# ECONOMETRICA

## MONTE CARLO CONFIDENCE SETS FOR IDENTIFIED SETS

Xiaohong Chen
*Cowles Foundation for Research in Economics, Yale University*

Timothy M. Christensen
*Department of Economics, New York University*

Elie Tamer
*Department of Economics, Harvard University*

# MONTE CARLO CONFIDENCE SETS FOR IDENTIFIED SETS

XIAOHONG CHEN
Cowles Foundation for Research in Economics, Yale University

TIMOTHY M. CHRISTENSEN
Department of Economics, New York University

ELIE TAMER
Department of Economics, Harvard University

It is generally difficult to know whether the parameters in nonlinear econometric models are point-identified. We provide computationally attractive procedures to construct confidence sets (CSs) for identified sets of the full parameter vector and of subvectors in models defined through a likelihood or a vector of moment equalities or inequalities. The CSs are based on level sets of "optimal" criterion functions (such as likelihoods, optimally-weighted or continuously-updated GMM criterions). The level sets are constructed using cutoffs that are computed via Monte Carlo (MC) simulations from the quasi-posterior distribution of the criterion. 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 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 whose local tangent spaces are convex cones. Further, our MC CSs for identified sets of subvectors are shown to have *exact* asymptotic coverage in models with singularities. We provide local power properties and uniform validity of our CSs over classes of DGPs that include point- and partially-identified models. Finally, we present two simulation experiments and two empirical examples: an airline entry game and a model of trade flows.

KEYWORDS: Partial identification, likelihood, moment (in)equality restrictions, confidence sets, subvector, (profile) quasi-likelihood ratio, quasi-posterior, convex cone, sequential Monte Carlo.

## 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 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

https://doi.org/10.3982/ECTA14525

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 sets (CSs) for the identified set $\Theta_I$ of the full parameter vector $\theta \equiv (\mu, \eta) \in \Theta$,[1] and for the identified sets $M_I$ of subvectors $\mu$. 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 $\Theta_I$ and contour sets of the sample profile criterion are used as CSs for $M_I$. We compute critical values differently from those in the existing literature, however. In two of our proposed CS constructions, we estimate critical values using quantiles of the sample criterion (or profile criterion) that are simulated from a quasi-posterior distribution, which is formed by combining the sample criterion with a prior over the model parameter space $\Theta$.[2]

We propose three procedures for constructing various CSs. To cover the identified set $\Theta_I$, Procedure 1 draws a sample $\{\theta^1, \ldots, \theta^B\}$ from the quasi-posterior, computes the $\alpha$ *quantile of the sample criterion evaluated at the draws*, and then defines our CS $\widehat{\Theta}_\alpha$ for $\Theta_I$ as the contour set at the said $\alpha$ quantile. Simulating from a quasi-posterior is a well-researched and understood area in the literature on Bayesian computation (see, e.g., Liu (2004), Robert and Casella (2004)). Many Monte Carlo (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 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 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 $\alpha$ quantile of a chi-squared 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 likelihoods or sandwich quasi-likelihoods, GMMs with optimally-weighted or continuously-updated or GEL (generalized empirical likelihood) criterions.[3] For point-identified regular models, optimal criterions are those that satisfy a generalized information equality. However, we also allow for some non-regular (or non-standard) models, such as models in which the local tangent space is a convex cone, models with singularities, and models with parameter-dependent support. Our Procedure 1

---

[1]Following the literature, the identified set $\Theta_I$ is the argmax of a population criterion over the whole parameter space $\Theta$. A model is point-identified if $\Theta_I$ 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 $\Theta$. We refer to the distribution as a quasi-posterior because we accommodate non-likelihood based models such as GMM.

[3]Moment inequality models are special cases of moment equality models as one can add nuisance parameters to transform moment inequalities into moment equalities. Notice that, although moment inequality models are allowed, our criterion differs from the popular generalized moment selection (GMS) criterion for moment inequality models in Andrews and Soares (2010) and others. See Sections 3.1.1, 5.2.1, and 5.3.3.

and 2 CSs, $\widehat{\Theta}_\alpha$ and $\widehat{M}_\alpha$, are shown to have *exact* asymptotic coverage for $\Theta_I$ and $M_I$ and have non-trivial local power in point- or partially-identified regular models. They are also shown to be valid but possibly conservative in point- or partially-identified models whose local tangent spaces are convex cones (e.g., models with reduced-form parameters on the boundary). Our Procedure 1 and 2 CSs are further shown to be uniformly valid over DGPs that include both point- and partially-identified models. 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 very close to point-identified (see the missing data example). Our Procedure 3 CS for $M_I$ of a scalar subvector is proved to be theoretically slightly conservative in partially-identified models, but performs well in our simulations and empirical examples. Finally, all of our three procedures are asymptotically efficient in regular models that happen to be point-identified.

Procedure 1 and 2 CSs are MC based. To establish their frequentist 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 (misspecified) regular models and several important non-regular cases as mentioned above. These theorems establish that the (quasi-)posterior distributions of the QLR and profile QLR converge to their frequentist counterparts in regular models, and asymptotically stochastic dominate their frequentist counterparts in non-regular models where the local tangent spaces are convex cones. Section 4 and Appendix C present similar results in other non-regular cases, such as models with singularities and models with parameter-dependent support. As an illustration, we briefly mention some results for Procedure 1 here: Section 4 presents conditions under which both the sample QLR statistic and the (quasi-)posterior distribution of the QLR converge to a chi-squared distribution with unknown degree of freedom in partially-identified regular models.[4] Appendix C shows that both the QLR and the (quasi-)posterior of the QLR 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 CSs to have exact asymptotic coverage for $\Theta_I$. Section 4 presents similar results for subvector inference (Procedure 2).

We demonstrate the computational feasibility and good finite-sample coverage of our proposed methods in two simulation experiments: a missing data example and an entry game. 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 $\mu$ is by definition no shorter than that for $\mu$ 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 $\mu$ are negligible in this simulation design. Finally, our CS constructions are applied to two real data examples: an airline entry game and an empirical model of trade flows. The airline entry game example has $\dim(\theta) = 17$ partially-identified model parameters, including covariates-dependent equilibrium selection probability parameters. While the popular projection 95% CSs are $[0, 1]$ (totally uninformative) for several equilibrium

---

[4]In point-identified regular models, Wilks type results state that the degree of freedom equals $\dim(\theta)$ (the dimension of $\theta$) for QLR statistics. In partially-identified regular models, the degree of freedom is some $d^*$ that is typically less than or equal to $\dim(\theta)$. The true $d^*$ is difficult to infer from a complex model and is typically "unknown."

selection probability parameters, our Procedures 2 and 3 95% CSs show that the data are informative about some of them. The trade example has $\dim(\theta) = 46$ model parameters. Here, our Procedures 2 and 3 CSs are very similar to the conventional $t$-statistic based CSs, indicating that the model is still point-identified when we conduct a sensitivity analysis to some restrictive model assumptions.

*Literature Review.* Several papers have recently proposed Bayesian (or pseudo Bayesian) methods for constructing CSs for $\Theta_I$ that have correct frequentist coverage properties. See Section 3.3 in the 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 $\phi$, that is (globally) point-identified and $\sqrt{n}$-consistently and asymptotically normally estimable from the data, and is linked to the model structural parameter $\theta$ via a *known global* mapping. Second, a prior is placed on the reduced-form parameter $\phi$, and third, a classical Bernstein–von Mises theorem stating the asymptotic normality of the posterior distribution for $\phi$ 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 $\Theta_I$ 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 settings we accommodate.

We show that our procedures are valid irrespective of whether the model is separable or not. As we impose priors on the model parameter $\theta$ only, there is no need for the model to admit a known, finite-dimensional global reduced-form reparameterization. In contrast, the above-mentioned existing Bayesian methods require researchers to specify priors on the global reduced-form parameters that are supported on $\{\phi(\theta) : \theta \in \Theta\}$ (i.e., the set of reduced-form parameters consistent with the structural model). Specifying priors on $\phi$ consistent with this support could be difficult in some empirically relevant cases, such as the airline entry game application in Section 3.2. Although there is no need to find a global reduced-form reparameterization to implement our procedures, we show that a *local* reduced-form reparameterization exists for a broad class of partially-identified likelihood or moment-based models (see Section 5). This local reparameterization is only used 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. Moreover, our new Bernstein–von Mises (or Bayesian Wilks) 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 likelihood models with flat priors, our Procedure 1 CS for $\Theta_I$ is equivalent to highest posterior density (HPD) Bayesian credible set for $\theta$. Our theoretical results imply that HPD credible sets for $\theta$ give correct frequentist coverage in partially-identified regular models and conservative coverage in some non-regular 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.[5] In point-identified regular models satisfying a generalized information equality with $\sqrt{n}$-consistent and asymptotically normally estimable parameters $\theta \equiv (\mu, \eta)$, Chernozhukov and Hong (2003) (CH hereafter) proposed constructing CSs for scalar subvectors $\mu$ by taking the upper and lower quantiles of MCMC

---

[5]This is not a contradiction because their key Assumption 2 is violated in our setting; see Remark 4 below.

draws $\{\mu^1, \ldots, \mu^B\}$ where $(\mu^b, \eta^b) \equiv \theta^b$ for $b = 1, \ldots, B$. Our CS constructions for scalar subvectors are asymptotically equivalent to CH's CSs in such point-identified models, but they differ otherwise. Our CS constructions, which are based on quantiles of the *criterion* evaluated at the MC draws rather than of the raw parameter draws themselves, are valid irrespective of whether the model is point- or partially-identified.

There are several published works on frequentist CS constructions for $\Theta_I$; see, for example, CHT and Romano and Shaikh (2010) where subsampling based methods are used for general partially-identified models, Bugni (2010) and Armstrong (2014) where bootstrap methods are used for moment inequality models, and Beresteanu and Molinari (2008) where random set methods are used when $\Theta_I$ is strictly convex. For inference on identified sets $M_I$ of subvectors, the subsampling based papers of CHT and Romano and Shaikh (2010) deliver valid tests with a judicious choice of the subsample size for a profiled criterion. Both subsampling methods and our methods can handle general partially-identified likelihood and moment based models. Whereas subsampling methods may be sensitive to choice of subsample size, our methods typically have asymptotically correct coverage and computationally seem less demanding.[6]

The rest of the paper is organized as follows. Section 2 describes our CS constructions. Section 3 presents simulations and real data applications. Section 4 first establishes new Bernstein–von Mises (BvM, or Bayesian Wilks) results for the QLR and profile QLR in partially-identified regular models and some non-regular cases. 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 adaptive SMC algorithm and implementation details for examples in Section 3. Appendix B presents results on local power. Appendix C presents results for partially-identified models with parameter-dependent support and establishes BvM results for this setting. Appendix D shows that our CSs for $\Theta_I$ and $M_I$ are valid uniformly over a class of DGPs. It also derives a uniform quadratic expansion of QLR statistics in partially-identified discrete probability models with increasing supports, which is of independent interest. The Supplemental Material (Chen, Christensen, and Tamer (2018)) contains verification of the main regularity conditions for uniform validity in the missing data and moment inequality examples (Appendix E) and proofs of all the new results in the paper (Appendix F). Appendix G in the Online Replication Material presents additional simulation results and implementation details for examples in Section 3.

## 2. DESCRIPTION OF OUR PROCEDURES

In this section, we first describe our CS constructions for the identified set $\Theta_I$ of the full parameter vector $\theta \equiv (\mu, \eta)$ and for the identified set $M_I$ of subvectors $\mu$ in general likelihood and moment-based models. We then present a simple method for constructing CSs for $M_I$ for scalar subvectors in certain situations.

---

[6]There is a large literature on frequentist approach for *inference on the true parameter* $\theta \in \Theta_I$ or $\mu \in M_I$; see, for example, Imbens and Manski (2004), Rosen (2008), Andrews and Guggenberger (2009), Stoye (2009), Andrews and Soares (2010), Canay (2010), Andrews and Barwick (2012), Wan (2013), Romano, Shaikh, and Wolf (2014), Bugni, Canay, and Shi (2017), and Kaido, Molinari, and Stoye (2017), among many others. Most of these works focus on uniform size control for moment inequality models and the resulting CSs for $\mu$ are generally conservative under point identification. Recently, Andrews (2017) considered identification-robust inference on $\mu \in M_I$ that is efficient in strongly point-identified regular models.

Let $L : \Theta \to \mathbb{R}$ be a population criterion associated with a parametric likelihood or a moment-based model. We assume that $L$ is an upper semicontinuous function of $\theta$ with $\sup_{\theta \in \Theta} L(\theta) < \infty$. The *identified set* for $\theta$ is the set of maximizers of $L$:

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

The set $\Theta_I$ is our first object of interest. Write $\theta \equiv (\mu, \eta)$ where $\mu$ is the subvector of interest and $\eta$ is a nuisance parameter. Our second object of interest is the identified set for the subvector $\mu$:

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

Let $\mathbf{X}_n = (X_1, \ldots, X_n)$ denote a sample of i.i.d. or stationary and ergodic data of size $n$. Given the data $\mathbf{X}_n$, we seek to construct computationally attractive CSs that cover $\Theta_I$ or $M_I$ with a prespecified probability (in repeated samples) as the sample size $n$ gets large.

To describe our approach, let $L_n$ denote an (upper semicontinuous) sample criterion that is a jointly measurable function of the data $\mathbf{X}_n$ and $\theta$. This criterion $L_n$ can be a natural sample analogue of $L$ associated with a model. To establish frequentist coverage guarantees, we require $L$ and $L_n$ to be "optimal," for example, to satisfy a generalized information equality (in regular models). The following are typical examples of "optimal" criterions associated with likelihood or moment-based models.

*Parametric Likelihood Models*: Consider a parametric model $\mathcal{P} = \{ p_\theta : \theta \in \Theta \}$ where $p_\theta(\cdot)$ is a probability density with respect to a common $\sigma$-finite dominating measure $\lambda$. Let $D_{\mathrm{KL}}(p \| q) = \int p \log(p/q) \, d\lambda$ denote the Kullback–Leibler divergence, and $p_0 \in \mathcal{P}$ be the true density of the data. The identified set is $\Theta_I = \{ \theta \in \Theta : D_{\mathrm{KL}}(p_0 \| p_\theta) = 0 \}$. A natural population criterion is $L(\theta) = E[\log p_\theta(X)]$. We let $L_n$ be the sample log-likelihood function:

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

Sandwich criterions could be used for misspecified parametric likelihood models; see Remark 3.

*Moment-Based Models*: Consider a class of moment functions $\{ \rho(X, \theta) : \theta \in \Theta \}$ such that the identified set is $\Theta_I = \{ \theta \in \Theta : E[\rho(X, \theta)] = 0 \}$. See Section 5.2.1 for converting moment inequality models into moment equality models. A popular population criterion is the continuously-updated GMM criterion of Hansen, Heaton, and Yaron (1996) which for i.i.d. data is

$$L(\theta) = -\frac{1}{2} g(\theta)' \left( E\left[ \rho(X, \theta) \rho(X, \theta)' \right] - g(\theta) g(\theta)' \right)^{-} g(\theta), \tag{4}$$

where $g(\theta) = E[\rho(X, \theta)]$ and the superscript $^{-}$ denotes generalized inverse. We could let $L_n$ be the sample continuously-updated GMM criterion for i.i.d. data:

$$L_n(\theta) = -\frac{1}{2} g_n(\theta)' \left( \frac{1}{n} \sum_{i=1}^{n} \rho(X_i, \theta) \rho(X_i, \theta)' - g_n(\theta) g_n(\theta)' \right)^{-} g_n(\theta), \tag{5}$$

where $g_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \rho(X_i, \theta)$. Given an optimal weighting matrix $\widehat{W}$, we could also use a sample optimally-weighted GMM criterion:

$$L_n(\theta) = -\frac{1}{2} g_n(\theta)' \widehat{W} g_n(\theta). \tag{6}$$

Various generalized empirical likelihood (GEL) criterions could also be used.

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 $\Pi$ over $\Theta$, the quasi-posterior distribution $\Pi_n$ for $\theta$ given $\mathbf{X}_n$ is defined as

$$\mathrm{d}\Pi_n(\theta|\mathbf{X}_n) = \frac{e^{nL_n(\theta)} \mathrm{d}\Pi(\theta)}{\displaystyle\int_{\Theta} e^{nL_n(\theta)} \mathrm{d}\Pi(\theta)}. \tag{7}$$

Procedures 1 and 2 require drawing a sample $\{\theta^1, \ldots, \theta^B\}$ from the quasi-posterior $\Pi_n$. Any MC sampler could, in principle, be used. In this paper, we use an adaptive Sequential Monte Carlo (SMC) algorithm which is known to be well-suited to drawing from irregular, multi-modal distributions. The SMC algorithm is described in detail in Appendix A.

