Journal of Econometrics 156 (2010) 408–425

Contents lists available at ScienceDirect

Journal of Econometrics

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

EL inference for partially identified models: Large deviations optimality and bootstrap validity

Ivan A. Canay *

Department of Economics, Northwestern University, USA

ARTICLE INFO

Article history:
Received 28 October 2008
Received in revised form
14 November 2009
Accepted 17 November 2009
Available online 2 December 2009

JEL classification:
C12
C14
C21

Keywords:
Empirical likelihood
Partial identification
Large deviations
Empirical likelihood bootstrap
Asymptotic optimality

ABSTRACT

This paper addresses the issue of optimal inference for parameters that are partially identified in models with moment inequalities. There currently exists a variety of inferential methods for use in this setting. However, the question of choosing optimally among contending procedures is unresolved. In this paper, I first consider a canonical large deviations criterion for optimality and show that inference based on the empirical likelihood ratio statistic is optimal. Second, I introduce a new empirical likelihood bootstrap that provides a valid resampling method for moment inequality models and overcomes the implementation challenges that arise as a result of non-pivotal limit distributions. Lastly, I analyze the finite sample properties of the proposed framework using Monte Carlo simulations. The simulation results are encouraging.

1. Introduction

Recently, there have been many papers proposing methodologies for estimation and inference in models where the parameter of interest is not uniquely defined by the economic model and the distribution of the observed data (see, among others, Chernozhukov et al., 2007; Pakes et al., 2005; Romano and Shaikh, 2008, forthcoming; Imbens and Manski, 2004; Rosen, 2008; Beresteanu and Molinari, 2008). When this situation arises the model is said to be partially identified. Given this expanding literature on various inferential methods, it is natural to wonder which method is optimal. This paper addresses the question of optimal inference and contains the following contributions. First, I consider a canonical large deviations criterion for optimality and show that inference based on the empirical likelihood ratio (ELR) statistic is optimal. Second, I introduce a simple and natural modification of the empirical likelihood bootstrap introduced by Brown and Newey (2002) that provides a valid bootstrap method for moment inequality models. This modified empirical likelihood bootstrap is important to overcome the implementation challenges associated with non-pivotal limit distributions in partially identified models. Third, I conduct a Monte Carlo experiment which suggests a finite sample performance advantage to the new bootstrap. These results firmly ground empirical likelihood as an attractive method for inference in moment inequality models.

The problem of optimal inference can be interpreted as a problem of optimal choice of a criterion function. Partially identified models are usually represented via a population objective function $Q(\theta, P_0)$ which does not have a unique minimizer, so that

$$\Theta_0(P_0) = \arg\min_{\theta\in\Theta} Q(\theta, P_0)$$

represents a set containing all the values of θ consistent with the economic model and the distribution P_0. The primary goal is to use a sample analog $\widehat{Q}$ of $Q(\theta, P_0)$ to construct confidence regions that cover each of the elements of $\Theta_0(P_0)$ with a given probability. Most of these models involve a moment inequality condition of the form $\mathbb{E}[m(z, \theta)] \geq 0$ in which case $\Theta_0(P_0)$ is the set of all θ that satisfy the moment condition.[1] In such cases, there are many different choices of $Q(\theta, P_0)$ that have $\Theta_0(P_0)$ as the minimizer set and each choice could lead to different sample analogs and thus different confidence sets. The question of interest is whether there is an optimal criterion function $Q^*(\theta, P_0)$, where optimal means

* Corresponding address: Department of Economics, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208, USA. Tel.: +1 847 491 2929.
E-mail address: iacanay@northwestern.edu.

[1] Some partially identified models cannot be represented as a moment inequality (see Santos, 2006; Chernozhukov et al., 2007).

0304-4076/$ – see front matter
doi:10.1016/j.jeconom.2009.11.009

that inference based on $\widehat{Q}^*$ is more *precise* than inference based on any other sample criterion function. This paper contributes to the growing literature on inference in partially identified models by giving an answer to this question. I introduce empirical likelihood (EL) as a new procedure for partially identified models and show that inference based on the empirical likelihood ratio (ELR) statistic is optimal in a large deviations sense.

The method of empirical likelihood is known to have several optimality properties for models with *equality* moment restrictions. In terms of point estimation, the EL estimator is semiparametrically efficient (i.e., attains the semiparametric efficiency bound derived by Chamberlain, 1987). In addition, this estimator exhibits desirable properties in terms of higher order comparisons (see Newey and Smith, 2004). Regarding inference, DiCiccioet al. (1991) proved that the ELR test admits Bartlett correction, which gives the same accuracy rate as the parametric case, while Kitamura (2001) showed that EL is uniformly most powerful in an Generalized Neyman–Pearson sense for testing moment restrictions. Additional optimality results are presented by Kitamura and Otsu (2005), Kitamura et al. (2009) and Canay and Otsu (2009). This is just a sample of the large list of papers that show some sort of optimality for EL. Kitamura (2006) and Owen (2001) provide additional discussions.

The search for an optimal test in partially identified models involves a number of complications that are not found in the point identified case. The fact that $\Theta_0(P_0)$ is not a singleton complicates the use of local asymptotic optimality notions since standard expansion tools are not as obviously available.[2] Another optimality notion that has been widely applied in point identified models is the large deviations approach. This approach has the virtue of translating naturally to the partially identified setting and is the criterion I pursue here.

