Simple Inference for Constrained Optima

Hiroaki Kaido, Francesca Molinari, Jörg Stoye

BOSTON UNIVERSITY

CORNELL UNIVERSITY FOUNDED A.D. 1865

EEA-ESEM Summer Meeting, Milan, 2022

Simple Inference for Constrained Optima

What do we want:

A confidence interval for $p'\theta$, where

- p is a known vector
  (for this talk – we can handle a known or estimated smooth function $g(\theta)$),
- θ is constrained through moment inequalities $E(m_j(X_i, \theta)) \leq 0, \; j = 1, ..., J$.

Original motivation:

Confidence intervals for components of a partially identified vector.

But the problem is considerably more general.

What's new?

That it's **simple** to compute:

- A "critical value" is computed only once (twice for two-sided interval).
- Only one global optimization per direction.
  The adjustment from estimate to CI happens locally.

We have:

- Written a generic Python implementation...
- with a Stata wrapper...
- and are in the process of applying it to past papers.

What's out there?

Method 1: Project Confidence Regions for Vectors
(our example: Andrews-Soares, Ecma 2010)

- This is the most generally valid method.
- It is also (potentially severely) conservative.
- A CV must be bootstrapped at each θ.
- A single bootstrap is cheap.

What's out there?

Method 1: Project Confidence Regions for Vectors

Method 2: Calibrated Projection
(Kaido-Molinari-Stoye, Ecma 2019)

- Less universally valid than Method 1.
- Much less conservative than Method 1.
- A CV must be bootstrapped at each θ.
- A single bootstrap is moderately expensive (resamples linear programs).

What's out there?

Method 1: Project Confidence Regions for Vectors

Method 2: Calibrated Projection

Method 3: Profiling
(Romano-Shaikh, JSPI 2008; Bugni-Canay-Shi, QE 2017)

- Less universally valid than Method 1.
- Much less conservative than Method 1.
- A CV must be bootstrapped at each γ.
- A single bootstrap is very expensive (resamples nonlinear programs).

What's out there?

Method 1: Project Confidence Regions for Vectors

Method 2: Calibrated Projection

Method 3: Profiling

Method 4: Support Function Bootstrap
(Pakes-Porter-Ho-Ishii, working paper 2011; see Ecma 2015)

- Much less universally valid than other methods.
- Excludes moment equalities and (local to) point identification.
- Does not studentize moment inequalities.
- Much less conservative than Method 1 (unless the previous point bites).
- A CV must be bootstrapped once.
- A single bootstrap is moderately expensive (resamples linear programs).

What's out there?

Method 1: Project Confidence Regions for Vectors

Method 2: Calibrated Projection

Method 3: Profiling

Method 4: Support Function Bootstrap

Simple Calibrated Projection
(This paper)

- More assumptions than 1, 2, or 3, but considerably fewer than 4.
- Allows for moment equalities and (local to) point identification.
- Much less conservative than Method 1.
- A CV must be bootstrapped once (twice for two-sided intervals).
- This bootstrap is moderately expensive (resamples linear programs).
- There is only one global optimization (twice for two-sided intervals).

What's out there?

Method 1: Project Confidence Regions for Vectors

Method 2: Calibrated Projection

Method 3: Profiling

Method 4: Support Function Bootstrap

Simple Calibrated Projection
(This paper)

More loosely related:
I. Andrews/Pakes/Roth, Cox/Shi, Belloni/Bugni/Chernozhukov, Gafarov, Cho/Russell,...

Explaining Simple Calibrated Projection

Setting the Stage

The setting is moment inequalities:

$$\text{"identified set"} = \Theta_I = \{\theta : E(m_j(X_i, \theta)) \leq 0, j = 1, \ldots, J\}$$

The object of interest is the projection (=support function)

$$s(p, \Theta_I) \equiv \max \; p'\theta \text{ s.t. } \theta \in \Theta_I.$$

Explaining Simple Calibrated Projection

Setting the Stage

The setting is moment inequalities:

$$\text{"identified set"} = \Theta_I = \{\theta : E(m_j(X_i, \theta)) \leq 0, j = 1, \ldots, J\}$$

The object of interest is the projection (=support function)

$$s(p, \Theta_I) \equiv \max\ p'\theta \text{ s.t. } \theta \in \Theta_I.$$

We define estimators

$$\begin{aligned} \hat{\Theta}_I &\equiv \arg\min \max_j \overline{m}_j(\theta) \\ s(p, \hat{\Theta}_I) &\equiv \max\ p'\theta \text{ s.t. } \theta \in \hat{\Theta}_I. \end{aligned}$$