### 2.1. *Confidence Sets for the Identified Set $\Theta_I$*

Here we seek a $100\alpha\%$ CS $\widehat{\Theta}_\alpha$ for $\Theta_I$ that has asymptotically exact coverage, that is,

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

PROCEDURE 1—Confidence Sets for the Identified Set:
1. Draw a sample $\{\theta^1, \ldots, \theta^B\}$ from the quasi-posterior distribution $\Pi_n$ in (7).
2. Calculate the $(1 - \alpha)$ quantile of $\{L_n(\theta^1), \ldots, L_n(\theta^B)\}$; call it $\zeta_{n,\alpha}^{\mathrm{mc}}$.
3. Our $100\alpha\%$ confidence set for $\Theta_I$ is then

$$\widehat{\Theta}_\alpha = \left\{ \theta \in \Theta : L_n(\theta) \geq \zeta_{n,\alpha}^{\mathrm{mc}} \right\}. \tag{8}$$

Note that no optimization of $L_n(\theta)$ over $\Theta$ is required to construct $\widehat{\Theta}_\alpha$. The MC draws should concentrate around $\Theta_I$ if the MC algorithm used to sample from the quasi-posterior $\Pi_n$ has converged (i.e., the MC draws are a representative sample from the $\Pi_n$) and the sample size is large. Thus one can, in many cases, avoid an exhaustive grid search over the entire parameter space to compute $\widehat{\Theta}_\alpha$.

CHT considered inference on the set of minimizers of a nonnegative population criterion $Q : \Theta \to \mathbb{R}_+$ using a sample analogue $Q_n$ of $Q$. Let $\xi_{n,\alpha}$ denote a consistent estimator of the $\alpha$ quantile of $\sup_{\theta \in \Theta_I} Q_n(\theta)$. The $100\alpha\%$ CS for $\Theta_I$ they proposed is $\widehat{\Theta}_\alpha^{\mathrm{CHT}} = \{\theta \in \Theta : Q_n(\theta) \leq \xi_{n,\alpha}\}$. In the existing literature, subsampling (or sometimes bootstrap) based methods have been used to compute $\xi_{n,\alpha}$. Instead, our procedure replaces $\xi_{n,\alpha}$ with a cut-off based on MC simulations. The next remark provides an equivalent approach to Procedure 1 but 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 $L_n$, that is,

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

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

$$Q_n(\theta) = 2n\big[L_n(\hat{\theta}) - L_n(\theta)\big]. \tag{9}$$

Let $\xi_{n,\alpha}^{\mathrm{mc}}$ denote the $\alpha$ quantile of $\{Q_n(\theta^1), \ldots, Q_n(\theta^B)\}$. The confidence set

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

is equivalent to $\widehat{\Theta}_\alpha$ defined in (8) because $L_n(\theta) \geq \zeta_{n,\alpha}^{\mathrm{mc}}$ if and only if $Q_n(\theta) \leq \xi_{n,\alpha}^{\mathrm{mc}}$.

In Procedure 1 and Remark 1 above, the posterior-like quantity involves the use of a prior distribution $\Pi$ over $\Theta$. This prior is user chosen and typically is a 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 values (which is avoided by using a uniform prior whenever $\Theta$ is compact).

The next lemma presents high-level conditions under which *any* $100\alpha\%$ criterion-based CS for $\Theta_I$ 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 $\alpha$ quantile of $F_W$. Let $(v_n)_{n \in \mathbb{N}}$ be a sequence of random variables. We say that $v_n \leq o_{\mathbb{P}}(1)$ if $\lim_{n \to \infty} \mathbb{P}(v_n > \epsilon) = 0$ for each $\epsilon > 0$ and that $v_n \geq o_{\mathbb{P}}(1)$ if $\lim_{n \to \infty} \mathbb{P}(v_n < -\epsilon) = 0$ for each $\epsilon > 0$.

LEMMA 2.1: *Let* (i) $\sup_{\theta \in \Theta_I} Q_n(\theta) \rightsquigarrow W$ *where* $W$ *has a distribution* $F_W$ *that is continuous at* $w_\alpha$, *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*:

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

*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 $\Theta_I$ are shown to be valid by verifying parts (i) and (ii) with $w_{n,\alpha} = \xi_{n,\alpha}^{\mathrm{mc}}$. To verify part (ii), we shall establish a new Bernstein–von Mises (BvM, or 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*

For $M_I$ defined in likelihood and moment-based models, 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 $\Theta_I$ into one for $M_I$. The projection CS

$$\widehat{M}_\alpha^{\mathrm{proj}} = \big\{\mu : (\mu, \eta) \in \widehat{\Theta}_\alpha \text{ for some } \eta\big\} \tag{10}$$

is a valid $100\alpha\%$ CS for $M_I$ whenever $\widehat{\Theta}_\alpha$ is a valid $100\alpha\%$ CS for $\Theta_I$. As is well documented, $\widehat{M}_\alpha^{\mathrm{proj}}$ is typically conservative, and especially so when the dimension of $\mu$ is small relative to the dimension of $\theta$. Indeed, our simulations below indicate that $\widehat{M}_\alpha^{\mathrm{proj}}$ 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

$$PL_n(M_I) \equiv \inf_{\mu \in M_I} \sup_{\eta \in H_\mu} L_n(\mu, \eta). \tag{11}$$

Let

$$M(\theta^b) = \{\mu : (\mu, \eta) \in \Delta(\theta^b) \text{ for some } \eta\},$$

where $\Delta(\theta^b) = \{\theta \in \Theta : D_{\mathrm{KL}}(p_{\theta^b} \parallel p_\theta) = 0\}$ for likelihood models, and $\Delta(\theta^b) = \{\theta \in \Theta : E[\rho(X, \theta)] = E[\rho(X, \theta^b)]\}$ for moment-based models. For partially-identified likelihood models and separable moment-based models, the sets $M(\theta^b)$ (or $\Delta(\theta^b)$) can be calculated numerically or, in some cases (e.g., the missing data example), in closed form.[7] Appendix A describes how we compute $M(\theta^b)$ in the entry game simulation and in both empirical applications. Finally, we define the profile criterion for $M(\theta^b)$ as

$$PL_n(M(\theta^b)) \equiv \inf_{\mu \in M(\theta^b)} \sup_{\eta \in H_\mu} L_n(\mu, \eta). \tag{12}$$

PROCEDURE 2—Confidence Sets for Subvectors:
1. Draw a sample $\{\theta^1, \ldots, \theta^B\}$ from the quasi-posterior distribution $\Pi_n$ in (7).
2. Calculate the $(1 - \alpha)$ quantile of $\{PL_n(M(\theta^b)) : b = 1, \ldots, B\}$; call it $\zeta_{n,\alpha}^{\mathrm{mc},p}$.
3. Our $100\alpha\%$ confidence set for $M_I$ is then

$$\widehat{M}_\alpha = \left\{\mu \in M : \sup_{\eta \in H_\mu} L_n(\mu, \eta) \geq \zeta_{n,\alpha}^{\mathrm{mc},p}\right\}. \tag{13}$$

By forming $\widehat{M}_\alpha$ in terms of the profile criterion, one can, in many cases, avoid having to do an exhaustive grid search over $\Theta$. 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 (9). We define the profile QLR for the set $M(\theta^b)$ analogously as

$$PQ_n(M(\theta^b)) \equiv 2n[L_n(\hat{\theta}) - PL_n(M(\theta^b))] = \sup_{\mu \in M(\theta^b)} \inf_{\eta \in H_\mu} Q_n(\mu, \eta). \tag{14}$$

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

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

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

---

[7]Computing $M(\theta^b)$ in non-separable moment-based models would require replacing expectations in the definition of $\Delta(\theta^b)$ by their sample analogues. We leave rigorous treatment of this case to future research.

Our Procedure 2 is different from taking quantiles of the MC parameter draws. A percentile CS (denoted as $\widehat{M}_\alpha^{\mathrm{perc}}$) for a scalar subvector $\mu$ is computed by taking the upper and lower $100(1-\alpha)/2$ percentiles of $\{\mu^1, \ldots, \mu^B\}$ where $(\mu^b, \eta^b) \equiv \theta^b$ for $b = 1, \ldots, B$. For point-identified regular models with $\sqrt{n}$-consistent and asymptotically normally estimable $\theta$, this approach is known to be valid for correctly-specified likelihood models in the standard Bayesian literature, and its validity for optimally-weighted GMM and GEL has been established by Chernozhukov and Hong (2003). However, in partially-identified models, this percentile CS 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 $W$ has a distribution $F_W$ that is continuous at its $\alpha$ quantile $w_\alpha$, 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 $w_{n,\alpha} = \xi_{n,\alpha}^{\mathrm{mc},p}$. To verify part (ii), we shall derive a new BvM type result for the quasi-posterior distribution 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 performs very favorably in simulations.

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

$$\widehat{M}_\alpha^\chi = \left\{ \mu \in M : \inf_{\eta \in H_\mu} Q_n(\mu, \eta) \leq \chi_{1,\alpha}^2 \right\}, \tag{15}$$

where $Q_n$ is the QLR in (9) and $\chi_{1,\alpha}^2$ denotes the $\alpha$ quantile of the $\chi_1^2$ distribution.

Procedure 3 above is justified when the limit distribution of the profile QLR for $M_I$ is (first-order) stochastically dominated by the $\chi_1^2$ distribution (i.e., $F_W(z) \geq F_{\chi_1^2}(z)$ for all $z \geq 0$ in Lemma 2.2). Unlike $\widehat{M}_\alpha$, the CS $\widehat{M}_\alpha^\chi$ for $M_I$ is typically asymptotically conservative and is only valid for scalar subvectors (see Section 4.3). But $\widehat{M}_\alpha^\chi$ is much less conservative than projection CS $\widehat{M}_\alpha^{\mathrm{proj}}$ for scalar subvectors. And $\widehat{M}_\alpha^\chi$ is asymptotically exact in point-identified regular models. 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 CSs $\widehat{M}_\alpha$ and $\widehat{M}_\alpha^\chi$ (that remain valid under partial identification); see Section 3.2.

## 3. SIMULATION EVIDENCE AND EMPIRICAL APPLICATIONS

### 3.1. *Simulation Evidence*

In this subsection, we investigate the finite-sample behavior of our proposed CSs in two popular examples of partially-identified models: missing data and an entry game with correlated payoff shocks. In both simulation designs, we use random 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 $B = 10{,}000$ draws from an adaptive SMC algorithm. See Appendix A for description of the SMC algorithm and implementation details for the simulations and empirical applications.

#### 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 $D_i$ are binary. The parameter of interest is the true mean $\mu_0 = \mathbb{E}[Y_i]$. Without further assumptions, $\mu_0$ 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, $\theta$ 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 $\Theta$ for $\theta$ is defined as

$$\Theta = \big\{(\mu, \eta_1, \eta_2) \in [0, 1]^3 : 0 \le \mu - \eta_1(1 - \eta_2) \le \eta_2\big\}. \tag{16}$$

The identified set for $\theta$ is

$$\Theta_I = \big\{(\mu, \eta_1, \eta_2) \in \Theta : \tilde{\gamma}_{00} = 1 - \eta_2, \tilde{\gamma}_{11} = \mu - \eta_1(1 - \eta_2)\big\}. \tag{17}$$

Here, $\eta_2$ is point-identified but only an affine combination of $\mu$ and $\eta_1$ is identified. The identified set for $\mu = \mathbb{E}[Y_i]$ is

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

and the identified set for the nuisance parameter $\eta_1$ 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 $\Theta$. The likelihood function of $(D_i, Y_i D_i) = (d, yd)$ is

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

In Appendix G of the Supplemental Material, we present additional results for a likelihood criterion with a curved prior and a continuously-updated GMM criterion based
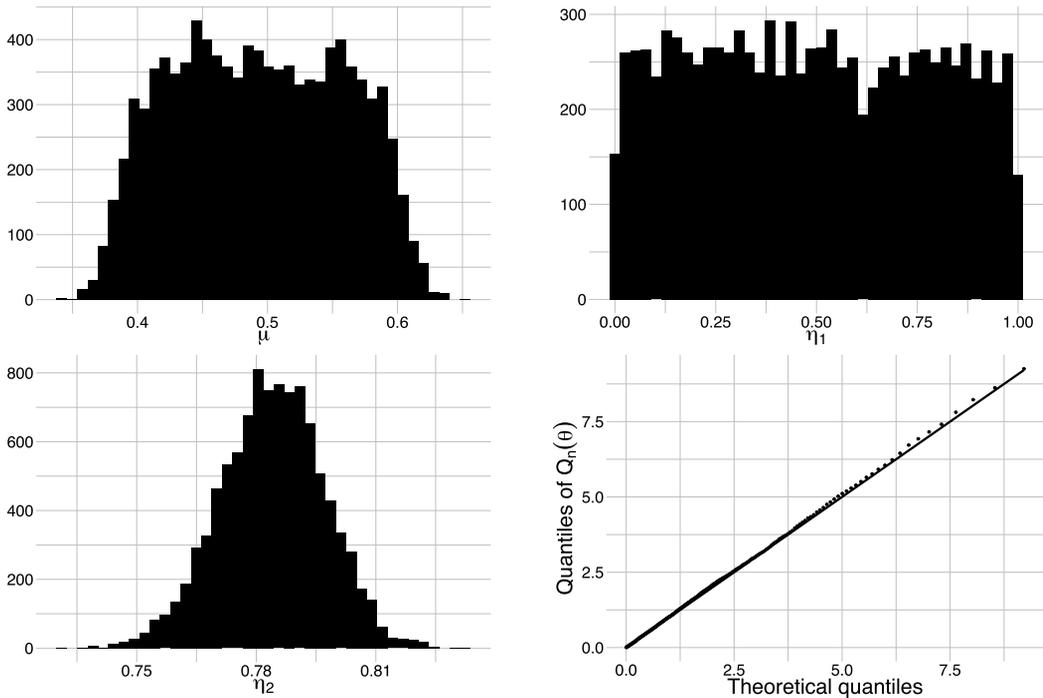
FIGURE 1.—Missing data example: histograms of the SMC draws for $\mu$ (top left), $\eta_1$ (top right), and $\eta_2$ (bottom left) and Q-Q plot of $Q_n(\theta)$ computed from the draws against $\chi_2^2$ quantiles (bottom right) for a sample of size $n = 1000$ with $\eta_2 = 0.8$. The identified sets for $\mu$ and $\eta_1$ are $[0.4, 0.6]$ and $[0, 1]$, respectively.

on the moments $E[\mathbb{1}\{D_i = 0\} - \tilde{\gamma}_{00}(\theta)] = 0$ and $E[\mathbb{1}\{(D_i, Y_iD_i) = (1, 1)\} - \tilde{\gamma}_{11}(\theta)] = 0$ 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.1. To illustrate sampling via the SMC algorithm and the resulting posterior of the QLR, Figure 1 displays histograms of the draws for $\mu$, $\eta_1$, and $\eta_2$ for one run of the SMC algorithm for a sample of size $n = 1000$ with $\eta_2 = 0.8$. Here $\mu$ is partially-identified with $M_I = [0.4, 0.6]$. The histograms in Figure 1 show that the draws for $\mu$ and $\eta_1$ are both approximately flat across their identified sets. In contrast, the draws for $\eta_2$, 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 $\chi_2^2$ distribution, as predicted by our theoretical results below (see Lemma 4.1).

*Confidence Sets for $\Theta_I$.* The top panel of Table I 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 probabilities are close even in samples as small as $n = 100$. When $\eta_2 = 1$, the CSs for $\Theta_I$ are conservative, as predicted by our Theorem 4.2 for models with singularities.

