.00]	1.000	[−1.94,0.00]
250	0.996	[−1.69,0.00]	0.999	[−1.72,0.00]	1.000	[−1.79,0.00]
500	0.996	[−1.60,0.00]	0.999	[−1.62,0.00]	1.000	[−1.67,0.00]
1000	0.989	[−1.54,0.00]	0.996	[−1.56,0.00]	0.999	[−1.59,0.00]
	$\widehat{M}^{perc}_\alpha$ (Percentiles)					
100	0.000	[−1.43,−0.06]	0.000	[−1.54,−0.03]	0.000	[−1.72,−0.01]
250	0.000	[−1.37,−0.06]	0.000	[−1.45,−0.03]	0.000	[−1.57,−0.01]
500	0.000	[−1.34,−0.05]	0.000	[−1.41,−0.03]	0.000	[−1.50,−0.01]
1000	0.000	[−1.33,−0.05]	0.000	[−1.39,−0.03]	0.000	[−1.46,−0.01]
	CSs for the identified set for β_1					
	$\widehat{M}_\alpha$ (Procedure 2)					
100	0.960	[−0.28,0.89]	0.974	[−0.32,0.94]	0.994	[−0.40,1.03]
250	0.935	[−0.18,0.81]	0.958	[−0.20,0.84]	0.995	[−0.26,0.89]
500	0.925	[−0.14,0.76]	0.958	[−0.16,0.78]	0.995	[−0.19,0.82]
1000	0.926	[−0.11,0.72]	0.970	[−0.12,0.74]	0.994	[−0.15,0.76]
	$\widehat{M}^{\chi}_\alpha$ (Procedure 3)					
100	0.918	[−0.26,0.87]	0.963	[−0.30,0.92]	0.992	[−0.38,1.01]
250	0.914	[−0.17,0.80]	0.953	[−0.20,0.83]	0.991	[−0.25,0.88]
500	0.912	[−0.13,0.75]	0.957	[−0.15,0.77]	0.990	[−0.19,0.81]
1000	0.917	[−0.11,0.72]	0.962	[−0.12,0.73]	0.993	[−0.14,0.76]
	$\widehat{M}^{proj}_\alpha$ (Projection)					
100	0.990	[−0.38,1.00]	0.997	[−0.41,1.05]	1.000	[−0.49,1.13]
250	0.989	[−0.24,0.88]	0.997	[−0.27,0.90]	1.000	[−0.32,0.96]
500	0.989	[−0.19,0.81]	0.996	[−0.20,0.82]	1.000	[−0.24,0.86]
1000	0.990	[−0.14,0.76]	0.998	[−0.15,0.77]	1.000	[−0.18,0.80]
	$\widehat{M}^{perc}_\alpha$ (Percentiles)					
100	0.395	[−0.11,0.71]	0.654	[−0.16,0.78]	0.937	[−0.26,0.90]
250	0.169	[−0.05,0.66]	0.478	[−0.09,0.71]	0.883	[−0.16,0.80]
500	0.085	[−0.04,0.63]	0.399	[−0.07,0.68]	0.840	[−0.12,0.74]
1000	0.031	[−0.03,0.62]	0.242	[−0.05,0.65]	0.803	[−0.09,0.70]

Table 3: Entry game example: average coverage probabilities for Θ_I and identified sets for Δ_1 and β_1 across MC replications and average lower and upper bounds of CSs for identified sets for Δ_1 and β_1 across MC replications using a likelihood criterion function and flat prior. The identified sets for Δ_1 and β_1 are approximately $[-1.42, 0]$ and $[-0.05, 0.66]$.

a binary game where the payoff for firm i from entering market m is

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

where the Δ_i are assumed to be negative (as usually the case in entry models). The data contain 7882 markets which are formally defined as trips between two airports irrespective of stopping and we examine the entry behavior of two kinds of firms: LC (low cost) firms,[10] and OA (other airlines) which includes all the other firms. The unconditional choice probabilities are $(.16, .61, .07, .15)$ which are respectively the probabilities that OA and LC serve a market, that OA and not LC serve a market, that LC and not OA serve a market, and finally whether no airline serve the market.

The regressors are *market presence* and *market size*. Market presence is a market- and airline-specific variable defined as follows: from a given airport, we compute the ratio of markets a given carrier (we take the maximum within the category OA or LC, as appropriate) serves divided by the total number of markets served from that given airport. The market presence variable MP is the average of the ratios from the two endpoints and it provides a proxy for an airline's presence in a given airport (See Berry (1992) for more on this variable). This variable acts as an excluded regressor: the market presence for OA only enters OA's payoffs, so MP is both market- and airline-specific. The second regressor we use is *market size* MS which is defined as the population at the endpoints, so this variable is market-specific. We discretize both MP and MS into binary variables that take the value of one if the variable is higher than its median (in the data) value and zero otherwise. The choice probabilities are $P(y_{OA}, y_{LC}|MS,\ MP_{OA},\ MP_{LC})$ are conditional on the three-dimensional vector (MS, MP_{OA}, MP_{LC}). We therefore have 4 choice probabilities for every value of the conditioning variables (and there are 8 values for these).[11] To use notation similar to that in Subsection 3.1.2, let OA be player 1 and firm LC be player 2. Denote $\beta_1(x_{mOA}) := \beta_{OA}^0 + \beta_{OA}' x_{mOA}$ and $\beta_2(x_{mLC}) := \beta_{LC}^0 + \beta_{LC}' x_{mLC}$ with $x_{mOA} = (MS_m, MP_{mOA})'$ and $x_{mLC} = (MS_m, MP_{mLC})'$. The likelihood for market m depends on the choice probabilities:

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

Here $s(x_m)$ is a nuisance parameter which corresponds to the various *aggregate* equilibrium selection probabilities. Here $s(\cdot)$ is defined on the support of x_m, so in the model this function takes $2^3 = 8$ values each belonging to $[0, 1]$. In the *full model* we make no assumptions on the equilibrium selection

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

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

mechanism. Therefore, the full model has 17 parameters: 4 parameters per profit function (namely Δ_i, β_i^0, β_i^{MS}, and β_i^{MP}), the correlation ρ between ϵ_{i1} and ϵ_{i2}, and the 8 parameters in the aggregate equilibrium choice probabilities $s(\cdot)$. We also estimate a restricted version of the model called *fixed* s in which we restrict the aggregate selection probabilities to be the same across markets, for a total of 10 parameters. Note that these are just one version of the econometric model for a game; a less parsimonious version would allow, for example, for the parameters to change with regressor values, or allow for the regressors' support to be richer (rather than binary). We analyze this case precisely to highlight the fact that our CSs provide coverage guarantees regardless of whether the parameter vector is point identified.

We again take a flat prior on Θ and implement the procedures using a likelihood criterion. We restrict the support of Δ_i to $[-2, 0]$, β_i to $[-1, 2]^3$, ρ to $[0, 1]$ and the selection probabilities to $[0, 1]$. We implement the procedure using the adaptive SMC algorithm as described in Appendix A.4 with $B = 10000$ draws. Histograms of the SMC draws for the selection probabilities are presented in Figure 3; histograms of draws for the profit function parameters and ρ are presented in Figures 7 and 8 in Appendix A.4. We construct CSs for each of the parameters using procedure 2 and procedure 3 and compare these to projection-based CSs (projecting $\widehat{\Theta}_\alpha$ using our procedure 1) and percentile CSs. The empirical findings are presented in Table 4 below. Appendix A.4 contains further details on computation of $M(\theta)$ for implementation of procedure 2. As with the game simulation, we do not explicitly use the reduced-form reparameterization $\theta \mapsto \tilde{\gamma}(\theta)$ when computing $M(\theta)$.

The results in Table 4 show that CSs computed via procedures 2 and 3 are generally similar (though there are some differences, with CSs via procedure 2, which is valid under weaker conditions than procedure 3, appearing wider for some of the selection probabilities in the full model). On the other hand, projection CSs are very wide, especially in the full model. For instance, the projection CS for s_{101} is $[0, 1]$ whereas CSs via procedures 2 and 3 are $[0.49, 0.92]$ and $[0.61, 0.84]$ respectively. As expected, percentile CSs are narrower than procedure 2 and 3 CSs, reflecting the fact that percentile CSs under-cover in partially identified models.

Starting with the full model results, we see that the estimates are meaningful economically and are inline with recent estimates obtained in the literature. For example, fixed costs (the intercepts) are positive and significant for the large airlines (OA) but are negative for the LC carriers. Typically, the presence of higher fixed costs can signal various barriers to entry prevent LCs from entering: the higher these fixed costs the less likely it is for LCs to enter. On the other hand, higher fixed costs of large airlines are associated with a bigger presence (such as a hub) and so OAs are more likely to enter. As expected, both market presence and market size are associated with a positive probability of entry for both OA and LC. Results for the fixed-s model are in agreement with the corresponding ones for the full model and tell a consistent story. Note also the very high correlation in the errors, which could indicate missing profitability variables whereby firms enter a particularly profitable markets regardless of competition.

	Full model				Fixed-s model			
	Procedure 2	Procedure 3	Projection	Percentile	Procedure 2	Procedure 3	Projection	Percentile
Δ_{OA}	[−1.599,−1.178]	[−1.539,−1.303]	[−1.707,−0.701]	[−1.515,−1.117]	[−1.563,−1.335]	[−1.543,−1.363]	[−1.655,−1.194]	[−1.536,−1.326]
Δ_{LC}	[−1.527,−1.218]	[−1.503,−1.246]	[−1.719,−1.018]	[−1.489,−1.225]	[−1.567,−1.343]	[−1.547,−1.367]	[−1.671,−1.222]	[−1.548,−1.339]
β_{OA}^{0}	[0.443,0.581]	[0.455,0.575]	[0.341,0.695]	[0.447,0.578]	[0.431,0.551]	[0.437,0.539]	[0.365,0.611]	[0.427,0.540]
β_{OA}^{MS}	[0.365,0.539]	[0.383,0.521]	[0.238,0.665]	[0.389,0.544]	[0.347,0.479]	[0.353,0.467]	[0.275,0.551]	[0.348,0.477]
β_{OA}^{MP}	[0.413,0.581]	[0.425,0.569]	[0.275,0.713]	[0.424,0.579]	[0.479,0.641]	[0.497,0.623]	[0.389,0.719]	[0.504,0.648]
β_{LC}^{0}	[−1.000,−0.729]	[−1.000,−0.723]	[−1.000,−0.453]	[−0.993,−0.751]	[−0.910,−0.627]	[−0.874,−0.657]	[−1.000,−0.507]	[−0.917,−0.655]
β_{LC}^{MS}	[0.226,0.431]	[0.238,0.419]	[0.064,0.599]	[0.220,0.405]	[0.299,0.443]	[0.305,0.431]	[0.220,0.527]	[0.303,0.442]
β_{LC}^{MP}	[1.591,1.868]	[1.633,1.832]	[1.423,1.988]	[1.615,1.821]	[1.573,1.790]	[1.597,1.760]	[1.489,1.880]	[1.590,1.776]
ρ	[0.874,0.986]	[0.910,0.978]	[0.713,0.998]	[0.867,0.977]	[0.938,0.990]	[0.948,0.986]	[0.886,0.998]	[0.935,0.986]
s	—	—	—	—	[0.926,0.980]	[0.932,0.976]	[0.888,0.992]	[0.927,0.977]
s_{000}	[0.587,0.964]	[0.679,0.950]	[0.000,1.000]	[0.572,0.934]	—	—	—	—
s_{001}	[0.812,1.000]	[0.854,1.000]	[0.439,1.000]	[0.797,0.995]	—	—	—	—
s_{010}	[0.000,1.000]	[0.000,0.906]	[0.000,1.000]	[0.018,0.828]	—	—	—	—
s_{100}	[0.637,0.998]	[0.794,0.998]	[0.000,1.000]	[0.612,0.990]	—	—	—	—
s_{011}	[0.916,1.000]	[0.930,1.000]	[0.804,1.000]	[0.915,0.999]	—	—	—	—
s_{101}	[0.491,0.920]	[0.607,0.842]	[0.000,1.000]	[0.449,0.799]	—	—	—	—
s_{110}	[0.000,1.000]	[0.000,1.000]	[0.000,1.000]	[0.042,0.986]	—	—	—	—
s_{111}	[0.942,1.000]	[0.966,1.000]	[0.856,1.000]	[0.941,0.999]	—	—	—	—

Table 4: Entry game application: 95% CSs for structural parameters computed via our Procedures 2 and 3 as well as via Projection and Percentile methods. The full model contains a general specification for equilibrium selection while the fixed-s model restricts the equilibrium selection probability to be the same across markets with different regressor values.

One interesting observation are the CSs for the selection probabilities (also see Figure 3). Consider s_{010} and s_{110}: these are the aggregate selection probabilities which, according to the results, are not identified. This is likely due to the rather small number of markets with small size, large presence for OA but small presence for LC (for s_{010}) and the small number of markets with large market size, large presence for OA but small presence for LC (for s_{110}). The strength of our approach is its *adaptivity* to lack of identification in a particular data set: for example, 95% CSs for the identified set for s_{010} are $[0, 1]$ (via procedure 2), indicating that the model (and data) has no information about this parameter, while the corresponding CS for the identified set for s_{111} is the narrow and informative interval $[0.94, 1.00]$.

3.2.2 An empirical model of trade flows

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

$$\log M_{ij} = \begin{cases} \beta_0 + \lambda_j + \chi_i - \nu' f_{ij} + \delta z^*_{ij} + u_{ij} & \text{if } z^*_{ij} > 0 \\ \text{not observed} & \text{if } z^*_{ij} \leq 0 \end{cases}$$
$$z^*_{ij} = \beta^*_0 + \lambda^*_j + \chi^*_i - \nu^{*\prime} f_{ij} + \eta^*_{ij}$$

in which λ_j, χ_i, λ^*_j and χ^*_i are exporting and importing continent fixed effects, f_{ij} is a vector of observable trade frictions between i and j, and u_{ij} and η^*_{ij} are error terms described below. Exclusion restrictions can be imposed by setting at least one of the elements of ν equal to zero.

There are three differences between our empirical model and that of Helpman et al. (2008). First, we let z^*_{ij} enter the outcome equation linearly instead of nonlinearly.[12] Second, we use continent fixed effect instead of country fixed effects. This reduces the number of parameters from over 400 to 46. Third, we allow for heteroskedasticity in the selection equation, which is known to be a problem in trade data. This illustrates the robustness approach we advocate which relaxes parametric assumptions on part of the model that is suspect (homoskedasticity) without worrying about loss of point identification.

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

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

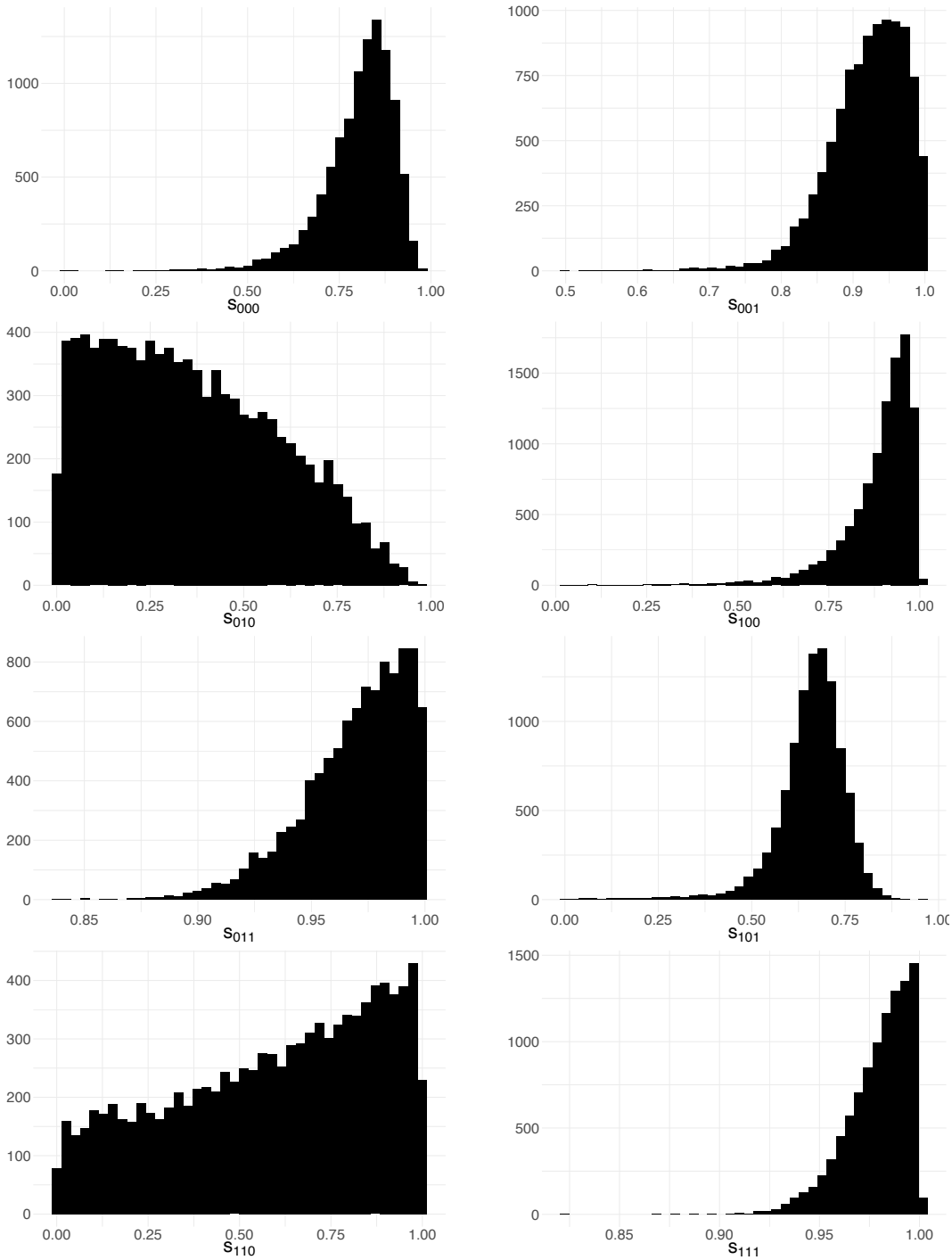

Figure 3: Entry game application: histograms of the SMC draws for selection probabilities s_{000}, s_{001}, s_{010}, s_{100}, s_{011}, s_{101}, s_{110}, and s_{111} for the **full model**.

	Homoskedastic		Heteroskedastic				
Variable	MLE	t-stat CI	MLE	t-stat CI	Procedure 2	Procedure 3	Percentile
Distance	2.352	[1.154,3.549]	0.314	[0.273,0.355]	[0.216,0.749]	[0.242,0.509]	[0.207,0.397]
Border	-5.191	[-7.077,-3.304]	-2.265	[-2.452,-2.077]	[-2.651,-1.859]	[-2.611,-1.898]	[-2.618,-1.816]
Island	-1.302	[-1.913,-0.691]	-0.728	[-0.868,-0.589]	[-1.060,-0.308]	[-1.060,-0.308]	[-0.983,-0.397]
Landlock	-7.275	[-10.769,-3.780]	-1.369	[-1.602,-1.137]	[-2.194,-0.914]	[-2.194,-0.890]	[-1.801,-0.954]
Legal	0.358	[0.002,0.715]	-0.122	[-0.183,-0.061]	[-0.254,0.004]	[-0.242,-0.009]	[-0.248,0.011]
Language	-4.098	[-6.430,-1.766]	-0.095	[-0.168,-0.021]	[-0.868,0.049]	[-0.868,0.026]	[-0.237,0.067]
Colonial	-17.378	[-26.002,-8.755]	-2.822	[-3.029,-2.615]	[-4.980,-2.373]	[-4.980,-2.461]	[-3.231,-2.298]
Currency	-1.550	[-2.780,-0.320]	-0.631	[-0.946,-0.315]	[-1.315,0.020]	[-1.282,-0.013]	[-1.274,0.062]
FTA	-19.540	[-29.783,-9.298]	-2.151	[-2.410,-1.892]	[-2.686,-1.589]	[-2.631,-1.616]	[-2.680,-1.577]

Table 5: Maximum likelihood estimates of the coefficients ν of the trade friction variables in the outcome equation (MLE) together with their 95% confidence sets based on inverting t-statistics (t-stat CI). Also shown are 95% CSs computed by our Procedures 2 and 3 as well as via Percentile methods.

ables is Normal with mean zero and covariance:

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

where X_{ij} denotes f_{ij}, the exporter's continent, and the importer's continent and where

$$\sigma_z(X_{ij}) = \exp(\varpi_1 \log(\text{distance}_{ij}) + \varpi_2 \log(\text{distance}_{ij})^2)\,.$$

We estimate the model from data on 24,649 country pairs in the selection equation and 11,146 country pairs in the outcome equation using the same data from 1986 as in Helpman et al. (2008). We also impose the exclusion restriction that the coefficient in ν corresponding to religion is equal to zero, else there is an exact linear relationship between the coefficients in the outcome and selection equation. This leaves a total of 46 parameters to be estimated. We only report estimates for the trade friction coefficients ν in the outcome equation as these are the most important. We estimate the model first by maximum likelihood under homoskedasticity and report conventional ML estimates for ν together with 95% CSs based on inverting t-statistics. We then re-estimate the model under heteroskedasticity and report conventional ML estimates together with confidence sets based on inverting t-statistics, percentile CSs (i.e. the Chernozhukov and Hong (2003) procedure under point identification), and our procedures 2 and 3. To implement our Procedure 2 and percentile CSs, we use the adaptive SMC algorithm as described in Appendix A.5 with $B = 10000$ draws.

The results are presented in Table 5.[13] Overall, though the model is sensitive to the presence of heteroskedasticity, the results for the heteroskedastic specification show that the confidence sets seem reasonably insensitive to the type of procedure used, which suggests that partial identification may not be an issue even allowing for heteroskedasticity. We also notice some difference in results relative to Helpman et al. (2008). For instance, they document strong positive effects of common legal systems and currency unions on trade flows, whereas we find much weaker evidence for this. We also find a positive effect of landlocked status on trade flows whereas they document a negative effect. Under heteroskedasticity, the magnitudes of coefficients of the trade friction variables are generally smaller than under homoskedasticity but of the same sign. The exception is the legal variable, whose coefficient is positive under homoskedasticity but negative under heteroskedasticity. A remaining question is whether the estimates are also sensitive to the normality assumption on the errors. This question can be examined within the context of our results by, for example, using a flexible form for the joint distribution of the errors.

[13] Note that the friction variables enter negatively in the outcome equation. The coefficient of distance is positive meaning that distance negatively affects trade flows; the remaining variables are dummy variables, so a negative coefficient of border means that sharing a border positively affects trade flows, and so forth.

4 Large Sample Properties

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

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

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

Assumption 4.1. *(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 a possibly unknown dimension $d^* \in [1,\infty)$, with $\gamma(\theta) = \gamma_0 \equiv 0$ if and only if $\theta \in \Theta_I$. Here $\gamma(\cdot)$ is merely a proof device and is only required to exist for θ in a fixed neighborhood of Θ_I. To accommodate situations in which the true reduced-form parameter value $\gamma_0 = 0$ may be "on the boundary" of Γ, we assume that the sets $T_{osn} \equiv \{\sqrt{n}\gamma(\theta) : \theta \in \Theta_{osn}\}$ *cover*[14] *a closed convex cone* $T \subseteq \mathbb{R}^{d^*}$. We note that this is trivially satisfied with $T = \mathbb{R}^{d^*}$ whenever each T_{osn} contains a ball of radius $k_n \to \infty$ centered at the origin. A similar approach is taken for point-identified models by Chernoff (1954), Geyer (1994), and Andrews (1999). Let $\|\gamma\|^2 := \gamma'\gamma$ and for any $v \in \mathbb{R}^{d^*}$, let $\mathbf{T}v = \arg\min_{t\in T} \|v-t\|^2$ denote the orthogonal (or metric) projection of v onto T.

Assumption 4.2. *(Local quadratic approximation)*
There exist sequences of random variables ℓ_n *and* $\mathbb{R}^{d^*}$*-valued random vectors* $\hat{\gamma}_n$ *(both measurable in* $\mathbf{X}_n$*) such that as* $n \to \infty$*:*

$$\sup_{\theta\in\Theta_{osn}} \left| nL_n(\theta) - \left(\ell_n + \frac{1}{2}\|\sqrt{n}\hat{\gamma}_n\|^2 - \frac{1}{2}\|\sqrt{n}(\hat{\gamma}_n - \gamma(\theta))\|^2\right)\right| = o_{\mathbb{P}}(1) \tag{18}$$

with $\sup_{\theta\in\Theta_{osn}} \|\gamma(\theta)\| \to 0$ *and* $\sqrt{n}\hat{\gamma}_n = \mathbf{T}\mathbb{V}_n$ *where* $\mathbb{V}_n \rightsquigarrow N(0,\Sigma)$*.*

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

[14] We say that a sequence of 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 4.3. *(Prior)*
(i) $\int_\Theta e^{nL_n(\theta)}\,\mathrm{d}\Pi(\theta) < \infty$ *almost surely;*
(ii) Π_Γ *has a continuous, strictly positive density* π_Γ *on* $B_\delta \cap \Gamma$ *for some* $\delta > 0$.

Discussion of Assumptions: Assumption 4.1(i) is a standard condition on any approximate extremum estimator, and Assumption 4.1(ii) is a standard posterior contraction condition. The choice of Θ_{osn} is deliberately general and will depend on the particular model under consideration. See Section 5 for verification of Assumption 4.1. Assumption 4.2 is a standard local quadratic expansion condition imposed on the local reduced form parameter around $\gamma = 0$. It is readily verified for likelihood and GMM models (see Section 5) with $\hat{\gamma}_n = \gamma(\hat{\theta})$ and $\mathbb{V}_n$ typically a normalized score function of the data. For these models with i.i.d. data the vector $\mathbb{V}_n$ is typically of the form: $\mathbb{V}_n = n^{-1/2}\sum_{i=1}^n v(X_i) + o_{\mathbb{P}}(1)$ with $\mathbb{E}[v(X_i)] = 0$ and $\mathrm{Var}[v(X_i)] = \Sigma$. In fact, Appendix B.1 shows that this quadratic expansion assumption is satisfied uniformly over a large class of DGPs in models with discrete data. Assumption 4.3(i) requires the quasi-posterior to be proper. Assumption 4.3(ii) is a standard prior mass and smoothness condition used to establish BvM theorems for identified parametric models (see, e.g., Section 10.2 of van der Vaart (2000)) but applied to Π_Γ. Under a flat prior on Θ and a continuous local mapping $\gamma : \Theta_I^N \mapsto \Gamma$, this assumption is easily satisfied (see its verification in examples of Section 5).

Assumptions 4.1(i) and 4.2 imply that the QLR statistic for Θ_I satisfies

$$\sup_{\theta\in\Theta_I} Q_n(\theta) = \|\mathbf{T}\mathbb{V}_n\|^2 + o_{\mathbb{P}}(1) \tag{19}$$

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

$$F_T(z) := \mathbb{P}_Z(\|\mathbf{T}Z\|^2 \le z) \tag{20}$$

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

Let $\mathbb{P}_{Z|\mathbf{X}_n}$ denote the distribution of a $N(0, I_{d^*})$ random vector Z (conditional on the data), and $T - v$ denote the convex cone T translated to have vertex at $-v$. The next lemma establishes the large sample behavior of the posterior distribution of the QLR statistic.

Lemma 4.1. *Let Assumptions 4.1, 4.2 and 4.3 hold. Then:*

$$\sup_z \left| \Pi_n\big(\{\theta : Q_n(\theta) \le z\} \big| \mathbf{X}_n\big) - \mathbb{P}_{Z|\mathbf{X}_n}\Big(\|Z\|^2 \le z \Big| Z \in T - \sqrt{n}\hat{\gamma}_n\Big)\right| = o_{\mathbb{P}}(1)\,. \tag{21}$$

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

This result shows that the posterior distribution of the QLR statistic is asymptotically $\chi^2_{d^*}$ when $T = \mathbb{R}^{d^*}$, which may be viewed as a Bayesian Wilks theorem for partially identified models, and asymptotically (first-order) stochastically dominates F_T when T is a closed convex cone. Note that Lemma 4.1 does not require the generalized information equality $\Sigma = I_{d^*}$ to hold. This lemma extends known BvM results for possibly misspecified likelihood models with point-identified $\sqrt{n}$-consistent and asymptotically normally estimable parameters (see Kleijn and van der Vaart (2012) and the references therein) to allow for other models with failure of $\Sigma = I_{d^*}$, with partially-identified parameters and/or parameters on a boundary.

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

Assumption 4.4. *(MC convergence)*
$\xi^{mc}_{n,\alpha} = \xi^{post}_{n,\alpha} + o_{\mathbb{P}}(1)$.

Lemma 4.1 and Assumption 4.4 together imply that our Procedure 1 CS $\widehat{\Theta}_\alpha$ is always a well-defined (quasi-)Bayesian credible set (BCS) regardless of whether $\Sigma = I_{d^*}$ holds or not. Further, together with Equation (19), they imply the following result.

Theorem 4.1. *Let Assumptions 4.1, 4.2, 4.3, and 4.4 hold with $\Sigma = I_{d^*}$. Then for any α such that $F_T(\cdot)$ is continuous at its α quantile, we have:*
(i) $\liminf_{n\to\infty} \mathbb{P}(\Theta_I \subseteq \widehat{\Theta}_\alpha) \ge \alpha$;
(ii) If $T = \mathbb{R}^{d^*}$ then: $\lim_{n\to\infty} \mathbb{P}(\Theta_I \subseteq \widehat{\Theta}_\alpha) = \alpha$.

Theorem 4.1 shows that we need the generalized information equality $\Sigma = I_{d^*}$ to hold so that our Procedure 1 CS $\widehat{\Theta}_\alpha$ has valid frequentist coverage for Θ_I in large samples.[15] This is because the asymptotic distribution of $\sup_{\theta\in\Theta_I} Q_n(\theta)$ is F_T only under $\Sigma = I_{d^*}$. It follows that, with a criterion satisfying $\Sigma = I_{d^*}$, our CS $\widehat{\Theta}_\alpha$ will be asymptotically exact (for Θ_I) when $T = \mathbb{R}^{d^*}$, and asymptotically valid but possibly conservative when T is a convex cone.

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

Remark 3. *Theorem 4.1 is still applicable to misspecified, separable partially-identified likelihood models. We can write the density in such models as* $p_\theta(\cdot) = q_{\tilde{\gamma}(\theta)}(\cdot)$ *where* $\tilde{\gamma}(\theta)$ *is an identifiable reduced-form parameter (see Section 5.1.1 below). Under misspecification the identified set is* $\Theta_I = \{\theta : \tilde{\gamma}(\theta) = \tilde{\gamma}^*\}$ *where* $\tilde{\gamma}^*$ *is the unique maximizer of* $E[\log q_{\tilde{\gamma}}(X_i)]$ *over* $\widetilde{\Gamma} = \{\tilde{\gamma}(\theta) : \theta \in \Theta\}$*. Following the insight of Müller (2013), we could base our inference on the sandwich log-likelihood function:*

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

where $\check{\gamma}$ *approximately maximizes* $\frac{1}{n}\sum_{i=1}^n \log q_\gamma(X_i)$ *over* $\widetilde{\Gamma}$ *and* $\widehat{\Sigma}_S$ *is the sandwich covariance matrix estimator for* $\check{\gamma}$*. If* $\sqrt{n}(\check{\gamma} - \tilde{\gamma}^*) \rightsquigarrow N(0, \Sigma_S)$ *and* $\hat{\Sigma}_S \to_p \Sigma_S$ *with* Σ_S *positive definite, then Assumption 4.2 will hold with* $\hat{\gamma}_n = \Sigma_S^{-1/2}(\check{\gamma} - \tilde{\gamma}^*)$ *where* $\sqrt{n}\hat{\gamma}_n \to_d N(0, I_{d^*})$ *and* $\gamma(\theta) = \Sigma_S^{-1/2}(\tilde{\gamma}(\theta) - \gamma^*)$*.*

Remark 4. *In correctly specified likelihood models with flat priors, one may interpret* $\widehat{\Theta}_\alpha$ *as a HPD* $100\alpha\%$ *BCS for* θ*. Moon and Schorfheide (2012) (MS hereafter) show that BCSs for a partially identified parameter* θ *(or subvectors) can under-cover asymptotically. As CSs for* Θ_I *should be larger than CSs for* θ*, MS's result might appear to suggest that our Procedure 1 CS* $\widehat{\Theta}_\alpha$ *would under-cover for* Θ_I*. The "apparent contradiction" is because a key regularity condition in MS's under-coverage result (their Assumption 2 on p. 767) is violated in our setting. For partially identified separable models, MS put a prior on the globally identified reduced-form parameter* γ*, say* $\Pi(\gamma)$*, and then a conditional prior, say* $\Pi(\theta|\gamma)$*, on the structural parameter* θ *given* γ*. The conditional prior* $\Pi(\cdot|\gamma)$ *needs to be supported on what would be the identified set for* θ *if* γ *were the true reduced form parameter. Their Assumption 2 requires that* $\Pi(\cdot|\gamma)$ *is (locally) Lipschitz in* γ*, which is violated in our setting. We only put a prior on* θ*. This prior on* θ *induces a prior on* $\gamma = \gamma(\theta)$ *and a conditional prior* $\Pi(\theta|\gamma)$ *that is supported on* $\{\theta \in \Theta : \gamma(\theta) = \gamma\}$*. Since* $\gamma(\theta) = 0$ *if and only if* $\theta \in \Theta_I$*, for any* $\bar{\gamma} \neq 0$ *our induced conditional prior satisfies*

$$\Pi(\{\theta \in \Theta : \gamma(\theta) = 0\}|\gamma = 0) - \Pi(\{\theta \in \Theta : \gamma(\theta) = 0\}|\gamma = \bar{\gamma}) = 1 - 0 = 1,$$

thereby violating MS's Lipschitz condition (their Assumption 2). See Remark 3 in MS for additional discussion of violation of their Lipschitz condition.

4.1.1 Models with singularities

In this subsection we consider (possibly) partially identified models with singularities.[16] In identifiable parametric models $\{P_\theta : \theta \in \Theta\}$, the standard notion of differentiability in quadratic mean requires that the mass of the part of P_θ that is singular with respect to the true distribution $P_0 = P_{\theta_0}$ vanishes faster than $\|\theta - \theta_0\|^2$ as $\theta \to \theta_0$ (Le Cam and Yang, 1990, section 6.2). If this

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

condition fails then the log-likelihood will not be locally quadratic at θ_0. By analogy with the identifiable case, we say a non-identifiable model has a singularity if it does not admit a local quadratic approximation (in the reduced-form reparameterization) like that in Assumption 4.2. One example is the missing data model under identification (see Subsection 5.3.1 below).

To allow for partially identified models with singularities, we first generalize the notion of the local reduced-form reparameterization to be of the form $\theta \mapsto (\gamma(\theta), \gamma_\perp(\theta))$ from Θ_I^N into $\Gamma \times \Gamma_\perp$ where $\Gamma \subseteq \mathbb{R}^{d^*}$ and $\Gamma_\perp \subseteq \mathbb{R}^{\dim(\gamma_\perp)}$ with $(\gamma(\theta), \gamma_\perp(\theta)) = 0$ if and only if $\theta \in \Theta_I$. The following regularity conditions generalize Assumptions 4.2 and 4.3 to allow for singularity.

Assumption 4.2′ *(Local quadratic approximation with singularity)*
(i) There exist sequences of random variables ℓ_n and $\mathbb{R}^{d^*}$-valued random vectors $\hat{\gamma}_n$ (both measurable in $\mathbf{X}_n$), and a sequence of functions $f_{n,\perp} : \Gamma_\perp \to \mathbb{R}_+$ (measurable in $\mathbf{X}_n$) with $f_{n,\perp}(0) = 0$ (almost surely), such that as $n \to \infty$:

$$\sup_{\theta \in \Theta_{osn}} \left| nL_n(\theta) - \left(\ell_n + \frac{1}{2}\|\sqrt{n}\hat{\gamma}_n\|^2 - \frac{1}{2}\|\sqrt{n}(\hat{\gamma}_n - \gamma(\theta))\|^2 - f_{n,\perp}(\gamma_\perp(\theta)) \right) \right| = o_{\mathbb{P}}(1) \tag{22}$$

with $\sup_{\theta \in \Theta_{osn}} \|(\gamma(\theta), \gamma_\perp(\theta))\| \to 0$ and $\sqrt{n}\hat{\gamma}_n = \mathbf{T}\mathbb{V}_n$ where $\mathbb{V}_n \rightsquigarrow N(0, \Sigma)$;
(ii) $\{(\gamma(\theta), \gamma_\perp(\theta)) : \theta \in \Theta_{osn}\} = \{\gamma(\theta) : \theta \in \Theta_{osn}\} \times \{\gamma_\perp(\theta) : \theta \in \Theta_{osn}\}$.

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

Discussion of Assumptions: Assumption 4.2' is generalizes of Assumption 4.2 to the singular case. Assumption 4.2' implies that the peak of the likelihood does not concentrate on sets of the form $\{\theta : f_{n,\perp}(\gamma_\perp(\theta)) > \epsilon > 0\}$. Recently, Bochkina and Green (2014) established a BvM result for *identifiable* parametric likelihood models with singularities. They assume the likelihood is locally quadratic in some parameters and locally linear in others (similar to Assumption 4.2'(i)) and that the local parameter space satisfies conditions similar to our Assumption 4.2'(ii). Assumption 4.3' generalizes Assumption 4.3 to the singular case. We impose no further restrictions on the set $\{\gamma_\perp(\theta) : \theta \in \Theta_I^N\}$.

The next lemma shows that the posterior distribution of the QLR asymptotically (first-order) stochastically dominates F_T in partially identified models with singularity.

Lemma 4.2. *Let Assumptions 4.1, 4.2' and 4.3' hold. Then:*

$$\sup_z \left(\Pi_n\big(\{\theta : Q_n(\theta) \le z\} \big| \mathbf{X}_n\big) - \mathbb{P}_{Z|\mathbf{X}_n}\left(\|Z\|^2 \le z \middle| Z \in T - \sqrt{n}\hat{\gamma}_n \right) \right) \le o_{\mathbb{P}}(1) \,. \tag{23}$$

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

Lemma 4.2 immediately implies the following result.