We are interested in the one-sided confidence interval

$$CI = (-\infty, \mathbf{???}].$$

Explaining Simple Calibrated Projection

Setting the Stage

The setting is moment inequalities:

$$\text{"identified set"} = \Theta_I = \{\theta : E(m_j(X_i,\theta)) \leq 0, j = 1,\dots,J\}$$

The object of interest is the projection (=support function)

$$s(p,\Theta_I) \equiv \max\ p'\theta \text{ s.t. } \theta \in \Theta_I.$$

We define estimators

$$\hat{\Theta}_I \equiv \arg\min \max_j \overline{m}_j(\theta)$$

$$s(p,\hat{\Theta}_I) \equiv \max\ p'\theta \text{ s.t. } \theta \in \hat{\Theta}_I.$$

We are interested in the one-sided confidence interval

$$CI = (-\infty, \mathbf{???}].$$

- The definition of $\hat{\Theta}_I$ ensures nonemptiness.
- We aim for coverage that is uniform over a reasonable class of DGPs.
- Two-sided inference is by intersection of one-sided confidence intervals (of size $(1-\alpha/2)$ or, depending on a pre-test, $(1-\alpha)$).

Explaining Simple Calibrated Projection

Simple Calibrated Projection: The Algorithm

Step 1: Pick an arbitrary $\hat{\theta}^* \in S(p, \hat{\Theta}_I) \equiv \arg\max_{\theta \in \hat{\Theta}_I} p'\theta$.

- This step is a by-product of estimating $s(p, \Theta_I)$.
- However, it is in practice the hardest.
- In other words, we incur minimal computational cost beyond estimation.

Explaining Simple Calibrated Projection

Simple Calibrated Projection: The Algorithm

Step 1: Pick an arbitrary $\hat{\theta}^* \in S(p, \hat{\Theta}_I) \equiv \arg\max_{\theta \in \hat{\Theta}_I} p'\theta$.

- This step is a by-product of estimating $s(p, \Theta_I)$.
- However, it is in practice the hardest.
- In other words, we incur minimal computational cost beyond estimation.

Intuitively, the next steps are:

- We replace the optimization problem with its local linear approximation at $\hat{\theta}^*$.
- For this approximation, we estimate by how much constraints would have to be relaxed so that the value of the relaxed sample problem exceeds the value of the unrelaxed population problem with probability $(1 - \alpha)$.
- We report the value of the correspondingly relaxed empirical problem.
- We retain the linearization in the last step, rendering it computationally trivial.

Explaining Simple Calibrated Projection

Simple Calibrated Projection: The Algorithm

Step 1: Pick an arbitrary $\hat{\theta}^* \in S(p, \hat{\Theta}_I) \equiv \arg\max_{\theta \in \hat{\Theta}_I} p'\theta$.

- This step is a by-product of estimating $s(p, \Theta_I)$.
- However, it is in practice the hardest.
- In other words, we incur minimal computational cost beyond estimation.

Intuitively, the next steps are:

- We replace the optimization problem with its local linear approximation at $\hat{\theta}^*$.
- For this approximation, we estimate by how much constraints would have to be relaxed so that the value of the relaxed sample problem exceeds the value of the unrelaxed population problem with probability $(1-\alpha)$.
- We report the value of the correspondingly relaxed empirical problem.
- We retain the linearization in the last step, rendering it computationally trivial.

The formal definition makes for one busy slide.

Step 2: Define the linear program

$$
\begin{aligned}
\psi(c, \boldsymbol{\mu}, \rho) \quad &\equiv \quad \max_{\vartheta \in \boldsymbol{R}^k} p'\vartheta \\
\text{s.t.} \quad & \hat{D}_j(\hat{\theta}^*)'\vartheta - \sqrt{n}\mu_j \leq c, j \in \mathcal{J}^* \\
& -\rho \leq e_j'\vartheta \leq \rho, j = 1, \ldots, k,
\end{aligned}
$$

where $(e_1, \ldots, e_k)$ is an orthonormal basis of $\boldsymbol{R}^k$ s.t. $e_1 = p/\|p\|$.

Step 2: Define the linear program

$$\begin{aligned}\psi(c,\boldsymbol{\mu},\rho) \quad &\equiv \quad \max_{\vartheta\in\boldsymbol{R}^k} p'\vartheta \\ \text{s.t.} \quad & \quad \hat{D}_j(\hat{\theta}^*)'\vartheta - \sqrt{n}\mu_j \leq c, j \in \mathcal{J}^* \\ & \quad -\rho \leq e_j'\vartheta \leq \rho, j = 1,\dots,k,\end{aligned}$$

where $(e_1,\dots,e_k)$ is an orthonormal basis of $\boldsymbol{R}^k$ s.t. $e_1 = p/\|p\|$.

