2011s-61

Set Coverage and Robust Policy

Marc Henry, Alexei Onatski

Série Scientifique
Scientific Series

Montréal
Octobre 2011

CIRANO
Allier savoir et décision

Centre interuniversitaire de recherche en analyse des organisations

Partenaire financier

Développement
économique, Innovation
et Exportation
Québec

Set Coverage and Robust Policy [*]

Marc Henry[†], Alexei Onatski[‡]

Résumé / *Abstract*

We show that confidence regions covering the identified set may be preferable to confidence regions covering each of its points in robust control applications.

Mots clés : partial identification, set coverage, maxmin expected utility, robust control.

Codes JEL : C44

[†] Departement de sciences économiques, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Québec, H3C 3J7, Canada. marc.henry@umontreal.ca.
[‡] Faculty of Economics, Cambridge University, Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DD, UK, ao319@cam.ac.uk.

2 MARC HENRY (UNIVERSITY OF MONTREAL) ALEXEI ONATSKI (UNIVERSITY OF CAMBRIDGE)

When conducting inference on partially identified parameters, Imbens and Manski (2004) pointed out that confidence regions may cover the whole identified set with a prescribed probability, to which we shall refer as *set coverage*, or they may cover each of its points with a prescribed probability, to which we shall refer as *point coverage*. Since set coverage implies point coverage, confidence regions satisfying point coverage are generally preferred on the grounds that they may be more informative. The object of this note is to describe a decision problem in which, contrary to received wisdom, point coverage is clearly undesirable.

Consider a random vector $s = (X, \varepsilon)$ on $\{1, \ldots, N\}$. Call realizations s_i, $i = 1, \ldots, N$, of this random vector states of the world, and call their collection $S = \{s_1, \ldots, s_N\}$. Suppose states of the world are partially observable, by which we mean that the realizations $\{x_1, \ldots, x_N\}$ of X can be observed over repeated experiments, but not the realizations $\{\varepsilon_1, \ldots, \varepsilon_N\}$ of ε. Call P_X the probability mass function of random vector X. Let Θ be a set of models for the states, defined by the fact that for each $\theta \in \Theta$, P_θ denotes a probability mass function for the random vector (X, ε). The identified set Θ_I is defined in the following way:

$$\Theta_I = \left\{ \theta \in \Theta : \sum_{j=1}^{N} P_\theta(x_i, \varepsilon_j) = P_X(x_i) \text{ for } i = 1, \ldots, N \right\}.$$

More generally, any additional a priori restriction on the joint distribution of (X, ε) can be incorporated in the definition of the identified set.

Suppose a decision maker may choose among actions in a set $\mathcal{A} = \{a_1, \ldots, a_K\}$. The actions may be treatments, as in Manski (2004), or policy controls as in Brainard (1967). Actions in $\mathcal{A}$ are defined as functions from S to real valued outcomes. Call $U(a, \theta)$ the ex-ante utility of the decision maker, when she knows P_θ to be the true data generating process for (X, ε). Typically, this will be von Neumann-Morgenstern expected utility $U(a, \theta) = \int a(s) dP_\theta(s)$. We shall consider two robust decision making procedures based on the identified set: (i) maxmin, where the decision maker maximizes the functional evaluation $V(a) = \min_{\Theta_I} U(a, \theta)$ over $\mathcal{A}$ and (ii) minmax regret, where the decision maker maximizes $V(a) = \min_{\Theta_I}[U(a, \theta) - \max_{a \in \mathcal{A}} U(a, \theta)]$ over $\mathcal{A}$. The arguments we make do not depend on which of the two options (i) or (ii) is chosen, so we shall concentrate on a maxmin decision maker.

The decision maker is supposed to have access to two types of confidence regions for Θ_I based on repeated sampling in the state space. A region covering the identified set called Θ_{SC} such that $\mathbb{P}(\Theta_I \in \Theta_{SC}) = 1 - \alpha$ and a region covering each point of the identified set called Θ_{PC} such that $\min_{\theta \in \Theta_I} \mathbb{P}(\theta \in \Theta_{PC}) = 1 - \alpha$. Without necessarily subscribing to the learning model of Epstein and

Schneider (2007), we appeal to Anderson, Hansen, and Sargent (2003) and "assume that our decision maker worries about alternative models that available data cannot readily dispose of". Hence, the decision maker considers two decision rules based on the two respective confidence regions. The decision rule based on Θ_{SC} consists in choosing $\hat{a}_{SC}$ in $\mathcal{A}$ to maximize $\min_{\theta \in \Theta_{SC}} U(a, \theta)$ and the decision rule based on Θ_{PC} consists in choosing $\hat{a}_{PC}$ in $\mathcal{A}$ that maximizes $\min_{\theta \in \Theta_{PC}} U(a, \theta)$. The decision rule based on Θ_{SC} is robust in the sense that

$$\mathbb{P}\left(\min_{\theta \in \Theta_I} U(\hat{a}_{SC}, \theta) \geq \min_{\theta \in \Theta_{SC}} U(\hat{a}_{SC}, \theta)\right) \geq 1 - \alpha,$$

so that $\min_{\theta \in \Theta_{SC}} U(a, \theta)$ provides a lower bound for the actual utility functional $V(\hat{a}_{SC})$ with probability at least as large as $1 - \alpha$. The decision rule based on Θ_{PC}, however, is not robust as will be shown with the following example that we contrived in the simplest possible way for expositional purposes.