The theory of large deviations deals with the behavior of estimators in a fixed neighborhood of the true value. Suppose that there is a statistic T_n that converges in probability to T and let A denote a set such that the closure of A does not contain T. For each n, $\Pr(T_n \in A) \to 0$. In typical cases, $\Pr(T_n \in A) \to 0$ at an exponential rate, i.e. there exists a constant $0 < \eta < \infty$ such that, $n^{-1}\log\Pr(T_n \in A) \to -\eta$. Notice the contrast with conventional local asymptotic theory where the focus is on the behavior of T_n in a shrinking neighborhood of the true parameter value, T. Here the neighborhood A is fixed. For example, let $X_1,\ldots,X_n$ be i.i.d. from $N(0,1)$ and consider the sample mean $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$. Since $\bar{X}_n$ is also normal with zero-mean and variance $1/n$, for any $\delta > 0$,

$$\Pr(|\bar{X}_n| \geq \delta) = 1 - (\sqrt{2\pi})^{-1}\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} e^{-x^2/2}dx \to 0$$

$$\Longrightarrow \frac{1}{n}\log\Pr(|\bar{X}_n| \geq \delta) \to -\frac{\delta^2}{2}. \quad (1.1)$$

Eq. (1.1) is an illustration of a large deviations statement: the typical value of $\bar{X}_n$ is of order $1/\sqrt{n}$, but with small probability (of the order $e^{-n\delta^2/2}$), $\bar{X}_n$ takes values outside a fixed bound. The large deviations behavior of the type I and type II error probabilities associated with a given test procedure gives insight on the good performance of the test: the higher the rate of decrease of these errors, the better the inference. Thus, while standard definitions of efficiency (e.g., Pitman relative efficiency) make the testing problem harder by considering alternative hypothesis that get closer to the null hypothesis as the sample size increases, the efficiency criteria based on large deviations make the problem harder by letting the type I and type II error probabilities go to zero asymptotically. Precise statements are postponed to Section 3, where I show that the ELR test achieves the fastest rates of decrease.

The second main contribution of this paper is related to the actual implementation of the new optimal procedure. I show that under the null hypothesis the ELR statistic converges to a well defined asymptotic distribution. However, whether this limit distribution is useful to calculate critical values depends on the case under consideration. The number of binding constraints – the components of the vector $m(z,\theta)$ with zero expectation – depends crucially on θ and this causes the asymptotic distribution to be non-pivotal. The non-pivotalness is not a barrier in some cases and then one can compute valid critical values using a simple and straightforward approximation. In more complicated setups though, these approximations could be really slack (see Wolak, 1991; Gourieroux et al., 2008) so that using a resampling technique could be desirable. One alternative that many authors have adopted in these types of models is the use of subsampling for the construction of critical values. Subsampling would in fact be valid to approximate the limit distribution of the ELR statistic. Yet, the contribution in this paper lies in a different alternative. I first show that the empirical likelihood bootstrap proposed by Brown and Newey (2002) is not asymptotically valid when applied to moment inequality models.[3] Then I propose a slight modification of that bootstrap, along the lines of the modified parametric bootstrap in Andrews (2000), that does work asymptotically. The modification involves changing the set of inequalities $\mathbb{E}[m(z,\theta)] \geq 0$ by $\mathbb{E}[m(z,\theta)] \geq \varrho_n$, where ϱ_n is a positive sequence that goes to zero asymptotically.[4]

Before proceeding any further, I mention the recent literature that has introduced different techniques to deal with partially identified models and is closely related to the tools presented here. Horowitz and Manski (1998, 2000), Manski and Tamer (2002) and Imbens and Manski (2004) developed methods for estimation and inference for the case where the identification region is defined by lower and upper bounds that can be estimated from the data. For an excellent exposition of such cases see Manski (2003). Going beyond these particular cases, Chernozhukov et al. (2004, 2007), were the first to extend the methodologies to more general setups, defining the identified set as the solution of the minimization of a criterion function and providing several results on estimation and inference on both θ_0 and $\Theta_0(P_0)$ based on subsampling, simulation and the bootstrap. See also Andrews et al. (2004) and Pakes et al. (2005). Romano and Shaikh (2008, forthcoming) carry out a further analysis of the validity of subsampling and present conditions under which the confidence regions cover the parameter of interest uniformly. For additional results on uniform coverage see Soares (2006) and the recent papers by Andrews and Guggenberger (2009) and Andrews and Soares (forthcoming). Rosen (2008) presents a connection between moment inequality models and the literature on one-sided hypothesis testing. As it will be noted in the next section, his Gaussian quasi-likelihood ratio (QLR) statistic is closely related to the empirical likelihood ratio statistic proposed here. Using a different line of analysis Beresteanu and Molinari (2008) propose an inference procedure for partially identified models that can be written as a transformation of an expectation of a set valued random variable. Galichon and Henry (2006a,b) address the choice of the criterion function suggesting the use

[2] This is related to the lack of an asymptotic distribution for sets. The methodology used by Beresteanu and Molinari (2008) is a promising direction for this type of analysis. Also, Andrews and Soares (forthcoming) use local asymptotics to compare the power properties of different critical values for a given criterion function.

[3] Brown and Newey (2002) developed an EL bootstrap for models comprised of moment *equalities*.

[4] This idea is related to the independently derived work by Andrews and Soares (forthcoming) and Bugni (forthcoming, 2009).