- The program is parameterized by relaxation c, (bootstrap) sample intercepts $\boldsymbol{\mu} \equiv (\mu_j)_{j\in\mathcal{J}^*}$ and regularization parameter ρ. We will return to these.
- The index set

$$\mathcal{J}^* \equiv \left\{j \in \{1,\dots,J\} : \overline{m}_j(\hat{\theta}^*)/\hat{\sigma}_j(\hat{\theta}^*) \leq \sqrt{\log(n)/n}\right\}$$

  restricts attention to constraints that are plausibly binding at $\hat{\theta}^*$ (=Generalized Moment Selection; e.g., see Andrews/Soares 2010).
- $\hat{D}_j(\hat{\theta}^*)$ estimates the gradient $D_j(\hat{\theta}^*) \equiv \nabla_\theta E(m_j(X_i,\hat{\theta}^*)/\sigma_j(\hat{\theta}^*))$.

Explaining Simple Calibrated Projection

Step 2: Define the linear program

$$
\begin{aligned}
\psi(c, \boldsymbol{\mu}, \rho) \quad &\equiv \quad \max_{\vartheta \in \boldsymbol{R}^k} p' \vartheta \\
\text{s.t.} \quad & \hat{D}_j(\hat{\theta}^*)' \vartheta - \sqrt{n} \mu_j \leq c, j \in \mathcal{J}^* \\
& -\rho \leq e_j' \vartheta \leq \rho, j = 1, \ldots, k,
\end{aligned}
$$

where $(e_1, \ldots, e_k)$ is an orthonormal basis of $\boldsymbol{R}^k$ s.t. $e_1 = p / \|p\|$.

- The slackness vector $\boldsymbol{\mu}$ collects intercepts of linearized constraints. Their sampling variation will be how we model estimation uncertainty.
- We also restrict ϑ to a hypercube scaled by ρ. This ensures that linear approximation is uniformly (asymptotically) valid where we need it.
- We give a suggestion for the tuning parameter ρ.
- Under stronger assumptions that resemble Pakes, Porter, Ho, and Ishii (2011), we can set $\rho = \infty$.
- This eliminates a tuning parameter and clarifies relation to earlier work.

Step 3: Let $\hat{c}$ be the smallest value of c s.t.

$$\Pr(\psi(c, \boldsymbol{\mu}^b, \rho) \geq 0) \geq 1 - \alpha,$$

where

$$\mu_j^b = \frac{\overline{m}_j^b(\hat{\theta}^*) - \overline{m}_j(\hat{\theta}^*)}{\hat{\sigma}_j(\hat{\theta}^*)}$$

and $\left(\overline{m}_j^b(\hat{\theta}^*)\right)_{j \in \mathcal{J}^*}$ is a bootstrap analog of $(\overline{m}_j(\hat{\theta}^*))_{j \in \mathcal{J}^*}$.

- In our linear bootstrap world, this is by how much we would have to relax empirical constraints so that the relaxed empirical maximization problem covers true problem's value.

Explaining Simple Calibrated Projection

Step 3: Let $\hat{c}$ be the smallest value of c s.t.

$$\Pr(\psi(c, \boldsymbol{\mu}^b, \rho) \geq 0) \geq 1 - \alpha,$$

where

$$\mu_j^b = \frac{\overline{m}_j^b(\hat{\theta}^*) - \overline{m}_j(\hat{\theta}^*)}{\hat{\sigma}_j(\hat{\theta}^*)}$$

and $\left(\overline{m}_j^b(\hat{\theta}^*)\right)_{j \in \mathcal{J}^*}$ is a bootstrap analog of $(\overline{m}_j(\hat{\theta}^*))_{j \in \mathcal{J}^*}$.

- In our linear bootstrap world, this is by how much we would have to relax empirical constraints so that the relaxed empirical maximization problem covers true problem's value.

Step 4: The confidence interval equals

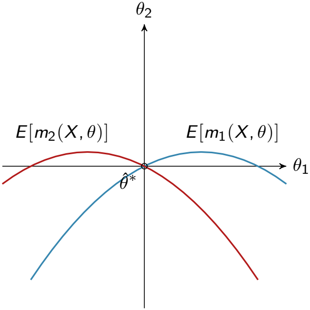
$$CI_\alpha \quad = \quad (-\infty, s(p, \hat{\Theta}_I) + \psi(\hat{c}, \mathbf{0}, \infty)/\sqrt{n}].$$

- We report the value of the empirical optimization problem but relaxed by $\hat{c}$.
- We retain the linearization and thereby avoid a second global optimization.
- Computationally, this is just one more bootstrap iteration (with the ρ-box dropped).