*Confidence Sets for $M_I$.* Table I also displays various CSs for the identified set $M_I$ for $\mu$. It clearly shows that the projection CS $\widehat{M}_\alpha^{\text{proj}}$ is very conservative. As the models with $c = 1, 2$ are close to point-identified, one might be tempted to report percentile CSs $\widehat{M}_\alpha^{\text{perc}}$

TABLE I

MISSING DATA EXAMPLE: AVERAGE COVERAGE PROBABILITIES FOR $\Theta_I$ AND $M_I$ AND AVERAGE LOWER AND UPPER BOUNDS OF CSS FOR $M_I$ ACROSS 5000 MC REPLICATIONS [a]

| | $\eta_2 = 1 - \frac{2}{\sqrt{n}}$ | | | $\eta_2 = 1 - \frac{1}{\sqrt{n}}$ | | | $\eta_2 = 1$ (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) for $\Theta_I$ | | | | | | | | | |
| 100 | — | 0.957 | 0.994 | 0.903 | 0.953 | 0.993 | 0.989 | 0.997 | 1.000 |
| 250 | — | 0.947 | 0.991 | 0.912 | 0.955 | 0.992 | 0.992 | 0.997 | 1.000 |
| 500 | — | 0.956 | 0.991 | 0.908 | 0.957 | 0.991 | 0.995 | 0.997 | 0.999 |
| 1000 | — | 0.958 | 0.992 | 0.911 | 0.958 | 0.994 | 0.997 | 0.999 | 1.000 |
| $\widehat{M}_\alpha$ (Procedure 2) for $M_I$ | | | | | | | | | |
| 100 | 0.920 [0.32, 0.68] | 0.969 [0.30, 0.70] | 0.997 [0.27, 0.73] | 0.918 [0.37, 0.63] | 0.964 [0.35, 0.65] | 0.994 [0.32, 0.68] | 0.911 [0.42, 0.59] | 0.958 [0.40, 0.60] | 0.990 [0.37, 0.63] |
| 250 | 0.917 [0.39, 0.61] | 0.961 [0.38, 0.62] | 0.992 [0.36, 0.64] | 0.920 [0.42, 0.58] | 0.963 [0.41, 0.59] | 0.991 [0.39, 0.61] | 0.915 [0.45, 0.55] | 0.959 [0.44, 0.56] | 0.991 [0.42, 0.58] |
| 500 | 0.914 [0.42, 0.58] | 0.961 [0.41, 0.59] | 0.993 [0.40, 0.60] | 0.914 [0.44, 0.56] | 0.958 [0.43, 0.57] | 0.992 [0.42, 0.58] | 0.916 [0.46, 0.54] | 0.959 [0.46, 0.54] | 0.990 [0.44, 0.56] |
| 1000 | 0.917 [0.44, 0.56] | 0.956 [0.44, 0.56] | 0.993 [0.43, 0.57] | 0.914 [0.46, 0.54] | 0.955 [0.45, 0.55] | 0.993 [0.44, 0.56] | 0.916 [0.47, 0.53] | 0.959 [0.47, 0.53] | 0.992 [0.46, 0.54] |
| $\widehat{M}_\alpha^{x}$ (Procedure 3) for $M_I$ | | | | | | | | | |
| 100 | 0.920 [0.32, 0.68] | 0.952 [0.31, 0.69] | 0.990 [0.28, 0.72] | 0.916 [0.37, 0.63] | 0.946 [0.36, 0.64] | 0.989 [0.33, 0.67] | 0.902 [0.42, 0.58] | 0.937 [0.41, 0.60] | 0.986 [0.38, 0.63] |
| 250 | 0.915 [0.39, 0.61] | 0.952 [0.38, 0.62] | 0.990 [0.36, 0.64] | 0.914 [0.42, 0.58] | 0.954 [0.41, 0.59] | 0.990 [0.39, 0.61] | 0.883 [0.45, 0.55] | 0.949 [0.44, 0.56] | 0.991 [0.42, 0.58] |
| 500 | 0.894 [0.42, 0.58] | 0.954 [0.41, 0.59] | 0.989 [0.40, 0.60] | 0.906 [0.44, 0.56] | 0.949 [0.44, 0.56] | 0.990 [0.42, 0.58] | 0.899 [0.46, 0.54] | 0.945 [0.46, 0.54] | 0.988 [0.44, 0.56] |
| 1000 | 0.909 [0.44, 0.56] | 0.950 [0.44, 0.56] | 0.993 [0.43, 0.57] | 0.904 [0.46, 0.54] | 0.954 [0.45, 0.55] | 0.989 [0.45, 0.55] | 0.906 [0.48, 0.52] | 0.946 [0.47, 0.53] | 0.991 [0.46, 0.54] |
| $\widehat{M}_\alpha^{\text{proj}}$ (Projection) for $M_I$ | | | | | | | | | |
| 100 | 0.972 [0.30, 0.70] | 0.990 [0.28, 0.71] | 0.999 [0.25, 0.75] | 0.969 [0.35, 0.65] | 0.989 [0.33, 0.67] | 0.998 [0.30, 0.70] | 0.989 [0.35, 0.63] | 0.997 [0.36, 0.64] | 1.000 [0.33, 0.67] |
| 250 | 0.971 [0.37, 0.63] | 0.986 [0.36, 0.64] | 0.998 [0.34, 0.66] | 0.976 [0.40, 0.60] | 0.988 [0.39, 0.61] | 0.998 [0.37, 0.63] | 0.992 [0.42, 0.58] | 0.997 [0.41, 0.59] | 1.000 [0.39, 0.61] |
| 500 | 0.972 [0.41, 0.59] | 0.985 [0.40, 0.60] | 0.999 [0.39, 0.61] | 0.972 [0.43, 0.57] | 0.989 [0.42, 0.58] | 0.999 [0.41, 0.59] | 0.995 [0.44, 0.56] | 0.997 [0.43, 0.57] | 0.999 [0.42, 0.58] |
| 1000 | 0.973 [0.44, 0.56] | 0.990 [0.43, 0.57] | 0.999 [0.42, 0.58] | 0.973 [0.45, 0.55] | 0.988 [0.45, 0.55] | 0.999 [0.44, 0.56] | 0.997 [0.45, 0.55] | 0.999 [0.45, 0.55] | 1.000 [0.44, 0.56] |
| $\widehat{M}_\alpha^{\text{perc}}$ (Percentile) for $M_I$ | | | | | | | | | |
| 100 | 0.416 [0.38, 0.62] | 0.676 [0.36, 0.64] | 0.945 [0.32, 0.68] | 0.661 [0.40, 0.59] | 0.822 [0.39, 0.61] | 0.963 [0.35, 0.65] | 0.896 [0.42, 0.58] | 0.946 [0.40, 0.60] | 0.989 [0.38, 0.63] |
| 250 | 0.402 [0.42, 0.58] | 0.669 [0.41, 0.59] | 0.917 [0.38, 0.62] | 0.662 [0.44, 0.56] | 0.822 [0.43, 0.57] | 0.960 [0.41, 0.59] | 0.899 [0.45, 0.55] | 0.950 [0.44, 0.56] | 0.990 [0.42, 0.58] |
| 500 | 0.400 [0.44, 0.56] | 0.652 [0.43, 0.57] | 0.914 [0.42, 0.58] | 0.652 [0.46, 0.54] | 0.812 [0.45, 0.55] | 0.955 [0.43, 0.57] | 0.903 [0.46, 0.54] | 0.953 [0.46, 0.54] | 0.988 [0.44, 0.56] |
| 1000 | 0.405 [0.46, 0.54] | 0.671 [0.45, 0.55] | 0.917 [0.44, 0.56] | 0.662 [0.47, 0.53] | 0.819 [0.46, 0.54] | 0.953 [0.45, 0.55] | 0.905 [0.47, 0.53] | 0.953 [0.47, 0.53] | 0.990 [0.46, 0.54] |
| GMS CSs for $\mu$ via moment inequalities | | | | | | | | | |
| 100 | 0.815 [0.34, 0.66] | 0.908 [0.32, 0.68] | 0.981 [0.29, 0.71] | 0.803 [0.39, 0.61] | 0.904 [0.37, 0.63] | 0.980 [0.34, 0.66] | 0.889 [0.42, 0.58] | 0.938 [0.40, 0.60] | 0.973 [0.39, 0.62] |
| 250 | 0.798 [0.40, 0.60] | 0.899 [0.39, 0.61] | 0.979 [0.36, 0.64] | 0.811 [0.43, 0.57] | 0.897 [0.42, 0.58] | 0.980 [0.40, 0.60] | 0.896 [0.45, 0.55] | 0.944 [0.44, 0.56] | 0.981 [0.42, 0.57] |
| 500 | 0.794 [0.43, 0.57] | 0.898 [0.42, 0.58] | 0.976 [0.40, 0.60] | 0.789 [0.45, 0.55] | 0.892 [0.44, 0.56] | 0.975 [0.43, 0.57] | 0.897 [0.46, 0.54] | 0.948 [0.46, 0.54] | 0.986 [0.45, 0.55] |
| 1000 | 0.802 [0.45, 0.55] | 0.900 [0.44, 0.56] | 0.978 [0.43, 0.57] | 0.812 [0.46, 0.54] | 0.900 [0.46, 0.54] | 0.978 [0.45, 0.55] | 0.898 [0.47, 0.53] | 0.949 [0.47, 0.53] | 0.990 [0.46, 0.54] |

[a] Procedures 1, 2, and 3, Projection, and Percentile are implemented using a likelihood criterion and flat prior.

for $M_I$, which take the upper and lower $100(1 - \alpha)/2$ quantiles of the draws for $\mu$.[8] Table I shows that $\widehat{M}_\alpha^{\mathrm{perc}}$ has correct coverage when $\mu$ is point-identified (i.e., $\eta_2 = 1$) but severely under-covers when $\mu$ is not point-identified (i.e., $\eta_2 < 1$).

In contrast, our Procedures 2 and 3 CSs for $M_I$ remain valid under partial identification. We show below (see Theorem 4.3) that the coverage probabilities of Procedure 2 CS $\widehat{M}_\alpha$ (for $M_I$) should be equal to their nominal values $\alpha$ when $n$ is large irrespective of whether the model is partially-identified (i.e., $\eta_2 < 1$) or point-identified (i.e., $\eta_2 = 1$). Table I shows that this is indeed the case, and that the coverage probabilities of Procedure 2 CS are close to their nominal level even for small values of $n$. In Section 5.3.1, we show that the asymptotic distribution of the profile QLR for $M_I$ is stochastically dominated by the $\chi_1^2$ distribution, verifying the validity of Procedure 3 in this design. Table I also presents results for Procedure 3 CS $\widehat{M}_\alpha^\chi$; the coverage results look remarkably close to their nominal values even for small sample sizes and for all values of $\eta_2$.

Finally, we compare the length of CSs for $M_I$ using our Procedures 2 and 3 with the length of CSs for the parameter $\mu$ constructed using the GMS procedure of Andrews and Soares (2010). We implement their procedure using the moment inequalities

$$E[\mu - Y_i D_i] \geq 0, \qquad E\big[Y_i D_i + (1 - D_i) - \mu\big] \geq 0, \tag{18}$$

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. GMS CSs are known to be asymptotically valid CSs for $\mu$ rather than for the set $M_I$, which is why their coverage probabilities for $M_I$ reported in Table I appear lower than nominal when $\eta_2 < 1$. Importantly, the average lower and upper bounds of our Procedures 2 and 3 CSs for $M_I$ are very close to those using GMS. On the other hand, the average lengths of projection CSs are larger (since they are conservative), and those of the percentile CSs are narrower (since they under-cover when $\eta_2 < 1$).

### 3.1.2. *Example 2: Entry Game*

We now consider the complete information entry game example described in Table II. We assume that $(\epsilon_1, \epsilon_2)$, observed by the players, are jointly normally distributed with variance 1 and correlation $\rho$, which we treat as an unknown parameter (some existing papers assume that $\rho$ is known to be zero). We assume that $\Delta_1$ and $\Delta_2$ 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 $\{(D_{00,i}, D_{10,i}, D_{01,i}, D_{11,i})\}_{i=1}^n$. The data provide 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 three non-redundant choice probabilities from which we need to learn about six parameters.

---

[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 $\mu$ using the particle weights in a manner similar to (30) in Appendix A.

TABLE II

PAYOFF MATRIX FOR THE BINARY ENTRY GAME[a]

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

[a]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 $\theta$ via

$$\tilde{\gamma}_{00}(\theta) := Q_\rho(\epsilon_1 \leq -\beta_1; \epsilon_2 \leq -\beta_2),$$

$$\tilde{\gamma}_{11}(\theta) := Q_\rho(\epsilon_1 \geq -\beta_1 - \Delta_1; \epsilon_2 \geq -\beta_2 - \Delta_2),$$

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

and $\tilde{\gamma}_{01}(\theta) = 1 - \tilde{\gamma}_{00}(\theta) - \tilde{\gamma}_{11}(\theta) - \tilde{\gamma}_{10}(\theta)$, where $Q_\rho$ denotes the joint probability distribution of $(\epsilon_1, \epsilon_2)$ indexed by the correlation parameter $\rho$. 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

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

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

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

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 $\Theta$ and implement the SMC algorithm as described in Appendix A.2. Figure 2 displays histograms of the marginal draws for $s$ for one run of the SMC algorithm with a sample of size $n = 1000$. As can be seen, the draws are reasonably flat across the identified set $[0, 1]$ for $s$. Figure 2 also shows that the quantiles of $Q_n(\theta)$ computed from the draws are very close to the $\chi_3^2$ quantiles, as predicted by our theoretical results below (see Lemma 4.1).

Table III reports average coverage probabilities and CSs for the various procedures across 1000 replications. We form CSs for $\Theta_I$ using Procedure 1, as well as CSs for the identified sets of scalar subvectors $\Delta_1$ and $\beta_1$ using Procedures 2 and 3.[9] Appendix A.2 provides details on computation of $M(\theta)$ for implementation of Procedure 2. We do not use the reduced-form reparameterization in terms of choice probabilities to compute $M(\theta)$. Coverage of $\widehat{\Theta}_\alpha$ for $\Theta_I$ is very good, even with the small sample size $n = 100$. Coverages of Procedures 2 and 3 CSs for the identified sets for $\Delta_1$ and $\beta_1$ are slightly conservative for the small sample size $n$, but close to nominal for $n = 1000$. As expected, projection CSs for identified sets for $\Delta_1$ and $\beta_1$ are valid but very conservative, whereas percentile-based CSs severely under-cover.

[9]As the parameterization is symmetric, the identified sets for $\Delta_2$ and $\beta_2$ are the same as for $\Delta_1$ and $\beta_1$ so we omit them. We also omit CSs for $\rho$ and $s$, whose identified sets are both $[0, 1]$.
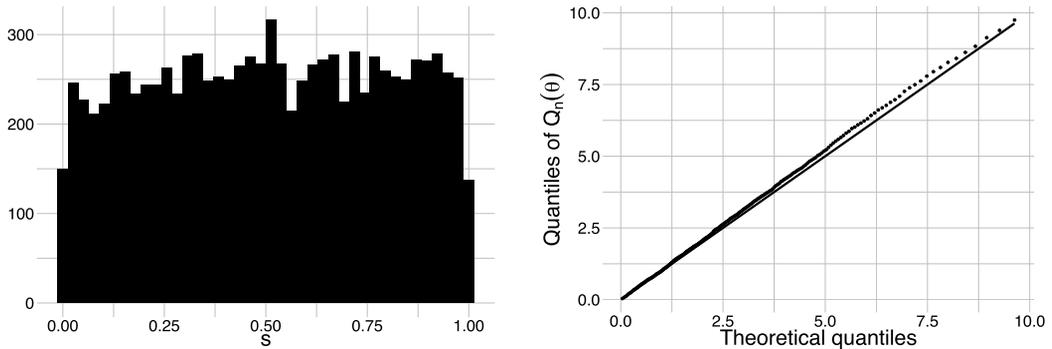
FIGURE 2.—Entry game example: histograms of the SMC draws for $s$ and Q-Q plot of $Q_n(\theta)$ computed from the draws against $\chi_3^2$ quantiles for a sample of size $n = 1000$.

## 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 U.S. 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 $\phi$ will be problematic in this example. This separable model has 24 non-redundant choice probabilities (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 specify a prior on $\phi$ and sample from the posterior for $\phi$. But, unless the posterior for $\phi$ is constrained to lie on $\{\phi(\theta) : \theta \in \Theta\}$ (i.e., the set of reduced-form choice probabilities consistent with the model, rather than the full 24-dimensional space), certain values of $\phi$ drawn from their posteriors for $\phi$ will not be consistent with the model.

The empirical trade example is a *non-separable* 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 $\theta$ 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 U.S. Airline Data*

This subsection estimates a version of the entry game that we study in Section 3.1.2 above. We use data from the second quarter of 2010's Airline Origin and Destination Survey (DB1B) to estimate a binary game, where the payoff for firm $i$ from entering market $m$ is

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

TABLE III

Entry Game Example: Average Coverage Probabilities for $\Theta_I$ and Identified Sets for $\Delta_1$ and $\beta_1$ Across 1000 MC Replications and Average Lower and Upper Bounds of CSs for Identified Sets for $\Delta_1$ and $\beta_1$ Across 1000 MC Replications Using a Likelihood Criterion and Flat Prior[a]

| | 0.90 | | 0.95 | | 0.99 | |
|---|---|---|---|---|---|---|
| | | CSs for the identified set $\Theta_I$ | | | | |
| | | $\widehat{\Theta}_\alpha$ (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 $\Delta_1$ | | | | |
| | | $\widehat{M}_\alpha$ (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 | $[-1.54, 0.00]$ |
| | | $\widehat{M}_\alpha^\chi$ (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 | $[-1.54, 0.00]$ |
| | | $\widehat{M}_\alpha^{\text{proj}}$ (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 | $[-1.59, 0.00]$ |
| | | $\widehat{M}_\alpha^{\text{perc}}$ (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 $\beta_1$ | | | | |
| | | $\widehat{M}_\alpha$ (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 | $[-0.15, 0.76]$ |
| | | $\widehat{M}_\alpha^\chi$ (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 | $[-0.14, 0.76]$ |
| | | $\widehat{M}_\alpha^{\text{proj}}$ (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 | $[-0.18, 0.80]$ |
| | | $\widehat{M}_\alpha^{\text{perc}}$ (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 | $[-0.09, 0.70]$ |

[a]The identified sets for $\Delta_1$ and $\beta_1$ are approximately $[-1.42, 0]$ and $[-0.05, 0.66]$.

where the $\Delta_i$ are assumed to be negative (as usually the case in entry models). The data set contains 7882 markets which are formally defined as trips between two airports irrespective of stopping. 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 $(0.16, 0.61, 0.07, 0.15)$ 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 serves the market.