Theorem 4.2. *Let Assumptions 4.1, 4.2', 4.3', and 4.4 hold with $\Sigma = I_{d^*}$. Then for any α such that $F_T(\cdot)$ is continuous at its α quantile, we have:* $\liminf_{n\to\infty} \mathbb{P}(\Theta_I \subseteq \widehat{\Theta}_\alpha) \ge \alpha$.

For non-singular models, Theorem 4.1 establishes that $\widehat{\Theta}_\alpha$ is asymptotically valid for Θ_I, with asymptotically exact coverage when T is linear and can be conservative when T is a closed convex cone. For singular models, Theorem 4.2 shows that $\widehat{\Theta}_\alpha$ is still asymptotically valid for Θ_I but can be conservative even when T is linear.[17] When applied to the missing data example, Theorems 4.1 and 4.2 imply that $\widehat{\Theta}_\alpha$ for Θ_I is asymptotically exact under partial identification but conservative under point identification. This is consistent with simulation results reported in Table 1; see Section 5.3.1 below for details.

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

Here we present conditions under which $\widehat{M}_\alpha$ has correct coverage for the identified set M_I of subvectors μ. Recall the definition of $M(\theta) \equiv \{\mu : (\mu, \eta) \in \Delta(\theta) \text{ for some } \eta\}$ from Section 2. The profile criterion $PL_n(M(\theta))$ for $M(\theta)$ and the profile QLR $PQ_n(M(\theta))$ for $M(\theta)$ are defined the same way as those in (11) and (13) respectively:

$$PL_n(M(\theta)) \equiv \inf_{\mu \in M(\theta)} \sup_{\eta \in H_\mu} L_n(\mu, \eta) \quad \text{and} \quad PQ_n(M(\theta)) \equiv 2n[L_n(\hat{\theta}) - PL_n(M(\theta))].$$

Assumption 4.5. *(Profile QL)*
There exists a measurable $f : \mathbb{R}^{d^*} \to \mathbb{R}_+$ such that:

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

with $\hat{\gamma}_n$ and $\gamma(\cdot)$ from Assumption 4.2 or 4.2'.

Discussion of Assumption 4.5: By definition of M_I (in display (2)) we have: $M_I = \{\mu : \gamma(\mu, \eta) = 0 \text{ for some } \eta \in H_\mu\}$ and also $M_I = M(\theta)$ for any $\theta \in \Theta_I$. Thus

$$PQ_n(M_I) = \sup_{\mu \in M_I} \inf_{\eta \in H_\mu} Q_n(\mu, \eta) = PQ_n(M(\theta)) \quad \text{for all } \theta \in \Theta_I \ .$$

[17] It might be possible to establish asymptotically exact coverage of $\widehat{\Theta}_\alpha$ for Θ_I in singular models where the singular part $f_{n,\perp}(\gamma_\perp(\theta))$ in Assumption 4.2' possesses some extra structure.

Assumption 4.5 imposes some structure on the profile QLR statistic for M_I over the local neighborhood Θ_{osn}. It implies that the profile QLR for M_I is of the form:

$$PQ_n(M_I) = f(\mathbf{T}\mathbb{V}_n) + o_{\mathbb{P}}(1)\,. \tag{24}$$

When $\Sigma = I_{d^*}$, the asymptotic distribution of $\sup_{\theta\in\Theta_I} PQ_n(M(\theta)) = PQ_n(M_I)$ becomes G_T:

$$G_T(z) := \mathbb{P}_Z(f(\mathbf{T}Z) \le z) \quad \text{where} \quad Z \sim N(0, I_{d^*})\ .$$

The functional form of f depends on the local reparameterization γ and the geometry of M_I. When M_I is a singleton and $T = \mathbb{R}^{d^*}$ then equation (24) is typically satisfied with $f(v) = \inf_{t\in T_1} \|v - t\|^2$ where $T_1 = \mathbb{R}^{d_1^*}$ with $d_1^* < d^*$ and the QLR statistic is asymptotically $\chi^2_{d^*-d_1^*}$. For a non-singleton set M_I, f will typically be more complex. In the missing data example, we show in Section 5 that $f(v) = \max_{\mu\in\{\underline{\mu},\overline{\mu}\}} \inf_{t\in T_\mu} \|v - t\|^2$ where $T_{\underline{\mu}}$ and $T_{\overline{\mu}}$ are halfspaces and $T = \mathbb{R}^{d^*}$. Here the resulting profile QLR statistic for M_I is asymptotically the maximum of two mixtures of χ^2 random variables. Note that the existence of f is merely a proof device, and one does not need to know its precise expression in the implementation of our Procedure 2 CS $\widehat{M}_\alpha$ for M_I.

The next lemma is a new BvM-type result for the posterior distribution of the profile QLR for M_I. Note that this result also allows for singular models.

Lemma 4.3. *Let Assumptions 4.1, 4.2, 4.3, and 4.5 or 4.1, 4.2', 4.3', and 4.5 hold. Then for any interval I such that $\mathbb{P}_Z(f(Z) \le z)$ is continuous on I, we have:*

$$\sup_{z\in I}\Big|\Pi_n\big(\{\theta : PQ_n(M(\theta)) \le z\}\,\big|\,\mathbf{X}_n\big) - \mathbb{P}_{Z|\mathbf{X}_n}\Big(f(Z) \le z\Big|Z \in \sqrt{n}\hat{\gamma}_n - T\Big)\Big| = o_{\mathbb{P}}(1)\,.$$

And hence we have:
(i) If $T \subsetneq \mathbb{R}^{d^*}$ and f is subconvex,[18] *then: $\Pi_n\big(\{\theta : Q_n(\theta) \le z\}\big|\,\mathbf{X}_n\big) \le G_T(z)$ for all $z \ge 0$.*
(ii) If $T = \mathbb{R}^{d^*}$ then: $\sup_z \big|\Pi_n\big(\{\theta : Q_n(\theta) \le z\}\,\big|\,\mathbf{X}_n\big) - \mathbb{P}_Z\big(f(Z) \le z\big)\big| = o_{\mathbb{P}}(1)$.

Let $\xi^{post,p}_{n,\alpha}$ denote the α quantile of the profile QLR $PQ_n(M(\theta))$ under the posterior distribution Π_n, and $\xi^{mc,p}_{n,\alpha}$ be given in Remark 2.

Assumption 4.6. *(MC convergence)*
$\xi^{mc,p}_{n,\alpha} = \xi^{post,p}_{n,\alpha} + o_{\mathbb{P}}(1)$.

The next theorem is an important consequence of Lemma 4.3.

[18] We say that $f : \mathbb{R}^{d^*} \to \mathbb{R}_+$ is quasiconvex if $f^{-1}(z) := \{v : f(v) \le z\}$ is convex for each $z \ge 0$ and subconvex if, in addition, $f(v) = f(-v)$ for all $v \in \mathbb{R}^{d^*}$. The conclusion of Lemma 4.3(i) remains valid under the weaker condition that (i) f is quasiconvex and (ii) $\mathbb{P}_Z(Z \in (f^{-1}(\xi_\alpha) - T^o)) \le \mathbb{P}_Z(f(\mathbf{T}Z) \le \xi_\alpha)$ holds, where ξ_α is the α quantile of G_T and T^o is the polar cone of T.

Theorem 4.3. *Let Assumptions 4.1, 4.2, 4.3, 4.5, and 4.6 or 4.1, 4.2', 4.3', 4.5, and 4.6 hold with $\Sigma = I_{d^*}$ and suppose that $G_T(\cdot)$ is continuous at its α quantile.*
(i) If $T \subsetneq \mathbb{R}^{d^*}$ and f is subconvex,[19] *then:* $\liminf_{n\to\infty} \mathbb{P}(M_I \subseteq \widehat{M}_\alpha) \geq \alpha$ *;*
(ii) If $T = \mathbb{R}^{d^*}$ then: $\lim_{n\to\infty} \mathbb{P}(M_I \subseteq \widehat{M}_\alpha) = \alpha$.

Theorem 4.3(ii) shows that our Procedure 2 CSs $\widehat{M}_\alpha$ for M_I can have asymptotically exact coverage if $T = \mathbb{R}^{d^*}$ even if the model is singular. In the missing data example, Theorem 4.3(ii) implies that $\widehat{M}_\alpha$ for M_I is asymptotically exact irrespective of whether the model is point-identified or not (see Subsection 5.3.1 below). Theorem 4.3(i) shows that the CSs $\widehat{M}_\alpha$ for M_I can have conservative coverage when T is a convex cone (see Appendix C.2 for a moment inequality example).

4.3 Coverage properties of $\widehat{M}^\chi_\alpha$ for M_I of scalar subvectors

This section presents one sufficient condition for validity of our Procedure 3 CS $\widehat{M}^\chi_\alpha$ for $M_I \subset \mathbb{R}$. We say a half-space is *regular* if it is of the form $\{v \in \mathbb{R}^{d^*} : a'v \leq 0\}$ for some $a \in \mathbb{R}^{d^*}$.

Assumption 4.7. *(Profile QLR, χ^2 bound)*
$PQ_n(M(\theta)) \rightsquigarrow W \leq \max_{i\in\{1,2\}} \inf_{t\in T_i} \|Z - t\|^2$ *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 regular half-spaces in $\mathbb{R}^{d^*}$.*

Theorem 4.4. *Let Assumption 4.7 hold and let the distribution of W be continuous at its α quantile. Then:* $\liminf_{n\to\infty} \mathbb{P}(M_I \subseteq \widehat{M}^\chi_\alpha) \geq \alpha$.

The following proposition presents a set of sufficient conditions for Assumption 4.7.

Proposition 4.1. *Let the following hold:*
(i) Assumptions 4.1(i), 4.2 or 4.2' hold with $\Sigma = I_{d^*}$ and $T = \mathbb{R}^{d^*}$;
(ii) $\inf_{\mu\in M_I} \sup_{\eta\in H_\mu} L_n(\mu,\eta) = \min_{\mu\in\{\underline{\mu},\overline{\mu}\}} \sup_{\eta\in H_\mu} L_n(\mu,\eta) + o_{\mathbb{P}}(n^{-1})$;
(iii) for each $\mu \in \{\underline{\mu}, \overline{\mu}\}$ there exists a sequence of sets $(\Gamma_{\mu,osn})_{n\in\mathbb{N}}$ with $\Gamma_{\mu,osn} \subseteq \Gamma$ for each n and a halfspace T_μ in $\mathbb{R}^{d^*}$ such that:

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

and $\inf_{\gamma\in\Gamma_{\mu,osn}} \|\sqrt{n}\gamma - \mathbb{V}_n\|^2 = \inf_{t\in T_\mu} \|t - \mathbb{V}_n\|^2 + o_{\mathbb{P}}(1)$.
Then: Assumption 4.7 holds with $W = \max_{i\in\{\underline{\mu},\overline{\mu}\}} \inf_{t\in T_i} \|Z - t\|^2$.

Suppose $M_I = [\underline{\mu}, \overline{\mu}] \subsetneq \mathbb{R}$ (which is true when Θ_I is connected and bounded). If $\sup_{\eta\in H_\mu} L_n(\mu,\eta)$ is strictly concave in μ then condition (ii) of Proposition 4.1 holds. The other conditions of Proposition 4.1 are easy to verify as in the missing data example (see Subsection 5.3.1 below).

[19] The conclusion of Theorem 4.3(i) remains valid under the weaker condition stated in footnote for Lemma 4.3(i).

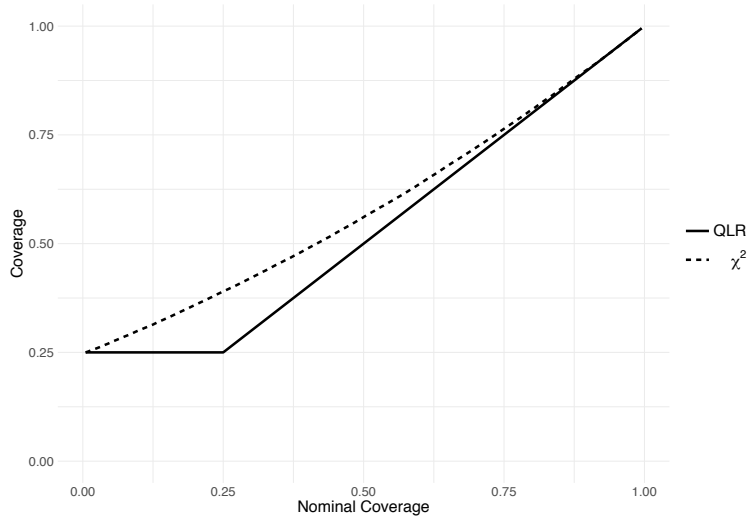

Figure 4: Missing data example: comparison of asymptotic coverage of $\widehat{M}_\alpha$ (QLR – solid kinked line) and $\widehat{M}_\alpha^\chi$ (χ^2 – dashed curved line) with their nominal coverage for models where $\widehat{M}_\alpha^\chi$ is valid for M_I but most conservative.

The exact distribution of $\max_{i\in\{1,2\}} \inf_{t\in T_i} \|Z-t\|^2$ depends on the geometry of T_1 and T_2. For the missing data example, the polar cones of T_1 and T_2 are at least 90° apart. The worst-case coverage (i.e., the case in which asymptotic coverage of $\widehat{M}_\alpha^\chi$ will be most conservative) will occur when the polar cones of T_1 and T_2 are orthogonal, in which case $\max_{i\in\{1,2\}} \inf_{t\in T_i} \|Z-t\|^2$ has the mixture distribution $W^* := \frac{1}{4}\delta_0 + \frac{1}{2}\chi_1^2 + \frac{1}{4}(\chi_1^2 \cdot \chi_1^2)$ where δ_0 is a point mass at zero and $\chi_1^2 \cdot \chi_1^2$ is the distribution of the product of two independent χ_1^2 random variables. The quantiles of the distribution of $\max_{i\in\{1,2\}} \inf_{t\in T_i} \|Z-t\|^2$ are continuous in α for all $\alpha > \frac{1}{4}$. For all configurations of T_1 and T_2 in this example, the distribution of $\max_{i\in\{1,2\}} \inf_{t\in T_i} \|Z-t\|^2$ (first-order) stochastically dominates F_{W^*} and is (first-order) stochastically dominated by $F_{\chi_1^2}$ (i.e., $F_{W^*}(w) \geq F_W(w) \geq F_{\chi_1^2}(w)$). Notice that this is different from the usual chi-bar-squared case encountered when testing whether a parameter belongs to the identified set on the basis of finitely many moment inequalities (Rosen, 2008).

To get an idea of the degree of conservativeness of $\widehat{M}_\alpha^\chi$, consider the class of models satisfying conditions for Proposition 4.1. Figure 4 plots the asymptotic coverage of $\widehat{M}_\alpha$ and $\widehat{M}_\alpha^\chi$ against nominal coverage for models in this class where $\widehat{M}_\alpha^\chi$ is most conservative for the missing data example (i.e., the worst-case coverage). For each model in this class, the asymptotic coverage of $\widehat{M}_\alpha$ and $\widehat{M}_\alpha^\chi$ is between the nominal coverage and the worst-case coverage. As can be seen, the coverage of $\widehat{M}_\alpha$ is exact at all levels $\alpha \in (0,1)$ for which the distribution of the profile QLR is continuous at its α quantile, as shown in Theorem 4.3(ii). 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.

Since empirical papers typically report CSs for scalar parameters, Theorem 4.4 will be very useful in applied work. One could generalize $\widehat{M}_\alpha^\chi$ to deal with vector-valued subvectors by allowing χ^2_d quantiles with higher degrees of freedom $d \in (1, \dim(\theta))$, but it might be difficult to provide sufficient conditions as those in Proposition 4.1 to establish results like Theorem 4.4.

5 Sufficient Conditions and Examples

This section provides sufficient conditions for the key regularity condition, Assumption 4.2, in possibly partially identified likelihood and moment-based models with i.i.d. data. See Appendix B.1 for low-level conditions to ensure that Assumption 4.2 holds uniformly over a large class of DGPs in discrete models. We also verify Assumptions 4.1, 4.2 (or 4.2'), 4.3 and 4.5) in examples.

We use standard empirical process notation: $P_0 g$ denotes the expectation of $g(X_i)$ under the true probability measure P_0, $\mathbb{P}_n g = n^{-1}\sum_{i=1}^n g(X_i)$ denotes expectation of $g(X_i)$ under the empirical measure, and $\mathbb{G}_n g = \sqrt{n}(\mathbb{P}_n - P_0)g$ denotes the empirical process.

5.1 Partially identified likelihood models

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

5.1.1 Separable likelihood models

For a large class of partially identified parametric likelihood models $\mathcal{P} = \{p_\theta : \theta \in \Theta\}$, there exists a function $\tilde{\gamma} : \Theta \to \widetilde{\Gamma} \subset \mathbb{R}^{d^*}$ for some possibly unknown $d^* \in [1, +\infty)$, such that $p_\theta(\cdot) = q_{\tilde{\gamma}(\theta)}(\cdot)$ for each $\theta \in \Theta$ and some densities $\{q_{\tilde{\gamma}(\theta)}(\cdot) : \tilde{\gamma} \in \widetilde{\Gamma}\}$. In this case we say that the model $\mathcal{P}$ is separable and admits a (global) reduced-form reparameterization. The reparameterization is assumed to be identifiable, i.e. $D_{KL}(q_{\tilde{\gamma}_0}\|q_{\tilde{\gamma}}) > 0$ for any $\tilde{\gamma} \neq \tilde{\gamma}_0$. The identified set is $\Theta_I = \{\theta \in \Theta : \tilde{\gamma}(\theta) = \tilde{\gamma}_0\}$ where $\tilde{\gamma}_0$ is the true parameter, i.e. $p_0 = q_{\tilde{\gamma}_0}$. Models with discrete choice probabilities (such as the missing data and entry game designs we used in simulations) fall into this framework, where the vector $\tilde{\gamma}$ maps the structural parameters θ to the model-implied probabilities of discrete outcomes and the true probabilities $\tilde{\gamma}_0 \in \widetilde{\Gamma}$ of discrete outcomes are point-identified.

The following result presents one set of sufficient conditions for Assumptions 4.1(ii) and 4.2 under

conventional smoothness assumptions.

Let $\ell_{\tilde{\gamma}}(\cdot) := \log q_{\tilde{\gamma}}(\cdot)$, let $\dot{\ell}_{\tilde{\gamma}}$ and $\ddot{\ell}_{\tilde{\gamma}}$ denote the score and Hessian, let $\mathbb{I}_0 := -P_0(\ddot{\ell}_{\tilde{\gamma}_0})$ and let $\gamma(\theta) = \mathbb{I}_0^{1/2}(\tilde{\gamma}(\theta) - \tilde{\gamma}_0)$ and $\Gamma = \{\mathbb{I}_0^{1/2}(\tilde{\gamma} - \tilde{\gamma}_0) : \tilde{\gamma} \in \widetilde{\Gamma}\}$.

Proposition 5.1. *Suppose that $\{q_{\tilde{\gamma}} : \tilde{\gamma} \in \widetilde{\Gamma}\}$ satisfies the following regularity conditions:*
(a) $X_1, \ldots, X_n$ is an i.i.d. sample from $q_{\tilde{\gamma}_0}$ with $\tilde{\gamma}_0$ identifiable and on the interior of $\widetilde{\Gamma}$;
(b) $\tilde{\gamma} \mapsto P_0\ell_{\tilde{\gamma}}$ is continuous and there is a neighborhood U of $\tilde{\gamma}_0$ on which $\ell_{\tilde{\gamma}}(x)$ is twice continuously differentiable for each x, with $\dot{\ell}_{\tilde{\gamma}_0} \in L^2(P_0)$ and $\sup_{\tilde{\gamma}\in U} \|\ddot{\ell}_{\tilde{\gamma}}(x)\| \leq \bar{\ell}(x)$ for some $\bar{\ell} \in L^2(P_0)$;
(c) $P_0\dot{\ell}_{\tilde{\gamma}} = 0$, $\mathbb{I}_0$ is non-singular, and $\mathbb{I}_0 = P_0(\dot{\ell}_{\tilde{\gamma}_0}\dot{\ell}'_{\tilde{\gamma}_0})$;
(d) $\widetilde{\Gamma}$ is compact and π_Γ is strictly positive and continuous on U.
Then: there exists a sequence $(r_n)_{n\in\mathbb{N}}$ with $r_n \to \infty$ and $r_n = o(n^{1/4})$ such that Assumptions 4.1(ii) and 4.2 hold for the average log-likelihood (3) over $\Theta_{osn} := \{\theta \in \Theta : \|\gamma(\theta)\| \leq r_n/\sqrt{n}\}$ with $\ell_n = n\mathbb{P}_n \log p_0$, $\sqrt{n}\hat{\gamma}_n = \mathbb{V}_n = \mathbb{I}_0^{-1/2}\mathbb{G}_n(\dot{\ell}_{\tilde{\gamma}_0})$, $\Sigma = I_{d^*}$ and $T = \mathbb{R}^{d^*}$.

5.1.2 General non-identifiable likelihood models

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

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

as $\epsilon \to 0$ and let $\overline{\mathcal{D}}_\epsilon = \mathcal{D}_\epsilon \cup \mathcal{D}$. The set $\mathcal{D}$ is the set of generalized Hellinger scores,[20] which consists of functions of X_i with mean zero and unit variance. The cone $\mathcal{T} = \{\tau d : \tau \geq 0, d \in \mathcal{D}\}$ 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 $g_p \in \mathcal{T}$ and remainders $R_p \in L^2(\lambda)$ such that:

$$\sqrt{p} - \sqrt{p_0} = g_p\sqrt{p_0} + h(p, p_0)R_p$$

with $\sup\{\|R_p\|_{L^2(\lambda)} : h(p, p_0) \leq \varepsilon\} \to 0$ as $\varepsilon \to 0$. If the linear hull $\mathrm{Span}(\mathcal{T})$ of $\mathcal{T}$ has finite dimension $d^* \geq 1$, then we can write each $g \in \mathcal{T}$ 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 $\mathrm{Span}(\mathcal{T})$ in $L^2(P_0)$. Let $\mathbb{T}$ denote the orthogonal

[20] 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.

projection[21] onto $\mathcal{T}$ and let $\gamma(\theta)$ be given by

$$\mathbb{T}(2(\sqrt{p_\theta/p_0}-1)) = \gamma(\theta)'\psi\,. \tag{25}$$

Proposition 5.2. *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, $\mathcal{T}$ is closed and convex and $\mathrm{Span}(\mathcal{T})$ 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 $r_n = o(n^{1/4})$, such that Assumption 4.2 holds for the average log-likelihood (3) over $\Theta_{osn} := \{\theta : h(p_\theta, p_0) \leq r_n/\sqrt{n}\}$ with $\ell_n = n\mathbb{P}_n \log p_0$, $\sqrt{n}\hat{\gamma}_n = \mathbb{V}_n = \mathbb{G}_n(\psi)$, $\Sigma = I_{d^*}$ and $\gamma(\theta)$ defined in (25).

Proposition 5.2 is a set of sufficient conditions for i.i.d. data; see Lemma F.4 in Appendix F for a more general result. Assumption 4.1(ii) can be verified under additional mild conditions (see, e.g., Theorem 5.1 of Ghosal, Ghosh, and van der Vaart (2000)).

5.2 GMM models

Consider the GMM model $\{\rho_\theta : \theta \in \Theta\}$ with $\rho : \mathscr{X} \times \Theta \to \mathbb{R}^{d_\rho}$. Let $g(\theta) = E[\rho_\theta(X_i)]$ and the identified set be $\Theta_I = \{\theta \in \Theta : g(\theta) = 0\}$ (we assume throughout this subsection that Θ_I is non-empty). When ρ is of higher dimension than θ, the set $\mathcal{G} = \{g(\theta) : \theta \in \Theta\}$ will not contain a neighborhood of the origin. But, if the map $\theta \mapsto g(\theta)$ is smooth (e.g. $\mathcal{G}$ is a smooth manifold) then $\mathcal{G}$ can typically be locally approximated at the origin by a closed convex cone $\mathcal{T} \subset \mathbb{R}^{d_\rho}$.

To simplify notation, we assume that for any $v \in \mathrm{Span}(\mathcal{T})$ we may partition $\Omega^{-1/2}v$ so that its upper d^* elements $[\Omega^{-1/2}v]_1$ are (possibly) non-zero and the remaining $d_\rho - d^*$ elements $[\Omega^{-1/2}v]_2 = 0$ (this can always be achieved by multiplying the moment functions by a suitable rotation matrix).[22] If $\mathcal{G}$ contains a neighborhood of the origin then we simply take $\mathcal{T} = \mathbb{R}^{d_\rho}$ and $[\Omega^{-1/2}v]_1 = \Omega^{-1/2}v$. Let $\mathbb{T}g(\theta)$ denote the projection of $g(\theta)$ onto $\mathcal{T} \subset \mathbb{R}^{d_\rho}$ and note that $[\Omega^{-1/2}\mathbb{T}g(\theta)]_2 = 0$. Finally, define $\Theta_I^\varepsilon = \{\theta \in \Theta : \|g(\theta)\| \leq \varepsilon\}$.

Proposition 5.3. *Suppose that $\{\rho_\theta : \theta \in \Theta\}$ satisfies the following regularity conditions:*
(a) there exists $\varepsilon_0 > 0$ such that $\{\rho_\theta : \theta \in \Theta_I^{\varepsilon_0}\}$ is Donsker;
(b) $E[\rho_\theta(X_i)\rho_\theta(X_i)'] = \Omega$ for each $\theta \in \Theta_I$ and Ω is positive definite;
(c) there exists $\theta^* \in \Theta_I$ such that $\sup_{\theta\in\Theta_I^\varepsilon} E[\|\rho_\theta(X_i) - \rho_{\theta^*}(X_i)\|^2] = o(1)$ as $\varepsilon \to 0$;

[21] If $\mathcal{T} \subseteq L^2(P_0)$ is a closed convex cone, the projection $\mathbb{T}f$ of any $f \in L^2(P_0)$ is defined as the unique element of $\mathcal{T}$ such that $\|f - \mathbb{T}f\|_{L^2(P_0)} = \inf_{t\in\mathcal{T}} \|f - t\|_{L^2(P_0)}$.

[22] See our July 2016 working paper version for details.

(d) there exists $\delta > 0$ such that $\sup_{\theta \in \Theta_I^\varepsilon} \|g(\theta) - \mathbb{T}g(\theta)\| = o(\varepsilon^{1+\delta})$ as $\varepsilon \to 0$.
Then: there exists a sequence $(r_n)_{n\in\mathbb{N}}$ with $r_n \to \infty$ and $r_n = o(n^{1/4})$ such that Assumption 4.2 holds for the CU-GMM criterion (4) over $\Theta_{osn} = \{\theta \in \Theta : \|g(\theta)\| \le r_n/\sqrt{n}\}$, where $\ell_n = -\frac{1}{2} Z_n' \Omega^{-1} Z_n$, $Z_n = \mathbb{G}_n(\rho_{\theta^*})$, $\gamma(\theta) = [\Omega^{-1/2}\mathbb{T}g(\theta)]_1$, and $\sqrt{n}\hat{\gamma}_n = \mathbb{V}_n = -[\Omega^{-1/2} Z_n]_1$ and $\Sigma = I_{d^*}$.
If $\mathcal{G}$ contains a neighborhood of the origin then $\gamma(\theta) = \Omega^{-1/2} g(\theta)$ and $\sqrt{n}\hat{\gamma}_n = \mathbb{V}_n = -\Omega^{-1/2} Z_n$.

Proposition 5.4. *Let all the conditions of Proposition 5.3 hold and let: (e) $\|\widehat{W} - \Omega^{-1}\| = o_{\mathbb{P}}(1)$.*
Then: the conclusions of Proposition 5.3 hold for the optimally-weighted GMM criterion (5).

5.2.1 Moment inequality models

Consider the moment inequality model $\{\tilde{\rho}(X_i, \mu) : \mu \in M\}$ where $\tilde{\rho}$ is a d_ρ vector of moments and the space is $M \subseteq \mathbb{R}^{d_\mu}$. The identified set for μ is $M_I = \{\mu \in M : E[\tilde{\rho}(X_i, \mu)] \le 0\}$ (the inequality is understood to hold element-wise). We may reformulate the moment inequality model as a moment equality model by augmenting the parameter vector with a vector of slackness parameters $\eta \in H = \mathbb{R}_+^{d_\rho}$. Thus we re-parameterize the model by $\theta = (\mu, \eta) \in \Theta = M \times H$ and write the inequality model as a GMM model with

$$E[\rho_\theta(X_i)] = 0 \text{ for } \theta \in \Theta_I, \quad \rho_\theta(X_i) = \tilde{\rho}(X_i, \mu) + \eta \ , \tag{26}$$

where the identified set for θ is $\Theta_I = \{\theta \in \Theta : E[\rho_\theta(X_i)] = 0\}$ and M_I is the projection of Θ_I onto M. Here the objective function would be as in display (4) or (5) using $\rho_\theta(X_i) = \tilde{\rho}(X_i, \mu) + \eta$. We may then apply Propositions 5.3 or 5.4 to the reparameterized GMM model (26).

As the parameter of interest is μ, one could use our Procedures 2 or 3 for inference on M_I. These procedures involve the profile criterion $\sup_{\eta \in H} L_n(\mu, \eta)$ which is simple to compute because the GMM objective function is quadratic in η for given μ (since the optimal weighting or continuous updating weighting matrix will typically not depend on η). See Example 3 in Subsection 5.3.3.

5.3 Examples

5.3.1 Example 1: missing data model in Subsection 3.1.1

We revisit the missing data example in Subsection 3.1.1, where the parameter space Θ for $\theta = (\mu, \eta_1, \eta_2)$ is given in (15), the identified set for θ is Θ_I given in (16), and the identified set for μ is $M_I = [\tilde{\gamma}_{11}, \tilde{\gamma}_{11} + \tilde{\gamma}_{00}]$.

Inference under partial identification: Consider the case in which the model is partially identified (i.e. $0 < \eta_2 < 1$). The likelihood of the i-th observation $(D_i, Y_i D_i) = (d, yd)$ is

$$p_\theta(d, yd) = [\tilde{\gamma}_{11}(\theta)]^{yd}[1 - \tilde{\gamma}_{11}(\theta) - \tilde{\gamma}_{00}(\theta)]^{d-yd}[\tilde{\gamma}_{00}(\theta)]^{1-d} = q_{\tilde{\gamma}(\theta)}(d, yd)$$

where:

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

with $\widetilde{\Gamma} = \{\tilde{\gamma}(\theta) : \theta \in \Theta\} = \{(g_{11} - \tilde{\gamma}_{11}, g_{00} - \tilde{\gamma}_{00}) : (g_{11}, g_{00}) \in [0,1]^2, 0 \leq g_{11} \leq 1 - g_{00}\}$. Conditions (a)-(b) of Proposition 5.1 hold and Assumption 4.2 is satisfied with $\gamma(\theta) = \mathbb{I}_0^{1/2}\tilde{\gamma}(\theta)$,

$$\mathbb{I}_0 = \begin{bmatrix} \frac{1}{\tilde{\gamma}_{11}} + \frac{1}{1-\tilde{\gamma}_{11}-\tilde{\gamma}_{00}} & \frac{1}{1-\tilde{\gamma}_{11}-\tilde{\gamma}_{00}} \\ \frac{1}{1-\tilde{\gamma}_{11}-\tilde{\gamma}_{00}} & \frac{1}{\tilde{\gamma}_{00}} + \frac{1}{1-\tilde{\gamma}_{11}-\tilde{\gamma}_{00}} \end{bmatrix} \qquad \sqrt{n}\hat{\gamma}_n = \mathbb{V}_n = \mathbb{I}_0^{-1/2}\mathbb{G}_n \begin{pmatrix} \frac{yd}{\tilde{\gamma}_{11}} - \frac{d-yd}{1-\tilde{\gamma}_{11}-\tilde{\gamma}_{00}} \\ \frac{1-d}{\tilde{\gamma}_{00}} - \frac{d-yd}{1-\tilde{\gamma}_{11}-\tilde{\gamma}_{00}} \end{pmatrix}$$

$\Sigma = I_2$ and $T = \mathbb{R}^2$. A flat prior on Θ in (15) induces a flat prior on Γ, which verifies Condition (c) of Proposition 5.1 and Assumption 4.3. Therefore, Theorem 4.1(ii) implies that our CSs $\widehat{\Theta}_\alpha$ for Θ_I has asymptotically exact coverage.

Now consider CSs for $M_I = [\tilde{\gamma}_{11}, \tilde{\gamma}_{11} + \tilde{\gamma}_{00}]$. Here $H_\mu = \{(\eta_1, \eta_2) \in [0,1]^2 : 0 \leq \mu - \eta_1(1 - \eta_2) \leq \eta_2\}$. By concavity in μ, the profile log-likelihood for M_I is:

$$PL_n(M_I) = \min_{\mu \in \{\underline{\mu}, \overline{\mu}\}} \sup_{\eta \in H_\mu} \mathbb{P}_n \log p_{(\mu,\eta)}$$

where $\underline{\mu} = \tilde{\gamma}_{11}$ and $\overline{\mu} = \tilde{\gamma}_{11} + \tilde{\gamma}_{00}$. The inner maximization problem is:

$$\sup_{\eta \in H_\mu} \mathbb{P}_n \log p_{(\mu,\eta)} = \sup_{\substack{0 \leq g_{11} \leq \mu \\ \mu \leq g_{11} + g_{00} \leq 1}} \mathbb{P}_n \Big(yd \log g_{11} + (d - yd) \log(1 - g_{11} - g_{00}) + (1 - d) \log g_{00} \Big).$$

Let $g = (g_{11}, g_{00})'$ and $\tilde{\gamma} = (\tilde{\gamma}_{11}, \tilde{\gamma}_{00})'$ and let:

$$T_\mu = \bigcup_{n \geq 1} \left\{ \sqrt{n}\mathbb{I}_0^{1/2}(g - \tilde{\gamma}) \,:\, 0 \leq g_{11} \leq \mu,\ \mu \leq g_{11} + g_{00} \leq 1,\ \|g - \tilde{\gamma}\|^2 \leq r_n^2/n \right\}$$

where r_n is from Proposition 5.1. It follows that:

$$nPL_n(M_I) = \ell_n + \frac{1}{2}\|\mathbb{V}_n\|^2 - \max_{\mu \in \{\underline{\mu}, \overline{\mu}\}} \frac{1}{2} \inf_{t \in T_\mu} \|\mathbb{V}_n - t\|^2 + o_{\mathbb{P}}(1)$$
$$PQ_n(M_I) = \max_{\mu \in \{\underline{\mu}, \overline{\mu}\}} \inf_{t \in T_\mu} \|\mathbb{V}_n - t\|^2 + o_{\mathbb{P}}(1)\,.$$

Equation (24) and Assumption 4.7 therefore hold with $f(v) = \max_{\mu \in \{\underline{\mu}, \overline{\mu}\}} \inf_{t \in T_\mu} \|v - t\|^2$ where $T_{\underline{\mu}}$ and $T_{\overline{\mu}}$ are regular halfspaces in $\mathbb{R}^2$. Theorem 4.4 implies that the CS $\widehat{M}_\alpha^\chi$ is asymptotically valid (but conservative) for M_I.

To verify Assumption 4.5, take n sufficiently large that $\gamma(\theta) \in \text{int}(\Gamma)$ for all $\theta \in \Theta_{osn}$. Then:

$$PL_n(M(\theta)) = \min_{\mu \in \{\tilde{\gamma}_{11}(\theta), \tilde{\gamma}_{11}(\theta)+\tilde{\gamma}_{00}(\theta)\}} \sup_{\eta \in H_\mu} \mathbb{P}_n \log p_{(\mu,\eta)} \,. \tag{27}$$

This is geometrically the same as the profile QLR for M_I up to a translation of the local parameter space from $(\tilde{\gamma}_{11}, \tilde{\gamma}_{00})'$ to $(\tilde{\gamma}_{11}(\theta), \tilde{\gamma}_{00}(\theta))'$. The local parameter spaces are approximated by $T_{\underline{\mu}}(\theta) = T_{\underline{\mu}} + \sqrt{n}\gamma(\theta)$ and $T_{\overline{\mu}}(\theta) = T_{\overline{\mu}} + \sqrt{n}\gamma(\theta)$. It follows that uniformly in $\theta \in \Theta_{osn}$,

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

verifying Assumption 4.5. Theorem 4.3(ii) implies that $\widehat{M}_\alpha$ has asymptotically exact coverage.

Inference under identification: Now consider the case in which the model is identified (i.e. $\eta_2 = 1$ and $\tilde{\gamma}_{00} = 0$) and $M_I = \{\mu_0\}$. Here each $D_i = 1$ so the likelihood of the i-th observation $(D_i, Y_i D_i) = (1, y)$ is

$$p_\theta(1, y) = [\tilde{\gamma}_{11}(\theta)]^y [1 - \tilde{\gamma}_{11}(\theta) - \tilde{\gamma}_{00}(\theta)]^{1-y} = q_{\tilde{\gamma}(\theta)}(1, y)$$

Lemma F.5 in Appendix F shows that with Θ as in (15) and a flat prior, the posterior Π_n concentrates on the local neighborhood $\Theta_{osn} = \{\theta : |\tilde{\gamma}_{11}(\theta) - \tilde{\gamma}_{11}| \le r_n/\sqrt{n}, \tilde{\gamma}_{00}(\theta) \le r_n/n\}$ for any positive sequence $(r_n)_{n\in\mathbb{N}}$ with $r_n \to \infty$, $r_n/\sqrt{n} = o(1)$.