Reminder: Calibrated Projection

The busy slide as a picture: We...

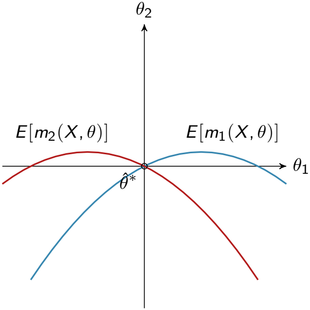
- choose an arbitrary empirical support point $\hat{\theta}^*$,

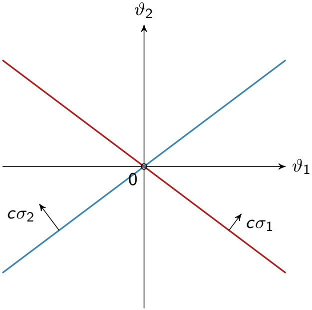


Reminder: Calibrated Projection

The busy slide as a picture: We...

- choose an arbitrary empirical support point $\hat{\theta}^*$,
- linearize,

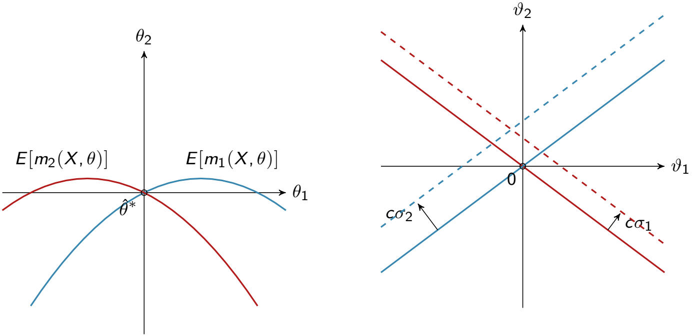




Reminder: Calibrated Projection

The busy slide as a picture: We...

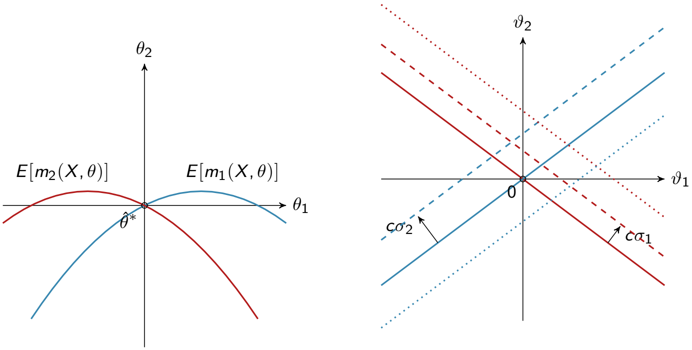
- choose an arbitrary empirical support point $\hat{\theta}^*$,
- linearize,
- relax by a candidate c,



Reminder: Calibrated Projection

The busy slide as a picture: We...

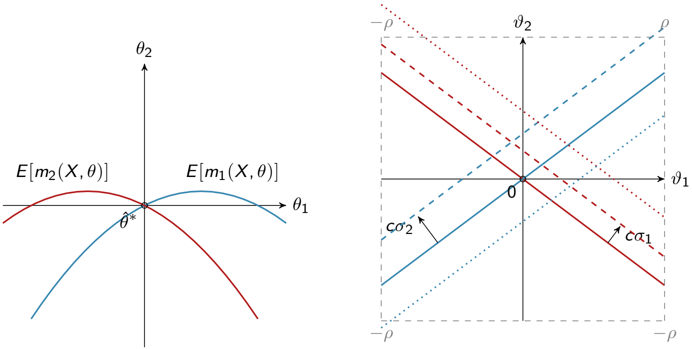
- choose an arbitrary empirical support point $\hat{\theta}^*$,
- linearize,
- relax by a candidate c,
- resample intercepts (one possible bootstrap draw shown),



Reminder: Calibrated Projection

The busy slide as a picture: We...

- choose an arbitrary empirical support point $\hat{\theta}^*$,
- linearize,
- relax by a candidate c,
- resample intercepts (one possible bootstrap draw shown),
- encase in the ρ-box if needed.



Justifying Simple Calibrated Projection

When does this work?

- The calibration of $\hat{c}$ at $\hat{\theta}^*$ builds on our previous work.
- The difference is we calibrate only at $\hat{\theta}^*$, not again and again for each θ.
- This makes the procedure much faster.
- The cost is that we need additional assumptions.

Justifying Simple Calibrated Projection

When does this work?