Let $\{1, \ldots, N\}$ be a population of individuals and let $X \in \{F, M\}$ be their gender and $\varepsilon \in \{T, NT\}$ be their talent (T for *talented* and NT for *not so talented*). Half the population is male and half the population is talented, but the correlation θ between talent and gender is unknown. The decision maker is a social planner who can offer an education opportunity to women only (action a_1), to men only (action a_2) or to everyone (action a_3). The net benefit of offering the education opportunity to a talented person is B. The net benefit of offering the education opportunity to a not so talented person is $-B$ (wasted resources). The net benefit of failing to offer the education opportunity to a not so talented person is zero. Finally, the net benefit of failing to offer the education opportunity to a talented person is $-B$ (wasted talent). Assume that the parameter set is equal to $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$, where under θ_1 all talent is male, and under θ_2 all talent is female, under θ_3 everyone is talented and under θ_4 no one is talented. Given the a priori constraints on the joint distribution of gender and talent, the identified set is $\Theta_I = \{\theta_1, \theta_2\}$. Now, with von Neumann-Morgenstern expected utility, we have:

$$U(a, \theta_1) = \frac{1}{2} U(a(F, NT)) + \frac{1}{2} U(a(M, T)) = \begin{cases} -B & \text{if} \quad a = a_1 \\ B/2 & \text{if} \quad a = a_2 \\ 0 & \text{if} \quad a = a_3 \end{cases}$$

and

$$U(a, \theta_2) = \frac{1}{2} U(a(F, T)) + \frac{1}{2} U(a(M, NT)) = \begin{cases} B/2 & \text{if} \quad a = a_1 \\ -B & \text{if} \quad a = a_2 \\ 0 & \text{if} \quad a = a_3 \end{cases}$$

MARC HENRY (UNIVERSITY OF MONTREAL) ALEXEI ONATSKI (UNIVERSITY OF CAMBRIDGE)

Note that $\mathbb{P}(\theta_1 \notin \Theta_{SC} \text{ or } \theta_2 \notin \Theta_{SC}) = \alpha$, whereas for Θ_{PC} we only require that either $\mathbb{P}(\theta_1 \notin \Theta_{PC}) \leq \alpha$ and $\mathbb{P}(\theta_2 \notin \Theta_{PC}) = \alpha$, or $\mathbb{P}(\theta_1 \notin \Theta_{PC}) = \alpha$ and $\mathbb{P}(\theta_2 \notin \Theta_{PC}) \leq \alpha$. If Θ_{PC} is more informative than Θ_{SC}, then $2\alpha \geq \mathbb{P}(\theta_1 \notin \Theta_{PC} \text{ or } \theta_2 \notin \Theta_{PC}) > \alpha$. Take the case $\theta_1 \notin \Theta_{PC}$ and suppose for clarity that $\Theta_{PC} = \{\theta_2\}$, then $\min_{\theta \in \Theta_{PC}} U(a, \theta) = B/2$ if $a = a_1$, $-B$ if $a = a_2$ and 0 if $a = a_3$; and symmetrically if $\theta_2 \notin \Theta_{PC}$. Hence, when $\theta_j \notin \Theta_{PC}$, the action that maximizes $\min_{\theta \in \Theta_{PC}} U(a, \theta)$ is $\hat{a}_{PC} = a_j$ and $\min_{\theta \in \Theta_I} U(\hat{a}_{PC}, \theta) = \min_{\theta \in \Theta_I} U(a_j, \theta) = -B$ which can be much smaller than $\min_{\Theta_{PC}} U(a_j, \theta) = B/2$. Hence the action taken on the basis of the region with point coverage yields a utility that may be much smaller than it appears with a probability strictly larger than α. In contrast $\hat{a}_{SC} = a_3$ with probability at least $1 - \alpha$ so $\min_{\Theta_{SC}} U(\hat{a}_{SC}, \theta) = 0$ and $\min_{\Theta_I} U(\hat{a}_{SC}, \theta) = 0$ with probability at least $1 - \alpha$, so that decision based on the region providing set coverage does not suffer from the same lack of robustness.

REFERENCES

ANDERSON, E., L. HANSEN, and T. SARGENT (2003): "A quartet of semigroups for model specification, robustness, prices of risk, and model detection," *Journal of the European Economic Association*, pp. 68–123.

BRAINARD, W. (1967): "Uncertainty and the Effectiveness of Policy," *American Economic Review*, pp. 411–425.

EPSTEIN, L., and M. SCHNEIDER (2007): "Learning under ambiguity," *Review of Economic Studies*, 74, 1275–1303.

IMBENS, G., and C. MANSKI (2004): "Confidence intervals for partially identified parameters," *Econometrica*, 72, 1845–1859.

MANSKI, C. (2004): "Statistical treatment rules for heterogeneous populations," *Econometrica*, pp. 1221–1246.