In this case, the reduced-form parameter is $\tilde{\gamma}_{11}(\theta)$ and the singular part is $\gamma_\perp(\theta) = \tilde{\gamma}_{00}(\theta) \ge 0$. Uniformly over Θ_{osn} we obtain:

$$nL_n(\theta) = \ell_n - \frac{1}{2}\frac{(\sqrt{n}(\tilde{\gamma}_{11}(\theta) - \tilde{\gamma}_{11}))^2}{\tilde{\gamma}_{11}(1 - \tilde{\gamma}_{11})} + \frac{\sqrt{n}(\tilde{\gamma}_{11}(\theta) - \tilde{\gamma}_{11})}{\tilde{\gamma}_{11}(1 - \tilde{\gamma}_{11})}\mathbb{G}_n(y) - n\tilde{\gamma}_{00}(\theta) + o_{\mathbb{P}}(1)$$

which verifies Assumption 4.2'(i) with

$$\gamma(\theta) = \frac{\tilde{\gamma}_{11}(\theta) - \tilde{\gamma}_{11}}{\sqrt{\tilde{\gamma}_{11}(1 - \tilde{\gamma}_{11})}} \qquad \sqrt{n}\hat{\gamma}_n = \mathbb{V}_n = \frac{\mathbb{G}_n(y)}{\sqrt{\tilde{\gamma}_{11}(1 - \tilde{\gamma}_{11})}} \qquad f_{n,\perp}(\gamma_\perp(\theta)) = n\gamma_\perp(\theta)$$

and $T = \mathbb{R}$. The remaining parts of Assumption 4.2' are easily shown to be satisfied. Therefore, Theorem 4.2 implies that $\widehat{\Theta}_\alpha$ for Θ_I will be asymptotically valid but conservative.

For inference on $M_I = \{\mu_0\}$, the profile LR statistic is asymptotically χ^2_1 and equation (24) holds with $f(v) = v^2$ and $T = \mathbb{R}$. To verify Assumption 4.5, for each $\theta \in \Theta_{osn}$ we need to solve

$$\sup_{\eta \in H_\mu} \mathbb{P}_n \log p_{(\mu,\eta)} = \sup_{\substack{0 \le g_{11} \le \mu \\ \mu \le g_{11} + g_{00} \le 1}} \mathbb{P}_n\Big(y \log g_{11} + (1 - y)\log(1 - g_{11} - g_{00})\Big)$$

at $\mu = \tilde{\gamma}_{11}(\theta)$ and $\mu = \tilde{\gamma}_{11}(\theta) + \tilde{\gamma}_{00}(\theta)$. The maximum is achieved when g_{00} is as small as possible,

i.e., when $g_{00} = \mu - g_{11}$. Substituting in and maximizing with respect to g_{11}:

$$\sup_{\eta \in H_\mu} \mathbb{P}_n \log p_{(\mu,\eta)} = \mathbb{P}_n \big(y \log \mu + (1-y) \log(1-\mu) \big) .$$

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

$$\begin{aligned} nPL_n(M(\theta)) &= \ell_n + \frac{1}{2}\mathbb{V}_n^2 - \frac{1}{2}\Big(\big(\mathbb{V}_n - \sqrt{n}\gamma(\theta)\big)^2 \vee \big(\mathbb{V}_n - \sqrt{n}(\gamma(\theta) + \tilde{\gamma}_{00}(\theta))\big)^2 \Big) + o_{\mathbb{P}}(1) \\ &= \ell_n + \frac{1}{2}\mathbb{V}_n^2 - \frac{1}{2}\big(\mathbb{V}_n - \sqrt{n}\gamma(\theta)\big)^2 + o_{\mathbb{P}}(1) \end{aligned}$$

where the last equality holds because $\sup_{\theta \in \Theta_{osn}} \tilde{\gamma}_{00}(\theta) \leq r_n/n = o(n^{-1/2})$. This verifies that Assumption 4.5 holds with $f(v) = v^2$. Thus Theorem 4.3(ii) implies that $\widehat{M}_\alpha$ has asymptotically exact coverage for M_I, even though $\widehat{\Theta}_\alpha$ is conservative for Θ_I in this case.

5.3.2 Example 2: entry game with correlated shocks in Subsection 3.1.2

Consider the bivariate discrete game with payoffs described in Subsection 3.1.2. Here we consider a slightly more general setting, in which Q_ρ denotes a general joint distribution (not just bivariate Gaussian) for (ϵ_1, ϵ_2) indexed by a parameter ρ. This model falls into the class of models dealt with in Proposition 5.1. Conditions (a)-(b) and (d) of Proposition 5.1 hold with $\tilde{\gamma}(\theta) = (\tilde{\gamma}_{00}(\theta), \tilde{\gamma}_{10}(\theta), \tilde{\gamma}_{11}(\theta))'$ and $\widetilde{\Gamma} = \{\tilde{\gamma}(\theta) : \theta \in \Theta\}$ under very mild conditions on the parameterization $\theta \mapsto \tilde{\gamma}(\theta)$ (which, in turn, is determined by the specification of Q_ρ). Assumption 4.2 is therefore satisfied with:

$$\mathbb{I}_0 = \begin{bmatrix} \frac{1}{\tilde{\gamma}_{00}} & 0 & 0 \\ 0 & \frac{1}{\tilde{\gamma}_{10}} & 0 \\ 0 & 0 & \frac{1}{\tilde{\gamma}_{11}} \end{bmatrix} + \frac{1}{1 - \tilde{\gamma}_{00} - \tilde{\gamma}_{10} - \tilde{\gamma}_{11}} \mathbf{1}_{3\times 3}$$

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

$$\sqrt{n}\hat{\gamma}_n = \mathbb{V}_n = \mathbb{I}_0^{-1/2} \mathbb{G}_n \begin{pmatrix} \frac{d_{00}}{\tilde{\gamma}_{00}} - \frac{1-d_{00}-d_{10}-d_{11}}{1-\tilde{\gamma}_{00}-\tilde{\gamma}_{10}-\tilde{\gamma}_{11}} \\ \frac{d_{01}}{\tilde{\gamma}_{10}} - \frac{1-d_{00}-d_{10}-d_{11}}{1-\tilde{\gamma}_{00}-\tilde{\gamma}_{10}-\tilde{\gamma}_{11}} \\ \frac{d_{11}}{\tilde{\gamma}_{11}} - \frac{1-d_{00}-d_{10}-d_{11}}{1-\tilde{\gamma}_{00}-\tilde{\gamma}_{10}-\tilde{\gamma}_{11}} \end{pmatrix} \rightsquigarrow N(0, I_3)$$

and $T = \mathbb{R}^3$. Condition (c) of Proposition 5.1 and Assumption 4.3 can be verified under mild conditions on the map $\theta \mapsto \tilde{\gamma}(\theta)$ and the prior Π. For instance, consider the parameterization $\theta = (\Delta_1, \Delta_2, \beta_1, \beta_2, \rho, s)$ where the joint distribution of (ϵ_1, ϵ_2) is a bivariate Normal with mean zero, standard deviations one and positive correlation $\rho \in [0, 1]$. The parameter space is

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

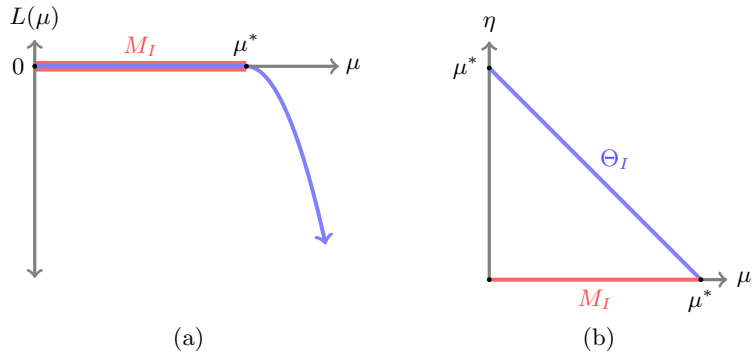

Figure 5: Panel (a): identified set M_I for μ as the argmax of the population (moment inequality) criterion $L(\mu) = -\frac{1}{2}((\mu - \mu^*) \vee 0)^2$. Panel (b): identified set Θ_I for $\theta = (\mu, \eta)$ for the moment equality model $E[\mu + \eta - X] = 0$.

where $-\infty < \underline{\Delta} < \overline{\Delta} < 0$ and $-\infty < \underline{\beta} < \overline{\beta} < \infty$. The image measure Π_Γ of a flat prior on Θ is positive and continuous on a neighborhood of the origin, which verifies Condition (c) of Proposition 5.1 and Assumption 4.3. Therefore, Theorem 4.1(ii) implies that our MC CSs for Θ_I will have asymptotically exact coverage.

5.3.3 Example 3: a moment inequality model

As a simple illustration, suppose that $\mu \in M = \mathbb{R}_+$ is identified by the inequality $\mathbb{E}[\mu - X_i] \leq 0$ where $X_1, \ldots, X_n$ are i.i.d. with unknown mean $\mu^* \in \mathbb{R}_+$ and unit variance. The identified set for μ is $M_I = [0, \mu^*]$, which is the argmax of the population criterion function $L(\mu) = -\frac{1}{2}((\mu - \mu^*) \vee 0)^2$ (see Figure 5). The sample criterion $-\frac{1}{2}((\mu - \bar{X}_n) \vee 0)^2$ is typically used in the moment inequality literature but violates our Assumption 4.2. However, we can rewrite the model as the moment equality model: $\mathbb{E}[\mu + \eta - X_i] = 0$ where $\eta \in H = \mathbb{R}_+$ is a slackness parameter. The parameter space for $\theta = (\mu, \eta)$ is $\Theta = \mathbb{R}_+^2$. The identified set for θ is $\Theta_I = \{(\mu, \eta) \in \Theta : \mu + \eta = \mu^*\}$ and the identified set for μ is M_I (see Figure 5). The GMM objective function is then:

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

It is straightforward to show that $2nL_n(\hat{\mu}, \hat{\eta}) = -((\mathbb{V}_n + \sqrt{n}\mu^*) \wedge 0)^2$ where $\mathbb{V}_n = \sqrt{n}(\bar{X}_n - \mu^*)$. Moreover, $\sup_{\eta \in H_\mu} 2nL_n(\mu, \eta) = -((\mathbb{V}_n + \sqrt{n}(\mu^* - \mu)) \wedge 0)^2$ and so the profile QLR for M_I is $PQ_n(M_I) = (\mathbb{V}_n \wedge 0)^2 - ((\mathbb{V}_n + \sqrt{n}\mu^*) \wedge 0)^2$.

For the posterior of the profile QLR, we also have $\Delta(\theta^b) = \{\theta \in \Theta : \mu + \eta = \mu^b + \eta^b\}$ and $M(\theta^b) = [0, \mu^b + \eta^b]$. The profile QLR for $M(\theta^b)$ is

$$PQ_n(M(\theta^b)) = ((\mathbb{V}_n - \sqrt{n}(\mu^b + \eta^b - \mu^*)) \wedge 0)^2 - ((\mathbb{V}_n + \sqrt{n}\mu^*) \wedge 0)^2$$

This maps into our framework with the local reduced-form parameter $\gamma(\theta) = \mu + \eta - \mu^*$. Consider the case $\mu^* \in (cn^{\alpha-1/2}, \infty)$ where $c > 0$ and $\alpha \in (0, \frac{1}{2}]$ are positive constants (we consider this case for the moment just to illustrate verification of our conditions). Here $T = \mathbb{R}$ and a positive continuous prior on μ and η induces a prior on γ that is positive and continuous at the origin. Moreover, Assumption 4.5 holds with $f(\kappa) = (\kappa \wedge 0)^2$. The regularity conditions of Theorem 4.3 hold, and hence $\widehat{M}_\alpha$ has asymptotically exact coverage for M_I.

More generally, Appendix C.2 shows that under very mild conditions our CS $\widehat{M}_\alpha$ is uniformly valid over a class of DGPs $\mathbf{P}$, i.e.:

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

where $M_I(\mathbb{P}) = [0, \mu^*(\mathbb{P})]$ and the set $\mathbf{P}$ allows for any mean $\mu^*(\mathbb{P}) \in \mathbb{R}_+$ (encompassing, in particular, point-identified, partially identified, and drifting-to-point identified cases). In contrast, we construct sequences of DGPs $(\mathrm{P}_n)_{n\in\mathbb{N}} \subset \mathbf{P}$ along which bootstrap-based CSs $\widehat{M}_\alpha^{boot}$ fail to cover with the prescribed coverage probability, i.e.:

$$\limsup_{n\to\infty} \mathrm{P}_n(\mathbb{M}_I(\mathrm{P}_n) \subseteq \widehat{M}_\alpha^{boot}) < \alpha\,.$$

This reinforces the fact that our MC CSs for M_I have very different asymptotic properties from bootstrap-based CSs for M_I.

6 Conclusion

We propose new methods for constructing CSs for identified sets in partially-identified econometric models. Our CSs are relatively simple to compute and have asymptotically valid 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 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 MC CSs for Θ_I may be slightly conservative, but our MC CSs for identified sets of subvectors could still be asymptotically exact. Simulation experiments demonstrate the good finite-sample coverage properties of our proposed CS constructions in standard difficult situations. We also illustrate our proposed CSs in two realistic empirical examples.

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

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.

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

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

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

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

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*.

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. and F. Gao (2017). A reverse gaussian correlation inequality by adding cones. *Statistics and Probability Letters 123*, 84–87.

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

Chernoff, H. (1954). On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics 25*(3), pp. 573–578.

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.

Chopin, N. (2002). A sequential particle filter method for static models. *Biometrika 89*(3), 539–552.

Chopin, N. (2004, 12). Central limit theorem for sequential monte carlo methods and its application to bayesian inference. *Ann. Statist. 32*(6), 2385–2411.

Del Moral, P., A. Doucet, and A. Jasra (2006). Sequential monte carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68*(3), 411–436.

Del Moral, P., A. Doucet, and A. Jasra (2012, 02). On adaptive resampling strategies for sequential monte carlo methods. *Bernoulli 18*(1), 252–278.

Durham, G. and J. Geweke (2014). *Adaptive Sequential Posterior Simulators for Massively Parallel Computing Environments*, Chapter 1, pp. 1–44.

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.

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

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

Geyer, C. J. (1994). On the asymptotics of constrained M-estimation. *The Annals of Statistics 22*(4), 1993–2010.

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

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

Herbst, E. and F. Schorfheide (2014). Sequential monte carlo sampling for DSGE models. *Journal of Applied Econometrics 29*(7), 1073–1098.

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

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 (2016). Bayesian inference in a class of partially identified models. *Quantitative Economics 6*, 329–366.

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

Liao, Y. and A. Simoni (2016). Posterior properties of the support function for set inference. Working paper, Rutgers University and CREST.

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

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

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

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

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

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

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

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

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

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

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

Schwartz, L. (1965, Mar). On bayes procedures. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 4*(1), 10–26.

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

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 details for the simulations and applications

A.1 An adaptive Sequential Monte Carlo algorithm

We use an adaptive Sequential Monte Carlo (SMC) algorithm to sample from the quasi-posterior in (6). Conventional MCMC algorithms such as the Metropolis-Hastings algorithm may fail to generate representative samples from the quasi-posterior in partially identified models or, more generally, models with multi-modal quasi-posteriors. For instance, the MCMC chain may get stuck exploring a single mode and fail to explore other modes if there is insufficient mass bridging the modes. In contrast, the SMC algorithm we use propagates large clouds of draws, in parallel, over a sequence of tempered distributions which begins with the prior, slowly incorporates information from the criterion, and ends with the quasi-posterior. The algorithm sequentially discards draws with relatively low mass as information is added, duplicates those with relatively high mass, then mutates the draws via a MCMC step to generate new draws (preventing particle impoverishment). Moreover, the algorithm is adaptive, i.e., the tuning parameters for the sequence of proposal distributions in the MCMC step are determined in a data-driven way.

The algorithm we use and its exposition below closely follows Herbst and Schorfheide (2014) who adapt a generic adaptive SMC algorithm to deal with large-scale DSGE models.[23] A similar algorithm is proposed by Durham and Geweke (2014), who emphasize its parallelizability. Let J and K be positive integers and let $\phi_1, \ldots, \phi_J$ be an increasing sequence with $\phi_1 = 0$ and $\phi_J = 1$. Set $w_1^b = 1$ for $b = 1, \ldots, B$ and draw $\theta_1^1, \ldots, \theta_1^B$ from the prior $\Pi(\theta)$. Then for $j = 2, \ldots, J$:

1. Correction: Let $v_j^b = e^{(\phi_j - \phi_{j-1}) n L_n(\theta_{j-1}^b)}$ and $w_j^b = (v_j^b w_{j-1}^b)/(\frac{1}{B}\sum_{b=1}^B v_j^b w_{j-1}^b)$.
2. Selection: Compute the effective sample size $ESS_j = B/(\frac{1}{B}\sum_{b=1}^B (w_j^b)^2)$. Then:
   (a) If $ESS_j > \frac{B}{2}$: set $\vartheta_j^b = \theta_{j-1}^b$ for $b = 1, \ldots, B$; or
   (b) If $ESS_j \leq \frac{B}{2}$: draw an i.i.d. sample $\vartheta_j^1, \ldots, \vartheta_j^B$ from the multinomial distribution with support $\theta_{j-1}^1, \ldots, \theta_{j-1}^B$ and weights $w_j^1, \ldots, w_j^B$, then set $w_j^b = 1$ for $b = 1, \ldots, B$.
3. Mutation: Run B separate and independent MCMC chains of length K using the random-walk Metropolis-Hastings algorithm initialized at each ϑ_j^b for the tempered quasi-posterior $\Pi_j(\theta|\mathbf{X}_n) \propto e^{\phi_j n L_n(\theta)}\Pi(\theta)$ and let θ_j^b be the final draw of the bth chain.

The resulting sample is $\theta^b = \theta_J^b$ for $b = 1, \ldots, B$. Multinomial resampling (step 2) and the B independent MCMC chains (step 3) can both be computed in parallel, so the additional computational time relative to conventional MCMC methods is modest.

[23] See Chopin (2002, 2004) and Del Moral, Doucet, and Jasra (2006) for the generic SMC algorithm for estimating static model parameters, Del Moral et al. (2012) and references therein for adaptive selection of tuning parameters with a SMC framework.

In practice, we take $J = 200$, $J = 1$, 4 or 8 (see below for the specific K used in the simulations and empirical applications), and $\phi_j = (\frac{j-1}{J-1})^\lambda$ with $\lambda = 2$. When the dimension of θ is low, in step 3 we use a $N(0, \sigma_j^2 I)$ proposal density (all parameters are transformed to have full support) where σ_j is chosen adaptively to target an acceptance ratio ≈ 0.35 by setting $\sigma_2 = 1$ and

$$\sigma_j = \sigma_{j-1}\Big(0.95 + 0.10\frac{e^{16(A_{j-1}-0.35)}}{1+e^{16(A_{j-1}-0.35)}}\Big)$$

for $j > 2$, where A_{j-1} is the acceptance ratio from the previous iteration. If the dimension of θ is large, we partition ϑ_j^b into L random blocks (we assign each element of ϑ_j^b to a block by drawing from the uniform distribution on $\{1, \ldots, L\}$) then apply a blockwise random-walk Metropolis-Hastings (i.e. Metropolis-within-Gibbs) algorithm. Here the proposal density we use for block $l \in \{1, \ldots, L\}$ is $N(0, \sigma_j^2 \Sigma_{j-1}^l)$ where σ_j is chosen as before, Σ_{j-1} is the covariance of the draws from iteration $j-1$, and Σ_j^l is the sub-matrix of Σ_j corresponding to block l.

As the SMC procedure uses a particle approximation to the posterior, in practice compute quantiles for procedure 1 using:

$$\Pi(\{\theta : Q_n(\theta) \leq z\}|\mathbf{X}_n) = \frac{1}{B}\sum_{b=1}^{B} w_J^b 1\!\!1\{Q_n(\theta^b) \leq z\} \tag{28}$$

and similarly for the profile QLR for procedure 2.

A.2 Example 1: missing data

SMC algorithm: We implement the SMC algorithm with $K = 1$ and a $N(0, \sigma_j^2 I)$ proposal in the mutation step for all simulations for this example.

Additional simulation results: Here we present additional simulation results for the missing data example using (i) a likelihood criterion and curved prior and (ii) a continuously-updated GMM criterion and flat prior. For the "curved" prior, we take $\pi(\mu, \eta_1, \eta_2) = \pi(\mu|\eta_1, \eta_2)\pi(\eta_1)\pi(\eta_2)$ with $\pi(\eta_1) = \text{Beta}(3, 8)$, $\pi(\eta_2) = \text{Beta}(8, 1)$, and $\pi(\mu|\eta_1, \eta_2) = U[\eta_1(1-\eta_2), \eta_2 + \eta_1(1-\eta_2)]$. Figure 6 plots the marginal curved priors for η_1 and η_2.

Results for the likelihood criterion with curved prior are presented in Table 6, and are very similar to those presented in Table 1, though the coverage of percentile-based CSs is worse here for the partially identified cases ($c = 1, 2$). Results for the CU-GMM criterion and flat prior are presented in Table 7. Results for Procedures 2 and 3 are very similar to the results with a likelihood criterion and show coverage very close to nominal coverage in point point- and partially-identified cases. Here procedure 1 does not over-cover in the point-identified case because the weighting matrix is singular when the model is identified, which forces the draws to concentrate on the region in which

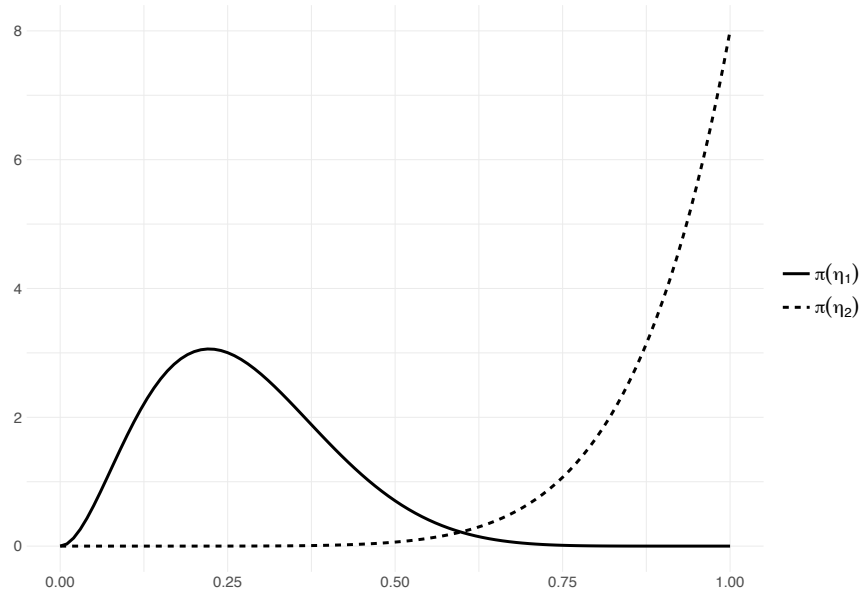

Figure 6: Missing data example: Marginal "curved" priors for η_1 (solid line) and η_2 (dashed line).

$\eta_2 = 1$. This, in turn, means projection is no longer conservative in the point-identified case, though it is still very conservative in the partially-identified cases. Percentile CSs again under-cover badly in the partially-identified case.

A.3 Example 2: entry game with correlated shocks

SMC Algorithm: As there are 6 partially-identified parameters here instead of 2 in the previous example, we initially increased J to reduce the distance between the successive tempered distributions. Like Herbst and Schorfheide (2014), whose DSGE examples use $(J, K) = (500, 1)$, we also found the effect of increasing K similar to the effect of increasing J. We therefore settled on $(J, K) = (200, 4)$ which was more computationally efficient than using larger J. We again use a $N(0, \sigma_j^2 I)$ proposal in the mutation step for all simulations for this example.

Procedure 2: Unlike the missing data example, where $M(\theta)$ is known in closed form, here the set $M(\theta)$ is no longer known in closed form if $\rho \neq 0$. We therefore calculate $M(\theta^b)$ for $b = 1, \ldots, B$ numerically in order to implement procedure 2 for $\mu = \Delta_1$ (in which case $\eta = (\Delta_2, \beta_1, \beta_2, \rho, s)$) and $\mu = \beta_1$ (in which case $\eta = (\Delta_1, \Delta_2, \beta_2, \rho, s)$) . Let $D_{KL}(p_\theta \| p_\vartheta)$ denote the KL distance between p_θ and p_ϑ or any $\theta, \vartheta \in \Theta$, which is given by

$$D_{KL}(p_\theta \| p_\vartheta) = \sum_{\{i,j\} \in \{0,1\}^2} p_\theta(a_1 = i, a_2 = j) \log \left(\frac{p_\theta(a_1 = i, a_2 = j)}{p_\vartheta(a_1 = i, a_2 = j)} \right)$$

	$\eta_2 = 1 - \frac{2}{\sqrt{n}}$						$\eta_2 = 1 - \frac{1}{\sqrt{n}}$						$\eta_2 = 1$ (Point ID)					
	0.90		0.95		0.99		0.90		0.95		0.99		0.90		0.95		0.99	
	$\widehat{\Theta}_\alpha$ (Procedure 1)																	
100	.911	—	.957	—	.990	—	.898	—	.950	—	.989	—	.985	—	.994	—	.999	—
250	.906	—	.954	—	.993	—	.899	—	.951	—	.991	—	.992	—	.997	—	1.000	—
500	.908	—	.956	—	.992	—	.910	—	.957	—	.991	—	.994	—	.998	—	1.000	—
1000	.896	—	.948	—	.989	—	.905	—	.954	—	.989	—	.996	—	.999	—	1.000	—
	$\widehat{M}_\alpha$ (Procedure 2)																	
100	.890	[.33,.67]	.948	[.31,.69]	.990	[.29,.71]	.911	[.37,.63]	.952	[.36,.64]	.992	[.33,.67]	.912	[.41,.58]	.961	[.40,.60]	.990	[.37,.63]
250	.905	[.39,.61]	.953	[.38,.62]	.991	[.36,.64]	.913	[.42,.58]	.957	[.41,.59]	.992	[.39,.61]	.917	[.45,.55]	.962	[.44,.56]	.993	[.42,.58]
500	.913	[.42,.58]	.957	[.41,.59]	.992	[.40,.60]	.915	[.44,.56]	.955	[.43,.57]	.993	[.42,.58]	.919	[.46,.54]	.959	[.46,.54]	.993	[.44,.56]
1000	.898	[.44,.56]	.948	[.44,.56]	.988	[.43,.57]	.898	[.46,.54]	.948	[.45,.55]	.990	[.44,.56]	.913	[.47,.53]	.954	[.47,.53]	.992	[.46,.54]
	$\widehat{M}_\alpha^\chi$ (Procedure 3)																	
100	.922	[.32,.68]	.950	[.31,.69]	.989	[.28,.72]	.912	[.37,.63]	.942	[.36,.64]	.989	[.33,.67]	.909	[.42,.58]	.941	[.40,.59]	.985	[.38,.62]
250	.910	[.39,.61]	.949	[.38,.62]	.989	[.36,.64]	.913	[.42,.58]	.951	[.41,.59]	.990	[.39,.61]	.891	[.45,.55]	.952	[.44,.56]	.992	[.42,.58]
500	.898	[.42,.58]	.952	[.41,.59]	.992	[.40,.60]	.910	[.44,.56]	.947	[.44,.56]	.990	[.42,.58]	.911	[.46,.54]	.948	[.46,.54]	.992	[.44,.56]
1000	.892	[.44,.56]	.944	[.44,.56]	.989	[.43,.57]	.892	[.46,.54]	.944	[.45,.55]	.986	[.45,.55]	.902	[.48,.52]	.941	[.47,.53]	.987	[.46,.54]
	$\widehat{M}_\alpha^{proj}$ (Projection)																	
100	.971	[.30,.70]	.988	[.29,.71]	.997	[.26,.74]	.966	[.35,.65]	.987	[.33,.67]	.997	[.31,.69]	.985	[.38,.62]	.994	[.36,.64]	.999	[.34,.66]
250	.971	[.37,.63]	.987	[.36,.64]	.999	[.34,.66]	.969	[.40,.60]	.986	[.39,.61]	.998	[.37,.62]	.992	[.42,.58]	.997	[.41,.59]	1.000	[.39,.61]
500	.975	[.41,.59]	.988	[.40,.60]	.998	[.39,.61]	.972	[.43,.57]	.989	[.42,.58]	.998	[.41,.59]	.994	[.44,.56]	.998	[.44,.57]	1.000	[.42,.58]
1000	.965	[.44,.56]	.983	[.43,.57]	.997	[.42,.58]	.966	[.45,.55]	.985	[.45,.55]	.998	[.44,.56]	.996	[.45,.55]	.999	[.45,.55]	1.000	[.44,.56]
	$\widehat{M}_\alpha^{perc}$ (Percentile)																	
100	.000	[.37,.54]	.037	[.36,.56]	.398	[.33,.59]	.458	[.40,.56]	.646	[.38,.57]	.866	[.35,.60]	.902	[.42,.58]	.951	[.40,.59]	.989	[.37,.62]
250	.000	[.41,.53]	.075	[.40,.54]	.438	[.38,.56]	.480	[.43,.54]	.653	[.42,.55]	.867	[.40,.57]	.909	[.45,.55]	.954	[.44,.56]	.992	[.42,.58]
500	.000	[.44,.52]	.098	[.43,.53]	.468	[.42,.54]	.488	[.45,.53]	.660	[.44,.54]	.878	[.43,.55]	.910	[.46,.54]	.955	[.46,.54]	.991	[.44,.56]
1000	.000	[.46,.52]	.107	[.45,.52]	.472	[.44,.53]	.483	[.47,.52]	.655	[.46,.53]	.866	[.45,.54]	.901	[.47,.53]	.948	[.47,.53]	.989	[.46,.54]
	Comparison with GMS CSs for μ via moment inequalities																	
100	.810	[.34,.66]	.909	[.32,.68]	.981	[.29,.71]	.806	[.39,.61]	.896	[.37,.63]	.979	[.34,.66]	.894	[.42,.58]	.943	[.40,.60]	.974	[.39,.61]
250	.800	[.40,.60]	.897	[.39,.61]	.978	[.36,.63]	.798	[.43,.57]	.897	[.42,.58]	.976	[.40,.60]	.903	[.45,.55]	.951	[.44,.56]	.984	[.43,.57]
500	.795	[.43,.57]	.902	[.42,.58]	.980	[.40,.60]	.788	[.45,.55]	.894	[.44,.56]	.977	[.43,.57]	.905	[.46,.54]	.950	[.46,.54]	.987	[.45,.55]
1000	.785	[.45,.55]	.885	[.44,.56]	.973	[.43,.57]	.785	[.46,.54]	.884	[.46,.54]	.973	[.45,.55]	.895	[.47,.53]	.943	[.47,.53]	.986	[.46,.54]

Table 6: Missing data example: average coverage probabilities for Θ_I and M_I and average lower and upper bounds of CSs for M_I across 5000 MC replications. Procedures 1–3, Projection and Percentile are implemented using a likelihood criterion and **curved prior**.

	$\eta_2 = 1 - \frac{2}{\sqrt{n}}$						$\eta_2 = 1 - \frac{1}{\sqrt{n}}$						$\eta_2 = 1$ (Point ID)					
	0.90		0.95		0.99		0.90		0.95		0.99		0.90		0.95		0.99	
	$\widehat{\Theta}_\alpha$ (Procedure 1)																	
100	.910	—	.951	—	.989	—	.888	—	.938	—	.972	—	.914	—	.945	—	.990	—
250	.908	—	.958	—	.994	—	.901	—	.952	—	.988	—	.904	—	.950	—	.990	—
500	.912	—	.957	—	.992	—	.923	—	.965	—	.993	—	.904	—	.954	—	.990	—
1000	.909	—	.955	—	.989	—	.912	—	.963	—	.995	—	.911	—	.953	—	.989	—
	$\widehat{M}_\alpha$ (Procedure 2)																	
100	.921	[.32,.68]	.965	[.30,.70]	.993	[.27,.73]	.913	[.37,.63]	.949	[.35,.65]	.991	[.32,.68]	.914	[.42,.58]	.945	[.40,.60]	.990	[.37,.63]
250	.911	[.39,.61]	.956	[.38,.62]	.993	[.36,.64]	.913	[.42,.58]	.953	[.41,.59]	.993	[.39,.61]	.904	[.45,.55]	.950	[.44,.56]	.990	[.42,.58]
500	.908	[.42,.58]	.954	[.41,.59]	.991	[.40,.60]	.912	[.44,.56]	.953	[.44,.57]	.991	[.42,.58]	.904	[.46,.54]	.954	[.46,.54]	.990	[.44,.56]
1000	.904	[.44,.56]	.953	[.44,.56]	.989	[.43,.57]	.902	[.46,.54]	.950	[.45,.55]	.990	[.44,.56]	.911	[.48,.52]	.953	[.47,.53]	.989	[.46,.54]
	$\widehat{M}_\alpha^\chi$ (Procedure 3)																	
100	.921	[.32,.68]	.949	[.31,.69]	.989	[.28,.72]	.913	[.37,.63]	.946	[.35,.64]	.989	[.32,.68]	.914	[.42,.58]	.945	[.40,.60]	.990	[.37,.63]
250	.911	[.39,.61]	.951	[.38,.62]	.992	[.36,.64]	.914	[.42,.58]	.952	[.41,.59]	.993	[.39,.61]	.886	[.45,.55]	.950	[.44,.56]	.991	[.42,.58]
500	.909	[.42,.58]	.950	[.41,.59]	.990	[.40,.60]	.913	[.44,.56]	.948	[.44,.57]	.990	[.42,.58]	.904	[.46,.54]	.947	[.46,.54]	.989	[.44,.56]
1000	.902	[.44,.56]	.950	[.44,.56]	.989	[.43,.57]	.901	[.46,.54]	.950	[.45,.55]	.988	[.45,.55]	.902	[.48,.52]	.948	[.47,.53]	.988	[.46,.54]
	$\widehat{M}_\alpha^{proj}$ (Projection)																	
100	.977	[.29,.71]	.990	[.28,.72]	.999	[.24,.76]	.970	[.34,.66]	.988	[.33,.67]	.998	[.30,.70]	.914	[.42,.58]	.945	[.40,.60]	.990	[.37,.63]
250	.972	[.37,.63]	.992	[.36,.64]	.999	[.34,.66]	.976	[.40,.60]	.992	[.39,.61]	.999	[.37,.64]	.904	[.45,.55]	.950	[.44,.56]	.990	[.42,.58]
500	.975	[.41,.59]	.988	[.40,.60]	.998	[.39,.61]	.979	[.43,.57]	.992	[.42,.58]	1.000	[.40,.60]	.904	[.46,.54]	.954	[.46,.54]	.990	[.44,.56]
1000	.972	[.44,.57]	.986	[.43,.57]	.998	[.42,.58]	.975	[.45,.55]	.991	[.44,.56]	.999	[.43,.57]	.911	[.48,.52]	.953	[.47,.53]	.989	[.46,.54]
	$\widehat{M}_\alpha^{perc}$ (Percentile)																	
100	.399	[.38,.62]	.665	[.36,.64]	.939	[.31,.68]	.642	[.41,.59]	.800	[.39,.61]	.954	[.35,.65]	.912	[.42,.58]	.945	[.40,.60]	.990	[.37,.63]
250	.386	[.42,.58]	.642	[.41,.59]	.914	[.38,.62]	.641	[.44,.56]	.804	[.43,.57]	.954	[.41,.59]	.903	[.45,.55]	.950	[.44,.56]	.990	[.42,.58]
500	.384	[.44,.56]	.639	[.44,.57]	.911	[.42,.58]	.638	[.46,.54]	.803	[.45,.55]	.952	[.43,.57]	.905	[.46,.54]	.953	[.46,.54]	.989	[.44,.56]
1000	.392	[.46,.54]	.647	[.45,.55]	.908	[.44,.56]	.651	[.47,.53]	.803	[.46,.54]	.950	[.45,.55]	.909	[.47,.53]	.954	[.47,.53]	.989	[.46,.54]
	Comparison with GMS CSs for μ via moment inequalities																	
100	.806	[.34,.66]	.908	[.32,.68]	.978	[.29,.71]	.803	[.39,.61]	.900	[.37,.63]	.976	[.34,.66]	.899	[.42,.58]	.946	[.40,.60]	.975	[.39,.62]
250	.787	[.40,.60]	.898	[.39,.61]	.979	[.37,.64]	.798	[.43,.57]	.898	[.42,.58]	.978	[.40,.60]	.896	[.45,.55]	.946	[.44,.56]	.981	[.43,.57]
500	.788	[.43,.57]	.896	[.42,.58]	.978	[.40,.60]	.786	[.45,.55]	.898	[.44,.56]	.976	[.43,.57]	.898	[.46,.54]	.950	[.46,.54]	.985	[.45,.56]
1000	.789	[.45,.55]	.892	[.44,.56]	.979	[.43,.57]	.800	[.46,.54]	.893	[.46,.54]	.977	[.45,.55]	.906	[.47,.53]	.950	[.47,.53]	.986	[.46,.54]

Table 7: Missing data example: average coverage probabilities for Θ_I and M_I and average lower and upper bounds of CSs for M_I across 5000 MC replications. Procedures 1–3, Projection and Percentile are implemented using a **CU-GMM criterion** and flat prior.

where $p_\theta(a_1 = i, a_2 = j)$ denotes the probability that player 1 takes action i and player 2 takes action j when the true structural parameter is θ. Clearly $\vartheta \in \Delta(\theta)$ if and only if $D_{KL}(p_\theta \| p_\vartheta) = 0$. We compute the endpoints of the interval $M(\theta^b)$ by solving

$$\min / \max \mu \quad \text{such that} \quad \inf_{\eta \in H_\mu} D_{KL}(p_{\theta^b} \| p_{(\mu,\eta)}) = 0 \tag{29}$$

where $H_\mu = [-2, 0] \times [-1, 2]^2 \times [0, 1]^2$ for $\mu = \Delta_1$ and $H_\mu = [-2, 0]^2 \times [-1, 2] \times [0, 1]^2$ for $\mu = \beta_1$. The profiled distance $\inf_{\eta \in H_\mu} D_{KL}(p_\theta \| p_{(\mu,\eta)})$ is independent of the data and is very fast to compute. Note that we do *not* make explicit use of the separable reparameterization in terms of reduced-form choice probabilities when computing $M(\theta^b)$. Moreover, computation of $M(\theta^b)$ can be run in parallel for $b = 1, \ldots, B$ once the draws $\theta^1, \ldots, \theta^B$ have been generated.

To accommodate a small amount of optimization error, in practice we replace the equality in (29) by a small tolerance $D_{KL}(p_{\theta^b} \| p_{(\mu,\eta)}) < 10^{-7}$. The effect of this slight relaxation is to make our CSs computed via procedure 2 slightly more conservative than if the interval $M(\theta^b)$ were known in closed form.