- The calibration of $\hat{c}$ at $\hat{\theta}^*$ builds on our previous work.
- The difference is we calibrate only at $\hat{\theta}^*$, not again and again for each θ.
- This makes the procedure much faster.
- The cost is that we need additional assumptions.

High-Level New Condition:

$\hat{\theta}^*$ is $O(n^{-1/2})$-close to **some**, **not necessarily constant** true support point.

Equivalently, its point-set distance from the true support set is of order $O(n^{-1/2})$.
We next formalize this and provide sufficient low-level conditions.

Justifying Simple Calibrated Projection

Preliminaries

Recall we want to verify coverage of

$$s(p, \Theta_I) \equiv \max p'\theta \text{ s.t. } \theta \in \Theta_I.$$

We impose assumptions from previous work, notably:

- The functions m_j are known.
- $E(m_j(X_i, \theta))$ as function of θ is smooth.
- $E(m_j(X_i, \theta))/\sigma_j(\theta)$ is bounded.
- $\mathbb{G}_j(X_i, \theta) = \frac{1}{\sqrt{n}\sigma_j(\theta)} \sum_i \big(m_j(X_i, \theta) - E(m_j(X_i, \theta))\big)$ can be bootstrapped including if an estimator $\hat{\theta}$ is plugged in.
- A restriction on covariances that precludes superconsistent estimation.

Justifying Simple Calibrated Projection

Condition 1

$$\max_{\theta \in S(p,\hat{\Theta}_I)} d(\theta, S(p, \Theta_I)) = O_P(n^{-1/2}),$$

where $d(\cdot)$ is point-set distance.

Note relation to Hausdorff distance:

- This is "directed" Hausdorff distance. Standard Hausdorff distance would be

$$\max\left\{ \max_{\theta \in S(p,\hat{\Theta}_I)} d(\theta, S(p, \Theta_I)), \max_{\theta \in S(p,\Theta_I)} d(\theta, S(p, \hat{\Theta}_I)) \right\}.$$

- The estimated support set must hit the true one, but need not explore it.
- This ensures the desired property of any selection $\hat{\theta}^*$ from $s(p, \hat{\Theta}_I)$.
- It is also **much** easier to enforce by assumption.

Justifying Simple Calibrated Projection

Condition 1

$$\max_{\theta \in S(p,\hat{\Theta}_I)} d(\theta, S(p, \Theta_I)) = O_P(n^{-1/2}),$$

where $d(\cdot)$ is point-set distance.

Result 1:

If Condition 1 holds, SCP achieves uniform asymptotic size control.

Why?

- All accumulation points of $\hat{\theta}^*$ are in $S(p, \Theta_I) \equiv \arg\max_{\theta \in \Theta_I} p'\theta$.
- Furthermore, they are approached at $\sqrt{n}$-rate.
- Can then leverage existing proofs to show: $\hat{c}_n(\hat{\theta}^*)$ ensures correct coverage of $p'\theta$ for some $\theta \in S(p, \Theta_I)$.
- But that is coverage of $s(p, \Theta_I)$.

Justifying Simple Calibrated Projection

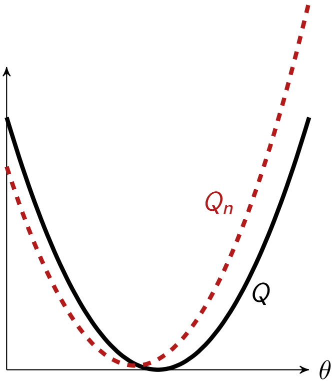
We next identify lower-level conditions that imply Condition 1.

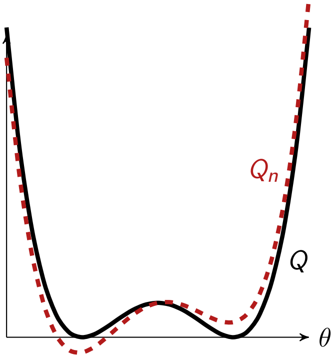
To do so, relate our problem to first-year PhD Extremum (or m-) Estimation.

$$
\begin{aligned}
S(p,\Theta_I) &= \arg\min_{\theta\in\Theta} Q(\theta), \\
Q(\theta) &\equiv \max_j [E(m_j(X_i,\theta))/\sigma(\theta)]_+ + |p'\theta - s(p,\Theta_I)| \\
S(p,\hat{\Theta}_I) &= \arg\min_{\theta\in\Theta} Q_n(\theta), \\
Q_n(\theta) &\equiv \max_j [\bar{m}_j(\theta)/\hat{\sigma}(\theta)]_+ - \min_{\theta\in\Theta}\max_j [\bar{m}_j(\theta)/\hat{\sigma}(\theta)]_+ + |p'\theta - s(p,\hat{\Theta}_I)|.
\end{aligned}
$$

If standard m-estimator consistency conditions apply to these objects, then $\max_{\theta\in S(p,\hat{\Theta}_I)} d(\theta, S(p,\Theta_I)) \xrightarrow{p} 0$ is implied.

