Econometrics and Statistics 000 (2017) 1–13

Contents lists available at ScienceDirect

Econometrics and Statistics

journal homepage: www.elsevier.com/locate/ecosta

On limiting distribution of quasi-posteriors under partial identification

Wenxin Jiang[1,2]

Northwestern University, 2006 Sheridan Rd, Evanston, IL 60208, United States

ARTICLE INFO

Article history:
Received 6 October 2016
Revised 19 March 2017
Accepted 21 March 2017
Available online xxx

Keywords:
Generalized method of moments
Interval data
Moment inequalities
Partial identification
Quasi-posterior
Total variation

ABSTRACT

The limiting distribution (in total variation) is established for the quasi posteriors based on moment conditions, which only partially identify the parameters of interest. Some examples are discussed.

© 2017 EcoSta Econometrics and Statistics. Published by Elsevier B.V. All rights reserved.

1. Introduction

Our paper studies theoretically the large sample behavior of certain Bayesian procedures which are of mutual interest to econometricians and statisticians. In Bayesian procedures, it is well known that the posterior distributions can often be approximated in total variation by normal distributions centered at the frequentist maximum likelihood estimates (see, e.g., a popular textbook account of the Berstein-von Mises theorem in Chapter 10 of van der Vaart, 2000). Econometricians have studied quasi-Bayesian approaches which require less assumptions by only assuming some moment conditions rather than a likelihood function (see, e.g., Kim 2002 and Chernozhukov and Hong, 2003). In this context, similar and more general limiting results in, e.g., Chernozhukov and Hong (2003), suggest that the limiting posterior distribution is typically normal and centered at a corresponding frequentist extremum estimator such as one from the Generalized Method of Moments.

In the models with partial identification, where parameters are not point identified, so that the frequentist extremum estimator is not unique, the asymptotic normal limiting results mentioned before may fail. Such situations of partial identification have generated much interest recently both in statistics (e.g., Gustafson, 2005; 2007; 2015 and in econometrics (e.g., Poirier, 1998; Chernozhukov et al., 2007; Moon and Schorfheide, 2012). The literature either focuses on the inference about the location of the partially identified parameter (e.g., Moon and Schorfheide, 2012; Gustafson, 2015), or on the fully identified set of all possible locations of the parameter (e.g., Chernozhukov et al., 2007). An incomplete sample of applications include missing data, interval censoring, (e.g., Manski and Tamer, 2002; Manski, 2003), game-theoretic models

E-mail address: mwj60201@gmail.com
[1] Taishan Scholar Overseas Distinguished Specialist Adjunct Professor at Shandong University in China.
[2] Professor of Statistics at Northwestern University in U.S.A.

http://dx.doi.org/10.1016/j.ecosta.2017.03.006
2452-3062/© 2017 EcoSta Econometrics and Statistics. Published by Elsevier B.V. All rights reserved.

with multiple equilibria (e.g., Bajari et al., 2007; Ciliberto and Tamer, 2009), auctions (Haile and Tamer, 2003), noncompliance of randomized clinical trials (Gustafson, 2015), and gene environment interactions (Gustafson, 2015).

Our current paper derives and rigorously proves some results on the limiting posterior distribution in presence of partial identification. In addition, we allow quasi-Bayes procedures based on moment conditions. The limit is in the total variation sense, which is a Bernstein-von Mises type result, but not generally asymptotic normal. When data are informative enough only to determine an identification region, instead of a point parameter, our result says that the limiting posterior is related to the prior distribution truncated in (a frequentist estimate of) this identification region.

Our result connects the literature on the inference about the identified parameter set to the inference of the unidentified parameter point, in the sense that one can easily convert a set estimate and combine it with a prior distribution to obtain a large sample approximation of the posterior distribution of the point parameter. This connection may have several meaningful applications.