A.4 Airline entry game application

SMC algorithm: We implement the adaptive SMC algorithm with $J = 200$ iterations, $K = 4$ blocked random-walk Metropolis-Hastings steps per iteration with $L = 4$ blocks for the full model and 2 blocks for the fixed-s model

Procedure 2: To implement procedure 2 here with any scalar subvector μ we calculate $M(\theta^b)$ numerically (in parallel), analogously to the entry game simulation example. We again compute the endpoints of $M(\theta^b)$ by solving (29) for the subvector of interest.

As the log-likelihood is conditional upon regressors, we replace $D_{KL}(p_{\theta^b} \| p_{(\mu,\eta)})$ by the sum of the KL distances between the conditional distributions of outcomes given regressors, namely:

$$\sum_{\{MS, MP_{OA}, MP_{LC}\} \in \{0,1\}^3} D_{KL}(p_{\theta^b}(\cdot \,|MS, MP_{OA}, MP_{LC}) \| p_{(\mu,\eta)}(\cdot \,|MS, MP_{OA}, MP_{LC}))$$

where $p_\theta(\cdot \,|MS, MP_{OA}, MP_{LC})$ denotes the probabilities of market outcomes conditional upon regressors when the structural parameter is θ.

A.5 Trade flow application

Priors: We use the change of variables $2\text{arctanh}(\rho)$ and $\log \sigma_m^2$ and assume that the transformed

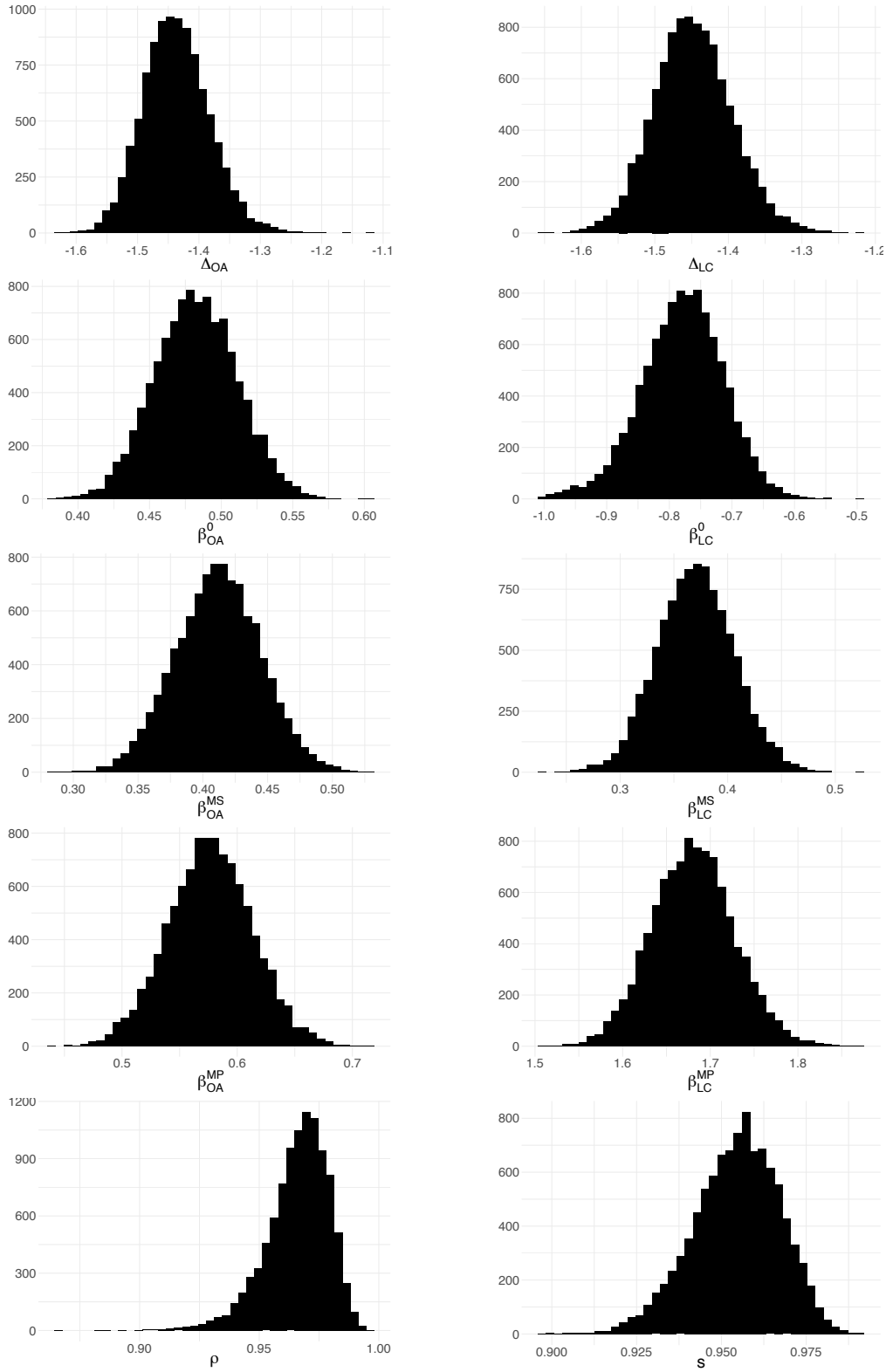

Figure 7: Entry game application: histograms of the SMC draws for Δ_{OA}, Δ_{LC}, β^0_{OA}, β^0_{LC}, β^{MS}_{OA}, β^{MS}_{LC}, β^{MP}_{OA}, β^{MP}_{LC}, ρ and s for the **fixed-s model**.

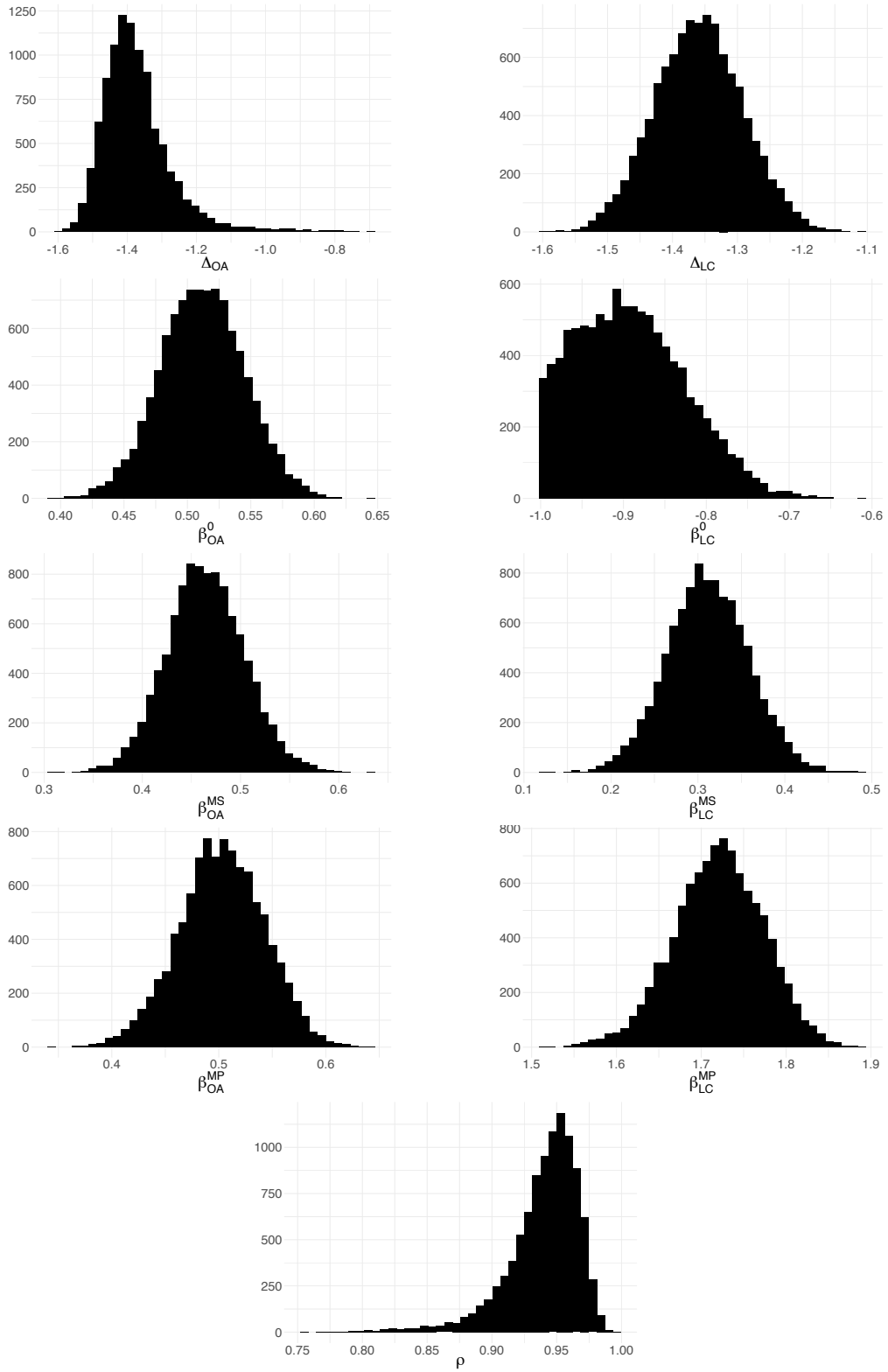

Figure 8: Entry game application: histograms of the SMC draws for Δ_{OA}, Δ_{LC}, β^0_{OA}, β^0_{LC}, β^{MS}_{OA}, β^{MS}_{LC}, β^{MP}_{OA}, β^{MP}_{LC}, ρ for the **full model**.

correlation and variance all have full support. We specify and independent $N(0, 100^2)$ priors on each of these 46 parameters.

SMC Algorithm: Given the high dimensionality of the parameter vector and the lack of a natural restriction of the parameter space for many of the parameters, we use a slight modification of the SMC algorithm described as follows.

We initialize the procedure from drawing from the $N(\hat{\theta}, -\hat{I}(\hat{\theta})^{-1})$ distribution, where $\hat{\theta}$ is the MLE and $-I(\hat{\theta})^{-1}$ is the inverse negative hessian at the MLE.

There are two more minor modifications which need to be made to correct the particle weights from initializing the algorithm in this manner. First, in the correction step, we replace v_j^b by $v_j^b = e^{(nL_n(\theta_{j-1}^b)}\Pi(\theta_{j-1}^b)/\mathcal{N}_n(\theta^b))^{\phi_j - \phi_{j-1}}$ where $\mathcal{N}_n(\theta^b)$ denotes the $N(\hat{\theta}, -I(\hat{\theta})^{-1})$ density evaluated at θ^b. Second, we use the tempered quasi-posterior $\Pi_j(\theta|\mathbf{X}_n) \propto (e^{nL_n(\theta)}\Pi(\theta))^{\phi_j}\mathcal{N}_n(\theta)^{1-\phi_j}$ in the updating step.

With these modifications, the algorithm is implemented with $K = 8$ block random-walk Metropolis-Hastings steps per iteration and $L = 6$ blocks.

Procedure 2: To implement procedure 2 here with any scalar subvector μ we calculate $M(\theta^b)$ numerically. We find the smallest and largest values of μ for which the average (across regressors) KL divergence, namely

$$\frac{1}{n}\sum_{ij}^{n} D_{KL}(p_{\theta^b}(\,\cdot\,|X_{ij})\|p_{(\mu,\eta)}(\,\cdot\,|X_{ij}))$$

is approximately zero. We then set $M(\theta^b) = [\underline{\mu}(\theta^b), \overline{\mu}(\theta^b)]$ where $\underline{\mu}(\theta^b)$ and $\overline{\mu}(\theta^b)$ denote the smallest and largest such values of μ for which the average KL divergence is minimized. If $M(\theta^b)$ is not an interval then the interval $[\underline{\mu}(\theta^b), \overline{\mu}(\theta^b)]$ will be a superset of $M(\theta^b)$ and the resulting CSs will be slightly conservative.

To compute $D_{KL}(p_{\theta^b}(\,\cdot\,|X_{ij})\|p_{(\mu,\eta)}(\,\cdot\,|X_{ij}))$, let d_{ij} be a dummy variable denoting exports from j to i and let $m_{ij} = \log M_{ij}$. We may write the model more compactly as:

$$d_{ij}m_{ij} = \begin{cases} X_{ij}'(\beta_m + \delta\beta_z) + (\delta\eta_{ij}^* + u_{ij}) & \text{if } d_{ij} = 1 \\ 0 & \text{if } d_{ij} = 0 \end{cases}$$

$$d_{ij} = \mathbb{1}\{X_{ij}'\beta_z + \eta_{ij}^* > 0\}$$

where X_{ij} collects the trade friction variables f_{ij} and dummy variables for importer and exporter's continent and β_z and β_m collect all coefficients in the selection and outcome equations, respectively. Therefore,

$$\Pr(d_{ij} = 1|X_{ij}) = \Phi\Big(\frac{X_{ij}'\beta_z}{\sigma_z(X_{ij})}\Big).$$

The likelihood is

$$p_\theta(d_{ij}, d_{ij}m_{ij}|X_{ij}) = \left(1 - \Phi\left(\frac{X_{ij}'\beta_z}{\sigma_z(X_{ij})}\right)\right)^{1-d_{ij}} \left(\Phi\left(\frac{\frac{X_{ij}'\beta_z}{\sigma_z(X_{ij})} + r(X_{ij})\frac{d_{ij}m_{ij} - X_{ij}'(\beta_m + \delta\beta_z)}{\sigma_v(X_{ij})}}{\sqrt{1 - r^2(X_{ij})}}\right)\right.$$
$$\left. \times \frac{1}{\sigma_v(X_{ij})}\phi\left(\frac{d_{ij}m_{ij} - X_{ij}'(\beta_m + \delta\beta_z)}{\sigma_v(X_{ij})}\right)\right)^{d_{ij}}$$

where

$$\sigma_v^2(X_{ij}) = \sigma_m^2 + 2\delta\rho\sigma_m\sigma_z(X_{ij}) + \delta^2\sigma_z^2(X_{ij}) \qquad r(X_{ij}) = \frac{\rho\sigma_m\sigma_z(X_{ij}) + \delta\sigma_z^2(X_{ij})}{\sigma_v(X_{ij})\sigma_z(X_{ij})}.$$

The conditional KL divergence between p_{θ^b} and $p_{(\mu,\eta)}$ is then straightforward to compute numerically (e.g. via Gaussian quadrature). Note also that the sets $M(\theta^b)$ for $b = 1, \ldots, B$ and for each subvector of interest can be computed in parallel once the draws $\theta^1, \ldots, \theta^B$ have been generated.

B Uniformity

Here we present conditions under which our CSs $\widehat{\Theta}_\alpha$ (Procedure 1) and $\widehat{M}_\alpha$ (Procedure 2) are uniformly valid over a class of DGPs $\mathbf{P}$. For each $\mathbb{P} \in \mathbf{P}$, let $L(\theta; \mathbb{P})$ denote the population objective function under $\mathbb{P}$. We assume that for each $\mathbb{P} \in \mathbf{P}$, $L(\cdot; \mathbb{P})$ and L_n are upper semicontinuous and $\sup_{\theta\in\Theta} L(\theta; \mathbb{P}) < \infty$. 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 subvector μ is $M_I(\mathbb{P}) = \{\mu : (\mu, \eta) \in \Theta_I(\mathbb{P}) \text{ for some } \eta\}$.

We now show that, under a natural extension of the assumptions in Section 4, the CSs $\widehat{\Theta}_\alpha$ and $\widehat{M}_\alpha$ are uniformly valid i.e.:

$$\liminf_{n\to\infty} \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\Theta_I(\mathbb{P}) \subseteq \widehat{\Theta}_\alpha) \geq \alpha \tag{30}$$

$$\liminf_{n\to\infty} \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(M_I(\mathbb{P}) \subseteq \widehat{M}_\alpha) \geq \alpha \tag{31}$$

both hold. The following Lemmas are straightforward extensions of Lemmas 2.1 and 2.2, but are helpful to organize ideas. Let $(\upsilon_n)_{n\in\mathbb{N}}$ be a sequence of random variables. We say that $\upsilon_n = o_{\mathbb{P}}(1)$ uniformly in $\mathbb{P}$ if $\lim_{n\to\infty} \sup_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(|\upsilon_n| > \epsilon) = 0$ for each $\epsilon > 0$, and that $\upsilon_n \leq o_{\mathbb{P}}(1)$ uniformly in $\mathbb{P}$ if $\lim_{n\to\infty} \sup_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\upsilon_n > \epsilon) = 0$ for each $\epsilon > 0$. Uniform $O_{\mathbb{P}}(1)$ statements are defined analogously.

Lemma B.1. *Let there exist sequences of random variables $(W_n, v_{\alpha,n})_{n\in\mathbb{N}}$ such that:*
(i) $\sup_{\theta\in\Theta_I(\mathbb{P})} Q_n(\theta) - W_n \leq o_{\mathbb{P}}(1)$ uniformly in $\mathbb{P}$; and
(ii) $\liminf_{n\to\infty} \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(W_n \leq v_{\alpha,n} - \varepsilon_n) \geq \alpha$ for any positive sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ with $\varepsilon_n = o(1)$.
Then: (30) holds for $\widehat{\Theta}_\alpha = \{\theta \in \Theta : Q_n(\theta) \leq v_{\alpha,n}\}$.

Lemma B.2. *Let there exist sequences of random variables $(W_n, v_{\alpha,n})_{n\in\mathbb{N}}$ such that:*
(i) $PQ_n(M_I(\mathbb{P})) - W_n \leq o_{\mathbb{P}}(1)$ uniformly in $\mathbb{P}$; and
(ii) $\liminf_{n\to\infty} \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(W_n \leq v_{\alpha,n} - \varepsilon_n) \geq \alpha$ for any positive sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ with $\varepsilon_n = o(1)$.
Then: (31) holds for $\widehat{M}_\alpha = \{\mu \in M : \inf_{\eta\in H_\mu} Q_n(\mu,\eta) \leq v_{\alpha,n}\}$.

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 $\Theta_I(\mathbb{P}) \subseteq \Theta_{osn}(\mathbb{P})$ for each n and for each $\mathbb{P} \in \mathbf{P}$. In what follows we omit the dependence of $\Theta_{osn}(\mathbb{P})$ on $\mathbb{P}$ to simplify notation.

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

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

To simply notation, in what follows we omit dependence of d^*, γ, $\gamma_\perp$, Γ, $\Gamma_\perp$, k_n, ℓ_n, T, $\mathbf{T}$, T_{osn}, τ, Θ_I^N, $\mathbb{V}_n$, Σ, and $f_{n,\perp}$ on $\mathbb{P}$.

Assumption B.2. *(Local quadratic approximation)*
(i) For each $\mathbb{P} \in \mathbf{P}$, there exist vectors $\tau \in T$, sequences of random variables ℓ_n and $\mathbb{R}^{d^*}$-valued random vectors $\hat{\gamma}_n$, and a sequence of non-negative measurable functions $f_{n,\perp} : \Gamma_\perp \to \mathbb{R}$ with $f_{n,\perp}(0) = 0$ (we take $\gamma_\perp \equiv 0$ and $f_{n,\perp} \equiv 0$ for nonsingular $\mathbb{P}$), such that as $n \to \infty$:

$$\sup_{\theta\in\Theta_{osn}} \left| nL_n(\theta) - \left(\ell_n + \frac{1}{2}\|\sqrt{n}(\hat{\gamma}_n - \tau)\|^2 - \frac{1}{2}\|\sqrt{n}(\hat{\gamma}_n - \tau - \gamma(\theta))\|^2 - f_{n,\perp}(\gamma_\perp(\theta)) \right) \right| = o_{\mathbb{P}}(1) \quad (32)$$

uniformly in $\mathbb{P}$, with $\sup_{\mathbb{P}\in\mathbf{P}} \sup_{\theta\in\Theta_{osn}} \|(\gamma(\theta), \gamma_\perp(\theta))\| \to 0$, $\sqrt{n}\hat{\gamma}_n = \mathbf{T}(\mathbb{V}_n + \sqrt{n}\tau)$ and $\|\mathbb{V}_n\| = O_{\mathbb{P}}(1)$ (uniformly in $\mathbb{P}$);
(ii) $\{\sqrt{n}\gamma(\theta) : \theta \in \Theta_{osn}\} \cap B_{k_n} = (T - \sqrt{n}\tau) \cap B_{k_n}$ where $\inf_{\mathbb{P}\in\mathbf{P}} k_n \to \infty$ and $\inf_{\mathbb{P}\in\mathbf{P}} \nu_{d^*}(T) > 0$;
(iii) for each singular $\mathbb{P} \in \mathbf{P}$: $\{(\gamma(\theta), \gamma_\perp(\theta)) : \theta \in \Theta_{osn}\} = \{\gamma(\theta) : \theta \in \Theta_{osn}\} \times \{\gamma_\perp(\theta) : \theta \in \Theta_{osn}\}$.

Let Π_{Γ^*} denote the image measure of Π under the map $\theta \mapsto \gamma(\theta)$ if $\mathbb{P}$ is nonsingular and $\theta \mapsto (\gamma(\theta), \gamma_\perp(\theta))$ if $\mathbb{P}$ is singular. We omit dependence of δ, Π_{Γ^*} and π_{Γ^*} on $\mathbb{P}$ in what follows.

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

The next lemma is a uniform-in-$\mathbb{P}$ extension of Lemmas 4.1 and 4.2. Recall that $\mathbb{P}_{Z|\mathbf{X}_n}$ is the distribution of a $N(0, I_{d^*})$ random vector Z (conditional on data).

Lemma B.3. *Let Assumptions B.1, B.2 and B.3 hold. Then:*

$$\sup_z \Big(\Pi_n\big(\{\theta : Q_n(\theta) \le z\} \,\big|\, \mathbf{X}_n\big) - \mathbb{P}_{Z|\mathbf{X}_n}\left(\|Z\|^2 \le z | Z \in T - \sqrt{n}\hat{\gamma}_n\right) \Big) \le o_{\mathbb{P}}(1)$$

uniformly in $\mathbb{P}$*. If no* $\mathbb{P} \in \mathbf{P}$ *is singular, then:*

$$\sup_z \Big| \Pi_n\big(\{\theta : Q_n(\theta) \le z\} \,\big|\, \mathbf{X}_n\big) - \mathbb{P}_{Z|\mathbf{X}_n}\left(\|Z\|^2 \le z | Z \in T - \sqrt{n}\hat{\gamma}_n\right) \Big| = o_{\mathbb{P}}(1)\,.$$

uniformly in $\mathbb{P}$*.*

As in Section 4, we let $\xi^{post}_{n,\alpha}$ denote the α quantile of $Q_n(\theta)$ under the posterior distribution Π_n.

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

The following result is a uniform-in-$\mathbb{P}$ extension of Theorems 4.1 and 4.2. Recall that $F_T(z) = \mathbb{P}_Z(\|\mathbf{T}Z\|^2 \le z)$ where $\mathbb{P}_Z$ denotes the distribution of a $N(0, I_{d^*})$ random vector. We say that the distributions $\{F_T : \mathbb{P} \in \mathbf{P}\}$ are equicontinuous at their α quantiles (denoted $\xi_{\alpha,\mathbb{P}}$) if for each $\epsilon > 0$ there is $\delta > 0$ such that $F_T(\xi_{\alpha,\mathbb{P}} - \epsilon) < \alpha - \delta$ for each $\mathbb{P} \in \mathbf{P}$ and $\inf_{\mathbb{P}\in\mathbf{P}} F_T(\xi_{\alpha,\mathbb{P}} - \epsilon) \to \alpha$ as $\epsilon \to 0$. This is trivially true if $T = \mathbb{R}^{d^*}$ for each $\mathbb{P} \in \mathbf{P}$ and $\sup_{\mathbb{P}\in\mathbf{P}} d^* < \infty$.

Theorem B.1. *Let Assumptions B.1, B.2, B.3 and B.4 hold, and let*

$$\sup_{\mathbb{P}\in\mathbf{P}} \sup_z |\mathbb{P}(\|\mathbf{T}\mathbb{V}_n\|^2 \le z) - \mathbb{P}_Z(\|\mathbf{T}Z\|^2 \le z)| = o(1)\,.$$

(i) If $\|\mathbf{T}(\mathbb{V}_n + \sqrt{n}\tau) - \sqrt{n}\tau\|^2 \le \|\mathbf{T}\mathbb{V}_n\|^2$ *(almost surely) for each* $\mathbb{P} \in \mathbf{P}$*, then: (30) holds.*
(ii) If no $\mathbb{P} \in \mathbf{P}$ *is singular and* $T = \mathbb{R}^{d^*}$ *for each* $\mathbb{P}$*, then: (30) holds with equality.*

To establish (31) we require a uniform version of Assumptions 4.5 and 4.6. In what follows, we omit dependence of f on $\mathbb{P}$ to simplify notation.

Assumption B.5. *(Profile QL)*

(i) For each $\mathbb{P} \in \mathbf{P}$, there exists a measurable function $f : \mathbb{R}^{d^*} \to \mathbb{R}$ such that:

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

uniformly in $\mathbb{P}$, with $\hat{\gamma}_n$, ℓ_n, τ and $\gamma(\cdot)$ from Assumption B.2;
(ii) $f(\mathbf{T}(\mathbb{V}_n + \sqrt{n}\tau) - \sqrt{n}\tau) \leq f(\mathbb{V}_n)$ (almost surely) for each $\mathbb{P} \in \mathbf{P}$;
(iii) $\sup_z(\mathbb{P}_Z(f(Z) \leq z | Z \in v - T) - \mathbb{P}_Z(f(Z) \leq z)) \leq 0$ for all $v \in T$.

Note that parts (ii) and (iii) of Assumption B.5 automatically hold with equality if $T = \mathbb{R}^{d^*}$. These conditions are not needed in the following result that is a uniform-in-$\mathbb{P}$ extension of Lemma 4.3.

Lemma B.4. *Let Assumptions B.1, B.2, B.3 and B.5(i) hold. Then for any interval $I = I(\mathbb{P}) \subseteq \mathbb{R}$ such that $\mathbb{P}_Z(f(Z) \leq z)$ is uniformly continuous on I (in both z and $\mathbb{P}$):*

$$\sup_{z \in I} \left| \Pi_n\left(\{\theta : PQ_n(M(\theta)) \leq z\} \,\middle|\, \mathbf{X}_n\right) - \mathbb{P}_{Z|\mathbf{X}_n}(f(Z) \leq z | Z \in \sqrt{n}\hat{\gamma}_n - T) \right| = o_{\mathbb{P}}(1)\,.$$

uniformly in $\mathbb{P}$.

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

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

The following result is a uniform-in-$\mathbb{P}$ extension of Theorem 4.3.

Theorem B.2. *Let Assumptions B.1, B.2, B.3, B.5 and B.6 hold, and let*

$$\sup_{\mathbb{P} \in \mathbf{P}} \sup_z |\mathbb{P}(f(\mathbb{V}_n) \leq z) - \mathbb{P}_Z(f(Z) \leq z)| = o(1)$$

where the distributions $\{\mathbb{P}_Z(f(Z) \leq z) : \mathbb{P} \in \mathbf{P}\}$ are equicontinuous at their α quantiles.
(i) Then: (31) holds.
(ii) If Assumption B.5(ii)(iii) holds with equality for all $\mathbb{P} \in \mathbf{P}$, then: (31) holds with equality.

B.1 A uniform quadratic expansion for discrete distributions

In this subsection we present low-level conditions that show the uniform quadratic expansion assumption is satisfied over a large class of DGPs in discrete models. Let $\mathbf{P}$ (possibly depending on n) be a class of distributions such that for each $\mathbb{P}_\theta \in \mathbf{P}$, $X_1, \ldots, X_n$ are i.i.d. discretely distributed on sample space $\{1, \ldots, k\}$ where $k \geq 2$. Let the k-vector p_θ denote the probabilities

$p_\theta(j) = \mathbb{P}_\theta(X_i = j)$ for $j = 1, \ldots, k$ and write $p_\theta > 0$ if $p_\theta(j) > 0$ for all $1 \le j \le k$. We identify a vector $\mathbb{P}_\theta$ with its probability vector p_θ and a generic distribution $\mathbb{P} \in \mathbf{P}$ with the k-vector p.

Our uniform quadratic approximation result encompasses a large variety of drifting sequence asymptotics, allowing $p(j)$ to drift towards 0 at rate up to (but not including) n^{-1}. That is, the first set of results concern any class of distributions $\mathbf{P}$ for which

$$\sup_{\mathbb{P}\in\mathbf{P}} \max_{1\le j\le k} \frac{1}{p(j)} = o(n)\,. \tag{33}$$

For any $\mathbb{P} \in \mathbf{P}$ with $p > 0$ and any θ, define the (squared) chi-square distance of $\mathbb{P}_\theta$ from $\mathbb{P}$ as

$$\chi^2(p_\theta; p) = \sum_{j=1}^{k} \frac{(p_\theta(j) - p(j))^2}{p(j)}\,.$$

For each $\mathbb{P}$, let $\Theta_{osn}(\mathbb{P}) = \{\theta : p_\theta > 0, \chi^2(p_\theta; p) \le r_n^2 n^{-1}\}$ where $(r_n)_{n\in\mathbb{N}}$ is a positive sequence to be defined below. Also let e_x denote a k-vector with 1 in its xth entry and 0 elsewhere, let $\mathbb{J}_p = \mathrm{diag}(p(1)^{-1/2}, \ldots, p(k)^{-1/2})$, and let $\sqrt{p} = (\sqrt{p(1)}, \ldots, \sqrt{p(k)})'$.

Lemma B.5. *Let (33) hold. Then: there exists a positive sequence $(r_n)_{n\in\mathbb{N}}$ with $r_n \to \infty$ as $n \to \infty$ such that:*

$$\sup_{\theta\in\Theta_{osn}(\mathbb{P})} \left| nL_n(p_\theta) - \left(\ell_n - \frac{1}{2}\|\sqrt{n}\tilde{\gamma}_{\theta;p}\|^2 + (\sqrt{n}\tilde{\gamma}_{\theta;p})'\tilde{\mathbb{V}}_{n;p}\right)\right| = o_{\mathbb{P}}(1)$$

uniformly in $\mathbb{P}$, where for each $\mathbb{P} \in \mathbf{P}$:

$$\ell_n = \ell_n(\mathbb{P}) = nL_n(p) \qquad \tilde{\gamma}_{\theta;p} = \begin{bmatrix} \frac{p_\theta(1)-p(1)}{\sqrt{p(1)}} \\ \vdots \\ \frac{p_\theta(k)-p(k)}{\sqrt{p(k)}} \end{bmatrix} \qquad \tilde{\mathbb{V}}_{n;p} = \mathbb{G}_n(\mathbb{J}_p e_x) \overset{\mathbb{P}}{\rightsquigarrow} N(0, I - \sqrt{p}\sqrt{p}')\,.$$

We are not quite done, as the covariance matrix is a rank $k-1$ orthogonal projection matrix. Let $v_{1,p}, \ldots, v_{k-1,p}$ denote an orthonormal basis for $\{v \in \mathbb{R}^k : v'\sqrt{p} = 0\}$ and define the matrix V_p by $V_p' = [v_{1,p} \ \cdots \ v_{k-1,p} \ \sqrt{p}]$. Notice that V_p is orthogonal (i.e. $V_p V_p' = V_p' V_p = I$) and

$$V_p\tilde{\gamma}_{\theta;p} = \begin{bmatrix} v_{1,p}'\tilde{\gamma}_{\theta;p} \\ \vdots \\ v_{k-1,p}'\tilde{\gamma}_{\theta;p} \\ 0 \end{bmatrix} \qquad V_p\mathbb{G}_n(\mathbb{J}_p e_x) = \begin{bmatrix} v_{1,p}'\mathbb{G}_n(\mathbb{J}_p e_x) \\ \vdots \\ v_{k-1,p}'\mathbb{G}_n(\mathbb{J}_p e_x) \\ 0 \end{bmatrix}. \tag{34}$$

Let $\gamma(\theta) = \gamma(\theta; \mathbb{P})$ and $\mathbb{V}_n = \mathbb{V}_n(\mathbb{P})$ denote the upper $k-1$ entries of $V_p \tilde{\gamma}_{\theta;p}$ and $V_p \mathbb{G}_n(\mathbb{J}_p e_x)$:

$$\gamma(\theta) = \begin{bmatrix} v'_{1,p} \tilde{\gamma}_{\theta;p} \\ \vdots \\ v'_{k-1,p} \tilde{\gamma}_{\theta;p} \end{bmatrix} \qquad \mathbb{V}_n = \begin{bmatrix} v'_{1,p} \mathbb{G}_n(\mathbb{J}_p e_x) \\ \vdots \\ v'_{k-1,p} \mathbb{G}_n(\mathbb{J}_p e_x) \end{bmatrix}. \tag{35}$$

We say that $\mathbb{V}_n \overset{\mathbb{P}}{\rightsquigarrow} N(0, I_{k-1})$ uniformly in $\mathbb{P}$ if $\sup_{\mathbb{P} \in \mathbf{P}} d_\pi(\mathbb{V}_n, N(0, I_{k-1})) \to 0$ where d_π denotes the distance (in the Prokhorov metric) between the sampling distribution of $\mathbb{V}_n$ and the $N(0, I_{k-1})$ distribution.

Proposition B.1. *Let (33) hold and $\Theta_{osn}(\mathbb{P})$ be as described in Lemma B.5. Then:*

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

uniformly in $\mathbb{P}$, where $\mathbb{V}_n \overset{\mathbb{P}}{\rightsquigarrow} N(0, I_{k-1})$ uniformly in $\mathbb{P}$.

We may generalize Proposition B.1 to allow for the support $k = k(n) \to \infty$ as $n \to \infty$ under a very mild condition on the growth rate of k. This result would be very useful in extending our procedures to semi/nonparametric models via discrete approximations of growing dimension. As before, let $\Theta_{osn}(\mathbb{P}) = \{\theta : p_\theta > 0, \chi^2(p_\theta; p) \leq r_n^2 n^{-1}\}$ where $(r_n)_{n \in \mathbb{N}}$ is a positive sequence to be defined below.

Proposition B.2. *Let $\sup_{\mathbb{P} \in \mathbf{P}} \max_{1 \leq j \leq k} (1/p(j)) = o(n/\log k)$. Then: there exists a positive sequence $(r_n)_{n \in \mathbb{N}}$ with $r_n \to \infty$ as $n \to \infty$ such that:*

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

uniformly in $\mathbb{P}$.

We now present two lemmas which are helpful in verifying the other conditions of Assumptions B.2 and B.5, respectively. Often, models may be parametrized such that $\{p_\theta : \theta \in \Theta, p_\theta > 0\} = \text{int}(\Delta^{k-1})$ where Δ^{k-1} denotes the unit simplex in $\mathbb{R}^k$. The following result shows that the sets $\{\sqrt{n}\gamma(\theta) : \theta \in \Theta_{osn}(\mathbb{P})\}$ each cover a ball of radius ρ_n (not depending on $\mathbb{P}$) with $\rho_n \to \infty$.

Lemma B.6. *Let (33) hold, $\{p_\theta : \theta \in \Theta, p_\theta > 0\} = \text{int}(\Delta^{k-1})$ and $\Theta_{osn}(\mathbb{P})$ be as described in Lemma B.5. Then: for each $\mathbb{P} \in \mathbf{P}$, $\{\sqrt{n}\gamma(\theta) : \theta \in \Theta_{osn}(\mathbb{P})\}$ covers a ball of radius $\rho_n \to \infty$ (with ρ_n not depending on $\mathbb{P}$) as $n \to \infty$.*

For the next result, let $\Theta'_{osn}(\mathbb{P}) = \{\theta : p_\theta > 0, \chi^2(p_\theta; p) \leq (r'_n)^2 n^{-1}\}$ where $(r'_n)_{n \in \mathbb{N}}$ is a positive sequence to be defined below.

Lemma B.7. *Let (33) hold. Then: there exists a positive sequence $(r'_n)_{n\in\mathbb{N}}$ with $r'_n \to \infty$ as $n \to \infty$ such that:*

$$\sup_{\theta\in\Theta'_{osn}(\mathbb{P})} \sup_{\mu\in M(\theta)} \left| \sup_{\eta\in H_\mu} nL_n(p_{\mu,\eta}) - \sup_{\eta\in H_\mu:(\mu,\eta)\in\Theta'_{osn}(\mathbb{P})} nL_n(p_{\mu,\eta}) \right| = o_{\mathbb{P}}(1)$$

uniformly in $\mathbb{P}$.

C Verification of main conditions for uniformity in examples

C.1 Example 1: uniform validity for missing data

Here we apply Proposition B.1 to establish uniform validity of our procedures. To make the missing data example fit the preceding notation, let $p_\theta = (\tilde{\gamma}_{11}(\theta), \tilde{\gamma}_{00}(\theta), 1 - \tilde{\gamma}_{00}(\theta) - \tilde{\gamma}_{11}(\theta))'$ and let $p = (\tilde{\gamma}_{11}, \tilde{\gamma}_{00}, 1 - \tilde{\gamma}_{00} - \tilde{\gamma}_{11})'$ denote the true probabilities under $\mathbb{P}$. The only requirement on $\mathbf{P}$ is that (33) holds. Therefore, Proposition B.1 holds uniformly over a set of DGPs under which the probability of missing data can drift to zero at rate up to n^{-1}. As $\{p_\theta : \theta \in \Theta, p_\theta > 0\} = \text{int}(\Delta^2)$, Lemma B.6 implies that $\{\sqrt{n}\gamma(\theta) : \theta \in \Theta_{osn}(\mathbb{P})\}$ covers a ball of radius ρ_n (independently of $\mathbb{P}$) with $\rho_n \to \infty$ as $n \to \infty$. This verifies parts (i)–(iv) of Assumption B.2.