Justifying Simple Calibrated Projection

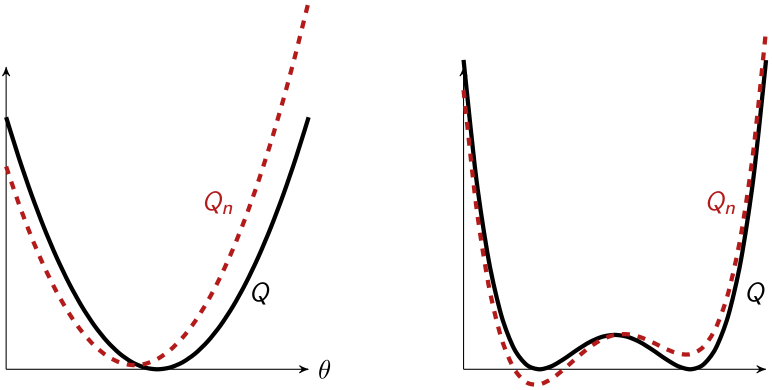




Left panel:

- Well-behaved extremum estimation.
- The minimum of Q_n estimates the minimum of Q.

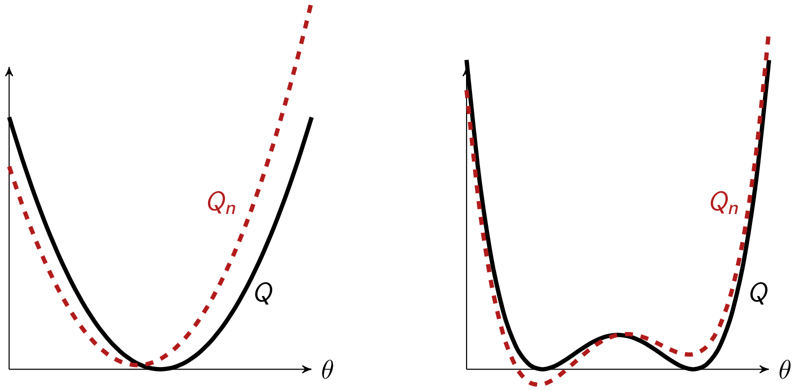
Justifying Simple Calibrated Projection



Right panel:

- Well-behaved extremum estimation except for identification failure.
- The minimum of Q_n "inner approximates" the minimum of Q.
  This insight goes back at least to Redner 81. See also Newey/McFadden 94.

Justifying Simple Calibrated Projection



- We are not quite done because we also need the $\sqrt{n}$-rate.
- Wanted:
  A result between consistency (too weak) and $\sqrt{n}$-asy normality (not true).
- We will adapt the "Argmax Theorem" of van der Vaart and Wellner.

Justifying Simple Calibrated Projection

A Set-Valued Argmax-Theorem

Define $\nu_n(\theta) \equiv \sqrt{n}(Q_n(\theta) - Q(\theta))$ and $\Theta^* \equiv \arg\min Q(\theta)$.

Suppose that:

1. $Q(\theta) - \min_{\theta \in \Theta} Q(\theta) \geq C \min\{d^2(\theta, \Theta^*), \delta\}$,
2. $\exists \epsilon > 0, \forall \delta \leq \epsilon : E\left(\sup_{\theta \in \Theta, \theta^* \in \Theta^* : \|\theta - \theta^*\| \leq \delta} |\nu_n(\theta) - \nu_n(\theta^*)|\right) \leq \delta$.

Then $\max_{\theta \in \hat{\Theta}^*} d(\theta, \Theta^*)) = O_P(n^{-1/2})$, where $\hat{\Theta}^* \equiv \arg\min Q_n(\theta)$.

Justifying Simple Calibrated Projection

A Set-Valued Argmax-Theorem

Define $\nu_n(\theta) \equiv \sqrt{n}(Q_n(\theta) - Q(\theta))$ and $\Theta^* \equiv \arg\min Q(\theta)$.

Suppose that:

1. $Q(\theta) - \min_{\theta \in \Theta} Q(\theta) \geq C \min\{d^2(\theta, \Theta^*), \delta\}$,
2. $\exists \epsilon > 0, \forall \delta \leq \epsilon : E\left(\sup_{\theta \in \Theta, \theta^* \in \Theta^* : \|\theta - \theta^*\| \leq \delta} |\nu_n(\theta) - \nu_n(\theta^*)|\right) \leq \delta$.

