Statistical Inference for Partially Identified Parameters

Jörg Stoye
New York University
ISIPTA 2009 – Durham, UK

July, 2009

Partial Identification: A School of Thought in Econometrics

Charles Manski (Northwestern)

Elie Tamer (Northwestern)

Francesca Molinari (Cornell)

Adam Rosen (University College London)

Jörg Stoye (NYU)

Marc Henry (Montréal)

Alfred Galichon (Polytechnique)

Thierry Magnac (Toulouse)

Identification (Identifiability) in Statistics

Identifiability: What does the sampling process reveal?

A parameter value is identified if it is a function of the distribution of observables.

If a parameter is not identifiable, estimation and inference are badly posed problems.

But:

Even if a parameter is not identifiable, we might be able to learn something from the data?

Partial Identification:

Nontrivial (identified) bounds on the parameter as a middle ground between identified and not identified.

What's new about this?

Identifiable bounds on unidentified quantities go back at least to Fisher.

The new idea is to generally **rethink identification as a non-binary event**, with "point identification" as a special case.

Partial Identification:

Nontrivial (identified) bounds on the parameter as a middle ground between identified and not identified.

What's new about this?

Identifiable bounds on unidentified quantities go back at least to Fisher.

The new idea is to generally **rethink identification as a non-binary event**, with "point identification" as a special case.

Many natural applications in economics and social sciences:

- missing data,
- observational studies (selection bias),
- problems where joint distributions are of interest but only marginal distributions are observed (missing copulae),
- games with multiple equilibria....

Partial Identification:

Nontrivial (identified) bounds on the parameter as a middle ground between identified and not identified.

What's new about this?

Identifiable bounds on unidentified quantities go back at least to Fisher.

The new idea is to generally **rethink identification as a non-binary event**, with "point identification" as a special case.

Many natural applications in economics and social sciences:

- missing data.
- observational studies (selection bias),
- problems where joint distributions are of interest but only marginal distributions are observed (missing copulae),
- games with multiple equilibria....

Ideally, such analysis requires to adapt and extend the complete toolkit of estimation and inference.

This is an active project. The poster showcases some contribution to it.

Relation to Interval Probabilities

Partial Identification captures limits to knowledge in the sense that the identified object – what we could ideally learn about a random process – is not a probability but a set of probabilities.

This set of probabilities has an **objective interpretation**; it is not "degrees of belief."

Conditional on this limitation, partial identification analysis aims to perform standard statistics. For example, sampling distributions are conventional probabilities.

A Simple Example: Mean with Missing Data

Want to know survival probability from a treatment.

10% of treatment subject vanish from the sample. 80% of observable subjects survive. Do not want to make assumptions like "ignorable attrition."

Ignoring sampling variation, we then know that expected survival lies in $[72\%, 82\%]$.

A Simple Example: Mean with Missing Data

Want to know survival probability from a treatment.

10% of treatment subject vanish from the sample. 80% of observable subjects survive. Do not want to make assumptions like "ignorable attrition."

Ignoring sampling variation, we then know that expected survival lies in $[72\%, 82\%]$.

In this example:

- true parameter value of interest = θ_0 = aggregate survival probability (might be 75%,say)
- identified set = $\Theta_0 = [72\%, 82\%]$

The identified set is learnable in the sense that a consistent estimator exists. The true parameter value is not. We can define a confidence region though.

Generalizing Estimation and Inference

Estimation is conceptually simple:

- An estimator of Θ_0 must be consistent in the Hausdorff metric.
  (Achieving this can be tricky, though not in the present example.)

- An estimator of θ_0 cannot be consistent in an interesting way.

Generalizing Estimation and Inference

Inference raises new conceptual questions: How to define confidence regions?

Should we cover the identified set:

$$\Pr(\Theta_0 \subseteq CI) \to 95\%$$

or the partially identified parameter

$$\inf_{\theta_0 \in \Theta_0} \Pr(\theta_0 \in CI) \to 95\%.$$

This depends on what is seen as quantity of fundamental interest.

Generalizing Estimation and Inference

Inference raises new conceptual questions: How to define confidence regions?

Should we cover the identified set:

$$\Pr(\Theta_0 \subseteq CI) \to 95\%$$

or the partially identified parameter

$$\inf_{\theta_0 \in \Theta_0} \Pr(\theta_0 \in CI) \to 95\%.$$

This depends on what is seen as quantity of fundamental interest.

- "Partial Identification as limits to knowledge" $\to$ cover the set
- "Partial Identification as data constraints" $\to$ cover the parameter

But there are also technical issues.

A Naive Approach

Let $\Theta_0 = [\theta_L, \theta_U]$, where estimators $\left(\widehat{\theta}_L, \widehat{\theta}_U\right)$ exist and have standard errors (σ_L, σ_U).

Then

$$\inf_{\theta_0 \in \Theta_0} \Pr\left(\theta_0 \in \left[\widehat{\theta}_L - 1.64\frac{\sigma_L}{\sqrt{n}}, \widehat{\theta}_U + 1.64\frac{\sigma_U}{\sqrt{n}}\right]\right) \to 95\%.$$

A construction that would usually be thought of as 90% CI is 95%!

A Naive Approach

Let $\Theta_0 = [\theta_L, \theta_U]$, where estimators $\left(\widehat{\theta}_L, \widehat{\theta}_U\right)$ exist and have standard errors (σ_L, σ_U).

Then

$$\inf_{\theta_0 \in \Theta_0} \Pr\left(\theta_0 \in \left[\widehat{\theta}_L - 1.64\frac{\sigma_L}{\sqrt{n}}, \widehat{\theta}_U + 1.64\frac{\sigma_U}{\sqrt{n}}\right]\right) \to 95\%.$$

A construction that would usually be thought of as 90% CI is 95%!

Reason: As $n \to \infty$, noncoverage risk is one-sided. Whatever the true value is, at least one end of the identified set is "far away" in the limit.

A Naive Approach

Let $\Theta_0 = [\theta_L, \theta_U]$, where estimators $\left(\widehat{\theta}_L, \widehat{\theta}_U\right)$ exist and have standard errors (σ_L, σ_U).

Then

$$\inf_{\theta_0 \in \Theta_0} \Pr\left(\theta_0 \in \left[\widehat{\theta}_L - 1.64\frac{\sigma_L}{\sqrt{n}}, \widehat{\theta}_U + 1.64\frac{\sigma_U}{\sqrt{n}}\right]\right) \to 95\%.$$

A construction that would usually be thought of as 90% CI is 95%!

Reason: As $n \to \infty$, noncoverage risk is one-sided. Whatever the true value is, at least one end of the identified set is "far away" in the limit.

But: This is true pointwise, i.e. holding $\theta_L < \theta_U$ fixed and letting $n \to \infty$.

It is not true uniformly: With $\theta_U - \theta_L = O(1/\sqrt{n})$, intuition and result fail.

This uniformity failure is bad: When point identification is approached, the interval fails.

How to Fix This

To fix the problem, must copy tricks from the pre-testing/model selection literature.

Intuitively, use data to select between a fully identified and a partially identified model.

Catch: This pre-test cannot have sufficient accuracy to vanish from the asymptotics. Therefore, must induce a conservative bias.

This author’s contribution is to show exactly how.