By concavity, the infimum in the definition of the profile likelihood $PL_n(M(\theta))$ is attained at either the lower or upper bound of $M_I(\theta) = [\tilde{\gamma}_{11}(\theta), \tilde{\gamma}_{11}(\theta) + \tilde{\gamma}_{00}(\theta)]$. Moreover, at both $\mu = \tilde{\gamma}_{11}(\theta)$ and $\mu = \tilde{\gamma}_{11}(\theta) + \tilde{\gamma}_{00}(\theta)$, the profile likelihood is

$$\sup_{\substack{0\le g_{11}\le\mu \\ \mu\le g_{11}+g_{00}\le 1}} \Big(n\mathbb{P}_n 1\!\!1\{yd = 1\} \log g_{11} + n\mathbb{P}_n 1\!\!1\{1 - d = 1\} \log g_{00} + n\mathbb{P}_n 1\!\!1\{d - yd = 1\} \log(1 - g_{11} - g_{00}) \Big).$$

The constraint $g_{11} \le \mu$ will be the binding constraint at the lower bound and the constraint $\mu \le g_{11} + g_{00}$ will be the binding constraint at the upper bound (wpa1, uniformly in $\mathbb{P}$). These constraints are equivalent to $a'_1(\gamma - \gamma(\theta)) \le 0$ and $a'_2(\gamma - \gamma(\theta)) \le 0$ for some $a_1 = a_1(\mathbb{P}) \in \mathbb{R}^2$ and $a_2 = a_2(\mathbb{P}) \in \mathbb{R}^2$. It now follows from Proposition B.1 and Lemmas B.6 and B.7 that

$$\left| nPL_n(M_I) - \min_{j\in\{1,2\}} \sup_{\gamma: a'_j\gamma\le 0} \left(\ell_n - \frac{1}{2}\|\sqrt{n}\gamma\|^2 + (\sqrt{n}\gamma)'\mathbb{V}_n \right) \right| = o_{\mathbb{P}}(1)$$

and

$$\sup_{\theta\in\Theta'_{osn}(\mathbb{P})} \left| nPL_n(M(\theta)) - \min_{j\in\{1,2\}} \sup_{\gamma: a'_j(\gamma-\gamma(\theta))\le 0} \left(\ell_n - \frac{1}{2}\|\sqrt{n}\gamma\|^2 + (\sqrt{n}\gamma)'\mathbb{V}_n \right) \right| = o_{\mathbb{P}}(1)$$

uniformly in $\mathbb{P}$. Let T_j denote the closed convex cone in $\mathbb{R}^2$ defined by the inequality $a'_j\gamma \le 0$ for

$j = 1, 2$. We may write the above as

$$\left| nPL_n(M_I) - \left(\ell_n + \frac{1}{2}\|\mathbb{V}_n\|^2 - \max_{j\in\{1,2\}} \inf_{t\in T_j} \|\mathbb{V}_n - t\|^2 \right) \right| = o_{\mathbb{P}}(1)$$

$$\sup_{\theta\in\Theta'_{osn}(\mathbb{P})} \left| nPL_n(M(\theta)) - \left(\ell_n + \frac{1}{2}\|\mathbb{V}_n\|^2 - \max_{j\in\{1,2\}} \inf_{t\in T_j} \|(\mathbb{V}_n - \sqrt{n}\gamma(\theta)) - t\|^2 \right) \right| = o_{\mathbb{P}}(1)$$

uniformly in $\mathbb{P}$. This verifies the uniform expansion of the profile criterion.

C.2 Example 3: uniform validity of Procedure 2 vs the bootstrap

We return to Example 3 considered in Subsection 5.3.3 and show that our MC CSs (based on the posterior distribution of the profile QLR) are uniformly valid under very mild conditions while bootstrap-based CSs (based on the bootstrap distribution of the profile QLR) can undercover along certain sequences of DGPs. This reinforces the fact that our MC CSs and bootstrap-based CSs have different asymptotic properties.

Recall that $X_1, \ldots, X_n$ are i.i.d. with unknown mean $\mu^* \in \mathbb{R}_+$ and $\mu \in \mathbb{R}_+$ is identified by the moment inequality $\mathbb{E}[\mu - X_i] \leq 0$. The identified set for μ is $M_I = [0, \mu^*]$. We consider coverage of the CS for $M_I = [0, \mu^*]$ We introduce a slackness parameter $\eta \in \mathbb{R}_+$ to write this model as a moment equality model $\mathbb{E}[\mu + \eta - X_i] = 0$. The parameter space for $\theta = (\mu, \eta)$ is $\Theta = \mathbb{R}_+^2$. The GMM objective function and profile QLR are

$$\begin{aligned} L_n(\mu, \eta) &= -\frac{1}{2}(\mu + \eta - \bar{X}_n)^2 \\ PQ_n(M_I) &= (\mathbb{V}_n \wedge 0)^2 - ((\mathbb{V}_n + \sqrt{n}\mu^*) \wedge 0)^2 \\ PQ_n(M(\theta)) &= ((\mathbb{V}_n - \sqrt{n}\gamma(\theta)) \wedge 0)^2 - ((\mathbb{V}_n + \sqrt{n}\mu^*) \wedge 0)^2 \end{aligned} \tag{36}$$

where $\sqrt{n}\gamma(\theta) = \sqrt{n}(\mu + \eta - \mu^*) \in [-\sqrt{n}\mu^*, \infty)$.

C.2.1 Uniform validity of Procedures 2 and 3

Let $\mathbf{P}$ be the family of distributions under which the X_i are i.i.d. with mean $\mu^* = \mu^*(\mathbb{P}) \in \mathbb{R}_+$ and unit variance and for which

$$\lim_{n\to\infty} \sup_{\mathbb{P}\in\mathbf{P}} \sup_{z\in\mathbb{R}} |\mathbb{P}(\mathbb{V}_n \leq z) - \Phi(z)| = 0 \tag{37}$$

holds, where $\mathbb{V}_n = \mathbb{V}_n(\mathbb{P}) = \sqrt{n}(\bar{X}_n - \mu^*)$. We first consider uniform coverage of our MC CSs $\widehat{M}_\alpha$ for the identified set $M_I = M_I(\mathbb{P}) = [0, \mu^*(\mathbb{P})]$.

To focus solely on the essential ideas, assume the prior on θ induces a uniform prior on γ (the posterior is still proper); this could be relaxed at the cost of more cumbersome notation without changing the results that follow. Letting $z \geq 0$, $\kappa = \sqrt{n}\gamma$ and $v_n = v_n(\mathbb{P}) = \mathbb{V}_n + \sqrt{n}\mu^*$, we have:

$$\Pi_n(\{\theta : PQ_n(M(\theta)) \leq z\}|\mathbf{X}_n) = \frac{\int_{-\sqrt{n}\mu^*}^{\infty} 1\!\!1\{((\mathbb{V}_n - \kappa) \wedge 0)^2 - (v_n \wedge 0)^2 \leq z\} e^{-\frac{1}{2}(\mathbb{V}_n - \kappa)^2} \mathrm{d}\kappa}{\int_{-\sqrt{n}\mu^*}^{\infty} e^{-\frac{1}{2}(\mathbb{V}_n - \kappa)^2} \mathrm{d}\kappa}.$$

A change of variables with $x = \mathbb{V}_n - \kappa$ yields:

$$\begin{aligned}\Pi_n(\{\theta : PQ_n(M(\theta)) \leq z\}|\mathbf{X}_n) &= \frac{\int_{-\infty}^{v_n} 1\!\!1\{(x \wedge 0)^2 \leq z + (v_n \wedge 0)^2\} e^{-\frac{1}{2}x^2} \mathrm{d}x}{\int_{-\infty}^{v_n} e^{-\frac{1}{2}x^2} \mathrm{d}x} \\ &= \mathbb{P}_{Z|\mathbf{X}_n}(-\sqrt{z + (v_n \wedge 0)^2} \leq Z | Z \leq v_n) = G(v_n; z)\,.\end{aligned}$$

As we have an explicit form for the posterior distribution of the profile QLR, we can compute the posterior critical value directly rather than resorting to MC sampling. Therefore, Assumption B.6 is not required here (since we can trivially set $\xi_{n,\alpha}^{post,p} = \xi_{n,\alpha}^{mc,p}$). If MC sampling were to be used, we would require that Assumption B.6 holds.

For $v_n \geq 0$, we have

$$G(v_n; z) = \mathbb{P}_{Z|\mathbf{X}_n}(-\sqrt{z} \leq Z | Z \leq v_n)$$

and so the posterior α-critical value $\xi_{n,\alpha}^{post,p} = \Phi^{-1}((1-\alpha)\Phi(v_n))^2$. Therefore,

$$\begin{aligned}\mathbb{P}(PQ_n(M_I) \leq \xi_{n,\alpha}^{post,p} | v_n \geq 0) &= \mathbb{P}((\mathbb{V}_n \wedge 0)^2 \leq \Phi^{-1}((1-\alpha)\Phi(v_n))^2 | v_n \geq 0) \\ &= \mathbb{P}(\Phi^{-1}((1-\alpha)\Phi(v_n)) \leq \mathbb{V}_n | v_n \geq 0)\,. \qquad (38)\end{aligned}$$

Now suppose that $v_n < 0$. Here we have

$$G(v_n; z) = \mathbb{P}_{Z|\mathbf{X}_n}(-\sqrt{z + v_n^2} \leq Z | Z \leq v_n) = \frac{\Phi(v_n) - \Phi(-\sqrt{z + v_n^2})}{\Phi(v_n)}$$

from which it follows that $\xi_{n,\alpha}^{post,p} = \Phi^{-1}((1-\alpha)\Phi(v_n))^2 - v_n^2$ and hence:

$$\begin{aligned}\mathbb{P}(PQ_n(M_I) \leq \xi_{n,\alpha}^{post,p} | v_n < 0) &= \mathbb{P}((\mathbb{V}_n \wedge 0)^2 \leq \Phi^{-1}((1-\alpha)\Phi(v_n))^2 | v_n < 0) \\ &= \mathbb{P}(\Phi^{-1}((1-\alpha)\Phi(v_n)) \leq \mathbb{V}_n | v_n < 0)\,. \qquad (39)\end{aligned}$$

Combining (38) and (39), we obtain:

$$\mathbb{P}(PQ_n(M_I) \leq \xi_{n,\alpha}^{post,p}) = \mathbb{P}((1-\alpha)\Phi(v_n) \leq \Phi(\mathbb{V}_n)) \geq \mathbb{P}((1-\alpha) \leq \Phi(\mathbb{V}_n))$$

which, together with (37), delivers the uniform coverage result for Procedure 2:

$$\liminf_{n\to\infty} \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\mathbb{M}_I(\mathbb{P}) \subseteq \widehat{M}_\alpha) \geq \alpha\,.$$

For uniform validity of Procedure 3, first note that (36) implies that the inequality

$$\mathbb{P}(PQ_n(M_I) \leq \chi^2_{1,\alpha}) \geq \mathbb{P}((\mathbb{V}_n \wedge 0)^2 \leq \chi^2_{1,\alpha})$$

holds uniformly in $\mathbb{P}$. It follows by (37) that:

$$\liminf_{n\to\infty} \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\mathbb{M}_I(\mathbb{P}) \subseteq \widehat{M}^\chi_\alpha) > \alpha\,.$$

C.2.2 Lack of uniformity of the bootstrap

We now show that bootstrap-based CSs for M_I are not uniformly valid when the standard (i.e. nonparametric) bootstrap is used. The bootstrap criterion function $L^\star_n(\mu,\eta)$ is

$$L^\star_n(\mu,\eta) = -\frac{1}{2}(\mu+\eta-\bar{X}^\star_n)^2$$

where $\bar{X}^\star_n$ is the bootstrap sample mean. Let $\widehat{M}_I = [0,(\bar{X}_n \vee 0)]$. Consider a subsequence $(\mathrm{P}_n)_{n\in\mathbb{N}} \subset \mathbf{P}$ with $\mu^*(\mathrm{P}_n) = c/\sqrt{n}$ for some $c>0$ (chosen below). By similar calculations to Subsection 5.3.3, along this sequence of DGPs, the bootstrapped profile QLR statistic for M_I is:

$$\begin{aligned} PQ^\star_n(M_I) &= 2nL^\star_n(\hat{\mu}^\star,\hat{\eta}^\star) - \inf_{\mu\in\widehat{M}_I} \sup_{\eta\in H_\mu} 2nL^\star_n(\mu,\eta) \\ &= ((\mathbb{V}^\star_n + ((\mathbb{V}_n + c)\wedge 0))\wedge 0)^2 - ((\mathbb{V}^\star_n + \mathbb{V}_n + c)\wedge 0)^2\,. \end{aligned}$$

Let $\xi^{boot,p}_{n,\alpha}$ denote the α-quantile of the distribution of $PQ^\star_n(M_I)$. Consider

$$\widehat{M}^{boot}_\alpha = \{\mu : \sup_{\eta\in H_\mu} Q_n(\mu,\eta) \leq \xi^{boot,p}_{n,\alpha}\}$$

We now show that for any $\alpha \in (\frac{1}{2},1)$ we may choose $c>0$ in the definition of $(\mathrm{P}_n)_{n\in\mathbb{N}}$ such that the asymptotic coverage of $\widehat{M}^{boot}_\alpha$ is strictly less than α along this sequence of DGPs. Since

$$PQ^\star_n(M_I) = ((\mathbb{V}^\star_n \wedge 0)^2 - ((\mathbb{V}^\star_n + \mathbb{V}_n + c)\wedge 0)^2)\mathbb{1}\{\mathbb{V}_n + c \geq 0\}$$

it follows that whenever $\mathbb{V}_n + c < 0$ the bootstrap distribution of the profile QLR for M_I is point mass at the origin, and the α-quantile of the bootstrap distribution is $\xi^{boot,p}_{n,\alpha} = 0$. However, the QLR statistic for M_I is $PQ_n(M_I) = (\mathbb{V}_n \wedge 0)^2 - ((\mathbb{V}_n + c)\wedge 0)^2$. So whenever $\mathbb{V}_n + c < 0$ we also

have that $PQ_n(M_I) = \mathbb{V}_n^2 - (\mathbb{V}_n + c)^2 > 0$. Therefore,

$$\mathrm{P}_n(M_I(\mathrm{P}_n) \subseteq \widehat{M}_\alpha^{boot} | \mathbb{V}_n + c < 0) = 0\,.$$

It follows by (37) that for any c for which $\Phi(c) < \alpha$, we have:

$$\limsup_{n\to\infty} \mathrm{P}_n(M_I(\mathrm{P}_n) \subseteq \widehat{M}_\alpha^{boot}) \leq \lim_{n\to\infty} \mathrm{P}_n(\mathbb{V}_n + c \geq 0) < \alpha\,.$$

C.2.3 An alternative recentering

An alternative is to recenter the criterion function at $(\bar{X}_n \vee 0)$, that is, one could use instead

$$L_n(\mu, \eta) = -\frac{1}{2}(\mu + \eta - (\bar{X}_n \vee 0))^2$$

similar to the idea of a sandwich (quasi-)likelihood with $(\bar{X}_n \vee 0) = \hat{\gamma}_n$. This maps into the setup described in Appendix B, where

$$nL_n(\theta) = \ell_n - \frac{1}{2}(\sqrt{n}\gamma(\theta))^2 + \sqrt{n}(\gamma(\theta))(\sqrt{n}(\hat{\gamma}_n - \tau))$$

where $\ell_n = -\frac{1}{2}(\sqrt{n}(\hat{\gamma}_n - \tau))^2$, $\theta = (\mu, \eta)$ and

$$\gamma(\theta) = \mu + \eta - \mu^* \qquad \tau = \mu^* \qquad \hat{\gamma}_n = (\bar{X}_n \vee 0) \qquad \sqrt{n}(\hat{\gamma}_n - \tau) = (\mathbb{V}_n \wedge -\sqrt{n}\mu^*)$$

where $\mathbb{V}_n = \sqrt{n}(\bar{X}_n - \mu^*)$, $\gamma(\theta) \in [-\mu^*, \infty)$, and $\mu^* \in \mathbb{R}_+$.

Assumption B.1 and B.2(i)–(iii) hold with $\Theta_{osn} = \Theta$, $k_n = +\infty$, $T = \mathbb{R}_+$, and $\mathbf{T}v = (v \vee 0)$ (none of the models are singular). We again take a prior on θ that induces a flat prior on γ to concentrate on the essential ideas, verifying Assumption B.3.

For inference on $M_I = [0, \mu^*(\mathbb{P})]$, observe that

$$PQ_n(M(\theta)) = f(\sqrt{n}(\hat{\gamma}_n - \tau) - \sqrt{n}\gamma(\theta)) \qquad PQ_n(M_I) = f(\sqrt{n}(\hat{\gamma}_n - \tau))$$

where $f(v) = (v \wedge 0)^2$ for each $\mathbb{P}$, verifying Assumption B.5(i). Assumption B.5(ii) also holds for this f. Finally, for Assumption B.5(iii), for any $z, v \geq 0$ we have

$$\mathbb{P}_Z(f(Z) \leq z | Z \in v - T) = \frac{\Phi(v) - \Phi(-\sqrt{z})}{\Phi(v)} \leq 1 - \Phi(-\sqrt{z}) = \mathbb{P}_Z(f(Z) \leq z)\,.$$

Theorem B.2, together with (37), delivers uniform coverage for Procedure 2.

Similarly, for uniform validity of Procedure 3 we have:

$$\mathbb{P}(PQ_n(M_I) \leq \chi^2_{1,\alpha}) \geq \mathbb{P}((\mathbb{V}_n \wedge 0)^2 \leq \chi^2_{1,\alpha})$$

which, together with (37), delivers uniform coverage for Procedure 3.

Now consider bootstrap-based inference. As before, let $\widehat{M}_I = [0, (\bar{X}_n \vee 0)]$ and consider a subsequence $(\mathrm{P}_n)_{n\in\mathbb{N}} \subset \mathbf{P}$ with $\mu^*(\mathrm{P}_n) = c/\sqrt{n}$ for some $c > 0$. Under P_n, we then have:

$$\begin{aligned} L_n^\star(\mu, \eta) &= -\frac{1}{2}(\mu + \eta - (\bar{X}_n^\star \vee 0))^2 \\ PQ_n^\star(M_I) &= ([((\mathbb{V}_n^\star + \mathbb{V}_n) \vee -c) - (\mathbb{V}_n \vee -c)] \wedge 0)^2 \end{aligned}$$

and the true QLR statistic is $PQ_n(M_I) = ((\mathbb{V}_n \vee -c) \wedge 0)^2$. We again show that for any $\alpha \in (\frac{1}{2}, 1)$ we may choose $c > 0$ in the definition of $(\mathrm{P}_n)_{n\in\mathbb{N}}$ such that the asymptotic coverage of $\widehat{M}_\alpha^{boot}$ is strictly less than α along this sequence of DGPs. Observe that when $\mathbb{V}_n < -c$ we have $PQ_n(M_I) = c^2 > 0$ and $PQ_n^\star(M_I) = 0$. Therefore,

$$\mathrm{P}_n(M_I(\mathrm{P}_n) \subseteq \widehat{M}_\alpha^{boot} | \mathbb{V}_n + c < 0) = 0\,.$$

It follows by (37) that for any c for which $\Phi(c) < \alpha$, we again have:

$$\limsup_{n\to\infty} \mathrm{P}_n(M_I(\mathrm{P}_n) \subseteq \widehat{M}_\alpha^{boot}) \leq \lim_{n\to\infty} \mathrm{P}_n(\mathbb{V}_n + c \geq 0) < \alpha\,.$$

D Local power

In this appendix we study the behavior of the CSs $\widehat{\Theta}_\alpha$ and $\widehat{M}_\alpha$ under $n^{-1/2}$-local (contiguous) alternatives. We maintain the same setup as in Section 4. Fix $a \in \mathbb{R}^{d^*}$.

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

$$\sup_{\theta\in\Theta_{osn}} \left| nL_n(\theta) - \left(\ell_n + \frac{1}{2}\|\sqrt{n}\hat{\gamma}_n\|^2 - \frac{1}{2}\|\sqrt{n}(\hat{\gamma}_n - \gamma(\theta))\|^2 \right) \right| = o_{\mathrm{P}_{n,a}}(1) \tag{40}$$

with $\sup_{\theta\in\Theta_{osn}} \|\gamma(\theta)\| \to 0$, $\sqrt{n}\hat{\gamma}_n = \mathbb{V}_n$ where $\mathbb{V}_n \overset{\mathrm{P}_{n,a}}{\rightsquigarrow} N(a, I_{d^*})$ and $T = \mathbb{R}^{d^*}$;
(iv) $\int_\Theta e^{nL_n(\theta)}\, \mathrm{d}\Pi(\theta) < \infty$ holds $\mathrm{P}_{n,a}$-almost surely;
(v) Π_Γ has a continuous, strictly positive density π_Γ on $B_\delta \cap \Gamma$ for some $\delta > 0$;

(vi) $\xi_{n,\alpha}^{mc} = \xi_{n,\alpha}^{post} + o_{\mathrm{P}_{n,a}}(1)$.

Assumption D.1 is essentially a restatement of Assumptions 4.1 to 4.4 with a modified quadratic expansion. Notice that with $a = 0$ we obtain $\mathrm{P}_{n,a} = \mathbb{P}$ and Assumption D.1 corresponds to Assumptions 4.1 to 4.4 with generalized information equality $\Sigma = I_{d^*}$ and $T = \mathbb{R}^{d^*}$.

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

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

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

if further Assumption D.1(ii)(iv)(v) holds, then:

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

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

$$\lim_{n\to\infty} \mathrm{P}_{n,a}(\Theta_I \subseteq \widehat{\Theta}_\alpha) = F_{\chi^2_{d^*}(a'a)}(\chi^2_{d^*,\alpha}) < \alpha \quad \textit{whenever} \quad a \neq 0.$$

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

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

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

with $\mathbb{V}_n$ *from Assumption D.1(iii).*
(vi′) $\xi_{n,\alpha}^{mc,p} = \xi_{n,\alpha}^{post,p} + o_{\mathrm{P}_{n,a}}(1)$.

Assumption D.1(vii) and (vi′) are essentially Assumptions 4.5 and 4.6.

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

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

$$PQ_n(M_I) \overset{\mathrm{P}_{n,a}}{\rightsquigarrow} f(Z + a)\ ;$$

if further Assumption D.1(ii)(iv)(v) holds, then for a neighborhood I of z_α:

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

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

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

It follows from Anderson's lemma (van der Vaart, 2000, Lemma 8.5) that

$$\lim_{n\to\infty} \mathrm{P}_{n,a}(M_I \subseteq \widehat{M}_\alpha) \le \alpha$$

whenever f is subconvex. In particular, this includes the case in which M_I is a singleton.

E Parameter-dependent support

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

We again presume the existence of a local reduced-form parameter γ such that $\gamma(\theta) = 0$ if and only if $\theta \in \Theta_I$. In what follows we assume without loss of generality that $L_n(\hat\theta) = \sup_{\theta\in\Theta_{osn}} L_n(\theta)$ since $\hat\theta$ is not required in order to compute the confidence set. We replace Assumption 4.2 (local quadratic approximation) with the following assumption, which permits the support of the data to depend on certain components of the local reduced-form parameter γ.

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

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

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

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

Lemma E.1. *Let Assumptions 4.1, E.2, and 4.3 hold. Then:*

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

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

Theorem E.1. *Let Assumptions 4.1, E.2, 4.3, and 4.4 hold and* $\sup_{\theta\in\Theta_I} Q_n(\theta) \rightsquigarrow F_\Gamma$. *Then:*

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

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

Then we use the reduced-form parameter $\gamma(\theta) = (\theta_1 \vee \theta_2) - \tilde{\gamma}$. Let $\hat{\gamma}_n = \max_{1\leq i\leq n} X_i - \tilde{\gamma}$. 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 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 from Theorem E.1.

F Proofs and Additional Results

F.1 Proofs and Additional Lemmas for Sections 2 and 4

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

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

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

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

Lemma F.1. *Let Assumptions 4.1(i) and 4.2 hold. Then:*

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

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

Proof of Lemma F.1. By Assumptions 4.1(i) and 4.2, we obtain:

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

where $\inf_{t\in T} \|t - \mathbf{T}\mathbb{V}_n\|^2 = 0$ because $\mathbf{T}\mathbb{V}_n \in T$. Now by Assumption 4.2,

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

where the $o_{\mathbb{P}}(1)$ term holds uniformly over Θ_{osn}. This proves expression (41). Finally, since $\gamma(\theta) = 0$ for $\theta \in \Theta_I$, we have $\sup_{\theta\in\Theta_I} Q_n(\theta) = \|\mathbf{T}\mathbb{V}_n\|^2 + o_{\mathbb{P}}(1)$. ■

Proof of Lemma 4.1. We first prove equation (21). Since $|\Pr(A) - \Pr(A \cap B)| \leq \Pr(B^c)$, we

have:

$$\sup_z \left|\Pi_n(\{\theta : Q_n(\theta) \le z\}|\mathbf{X}_n) - \Pi_n(\{\theta : Q_n(\theta) \le z\} \cap \Theta_{osn}|\mathbf{X}_n)\right| \le \Pi_n(\Theta_{osn}^c|\mathbf{X}_n) = o_{\mathbb{P}}(1) \tag{43}$$

by Assumption 4.1(ii). Moreover by Assumptions 4.1(ii) and 4.3(i),

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

and hence:

$$\sup_z \left|\Pi_n(\{\theta : Q_n(\theta) \le z\} \cap \Theta_{osn} \,|\, \mathbf{X}_n) - \frac{\int_{\{\theta:Q_n(\theta)\le z\}\cap\Theta_{osn}} e^{nL_n(\theta)}\mathrm{d}\Pi(\theta)}{\int_{\Theta_{osn}} e^{nL_n(\theta)}\mathrm{d}\Pi(\theta)}\right| = o_{\mathbb{P}}(1)\,. \tag{44}$$

In view of (43) and (44), it suffices to characterize the large-sample behavior of:

$$R_n(z) := \frac{\int_{\{\theta:Q_n(\theta)\le z\}\cap\Theta_{osn}} e^{nL_n(\theta)-\ell_n-\frac{1}{2}\|\mathbf{T}\mathbb{V}_n\|^2}\mathrm{d}\Pi(\theta)}{\int_{\Theta_{osn}} e^{nL_n(\theta)-\ell_n-\frac{1}{2}\|\mathbf{T}\mathbb{V}_n\|^2}\mathrm{d}\Pi(\theta)}\,. \tag{45}$$

Lemma F.1 and Assumption 4.2 imply that there exists a positive sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ independent of z with $\varepsilon_n = o(1)$ such that the inequalities:

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

both hold wpa1. Therefore, wpa1 we have:

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

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

$$\begin{aligned} &e^{-2\varepsilon_n}\frac{\int_{\{\gamma:\|\sqrt{n}\gamma-\mathbf{T}\mathbb{V}_n\|^2\le z-\varepsilon_n\}\cap\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma-\mathbf{T}\mathbb{V}_n\|^2}\mathrm{d}\Pi_\Gamma(\gamma)}{\int_{\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma-\mathbf{T}\mathbb{V}_n\|^2}\mathrm{d}\Pi_\Gamma(\gamma)} \\ &\le R_n(z) \le e^{2\varepsilon_n}\frac{\int_{\{\gamma:\|\sqrt{n}\gamma-\mathbf{T}\mathbb{V}_n\|^2\le z+\varepsilon_n\}\cap\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma-\mathbf{T}\mathbb{V}_n\|^2}\mathrm{d}\Pi_\Gamma(\gamma)}{\int_{\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma-\mathbf{T}\mathbb{V}_n\|^2}\mathrm{d}\Pi_\Gamma(\gamma)}\,. \end{aligned} \tag{46}$$

Recall B_δ from Assumption 4.3(ii). The inclusion $\Gamma_{osn} \subset B_\delta \cap \Gamma$ holds for all n sufficiently large

by Assumption 4.2. Taking n sufficiently large and using Assumption 4.3(ii), we may deduce that there exists a positive sequence $(\bar{\varepsilon}_n)_{n\in\mathbb{N}}$ with $\bar{\varepsilon}_n = o(1)$ such that:

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

for each n. Substituting into (46):

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

uniformly in z, where "$\mathrm{d}\gamma$" denotes integration with respect to Lebesgue measure on $\mathbb{R}^{d^*}$.

Let $T_{osn} = \{\sqrt{n}\gamma : \gamma \in \Gamma_{osn}\}$ and let B_z denote a ball of radius z in $\mathbb{R}^{d^*}$ centered at the origin. Using the change of variables $\sqrt{n}\gamma - \mathbf{T}\mathbb{V}_n \mapsto \kappa$, we can rewrite the preceding inequalities as:

$$(1-\bar{\varepsilon}_n)e^{-2\varepsilon_n}\frac{\int_{B_{\sqrt{z-\varepsilon_n}}\cap(T_{osn}-\mathbf{T}\mathbb{V}_n)} e^{-\frac{1}{2}\|\kappa\|^2}\mathrm{d}\kappa}{\int_{(T_{osn}-\mathbf{T}\mathbb{V}_n)} e^{-\frac{1}{2}\|\kappa\|^2}\mathrm{d}\kappa} \leq R_n(z) \leq (1+\bar{\varepsilon}_n)e^{2\varepsilon_n}\frac{\int_{B_{\sqrt{z+\varepsilon_n}}\cap(T_{osn}-\mathbf{T}\mathbb{V}_n)} e^{-\frac{1}{2}\|\kappa\|^2}\mathrm{d}\kappa}{\int_{(T_{osn}-\mathbf{T}\mathbb{V}_n)} e^{-\frac{1}{2}\|\kappa\|^2}\mathrm{d}\kappa}$$

with the understanding that $B_{\sqrt{z-\varepsilon_n}}$ is empty if $\varepsilon_n > z$.

Let $\nu_{d^*}(A) = (2\pi)^{-d^*/2}\int_A e^{-\frac{1}{2}\|\kappa\|^2}\,\mathrm{d}\kappa$ denote Gaussian measure. We now show that:

$$\sup_z\left|\frac{\nu_{d^*}(B_{\sqrt{z\pm\varepsilon_n}}\cap(T_{osn}-\mathbf{T}\mathbb{V}_n))}{\nu_{d^*}(T_{osn}-\mathbf{T}\mathbb{V}_n)} - \frac{\nu_{d^*}(B_{\sqrt{z\pm\varepsilon_n}}\cap(T-\mathbf{T}\mathbb{V}_n))}{\nu_{d^*}(T-\mathbf{T}\mathbb{V}_n)}\right| = o_{\mathbb{P}}(1) \tag{47}$$

$$\sup_z\left|\frac{\nu_{d^*}(B_{\sqrt{z\pm\varepsilon_n}}\cap(T-\mathbf{T}\mathbb{V}_n))}{\nu_{d^*}(T-\mathbf{T}\mathbb{V}_n)} - \frac{\nu_{d^*}(B_{\sqrt{z}}\cap(T-\mathbf{T}\mathbb{V}_n))}{\nu_{d^*}(T-\mathbf{T}\mathbb{V}_n)}\right| = o_{\mathbb{P}}(1)\,. \tag{48}$$

Consider (47). To simplify presentation, we assume wlog that $T_{osn} \subseteq T$. As

$$\left|\frac{\Pr(A\cap B)}{\Pr(B)} - \frac{\Pr(A\cap C)}{\Pr(C)}\right| \leq 2\frac{\Pr(C\setminus B)}{\Pr(C)} \tag{49}$$

holds for events A, B, C with $B \subseteq C$, we have:

$$\sup_z\left|\frac{\nu_{d^*}(B_{\sqrt{z\pm\varepsilon_n}}\cap(T_{osn}-\mathbf{T}\mathbb{V}_n))}{\nu_{d^*}(T_{osn}-\mathbf{T}\mathbb{V}_n)} - \frac{\nu_{d^*}(B_{\sqrt{z\pm\varepsilon_n}}\cap(T-\mathbf{T}\mathbb{V}_n))}{\nu_{d^*}(T-\mathbf{T}\mathbb{V}_n)}\right| \leq 2\frac{\nu_{d^*}((T\setminus T_{osn})-\mathbf{T}\mathbb{V}_n)}{\nu_{d^*}(T-\mathbf{T}\mathbb{V}_n)}$$

As $\mathbb{V}_n$ is tight and $T \subseteq \mathbb{R}^{d^*}$ has positive volume, we may deduce that

$$1/\nu_{d^*}(T-\mathbf{T}\mathbb{V}_n) = O_{\mathbb{P}}(1)\,. \tag{50}$$

It also follows by tightness of $\mathbb{V}_n$ and Assumption 4.2 that $\nu_{d^*}((T \setminus T_{osn}) - \mathbf{T}\mathbb{V}_n) = o_{\mathbb{P}}(1)$, which proves (47). Result (48) now follows by (50) and the fact that:

$$\sup_z |\nu_{d^*}(B_{\sqrt{z \pm \varepsilon_n}} \cap (T_{osn} - \mathbf{T}\mathbb{V}_n)) - \nu_{d^*}(B_{\sqrt{z}} \cap (T_{osn} - \mathbf{T}\mathbb{V}_n))| \leq \sup_z |F_{\chi^2_{d^*}}(z \pm \varepsilon_n) - F_{\chi^2_{d^*}}(z)| = o(1)$$

since $\nu_{d^*}(B_{\sqrt{z}}) = F_{\chi^2_{d^*}}(z)$. This completes the proof of result (21).

Part (i) follows by combining (21) and the inequality:

$$\sup_z \left(\mathbb{P}_Z \left(\|Z\|^2 \leq z \middle| Z \in T - \mathbf{T}v \right) - \mathbb{P}_Z(\|\mathbf{T}Z\|^2 \leq z) \right) \leq 0 \quad \text{for all } v \in \mathbb{R}^{d^*} \tag{51}$$

(see Theorem 2 in Chen and Gao (2017)). Part (ii) also follows from (21) by observing that if $T = \mathbb{R}^{d^*}$ then $T - \mathbb{V}_n = \mathbb{R}^{d^*}$. ■

Proof of Theorem 4.1. We verify the conditions of Lemma 2.1. We may assume without loss of generality that $L_n(\hat{\theta}) = \sup_{\theta \in \Theta_{osn}} L_n(\theta) + o_{\mathbb{P}}(n^{-1})$ because $\widehat{\Theta}_\alpha$ does not depend on the precise $\hat{\theta}$ used (cf. Remark 1). By Lemma F.1 we have:

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

with $Z \sim N(0, I_{d^*})$ when $\Sigma = I_{d^*}$. Let z_α denote the α quantile of the distribution of $\|\mathbf{T}Z\|^2$.

For part (i), Lemma 4.1(i) shows that the posterior distribution of the QLR asymptotically (first-order) stochastically dominates the distribution of $\|\mathbf{T}Z\|^2$ which implies that $\xi^{post}_{n,\alpha} \geq z_\alpha + o_{\mathbb{P}}(1)$. Therefore:

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

where the final equality is by Assumption 4.4.

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

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

Further:

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

by Lemma 4.1(ii) and Assumption 4.4. ■

Lemma F.2. *Let Assumptions 4.1(i) and 4.2' hold. Then:*

$$\sup_{\theta\in\Theta_{osn}} \left|Q_n(\theta) - \left(\|\sqrt{n}\gamma(\theta) - \mathbf{T}\mathbb{V}_n\|^2 + 2f_{n,\perp}(\gamma_\perp(\theta))\right)\right| = o_{\mathbb{P}}(1)\,. \tag{52}$$

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

Proof of Lemma F.2. Using Assumptions 4.1(i) and 4.2', we obtain:

$$\begin{aligned}
2nL_n(\hat{\theta}) &= \sup_{\theta\in\Theta_{osn}} \left(2\ell_n + \|\mathbf{T}\mathbb{V}_n\|^2 - \|\sqrt{n}\gamma(\theta) - \mathbf{T}\mathbb{V}_n\|^2 - 2f_{n,\perp}(\gamma_\perp(\theta))\right) + o_{\mathbb{P}}(1) \\
&= 2\ell_n + \|\mathbf{T}\mathbb{V}_n\|^2 - \inf_{t\in T_{osn}} \|t - \mathbf{T}\mathbb{V}_n\|^2 - \inf_{\theta\in\Theta_{osn}} 2f_{n,\perp}(\gamma_\perp(\theta)) + o_{\mathbb{P}}(1) \\
&= 2\ell_n + \|\mathbf{T}\mathbb{V}_n\|^2 + o_{\mathbb{P}}(1)\,,
\end{aligned} \tag{53}$$

because $\mathbf{T}\mathbb{V}_n \in T$ and $f_{n,\perp}(\cdot) \geq 0$ with $f_{n,\perp}(0) = 0$, $\gamma_\perp(\theta) = 0$ for all $\theta \in \Theta_I$ thus:

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

Then by Assumption 4.2'(i) and definition of Q_n, we obtain:

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

where the $o_{\mathbb{P}}(1)$ term holds uniformly over Θ_{osn}. This proves expression (52).