Then $\max_{\theta \in \hat{\Theta}^*} d(\theta, \Theta^*)) = O_P(n^{-1/2})$, where $\hat{\Theta}^* \equiv \arg\min Q_n(\theta)$.

- This novel result clarifies general conditions under which a sample arg max approaches its population counterpart at "parametric" rate.
- Or at other rates, depending on modulus of continuity of $\nu_n(\cdot)$. The above statement extracts what we need today.
- We "morally" impose the (much stronger) conditions for $\sqrt{n}$-Hausdorff consistency on $\Theta^* = \Theta_I$ because we need $\sqrt{n}$-consistency of $s(p, \hat{\Theta}_I)$.
- But we do not, and would not want to, impose them on $\Theta^* = S(p, \Theta_I)$.

Justifying Simple Calibrated Projection

Condition 2

We have that

$$s(p, \hat{\Theta}_I) - s(p, \Theta_I) = O_P(n^{-1/2})$$

and

$$\theta \in H(p, \Theta_I) \implies Q(\theta) \geq C \min\{d^2(\theta, S(p, \Theta_I)), \delta\}$$

$\uparrow$ supporting hyperplane.

- The sample analog of the support function is $\sqrt{n}$-consistent.
  Sufficient condition: $d_H(\hat{\Theta}_I, \Theta_I) = O_P(n^{-1/2})$.
  Counterexample: Weak identification.
- $S(p, \Theta_I)$ is a well-separated minimum of Q *on the supporting hyperplane*.
  Counterexamples: Nearly flat face, smooth maximum with singular Hessian.

Justifying Simple Calibrated Projection

Condition 2

We have that

$$s(p, \hat{\Theta}_I) - s(p, \Theta_I) = O_P(n^{-1/2})$$

and

$$\theta \in H(p, \Theta_I) \implies Q(\theta) \geq C \min\{d^2(\theta, S(p, \Theta_I)), \delta\}$$

$\uparrow$ supporting hyperplane.

Result 2:
If Condition 2 holds, SCP achieves uniform asymptotic size control.

Why?
The conditions allow to invoke the generalized arg max-theorem.

Justifying Simple Calibrated Projection

Condition 2

We have that

$$s(p, \hat{\Theta}_I) - s(p, \Theta_I) = O_P(n^{-1/2})$$

and

$$\theta \in H(p, \Theta_I) \implies Q(\theta) \geq C \min\{d^2(\theta, S(p, \Theta_I)), \delta\}$$

↑ supporting hyperplane.

Result 2a:
The first assumption above is implied by the **degeneracy** and **polynomial minorant** conditions for moment inequality models in Chernozhukov/Hong/Tamer (2007) and also by similar conditions in Pakes/Porter/Ho/Ishii (2011).

Justifying Simple Calibrated Projection

Condition 2

We have that

$$s(p, \hat{\Theta}_I) - s(p, \Theta_I) = O_P(n^{-1/2})$$

and

$$\theta \in H(p, \Theta_I) \implies Q(\theta) \geq C \min\{d^2(\theta, S(p, \Theta_I)), \delta\}$$

$\uparrow$ supporting hyperplane.

Result 2b:
The second assumption above is considerably weaker than a similar assumption in Bugni/Canay/Shi (2017) and also Pakes/Porter/Ho/Ishii (2011).
In particular, both imply the above but with $Q(\theta) \geq C \min\{d(\theta, S(p, \Theta_I)), \delta\}$.

For detailed analysis of how these restrictions relate to each other and to **constraint qualifications**, see Kaido/Molinari/Stoye (2022).

Justifying Simple Calibrated Projection

How restrictive is Condition 2?

Condition 2 does exclude cases of interest, e.g. weak identification of Θ_I.
(Both calibrated projection and profiling allow weak identification of Θ_I.)

However, it accommodates some interesting "tricky" cases.

- **Smooth maximum:**
  The support set is (at least partially) characterized by a FOC.
  (This case is excluded for support function bootstrap and profiling.)
- **Multiple solutions:**
  Distinct, locally well-identified global maxima.
  (This case is excluded for support function bootstrap.)
- **Thin identified set:**
  This occurs whenever there are some moment equalities, including the boundary case of GMM.
  (This case is excluded for support function bootstrap.)

Justifying Simple Calibrated Projection

Two-Sided Confidence Intervals

We propose to always intersect two one-sided intervals.

If the projection of Θ_I is "long" in the sense of

$$\Delta \equiv s(p, \Theta_I) - s(-p, \Theta_I) \gg \text{sampling uncertainty},$$

being much larger than sampling uncertainty, intersect two $(1-\alpha)$-CI's.