The regressors are *market presence (MP)* and *market size (MS)*. *MP* 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 *MP* variable 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 *MP* for OA only enters OA's payoffs, so *MP* is both market- and airline-specific. The second regressor *MS* is defined as the population of the endpoints, so this variable is market-specific. We discretize both *MP* and *MS* into binary variables that take the value of 1 if the variable is higher than its median (in the data) value and zero otherwise. Let $P(y_{\mathrm{OA}}, y_{\mathrm{LC}} | MS, MP_{\mathrm{OA}}, MP_{\mathrm{LC}})$ denote the conditional choice probabilities. We therefore have four choice probabilities for every value of the conditioning variables (and there are eight values for these).[11] To use notation similar to that in Section 3.1.2, let OA be player 1 and firm LC be player 2. Denote $\beta_1(x_{m\mathrm{OA}}) := \beta_{\mathrm{OA}}^0 + \beta_{\mathrm{OA}}' x_{m\mathrm{OA}}$ and $\beta_2(x_{m\mathrm{LC}}) := \beta_{\mathrm{LC}}^0 + \beta_{\mathrm{LC}}' x_{m\mathrm{LC}}$ with $x_{m\mathrm{OA}} = (MS_m, MP_{m\mathrm{OA}})'$ and $x_{m\mathrm{LC}} = (MS_m, MP_{m\mathrm{LC}})'$. The likelihood for market $m$ depends on the (conditional) choice probabilities:

$$\tilde{\gamma}_{11}(\theta; x_m) := P(\epsilon_{1m} \geq -\beta_1(x_{m\mathrm{OA}}) - \Delta_{\mathrm{OA}}; \epsilon_{2m} \geq -\beta_2(x_{m\mathrm{LC}}) - \Delta_{\mathrm{LC}}),$$

$$\tilde{\gamma}_{00}(\theta; x_m) := P(\epsilon_{1m} \leq -\beta_1(x_{m\mathrm{OA}}); \epsilon_{2m} \leq -\beta_2(x_{m\mathrm{LC}})),$$

$$\tilde{\gamma}_{10}(\theta; x_m) := s(x_m) \times P(-\beta_1(x_{m\mathrm{OA}}) \leq \epsilon_{1m} \leq -\beta_1(x_{m\mathrm{OA}}) - \Delta_{\mathrm{OA}};$$
$$-\beta_2(x_{m\mathrm{LC}}) \leq \epsilon_{2m} \leq -\beta_2(x_{m\mathrm{LC}}) - \Delta_{\mathrm{LC}})$$
$$+ P(\epsilon_{1m} \geq -\beta_1(x_{m\mathrm{OA}}); \epsilon_{2m} \leq -\beta_2(x_{m\mathrm{LC}}))$$
$$+ P(\epsilon_{1m} \geq -\beta_1(x_{m\mathrm{OA}}) - \Delta_{\mathrm{OA}}; -\beta_2(x_{m\mathrm{LC}}) \leq \epsilon_{2m} \leq -\beta_2(x_{m\mathrm{LC}}) - \Delta_{\mathrm{LC}}).$$

Here $s(x_m)$ corresponds to the various *aggregate* equilibrium selection probabilities. Note that $s(\cdot)$ is a mapping from the support of $x_m$ to $[0, 1]$, 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 mechanism. Therefore, the full model has 17 parameters: four parameters per profit function (namely, $\Delta_i$, $\beta_i^0$, $\beta_i^{\mathrm{MS}}$, and $\beta_i^{\mathrm{MP}}$), the correlation $\rho$ between $\epsilon_{i1}$ and $\epsilon_{i2}$, and the eight 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. Both are popular versions of econometric models for a discrete game.

---

[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 $(MS, MP_{\mathrm{OA}}, MP_{\mathrm{LC}})$ takes 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).

TABLE IV

AIRLINE ENTRY GAME: 95% CSs FOR MODEL PARAMETERS COMPUTED VIA OUR PROCEDURES 2 AND 3 AS WELL AS VIA PROJECTION AND PERCENTILE METHODS[a]

| | Full model | | | | Fixed-s model | | | |
|---|---|---|---|---|---|---|---|---|
| | Procedure 2 | Procedure 3 | Projection | Percentile | Procedure 2 | Procedure 3 | Projection | Percentile |
| $\Delta_{OA}$ | [−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] |
| $\Delta_{LC}$ | [−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] |
| $\beta^0_{OA}$ | [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] |
| $\beta^{MS}_{OA}$ | [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] |
| $\beta^{MP}_{OA}$ | [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] |
| $\beta^0_{LC}$ | [−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] |
| $\beta^{MS}_{LC}$ | [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] |
| $\beta^{MP}_{LC}$ | [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] |
| $\rho$ | [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] | — | — | — | — |

[a]The full model is fully flexible in the equilibrium selection probabilities, while the fixed-s model restricts the equilibrium selection probability to be the same across markets with different regressor values.

We take a flat prior on $\Theta$ and implement the procedures using a likelihood criterion. We restrict the supports of $\Delta_i$ to $[-2, 0]$, $\beta_i$ to $[-1, 2]^3$, $\rho$ to $[0, 1]$, and $s(\ )$ to $[0, 1]$. We implement the procedure using the adaptive SMC algorithm as described in Appendix A.3 with $B = 10{,}000$ draws. Histograms of the SMC draws for the selection probabilities $s(\ )$ are presented in Figure 3; histograms of draws for the profit function parameters and $\rho$ are presented in Figures 8 and 9 in Appendix G.3 of the Supplemental Material. To illustrate convergence of the SMC algorithm, we present Q-Q plots of the profile QLR $PQ_n(M(\theta^b))$ for each parameter against the average quantiles across independent runs of the algorithm (see Figures 10 and 11 in Appendix G.3 of the Supplemental Material). The Q-Q plots show the profile QLR draws used to compute the critical values for Procedure 2 CSs align closely with draws obtained from independent runs of the algorithm. Table 8 in Appendix G.3 shows that recomputing Procedure 2 CSs using the independent runs of the SMC algorithm adjusts the endpoints by at most $10^{-3}$.

We construct CSs for each of the parameters using our Procedures 2 and 3, and compare these to projection-based CSs (projecting $\widehat{\Theta}_\alpha$ using our Procedure 1) and percentile CSs. (See Appendix A.3 for details on computation of $M(\theta)$ for implementation of Procedure 2.) The empirical findings are presented in Table IV. The results in Table IV show that Procedures 2 and 3 CSs are generally similar (though there are some differences, with Procedure 2 CSs 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 Procedure 2 CS is $[0.49, 0.92]$. As expected, percentile CSs are narrower than Procedure 2 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 in line 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 that 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 positive correlation in the payoff shocks, which could indicate missing profitability variables whereby firms enter a particularly profitable market regardless of competition.

Our Procedures 2 and 3 CSs for the selection probabilities are interesting (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 sets for $s_{010}$ and $s_{110}$ are $[0, 1]$ (via Procedure 2), indicating that the model (and data) has no information about these parameters, while the 95% CS for the identified set for $s_{111}$ is the narrow and informative interval $[0.94, 1.00]$ (via Procedure 2).

### 3.2.2. *An Empirical Model of Trade Flows*

In an influential paper, Helpman, Melitz, and Rubinstein (2008) examined 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
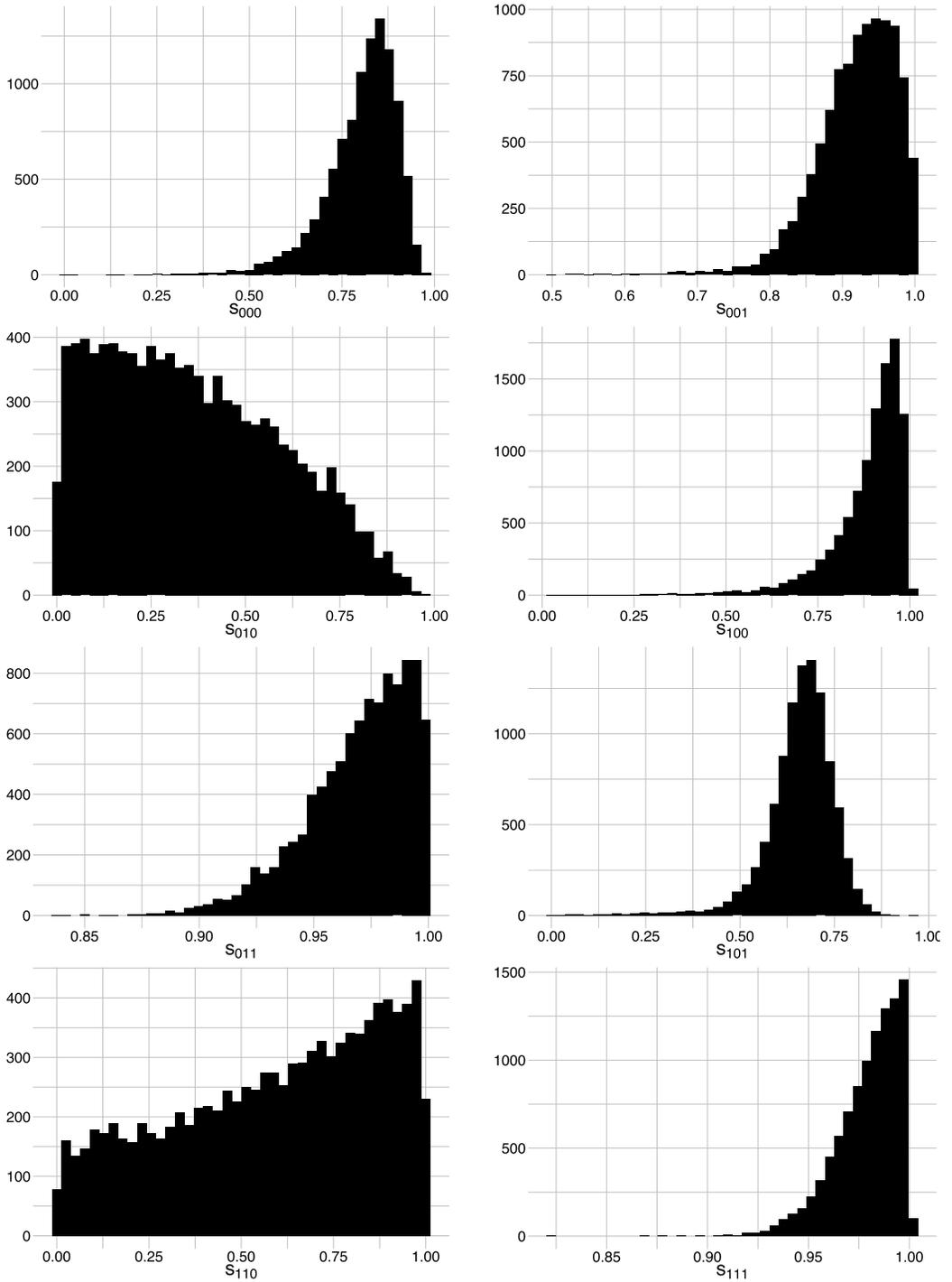
FIGURE 3.—Airline entry game: histograms of the SMC draws for selection probabilities $s_{000}$, $s_{001}$, $s_{010}$, $s_{100}$, $s_{011}$, $s_{101}$, $s_{110}$, and $s_{111}$ for the *full model*.

country $i$'s imports from country $j$, which is only observed if country $j$ exports to country $i$. Let $m_{ij} \equiv \log M_{ij}$. If a random draw for productivity from country $j$ to $i$ is sufficiently high, then $j$ will export to $i$. To model this, Helpman, Melitz, and Rubinstein (2008) introduced 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

$$m_{ij} = \begin{cases} \beta_0 + \lambda_j + \chi_i - \nu' f_{ij} + \delta z_{ij}^* + u_{ij}, & \text{if } z_{ij}^* > 0, \\ \text{not observed,} & \text{if } z_{ij}^* \leq 0, \end{cases}$$

$$z_{ij}^* = \beta_0^* + \lambda_j^* + \chi_i^* - \nu^{*'} f_{ij} + \eta_{ij}^*,$$

in which $\lambda_j$, $\chi_i$, $\lambda_j^*$, and $\chi_i^*$ are exporting and importing continent fixed effects, $f_{ij}$ is a vector of observable trade frictions between $i$ and $j$, and $u_{ij}$ and $\eta_{ij}^*$ are error terms described below. Exclusion restrictions can be imposed by setting at least one of the elements of $\nu$ equal to zero.

There are three differences between our empirical model and that of Helpman, Melitz, and Rubinstein (2008). First, we let $z_{ij}^*$ enter the outcome equation linearly instead of nonlinearly.[12] Second, we use continent fixed effects instead of country fixed effects. This reduces the number of parameters from over 400 to 46. Third, we allow for heteroscedasticity 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 (homoscedasticity) without worrying about loss of point identification.

To allow for heteroscedasticity, we suppose that the distribution of $(u_{ij}, \eta_{ij}^*)$ conditional on observables 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

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

We estimate the model from data on 24,649 country pairs in the selection equation and 11,156 in the outcome equation using the same data from 1986 as in Helpman, Melitz, and Rubinstein (2008). We also impose the exclusion restriction that the coefficient in $\nu$ 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 $\nu$ in the outcome equation as these are the most important. We estimate the model first by maximum likelihood under homoscedasticity and report conventional ML estimates for $\nu$ together with 95% CSs based on inverting $t$-statistics. We then re-estimate the model under heteroscedasticity and report conventional ML estimates together with CSs based on inverting $t$-statistics, percentile CS, and our Procedures 2 and 3 CSs. To implement our Procedure 2 and percentile CSs, we use the adaptive SMC algorithm (in Appendix A.4) with $B = 10,000$ draws.

---

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

The results are presented in Table V.[13] Overall, the CSs based on different methods are similar under the heteroscedastic specification, which suggests that partial identification may not be an issue even allowing for heteroscedasticity. Table V does show that the model is sensitive to the presence of heteroscedasticity. Under heteroscedasticity, the magnitudes of coefficients of the trade friction variables are generally smaller than under homoscedasticity, but of the same sign. The exception is the legal variable, whose coefficient is negative under heteroscedasticity but positive under homoscedasticity. We also notice some difference in our results under heteroscedasticity relative to those of Helpman, Melitz, and Rubinstein (2008) who assumed homoscedastic errors. For instance, they documented strong positive effects of common legal systems and currency unions, and a negative effect of landlocked status on trade flows, whereas we find much weaker evidence for common legal systems and currency unions, and a positive effect of landlocked status on trade flows.

## 4. LARGE-SAMPLE PROPERTIES

This section provides 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 $\Theta_I$ and $M_I$. The main new theoretical contributions are the derivations of the large-sample (quasi)-posterior distributions of the QLR for $\Theta_I$ and of the profile QLR for $M_I$ under loss of identifiability.

### 4.1. *Coverage Properties of $\widehat{\Theta}_\alpha$ for $\Theta_I$*

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

ASSUMPTION 4.1—Posterior Contraction:
  (i) $L_n(\hat{\theta}) = \sup_{\theta \in \Theta_{\mathrm{osn}}} L_n(\theta) + o_{\mathbb{P}}(n^{-1})$, with $(\Theta_{\mathrm{osn}})_{n \in \mathbb{N}}$ a sequence of local neighborhoods of $\Theta_I$;
  (ii) $\Pi_n(\Theta_{\mathrm{osn}}^c | \mathbf{X}_n) = o_{\mathbb{P}}(1)$, where $\Theta_{\mathrm{osn}}^c = \Theta \setminus \Theta_{\mathrm{osn}}$.

We presume the existence of a fixed neighborhood $\Theta_I^N$ of $\Theta_I$ (with $\Theta_{\mathrm{osn}} \subset \Theta_I^N$ for all $n$ sufficiently large) upon which there exists a *local* reduced-form reparameterization $\theta \mapsto \gamma(\theta)$ from $\Theta_I^N$ 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 $\theta$ in a fixed neighborhood of $\Theta_I$. The restriction that $d^*$ is finite and does not vary with $\gamma$ near zero might fail to hold in some models. To accommodate situations in which the true reduced-form parameter value $\gamma_0 = 0$ may be "on the boundary" of $\Gamma$, a relevant case in applications, we assume that the sets $T_{\mathrm{osn}} \equiv \{\sqrt{n}\gamma(\theta) : \theta \in \Theta_{\mathrm{osn}}\}$ cover[14] a *closed convex cone* $T \subseteq \mathbb{R}^{d^*}$ that has a positive volume. We note that this is trivially satisfied with $T = \mathbb{R}^{d^*}$ whenever each $T_{\mathrm{osn}}$ contains a ball of radius $k_n \to \infty$ centered at the origin. A similar approach was 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 $T$.

---

[13]Note that the friction variables enter negatively in the outcome equation. A positive coefficient of distance means 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.

[14]We say that a sequence of sets $A_n \subseteq \mathbb{R}^{d^*}$ *covers* a set $A \subseteq \mathbb{R}^{d^*}$ if there is a sequence of closed balls $B_{k_n}$ of radius $k_n \to \infty$ centered at the origin such that $A_n \cap B_{k_n} = A \cap B_{k_n}$ wpa1.