Since $\gamma(\theta) = 0$ and $\gamma_\perp(\theta) = 0$ for $\theta \in \Theta_I$, and $f_{n,\perp}(0) = 0$ (almost surely), we therefore have $\sup_{\theta\in\Theta_I} Q_n(\theta) = \|\mathbf{T}\mathbb{V}_n\|^2 + o_{\mathbb{P}}(1)$. ■

Proof of Lemma 4.2. We first show that relation (23) holds. By identical arguments to the proof of Lemma 4.1, it is enough to characterize the large-sample behavior of $R_n(z)$ defined in (45). By Lemma F.2 and Assumption 4.2', there exists a positive sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ independent of z with $\varepsilon_n = o(1)$ such that:

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

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

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

$$\left\{\theta \in \Theta_{osn} : \|\sqrt{n}\gamma(\theta) - \mathbf{T}\mathbb{V}_n\|^2 + 2f_{n,\perp}(\gamma_\perp(\theta)) \pm \varepsilon_n \leq z\right\} \subseteq \left\{\theta \in \Theta_{osn} : \|\sqrt{n}\gamma(\theta) - \mathbf{T}\mathbb{V}_n\|^2 \pm \varepsilon_n \leq z\right\}$$

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

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

uniformly in z. Define $\Gamma_{osn} = \{\gamma(\theta) : \theta \in \Theta_{osn}\}$ and $\Gamma_{\perp,osn} = \{\gamma_\perp(\theta) : \theta \in \Theta_{osn}\}$. By similar arguments to the proof of Lemma 4.1, Assumption 4.3'(ii) and a change of variables yield:

$$R_n(z) \leq e^{2\varepsilon_n}(1+\bar{\varepsilon}_n) \frac{\int_{(\{\gamma:\|\sqrt{n}\gamma-\mathbf{T}\mathbb{V}_n\|^2 \leq z+\varepsilon_n\}\cap\Gamma_{osn})\times\Gamma_{\perp,osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma-\mathbf{T}\mathbb{V}_n\|^2 - f_{n,\perp}(\gamma_\perp)} \mathrm{d}(\gamma,\gamma_\perp)}{\int_{\Gamma_{osn}\times\Gamma_{\perp,osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma-\mathbf{T}\mathbb{V}_n\|^2 - f_{n,\perp}(\gamma_\perp)} \mathrm{d}(\gamma,\gamma_\perp)}$$

which holds uniformly in z (wpa1) for some $\bar{\varepsilon}_n = o(1)$. By Tonelli's theorem and Assumption 4.2'(ii), the preceding inequality becomes:

$$R_n(z) \leq e^{2\varepsilon_n}(1+\bar{\varepsilon}_n) \frac{\int_{(\{\gamma:\|\sqrt{n}\gamma-\mathbf{T}\mathbb{V}_n\|^2 \leq z+\varepsilon_n)\cap\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma-\mathbf{T}\mathbb{V}_n\|^2} \mathrm{d}\gamma}{\int_{\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma-\mathbf{T}\mathbb{V}_n\|^2} \mathrm{d}\gamma}.$$

The rest of the proof of inequality (23) follows by similar arguments to the proof of Lemma 4.1. The conclusion now follows by combining inequalities (23) and (51). ■

Proof of Theorem 4.2. We verify the conditions of Lemma 2.1. Again, we assume wlog that $L_n(\hat{\theta}) = \sup_{\theta\in\Theta_{osn}} L_n(\theta) + o_{\mathbb{P}}(n^{-1})$. By Lemma F.2, when $\Sigma = I_{d^*}$, we have:

$$\sup_{\theta\in\Theta_I} Q_n(\theta) = \|\mathbf{T}\mathbb{V}_n\|^2 + o_{\mathbb{P}}(1) \rightsquigarrow \|\mathbf{T}Z\|^2 \tag{54}$$

where $Z \sim N(0, I_{d^*})$. Lemma 4.2 shows that the posterior distribution of the QLR asymptotically (first-order) stochastically dominates the F_T distribution. The result follows by the same arguments as the proof of Theorem 4.1(i). ■

Lemma F.3. *Let Assumptions 4.1(i) and 4.2 or 4.2' and 4.5 hold. Then:*

$$\sup_{\theta\in\Theta_{osn}} \left|PQ_n(M(\theta)) - f\left(\mathbf{T}\mathbb{V}_n - \sqrt{n}\gamma(\theta)\right)\right| = o_{\mathbb{P}}(1).$$

Proof of Lemma F.3. By display (42) in the proof of Lemma F.1 or display (53) in the proof of Lemma F.2 and Assumption 4.5, we obtain:

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

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

Proof of Lemma 4.3. We prove the result under Assumptions 4.1, 4.2', 4.3', and 4.5. The proof under Assumptions 4.1, 4.2, 4.3, and 4.5 follows similarly. By the same arguments as the proof of Lemma 4.1, it suffices to characterize the large-sample behavior of:

$$R_n(z) := \frac{\int_{\{\theta: PQ_n(M(\theta)) \leq z\} \cap \Theta_{osn}} e^{nL_n(\theta)} \, \mathrm{d}\Pi(\theta)}{\int_{\Theta_{osn}} e^{nL_n(\theta)} \, \mathrm{d}\Pi(\theta)} . \tag{55}$$

By Lemma F.3 and Assumption 4.2', there exists a positive sequence $(\varepsilon_n)_{n \in \mathbb{N}}$ independent of z with $\varepsilon_n = o(1)$ such that the inequalities:

$$\sup_{\theta \in \Theta_{osn}} \left| PQ_n(M(\theta)) - f(\mathbf{T}\mathbb{V}_n - \sqrt{n}\gamma(\theta)) \right| \leq \varepsilon_n$$

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

both hold wpa1. Therefore, wpa1 we have:

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

uniformly in z. By similar arguments to the proof of Lemma 4.2, we may use the change of variables $\theta \mapsto (\gamma(\theta), \gamma_\perp(\theta))$, continuity of π_{Γ^*} (Assumption 4.3'(ii)), and Tonelli's theorem to restate the preceding inequalities as:

$$(1 - \bar{\varepsilon}_n) e^{-2\varepsilon_n} \frac{\int_{\{\gamma: f(\mathbf{T}\mathbb{V}_n - \sqrt{n}\gamma) \leq z - \varepsilon_n\} \cap \Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma - \mathbf{T}\mathbb{V}_n\|^2} \mathrm{d}\gamma}{\int_{\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma - \mathbf{T}\mathbb{V}_n\|^2} \mathrm{d}\gamma}$$
$$\leq R_n(z) \leq (1 + \bar{\varepsilon}_n) e^{2\varepsilon_n} \frac{\int_{\{\gamma: f(\mathbf{T}\mathbb{V}_n - \sqrt{n}\gamma) \leq z + \varepsilon_n\} \cap \Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma - \mathbf{T}\mathbb{V}_n\|^2} \mathrm{d}\gamma}{\int_{\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma - \mathbf{T}\mathbb{V}_n\|^2} \mathrm{d}\gamma}$$

which holds (wpa1) for some $\bar{\varepsilon}_n = o(1)$. Let $f^{-1}(z) = \{\kappa \in \mathbb{R}^{d^*} : f(\kappa) \leq z\}$. A second change of variables $\mathbf{T}\mathbb{V}_n - \sqrt{n}\gamma \mapsto \kappa$ yields:

$$(1 - \bar{\varepsilon}_n) e^{-2\varepsilon_n} \frac{\nu_{d^*}((f^{-1}(z - \varepsilon_n)) \cap (\mathbf{T}\mathbb{V}_n - T_{osn}))}{\nu_{d^*}(\mathbf{T}\mathbb{V}_n - T_{osn})}$$
$$\leq R_n(z) \leq (1 + \bar{\varepsilon}_n) e^{2\varepsilon_n} \frac{\nu_{d^*}((f^{-1}(z + \varepsilon_n)) \cap (\mathbf{T}\mathbb{V}_n - T_{osn}))}{\nu_{d^*}(\mathbf{T}\mathbb{V}_n - T_{osn})}$$

uniformly in z, where it should be understood that $\mathbf{T}\mathbb{V}_n - T_{osn}$ is the Minkowski sum $\mathbf{T}\mathbb{V}_n + (-T_{osn})$ with $-T_{osn} = \{-\kappa : \kappa \in T_{osn}\}$.

The remainder of the proof follows by similar arguments to the proof of Lemma 4.1, noting that

$$\sup_{z\in I}\left|\frac{\nu_{d^*}((f^{-1}(z\pm\varepsilon_n))\cap(\mathbf{T}\mathbb{V}_n-T))}{\nu_{d^*}(\mathbf{T}\mathbb{V}_n-T)}-\mathbb{P}_{Z|\mathbf{X}_n}\Big(f(Z)\le z\Big|Z\in\mathbf{T}\mathbb{V}_n-T\Big)\right|$$
$$\le\sup_{z\in I}|\nu_{d^*}(f^{-1}(z\pm\varepsilon_n))-\nu_{d^*}(f^{-1}(z))|=o(1)$$

where the final equality is by uniform continuity of bounded, monotone continuous functions. ■

Proof of Theorem 4.3. We verify the conditions of Lemma 2.2. Again, we assume wlog that $L_n(\hat{\theta})=\sup_{\theta\in\Theta_{osn}}L_n(\theta)+o_{\mathbb{P}}(n^{-1})$.

To prove Theorem 4.3(i), let ξ_α denote the α quantile of $f(\mathbf{T}Z)$. By Lemma 4.3 we need to show that

$$\mathbb{P}_Z(f(Z)\le z|Z\in\mathbf{T}v-T)\le\mathbb{P}_Z(f(\mathbf{T}Z)\le z)$$

holds for all $v\in\mathbb{R}^{d^*}$ and all z in a neighborhood of ξ_α. To prove this, i is sufficient to show that

$$\nu_{d^*}(f^{-1}(z)\cap(\mathbf{T}v-T))\le\nu_{d^*}(\mathbf{T}v-T)\times\nu_{d^*}(\{\kappa\in\mathbb{R}^{d^*}:f(\mathbf{T}\kappa)\le z\})$$

holds for each z and each $v\in\mathbb{R}^{d^*}$. But notice that

$$\nu_{d^*}(f^{-1}(z)\cap(\mathbf{T}v-T))\le\nu_{d^*}((f^{-1}(z)-T^o)\cap(\mathbf{T}v-T))\le\nu_{d^*}(f^{-1}(z)-T^o)\times\nu_{d^*}(\mathbf{T}v-T)$$

where the first inequality is because $f^{-1}(z)\subseteq f^{-1}(z)-T^o=\{\kappa_1+\kappa_2:\kappa_1\in f^{-1}(z),-\kappa_2\in T^o\}$ since $0\in T^o$ and the second inequality is by Theorem 1 of Chen and Gao (2017) (taking $A=\{\mathbf{T}v\}$, $B=f^{-1}(z)$, $C=-T$ and $D=-T^o$ in their notation).

Whenever $\nu_{d^*}(f^{-1}(z)-T^o)\le\nu_{d^*}(\{\kappa\in\mathbb{R}^{d^*}:f(\mathbf{T}\kappa)\le z\})$ holds (which it does, in particular, when f is subconvex), we therefore have:

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

where the final equality is by Assumption 4.6.

To prove Theorem 4.3(ii), when $T=\mathbb{R}^{d^*}$ we have $PQ_n(M_I)\rightsquigarrow f(Z)$. Let ξ_α denote the α quantile of $f(Z)$. Then:

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

by Lemma 4.3 and Assumption 4.6. ■

Proof of Theorem 4.4. By Lemma 2.2, it is enough to show that $\Pr(W^*\le w)\ge F_{\chi^2_1}(w)$ holds for $w\ge 0$, where $W^*=\max_{i\in\{1,2\}}\inf_{t\in T_i}\|Z-t\|^2$.

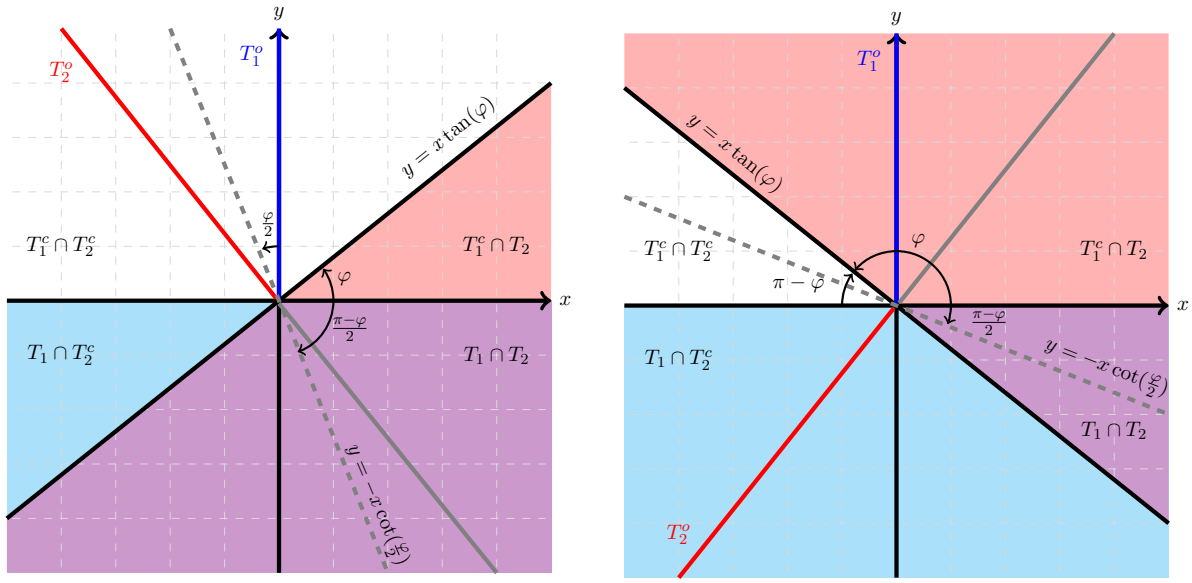

Figure 9: Cones and polar cones for the proof of Theorem 4.4.

Case 1: $d^* = 1$**.** Wlog let $T_1 = [0, \infty)$ and $T_1^o = (-\infty, 0]$. If $T_2 = T_1$ then $\mathbf{T}_1^o Z = \mathbf{T}_2^o Z = (Z \wedge 0)$ so $W^* = (Z \wedge 0)^2 \leq Z^2 \sim \chi_1^2$. If $T_2 = T_1^o$ then $\mathbf{T}_1^o Z = (Z \wedge 0)$ and $\mathbf{T}_2^o Z = (Z \vee 0)$, so $W^* = Z^2 \sim \chi_1^2$. In either case, we have: $\Pr(W^* \leq w) \geq F_{\chi_1^2}(w)$ for any $w \geq 0$.

Case 2: $d^* = 2$**.** Wlog let $T_1 = \{(x, y) : y \leq 0\}$ then T_1^o is the positive y-axis. Let $Z = (X, Y)'$. If $T_1 = T_2$ then $\mathbf{T}_1^o Z = \mathbf{T}_2^o Z = (Y \vee 0)$, so $W^* = (Y \vee 0)^2 \leq Y^2 \sim \chi_1^2$. If $T_2 = \{(x, y) : y \geq 0\}$ then T_2^o is the negative y-axis. So, in this case, $\mathbf{T}_1^o Z = (Y \vee 0)$, $\mathbf{T}_2^o = (Y \wedge 0)$ and so $W^* = Y^2 \sim \chi_1^2$.

Now let T_2 be the rotation of T_1 by $\varphi \in (0, \pi)$ radians. This is plotted in Figure 9 for $\varphi \in (0, \pi/2)$ (left panel) and $\varphi \in (\frac{\pi}{2}, \pi)$ (right panel). The axis of symmetry is the line $y = -x\cot(\frac{\varphi}{2})$, which bisects the angle between T_1^o and T_2^o.

Suppose $Z = (X, Y)'$ lies in the half-space $Y \geq -X\cot(\frac{\varphi}{2})$. There are three options:

- $Z \in (T_1 \cap T_2)$ (purple region): $\mathbf{T}_1^o Z = 0$, $\mathbf{T}_2^o Z = 0$, so $W^* = 0$
- $Z \in (T_1^c \cap T_2)$ (red region): $\mathbf{T}_1^o Z = (0, Y)'$, $\mathbf{T}_2^o Z = 0$, so $W^* = Y^2$
- $Z \in (T_1^c \cap T_2^c)$ (white region): $\mathbf{T}_1^o Z = (0, Y)'$. To calculate $\mathbf{T}_2^o Z$, observe that if we rotate about the origin by $-\varphi$ then the polar cone $\mathbf{T}_2^o$ becomes the positive y axis. Under the rotation, $\mathbf{T}_2^o Z = (0, Y^*)$ where Y^* is the y-value of the rotation of (X, Y) by negative φ. The point (X, Y) rotates to $(X\cos\varphi + Y\sin\varphi, Y\cos\varphi - X\sin\varphi)$, so we get $\|\mathbf{T}_2^o Z\|^2 = (Y\cos\varphi - X\sin\varphi)^2$. We assumed $Y \geq -X\cot(\frac{\varphi}{2})$. By the half-angle formula $\cot(\frac{\varphi}{2}) = \frac{\sin\varphi}{1-\cos\varphi}$, this means that $Y \geq Y\cos\varphi - X\sin\varphi$. But $Y\cos\varphi - X\sin\varphi \geq 0$ as $Y \geq X\tan\varphi$. Therefore, $(Y\cos\varphi - X\sin\varphi)^2 \leq Y^2$ and so $W^* = Y^2$.

We have shown that $W^* \leq Y^2$ whenever $Y \geq -X\cot(\frac{\varphi}{2})$. Now, for any $w \geq 0$:

$$\Pr(W^* \leq w | Y \geq -X\cot(\tfrac{\varphi}{2})) \geq \Pr(Y^2 \leq w | Y \geq -X\cot(\tfrac{\varphi}{2})) = \Pr(Y^2 \leq w | V \geq 0) \tag{56}$$

where $V = Y\sin(\frac{\varphi}{2}) + X\cos(\frac{\varphi}{2})$. Note that Y and V are jointly normal with mean 0, unit variance, and correlation $\rho = \sin(\frac{\varphi}{2})$. The pdf of Y given $V \geq 0$ is:

$$f(y|V \geq 0) = \frac{\int_0^\infty f_{Y|V}(y|v) f_V(v)\mathrm{d}v}{\int_0^\infty f_V(v)\mathrm{d}v} = 2f_Y(y)(1 - F_{V|Y}(0|y))\,.$$

As $V|Y = y \sim N(\rho y, (1-\rho^2))$, we have:

$$F_{V|Y}(0|y) = \Phi\Big(\frac{-\rho y}{\sqrt{1-\rho^2}}\Big) = 1 - \Phi\Big(\frac{\rho}{\sqrt{1-\rho^2}}y\Big)$$

and so

$$f(y|V \geq 0) = 2\phi(y)\Phi\Big(\frac{\rho}{\sqrt{1-\rho^2}}y\Big)\,.$$

Therefore:

$$\Pr(Y^2 \leq w | V \geq 0) = \Pr(-\sqrt{w} \leq y \leq \sqrt{w} | V \geq 0) = \int_{-\sqrt{w}}^{\sqrt{w}} 2\phi(y)\Phi\Big(\frac{\rho}{\sqrt{1-\rho^2}}y\Big)\,\mathrm{d}y\,. \tag{57}$$

But differentiating the right-hand side of (57) with respect to ρ gives:

$$\frac{\mathrm{d}}{\mathrm{d}\rho}\int_{-\sqrt{w}}^{\sqrt{w}} 2\phi(y)\Phi\Big(\frac{\rho}{\sqrt{1-\rho^2}}y\Big)\,\mathrm{d}y = \frac{1}{(1-\rho^2)^{3/2}}\int_{-\sqrt{w}}^{\sqrt{w}} 2y\phi(y)\phi\Big(\frac{\rho}{\sqrt{1-\rho^2}}y\Big)\,\mathrm{d}y = 0$$

for any $\rho \in (-1, 1)$, because $y\phi(y)\phi\big(\rho y/\sqrt{1-\rho^2}\big)$ is an odd function. Therefore, the probability in display (57) doesn't depend on the value of ρ. Setting $\rho = 0$, we obtain:

$$\Pr(Y^2 \leq w | V \geq 0) = \int_{-\sqrt{w}}^{\sqrt{w}} 2\phi(y)\Phi(0)\,\mathrm{d}y = \Phi(\sqrt{w}) - \Phi(-\sqrt{w}) = F_{\chi_1^2}(w)\,.$$

Therefore, by inequality (56) we have:

$$\Pr(W^* \leq w | Y \geq -X\cot(\tfrac{\varphi}{2})) \geq F_{\chi_1^2}(w)\,.$$

By symmetry, we also have $\Pr(W^* \leq w | Y < -X\cot(\frac{\varphi}{2})) \geq F_{\chi_1^2}(w)$. Therefore, we have shown that $\Pr(W^* \leq w) \geq F_{\chi_1^2}(w)$ holds for each $w \geq 0$. A similar argument applies when T_2 is the rotation of T_1 by $\varphi \in (-\pi, 0)$ radians. This completes the proof of the case $d^* = 2$.

Case 3: $d^* \geq 3$**.** As T_1 and T_2 are closed half-spaces we have $T_1 = \{z \in \mathbb{R}^{d^*} : a'z \leq 0\}$ and $T_2 = \{z \in \mathbb{R}^{d^*} : b'z \leq 0\}$ for some $a, b \in \mathbb{R}^{d^*} \setminus \{0\}$. The polar cones are the rays $T_1^o = \{sa : s \geq 0\}$ and $T_2^o = \{sb : s \geq 0\}$. There are three sub-cases to consider.

Case 3a: $a = sb$ for some $s > 0$. Let $u_a = \frac{a}{\|a\|}$. Here $T_1 = T_2$, $\mathbf{T}_1^o Z = \mathbf{T}_2^o Z = 0$ if $Z \in T_1$, and

$$\mathbf{T}_1^o Z = \mathbf{T}_2^o Z = u_a(Z'u_a) \quad \text{if } Z \notin T_1 \text{ (i.e. if } Z'u_a > 0)\,.$$

Therefore, $W^* = (Z'u_a \vee 0)^2 \leq (Z'u_a)^2 \sim \chi_1^2$.

Case 3b: $a = sb$ for some $s < 0$. Here $T_1 = -T_2$ and $T_1^o = -T_2^o$, so $\mathbf{T}_1^o Z = 0$ and $\mathbf{T}_2^o Z = u_a(Z'u_a)$ if $Z \in T_1$ (i.e. if $Z'u_a \leq 0$) and $\mathbf{T}_1^o Z = u_a(Z'u_a)$ and $\mathbf{T}_2^o Z = 0$ if $Z \notin T_1$ (i.e. if $Z'u_a > 0$). Therefore $W^* = (Z'u_a)^2 \sim \chi_1^2$.

Case 3c: a and b are linearly independent. Without loss of generality,[24] we can take T_1^o to be the positive y-axis (i.e. $a = (0, a_2, 0, \ldots, 0)'$ for some $a_2 > 0$) and take T_2^o to lie in the (x, y)-plane (i.e. $b = (b_1, b_2, 0, \ldots, 0)'$ for some $b_1 \neq 0$).

Now write $Z = (X, Y, U)$ where $U \in \mathbb{R}^{d^*-2}$. Note that $a'Z = a_2 Y$ and $b'Z = b_1 X + b_2 Y$. So only the values of X and Y matter in determining whether or not Z belongs to T_1 and T_2.

Without loss of generality we may assume that $(b_1, b_2)'$ is, up to scale, a rotation of $(0, a_2)'$ by $\varphi \in (0, \pi)$ (the case $(-\pi, 0)$ can be handled by similar arguments, as in Case 2).

Suppose that $Y \geq -X \cot(\frac{\varphi}{2})$. As in Case 2, there are three options:

- $Z \in (T_1 \cap T_2)$: $\mathbf{T}_1^o Z = 0$, $\mathbf{T}_2^o Z = 0$, so $W^* = 0$
- $Z \in (T_1^c \cap T_2)$: $\mathbf{T}_1^o Z = (0, Y, 0, \ldots, 0)'$, $\mathbf{T}_2^o Z = 0$, so $W^* = Y^2$
- $Z \in (T_1^c \cap T_2^c)$: $\|\mathbf{T}_1^o Z\|^2 = Y^2$ and $\|\mathbf{T}_2^o Z\|^2 = (Y \cos\varphi - X \sin\varphi)^2 \leq Y^2$, so $W^* = Y^2$.

Arguing as in Case 2, we obtain $\Pr(W^* \leq w | Y \geq -X \cot(\frac{\varphi}{2})) \geq F_{\chi_1^2}(w)$. By symmetry, we also have $\Pr(W^* \leq w | Y < -X \cot(\frac{\varphi}{2})) \geq F_{\chi_1^2}(w)$. Therefore, $\Pr(W^* \leq w) \geq F_{\chi_1^2}(w)$. ■

Proof of Proposition 4.1. It follows from condition (i) and display (42) or display (53) that:

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

Moreover, applying conditions (ii) and (iii), we obtain:

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

[24] By Gram-Schmidt, we can always define a new set of coordinate vectors $e_1, e_2, \ldots, e_{d^*}$ for $\mathbb{R}^{d^*}$ with $e_2 = u_a$ and such that b is in the span of e_1 and e_2.

Therefore:

$$\sup_{\mu\in M_I}\inf_{\eta\in H_\mu} Q_n(\mu,\eta) = \max_{\mu\in\{\underline{\mu},\overline{\mu}\}}\inf_{t\in T_\mu}\|\mathbb{V}_n - t\|^2 + o_{\mathbb{P}}(1)\,.$$

The result now follows from $\Sigma = I_{d^*}$. ■

F.2 Proofs and Additional Lemmas for Section 5

Proof of Proposition 5.1. Wlog we can take $\tilde{\gamma}_0 = 0$. Also take n large enough that $\{\tilde{\gamma} : \|\tilde{\gamma}\| \leq n^{-1/4}\} \subseteq U$. Then by condition (b), for any such $\tilde{\gamma}$ we have:

$$nL_n(\tilde{\gamma}) = nL_n(\tilde{\gamma}_0) + (\sqrt{n}\tilde{\gamma})'(\sqrt{n}\mathbb{P}_n\dot{\ell}_{\tilde{\gamma}_0}) + \frac{1}{2}(\sqrt{n}\tilde{\gamma})'(\mathbb{P}_n\ddot{\ell}_{\tilde{\gamma}^*})(\sqrt{n}\tilde{\gamma})$$

where $\tilde{\gamma}^*$ is in the segment between $\tilde{\gamma}$ and $\tilde{\gamma}_0$ for each element of $\mathbb{P}_n\ddot{\ell}_{\tilde{\gamma}^*}$. We may deduce from Lemma 2.4 of Newey and McFadden (1994) that $\sup_{\tilde{\gamma}:\|\tilde{\gamma}\|\leq n^{-1/4}}\|(\mathbb{P}_n\ddot{\ell}_{\tilde{\gamma}^*}) - P_0(\ddot{\ell}_{\gamma_0})\| = o_{\mathbb{P}}(1)$ holds under conditions (a) and (b). Since this term is $o_{\mathbb{P}}(1)$, we can choose a positive sequence $(r_n)_{n\in\mathbb{N}}$ with $r_n \to \infty$, $r_n = o(n^{-1/4})$ such that $r_n^2\sup_{\tilde{\gamma}:\|\tilde{\gamma}\|\leq n^{1/4}}\|(\mathbb{P}_n\ddot{\ell}_{\tilde{\gamma}^*}) - P_0(\ddot{\ell}_{\tilde{\gamma}_0})\| = o_{\mathbb{P}}(1)$. Assumption 4.2 then holds over $\Theta_{osn} = \{\theta \in \Theta : \|\tilde{\gamma}(\theta)\| \leq r_n/\sqrt{n}\}$ with $\ell_n = nL_n(\tilde{\gamma}_0)$, $\gamma(\theta) = \mathbb{I}_{\tilde{\gamma}_0}^{1/2}\tilde{\gamma}(\theta)$, $\sqrt{n}\hat{\gamma}_n = \mathbb{V}_n = \mathbb{I}_{\tilde{\gamma}_0}^{-1/2}\mathbb{P}_n(\dot{\ell}_{\gamma_0})$, $\Sigma = I_{d^*}$.

It remains to show that the posterior concentrates on Θ_{osn}. Choose ε sufficiently small that $U_\varepsilon = \{\tilde{\gamma} : \|\tilde{\gamma}\| < \varepsilon\} \subseteq U$. By a similar expansion to the above and condition (c), we have $D_{KL}(p_0\|q_{\tilde{\gamma}}) = -\frac{1}{2}\tilde{\gamma}'P_0(\ddot{\ell}_{\tilde{\gamma}^*})\tilde{\gamma}$ where $\tilde{\gamma}^*$ is in the segment between $\tilde{\gamma}^*$ and $\tilde{\gamma}_0$. Since $\|P_0(\ddot{\ell}_{\tilde{\gamma}^*}) + \mathbb{I}_{\tilde{\gamma}_0}\| \to 0$ as $\|\tilde{\gamma}\| \to 0$, we may reduce ε so that $\inf_{\tilde{\gamma}\in U_\varepsilon}\|P_0(\ddot{\ell}_{\tilde{\gamma}^*}) + \mathbb{I}_{\tilde{\gamma}_0}\| \leq \frac{1}{2}\lambda_{\min}(\mathbb{I}_{\tilde{\gamma}_0})$. On U_ε we then have that there exist finite positive constants $\underline{c}$ and $\overline{c}$ such that $\underline{c}\|\tilde{\gamma}\|^2 \leq D_{KL}(p_0\|q_{\tilde{\gamma}}) \leq \overline{c}\|\tilde{\gamma}\|^2$. Also note that $\inf_{\tilde{\gamma}\in\widetilde{\Gamma}\setminus U_\varepsilon} D_{KL}(p_0\|q_{\tilde{\gamma}}) =: \delta$ with $\delta > 0$ by identifiability of $\tilde{\gamma}_0$, continuity of the map $\tilde{\gamma} \mapsto P_0\ell_{\tilde{\gamma}}$, and compactness of $\widetilde{\Gamma}$. Standard consistency arguments (e.g. the Corollary to Theorem 6.1 in Schwartz (1965))) then imply that $\Pi_n(U_\varepsilon|\mathbf{X}_n) \to_{a.s.} 1$. Since the posterior concentrates on U_ε and $\Theta_{osn} \subset U_\varepsilon$ for all n sufficiently large, it's enough to confine attention to U_ε. We have shown that $\underline{c}\|\tilde{\gamma}\|^2 \leq D_{KL}(p_0\|q_{\tilde{\gamma}}) \leq \overline{c}\|\tilde{\gamma}\|^2$ holds on U_ε. It now follows by the parametric Bernstein-von Mises theorem (e.g. Theorem 10.1 in van der Vaart (2000)) that the posterior contracts at a $\sqrt{n}$-rate, verifying Assumption 4.1(ii). ■

For the following lemma, let $(r_n)_{n\in\mathbb{N}}$ be a positive sequence with $r_n \to \infty$ and $r_n = o(n^{1/2})$, $\mathcal{P}_{osn} = \{p \in \mathcal{P} : h(p,p_0) \leq r_n/\sqrt{n}\}$ and $\Theta_{osn} = \{\theta \in \Theta : h(p_\theta,p_0) \leq r_n/\sqrt{n}\}$. For each $p \in \mathcal{P}$ with $p \neq p_0$, define $S_p = \sqrt{p/p_0} - 1$ and $s_p = S_p/h(p,p_0)$. Recall the definitions of $\overline{\mathcal{D}}_\varepsilon$, the tagent cone $\mathcal{T}$ and the projection $\mathbb{T}$ from Section 5.1.2. We say $\mathcal{P}$ is r_n-DQM if each p is absolutely continuous with respect to p_0 and for each $p \in \mathcal{P}$ there are $g_p \in \mathcal{T}$ and $R_p \in L^2(\lambda)$ such that:

$$\sqrt{p} - \sqrt{p_0} = g_p\sqrt{p_0} + h(p,p_0)R_p$$

with $\sup\{r_n\|R_p\|_{L^2(\lambda)} : h(p,p_0) \le r_n/\sqrt{n}\} \to 0$ as $n \to \infty$. Let $\overline{\mathcal{D}}_\varepsilon^2 = \{d^2 : d \in \overline{\mathcal{D}}_\varepsilon\}$.

Lemma F.4. *Let the following conditions hold.*
(i) $\mathcal{P}$ is r_n-DQM
(ii) there exists $\varepsilon > 0$ such that $\overline{\mathcal{D}}_\varepsilon^2$ is P_0-Glivenko Cantelli and $\overline{\mathcal{D}}_\varepsilon$ has envelope $D \in L^2(P_0)$ with $\max_{i \le i \le n} D(X_i) = o_{\mathbb{P}}(\sqrt{n}/r_n^3)$
(iii) $\sup_{p \in \mathcal{P}_{osn}} |\mathbb{G}_n(S_p - \mathbb{T}S_p)| = o_{\mathbb{P}}(n^{-1/2})$
(iv) $\sup_{p \in \mathcal{P}_{osn}} |(\mathbb{P}_n - P_0)S_p^2| = o_{\mathbb{P}}(n^{-1})$.
Then:

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

If, in addition, $\mathrm{Span}(\mathcal{T})$ has finite dimension $d^* \ge 1$ then Assumption 4.2 holds over Θ_{osn} with $\ell_n = n\mathbb{P}_n \log p_0$, $\sqrt{n}\hat{\gamma}_n = \mathbb{V}_n = \mathbb{G}_n(\psi)$, $\Sigma = I_{d^*}$ and $\gamma(\theta)$ defined in (25).

Proof of Lemma F.4. We first prove

$$\sup_{p \in \mathcal{P}_{osn}} \left| n\mathbb{P}_n \log(p/p_0) - 2n\mathbb{P}_n(S_p - P_0(S_p)) + n(\mathbb{P}_n S_p^2 + h^2(p,p_0)) \right| = o_{\mathbb{P}}(1) \tag{58}$$

by adapting arguments used in Theorem 1 of Azaïs et al. (2009), Theorem 3.1 in Gassiat (2002), and Theorem 2.1 in Liu and Shao (2003).

Take n large enough that $r_n/\sqrt{n} \le \varepsilon$. Then for each $p \in \mathcal{P}_{osn} \setminus \{p_0\}$:

$$n\mathbb{P}_n \log(p/p_0) = 2n\mathbb{P}_n S_p - n\mathbb{P}_n S_p^2 + 2n\mathbb{P}_n S_p^2 r(S_p) \tag{59}$$

where $r(u) = (\log(1+u) - u - \frac{1}{2}u^2)/u^2$ and $\lim_{u \to 0} |r(u)/(\frac{1}{3}u) - 1| = 0$. By condition (ii), $\max_{1 \le i \le n} |S_p(X_i)| \le r_n/\sqrt{n} \times \max_{1 \le i \le n} D(X_i) = o_{\mathbb{P}}(r_n^{-2})$ uniformly for $p \in \mathcal{P}_{osn}$. This implies that $\sup_{p \in \mathcal{P}_{osn}} \max_{1 \le i \le n} |r(S_p(X_i))| = o_{\mathbb{P}}(r_n^{-2})$. Therefore, by the Glivenko-Cantelli condition in (ii):

$$\sup_{p \in \mathcal{P}_{osn}} |2n\mathbb{P}_n S_p^2 r(S_p)| \le 2r_n^2 \times o_{\mathbb{P}}(r_n^{-2}) \times \sup_{p \in \mathcal{P}_{osn}} \mathbb{P}_n s_p^2 = o_{\mathbb{P}}(1) \times (1 + o_{\mathbb{P}}(1)) = o_{\mathbb{P}}(1)\,.$$

Display (58) now follows by adding and subtracting $2nP_0(S_p) = -nh^2(p,p_0)$ to (59).

Each element of $\mathcal{T}$ has mean zero and so $P_0(\mathbb{T}S_p) = 0$ for each p. By Condition (iii):

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

It remains to show:

$$\sup_{p\in\mathcal{P}_{osn}} \left|\mathbb{P}_n(S_p^2) + h^2(p,p_0) - 2P_0((\mathbb{T}S_p)^2)\right| = o_{\mathbb{P}}(n^{-1})\,. \tag{60}$$

By condition (iv) and $P_0(S_p^2) = h^2(p,p_0)$, to establish (60) it is enough to show:

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

Observe by definition of $\mathbb{T}$ and condition (i), for each $p \in \mathcal{P}$ there is a $g_p \in \mathcal{T}$ and remainder $R_p^* = R_p/\sqrt{p_0}$ such that $S_p = g_p + h(p,p_0)R_p^*$, and so:

$$\|S_p - \mathbb{T}S_p\|_{L^2(P_0)} \le \|S_p - g_p\|_{L^2(P_0)} = h(p,p_0)\|R_p^*\|_{L^2(P_0)} = h(p,p_0)\|R_p\|_{L^2(\lambda)} \tag{61}$$

By Moreau's decomposition theorem and inequality (61), we may deduce:

$$\sup_{p\in\mathcal{P}_{osn}} |P_0(S_p^2) - P_0((\mathbb{T}S_p)^2)| = \sup_{p\in\mathcal{P}_{osn}} \|S_p - \mathbb{T}S_p\|^2_{L^2(P_0)} \le \sup_{p\in\mathcal{P}_{osn}} h(p,p_0)^2\|R_p\|^2_{L^2(\lambda)}$$

which is $o_{\mathbb{P}}(n^{-1})$ by condition (i) and definition of $\mathcal{P}_{osn}$. This proves the first result.

The second result is immediate by defining $\mathbb{V}_n = \mathbb{G}_n(\psi)$ with $\psi = (\psi_1,\ldots,\psi_{d^*})'$ where $\psi_1,\ldots,\psi_{d^*}$ is an orthonormal basis for $\mathrm{Span}(\mathcal{T})$, and $\gamma(\theta)$ as in (25), then noting that $P_0((\mathbb{T}(2S_{p_\theta}))^2) = \gamma(\theta)'P_0(\psi\psi')\gamma(\theta) = \|\gamma(\theta)\|^2$. ■

Proof of Proposition 5.2. We verify the conditions of Lemma F.4. By DQM (condition (b)) we have $\sup\{\|R_p\|_{L^2(\lambda)} : h(p,p_0) \le n^{-1/4}\} \to 0$ as $n \to \infty$. Therefore, we may choose a sequence $(a_n)_{n\in\mathbb{N}}$ with $a_n \le n^{1/4}$ but $a_n \to \infty$ slowly enough that

$$\sup\{a_n\|R_p\|_{L^2(\lambda)} : h(p,p_0) \le a_n/\sqrt{n}\} \to 0 \quad \text{as } n \to \infty$$

and hence $\sup\{r_n\|R_p\|_{L^2(\lambda)} : h(p,p_0) \le r_n/\sqrt{n}\} \to 0$ as $n \to \infty$ for any slowly diverging positive sequence $(r_n)_{n\in\mathbb{N}}$ with $r_n \le a_n$. This verifies condition (i) of Lemma F.4.

For condition (ii), $\overline{\mathcal{D}}_\varepsilon^2$ is Glivenko-Cantelli by condition (c) and Lemma 2.10.14 of van der Vaart and Wellner (1996). Moreover, it follows from the envelope condition (in condition (c)) that $\max_{1\le i\le n} D(X_i) = o_{\mathbb{P}}(n^{1/2})$. We can therefore choose a positive sequence $(c_n)_{n\in\mathbb{N}}$ with $c_n \to \infty$ such that $c_n^3 \max_{1\le i\le n} D(X_i) = o_{\mathbb{P}}(n^{1/2})$ and so $\max_{1\le i\le n} D(X_i) = o_{\mathbb{P}}(n^{1/2}/r_n^3)$ for any $0 < r_n \le c_n$.

