EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity
SUPPLEMENTARY APPENDIX ☆

Ivan A. Canay[a,*]

[a] *Department of Economics, Northwestern University, USA*

Abstract

This appendix contains a few results that were not included in the main paper.

Key words: empirical likelihood, partial identification, large deviations, empirical likelihood bootstrap, asymptotic optimality

JEL Classification: C12, C14, C21

☆ First Draft: April 2006. Revised: June 2009

* Corresponding address: Department of Economics, Northwestern University 2001 Sheridan Road, Evanston, Illinois 60208, USA. Tel: +1 847 491 2929

Email addresses: iacanay@northwestern.edu (Ivan A. Canay)

Appendices

Throughout the appendix let $m(z_i,\theta)' = [m_b(z_i,\theta)'\ m_s(z_i,\theta)']$ where $m_b(z_i,\theta)$ is a $b(\theta)\times 1$ vector of moments with zero mean, $\mathbb{E}_{P_0}[m_b(z_i,\theta)] = 0$, and $m_s(z_i,\theta)$ is a $s(\theta)\times 1$ vector of moments with positive mean, $\mathbb{E}_{P_0}[m_s(z_i,\theta)] > 0$. I use $\mathbb{E}\equiv\mathbb{E}_{P_0}$ in those cases where P_0 is understood.

Appendix D Parameter on the Boundary Example

Consider the example in Andrews (2000, page 401) where $X_i \sim N(\mu,1)$ and $\mu\geq 0$. The EL estimator of μ that solves,

$$l^r_{EL} = \sup_{p_1,\dots,p_n}\left\{\sum_{i=1}^n \log(p_i)\,\middle|\, p_i>0; \sum_{i=1}^n p_i = 1; \sum_{i=1}^n p_i x_i \geq 0\right\} \tag{D-1}$$

is $\widetilde{\mu}_n = \sum \widetilde{p}_i X_i = \max(\overline{X}_n, 0)$ where $\widetilde{P}_n = (\widetilde{p}_1,\dots,\widetilde{p}_n)$ are the EL probabilities. Thus,

$$\sqrt{n}(\widetilde{\mu}_n - \mu) \rightsquigarrow \begin{cases} \mathbb{Z} & \text{if } \mu>0 \\ \max(\mathbb{Z},0) & \text{if } \mu = 0\end{cases} \tag{D-2}$$

where $\mathbb{Z}\sim N(0,1)$. The goal is to use the bootstrap to approximate the distribution of $\sqrt{n}(\widetilde{\mu}_n-\mu)$. Andrews shows that the nonparametric and parametric bootstrap are invalid asymptotically. In this appendix I show that while the standard EL bootstrap is not consistent either, the modified EL bootstrap introduced in section 4 of the paper is.

Let X^*_i be i.i.d. according to $\widetilde{P}_n$. The bootstrap EL estimator that solves (D-1) with X replaced by X^* is $\widetilde{\mu}^*_n = \sum\widetilde{p}^*_i X^*_i = \max(\overline{X}^*_n, 0)$. Note that $\sqrt{n}(\overline{X}^*_n - \widetilde{\mu}_n)\rightsquigarrow\mathbb{Z}$ by a triangular array CLT. As in Andrews, let $B_c = \{\omega : \limsup_{n\to\infty}\sqrt{n}\overline{X}_n > c\}$ and note that by the law of iterated logarithm, $P(B_c) = 1$. Now, let $\mu = 0$ and w.l.o.g consider a sequence such that $\sqrt{n}\overline{X}_n\geq c$ for all n,

$$\begin{aligned}
\sqrt{n}(\widetilde{\mu}^*_n - \widetilde{\mu}_n) &= \max(\sqrt{n}\overline{X}^*_n, 0) - \sqrt{n}\widetilde{\mu}_n \\
&= \max(\sqrt{n}(\overline{X}^*_n - \widetilde{\mu}_n), -\max(\sqrt{n}\overline{X}_n, 0)) \\
&\leq \max(\sqrt{n}(\overline{X}^*_n - \widetilde{\mu}_n), -c) \\
&\rightsquigarrow \max(\mathbb{Z}, -c) \quad \text{conditional on } \{\widetilde{P}_n : n\geq 1\} \\
&\leq \max(\mathbb{Z}, 0)\ .
\end{aligned}$$

Thus, the asymptotic distribution of $\sqrt{n}(\widetilde{\mu}^*_n - \widetilde{\mu}_n)$ is inconsistent with probability 1 and the EL bootstrap proposed by Brown and Newey (2002) is not asymptotically valid.