TABLE V

MAXIMUM LIKELIHOOD ESTIMATES OF THE COEFFICIENTS $\nu$ OF THE TRADE FRICTION VARIABLES IN THE OUTCOME EQUATION (MLE) TOGETHER WITH THEIR 95% CONFIDENCE SETS BASED ON INVERTING $t$-STATISTICS ($t$-STAT CI)[a]

| Variable | Homoscedastic | | Heteroscedastic | | | | |
|---|---|---|---|---|---|---|---|
| | MLE | $t$-stat CI | MLE | $t$-stat CI | Procedure 2 | Procedure 3 | Percentile |
| Distance | 2.352 | [−0.120, 4.823] | 0.314 | [0.229, 0.399] | [0.216, 0.749] | [0.242, 0.509] | [0.207, 0.397] |
| Border | −5.191 | [−9.084, −1.298] | −2.265 | [−2.651, −1.878] | [−2.651, −1.859] | [−2.611, −1.898] | [−2.618, −1.816] |
| Island | −1.302 | [−2.563, −0.041] | −0.728 | [−1.016, −0.441] | [−1.060, −0.308] | [−1.060, −0.308] | [−0.983, −0.397] |
| Landlock | −7.275 | [−14.484, −0.065] | −1.369 | [−1.849, −0.889] | [−2.194, −0.914] | [−2.194, −0.890] | [−1.801, −0.954] |
| Legal | 0.358 | [−0.377, 1.093] | −0.122 | [−0.248, 0.004] | [−0.254, 0.004] | [−0.242, −0.009] | [−0.248, 0.011] |
| Language | −4.098 | [−8.909, 0.713] | −0.095 | [−0.246, 0.057] | [−0.868, 0.049] | [−0.868, 0.026] | [−0.237, 0.067] |
| Colonial | −17.378 | [−35.169, 0.413] | −2.822 | [−3.249, −2.395] | [−4.980, −2.373] | [−4.980, −2.461] | [−3.231, −2.298] |
| Currency | −1.550 | [−4.088, 0.988] | −0.631 | [−1.282, 0.020] | [−1.315, 0.020] | [−1.282, −0.013] | [−1.274, 0.062] |
| FTA | −19.540 | [−40.672, 1.592] | −2.151 | [−2.686, −1.616] | [−2.686, −1.589] | [−2.631, −1.616] | [−2.68, −1.577] |

[a] Also shown are 95% CSs computed by our Procedures 2 and 3 as well as via Percentile methods.

ASSUMPTION 4.2—Local Quadratic Approximation: *There exist sequences of random variables $\ell_n$ and $\mathbb{R}^{d^*}$-valued random vectors $\hat{\gamma}_n$ (both measurable in $\mathbf{X}_n$) such that, as $n \to \infty$,*

$$\sup_{\theta \in \Theta_{\text{osn}}} \left| nL_n(\theta) - \left( \ell_n + \frac{1}{2} \| \sqrt{n} \hat{\gamma}_n \|^2 - \frac{1}{2} \| \sqrt{n} (\hat{\gamma}_n - \gamma(\theta)) \|^2 \right) \right| = o_{\mathbb{P}}(1), \qquad (19)$$

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

Let $\Pi_\Gamma$ denote the image measure (under the map $\theta \mapsto \gamma(\theta)$) of the prior $\Pi$ on $\Theta_I^N$, 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 $\delta$ centered at the origin.

ASSUMPTION 4.3—Prior:
  (i)  $\int_\Theta e^{nL_n(\theta)} \, d\Pi(\theta) < \infty$ *almost surely;*
  (ii) $\Pi_\Gamma$ *has a continuous, strictly positive density $\pi_\Gamma$ 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 $\Theta_{\text{osn}}$ 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 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). 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 $\text{Var}[v(X_i)] = \Sigma$. In fact, Appendix D.1 shows that this quadratic expansion assumption is satisfied uniformly over a large class of DGPs in models of discrete random variables with increasing support. Assumption 4.3(i) requires the quasi-posterior to be proper. Assumption 4.3(ii) is a prior mass and smoothness condition used to establish BvM theorems for point-identified parametric models (see, e.g., Section 10.2 of van der Vaart (2000)) but applied to $\Pi_\Gamma$. We verify this condition in examples of Section 5.

Assumptions 4.1(i) and 4.2 imply that the QLR statistic for $\Theta_I$ satisfies

$$\sup_{\theta \in \Theta_I} Q_n(\theta) = \| \mathbf{T} \mathbb{V}_n \|^2 + o_{\mathbb{P}}(1) \qquad (20)$$

(see Lemma F.1). Therefore, under the *generalized information equality* $\Sigma = I_{d^*}$, which holds for correctly-specified likelihood, 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

$$F_T(z) := \mathbb{P}_Z \big( \| \mathbf{T} Z \|^2 \leq z \big), \qquad (21)$$

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-squared random variable with $d^*$ degrees 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*:

$$\sup_z \left| \Pi_n\left(\{\theta : Q_n(\theta) \leq z\} | \mathbf{X}_n\right) - \mathbb{P}_{Z|\mathbf{X}_n}\left(\|Z\|^2 \leq z | Z \in T - \sqrt{n}\hat{\gamma}_n\right)\right| = o_{\mathbb{P}}(1). \quad (22)$$

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

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 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 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. Lemma 4.1(i) for the convex cone case could be of independent interest to Bayesian QLR tests for shape restrictions even in point-identified models; see, for example, Wei, Wainwright, and Guntuboyina (2018).

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

ASSUMPTION 4.4—MC Convergence: $\xi_{n,\alpha}^{\text{mc}} = \xi_{n,\alpha}^{\text{post}} + o_{\mathbb{P}}(1)$.

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 (20), 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* $\alpha$ *such that* $F_T(\cdot)$ *is continuous at its* $\alpha$ *quantile, we have*:
   (i) $\liminf_{n\to\infty} \mathbb{P}(\Theta_I \subseteq \widehat{\Theta}_\alpha) \geq \alpha$;
   (ii) *If* $T = \mathbb{R}^{d^*}$, *then*: $\lim_{n\to\infty} \mathbb{P}(\Theta_I \subseteq \widehat{\Theta}_\alpha) = \alpha$.

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 $\Theta_I$ 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 $\Theta_I$) when $T = \mathbb{R}^{d^*}$, and asymptotically valid but possibly conservative when $T$ is a convex cone.

REMARK 3: Theorem 4.1 is still applicable to a broad class of separable partially-identified parametric likelihood models that are misspecified. We can write the density

---

[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)).

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 $\mathbb{E}[\log q_{\tilde\gamma}(X_i)]$ 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:

$$L_n(\theta) = -\frac{1}{2}\big(\check\gamma - \tilde\gamma(\theta)\big)'(\widehat\Sigma_S)^{-1}\big(\check\gamma - \tilde\gamma(\theta)\big),$$

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 $\Sigma_S$ 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 $\gamma(\theta) = \Sigma_S^{-1/2}(\tilde\gamma(\theta) - \tilde\gamma^*)$.

REMARK 4: In likelihood models with flat priors, our Procedure 1 CS $\widehat\Theta_\alpha$ is a highest posterior density (HPD) $100\alpha\%$ Bayesian credible set (BCS) for $\theta$. Moon and Schorfheide (2012) (MS hereafter) showed that HPD BCSs for partially-identified parameters can, under some conditions, under-cover (in a frequentist sense) asymptotically. However, a key regularity condition underlying MS's result is violated in our setting. MS put a conditional prior on the model parameter $\theta$ given their globally identified reduced-form parameter $\gamma$. Their Assumption 2 imposes a Lipschitz condition on this conditional prior. We put a prior on $\theta$ only, which induces a prior on the reduced-form parameter $\gamma$. The induced prior necessarily violates MS's Assumption 2. Further, MS's Assumption 2 is violated whenever $\Theta_I$ lies in a lower-dimensional subset of $\Theta$ (see Remark 3 in MS).

### 4.1.1. *Models With Singularities*

In this subsection, we consider 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 $P_\theta$ that is singular with respect to the true distribution $P_0 = P_{\theta_0}$ vanishes faster than $\|\theta - \theta_0\|^2$ as $\theta \to \theta_0$ (Le Cam and Yang (1990, Section 6.2)). If this condition fails, then the log-likelihood will not be locally quadratic at $\theta_0$. 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 Section 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 $\Theta_I^N$ 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 singularities.

ASSUMPTION 4.2′—Local Quadratic Approximation With Singularity:
(i) *There exist sequences of random variables $\ell_n$ 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 $n \to \infty$,*

$$\sup_{\theta \in \Theta_{\mathrm{osn}}} \left| nL_n(\theta) - \left(\ell_n + \frac{1}{2}\|\sqrt{n}\hat\gamma_n\|^2 - \frac{1}{2}\big\|\sqrt{n}\big(\hat\gamma_n - \gamma(\theta)\big)\big\|^2 - f_{n,\perp}\big(\gamma_\perp(\theta)\big)\right)\right| = o_\mathbb{P}(1), \quad (23)$$

---

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

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

Let $\Pi_{\Gamma*}$ denote the image of the measure $\Pi$ 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)  $\int_\Theta e^{nL_n(\theta)}\, d\Pi(\theta) < \infty$ *almost surely*;
(ii)  $\Pi_{\Gamma*}$ *has a continuous, strictly positive density* $\pi_{\Gamma*}$ *on* $B_\delta^* \cap (\Gamma \times \Gamma_\perp)$ *for some* $\delta > 0$.

*Discussion of Assumptions.* Assumption 4.2′ generalizes 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 assumed 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 stochastically dominates $F_T$ in models with singularities.

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

$$\sup_z \big(\Pi_n(\{\theta : Q_n(\theta) \le z\}|\mathbf{X}_n) - \mathbb{P}_{Z|\mathbf{X}_n}\big(\|Z\|^2 \le z | Z \in T - \sqrt{n}\hat{\gamma}_n\big)\big) \le o_\mathbb{P}(1). \qquad (24)$$

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

Lemma 4.2 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* $\alpha$ *such that* $F_T(\cdot)$ *is continuous at its* $\alpha$ *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 $\Theta_I$, 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 $\Theta_I$ 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 $\Theta_I$ is asymptotically exact under partial identification but conservative under point identification. This is consistent with simulation results reported in Table I; 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 $\mu$ in likelihood and moment-based models. Recall the definition of $M(\theta) \equiv \{\mu : (\mu, \eta) \in \Delta(\theta) \text{ for some } \eta\}$ from Section 2.2. The profile criterion

---

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

$PL_n(M(\theta))$ for $M(\theta)$ and the profile QLR $PQ_n(M(\theta))$ for $M(\theta)$ are defined as

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

ASSUMPTION 4.5—Profile QL: *There exists* $f : \mathbb{R}^{d^*} \to \mathbb{R}_+$ *such that*

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

*with* $\hat\gamma_n$ *and* $\gamma(\cdot)$ *from Assumption* 4.2 *or* 4.2′.

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

$$PQ_n(M_I) = f(\mathbf{T}\mathbb{V}_n) + o_{\mathbb{P}}(1). \tag{25}$$

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

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

The functional form of $f$ depends on the local reparameterization $\gamma$ and the geometry of $M_I$. When $M_I$ is a singleton and $T = \mathbb{R}^{d^*}$, then equation (25) 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 profile QLR for $M_I$ is asymptotically $\chi^2_{d^*-d_1^*}$. For a non-singleton set $M_I$, $f$ will typically be more complex. For instance, when $M_I$ is an identified set for scalar subvectors, Proposition 4.1 below presents sufficient conditions so that $f(\mathbf{T}Z)$ becomes a maximum of two mixtures of $\chi^2$ random variables. Luckily, the existence of $f$ is merely a proof device, and one does not need to know its precise expression to implement Procedure 2.

In the following, a function $f : \mathbb{R}^{d^*} \to \mathbb{R}_+$ is said to be quasiconvex if $f^{-1}(z) := \{v : f(v) \le z\}$ is convex for each $z \ge 0$. A function $f$ is said to be subconvex if it is quasiconvex and symmetric at zero (i.e., $f(v) = f(-v)$ for all $v \in \mathbb{R}^{d^*}$). 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 a neighborhood of* $I$, *we have*

$$\sup_{z \in I} \big| \Pi_n\big(\{\theta : PQ_n\big(M(\theta)\big) \le z\} | \mathbf{X}_n\big) - \mathbb{P}_{Z|\mathbf{X}_n}\big(f(Z) \le z | Z \in \sqrt{n}\hat\gamma_n - T\big)\big| = o_{\mathbb{P}}(1). \tag{26}$$

*And hence we have*:
   (i) *If* $T \subsetneq \mathbb{R}^{d^*}$ *and* $f$ *is subconvex, then*: $\sup_{z \in I}(\Pi_n(\{\theta : PQ_n(M(\theta)) \le z\} | \mathbf{X}_n) - G_T(z)) \le o_{\mathbb{P}}(1)$.
   (ii) *If* $T = \mathbb{R}^{d^*}$, *then*: $\sup_{z \in I} |\Pi_n(\{\theta : PQ_n(M(\theta)) \le z\} | \mathbf{X}_n) - \mathbb{P}_Z(f(Z) \le z)| = o_{\mathbb{P}}(1)$.

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

ASSUMPTION 4.6—MC Convergence: $\xi_{n,\alpha}^{\mathrm{mc},p} = \xi_{n,\alpha}^{\mathrm{post},p} + o_{\mathbb{P}}(1)$.

The next theorem is an important consequence of Lemma 4.3.

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* $\alpha$ *quantile.*
  (i) *If* $T \subsetneq \mathbb{R}^{d^*}$ *and* $f$ *is subconvex, then:* $\liminf_{n\to\infty} \mathbb{P}(M_I \subseteq \widehat{M}_\alpha) \geq \alpha$.[18]
  (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 Section 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 E.2 for a moment inequality example).

Procedure 2 CS $\widehat{M}_\alpha$ does not have an interpretation as a HPD BCS for $\mu$. For subvector inference, we eliminate nuisance parameters $\eta$ via profiling and work with the posterior of the profile QLR. A more conventional Bayesian approach would integrate out nuisance parameters and work with the marginal posterior of the subvector $\mu$. HPD BCSs formed from the marginal posterior would be more susceptible to Moon and Schorfheide's (2012) under-coverage result, explaining the under-coverage of percentile CSs in the partially-identified designs in the simulations.

### 4.3. *Coverage Properties of $\widehat{M}_\alpha^\chi$ for $M_I$ for Scalar Subvectors*

This section presents one sufficient condition for validity of Procedure 3 CS $\widehat{M}_\alpha^\chi$ 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, $\chi^2$ Bound: $PQ_n(M_I) \rightsquigarrow W \leq \max_{i\in\{1,2\}} \inf_{t\in T_i} \|Z - t\|^2$, *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* $\alpha$ *quantile. Then:* $\liminf_{n\to\infty} \mathbb{P}(M_I \subseteq \widehat{M}_\alpha^\chi) \geq \alpha$.

We present one set of sufficient conditions for Assumption 4.7 (and hence Theorem 4.4).

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) $\inf_{\mu\in M_I} \sup_{\eta\in H_\mu} L_n(\mu, \eta) = \min_{\mu\in\{\underline{\mu},\overline{\mu}\}} \sup_{\eta\in H_\mu} L_n(\mu, \eta) + o_{\mathbb{P}}(n^{-1})$;

---

[18]The conclusion of Theorem 4.3(i) remains valid under the weaker condition that (i) $f$ is quasiconvex and (ii) $\mathbb{P}_Z(Z \in (f^{-1}(\xi_\alpha) - T^o)) \leq G_T(\xi_\alpha)$, where $\xi_\alpha$ is the $\alpha$ quantile of $G_T$ and $T^o := \{s \in \mathbb{R}^{d^*} : s't \leq 0 \text{ for all } t \in T\}$ is the polar cone of $T$. Similarly, the conclusion of Lemma 4.3(i) remains valid if (i) $f$ is quasiconvex and (ii) $\mathbb{P}_Z(Z \in (f^{-1}(z) - T^o)) \leq G_T(z)$ for each $z \in I$.

(iii) *for each $\mu \in \{\underline{\mu}, \overline{\mu}\}$, there exists a sequence of sets $(\Gamma_{\mu,\text{osn}})_{n\in\mathbb{N}}$ with $\Gamma_{\mu,\text{osn}} \subseteq \Gamma$ for each n and a regular half-space $T_\mu$ in $\mathbb{R}^{d^*}$ such that*

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

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

In many empirical applications, $\Theta_I$ is a connected and bounded subset of $\Theta$, and then $M_I$ for a scalar $\mu$ becomes a finite interval: $M_I = [\underline{\mu}, \overline{\mu}]$ with $-\infty < \underline{\mu} \leq \overline{\mu} < +\infty$. If $\sup_{\eta \in H_\mu} L_n(\mu, \eta)$ is strictly concave in $\mu$, 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 Section 5.3.1). Nevertheless, conditions of Proposition 4.1 could still be satisfied even when $M_I$ is not an interval, as illustrated by the following simple example. Let $(Y_1, Z_1), \ldots, (Y_n, Z_n)$ be i.i.d. $N(\gamma(\theta), I_2)$ with $\theta = (\mu, \eta) \in [0, 1] \times \mathbb{R}$ and $\gamma(\theta) = (\mu^2, \eta)'$. Let $E[Y_i] = 1$. The model is partially identified with $\Theta_I = \{-1, 1\} \times \{E[Z_i]\}$ and $M_I = \{-1, 1\}$. Condition (ii) of Proposition 4.1 is satisfied. The criterion $nL_n(\theta)$ has the required local quadratic form, and the local reduced-form reparameterization is $\gamma(\theta) = ((\mu^2 - 1), \eta - \mathbb{E}[Z_i])' \in [-1, 0] \times \mathbb{R}$. It follows that condition (iii) holds with $\Gamma_{\mu,\text{osn}} = [-n, 0] \times \mathbb{R}$ and $\mathbb{T}_\mu = \mathbb{R}_- \times \mathbb{R}$ for each $\mu \in \{-1, 1\}$.