For condition (iv), since $\overline{\mathcal{D}}_\varepsilon^2$ is Glivenko-Cantelli we may choose a positive sequence $(b_n)_{n\in\mathbb{N}}$ with $b_n \to \infty$ such that $b_n^2 \sup_{s_p\in\overline{\mathcal{D}}_\varepsilon} |(\mathbb{P}_n - P_0)s_p^2| = o_{\mathbb{P}}(1)$. Therefore, for any $0 < r_n \le b_n$ we have:

$$\sup_{p:h(p,p_0)\le r_n/\sqrt{n}} |(\mathbb{P}_n - P_0)S_p^2| \le \sup_{p:h(p,p_0)\le r_n/\sqrt{n}} r_n^2|(\mathbb{P}_n - P_0)s_p^2|/n = o_{\mathbb{P}}(n^{-1})\,.$$

Finally, for condition (iii), note that condition (c) implies that $\overline{\mathcal{D}}_\varepsilon^o := \{s_p - \mathbb{T}s_p : s_p \in \overline{\mathcal{D}}_\varepsilon\}$ is Donsker. Also note that the singleton $\{0\}$ is the only limit point of $\overline{\mathcal{D}}_\varepsilon^o$ as $\varepsilon \searrow 0$ because:

$$\sup\{\|s_p - \mathbb{T}s_p\|_{L^2(P_0)} : h(p,p_0) \le \varepsilon\} \le \sup\{\|R_p\|_{L^2(\lambda)} : h(p,p_0) \le \varepsilon\} \to 0 \quad (\text{as } \varepsilon \to 0)$$

by DQM (condition (b)). Asymptotic equicontinuity of $\mathbb{G}_n$ on $\overline{\mathcal{D}}_\varepsilon^o$ then implies that

$$\sup_{p:h(p,p_0)\le n^{-1/4}} |\mathbb{G}_n(s_p - \mathbb{T}s_p)| = o_{\mathbb{P}}(1)\,.$$

We can therefore choose a positive sequence $(d_n)_{n\in\mathbb{N}}$ with $d_n \le n^{1/4}$ but $d_n \to \infty$ slowly enough that $d_n \sup_{p:h(p,p_0)\le n^{-1/4}} |\mathbb{G}_n(s_p - \mathbb{T}s_p)| = o_{\mathbb{P}}(1)$ and so for any $0 < r_n \le d_n$:

$$\sup_{p:h(p,p_0)\le r_n/\sqrt{n}} |\mathbb{G}_n(S_p - \mathbb{T}S_p)| \le \frac{r_n}{\sqrt{n}} \sup_{p:h(p,p_0)\le n^{-1/4}} \mathbb{G}_n(s_p - \mathbb{T}s_p) = o_{\mathbb{P}}(n^{-1/2})\,.$$

The result follows by taking $r_n = (a_n \wedge b_n \wedge c_n \wedge d_n)$. ■

Proof of Proposition 5.3. We first show that:

$$\sup_{\theta:\|g(\theta)\|\le r_n/\sqrt{n}} \left| nL_n(\theta) - \left(-\frac{1}{2}(\mathbb{T}(\sqrt{n}g(\theta)) + Z_n)'\Omega^{-1}(\mathbb{T}(\sqrt{n}g(\theta)) + Z_n) \right) \right| = o_{\mathbb{P}}(1) \tag{62}$$

holds for a positive sequence $(r_n)_{n\in\mathbb{N}}$ with $r_n \to \infty$ with $Z_n = \mathbb{G}_n(\rho_{\theta^*})$. Take n large enough that $n^{-1/4} \le \varepsilon_0$. By conditions (a)–(c) and Lemma 2.10.14 of van der Vaart and Wellner (1996), we have that $\sup_{\theta:\|g(\theta)\|\le n^{-1/4}} \|\mathbb{P}_n(\rho_\theta\rho_\theta') - \Omega\| = o_{\mathbb{P}}(1)$. Therefore, we may choose a positive sequence $(a_n)_{n\in\mathbb{N}}$ with $a_n \to \infty$, $a_n = o(n^{1/4})$ such that $\sup_{\theta:\|g(\theta)\|\le n^{-1/4}} a_n^2\|\mathbb{P}_n(\rho_\theta\rho_\theta') - \Omega\| = o_{\mathbb{P}}(1)$ and hence:

$$\sup_{\theta:\|g(\theta)\|\le r_n/\sqrt{n}} \|\mathbb{P}_n(\rho_\theta\rho_\theta') - \Omega\| = o_{\mathbb{P}}(r_n^{-2}) \tag{63}$$

for any $0 < r_n \le a_n$.

Notice that $Z_n \rightsquigarrow N(0,\Omega)$ by condition (a) and that the covariance of each element of $\rho_\theta(X_i) - \rho_{\theta^*}(X_i)$ vanishes uniformly over Θ_I^ε as $\varepsilon \to 0$ by condition (c). Asymptotic equicontinuity of $\mathbb{G}_n$ (which holds under (a)) then implies that $\sup_{\theta:\|g(\theta)\|\le n^{-1/4}} \|\mathbb{G}_n(\rho_\theta) - Z_n\| = o_{\mathbb{P}}(1)$. We can therefore choose a positive sequence $(b_n)_{n\in\mathbb{N}}$ with $b_n \to \infty$, $b_n = o(n^{1/4})$ as $n \to \infty$ such that $b_n \sup_{\theta:\|g(\theta)\|\le b_n/\sqrt{n}} \|\mathbb{G}_n(\rho_\theta) - Z_n\| = o_{\mathbb{P}}(1)$ and hence:

$$\sup_{\theta:\|g(\theta)\|\le r_n/\sqrt{n}} |\sqrt{n}\mathbb{P}_n\rho_\theta - (\sqrt{n}g(\theta) + Z_n)| = o_{\mathbb{P}}(r_n^{-1})\,. \tag{64}$$

for any $0 < r_n \le b_n$.

Condition (d) implies that we may choose a sequence $(c_n)_{n\in\mathbb{N}}$ with $c_n \to \infty$, $c_n = o(n^{1/4})$ such that $\sup_{\theta:\|g(\theta)\|\le c_n/\sqrt{n}} \sqrt{n}\|g(\theta) - \mathbb{T}g(\theta)\| = o(c_n^{-1})$ and so:

$$\sup_{\theta:\|g(\theta)\|\le r_n/\sqrt{n}} \|\sqrt{n}g(\theta) - \mathbb{T}(\sqrt{n}g(\theta))\| = o(r_n^{-1}) \tag{65}$$

for any $0 < r_n \le c_n$.

Result (62) now follows by taking $r_n = (a_n \wedge b_n \wedge c_n)$ and using (63), (64) and (65). To complete the proof, expanding the quadratic in (62) we obtain:

$$\begin{aligned}-\frac{1}{2}(\mathbb{T}(\sqrt{n}g(\theta)) + Z_n)'\Omega^{-1}(\mathbb{T}(\sqrt{n}g(\theta)) + Z_n) = &-\frac{1}{2}Z_n'\Omega^{-1}Z_n - \frac{1}{2}\|[\Omega^{-1/2}\mathbb{T}(\sqrt{n}g(\theta))]_1\|^2 \\ &- [\Omega^{-1/2}Z_n]_1'[\Omega^{-1/2}\mathbb{T}(\sqrt{n}g(\theta))]_1\end{aligned}$$

and the result follows with $\ell_n = Z_n'\Omega^{-1}Z_n$, $\gamma(\theta) = [\Omega^{-1/2}\mathbb{T}g(\theta)]_1$, and $\mathbb{V}_n = -[\Omega^{-1/2}Z_n]_1$. ■

Proof of Proposition 5.4. Follows by similar arguments to the proof of Proposition 5.3, noting that by condition (e) we may choose a positive sequence $(a_n)_{n\in\mathbb{N}}$ with $a_n \to \infty$ slowly such that $a_n^2\|\widehat{W} - \Omega^{-1}\| = o_{\mathbb{P}}(1)$. Therefore $\|\widehat{W} - \Omega^{-1}\| = o_{\mathbb{P}}(r_n^{-2})$ holds for any $0 < r_n \le a_n$. ■

Lemma F.5. *Consider the missing data model with a flat prior on Θ. Suppose that the model is point identified (i.e. the true $\eta_2 = 1$). Then Assumption 4.1(ii) holds for*

$$\Theta_{osn} = \{\theta : |\tilde{\gamma}_{11}(\theta) - \tilde{\gamma}_{11}| \le r_n/\sqrt{n}, \tilde{\gamma}_{00}(\theta) \le r_n/n\}$$

for any positive sequence $(r_n)_{n\in\mathbb{N}}$ with $r_n \to \infty$, $r_n/\sqrt{n} = o(1)$

Proof of Lemma F.5. The flat prior on Θ induces a flat prior on $\{(a,b) \in [0,1] : 0 \le a \le 1-b\}$ under the map $\theta \mapsto (\tilde{\gamma}_{11}(\theta), \tilde{\gamma}_{00}(\theta))$. Take n large enough that $[\tilde{\gamma}_{11} - r_n/\sqrt{n}, \tilde{\gamma}_{11} + r_n/\sqrt{n}] \subseteq [0,1]$ and $r_n/n < 1$. Then with $S_n := \sum_{i=1}^n Y_i$, we have:

$$\begin{aligned}\Pi_n(\Theta_{osn}^c|\mathbf{X}_n) = &\frac{\int_{[0,\tilde{\gamma}_{11}-r_n/\sqrt{n}]\cup[\tilde{\gamma}_{11}+r_n/\sqrt{n}]}\int_0^{1-a}(a)^{S_n}(1-a-b)^{n-S_n}\,\mathrm{d}b\,\mathrm{d}a}{\int_0^1\int_0^{1-a}(a)^{S_n}(1-a-b)^{n-S_n}\,\mathrm{d}b\,\mathrm{d}a} \\ &+ \frac{\int_{\kappa_{11}-k_n/\sqrt{n}}^{\kappa_{11}+k_n/\sqrt{n}}\int_{k_n/n}^{1-a}(a)^{S_n}(1-a-b)^{n-S_n}\,\mathrm{d}b\,\mathrm{d}a}{\int_0^1\int_0^{1-a}(a)^{S_n}(1-a-b)^{n-S_n}\,\mathrm{d}b\,\mathrm{d}a} =: I_1 + I_2\,.\end{aligned}$$

Integrating I_1 first with respect to b yields:

$$I_1 = \frac{\int_{[0,\tilde{\gamma}_{11}-r_n/\sqrt{n}]\cup[\tilde{\gamma}_{11}+r_n/\sqrt{n}]}\int_0^{1-a}(a)^{S_n}(1-a)^{n-S_n+1}\,\mathrm{d}a}{\int_0^1(a)^{S_n}(1-a)^{n-S_n+1}\,\mathrm{d}a} = \mathbb{P}_{U|S_n}(|U - \tilde{\gamma}_{11}| > r_n/\sqrt{n})$$

where $U|S_n \sim \text{Beta}(S_n+1, n-S_n+2)$. Note that this implies:

$$\mathbb{E}[U|S_n] = \frac{S_n+1}{n+3} \qquad\qquad \text{Var}[U|S_n] = \frac{(S_n+1)(n-S_n+2)}{(n+3)^2(n+4)}\,.$$

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

$$\begin{aligned}
I_1 &\leq \mathbb{P}_{U|S_n}\Big(|U - \mathbb{E}[U|S_n]| > r_n/(2\sqrt{n})\Big) + \mathbb{1}\Big\{|\mathbb{E}[U|S_n] - \tilde{\gamma}_{11}| > r_n/(2\sqrt{n})\Big\} \\
&= \mathbb{P}_{U|S_n}\Big(|U - \mathbb{E}[U|S_n]| > r_n/(2\sqrt{n})\Big) + o_{\mathbb{P}}(1) \\
&\leq \frac{4}{r_n^2}\frac{(\frac{S_n}{n}+\frac{1}{n})(1-\frac{S_n}{n}+\frac{2}{n})}{(1+\frac{3}{n})^2(1+\frac{4}{n})} + o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1)\,.
\end{aligned}$$

Similarly:

$$I_2 = \frac{\int_{\tilde{\gamma}_{11}-r_n/\sqrt{n}}^{\tilde{\gamma}_{11}+r_n/\sqrt{n}} (a)^{S_n}(1-a-(r_n/n))^{n-S_n+1}\,\mathrm{d}a}{\int_0^1 (a)^{S_n}(1-a)^{n-S_n+1}\,\mathrm{d}a} \leq \frac{\int_0^{1-r_n/n} (a)^{S_n}(1-a-(r_n/n))^{n-S_n+1}\,\mathrm{d}a}{\int_0^1 (a)^{S_n}(1-a)^{n-S_n+1}\,\mathrm{d}a}\,.$$

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

$$I_2 \leq (1-(r_n/n))^{n+2}\frac{\int_0^1 (1-c)^{S_n}(c)^{n-S_n+1}\,\mathrm{d}c}{\int_0^1 (a)^{S_n}(1-a)^{n-S_n+1}\,\mathrm{d}a} = (1-(r_n/n))^{n+2} \to 0\,.$$

Therefore, $\Pi_n(\Theta_{osn}^c|\mathbf{X}_n) = o_{\mathbb{P}}(1)$, as required. ■

F.3 Proofs and Additional Lemmas for Appendix B

Proof of Lemma B.1. By condition (i), there exists a positive sequence $(\varepsilon_n)_{n\in\mathbb{N}}$, $\varepsilon_n = o(1)$ such that $\sup_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\sup_{\theta\in\Theta_I(\mathbb{P})} Q_n(\theta) - W_n > \varepsilon_n) = o(1)$. Let $\mathcal{A}_{n,\mathbb{P}}$ denote the event on which $\sup_{\theta\in\Theta_I(\mathbb{P})} Q_n(\theta) - W_n \leq \varepsilon_n$. Then:

$$\begin{aligned}
\inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\Theta_I(\mathbb{P}) \subseteq \widehat{\Theta}_\alpha) &\geq \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\{\Theta_I(\mathbb{P}) \subseteq \widehat{\Theta}_\alpha\} \cap \mathcal{A}_{n,\mathbb{P}}) \\
&= \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\{\sup_{\theta\in\Theta_I(\mathbb{P})} Q_n(\theta) \leq v_{\alpha,n}\} \cap \mathcal{A}_{n,\mathbb{P}}) \\
&\geq \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\{W_n \leq v_{\alpha,n} - \varepsilon_n\} \cap \mathcal{A}_{n,\mathbb{P}})\,,
\end{aligned}$$

where the second equality is by the definition of $\widehat{\Theta}_\alpha$. Since $\mathbb{P}(A \cap B) \geq 1 - \mathbb{P}(A^c) - \mathbb{P}(B^c)$, we therefore have:

$$\begin{aligned}
\inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\Theta_I(\mathbb{P}) \subseteq \widehat{\Theta}_\alpha) &\geq 1 - \sup_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(W_n > v_{\alpha,n} - \varepsilon_n) - \sup_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\mathcal{A}^c_{n,\mathbb{P}}) \\
&= 1 - (1 - \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(W_n \leq v_{\alpha,n} - \varepsilon_n)) - o(1) \\
&= \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(W_n \leq v_{\alpha,n} - \varepsilon_n) - o(1) \\
&\geq \alpha - o(1)\,,
\end{aligned}$$

where the final line is by condition (ii) and definition of $\mathcal{A}_{n,\mathbb{P}}$. ■

Proof of Lemma B.2. Follows by similar arguments to the proof of Lemma B.1. ■

We use the next Lemma several times in the following proofs.

Lemma F.6. *Let $T \subseteq \mathbb{R}^d$ be a closed convex cone and let $\mathbf{T}$ denote the projection onto T. Then:*

$$\|\mathbf{T}(x+t) - t\| \leq \|x\|$$

for any $x \in \mathbb{R}^d$ and $t \in T$.

Proof of Lemma F.6. Let $\mathbf{T}^o$ denote the projection onto the polar cone T^o of T. Since $u't \leq 0$ holds for any $u \in T^o$ and $\|\mathbf{T}v\| \leq \|v\|$ holds for any $v \in \mathbb{R}^d$, we obtain:

$$\|\mathbf{T}(x+t)\|^2 + 2(\mathbf{T}^o(x+t))'t \leq \|\mathbf{T}(x+t)\|^2 \leq \|x+t\|^2\,.$$

Subtracting $2(x+t)'t$ from both sides and using the fact that $v = \mathbf{T}v + \mathbf{T}^o v$ yields:

$$\|\mathbf{T}(x+t)\|^2 - 2(\mathbf{T}(x+t))'t \leq \|x+t\|^2 - 2(x+t)'t\,.$$

Adding $\|t\|^2$ to both sides and completing the square gives $\|\mathbf{T}(x+t)-t\|^2 \leq \|x+t-t\|^2 = \|x\|^2$. ■

In view of Lemma F.6 and Assumption B.2(i), for each $\mathbb{P} \in \mathbf{P}$ we have:

$$\|\sqrt{n}(\hat{\gamma}_n - \tau)\| \leq \|\mathbb{V}_n\|\,. \tag{66}$$

Lemma F.7. *Let Assumptions B.1(i) and B.2 hold. Then:*

$$\sup_{\theta\in\Theta_{osn}} \left| Q_n(\theta) - \|\sqrt{n}\gamma(\theta) - \sqrt{n}(\hat{\gamma}_n - \tau)\|^2 - 2f_{n,\perp}(\gamma_\perp(\theta)) \right| = o_{\mathbb{P}}(1) \tag{67}$$

uniformly in $\mathbb{P}$.

If, in addition, Assumption B.5(i) holds, then:

$$\sup_{\theta\in\Theta_{osn}} \left|PQ_n(M(\theta)) - f\left(\sqrt{n}(\hat{\gamma}_n - \tau) - \sqrt{n}\gamma(\theta)\right)\right| = o_{\mathbb{P}}(1) \tag{68}$$

uniformly in $\mathbb{P}$.

Proof of Lemma F.7. To show (67), by Assumptions B.1(i) and B.2(i)(iii):

$$\begin{aligned} nL_n(\hat{\theta}) &= \sup_{\theta\in\Theta_{osn}} \left(\ell_n + \frac{n}{2}\|\hat{\gamma}_n - \tau\|^2 - \frac{1}{2}\|\sqrt{n}\gamma(\theta) - \sqrt{n}(\hat{\gamma}_n - \tau)\|^2 - f_{n,\perp}(\gamma_\perp(\theta))\right) + o_{\mathbb{P}}(1) \\ &= \ell_n + \frac{n}{2}\|\hat{\gamma}_n - \tau\|^2 - \inf_{\theta\in\Theta_{osn}} \frac{1}{2}\|\sqrt{n}\gamma(\theta) - \sqrt{n}(\hat{\gamma}_n - \tau)\|^2 + o_{\mathbb{P}}(1) \end{aligned} \tag{69}$$

uniformly in $\mathbb{P}$. But observe that by Assumption B.2(i)(ii), for any $\epsilon > 0$:

$$\begin{aligned} &\sup_{\mathbb{P}\in\mathbf{P}} \mathbb{P}\left(\inf_{\theta\in\Theta_{osn}} \|\sqrt{n}\gamma(\theta) - \sqrt{n}(\hat{\gamma}_n - \tau)\|^2 > \epsilon\right) \\ &\leq \sup_{\mathbb{P}\in\mathbf{P}} \mathbb{P}\left(\left\{\inf_{t\in(T-\sqrt{n}\tau)\cap B_{k_n}} \|t - \sqrt{n}(\hat{\gamma}_n - \tau)\|^2 > \epsilon\right\} \cap \left\{\|\hat{\gamma}_n - \tau\| < \frac{k_n}{\sqrt{n}}\right\}\right) \\ &\quad + \sup_{\mathbb{P}\in\mathbf{P}} \mathbb{P}\left(\|\hat{\gamma}_n - \tau\| \geq \frac{k_n}{\sqrt{n}}\right) \end{aligned}$$

where $\inf_{t\in(T-\sqrt{n}\tau)\cap B_{k_n}} \|t - \sqrt{n}(\hat{\gamma}_n - \tau)\|^2 = 0$ whenever $\|\sqrt{n}(\hat{\gamma}_n - \tau)\| < k_n$ (because $\sqrt{n}\hat{\gamma}_n \in T$). Notice $\|\sqrt{n}(\hat{\gamma}_n - \tau)\| = o_{\mathbb{P}}(k_n)$ uniformly in $\mathbb{P}$ by (66) and the condition $\|\mathbb{V}_n\| = O_{\mathbb{P}}(1)$ (uniformly in $\mathbb{P}$). This proves (67). Result (68) follows by Assumption B.5(i). ■

Proof of Lemma B.3. We only prove the case with singularity; the case without singularity follows similarly. By identical arguments to the proof of Lemma 4.2, it is enough to characterize the large-sample behavior of $R_n(z)$ defined in equation (45) uniformly in $\mathbb{P}$. By Lemma F.7 and Assumption B.2(i)–(iii), there exist a positive sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ independent of z with $\varepsilon_n = o(1)$ and a sequence of events $(\mathcal{A}_n)_{n\in\mathbb{N}} \subset \mathcal{F}$ with $\inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\mathcal{A}_n) = 1 - o(1)$ such that:

$$\sup_{\theta\in\Theta_{osn}} \left|Q_n(\theta) - \left(\|\sqrt{n}\gamma(\theta) - \sqrt{n}(\hat{\gamma}_n - \tau)\|^2 + 2f_{n,\perp}(\gamma_\perp(\theta))\right)\right| \leq \varepsilon_n$$

$$\sup_{\theta\in\Theta_{osn}} \left|nL_n(\theta) - \ell_n - \frac{n}{2}\|\hat{\gamma}_n - \tau\|^2 + \frac{1}{2}\|\sqrt{n}\gamma(\theta) - \sqrt{n}(\hat{\gamma}_n - \tau)\|^2 + f_{n,\perp}(\gamma_\perp(\theta))\right| \leq \varepsilon_n$$

both hold on $\mathcal{A}_n$ for all $\mathbb{P} \in \mathbf{P}$. Also note that for any $z \in \mathbb{R}$ and any singular $\mathbb{P} \in \mathbf{P}$, we have

$$\begin{aligned} &\left\{\theta \in \Theta_{osn} : \|\sqrt{n}\gamma(\theta) - \sqrt{n}(\hat{\gamma}_n - \tau)\|^2 + 2f_{n,\perp}(\gamma_\perp(\theta)) \leq z + \varepsilon_n\right\} \\ &\subseteq \left\{\theta \in \Theta_{osn} : \|\sqrt{n}\gamma(\theta) - \sqrt{n}(\hat{\gamma}_n - \tau)\|^2 \leq z + \varepsilon_n\right\} \end{aligned}$$

because $f_{n,\perp} \geq 0$. Therefore, on $\mathcal{A}_n$ we have:

$$R_n(z) \leq e^{2\varepsilon_n} \frac{\int_{\{\theta:\|\sqrt{n}\gamma(\theta)-\sqrt{n}(\hat{\gamma}_n-\tau)\|^2 \leq z+\varepsilon_n\}\cap\Theta_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma(\theta)-\sqrt{n}(\hat{\gamma}_n-\tau)\|^2 - f_{n,\perp}(\gamma_\perp(\theta))} \mathrm{d}\Pi(\theta)}{\int_{\Theta_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma(\theta)-\sqrt{n}(\hat{\gamma}_n-\tau)\|^2 - f_{n,\perp}(\gamma_\perp(\theta))} \mathrm{d}\Pi(\theta)}$$

uniformly in z, for all $\mathbb{P} \in \mathbf{P}$.

Define $\Gamma_{osn} = \{\gamma(\theta) : \theta \in \Theta_{osn}\}$ and $\Gamma_{\perp,osn} = \{\gamma_\perp(\theta) : \theta \in \Theta_{osn}\}$ (if $\mathbb{P}$ is singular). The condition $\sup_{\mathbb{P}\in\mathbf{P}} \sup_{\theta\in\Theta_{osn}} \|(\gamma(\theta), \gamma_\perp(\theta))\| \to 0$ in Assumption B.2(i) implies that for all n sufficiently large we have $\Gamma_{osn} \times \Gamma_{\perp,osn} \subset B^*_\delta$ for all $\mathbb{P} \in \mathbf{P}$. By similar arguments to the proof of Lemma 4.2, we use Assumption B.3(ii), a change of variables and Tonelli's theorem to obtain:

$$R_n(z) \leq e^{2\varepsilon_n}(1+\bar{\varepsilon}_n) \frac{\int_{(\{\gamma:\|\sqrt{n}\gamma-\sqrt{n}(\hat{\gamma}_n-\tau)\|^2 \leq z+\varepsilon_n)\cap\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma-\sqrt{n}(\hat{\gamma}_n-\tau)\|^2} \mathrm{d}\gamma}{\int_{\Gamma_{osn}} e^{-\frac{1}{2}\|\sqrt{n}\gamma-\sqrt{n}(\hat{\gamma}_n-\tau)\|^2} \mathrm{d}\gamma}$$

which holds uniformly in z for all $\mathbb{P} \in \mathbf{P}$ (on $\mathcal{A}_n$ with n sufficiently large) for some sequence $(\bar{\varepsilon}_n)_{n\in\mathbb{N}}$ with $\bar{\varepsilon}_n = o(1)$. A second change of variables with $\sqrt{n}\gamma - \sqrt{n}(\hat{\gamma}_n - \tau) \mapsto \kappa$ yields:

$$R_n(z) \leq e^{2\varepsilon_n}(1+\bar{\varepsilon}_n) \frac{\nu_{d^*}(\{\kappa : \|\kappa\|^2 \leq z+\varepsilon_n\} \cap (T_{osn} - \sqrt{n}(\hat{\gamma}_n - \tau)))}{\nu_{d^*}(T_{osn} - \sqrt{n}(\hat{\gamma}_n - \tau))}$$

where $T_{osn} = \{\sqrt{n}\gamma : \gamma \in \Gamma_{osn}\} = \{\sqrt{n}\gamma(\theta) : \theta \in \Theta_{osn}\}$.

Recall that $B_\delta \subset \mathbb{R}^{d^*}$ denotes a ball of radius δ centered at zero. To complete the proof, it is enough to show that:

$$\sup_z \left| \frac{\nu_{d^*}(B_{\sqrt{z+\varepsilon_n}} \cap (T_{osn} - \sqrt{n}(\hat{\gamma}_n - \tau)))}{\nu_{d^*}(T_{osn} - \sqrt{n}(\hat{\gamma}_n - \tau))} - \frac{\nu_{d^*}(B_{\sqrt{z}} \cap (T - \sqrt{n}\hat{\gamma}_n))}{\nu_{d^*}(T - \sqrt{n}\hat{\gamma}_n)} \right| = o_{\mathbb{P}}(1) \tag{70}$$

uniformly in $\mathbb{P}$. We split this into three parts. First note that

$$\begin{aligned}
&\sup_z \left| \frac{\nu_{d^*}(B_{\sqrt{z+\varepsilon_n}} \cap (T_{osn} - \sqrt{n}(\hat{\gamma}_n - \tau)))}{\nu_{d^*}(T_{osn} - \sqrt{n}(\hat{\gamma}_n - \tau))} - \frac{\nu_{d^*}(B_{\sqrt{z+\varepsilon_n}} \cap (T_{osn} \cap B_{k_n} - \sqrt{n}(\hat{\gamma}_n - \tau)))}{\nu_{d^*}(T_{osn} \cap B_{k_n} - \sqrt{n}(\hat{\gamma}_n - \tau))} \right| \\
&\leq 2\frac{\nu_{d^*}(((T_{osn} \setminus B_{k_n}) - \sqrt{n}(\hat{\gamma}_n - \tau)))}{\nu_{d^*}(T_{osn} - \sqrt{n}(\hat{\gamma}_n - \tau))} \\
&\leq 2\frac{\nu_{d^*}(B^c_{k_n} - \sqrt{n}(\hat{\gamma}_n - \tau))}{\nu_{d^*}(T_{osn} - \sqrt{n}(\hat{\gamma}_n - \tau))}
\end{aligned} \tag{71}$$

where the first inequality is by (49) and the second is by the inclusion $(T_{osn} \setminus B_{k_n}) \subseteq B^c_{k_n}$. Since $\|\sqrt{n}(\hat{\gamma}_n - \tau)\| \leq \|\mathbb{V}_n\|$ (by 66) where $\|\mathbb{V}_n\| = O_{\mathbb{P}}(1)$ uniformly in $\mathbb{P}$ and $\inf_{\mathbb{P}\in\mathbf{P}} k_n(\mathbb{P}) \to \infty$ and $d^* = d^*(\mathbb{P}) \leq \bar{d} < \infty$, we have

$$\nu_{d^*}(B^c_{k_n} - \sqrt{n}(\hat{\gamma}_n - \tau)) = o_{\mathbb{P}}(1)$$

uniformly in $\mathbb{P}$. Also notice that, by Assumption B.2(ii),

$$\frac{\nu_{d^*}(B_{\sqrt{z+\varepsilon_n}} \cap (T_{osn} \cap B_{k_n} - \sqrt{n}(\hat{\gamma}_n - \tau)))}{\nu_{d^*}(T_{osn} \cap B_{k_n} - \sqrt{n}(\hat{\gamma}_n - \tau))} = \frac{\nu_{d^*}(B_{\sqrt{z+\varepsilon_n}} \cap ((T - \sqrt{n}\tau) \cap B_{k_n} - \sqrt{n}(\hat{\gamma}_n - \tau)))}{\nu_{d^*}((T - \sqrt{n}\tau) \cap B_{k_n} - \sqrt{n}(\hat{\gamma}_n - \tau))}$$

where, by similar arguments to (71),

$$\begin{aligned}
&\sup_z \left| \frac{\nu_{d^*}(B_{\sqrt{z+\varepsilon_n}} \cap ((T - \sqrt{n}\tau) \cap B_{k_n} - \sqrt{n}(\hat{\gamma}_n - \tau)))}{\nu_{d^*}((T - \sqrt{n}\tau) \cap B_{k_n} - \sqrt{n}(\hat{\gamma}_n - \tau))} - \frac{\nu_{d^*}(B_{\sqrt{z+\varepsilon_n}} \cap (T - \sqrt{n}\hat{\gamma}_n)}{\nu_{d^*}(T - \sqrt{n}\hat{\gamma}_n)} \right| \\
&\leq 2 \frac{\nu_{d^*}(((T - \sqrt{n}\tau) \setminus B_{k_n}) - \sqrt{n}(\hat{\gamma}_n - \tau))}{\nu_{d^*}(T - \sqrt{n}\hat{\gamma}_n)} \qquad (72) \\
&\leq 2 \frac{\nu_{d^*}(B_{k_n}^c - \sqrt{n}(\hat{\gamma}_n - \tau))}{\nu_{d^*}(T - \sqrt{n}\hat{\gamma}_n)} . \qquad (73)
\end{aligned}$$

A sufficient condition for the right-hand side of display (73) to be $o_{\mathbb{P}}(1)$ (uniformly in $\mathbb{P}$) is that

$$1/\nu_{d^*}(T - \sqrt{n}\hat{\gamma}_n) = O_{\mathbb{P}}(1) \quad \text{(uniformly in } \mathbb{P}). \qquad (74)$$

But notice that $T - \sqrt{n}\hat{\gamma}_n = (T - \sqrt{n}\tau) - \sqrt{n}(\hat{\gamma}_n - \tau)$ where the $\sqrt{n}(\hat{\gamma}_n - \tau)$ are uniformly tight (by (66) and the condition $\|\mathbb{V}_n\| = O_{\mathbb{P}}(1)$ uniformly in $\mathbb{P}$) and $T - \sqrt{n}\tau \supseteq T$. We may therefore deduce by the condition $\inf_{\mathbb{P}\in\mathbf{P}} \nu_{d^*}(T) > 0$ in Assumption B.2(ii) that (74) holds, and so:

$$\sup_z \left| \frac{\nu_{d^*}(B_{\sqrt{z+\varepsilon_n}} \cap ((T - \sqrt{n}\tau) \cap B_{k_n} - \sqrt{n}(\hat{\gamma}_n - \tau)))}{\nu_{d^*}((T - \sqrt{n}\tau) \cap B_{k_n} - \sqrt{n}(\hat{\gamma}_n - \tau))} - \frac{\nu_{d^*}(B_{\sqrt{z+\varepsilon_n}} \cap (T - \sqrt{n}\hat{\gamma}_n)}{\nu_{d^*}(T - \sqrt{n}\hat{\gamma}_n)} \right| = o_{\mathbb{P}}(1)$$

uniformly in $\mathbb{P}$. To see that the right-hand side of (71) is $o_{\mathbb{P}}(1)$ (uniformly in $\mathbb{P}$), first note that:

$$\frac{1}{\nu_{d^*}(T_{osn} - \sqrt{n}(\hat{\gamma}_n - \tau))} = \frac{\frac{1}{\nu_{d^*}(T_{osn} \cap B_{k_n} - \sqrt{n}(\hat{\gamma}_n - \tau))}}{1 + \frac{\nu_{d^*}(T_{osn} \setminus B_{k_n} - \sqrt{n}(\hat{\gamma}_n - \tau))}{\nu_{d^*}(T_{osn} \cap B_{k_n} - \sqrt{n}(\hat{\gamma}_n - \tau))}}$$

where

$$\begin{aligned}
\frac{1}{\nu_{d^*}(T_{osn} \cap B_{k_n} - \sqrt{n}(\hat{\gamma}_n - \tau))} &= \frac{1}{\nu_{d^*}((T - \sqrt{n}\tau) \cap B_{k_n} - \sqrt{n}(\hat{\gamma}_n - \tau))} \\
&= \frac{1}{(1 - o_{\mathbb{P}}(1)) \times \nu_{d^*}(T - \sqrt{n}\tau))} = O_{\mathbb{P}}(1)
\end{aligned}$$

(uniformly in $\mathbb{P}$) by (74) and because the $o_{\mathbb{P}}(1)$ term holds uniformly in $\mathbb{P}$ by (72) and (73). It follows that $1/\nu_{d^*}(T_{osn} - \sqrt{n}(\hat{\gamma}_n - \tau)) = O_{\mathbb{P}}(1)$ (uniformly in $\mathbb{P}$) and so, by (71), we obtain:

$$\sup_z \left| \frac{\nu_{d^*}(B_{\sqrt{z+\varepsilon_n}} \cap (T_{osn} - \sqrt{n}(\hat{\gamma}_n - \tau)))}{\nu_{d^*}(T_{osn} - \sqrt{n}(\hat{\gamma}_n - \tau))} - \frac{\nu_{d^*}(B_{\sqrt{z+\varepsilon_n}} \cap (T_{osn} \cap B_{k_n} - \sqrt{n}(\hat{\gamma}_n - \tau)))}{\nu_{d^*}(T_{osn} \cap B_{k_n} - \sqrt{n}(\hat{\gamma}_n - \tau))} \right| = o_{\mathbb{P}}(1)$$

(uniformly in $\mathbb{P}$). To complete the proof of (70), it remains to show that

$$\sup_z \left|\nu_{d^*}(B_{\sqrt{z+\varepsilon_n}} \cap (T - \sqrt{n}\hat{\gamma}_n)) - \nu_{d^*}(B_{\sqrt{z}} \cap (T - \sqrt{n}\hat{\gamma}_n))\right| = o_{\mathbb{P}}(1)$$

holds uniformly in $\mathbb{P}$. But here we have:

$$\begin{aligned}
&\sup_z \left|\nu_{d^*}(B_{\sqrt{z+\varepsilon_n}} \cap (T - \sqrt{n}\hat{\gamma}_n)) - \nu_{d^*}(B_{\sqrt{z}} \cap (T - \sqrt{n}\hat{\gamma}_n))\right| \\
&\leq \sup_z \left|\nu_{d^*}(B_{\sqrt{z+\varepsilon_n}} \setminus B_{\sqrt{z}})\right| \\
&= \sup_z \left|F_{\chi^2_{d^*}}(z+\varepsilon_n) - F_{\chi^2_{d^*}}(z)\right| \to 0
\end{aligned}$$

by uniform equicontinuity of $\{F_{\chi^2_d} : d \leq \overline{d}\}$. ■

Proof of Theorem B.1. We first prove part (i) by verifying the conditions of Lemma B.1. We assume w.l.o.g. that $L_n(\hat{\theta}) = \sup_{\theta \in \Theta_{osn}} L_n(\theta) + o_{\mathbb{P}}(n^{-1})$ uniformly in $\mathbb{P}$. By display (67) in Lemma F.7 we have $\sup_{\theta \in \Theta_I(\mathbb{P})} Q_n(\theta) = \|\mathbf{T}(\mathbb{V}_n + \sqrt{n}\tau) - \sqrt{n}\tau\|^2 + o_{\mathbb{P}}(1)$ uniformly in $\mathbb{P}$. This verifies condition (i) with $W_n = \|\mathbf{T}(\mathbb{V}_n + \sqrt{n}\tau) - \sqrt{n}\tau\|^2$.

For condition (ii) let $\xi_{\alpha,\mathbb{P}}$ denote the α quantile of F_T under $\mathbb{P}$ and let $(\varepsilon_n)_{n\in\mathbb{N}}$ be a positive sequence with $\varepsilon_n = o(1)$. By the conditions $\|\mathbf{T}(\mathbb{V}_n + \sqrt{n}\tau) - \sqrt{n}\tau\|^2 \leq \|\mathbf{T}\mathbb{V}_n\|^2$ (almost surely) for each $\mathbb{P} \in \mathbf{P}$, $\sup_{\mathbb{P}\in\mathbf{P}} \sup_z |\mathbb{P}(\|\mathbf{T}\mathbb{V}_n\|^2 \leq z) - \mathbb{P}_Z(\|\mathbf{T}Z\|^2 \leq z)| = o(1)$ and the equicontinuity of $\{F_T : \mathbb{P} \in \mathbf{P}\}$ at their α quantiles, we have:

$$\begin{aligned}
\liminf_{n\to\infty} \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(W_n \leq \xi_{\alpha,\mathbb{P}} - \varepsilon_n) &\geq \liminf_{n\to\infty} \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\|\mathbf{T}\mathbb{V}_n\|^2 \leq \xi_{\alpha,\mathbb{P}} - \varepsilon_n) \\
&\geq \liminf_{n\to\infty} \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}_Z(\|\mathbf{T}Z\|^2 \leq \xi_{\alpha,\mathbb{P}} - \varepsilon_n) \\
&= \alpha.
\end{aligned}$$

By Condition B.4 it suffices to show that for each $\epsilon > 0$

$$\lim_{n\to\infty} \sup_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\xi_{\alpha,\mathbb{P}} - \xi^{post}_{n,\alpha} > \epsilon) = 0\,.$$

A sufficient condition is that

$$\lim_{n\to\infty} \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\Pi_n(\{\theta : Q_n(\theta) \leq \xi_{\alpha,\mathbb{P}} - \epsilon\}|\mathbf{X}_n) < \alpha) = 1\,.$$

By Lemma B.3 there exists a sequence of positive constants $(u_n)_{n\in\mathbb{N}}$ with $u_n = o(1)$ and a sequence of events $(\mathcal{A}_n)_{n\in\mathbb{N}}$ (possibly depending on $\mathbb{P}$) with $\inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\mathcal{A}_n) = 1 - o(1)$ such that:

$$\Pi_n\big(\{\theta : Q_n(\theta) \leq \xi_{\alpha,\mathbb{P}} - \epsilon\} \,\big|\, \mathbf{X}_n\big) \leq \mathbb{P}_{Z|\mathbf{X}_n}\left(\|Z\|^2 \leq \xi_{\alpha,\mathbb{P}} - \epsilon | Z \in T - \sqrt{n}\hat{\gamma}_n\right) + u_n$$

holds on $\mathcal{A}_n$ for each $\mathbb{P}$. But by Theorem 2 of Chen and Gao (2017) we also have:

$$\mathbb{P}_{Z|\mathbf{X}_n}\left(\|Z\|^2 \leq \xi_{\alpha,\mathbb{P}} - \epsilon | Z \in T - \sqrt{n}\hat{\gamma}_n\right) \leq F_T(\xi_{\alpha,\mathbb{P}} - \epsilon)$$

and hence

$$\Pi_n\big(\{\theta : Q_n(\theta) \leq \xi_{\alpha,\mathbb{P}} - \epsilon\} \,\big|\, \mathbf{X}_n\big) \leq F_T(\xi_{\alpha,\mathbb{P}} - \epsilon) + u_n$$

holds on $\mathcal{A}_n$ for each $\mathbb{P}$. Also note that by the equicontinuity of $\{F_T : \mathbb{P} \in \mathbf{P}\}$ at their α quantiles:

$$\limsup_{n\to\infty} \sup_{\mathbb{P}\in\mathbf{P}} F_T(\xi_{\alpha,\mathbb{P}} - \epsilon) + u_n < \alpha - \delta \tag{75}$$

for some $\delta > 0$.