Consider the modified EL bootstrap instead. Let ϱ_n satisfy (4.2) and define the modified EL probabilities $\overline{P}_n = (\overline{p}_1,\dots,\overline{p}_n)$ as those solving,

$$\widetilde{l}^r_{EL}(\varrho_n) = \sup_{p_1,\dots,p_n}\left\{\sum_{i=1}^n \log(p_i)\,\middle|\, p_i>0; \sum_{i=1}^n p_i = 1; \sum_{i=1}^n p_i x_i \geq \varrho_n\right\}$$

so that $\overline{\mu}_n = \max(\overline{X}_n, \varrho_n)$. Let X^*_i be i.i.d. according to $\overline{P}_n$ and $\overline{\mu}^*_n = \max(\overline{X}^*_n, \varrho_n)$ be the modified EL bootstrap estimator of μ. Equation (4.2) and the law of iterated logarithm imply that,

$$P\left(\limsup_{n\to\infty}(\overline{X}_n - \varrho_n)\leq 0\right) = \begin{cases} 0 & \text{if } \mu>0 \\ 1 & \text{if } \mu=0\end{cases},$$

and this is enough to get asymptotic validity for $\sqrt{n}(\overline{\mu}^*_n - \overline{\mu}_n)$. When $\mu = 0$ we have for n sufficiently large that $\overline{X}_n \le \varrho_n$ wp1 and hence $\overline{\mu}_n = \varrho_n$ wp1 so that,

$$\begin{aligned}\sqrt{n}(\overline{\mu}_n^* - \overline{\mu}_n) &= \max\{\sqrt{n}(\overline{X}_n^* - \overline{\mu}_n), \sqrt{n}(\varrho_n - \overline{\mu}_n)\} \\ &= \max\{\sqrt{n}(\overline{X}_n^* - \overline{\mu}_n), 0\} \\ &\rightsquigarrow \max\{\mathbb{Z}, 0\} \quad \text{conditional on } \{\overline{P}_n : n \ge 1\}.\end{aligned}$$

Therefore, the modified EL bootstrap is asymptotically valid when $\mu = 0$. In addition, when $\mu > 0$, $\overline{X}_n > \varrho_n$ for n large enough wp1 and then $\overline{\mu}_n = \overline{X}_n$ for n large wp1 so that

$$\begin{aligned}\sqrt{n}(\overline{\mu}_n^* - \overline{\mu}_n) &= \max\{\sqrt{n}(\overline{X}_n^* - \overline{\mu}_n), \sqrt{n}(\varrho_n - \overline{\mu}_n)\} \\ &= \max\{\sqrt{n}(\overline{X}_n^* - \overline{\mu}_n), -\sqrt{n}(\overline{X}_n - \varrho_n)\} \\ &\rightsquigarrow \mathbb{Z} \quad \text{conditional on } \{\overline{P}_n : n \ge 1\}.\end{aligned}$$

The last step follows from $-\sqrt{n}(\overline{X}_n - \varrho_n) \to -\infty$ wp1. Therefore, the modified EL bootstrap is also asymptotically valid when $\mu > 0$.