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 quantiles of the distribution of $\max_{i \in \{1,2\}} \inf_{t \in T_i} \|Z - t\|^2$ are continuous in $\alpha$ for all $\alpha > \frac{1}{4}$. Here $\widehat{M}_\alpha^\chi$ will be most conservative 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 distribution $\frac{1}{4} + \frac{1}{2}F_{\chi_1^2}(z) + \frac{1}{4}F_{\chi_1^2}(z)^2$, which is stochastically dominated by $F_{\chi_1^2}(z)$ for all $z \geq 0$. Note that this is different from the usual chi-bar-squared case encountered when testing whether a parameter $\mu$ belongs to the identified set $M_I$ on the basis of finitely many moment inequalities (Rosen (2008)). Figure 4 plots the asymptotic coverage of $\widehat{M}_\alpha$ and $\widehat{M}_\alpha^\chi$ against nominal coverage for the configuration in which $\widehat{M}_\alpha^\chi$ is *most* conservative for the missing data example. As can be seen, the coverage of $\widehat{M}_\alpha$ is exact at all levels $\alpha \in (\frac{1}{4}, 1)$ (cf. Theorem 4.3(ii)). On the other hand, $\widehat{M}_\alpha^\chi$ is asymptotically conservative, but the level of conservativeness decreases as $\alpha$ increases towards 1. Indeed, for levels of $\alpha$ in excess of 0.85, the level of conservativeness is negligible.

As empirical papers typically report CSs for scalar parameters, Theorem 4.4 and Procedure 3 can be useful in applied work. One could generalize $\widehat{M}_\alpha^\chi$ to deal with vector-valued subvectors by allowing $\chi_d^2$ quantiles with higher degrees of freedom $d \in (1, \dim(\theta))$, but it would be difficult to provide sufficient conditions as those in Proposition 4.1 to establish results like Theorem 4.4. Luckily, Theorem 4.3 and Procedure 2 CSs remain valid for general subvector inference in more complex partially-identified models.

## 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 D.1 for low-level conditions to ensure that Assumption 4.2 holds
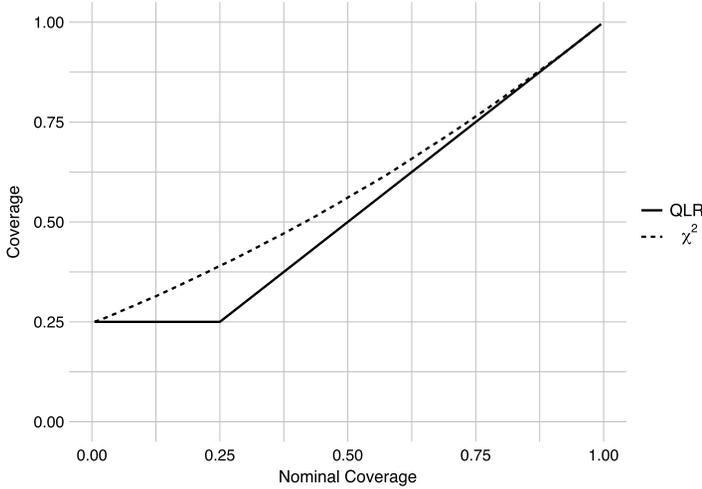
FIGURE 4.—Missing data example: comparison of asymptotic coverage of $\widehat{M}_\alpha$ (QLR—solid kinked line) and $\widehat{M}_\alpha^\chi$ ($\chi^2$—dashed curved line) with their nominal coverage for models where $\widehat{M}_\alpha^\chi$ is valid for $M_I$ but most conservative.

uniformly over a large class of DGPs in models of discrete distributions with increasing supports. 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 $\sigma$-finite dominating measure $\lambda$. Let $p_0 \in \mathcal{P}$ be the true density under the data-generating probability measure, $D_{\mathrm{KL}}(p \parallel q)$ denote the Kullback–Leibler divergence, and $h(p, q)^2 = \int (\sqrt{p} - \sqrt{q})^2 \, d\lambda$ denote the squared Hellinger distance between densities $p$ and $q$. The identified set is $\Theta_I = \{\theta \in \Theta : D_{\mathrm{KL}}(p_0 \parallel p_\theta) = 0\} = \{\theta \in \Theta : h(p_0, p_\theta) = 0\}$.

#### 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, that is, $D_{\mathrm{KL}}(q_{\tilde{\gamma}_0} \parallel q_{\tilde{\gamma}}) > 0$ 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, that is, $p_0 = q_{\tilde{\gamma}_0}$. 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 $\theta$ 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) $P_0 \dot{\ell}_{\tilde{\gamma}_0} = 0$ and $\mathbb{I}_0$ is non-singular;
(d) $\widetilde{\Gamma}$ is compact and $\pi_\Gamma$ 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 $\Theta_{\mathrm{osn}} := \{\theta \in \Theta : \|\gamma(\theta)\| \leq r_n/\sqrt{n}\}$ 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,[19] 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 metric projection[20] onto $\mathcal{T}$ and let $\gamma(\theta)$ be given by

$$\mathbb{T}\big(2(\sqrt{p_\theta/p_0} - 1)\big) = \gamma(\theta)'\psi. \tag{27}$$

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 and $\mathcal{T}$ is a linear space of finite dimension $d^* \geq 1$;
(c) there exists $\varepsilon > 0$ such that $\overline{\mathcal{D}}_\varepsilon$ is Donsker and has an envelope $D \in L^2(P_0)$.

---

[19]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.

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

*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* $\Theta_{\mathrm{osn}} := \{\theta : h(p_\theta, p_0) \leq r_n/\sqrt{n}\}$ *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* (27).

Proposition 5.2 is a set of sufficient conditions for i.i.d. data; see Lemma F.4 in Appendix F of the Supplemental Material 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 a class of (moment equality) functions $\{\rho(X_i, \theta) : \theta \in \Theta\}$ with $\rho : \mathcal{X} \times \Theta \to \mathbb{R}^{d_\rho}$. Let $g(\theta) = E[\rho(X_i, \theta)]$ and the identified set be $\Theta_I = \{\theta \in \Theta : g(\theta) = 0\}$ (we assume throughout this subsection that $\Theta_I$ is non-empty). When $\rho$ is of higher dimension than $\theta$, 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 linear subspace $\mathcal{T} \subset \mathbb{R}^{d_\rho}$.

Let $\rho_\theta = \rho(\cdot, \theta)$ and $\Omega$ be given in condition (b) of Proposition 5.3 below. We assume that, for any $v \in \mathcal{T}$, 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). 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 $[\Omega^{-1/2}\mathbb{T}g(\theta)]_2 = 0$. Let $\Theta_I^\varepsilon = \{\theta \in \Theta : \|g(\theta)\| \leq \varepsilon\}$.

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 and has an envelope* $D \in L^2(P_0)$;
(b) $E[\rho_\theta(X_i)\rho_\theta(X_i)'] = \Omega$ *for each* $\theta \in \Theta_I$ *and* $\Omega$ *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* $\varepsilon \to 0$;
(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* (5) *over* $\Theta_{\mathrm{osn}} = \{\theta \in \Theta : \|g(\theta)\| \leq r_n/\sqrt{n}\}$, *where* $\ell_n = -\frac{1}{2}Z_n'\Omega^{-1}Z_n$, $Z_n = \mathbb{G}_n(\rho_{\theta^*})$, $\gamma(\theta) = [\Omega^{-1/2}\mathbb{T}g(\theta)]_1$, $\sqrt{n}\hat{\gamma}_n = \mathbb{V}_n = -[\Omega^{-1/2}Z_n]_1$ *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* (6).

### 5.2.1. *Moment Inequality Models*

Consider a class of (moment inequality) functions $\{\tilde{\rho}(X_i, \mu) : \mu \in M\}$ with $\tilde{\rho} : \mathcal{X} \times M \to \mathbb{R}^{d_\rho}$ and a parameter space $M \subseteq \mathbb{R}^{d_\mu}$. The identified set for $\mu$ is $M_I = \{\mu \in M : E[\tilde{\rho}(X_i, \mu)] \leq 0\}$ (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 reparameterize

the model by $\theta = (\mu, \eta) \in \Theta = M \times H$ and write the inequality model as a GMM model with

$$E[\rho_\theta(X_i)] = 0 \quad \text{for } \theta \in \Theta_I, \rho_\theta(X_i) = \tilde{\rho}(X_i, \mu) + \eta, \tag{28}$$

where the identified set for $\theta$ is $\Theta_I = \{\theta \in \Theta : E[\rho_\theta(X_i)] = 0\}$ and $M_I$ is the projection of $\Theta_I$ onto $M$. Here, the objective function would be as in display (5) or (6) 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 (28).

As the parameter of interest is $\mu$, 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 $\eta$ for given $\mu$ (as the optimal weighting or continuous updating weighting matrix will typically not depend on $\eta$). See Example 3 in Section 5.3.3.

### 5.3. *Examples*

#### 5.3.1. *Example 1: Missing Data Model in Section 3.1.1*

We revisit the missing data example in Section 3.1.1, where the parameter space $\Theta$ for $\theta = (\mu, \eta_1, \eta_2)$ is given in (16), the identified set for $\theta$ is $\Theta_I$ given in (17), and the identified set for $\mu$ 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

$$\tilde{\gamma}(\theta) = \begin{pmatrix} \tilde{\gamma}_{11}(\theta) \\ \tilde{\gamma}_{00}(\theta) \end{pmatrix}, \qquad \tilde{\gamma}_0 = \begin{pmatrix} \tilde{\gamma}_{11} \\ \tilde{\gamma}_{00} \end{pmatrix}$$

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

$$\mathbb{I}_0 = \begin{bmatrix} \dfrac{1}{\tilde{\gamma}_{11}} + \dfrac{1}{1 - \tilde{\gamma}_{11} - \tilde{\gamma}_{00}} & \dfrac{1}{1 - \tilde{\gamma}_{11} - \tilde{\gamma}_{00}} \\ \dfrac{1}{1 - \tilde{\gamma}_{11} - \tilde{\gamma}_{00}} & \dfrac{1}{\tilde{\gamma}_{00}} + \dfrac{1}{1 - \tilde{\gamma}_{11} - \tilde{\gamma}_{00}} \end{bmatrix},$$

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

$\Sigma = I_2$ and $T = \mathbb{R}^2$. A flat prior on $\Theta$ in (16) induces a flat prior on $\Gamma$, which verifies condition (d) of Proposition 5.1 and Assumption 4.3. Therefore, Theorem 4.1(ii) implies that our CSs $\widehat{\Theta}_\alpha$ for $\Theta_I$ has asymptotically exact coverage.

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

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

Let $g = (g_{11}, g_{00})'$ and let

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

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 (25) 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 half-spaces 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 $\theta \in \Theta_{\text{osn}}$. Then,

$$PL_n(M(\theta)) = \min_{\mu \in \{\tilde{\gamma}_{11}(\theta), \tilde{\gamma}_{11}(\theta) + \tilde{\gamma}_{00}(\theta)\}} \sup_{\eta \in H_\mu} \mathbb{P}_n \log p_{(\mu, \eta)}. \tag{29}$$

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 $\theta \in \Theta_{\text{osn}}$,

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

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 of the Supplemental Material shows that with $\Theta$ as in (16) and a flat prior, the posterior $\Pi_n$ concentrates on the local neighborhood $\Theta_{\text{osn}} = \{\theta : |\tilde{\gamma}_{11}(\theta) - \tilde{\gamma}_{11}| \le r_n/\sqrt{n}, \tilde{\gamma}_{00}(\theta) \le r_n/n\}$ 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 $\gamma_\perp(\theta) = \tilde{\gamma}_{00}(\theta) \geq 0$. Uniformly over $\Theta_{\text{osn}}$, we obtain

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

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 $\Theta_I$ will be asymptotically valid but conservative.

For inference on $M_I = \{\mu_0\}$, the profile LR statistic is asymptotically $\chi_1^2$ and equation (25) holds with $f(v) = v^2$ and $T = \mathbb{R}$. To verify Assumption 4.5, for each $\theta \in \Theta_{\text{osn}}$ we need to solve

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

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

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

With $\gamma_{00}(\theta) = \tilde{\gamma}_{00}(\theta)/\sqrt{\tilde{\gamma}_{11}(1 - \tilde{\gamma}_{11})}$, we obtain the following expansion uniformly for $\theta \in \Theta_{\text{osn}}$:

$$nPL_n(M(\theta)) = \ell_n + \frac{1}{2}\mathbb{V}_n^2 - \frac{1}{2}\left(\left(\mathbb{V}_n - \sqrt{n}\gamma(\theta)\right)^2 \vee \left(\mathbb{V}_n - \sqrt{n}(\gamma(\theta) + \gamma_{00}(\theta))\right)^2\right) + o_\mathbb{P}(1)$$

$$= \ell_n + \frac{1}{2}\mathbb{V}_n^2 - \frac{1}{2}\left(\mathbb{V}_n - \sqrt{n}\gamma(\theta)\right)^2 + o_\mathbb{P}(1),$$

where the last equality holds because $\sup_{\theta \in \Theta_{\text{osn}}} \tilde{\gamma}_{00}(\theta) \leq r_n/n = o(n^{-1/2})$. 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 $\Theta_I$ in this case.

### 5.3.2. *Example 2: Entry Game With Correlated Shocks in Section 3.1.2*

Consider the bivariate discrete game with payoffs described in Section 3.1.2. Here, we consider a slightly more general setting, in which $Q_\rho$ denotes a general joint distribution (not just bivariate Gaussian) for $(\epsilon_1, \epsilon_2)$ indexed by a parameter $\rho$. This model falls into the class of models dealt with in Proposition 5.1. Conditions (a), (b), and (c) of Proposition 5.1 hold with $\tilde{\gamma}(\theta) = (\tilde{\gamma}_{00}(\theta), \tilde{\gamma}_{10}(\theta), \tilde{\gamma}_{11}(\theta))'$ and $\tilde{\Gamma} = \{\tilde{\gamma}(\theta) : \theta \in \Theta\}$ under very mild conditions on the parameterization $\theta \mapsto \tilde{\gamma}(\theta)$ (which, in turn, is determined by the

specification of $Q_\rho$). Assumption 4.2 is therefore satisfied with

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

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

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

and $T = \mathbb{R}^3$. Condition (d) 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 $\Pi$. For instance, consider the parameterization $\theta = (\Delta_1, \Delta_2, \beta_1, \beta_2, \rho, s)$ where the joint distribution of $(\epsilon_1, \epsilon_2)$ is a bivariate Normal with mean zero, standard deviations 1, and positive correlation $\rho \in [0, 1]$. The parameter space is

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

where $-\infty < \underline{\Delta} < \overline{\Delta} < 0$ and $-\infty < \underline{\beta} < \overline{\beta} < \infty$. The image measure $\Pi_\Gamma$ of a flat prior on $\Theta$ 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 $\widehat{\Theta}_\alpha$ has asymptotically exact coverage for $\Theta_I$.

### 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 $\mathbb{E}[\mu - X_i] \le 0$ where $X_1, \ldots, X_n$ are i.i.d. with unknown mean $\mu^* \in \mathbb{R}_+$ and unit variance. The identified set for $\mu$ 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 $\theta$ is $\Theta_I = \{(\mu, \eta) \in \Theta : \mu + \eta = \mu^*\}$ and the identified set for $\mu$ 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 $PQ_n(M_I) = (\mathbb{V}_n \wedge 0)^2 - ((\mathbb{V}_n + \sqrt{n}\mu^*) \wedge 0)^2$.
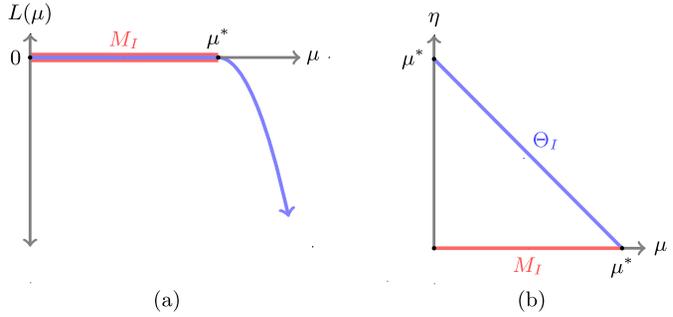
FIGURE 5.—Panel (a): Identified set $M_I$ for $\mu$ as the argmax of the population (moment inequality) criterion $L(\mu) = -\frac{1}{2}((\mu - \mu^*) \vee 0)^2$. Panel (b): Identified set $\Theta_I$ for $\theta = (\mu, \eta)$ for the moment equality model $E[\mu + \eta - X] = 0$.

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

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

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 $\mu$ and $\eta$ induces a prior on $\gamma$ 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 E.2 of the Supplemental Material shows that, under very mild conditions, our CS $\widehat{M}_\alpha$ is uniformly valid over a class of DGPs $\mathbf{P}$, that is,

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

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 $(P_n)_{n\in\mathbb{N}} \subset \mathbf{P}$ along which bootstrap-based CSs $\widehat{M}_\alpha^{\text{boot}}$ fail to cover with the prescribed coverage probability, that is,

$$\limsup_{n\to\infty} P_n\big(\mathbb{M}_I(P_n) \subseteq \widehat{M}_\alpha^{\text{boot}}\big) < \alpha.$$

This shows that our MC CSs for $M_I$ have very different asymptotic properties from bootstrap-based CSs for $M_I$ in some non-regular situations.

## 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 $\Theta_I$ may be slightly conservative, but our MC CSs for identified sets $M_I$ of subvectors could still be asymptotically exact. Simulation experiments demonstrate the good finite-sample coverage properties of our proposed CSs. Empirical examples illustrate that our proposed CSs are applicable in difficult situations where complicated structural models might not be point-identified.