1. (Simplifying computation) It can be used to avoid lengthy Markov chain Monte Carlo simulation that is typically involved in posterior computations, similar to using the normal approximation in the point-identified situation.
2. (Incorporating prior information.) This is also useful for incorporating prior information to improve the inference results from the more conservative set-based approach. The prior information can be from a same or different study with additional data that are either identifying or partially identifying the parameters of interest. For example, when a small part of the data are exact and all the rest are interval censored, it is obviously not advisable to use the interval data only to estimate an identification region. The exact part of the data can be used to derive a posterior, which can serve as a prior when further incorporating the interval data.
3. (Combining studies.) In the above discussion, we have applied the principle that the prior can be derived from a posterior based on independent data. Multiple applications of this principle can allow meta-analysis (combining results from different studies), sequential computation (for dynamic data flow) or parallel computation (subsetting big data when they are hard to be handled altogether). Our limiting posterior distribution suggests that combining inferences from subsets of data is equivalent to intersecting their resulting identification regions.

1.1. Related works.

A fundamental paper about two decades ago by Poirier (1998) has shown many applications of the Bayesian method in handling problems with partial identification, where data D are informative for only a subset of the parameters, say, λ, out of all the parameters (λ, θ) of the model. This decomposition of (λ, θ) does not have to be the most natural parameterization, but can be achieved by a clever re-parameterization. This situation is further illustrated by a sequence of works by Gustafson (e.g., Gustafson, 2005; 2007; 2015), with many interesting examples. The works of these authors have described that the limiting posterior distribution of $p(\lambda, \theta|D) = p(\lambda|D)p(\theta|\lambda)$ is the product of a usual asymptotic normal distribution on λ, and a conditional prior $p(\theta|\lambda)$ where λ either takes the true value or its maximum likelihood estimate. Recently this limiting result is rigorously proved in total variation distance by Moon and Schorfheide (2012, Theorem 1). The key for this line of existing work is the following assumption:

Assumption (*)(Conditional uninformativeness): There exists a parameterization decomposable into (λ, θ), *such that the likelihood* $p(D|\theta, \lambda) = p(D|\lambda)$ *depends only on* λ, *i.e., D and θ are independent given* λ.

Moon and Schorfheide (2012) call λ the "reduced form" parameter, and θ the "structural" parameter of interest.

The current work aims for generalizing the works of these previous authors and studying the limiting posteriors under partial identification. The generalization is in two important ways:

Generalization (i)(Posterior): We generalize the likelihood-based posterior to be quasi-likelihood-based quasi-posterior, according to a general framework described in Chernozhukov and Hong (2003).

Generalization (ii) (Partial identification): We also allow more general scenarios of partial identification, where no obvious decomposition (λ, θ) can satisfy Assumption (*), so that given λ, the data D are *not* "conditionally" uninformative / independent of the parameter of interest θ.

To be more specific: we allow quasi-likelihood of the form $e^{-nR_n(\lambda,\theta)}$, where R_n is a general empirical risk function that depends on data D, which was $(-1/n)$ times the log-likelihood function in the special case of the usual likelihood. We allow this quasi-likelihood to depend also on θ (unlike in Assumption (*) before), and only assume the following:

Assumption (†)(Marginal uninformativeness): There is a parameterization that can decompose into (λ, θ), *such that the marginal likelihood* $\int e^{-nR_n(\lambda,\theta)}d\lambda$ *is a constant in* θ *(and is therefore "marginally" uninformative).*

It is obvious that this Assumption (†) (marginal uninformativeness) contains Assumption (*)(conditional uninformativeness) as a special case, where $R_n = R_n(\lambda)$ had no dependence on θ; but we will show later that there indeed exist interesting examples of Assumption (†) which do not satisfy Assumption (*).

The current paper is closely related to the paper by Liao and Jiang (2010) and Liao (2010), who also consider a quasi-posterior with partial identification. These works can be viewed as a special case of this current paper (according to the example of Bayesian moment inequalities in Section 2.4), and they have not established the result of convergence in distribution in the total variation distance. Their approach is also fundamentally different: they use a specific (exponential) prior on λ to facilitate the integration over λ, and obtain an expression of the marginal quasi-posterior $p(\theta|data)$ for the parameter of interest θ. Their proof technique then starts with analyzing this specific expression of marginal quasi-posterior

Download English Version:

https://daneshyari.com/en/article/8919509

Download Persian Version:

https://daneshyari.com/article/8919509

Daneshyari.com