Appendix E Delta Optimality of Empirical Likelihood

This appendix shows an alternative way to show that empirical likelihood is optimal for testing moment inequalities in a Generalized Neyman Pearson (GNP) sense. As mentioned in the paper, the recent work by Kitamura, Santos, and Shaikh (2009) shows that in order to explore large deviations optimality in a semi-parametric context it is important to control for ill-behaved distributions. This is so because for most commonly-used tests, it is always possible to find a distribution in $\mathcal{P}_0(\theta)$ such that the rate at which the type I error probability goes to zero is arbitrarily small (see Kitamura, Santos, and Shaikh, 2009, for a discussion and examples). Taking this into consideration, one can introduce restrictions to the set of null distributions and proceed from there without changing the test statistic (and this is the approach taken in the paper). A completely different way involves introducing a δ-perturbation to the test, keeping the whole set of null distributions. This approach follows Zeitouni and Gutman (1991) and uses the notion of δ-optimality. According to this notion of optimality, the GNP optimal test is "based" on empirical likelihood. Therefore, one possible reading will be that the ELR statistic is arbitrarily close to the optimal test. However, in order to achieve the optimal bound, some perturbation to the ELR statistic is needed. As in the paper, I use the following assumptions.

Assumption E.1 *$\mathcal{Z}$ and Θ are compact subsets of $\mathbb{R}^d$ and $\mathbb{R}^k$.*

Assumption E.2 *$m(z,\theta) : \mathcal{Z} \times \Theta \mapsto \mathbb{R}^q$ is continuous in z for each $\theta \in \Theta$.*

Remark E.1 If $\mathcal{Z}$ is compact so is $\mathcal{M}(\mathcal{Z})$ and therefore any subset of probability measures is pre-compact, including in particular,

$$\{Q \in \mathcal{M} : \inf_{P \in \mathcal{P}(Q,\theta)} I_P(Q) < \eta\} \tag{E-1}$$

where $I_P(Q) = I(Q||P)$. Also, the set of null distributions $\mathcal{P}_0(\theta)$ is closed in this case (by the Portmanteau Lemma) and therefore compact.

To define the optimal test first define, we need some notation. For a subset $A \in \mathcal{X}$ define a ν-neighborhood of A as,

$$A^\nu \equiv \{x \in X : d(x,A) < \nu\} = \cup_{a\in A} B(a,\nu), \tag{E-2}$$

so that we can write the Levy metric as,

$$d(Q,P) \equiv \inf\{\nu \in \mathbb{R} : Q(F) \le P(F^\nu) + \nu, \quad \forall F \in \mathcal{X} \text{ closed}\}. \tag{E-3}$$

Then, let

$$\bar{B}(Q, 2\delta) \equiv \{J \in \mathcal{M} : d(Q, J) \leq 2\delta\} \tag{E-4}$$

be a closed ball around Q. We can now define,

$$I_\delta^\theta(Q) \equiv \inf_{P \in \mathcal{P}(J,\theta)} \inf_{J \in \bar{B}(Q,2\delta)} I_P(J), \tag{E-5}$$

where

$$\mathcal{P}(J, \theta) \equiv \{Q \in \mathcal{P}_0(\theta) : Q << J, J << Q\}. \tag{E-6}$$

Definition E.1 *The δ-smoothed Empirical Likelihood Ratio test is the test with rejection region,*

$$\Lambda_1^\theta \equiv \{Q \in \mathcal{M} : I_\delta^\theta(Q) \geq \eta^\theta\}. \tag{E-7}$$

Remark E.2 Note that if $\delta = 0$, we get the original ELR statistic as,

$$I_0^\theta(Q) = I^\theta(Q) \equiv \inf_{P \in \mathcal{P}(Q,\theta)} I_P(Q). \tag{E-8}$$

Finally, for any partition $\Omega_n^\theta \equiv (\Omega_{n,0}^\theta, \Omega_{n,1}^\theta)$ of the space $\mathcal{M}$ let $\Omega_n^{\theta,\delta}$ denote the δ-smoothing defined by

$$\Omega_{n,1}^{\theta,\delta} \equiv \{J \in \mathcal{M} : d(J, \Omega_{n,1}^\theta) < \delta\}, \quad \Omega_{n,0}^\theta \equiv \mathcal{M} \setminus \Omega_{n,1}^{\theta,\delta}. \tag{E-9}$$

The next theorem shows that the above test is asymptotically optimal in a GNP sense.