This paper's MC approach could be a useful alternative to the sieve MLE bootstrap inference on partially-identified likelihood models in Chen, Tamer, and Torgovitsky (2011). We are currently working on extensions to semiparametric probability models or conditional moment-based models involving partially-identified unknown functions.

## APPENDIX A: AN ADAPTIVE SEQUENTIAL MONTE CARLO ALGORITHM

We use an adaptive Sequential Monte Carlo (SMC) algorithm to sample from the quasi-posterior in (7). 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, that is, 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 follow Herbst and Schorfheide (2014) who adapted a generic adaptive SMC algorithm to deal with large-scale DSGE models.[21] A similar algorithm was proposed by Durham and Geweke (2014), who emphasized 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 $v_j^b = e^{(\phi_j - \phi_{j-1})nL_n(\theta_{j-1}^b)}$ 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 $\vartheta_j^b$ for the tempered quasi-posterior $\Pi_j(\theta|\mathbf{X}_n) \propto e^{\phi_j nL_n(\theta)}\Pi(\theta)$ and let $\theta_j^b$ be the final draw of the $b$th chain.

---

[21]See Chopin (2002, 2004) and Del Moral, Doucet, and Jasra (2006) for the generic SMC algorithm for estimating static model parameters. See Del Moral, Doucet, and Jasra (2012), Beskos, Jasra, Kantas, and Thiery (2016) and references therein for adaptive selection of tuning parameters with a SMC framework and theoretical analyses of the convergence properties of adaptive SMC algorithms. Creal (2012) provided a survey of applications of SMC methods in economics.

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.

In practice, we take $J = 200$, $K = 1$, 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 $\theta$ is low, in step 3 we use an $N(0, \sigma_j^2 I)$ proposal density (all parameters are transformed to have full support) where $\sigma_j$ is chosen adaptively to target an acceptance ratio $\approx 0.35$ by setting $\sigma_2 = 1$ and

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

for $j > 2$, where $A_{j-1}$ is the acceptance ratio from the previous iteration. If the dimension of $\theta$ is large, we partition $\vartheta_j^b$ into $L$ random blocks (we assign each element of $\vartheta_j^b$ 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 $\sigma_j$ is chosen as before, $\Sigma_{j-1}$ is the covariance of the draws from iteration $j - 1$, and $\Sigma_j^l$ is the submatrix of $\Sigma_j$ corresponding to block $l$.

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

$$\Pi\big(\{\theta : Q_n(\theta) \le z\}|\mathbf{X}_n\big) = \frac{1}{B}\sum_{b=1}^{B} w_J^b \mathbb{1}\big\{Q_n(\theta^b) \le z\big\} \tag{30}$$

and similarly for the profile QLR for Procedure 2.

### A.1. *Example 1: Missing Data*

*SMC Algorithm*: We implement the SMC algorithm with $(J, K) = (200, 1)$ and an $N(0, \sigma_j^2 I)$ proposal in the mutation step for all simulations for this example.

In Appendix G of the Supplemental Material, we present additional simulation results using a likelihood criterion with a curved prior and a continuously-updated GMM criterion with a flat prior.

### A.2. *Example 2: Entry Game With Correlated Shocks*

*SMC Algorithm*: As there are six partially-identified parameters here instead of two 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 computationally more efficient than using larger $J$. We again use an $N(0, \sigma_j^2 I)$ 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 $\rho \ne 0$. 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

$D_{\mathrm{KL}}(p_\theta \parallel p_\vartheta)$ denote the KL divergence between $p_\theta$ and $p_\vartheta$ or any $\theta, \vartheta \in \Theta$, which is given by

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

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 $\theta$. Clearly, $\vartheta \in \Delta(\theta)$ if and only if $D_{\mathrm{KL}}(p_\theta \parallel p_\vartheta) = 0$. We compute the endpoints of the interval $M(\theta^b)$ by solving

$$\min/\max \mu \quad \text{such that} \quad \inf_{\eta \in H_\mu} D_{\mathrm{KL}}(p_{\theta^b} \parallel p_{(\mu,\eta)}) = 0, \tag{31}$$

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 $\inf_{\eta \in H_\mu} D_{\mathrm{KL}}(p_\theta \parallel p_{(\mu,\eta)})$ 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 small optimization errors, in practice we replace the equality in (31) by a small tolerance: $D_{\mathrm{KL}}(p_{\theta^b} \parallel p_{(\mu,\eta)}) < 10^{-7}$. This slight relaxation makes our CSs computed via Procedure 2 slightly more conservative than those if the set $M(\theta^b)$ were known in closed form.

## A.3. *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. See Appendix G of the Supplemental Material for evidences illustrating convergence of the SMC algorithm for this application.

*Procedure 2*: To implement Procedure 2 here with any scalar subvector $\mu$, 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 (31) for the subvector of interest. To accommodate small numerical optimization errors, we again replace the equality in (31) by a small tolerance: $D_{\mathrm{KL}}(p_{\theta^b} \parallel p_{(\mu,\eta)}) < 10^{-5}$. This makes our Procedure 2 CSs slightly more conservative than those if $M(\theta^b)$ were known in closed form. 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.

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

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

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

### A.4. *Trade Flow Application*

*Priors*: We use the change of variables $2\arctanh(\rho)$ and $\log\sigma_m^2$ and assume that the transformed correlation and variance all have full support. We specify an 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 with $\theta_1^1, \ldots, \theta_1^B$ 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. For $j = 2, \ldots, J$, we make two more minor modifications to correct the particle weights from initializing the algorithm in this manner. First, in the correction step, we replace $v_j^b$ by $v_j^b = (e^{nL_n(\theta_{j-1}^b)}\Pi(\theta_{j-1}^b)/\mathcal{N}_n(\theta_{j-1}^b))^{\phi_j - \phi_{j-1}}$ where $\mathcal{N}_n(\theta^b)$ denotes the $N(\hat{\theta}, -I(\hat{\theta})^{-1})$ density evaluated at $\theta^b$. Second, in the mutation and updating step, we use the tempered quasi-posterior $\Pi_j(\theta|\mathbf{X}_n) \propto (e^{nL_n(\theta)}\Pi(\theta))^{\phi_j}\mathcal{N}_n(\theta)^{1-\phi_j}$.

With these modifications, the algorithm is implemented with $J = 200$, $K = 8$ block random-walk Metropolis–Hastings steps per iteration, and $L = 6$ blocks.

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

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

is approximately zero (in practice, we use a tolerance of $10^{-7}$). 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 $\mu$ 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 $D_{\mathrm{KL}}(p_{\theta^b}(\cdot|X_{ij})\|p_{(\mu,\eta)}(\cdot|X_{ij}))$, let $d_{ij}$ be a dummy variable denoting exports from $j$ to $i$. 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 $\beta_z$ and $\beta_m$ collect all coefficients in the selection and outcome equations, respectively. Therefore,

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

The likelihood is

$$p_\theta(d_{ij}, d_{ij}m_{ij}|X_{ij}) = \left(1 - \Phi\left(\frac{X'_{ij}\beta_z}{\sigma_z(X_{ij})}\right)\right)^{1-d_{ij}}$$

$$\times \left(\Phi\left(\frac{\dfrac{X'_{ij}\beta_z}{\sigma_z(X_{ij})} + r(X_{ij})\dfrac{d_{ij}m_{ij} - X'_{ij}(\beta_m + \delta\beta_z)}{\sigma_v(X_{ij})}}{\sqrt{1 - r^2(X_{ij})}}\right)\right.$$

$$\left.\times \frac{1}{\sigma_v(X_{ij})}\phi\left(\frac{d_{ij}m_{ij} - X'_{ij}(\beta_m + \delta\beta_z)}{\sigma_v(X_{ij})}\right)\right)^{d_{ij}},$$

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 $p_{\theta^b}$ 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.

## APPENDIX B: 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 B.1: *There exist sequences of distributions* $(P_{n,a})_{n\in\mathbb{N}}$ *such that as* $n \to \infty$:

(i) $L_n(\hat{\theta}) = \sup_{\theta\in\Theta_{\mathrm{osn}}} L_n(\theta) + o_{P_{n,a}}(n^{-1})$, *with* $(\Theta_{\mathrm{osn}})_{n\in\mathbb{N}}$ *a sequence of local neighborhoods of* $\Theta_I$;

(ii) $\Pi_n(\Theta_{\mathrm{osn}}^c|\mathbf{X}_n) = o_{P_{n,a}}(1)$;

(iii) *There exist sequences of random variables* $\ell_n$ *and* $\mathbb{R}^{d^*}$-*valued random vectors* $\hat{\gamma}_n$ (*both measurable in* $\mathbf{X}_n$) *such that*

$$\sup_{\theta\in\Theta_{\mathrm{osn}}} \left|nL_n(\theta) - \left(\ell_n + \frac{1}{2}\|\sqrt{n}\hat{\gamma}_n\|^2 - \frac{1}{2}\|\sqrt{n}(\hat{\gamma}_n - \gamma(\theta))\|^2\right)\right| = o_{P_{n,a}}(1) \tag{32}$$

*with* $\sup_{\theta\in\Theta_{\mathrm{osn}}} \|\gamma(\theta)\| \to 0$, $\sqrt{n}\hat{\gamma}_n = \mathbb{V}_n$ *where* $\mathbb{V}_n \overset{P_{n,a}}{\rightsquigarrow} N(a, I_{d^*})$, *and* $T = \mathbb{R}^{d^*}$;

(iv) $\int_\Theta e^{nL_n(\theta)}\,d\Pi(\theta) < \infty$ *holds* $P_{n,a}$-*almost surely*;

(v) $\Pi_\Gamma$ *has a continuous, strictly positive density* $\pi_\Gamma$ *on* $B_\delta \cap \Gamma$ *for some* $\delta > 0$;

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

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

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

THEOREM B.1: *Let Assumption* B.1(i), (iii) *hold. Then*:

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

*if, further, Assumption* B.1(ii), (iv), (v) *hold, then*

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

*and if, further, Assumption* B.1(vi) *holds, then*

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

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

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

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

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

Assumption B.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 $\alpha$ quantile, which we denote by $z_\alpha$.

THEOREM B.2: *Let Assumption* B.1(i), (iii), (vii) *hold. Then*

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

*if, further, Assumption* B.1(ii), (iv), (v) *hold, then, for a neighborhood* $I$ *of* $z_\alpha$,

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

*and if, further, Assumption* B.1(vi′) *holds, then*

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

When $f$ is subconvex, it follows from Anderson's lemma (van der Vaart (2000, Lemma 8.5)) that $\lim_{n \to \infty} P_{n,a}(M_I \subseteq \widehat{M}_\alpha) \le \alpha$, and from Lewandowski, Ryznar, and Zak (1995) that

$$\lim_{n \to \infty} P_{n,a}(M_I \subseteq \widehat{M}_\alpha) < \alpha \quad \text{whenever } a \neq 0.$$

In particular, this includes the case in which $M_I$ is a singleton.

## APPENDIX C: PARAMETER-DEPENDENT SUPPORT

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

We again presume the existence of a local reduced-form parameter $\gamma$ such that $\gamma(\theta) = 0$ if and only if $\theta \in \Theta_I$. In what follows, we assume without loss of generality that $L_n(\hat{\theta}) = \sup_{\theta \in \Theta_{\text{osn}}} L_n(\theta)$, as $\hat{\theta}$ 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 $\gamma$.

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

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

*with* $\sup_{\theta \in \Theta_{\text{osn}}} \|\gamma(\theta)\| \to 0$ *and* $\inf\{h(\gamma) : \|\gamma\| = 1\} > 0$;

(ii) *there exist* $r_1, \ldots, r_{d^*} > 0$ *such that* $th(\gamma) = h(t^{r_1}\gamma_1, t^{r_2}\gamma_2, \ldots, t^{r_{d^*}}\gamma_{d^*})$ *for each* $t > 0$;

(iii) *the sets* $K_{\text{osn}} = \{(b_n^{-r_1}(\gamma_1(\theta) - \hat{\gamma}_{n,1}), \ldots, b_n^{-r_{d^*}}(\gamma_{d^*}(\theta) - \hat{\gamma}_{n,d^*}))' : \theta \in \Theta_{\text{osn}}\}$ *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 and 3 in Fan, Hung, and Wong (2000) but has been modified to allow for non-identifiable parameters $\theta$. Let $F_\Gamma$ 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 $F_\Gamma$.

LEMMA C.1: *Let Assumptions* 4.1, *C.2, and* 4.3 *hold. Then*

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

By modifying appropriately the arguments in Fan, Hung, and Wong (2000), one can show that, under Assumption C.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 C.1: *Let Assumptions* 4.1, *C.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 $\Theta_I = \{\theta \in \Theta : \theta_1 \vee \theta_2 = \tilde{\gamma}\}$. 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 $\Theta_{\text{osn}} = \{\theta : (1 + \varepsilon_n)\hat{\gamma}_n \geq$

$\gamma(\theta) \geq \hat{\gamma}_n\}$ 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 $F_\Gamma$ 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

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

Notice that $r^* = 1$ and that the sets $K_{\mathrm{osn}} = \{n(\gamma(\theta) - \hat{\gamma}_n) : \theta \in \Theta_{\mathrm{osn}}\} = \{n(\gamma - \hat{\gamma}_n) : (1 + \varepsilon_n)\hat{\gamma} \geq \gamma \geq \hat{\gamma}_n\}$ cover $\mathbb{R}^+$. A smooth prior on $\Theta$ will induce a smooth prior on $\gamma(\theta)$, and the result follows from Theorem C.1.

## APPENDIX D: 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 **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 $\mu$ is $M_I(\mathbb{P}) = \{\mu : (\mu, \eta) \in \Theta_I(\mathbb{P})$ 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, that is,

$$\liminf_{n \to \infty} \inf_{\mathbb{P} \in \mathbf{P}} \mathbb{P}(\Theta_I(\mathbb{P}) \subseteq \widehat{\Theta}_\alpha) \geq \alpha, \tag{33}$$

$$\liminf_{n \to \infty} \inf_{\mathbb{P} \in \mathbf{P}} \mathbb{P}(M_I(\mathbb{P}) \subseteq \widehat{M}_\alpha) \geq \alpha \tag{34}$$

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

LEMMA D.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*: (33) *holds for* $\widehat{\Theta}_\alpha = \{\theta \in \Theta : Q_n(\theta) \leq v_{\alpha,n}\}$.

LEMMA D.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*: (34) *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 $(\Theta_{\mathrm{osn}}(\mathbb{P}))_{n \in \mathbb{N}}$ denote a sequence of local neighborhoods of $\Theta_I(\mathbb{P})$ such that $\Theta_I(\mathbb{P}) \subseteq \Theta_{\mathrm{osn}}(\mathbb{P})$ for each $n$ and for each $\mathbb{P} \in \mathbf{P}$. In what follows, we omit the dependence of $\Theta_{\mathrm{osn}}(\mathbb{P})$ on $\mathbb{P}$ to simplify notation.

ASSUMPTION D.1—Consistency, Posterior Contraction:
 (i) $L_n(\hat\theta) = \sup_{\theta \in \Theta_{\mathrm{osn}}} L_n(\theta) + o_{\mathbb{P}}(n^{-1})$ *uniformly in* $\mathbb{P}$.
 (ii) $\Pi_n(\Theta_{\mathrm{osn}}^c | \mathbf{X}_n) = o_{\mathbb{P}}(1)$ *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 $\Theta_I^N$ of $\Theta_I$. We require that $\Theta_{\mathrm{osn}}(\mathbb{P}) \subseteq \Theta_I^N(\mathbb{P})$ 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 $\gamma(\theta; \mathbb{P}) = 0$ 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 $B_\delta$ denote a ball of radius $\delta$ centered at the origin (the dimension will be obvious depending on the context) and let $\nu_{d^*}$ denote Gaussian measure on $\mathbb{R}^{d^*}$.

To simplify notation, in what follows we omit dependence of $d^*$, $\gamma$, $\gamma_\perp$, $\Gamma$, $\Gamma_\perp$, $k_n$, $\ell_n$, $T$, $\mathbf{T}$, $T_{\mathrm{osn}}$, $\tau$, $\Theta_I^N$, $\mathbb{V}_n$, $\Sigma$, and $f_{n,\perp}$ on $\mathbb{P}$.

ASSUMPTION D.2—Local Quadratic Approximation:
 (i) *For each* $\mathbb{P} \in \mathbf{P}$, *there exist vectors* $\tau \in T$, *sequences of random variables* $\ell_n$ *and* $\mathbb{R}^{d^*}$-*valued random vectors* $\hat\gamma_n$, *and a sequence of nonnegative 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* $n \to \infty$,

$$\sup_{\theta \in \Theta_{\mathrm{osn}}} \left| nL_n(\theta) - \left( \ell_n + \frac{1}{2} \left\| \sqrt{n}(\hat\gamma_n - \tau) \right\|^2 - \frac{1}{2} \left\| \sqrt{n}(\hat\gamma_n - \tau - \gamma(\theta)) \right\|^2 - f_{n,\perp}(\gamma_\perp(\theta)) \right) \right|$$
$$= o_{\mathbb{P}}(1) \tag{35}$$

*uniformly in* $\mathbb{P}$, *with* $\sup_{\mathbb{P} \in \mathbf{P}} \sup_{\theta \in \Theta_{\mathrm{osn}}} \|(\gamma(\theta), \gamma_\perp(\theta))\| \to 0$, $\sqrt{n}\hat\gamma_n = \mathbf{T}(\mathbb{V}_n + \sqrt{n}\tau)$, *and* $\|\mathbb{V}_n\| = O_{\mathbb{P}}(1)$ *uniformly in* $\mathbb{P}$;
 (ii) $\{\sqrt{n}\gamma(\theta) : \theta \in \Theta_{\mathrm{osn}}\} \cap B_{k_n} = (T - \sqrt{n}\tau) \cap B_{k_n}$ *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* $\mathbb{P} \in \mathbf{P}$: $\{(\gamma(\theta), \gamma_\perp(\theta)) : \theta \in \Theta_{\mathrm{osn}}\} = \{\gamma(\theta) : \theta \in \Theta_{\mathrm{osn}}\} \times \{\gamma_\perp(\theta) : \theta \in \Theta_{\mathrm{osn}}\}$.