Else, intersect two $(1-\alpha/2)$-CI's.

Of course, we do not know which obtains.

We propose a conservative pre-test, e.g. use the larger interval if

$$\hat{\Delta} \equiv s(p, \hat{\Theta}_I) - s(-p, \hat{\Theta}_I) \leq \kappa_n.$$

(This resembles Imbens/Manski (2004) and in particular Stoye (2009).
Can be generalized to shrinkage as in Andrews/Soares (2010).)

Detailed Comparison to Support Function Bootstrap

Superconsistency and Support Function Bootstrap

Compared to PPHI, we impose an additional restriction on the d.g.p.:

Moment conditions must not be (near) perfectly correlated without the researcher's knowledge.
(Known perfect correlations are no problem.)

$s(p, \hat{\Theta}_I)$ may otherwise be superconsistent for $s(p, \Theta_I)$.
A nonparametric bootstrap will then fail.

Detailed Comparison to Support Function Bootstrap

Superconsistency and Support Function Bootstrap

Compared to PPHI, we impose an additional restriction on the d.g.p.:

Moment conditions must not be (near) perfectly correlated without the researcher's knowledge.
(Known perfect correlations are no problem.)

$s(p, \hat{\Theta}_I)$ may otherwise be superconsistent for $s(p, \Theta_I)$.
A nonparametric bootstrap will then fail.

This observation also applies to support function bootstrap.
Our assumption must in fact be added to PPHI.
Auxiliary contribution: Rigorous justification of support function bootstrap.

Detailed Comparison to Support Function Bootstrap

To studentize or not to studentize?

Our presentation assumed studentization throughout.

The Support Function Bootstrap cannot studentize moment conditions.

This is expected to sacrifice power (Andrews and Soares, 2010).

Example: Minimum of two means

$$\begin{aligned} \theta &\leq \min\{E(X_i), E(Y_i)\} \\ X_i &\sim N(\mu_x, 1) \\ Y_i &\sim N(\mu_y, \sigma^2). \end{aligned}$$

We can then compare

$$\begin{aligned} \text{studentized } CI &= (-\infty, \min\{\bar{x} + 1.95/\sqrt{n}, \bar{y} + 1.95\sigma/\sqrt{n}\}] \\ \text{nonstudentized } CI &= (-\infty, \min\{\bar{x}, \bar{y}\} + 1.64\sigma/\sqrt{n}]. \end{aligned}$$

With large σ, the second interval is much larger in expectation.

Detailed Comparison to Support Function Bootstrap

To studentize or not to studentize?

Our presentation assumed studentization throughout.

The Support Function Bootstrap cannot studentize moment conditions.

This is expected to sacrifice power (Andrews and Soares, 2010).

On the other hand:

Studentization can add computational cost.

- This is first-order if nonstudentized constraints are linear.
- We may also not want to studentize when computing $\hat{\theta}^*$.

We therefore treat studentization as optional.

Wrapping up

Conclusion

We analyze **simple** inference methods for constrained optima.

Motivating application:
Projections of partially identified parameter vectors.

Core findings:

- We show that under reasonable (though not innocuous) conditions, our earlier **Calibrated Projection** method can be drastically simplified.
- Namely, a "critical value" gets computed at most twice.
- We provide an analogous simplification of a **Profiling** method.
- We rigorously justify another simple method (**Support Function Bootstrap**) that is used in applications.
- We explain trade-offs between the relatively simple methods thus justified.

Wrapping up

Conclusion

We analyze **simple** inference methods for constrained optima.

Motivating application:
Projections of partially identified parameter vectors.

Core findings:

- We show that under reasonable (though not innocuous) conditions, our earlier **Calibrated Projection** method can be drastically simplified.
- Namely, a "critical value" gets computed at most twice.
- We provide an analogous simplification of a **Profiling** method.
- We rigorously justify another simple method (**Support Function Bootstrap**) that is used in applications.
- We explain trade-offs between the relatively simple methods thus justified.

Did I mention the Stata command?

Wrapping up

Conclusion

We analyze **simple** inference methods for constrained optima.

Motivating application:
Projections of partially identified parameter vectors.

Core findings:

- We show that under reasonable (though not innocuous) conditions, our earlier **Calibrated Projection** method can be drastically simplified.
- Namely, a "critical value" gets computed at most twice.
- We provide an analogous simplification of a **Profiling** method.
- We rigorously justify another simple method (**Support Function Bootstrap**) that is used in applications.
- We explain trade-offs between the relatively simple methods thus justified.

Did I mention the Stata command?

Thank you!