Theorem E.1 *Let Assumptions (E.1) and (E.2) hold. For any $0 < \delta < 1/2$ and any $\eta^\theta > 0$ the following statements are true.*

1. *The δ-smoothed Empirical Likelihood Ratio test satisfies,*

$$\sup_{P \in \mathcal{P}_0(\theta)} \limsup_{n\to\infty} \frac{1}{n} \log P^n(\hat{P}_n \in \Lambda_1^{\theta,\delta}) \leq -\eta^\theta \tag{E-10}$$

2. *If the alternative test $\Omega_n^\theta \equiv (\Omega_{n.0}^\theta, \Omega_{n,1}^\theta)$ satisfies,*

$$\sup_{P \in \mathcal{P}_0(\theta)} \limsup_{n\to\infty} \frac{1}{n} \log P^n(\hat{P}_n \in \Omega_{n,1}^{\theta,c\delta}) \leq -\eta^\theta \tag{E-11}$$

for any $c > 1$, it follows that,

$$\limsup_{n\to\infty} \frac{1}{n} \log P^n(\hat{P}_n \in \Omega_{n,0}^{\theta,\delta}) \geq \limsup_{n\to\infty} \frac{1}{n} \log P^n(\hat{P}_n \in \Lambda_0^{\theta,\delta}) \tag{E-12}$$

for any $P_1 \in \mathcal{M}$.

Proof. To prove (1) we just need to use Sanov's Theorem. Define,

$$\eta^\delta(P) \equiv \inf_{Q \in \overline{\Lambda_1^{\theta,\delta}}} I_P(Q), \quad \forall P \in \mathcal{P}_0(\theta), \tag{E-13}$$

and note that,

$$\sup_{P \in \mathcal{P}_0(\theta)} \limsup_{n\to\infty} \frac{1}{n} \log P^n(\hat{P}_n \in \Lambda_1^{\theta,\delta}) \leq \sup_{P \in \mathcal{P}_0(\theta)} - \inf_{Q \in \overline{\Lambda_1^{\theta,\delta}}} I_P(Q) \tag{E-14}$$

$$= - \inf_{P \in \mathcal{P}_0(\theta)} \eta^\delta(P). \tag{E-15}$$

The result would then follow by proving that,

$$\inf_{P\in\mathcal{P}_0(\theta)} \eta^{\delta}(P) \geq \eta^{\theta}. \tag{E-16}$$

Take any $P \in \mathcal{P}_0(\theta)$. By Dupuis and Ellis (1997, Lemma 1.4.3(b)), $I_P(Q)$ is lower semi continuous and by Assumption (E.1), $\overline{\Lambda_1^{\theta,\delta}}$ is compact, so that there exists $J_P \in \overline{\Lambda_1^{\theta,\delta}}$ such that $\eta^{\delta}(P) = I_P(J_P)$. By the definition of $\Lambda_1^{\theta,\delta}$, there exists a $Q \in \mathcal{M}$ such that $J_P \in \bar{B}(Q, 2\delta)$ and $I_\delta^\theta(Q) \geq \eta^\theta$. It follows that,

$$\eta^{\delta}(P) = I_P(J_P) \geq \inf_{P\in\mathcal{P}(J,\theta)} \inf_{J\in\bar{B}(Q,2\delta)} I_P(J) = I_\delta^\theta(Q) \geq \eta, \tag{E-17}$$

for all $P \in \mathcal{P}_0(\theta)$. Since $\mathcal{P}_0(\theta)$ is closed, the result follows.

To prove (2) the basic idea is to prove that for each $\theta \in \Theta$ there exists $n(\delta) \in \mathbb{N}$ such that

$$\Lambda_0^{\theta,\delta} \subseteq \Omega_{n,0}^{\theta,\delta} \tag{E-18}$$

for all $n > n(\delta)$. Suppose it is not so. Then there exists an infinite sequence of measures $\{Q_n\}_{n\in\mathbb{N}}$ such that $Q_n \in \Lambda_0^{\theta,\delta}$ and $Q_n \in \Omega_{n,1}^{\theta,\delta}$. Since $\mathcal{M}$ is compact in the weak topology by Assumption (E.1), there exists a subsequence $\{n_k\}_{k\in\mathbb{N}}$ such that $Q_{n_k} \rightsquigarrow \bar{Q} \in \mathcal{M}$. For a given $c > 1$, let $\delta' = c\delta$ and $\tilde{\delta} = (\delta' - \delta)/2$. It follows that there exists k_0 such that for all $k \geq k_0$,

$$Q_{n_k} \in B(\bar{Q}, \tilde{\delta}) \subset \Omega_{n_k,1}^{\theta,\delta'}. \tag{E-19}$$