Let $\Pi_{\Gamma^*}$ denote the image measure of $\Pi$ 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 $\delta$, $\Pi_{\Gamma^*}$, and $\pi_{\Gamma^*}$ on $\mathbb{P}$ in what follows.

ASSUMPTION D.3—Prior:
 (i) $\int_\theta e^{nL_n(\theta)} \, d\Pi(\theta) < \infty$ $\mathbb{P}$-*almost surely for each* $\mathbb{P} \in \mathbf{P}$;
 (ii) *each* $\Pi_{\Gamma^*}$ *has a density* $\pi_{\Gamma^*}$ *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 an $N(0, I_{d^*})$ random vector $Z$ (conditional on the data).

LEMMA D.3: *Let Assumptions* D.1, D.2, *and* D.3 *hold. Then*

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

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

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

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

As in Section 4, we let $\xi_{n,\alpha}^{\text{post}}$ denote the $\alpha$ quantile of $Q_n(\theta)$ under the posterior distribution $\Pi_n$.

ASSUMPTION D.4—MC Convergence: $\xi_{n,\alpha}^{\text{mc}} = \xi_{n,\alpha}^{\text{post}} + o_{\mathbb{P}}(1)$ *uniformly in* $\mathbb{P}$.

The following result is a uniform-in-$\mathbb{P}$ extension of Theorems 4.1 and 4.2. Recall that $F_T(z) = \mathbb{P}_Z(\|\mathbf{T}Z\|^2 \leq z)$ where $\mathbb{P}_Z$ denotes the distribution of an $N(0, I_{d^*})$ random vector. We say that the distributions $\{F_T : \mathbb{P} \in \mathbf{P}\}$ are equicontinuous at their $\alpha$ 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 D.1: *Let Assumptions* D.1, D.2, D.3, *and* D.4 *hold, and let*

$$\sup_{\mathbb{P}\in\mathbf{P}} \sup_z \big|\mathbb{P}\big(\|\mathbf{T}\mathbb{V}_n\|^2 \leq z\big) - F_T(z)\big| = o(1),$$

*where the distributions* $\{F_T : \mathbb{P} \in \mathbf{P}\}$ *are equicontinuous at their* $\alpha$ *quantiles.*
  (i) *If* $T = \mathbb{R}^{d^*}$ *for each* $\mathbb{P}$, *then*: (33) *holds.*
  (ii) *If no* $\mathbb{P} \in \mathbf{P}$ *is singular and* $T = \mathbb{R}^{d^*}$ *for each* $\mathbb{P}$, *then*: (33) *holds with equality.*

To establish (34), 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 D.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*

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

*uniformly in* $\mathbb{P}$, *with* $\hat{\gamma}_n, \ell_n, \tau$, *and* $\gamma(\cdot)$ *from Assumption* D.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 D.5 automatically hold with equality if $T = \mathbb{R}^{d^*}$. These conditions are not needed in the following result which is a uniform-in-$\mathbb{P}$ extension of Lemma 4.3.

LEMMA D.4: *Let Assumptions* D.1, D.2, D.3, *and* D.5(i) *hold. Let $\epsilon$ be a small positive value that is independent of $\mathbb{P} \in \mathbf{P}$. Then, for any interval $I = I(\mathbb{P})$ such that $\mathbb{P}_Z(f(Z) \le z)$ is uniformly continuous on an $\epsilon$-neighborhood of $I$ (in both $z$ and $\mathbb{P}$),*

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

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

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

ASSUMPTION D.6—MC Convergence: $\xi_{n,\alpha}^{\mathrm{mc},p} = \xi_{n,\alpha}^{\mathrm{post},p} + o_{\mathbb{P}}(1)$ *uniformly in $\mathbb{P}$.*

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

THEOREM D.2: *Let Assumptions* D.1, D.2, D.3, D.5, *and* D.6 *hold, and let*

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

*where the distributions $\{\mathbb{P}_Z(f(Z) \le z) : \mathbb{P} \in \mathbf{P}\}$ are equicontinuous at their $\alpha$ quantiles.*
   (i) *Then*: (34) *holds*.
   (ii) *If Assumption* D.5(ii), (iii) *holds with equality for all $\mathbb{P} \in \mathbf{P}$, then*: (34) *holds with equality.*

## D.1. *A Uniform Quadratic Expansion for Discrete Distributions With Increasing Supports*

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 $k \ge 2$. Let the $k$-vector $p_\theta$ 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 $p_\theta$ 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 concerns any class of distributions $\mathbf{P}$ for which

$$\sup_{\mathbb{P} \in \mathbf{P}} \max_{1 \le j \le k} \frac{1}{p(j)} = o(n). \tag{36}$$

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

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

For each $\mathbb{P}$, let $\Theta_{\mathrm{osn}}(\mathbb{P}) = \{\theta : p_\theta > 0, \chi^2(p_\theta; p) \le r_n^2 n^{-1}\}$ 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 $x$th 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 D.5: *Let* (36) *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* $\mathbb{P} \in \mathbf{P}$,

$$\ell_n = \ell_n(\mathbb{P}) = nL_n(p), \qquad \tilde{\gamma}_{\theta;p} = \begin{bmatrix} \dfrac{p_\theta(1) - p(1)}{\sqrt{p(1)}} \\ \vdots \\ \dfrac{p_\theta(k) - p(k)}{\sqrt{p(k)}} \end{bmatrix},$$

$$\tilde{\mathbb{V}}_{n;p} = \mathbb{G}_n(\mathbb{J}_p e_x) \overset{\mathbb{P}}{\rightsquigarrow} N\big(0, I - \sqrt{p}\sqrt{p}'\big).$$

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

$$V_p \tilde{\gamma}_{\theta;p} = \begin{bmatrix} v_{1,p}' \tilde{\gamma}_{\theta;p} \\ \vdots \\ v_{k-1,p}' \tilde{\gamma}_{\theta;p} \\ 0 \end{bmatrix}, \qquad V_p \mathbb{G}_n(\mathbb{J}_p e_x) = \begin{bmatrix} v_{1,p}' \mathbb{G}_n(\mathbb{J}_p e_x) \\ \vdots \\ v_{k-1,p}' \mathbb{G}_n(\mathbb{J}_p e_x) \\ 0 \end{bmatrix}. \tag{37}$$

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 $V_p \mathbb{G}_n(\mathbb{J}_p e_x)$:

$$\gamma(\theta) = \begin{bmatrix} v_{1,p}' \tilde{\gamma}_{\theta;p} \\ \vdots \\ v_{k-1,p}' \tilde{\gamma}_{\theta;p} \end{bmatrix}, \qquad \mathbb{V}_n = \begin{bmatrix} v_{1,p}' \mathbb{G}_n(\mathbb{J}_p e_x) \\ \vdots \\ v_{k-1,p}' \mathbb{G}_n(\mathbb{J}_p e_x) \end{bmatrix}. \tag{38}$$

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 $d_\pi$ denotes the distance (in the Prokhorov metric) between the sampling distribution of $\mathbb{V}_n$ and the $N(0, I_{k-1})$ distribution.

PROPOSITION D.1: *Let* (36) *hold and* $\Theta_{osn}(\mathbb{P})$ *be as described in Lemma* D.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 D.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 $\Theta_{\mathrm{osn}}(\mathbb{P}) = \{\theta : p_\theta > 0, \chi^2(p_\theta; p) \leq r_n^2 n^{-1}\}$ where $(r_n)_{n \in \mathbb{N}}$ is a positive sequence to be defined below.

PROPOSITION D.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*

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

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

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

LEMMA D.6: *Let* (36) *hold*, $\{p_\theta : \theta \in \Theta, p_\theta > 0\} = \mathrm{int}(\Delta^{k-1})$, *and* $\Theta_{\mathrm{osn}}(\mathbb{P})$ *be as described in Lemma* D.5. *Then*: *for each* $\mathbb{P} \in \mathbf{P}$, $\{\sqrt{n} \gamma(\theta) : \theta \in \Theta_{\mathrm{osn}}(\mathbb{P})\}$ *covers a ball of radius* $\rho_n \to \infty$ (*with* $\rho_n$ *not depending on* $\mathbb{P}$) *as* $n \to \infty$.

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

LEMMA D.7: *Let* (36) *hold*. *Then*: *there exists a positive sequence* $(r'_n)_{n \in \mathbb{N}}$ *with* $r'_n \to \infty$ *as* $n \to \infty$ *such that*

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

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

## REFERENCES

ANDREWS, D. (1999): "Estimation When a Parameter Is on a Boundary," *Econometrica*, 67 (6), 1341–1383.[1987]

——— (2017): "Identification-Robust Subvector Inference," Cowles Foundation Discussion Paper no 3005, Yale. [1969]

ANDREWS, D., AND P. J. BARWICK (2012): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," *Econometrica*, 80 (6), 2805–2826.[1969]

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.[1969]

ANDREWS, D., AND G. SOARES (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78 (1), 119–157.[1966,1967,1969,1978]

ARMSTRONG, T. (2014): "Weighted KS Statistics for Inference on Conditional Moment Inequalities," *Journal of Econometrics*, 181 (2), 92–116.[1969]

AZAÏS, J.-M., É. 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.[1997]

BERESTEANU, A., AND F. MOLINARI (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76 (4), 763–814.[1969]

BERRY, S. T. (1992): "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, 60, 889–917.[1982]

BESKOS, A., A. JASRA, N. KANTAS, AND A. THIERY (2016): "On the Convergence of Adaptive Sequential Monte Carlo Methods," *The Annals of Applied Probability*, 26 (2), 1111–1146.[2004]

BOCHKINA, N. A., AND P. J. GREEN (2014): "The Bernstein–von Mises Theorem and Nonregular Models," *The Annals of Statistics*, 42 (5), 1850–1878.[1992]

BUGNI, F. (2010): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78 (2), 735–753.[1969]

BUGNI, F., I. CANAY, AND X. SHI (2017): "Inference for Subvectors and Other Functions of Partially Identified Parameters in Moment Inequality Models," *Quantitative Economics*, 8, 1–38.[1969]

CANAY, I. A. (2010): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," *Journal of Econometrics*, 156 (2), 408–425.[1969]

CHEN, X., T. M. CHRISTENSEN AND E. TAMER (2018): "Supplement to 'Monte Carlo Confidence Sets for Identified Sets'," *Econometrica Supplemental Material*, 86, https://doi.org/10.3982/ECTA14525. [1969]

CHEN, X., E. TAMER, AND A. TORGOVITSKY (2011): "Sensitivity Analysis in Semiparametric Likelihood Models," Cowles Foundation Discussion Paper no 1836, Yale and Northwestern. [2004]

CHERNOFF, H. (1954): "On the Distribution of the Likelihood Ratio," *The Annals of Mathematical Statistics*, 25 (3), 573–578.[1987]

CHERNOZHUKOV, V., AND H. HONG (2003): "An MCMC Approach to Classical Estimation," *Journal of Econometrics*, 115 (2), 293–346.[1968,1974,1990]

——— (2004): "Likelihood Estimation and Inference in a Class of Nonregular Econometric Models," *Econometrica*, 72 (5), 1445–1480.[2010]

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75 (5), 1243–1284.[1966]

CHOPIN, N. (2002): "A Sequential Particle Filter Method for Static Models," *Biometrika*, 89 (3), 539–552.[2004]

——— (2004): "Central Limit Theorem for Sequential Monte Carlo Methods and Its Application to Bayesian Inference," *The Annals of Statistics*, 32 (6), 2385–2411.[2004]

CREAL, D. (2012): "A Survey of Sequential Monte Carlo Methods for Economics and Finance," *Econometric Reviews*, 31 (3), 245–296.[2004]

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.[2004]

——— (2012): "On Adaptive Resampling Strategies for Sequential Monte Carlo Methods," *Bernoulli*, 18 (1), 252–278.[1966,2004]

DURHAM, G., AND J. GEWEKE (2014): "Adaptive Sequential Posterior Simulators for Massively Parallel Computing Environments," in *Bayesian Model Comparison*, Chapter 1, 1–44. [1966,2004]

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.[2010]

FLINN, C., AND J. HECKMAN (1982): "New Methods for Analyzing Structural Models of Labor Force Dynamics," *Journal of Econometrics*, 18 (1), 115–168.[2010]

GEYER, C. J. (1994): "On the Asymptotics of Constrained $M$-Estimation," *The Annals of Statistics*, 22 (4), 1993–2010.[1987]

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.[1998]

HANSEN, L. P., J. HEATON, AND A. YARON (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business & Economic Statistics*, 14 (3), 262–280.[1970]

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.[1980,1984,1986,1987]

HERBST, E., AND F. SCHORFHEIDE (2014): "Sequential Monte Carlo Sampling for DSGE Models," *Journal of Applied Econometrics*, 29 (7), 1073–1098.[1966,2004,2005]

HIRANO, K., AND J. PORTER (2003): "Asymptotic Efficiency in Parametric Structural Models With Parameter-Dependent Support," *Econometrica*, 71 (5), 1307–1338.[2010]

IMBENS, G. W., AND C. F. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72 (6), 1845–1857.[1969]

KAIDO, H., F. MOLINARI, AND J. STOYE (2017): "Confidence Intervals for Projections of Partially Identified Parameters," Working paper, Boston University and Cornell. [1969]

KITAGAWA, T. (2012): "Estimation and Inference for Set-Identified Parameters Using Posterior Lower Probability," Working paper, University College London. [1968]

KLEIJN, B. J. K., AND A. W. VAN DER VAART (2012): "The Bernstein–Von-Mises Theorem Under Misspecification," *Electronic Journal of Statistics*, 6, 354–381.[1990]

KLINE, B., AND E. TAMER (2016): "Bayesian Inference in a Class of Partially Identified Models," *Quantitative Economics*, 6, 329–366.[1968,1980]

LE CAM, L., AND G. L. YANG (1990): *Asymptotics in Statistics: Some Basic Concepts*. Springer.[1991]

LEWANDOWSKI, M., M. RYZNAR, AND T. ZAK (1995): "Anderson Inequality Is Strict for Gaussian and Stable Measures," Proceedings of the American Mathematical Society, 3875–3880.[2009]

LIAO, Y., AND A. SIMONI (2016): "Bayesian Inference for Partially Identified Convex Models: Is It Valid for Frequentist Inference?" Working paper, Rutgers University and CREST. [1968]

LIU, J. S. (2004): *Monte Carlo Strategies in Scientific Computing*. Springer.[1966]

LIU, X., AND Y. SHAO (2003): "Asymptotics for Likelihood Ratio Tests Under Loss of Identifiability," *The Annals of Statistics*, 31 (3), 807–832.[1997]

MOON, H. R., AND F. SCHORFHEIDE (2012): "Bayesian and Frequentist Inference in Partially Identified Models," *Econometrica*, 80 (2), 755–782.[1968,1980,1991,1994]

MÜLLER, U. K. (2013): "Risk of Bayesian Inference in Misspecified Models, and the Sandwich Covariance Matrix," *Econometrica*, 81 (5), 1805–1849.[1991]

NORETS, A., AND X. TANG (2014): "Semiparametric Inference in Dynamic Binary Choice Models," *The Review of Economic Studies*, 81 (3), 1229–1262.[1968]

ROBERT, C. P., AND G. CASELLA (2004): *Monte Carlo Statistical Methods*. New York: Springer.[1966,1990]

ROMANO, J. P., AND A. M. SHAIKH (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78 (1), 169–211.[1969,1974]

ROMANO, J. P., A. M. SHAIKH, AND M. WOLF (2014): "A Practical Two-Step Method for Testing Moment Inequalities," *Econometrica*, 82 (5), 1979–2002.[1969]

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.[1969,1995]

SANTOS SILVA, J. M. C., AND S. TENREYRO (2015): "Trading Partners and Trading Volumes: Implementing the Helpman–Melitz–Rubinstein Model Empirically," *Oxford Bulletin of Economics and Statistics*, 77 (1), 93–105.[1986]

STOYE, J. (2009): "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, 77 (4), 1299–1315.[1969]

VAN DER VAART, A. W. (2000): *Asymptotic Statistics*. Cambridge University Press.[1989,2009]

WAN, Y. (2013): "An Integration-Based Approach to Moment Inequality Models," Department of Economics, University of Toronto, unpublished working paper. [1969]

WEI, Y., M. WAINWRIGHT, AND A. GUNTUBOYINA (2018): "The Geometry of Hypothesis Testing Over Convex Cones: Generalized Likelihood Tests and Minimax Radii," *The Annals of Statistics* (forthcoming).[1990]

---

*Guest Editor Keisuke Hirano handled this manuscript.*