We therefore have:

$$\begin{aligned}
&\lim_{n\to\infty} \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\Pi_n(\{\theta : Q_n(\theta) \leq \xi_{\alpha,\mathbb{P}} - \epsilon\}|\mathbf{X}_n) < \alpha) \\
&\geq \liminf_{n\to\infty} \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}\Big(\Big\{\Pi_n(\{\theta : Q_n(\theta) \leq \xi_{\alpha,\mathbb{P}} - \epsilon\}|\mathbf{X}_n) < \alpha\Big\} \cap \mathcal{A}_n\Big) \\
&\geq \liminf_{n\to\infty} \inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}\Big(\Big\{F_T(\xi_{\alpha,\mathbb{P}} - \epsilon) + u_n < \alpha\Big\} \cap \mathcal{A}_n\Big) \\
&\geq 1 - \limsup_{n\to\infty} \sup_{\mathbb{P}\in\mathbf{P}} 1\!\!1\{F_T(\xi_{\alpha,\mathbb{P}} - \epsilon) + u_n \geq \alpha\} - \limsup_{n\to\infty} \sup_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\mathcal{A}_n^c) \\
&= 1
\end{aligned}$$

where the final line is by (75) and definition of $\mathcal{A}_n$.

The proof of part (ii) is similar. ■

Proof of Lemma B.4. It suffices to characterize the large-sample behavior of $R_n(z)$ defined in (55) uniformly in $\mathbb{P}$. By Lemma F.7 and Assumption B.2(i)–(iii), there exist a positive sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ independent of z with $\varepsilon_n = o(1)$ and a sequence of events $(\mathcal{A}_n)_{n\in\mathbb{N}} \subset \mathcal{F}$ with $\inf_{\mathbb{P}\in\mathbf{P}} \mathbb{P}(\mathcal{A}_n) = 1 - o(1)$ such that:

$$\sup_{\theta\in\Theta_{osn}} \left|PQ_n(M(\theta)) - f(\sqrt{n}(\hat{\gamma}_n - \tau) - \sqrt{n}\gamma(\theta))\right| \leq \varepsilon_n$$

$$\sup_{\theta\in\Theta_{osn}} \left|nL_n(\theta) - \ell_n - \frac{n}{2}\|\hat{\gamma}_n - \tau\|^2 + \frac{1}{2}\|\sqrt{n}\gamma(\theta) - \sqrt{n}(\hat{\gamma}_n - \tau)\|^2 + f_{n,\perp}(\gamma_\perp(\theta))\right| \leq \varepsilon_n$$

both hold on $\mathcal{A}_n$ for all $\mathbb{P} \in \mathbf{P}$. By similar arguments to the proof of Lemma 4.3, wpa1 we obtain:

$$\begin{aligned}
&(1 - \bar{\varepsilon}_n)e^{-2\varepsilon_n} \frac{\nu_{d^*}((f^{-1}(z - \varepsilon_n)) \cap (\sqrt{n}(\hat{\gamma}_n - \tau) - T_{osn}))}{\nu_{d^*}(\sqrt{n}(\hat{\gamma}_n - \tau) - T_{osn})} \\
&\leq R_n(z) \leq (1 + \bar{\varepsilon}_n)e^{2\varepsilon_n} \frac{\nu_{d^*}((f^{-1}(z + \varepsilon_n)) \cap (\sqrt{n}(\hat{\gamma}_n - \tau) - T_{osn}))}{\nu_{d^*}(\sqrt{n}(\hat{\gamma}_n - \tau) - T_{osn})}
\end{aligned}$$

uniformly in z for all $\mathbb{P} \in \mathbf{P}$, for some positive sequence $(\bar{\varepsilon}_n)_{n\in\mathbb{N}}$ with $\bar{\varepsilon}_n = o(1)$. To complete the proof, it remains to show that:

$$\sup_{z\in I}\left|\frac{\nu_{d^*}((f^{-1}(z+\varepsilon_n))\cap(\sqrt{n}(\hat{\gamma}_n-\tau)-T_{osn}))}{\nu_{d^*}(\sqrt{n}(\hat{\gamma}_n-\tau)-T_{osn})}-\frac{\nu_{d^*}(f^{-1}(z)\cap(\sqrt{n}\hat{\gamma}_n-T))}{\nu_{d^*}(\sqrt{n}\hat{\gamma}_n-T)}\right| = o_{\mathbb{P}}(1)$$

uniformly in $\mathbb{P}$. This follows by the uniform continuity condition on I in the statement of the lemma, using similar arguments to the proofs of Lemmas 4.3 and B.3. ■

Proof of Theorem B.2. We verify the conditions of Lemma B.2. We assume w.l.o.g. that $L_n(\hat{\theta}) = \sup_{\theta\in\Theta_{osn}} L_n(\theta) + o_{\mathbb{P}}(n^{-1})$ uniformly in $\mathbb{P}$. By display (68) in Lemma F.7 we have $PQ_n(M_I) = f(\sqrt{n}(\hat{\gamma}_n-\tau)) + o_{\mathbb{P}}(1)$ uniformly in $\mathbb{P}$. This verifies condition (i) with $W_n = f(\sqrt{n}(\hat{\gamma}_n-\tau)) = f(\mathbf{T}(\mathbb{V}_n+\sqrt{n}\tau)-\sqrt{n}\tau)$.

For condition (ii) let $\xi_{\alpha,\mathbb{P}}$ denote the α quantile of $f(Z)$ under $\mathbb{P}$ and let $(\varepsilon_n)_{n\in\mathbb{N}}$ be a positive sequence with $\varepsilon_n = o(1)$. By Assumption B.5(ii), the condition $\sup_{\mathbb{P}\in\mathbf{P}}\sup_z|\mathbb{P}(f(\mathbb{V}_n)\leq z)-\mathbb{P}_Z(f(Z)\leq z)| = o(1)$, and equicontinuity of $f(Z)$ at thier α quantiles, we have:

$$\begin{aligned}\liminf_{n\to\infty}\inf_{\mathbb{P}\in\mathbf{P}}\mathbb{P}(W_n\leq\xi_{\alpha,\mathbb{P}}-\varepsilon_n) &\geq \liminf_{n\to\infty}\inf_{\mathbb{P}\in\mathbf{P}}\mathbb{P}(f(\mathbb{V}_n)\leq\xi_{\alpha,\mathbb{P}}-\varepsilon_n)\\ &\geq \liminf_{n\to\infty}\inf_{\mathbb{P}\in\mathbf{P}}\mathbb{P}_Z(f(Z)\leq\xi_{\alpha,\mathbb{P}}-\varepsilon_n)\\ &= \alpha\,.\end{aligned}$$

By condition B.6 it suffices to show that for each $\epsilon > 0$:

$$\lim_{n\to\infty}\sup_{\mathbb{P}\in\mathbf{P}}\mathbb{P}(\xi_{\alpha,\mathbb{P}}-\xi_{n,\alpha}^{post,p}>\epsilon) = 0\,.$$

A sufficient condition is that

$$\lim_{n\to\infty}\inf_{\mathbb{P}\in\mathbf{P}}\mathbb{P}(\Pi_n(\{\theta : PQ_n(M(\theta))\leq\xi_{\alpha,\mathbb{P}}-\epsilon\}|\mathbf{X}_n)<\alpha) = 1\,.$$

By Lemma B.4 there exists a sequence of positive constants $(u_n)_{n\in\mathbb{N}}$ with $u_n = o(1)$ and a sequence of events $(\mathcal{A}_n)_{n\in\mathbb{N}}$ (possibly depending on $\mathbb{P}$) with $\inf_{\mathbb{P}\in\mathbf{P}}\mathbb{P}(\mathcal{A}_n) = 1-o(1)$ such that:

$$\Pi_n\big(\{\theta : PQ_n(M(\theta))\leq\xi_{\alpha,\mathbb{P}}-\epsilon\}\,\big|\,\mathbf{X}_n\big)\leq\mathbb{P}_{Z|\mathbf{X}_n}(f(Z)\leq\xi_{\alpha,\mathbb{P}}-\epsilon|Z\in\sqrt{n}\hat{\gamma}_n-T)+u_n$$

holds on $\mathcal{A}_n$ for each $\mathbb{P}$. But by Assumption B.5(iii) we may deduce that

$$\Pi_n\big(\{\theta : PQ_n(M(\theta))\leq\xi_{\alpha,\mathbb{P}}-\epsilon\}\,\big|\,\mathbf{X}_n\big)\leq\mathbb{P}_Z(f(Z)\leq\xi_{\alpha,\mathbb{P}}-\epsilon)+u_n$$

holds on $\mathcal{A}_n$ for each $\mathbb{P}$. By equicontinuity of the distribution of $\{f(Z) : \mathbb{P} \in \mathbf{P}\}$ we have:

$$\limsup_{n\to\infty} \sup_{\mathbb{P}\in\mathbf{P}} \mathbb{P}_Z(f(Z) \le \xi_{\alpha,\mathbb{P}} - \epsilon) + u_n < \alpha - \delta$$

for some $\delta > 0$. The result now follows by the same arguments as the proof of Theorem B.1. ■

Proof of Lemma B.5. To simplify notation, let $D_{\theta;p} = \sqrt{\chi^2(p_\theta; p)}$. Define the *generalized score* of $\mathbb{P}_\theta$ with respect to $\mathbb{P}$ as $S_{\theta;p}(x) = g'_{\theta;p} e_x$ where

$$g_{\theta;p} = \frac{1}{D_{\theta;p}} \begin{bmatrix} \frac{p_\theta(1)-p(1)}{p(1)} \\ \vdots \\ \frac{p_\theta(k)-p(k)}{p(k)} \end{bmatrix}.$$

Note that $PS_{\theta;p} = 0$ and $P(S^2_{\theta;p}) = 1$. Also define $u_{\theta;p} = \mathbb{J}_p^{-1} g_{\theta;p}$ and notice that $u_{\theta;p}$ is a unit vector (i.e. $\|u_{\theta;p}\| = 1$). Therefore,

$$|S_{\theta;p}(x)| \le 1/(\min_{1\le j\le k} \sqrt{p(j)}) \tag{76}$$

for each θ and $\mathbb{P} \in \mathbf{P}$.

For any $p_\theta > 0$, a Taylor series expansion of $\log(u+1)$ about $u = 0$ yields

$$\begin{aligned} nL_n(p_\theta) - nL_n(p) &= n\mathbb{P}_n \log(D_{\theta;p} S_{\theta;p} + 1) \\ &= nD_{\theta;p}\mathbb{P}_n S_{\theta;p} - \frac{nD^2_{\theta;p}}{2}\mathbb{P}_n S^2_{\theta;p} + nD^2_{\theta;p}\mathbb{P}_n(S^2_{\theta;p} R(D_{\theta;p} S_{\theta;p})) \end{aligned} \tag{77}$$

where $R(u) \to 0$ as $u \to 0$.

By (76), we may choose $(a_n)_{n\in\mathbb{N}}$ be a positive sequence with $a_n \to \infty$ as $n \to \infty$ such that $a_n \sup_{\theta:p_\theta>0} \max_{1\le i\le n} |S_{\theta;p}(X_i)| = o_{\mathbb{P}}(\sqrt{n})$ (uniformly in $\mathbb{P}$). Then, for any $r_n \le a_n$:

$$\sup_{\theta\in\Theta_{osn}(\mathbb{P})} \max_{1\le i\le n} |D_{\theta;p} S_{\theta;p}(X_i)| = o_{\mathbb{P}}(1) \quad \text{(uniformly in } \mathbb{P}). \tag{78}$$

By the two-sided Chernoff bound, for any $\delta \in (0,1)$:

$$\sup_{\mathbb{P}\in\mathbf{P}} \mathbb{P}\Big(\max_{1\le j\le k} \Big| \frac{\mathbb{P}_n \mathbb{1}\{x = j\}}{p(j)} - 1 \Big| > \delta \Big) \le 2ke^{-n(\inf_{\mathbb{P}\in P} \min_{1\le j\le k} p(j))\frac{\delta^2}{3}} \to 0 \tag{79}$$

because $\sup_{\mathbb{P}\in\mathbf{P}} \max_{1\le j\le k}(1/p(j)) = o(n)$. It follows that $\mathbb{P}_n(\mathbb{J}_p e_x e'_x \mathbb{J}_p) = I + o_{\mathbb{P}}(1)$ uniformly in $\mathbb{P}$.

Also notice that $S^2_{\theta;p}(x) = u'_{\theta;p}\mathbb{J}_p e_x e'_x \mathbb{J}_p u_{\theta;p}$ where each $u_{\theta;p}$ is a unit vector. Therefore,

$$\sup_{\theta:p_\theta>0} |\mathbb{P}_n S^2_{\theta;p} - 1| = o_{\mathbb{P}}(1) \quad \text{(uniformly in } \mathbb{P}). \tag{80}$$

Substituting (78) and (80) into (77) yields:

$$nL_n(p_\theta) - nL_n(p) = nD_{\theta;p}\mathbb{P}_n S_{\theta;p} - \frac{nD^2_{\theta;p}}{2} + nD^2_{\theta;p} \times o_{\mathbb{P}}(1)$$

where the $o_{\mathbb{P}}(1)$ term holds uniformly for all θ with $p_\theta > 0$, uniformly for all $\mathbb{P} \in \mathbf{P}$. We may therefore choose a positive sequence $(b_n)_{n\in\mathbb{N}}$ with $b_n \to \infty$ slowly such that b_n^2 times the $o_{\mathbb{P}}(1)$ term is still $o_{\mathbb{P}}(1)$ uniformly in $\mathbb{P}$. Letting $r_n = (a_n \wedge b_n)$, we obtain

$$\sup_{\theta\in\Theta_{osn}(\mathbb{P})} \left| nL_n(p_\theta) - nL_n(p) - nD_{\theta;p}\mathbb{P}_n S_{\theta;p} + \frac{nD^2_{\theta;p}}{2} \right| = o_{\mathbb{P}}(1) \quad \text{(uniformly in } \mathbb{P})$$

where $nD_{\theta;p}\mathbb{P}_n S_{\theta;p} = \sqrt{n}D_{\theta;p}\mathbb{G}_n(S_{\theta;p}) = \sqrt{n}\tilde{\gamma}_{\theta;p}\mathbb{G}_n(\mathbb{J}_p e_x)$ and $D^2_{\theta;p} = \|\tilde{\gamma}_{\theta;p}\|^2$. ■

Proof of Proposition B.1. The quadratic expansion follows from Lemma B.5 and (34) and (35), which give $\|\tilde{\gamma}_{\theta;p}\|^2 = \tilde{\gamma}'_{\theta;p}\tilde{\gamma}_{\theta;p} = \tilde{\gamma}'_{\theta;p}V'_pV_p\tilde{\gamma}_{\theta;p} = \gamma(\theta)'\gamma(\theta)$ and $\tilde{\gamma}'_{\theta;p}\tilde{\mathbb{V}}_{n,p} = \tilde{\gamma}'_{\theta;p}V'_pV_p\tilde{\mathbb{V}}_{n,p} = \gamma(\theta)'\mathbb{V}_n$.

Uniform convergence in distribution is by Proposition A.5.2 of van der Vaart and Wellner (1996), since $\sup_{\mathbb{P}\in\mathbf{P}} \max_{1\le j\le k}(1/p(j)) = o(n)$ implies $\sup_{\mathbb{P}\in\mathbf{P}} |v'_{j,p}\mathbb{J}_p e_x| \le 1/(\min_{1\le j\le k}\sqrt{p(j)}) = o(n^{1/2})$. ■

Proof of Proposition B.2. The condition $\sup_{\mathbb{P}\in\mathbf{P}} \max_{1\le j\le k}(1/p(j)) = o(n/\log k)$ ensures that display (79) holds with $k = k(n) \to \infty$. The rest of the proof follows that of Proposition B.1. ■

Proof of Lemma B.6. For any $\mathbb{P} \in \mathbf{P}$, the mapping $p_\theta \mapsto V_p\tilde{\gamma}_{\theta;p}$ is a homeomorphism because $p > 0$ and V_p is an orthogonal matrix. Recall that the upper $k-1$ elements of $V_p\tilde{\gamma}_{\theta;p}$ is the vector $\gamma(\theta) = \gamma(\theta;\mathbb{P})$ and the remaining kth element is zero. Therefore, for each $\mathbb{P} \in \mathbf{P}$ the mapping $p_\theta \mapsto \gamma(\theta)$ is a homeomorphism. Since $\{p_\theta : \theta \in \Theta, p_\theta > 0\} = \text{int}(\Delta^{k-1})$ and $p \in \text{int}(\Delta^{k-1})$ for each $\mathbb{P} \in \mathbf{P}$, it follows that $\{\gamma(\theta) : \theta \in \Theta, p_\theta > 0\}$ contains a ball of radius $\epsilon = \epsilon(\mathbb{P}) > 0$ for each $\mathbb{P} \in \mathbf{P}$ (because homeomorphisms map interior points to interior points).

Recall that $\theta \in \Theta_{osn}(\mathbb{P})$ if and only if $\|\gamma(\theta)\| \le r_n/\sqrt{n}$ (because $\|\gamma(\theta)\|^2 = \|\tilde{\gamma}_{\theta;p}\|^2 = \chi^2(p_\theta;p)$). Let $\epsilon(\mathbb{P}) = \sup\{\epsilon > 0 : B_\epsilon \subseteq \{\gamma(\theta) : \theta \in \Theta, p_\theta > 0\}\}$. It suffices to show that $\inf_{\mathbb{P}\in\mathbf{P}} \sqrt{n}\epsilon(\mathbb{P}) \to \infty$ as $n \to \infty$. We can map back from any $\gamma \in \mathbb{R}^{k-1}$ by the inverse mapping $q_{\gamma;p}$ given by

$$q_{\gamma;p}(j) = p(j) + \sqrt{p(j)}[V_p^{-1}((\gamma'\ 0)')]_j$$

for $1 \leq j \leq k$, where $[V_p^{-1}((\gamma' \ 0)')]_j$ denotes the jth element of $[V_p^{-1}((\gamma' \ 0)')]$. An equivalent definition of $\epsilon(\mathbb{P})$ is $\inf\{\epsilon > 0 : q_p(\gamma) \notin \text{int}(\Delta^{k-1})$ for some $\gamma \in B_\epsilon\}$. As $p > 0$ and $\sum_{j=1}^k q_{\gamma;p}(j) = 1$ for each γ by construction, we therefore need to find the smallest $\epsilon > 0$ for which $q_{\gamma;p}(j) \leq 0$ for some j, for some $\gamma \in B_\epsilon$. This is equivalent to finding the smallest $\epsilon > 0$ for which

$$\frac{1}{\sqrt{p(j)}} \geq \frac{1}{[V_p^{-1}((\gamma' \ 0)')]_j} \tag{81}$$

for some j, for some $\gamma \in B_\epsilon$. The left-hand side is $o(\sqrt{n})$ uniformly for $1 \leq j \leq k$ and uniformly in $\mathbb{P}$ under the condition $\sup_{\mathbb{P}\in\mathbf{P}} \max_{1\leq j\leq k}(1/p(j)) = o(n)$. Also notice that, since the ℓ^2 norm dominates the maximum norm and V_p is an orthogonal matrix, we have

$$\frac{1}{[V_p^{-1}((\gamma' \ 0)')]_j} \geq \frac{1}{\|V_p^{-1}((\gamma' \ 0)')\|} = \frac{1}{\|\gamma\|} \geq \frac{1}{\epsilon} \tag{82}$$

It follows from (81) and (82) that $\sqrt{n}\inf_{\mathbb{P}\in\mathbf{P}} \epsilon(\mathbb{P}) \geq \frac{\sqrt{n}}{o(\sqrt{n})} \to \infty$ as $n \to \infty$, as required. ■

Proof of Lemma B.7. Condition (79) implies that

$$\begin{aligned}
&\sup_{\theta\in\Theta_{osn}(\mathbb{P})} \sup_{\mu\in M(\theta)} \left| \sup_{\eta\in H_\mu} nL_n(p_{\mu,\eta}) - \sup_{\eta\in H_\mu:(\mu,\eta)\in\Theta_{osn}(\mathbb{P})} nL_n(p_{\mu,\eta}) \right| \\
&= \sup_{\theta\in\Theta_{osn}(\mathbb{P})} \sup_{\mu\in M(\theta)} \left| \inf_{\eta\in H_\mu} nD_{KL}(p\|p_{\mu,\eta}) - \inf_{\eta\in H_\mu:(\mu,\eta)\in\Theta_{osn}(\mathbb{P})} nD_{KL}(p\|p_{\mu,\eta}) \right| (1 + o_{\mathbb{P}}(1))
\end{aligned}$$

where the $o_{\mathbb{P}}(1)$ term holds uniformly in $\mathbb{P}$ and $D_{KL}(p\|p_\theta) = \sum_{j=1}^k p(j)\log(p(j)/p_\theta(j))$. By a Taylor expansion of $-\log(u+1)$ about $u = 0$, it is straightforward to deduce that

$$\lim_{\epsilon\to 0} \sup_{\mathbb{P}\in\mathbf{P}} \sup_{\theta\in\Theta:\chi^2(p_\theta;p)\leq\epsilon} \left| \frac{D_{KL}(p\|p_\theta)}{\frac{1}{2}\chi^2(p_\theta;p)} - 1 \right| = o(1)\,. \tag{83}$$

In particular, for any $\theta \in \Theta_{osn}(\mathbb{P})$ and any $\mu \in M(\theta)$, we have

$$\inf_{\eta\in H_\mu} D_{KL}(p\|p_{\mu,\eta}) \leq \inf_{\eta\in H_\mu:(\mu,\eta)\in\Theta_{osn}(\mathbb{P})} D_{KL}(p\|p_{\mu,\eta}) \leq \frac{\chi^2(p_\theta;p)}{2}(1 + o(1)) \tag{84}$$

uniformly in $\mathbb{P}$. We want to show that an equivalence (83) holds uniformly over shrinking KL-divergence neighborhoods (rather χ^2-divergence neighborhoods). By similar arguments to Lemma 3.1 in Liu and Shao (2003), we may deduce that

$$\frac{1}{\chi^2(p_\theta;p)}|4h^2(p_\theta,p) - \chi^2(p_\theta;p)| \leq \frac{3}{D_{\theta;p}} \max_x |S_{\theta,p}(x)| h^2(p_\theta,p)$$

where again $D_{\theta;p} = \sqrt{\chi^2(p_\theta;p)}$. But, $h(p_\theta,p) \leq D_{\theta;p}$. Moreover, the proof of Proposition B.1 also shows that $|S_{\theta;p}| \leq 1/(\min_{1\leq j\leq k}\sqrt{p(j)})$ holds for each θ and each $\mathbb{P} \in \mathbf{P}$ so $\max_x |S_{\theta,p}(x)| = o(\sqrt{n})$

uniformly in $\mathbb{P}$. This, together with the fact that $h(p_\theta, p) \leq \sqrt{D_{KL}(p\|p_\theta)}$, yields

$$\frac{1}{\chi^2(p_\theta; p)}|4h^2(p_\theta, p) - \chi^2(p_\theta; p)| \leq o(\sqrt{n}) \times \sqrt{D_{KL}(p\|p_\theta)}$$

where the $o(\sqrt{n})$ term holds uniformly for $\theta \in \Theta$ and $\mathbb{P} \in \mathbf{P}$. Let $(a_n)_{n\in\mathbb{N}}$ be a positive sequence with $a_n \leq r_n$ and $a_n \to \infty$ sufficiently slowly that a_n times the $o(\sqrt{n})$ term in the above display is still $o(\sqrt{n})$ (uniformly in θ and $\mathbb{P}$). We then have

$$\sup_{\mathbb{P}\in\mathbf{P}} \sup_{\theta: D_{KL}(p\|p_\theta) \leq \frac{a_n}{\sqrt{n}}} \frac{1}{\chi^2(p_\theta; p)}|4h^2(p_\theta, p) - \chi^2(p_\theta; p)| = o(1)\,.$$

Since $h^2(p_\theta, p) \leq D_{KL}(p\|p_\theta)$, this implies that

$$\sup_{\mathbb{P}\in\mathbf{P}} \sup_{\theta: D_{KL}(p\|p_\theta) \leq \frac{a_n}{\sqrt{n}}} \chi^2(p_\theta; p) = o(1)$$

and so, by (83), we obtain

$$\sup_{\mathbb{P}\in\mathbf{P}} \sup_{\theta\in\Theta: D_{KL}(p\|p_\theta) \leq \frac{a_n}{\sqrt{n}}} \left| \frac{D_{KL}(p\|p_\theta)}{\frac{1}{2}\chi^2(p_\theta; p)} - 1 \right| = o(1)\,.$$

It now follows by (83) that

$$\begin{aligned}
&\sup_{\theta: \chi^2(p_\theta;p) \leq \frac{a_n}{n}} \sup_{\mu\in M(\theta)} \left| \inf_{\eta\in H_\mu} nD_{KL}(p\|p_{\mu,\eta}) - \inf_{\eta\in H_\mu: \chi^2(p_{(\mu,\eta)};p) \leq \frac{a_n}{n}} nD_{KL}(p\|p_{\mu,\eta}) \right| (1 + o_{\mathbb{P}}(1)) \\
&= \frac{n}{2} \sup_{\theta: \chi^2(p_\theta;p) \leq \frac{a_n}{n}} \sup_{\mu\in M(\theta)} \left| \inf_{\eta\in H_\mu} \chi^2(p_{\mu,\eta}; p) - \inf_{\eta\in H_\mu: \chi^2(p_{\mu,\eta};p) \leq \frac{a_n}{n}} \chi^2(p_{\mu,\eta}; p) \right| (1 + o_{\mathbb{P}}(1))
\end{aligned}$$

where $\inf_{\eta\in H_\mu} \chi^2(p_{\mu,\eta}; p) - \inf_{\eta\in H_\mu: \chi^2(p_{\mu,\eta};p) \leq \frac{a_n}{n}} \chi^2(p_{\mu,\eta}; p) = 0$ because $\chi^2(p_\theta; p) \leq \frac{a_n}{n}$ and $\mu \in M(\theta)$ implies that there exists an $\eta \in H(\mu)$ with $p_\theta = p_{\mu,\eta}$, so the constraint $\chi^2(p_{\mu,\eta}; p) \leq \frac{a_n}{n}$ is never violated for any $\mu \in M(\theta)$, for any such θ. The result follows by taking $r'_n = a_n$. ■

F.4 Proofs for Appendix D

Proof of Theorem D.1. We first derive the asymptotic distribution of $\sup_{\theta\in\Theta_I} Q_n(\theta)$ under $\mathrm{P}_{n,a}$. By similar arguments to the proof of Theorem 4.1, we have:

$$\sup_{\theta\in\Theta_I} Q_n(\theta) = \|\mathbb{V}_n\|^2 + o_{\mathrm{P}_{n,a}}(1) \overset{\mathrm{P}_{n,a}}{\rightsquigarrow} \chi^2_{d^*}(a'a)\,.$$

Identical arguments to the proof of Lemma 4.1 yield:

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

Therefore, $\xi^{mc}_{n,\alpha} = \chi^2_{d^*,\alpha} + o_{\mathrm{P}_{n,a}}(1)$ and we obtain:

$$\mathrm{P}_{n,a}(\Theta_I \subseteq \widehat{\Theta}_\alpha) = \Pr(\chi^2_{d^*}(a'a) \leq \chi^2_{d^*,\alpha}) + o(1)$$

as required. ■

Proof of Theorem D.2. By similar arguments to the proof of Theorem 4.3, we have:

$$PQ_n(M_I) = f(\mathbb{V}_n) + o_{\mathrm{P}_{n,a}}(1) \overset{\mathrm{P}_{n,a}}{\rightsquigarrow} f(Z + a)$$

where $Z \sim N(0, I_{d^*})$. Identical arguments to the proof of Lemma 4.3 yield:

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

for a neighborhood I of z_α. Therefore, $\xi^{mc,p}_{n,\alpha} = z_\alpha + o_{\mathrm{P}_{n,a}}(1)$ and we obtain:

$$\mathrm{P}_{n,a}(M_I \subseteq \widehat{M}_\alpha) = \mathbb{P}_Z(f(Z + a) \leq z_\alpha) + o(1)$$

as required. ■

F.5 Proofs for Appendix E

Proof of Lemma E.1. By equations (43) and (44) in the proof of Lemma 4.1, it suffices to characterize the large-sample behavior of:

$$R_n(z) := \frac{\int_{\{\theta : Q_n(\theta) \leq z\} \cap \Theta_{osn}} e^{-\frac{1}{2}Q_n(\theta)} \mathrm{d}\Pi(\theta)}{\int_{\Theta_{osn}} e^{-\frac{1}{2}Q_n(\theta)} \mathrm{d}\Pi(\theta)}\,.$$

By Assumption E.2(i), there exists a positive sequence $(\varepsilon_n)_{n \in \mathbb{N}}$ with $\varepsilon_n = o(1)$ such that: $(1 - \varepsilon_n)h(\gamma(\theta) - \hat{\gamma}_n) \leq \frac{a_n}{2} Q_n(\theta) \leq (1 + \varepsilon_n)h(\gamma(\theta) - \hat{\gamma}_n)$ holds uniformly over Θ_{osn}. Therefore:

$$\begin{aligned}
&\frac{\int_{\{\theta : 2a_n^{-1}(1+\varepsilon_n)h(\gamma(\theta)-\hat{\gamma}_n) \leq z\} \cap \Theta_{osn}} e^{-a_n^{-1}(1+\varepsilon_n)h(\gamma(\theta)-\hat{\gamma}_n)} \mathrm{d}\Pi(\theta)}{\int_{\Theta_{osn}} e^{-a_n^{-1}(1-\varepsilon_n)h(\gamma(\theta)-\hat{\gamma}_n)} \mathrm{d}\Pi(\theta)} \\
&\leq R_n(z) \leq \frac{\int_{\{\theta : 2a_n^{-1}(1-\varepsilon_n)h(\gamma(\theta)-\hat{\gamma}_n) \leq z\} \cap \Theta_{osn}} e^{-a_n^{-1}(1-\varepsilon_n)h(\gamma(\theta)-\hat{\gamma}_n)} \mathrm{d}\Pi(\theta)}{\int_{\Theta_{osn}} e^{-a_n^{-1}(1+\varepsilon_n)h(\gamma(\theta)-\hat{\gamma}_n)} \mathrm{d}\Pi(\theta)}\,.
\end{aligned}$$

By similar arguments to the proof of Lemma 4.1, under Assumption 4.3 there exists a positive sequence $(\bar{\varepsilon}_n)_{n\in\mathbb{N}}$ with $\bar{\varepsilon}_n = o(1)$ such that for all n sufficiently large we have:

$$(1-\bar{\varepsilon}_n)\frac{\int_{\{\gamma:2a_n^{-1}(1+\varepsilon_n)h(\gamma-\hat{\gamma}_n)\le z\}\cap\Gamma_{osn}} e^{-a_n^{-1}(1+\varepsilon_n)h(\gamma-\hat{\gamma}_n)}\mathrm{d}\gamma}{\int_{\Gamma_{osn}} e^{-a_n^{-1}(1-\varepsilon_n)h(\gamma-\hat{\gamma}_n)}\mathrm{d}\gamma}$$
$$\le R_n(z) \le (1+\bar{\varepsilon}_n)\frac{\int_{\{\gamma:2a_n^{-1}(1-\varepsilon_n)h(\gamma-\hat{\gamma}_n)\le z\}\cap\Gamma_{osn}} e^{-a_n^{-1}(1-\varepsilon_n)h(\gamma-\hat{\gamma}_n)}\mathrm{d}\gamma}{\int_{\Gamma_{osn}} e^{-a_n^{-1}(1+\varepsilon_n)h(\gamma-\hat{\gamma}_n)}\mathrm{d}\gamma}.$$

under the change of variables $\theta \mapsto \gamma(\theta)$, where $\Gamma_{osn} = \{\gamma(\theta) : \theta \in \Theta_{osn}\}$.

Assumption E.2(ii) implies that:

$$a_n^{-1}(1\pm\varepsilon_n)h(\gamma-\hat{\gamma}_n) = h\Big(a_n^{-r_1}(1\pm\varepsilon_n)^{r_1}(\gamma_1-\hat{\gamma}_{n,1}),\ldots,a_n^{-r_{d^*}}(1\pm\varepsilon_n)^{r_{d^*}}(\gamma_{d^*}-\hat{\gamma}_{n,d^*})\Big).$$

Using a change of variables:

$$\gamma \mapsto \kappa_\pm(\gamma) = \big(a_n^{-r_1}(1\pm\varepsilon_n)^{r_1}(\gamma_1-\hat{\gamma}_{n,1}),\ldots,a_n^{-r_{d^*}}(1\pm\varepsilon_n)^{r_{d^*}}(\gamma_{d^*}-\hat{\gamma}_{n,d^*})\big)$$

(with choice of sign as appropriate) and setting $r^* = r_1 + \ldots + r_{d^*}$, we obtain:

$$(1-\bar{\varepsilon}_n)\frac{(1-\varepsilon_n)^{r^*}}{(1+\varepsilon_n)^{r^*}}\frac{\int_{\{\kappa:2h(\kappa)\le z\}\cap K^+_{osn}} e^{-h(\kappa)}\mathrm{d}\kappa}{\int e^{-h(\kappa)}\mathrm{d}\kappa}$$
$$\le R_n(z) \le (1+\bar{\varepsilon}_n)\frac{(1+\varepsilon_n)^{r^*}}{(1-\varepsilon_n)^{r^*}}\frac{\int_{\{\kappa:2h(\kappa)\le z\}} e^{-h(\kappa)}\mathrm{d}\kappa}{\int_{K^+_{osn}} e^{-h(\kappa)}\mathrm{d}\kappa} \tag{85}$$

uniformly in z, where $K^+_{osn} = \{\kappa_+(\gamma) : \gamma \in \Gamma_{osn}\}$.

We can use a change of variables for $\kappa \mapsto t = 2h(\kappa)$ to obtain:

$$\int_{\{\kappa:h(\kappa)\le z/2\}} e^{-h(\kappa)}\mathrm{d}\kappa = 2^{-r^*}\mathrm{V}(S)\int_0^z e^{-t/2}t^{r^*-1}dt \qquad \int e^{-h(\kappa)}\mathrm{d}\kappa = 2^{-r^*}\mathrm{V}(S)\int_0^\infty e^{-t/2}t^{r^*-1}dt \tag{86}$$

where $\mathrm{V}(S)$ denotes the volume of the set $S = \{\kappa : h(\kappa) = 1\}$.

For the remaining integrals over K^+_{osn} we first fix any $\omega \in \Omega$ so that $K^+_{osn}(\omega)$ becomes a deterministic sequence of sets. Let $C_n(\omega) = K^+_{osn}(\omega)\cap B_{k_n}$. Assumption E.2(iii) gives $\mathbb{R}^{d^*}_+ = \overline{\cup_{n\ge1}C_n(\omega)}$ for almost every ω. Now clearly:

$$\int e^{-h(\kappa)}\mathrm{d}\kappa \ge \int_{K^+_{osn}(\omega)} e^{-h(\kappa)}\mathrm{d}\kappa \ge \int \mathbb{1}\{\kappa \in C_n(\omega)\}e^{-h(\kappa)}\mathrm{d}\kappa \to \int e^{-h(\kappa)}\,\mathrm{d}\kappa$$

(by dominated convergence) for almost every ω. Therefore:

$$\int_{K^+_{osn}} e^{-h(\kappa)} \mathrm{d}\kappa \to_p 2^{-r^*} \mathrm{V}(S) \int_0^\infty e^{-t/2} t^{r^*-1} dt \,. \tag{87}$$

We may similarly deduce that:

$$\sup_z \left| \int_{\{\kappa : h(\kappa) \leq 2z\} \cap K^+_{osn}} e^{-h(\kappa)} \mathrm{d}\kappa - 2^{-r^*} \mathrm{V}(S) \int_0^z e^{-t/2} t^{r^*-1} dt \right| \mathrm{x} o_p 0 \,. \tag{88}$$

The result follows by substituting (86), (87), and (88) into (85). ■

Proof of Theorem E.1. We verify the conditions of Lemma 2.1. Lemma E.1 shows that the posterior distribution of the QLR is asymptotically $F_\Gamma = \Gamma(r^*, 1/2)$, and hence $\xi^{post}_{n,\alpha} = z_\alpha + o_{\mathbb{P}}(1)$, where z_α denotes the α quantile of the F_Γ. By Assumption $\sup_{\theta \in \Theta_I} Q_n(\theta) \rightsquigarrow F_\Gamma$. Then:

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

where the final equality is by Assumption 4.4. ■