We can then use Sanov's Theorem to note that,

$$\begin{aligned}
\sup_{P\in\mathcal{P}_0(\theta)} \limsup_{n\to\infty} n^{-1} \log P^n(\hat{P}_n \in \Omega_{n,1}^{\theta,\delta'}) &\geq \sup_{P\in\mathcal{P}_0(\theta)} \liminf_{n_k\to\infty} \frac{1}{n_k} \log P^{n_k}(\hat{P}_{n_k} \in \Omega_{n_k,1}^{\theta,\delta'}) \\
&\overset{(1)}{\geq} \sup_{P\in\mathcal{P}_0(\theta)} \liminf_{n_k\to\infty} \frac{1}{n_k} \log P^{n_k}(\hat{P}_{n_k} \in B(\bar{Q}, \tilde{\delta})) \\
&\overset{(2)}{\geq} \sup_{P\in\mathcal{P}_0(\theta)} - \inf_{J\in B(\bar{Q},\tilde{\delta})} I(J|P) \\
&\geq - \inf_{P\in\mathcal{P}_0(\theta)} I_P(Q_{n_{k_0}}),
\end{aligned} \tag{E-20}$$

where $\overset{(1)}{\geq}$ follows from $B(\bar{Q}, \tilde{\delta}) \subset \Omega_{n_k,1}^{\theta,\delta'}$ and $\overset{(2)}{\geq}$ follows from Sanov's theorem. Finally, since $Q_{n_{k_0}} \in \Lambda_0^{\theta,\delta}$ it follows that

$$\inf_{P\in\mathcal{P}_0(\theta)} I_P(Q_{n_{k_0}}) \leq \inf_{P\in\mathcal{P}(Q_{n_{k_0}},\theta)} I_P(Q_{n_{k_0}}) \leq \inf_{P\in\mathcal{P}(J,\theta)} \inf_{J\in\bar{B}(Q_{n_{k_0}},2\delta)} I_P(J) = I_\delta^\theta(Q_{n_{k_0}}) < \eta^\theta \tag{E-21}$$

which together with (E-20) contradicts

$$\limsup_{n\to\infty} n^{-1} \log P^n(\hat{P}_n \in \Omega_{n,1}^{\theta,\delta'}) \leq -\eta^\theta.$$

Therefore, $\Lambda_0^{\theta,\delta} \subseteq \Omega_{n,0}^{\theta,\delta}$ for all $n > n(\delta)$ and this implies,

$$\limsup_{n\to\infty} n^{-1} \log P^n(\hat{P}_n \in \Omega_{n,1}^{\theta,\delta'}) \geq \limsup_{n\to\infty} n^{-1} \log P^n(\hat{P}_n \in \Lambda_0^{\theta,\delta})$$

■

References

Andrews, D. W. K., 2000. Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. Econometrica 68 (2), 399–405.

Brown, B. W., Newey, W., 2002. Generalized method of moments, efficient bootstrapping, and improved inference. Journal of Business and Economics Statistics 20 (4), 507–517.

Dupuis, P., Ellis, R. S., 1997. A Weak Convergence Approach to the Theory of Large Deviations. Wiley Series in Probability and Statistics: New York.

Kitamura, Y., Santos, A., Shaikh, A., 2009. On the asymptotic optimality of empirical likelihood for testing moment restrictions, manuscript, University of Chicago.

Zeitouni, O., Gutman, M., 1991. On universal hypothesis testing via large deviations. IEEE Transactions on Information Theory 37 (2), 285–290.