Smoothed inference for moment inequality models

C. Bontemps[1,2], R. Kumar[3] and M. Lesellier[4]

1: Toulouse School of Economics
2: Ecole Nationale Aviation Civile
3: Indian Institute of Technology Delhi
4: Université de Montréal

August 28, 2024

Moment inequalities in economics

- Conditional and unconditional moment inequalities appear naturally in economic/econometric models.
  - Measurement issues/incomplete data: interval-censored data
  - Sample selection: treatment effect models with endogenous selection
  - Games with multiple equilibria: entry games, network formation games

Moment inequalities in economics

- Conditional and unconditional moment inequalities appear naturally in economic/econometric models.
  - Measurement issues/incomplete data: interval-censored data
  - Sample selection: treatment effect models with endogenous selection
  - Games with multiple equilibria: entry games, network formation games
- Two solutions for empirical research:
  1. Add more structure to recover moment equalities $\implies$ standard estimation procedures.
  2. Directly use the moment inequalities to estimate the set of parameters that can generate the observed data.

Moment inequalities in economics

- Conditional and unconditional moment inequalities appear naturally in economic/econometric models.
  - Measurement issues/incomplete data: interval-censored data
  - Sample selection: treatment effect models with endogenous selection
  - Games with multiple equilibria: entry games, network formation games
- Two solutions for empirical research:
  1. Add more structure to recover moment equalities $\Longrightarrow$ standard estimation procedures.
  2. Directly use the moment inequalities to estimate the set of parameters that can generate the observed data.
- **This paper:** we propose a new inference procedure for moment inequalities that combines **good statistical properties** and **ease of implementation**.

Challenges in the estimation of models featuring moment inequalities

1. **Selection and derivation of the moment inequalities**

Challenges in the estimation of models featuring moment inequalities

1. **Selection and derivation of the moment inequalities**

2. **Inference for unconditional moment inequalities:**

   - The asymptotic distribution of the test statistic under the null depends on the set of binding moments, which is unknown.

   - The commonly used test statistics are non-pivotal $\Longrightarrow$ complicates the derivation of the critical value.

Challenges in the estimation of models featuring moment inequalities

1. **Selection and derivation of the moment inequalities**
2. **Inference for unconditional moment inequalities:**
   - The asymptotic distribution of the test statistic under the null depends on the set of binding moments, which is unknown.
   - The commonly used test statistics are non-pivotal $\Longrightarrow$ complicates the derivation of the critical value.
   - The current methods rely on simulation-based methods (sub-sampling, GMS, bootstrap) or upper bounds on the test statistic.
   - Most methods are both conservative and computationally intensive.

other challenges

Challenges in the estimation of models featuring moment inequalities (II)

3. **Inference for conditional moment inequalities:**
   - The asymptotic distribution of the test statistic depends on the set of binding conditional moments for each $x \in \mathcal{X}$
   - If conditioning variable X is continuous, the identified set is characterized by an infinite number of inequalities.
   - Conditional moment inequalities are non-parametric objects that are harder to estimate with slower convergence rates and a usually unknown asymptotic distribution.

Challenges in the estimation of models featuring moment inequalities (II)

3. **Inference for conditional moment inequalities:**
   - The asymptotic distribution of the test statistic depends on the set of binding conditional moments for each $x \in \mathcal{X}$
   - If conditioning variable X is continuous, the identified set is characterized by an infinite number of inequalities.
   - Conditional moment inequalities are non-parametric objects that are harder to estimate with slower convergence rates and a usually unknown asymptotic distribution.

4. **Implementation:** the estimation of the confidence region is done by inverting a test over a grid $\implies$ curse of dimensionality
   - One needs to repeat the test for each point in a grid of tested points!

Challenges in the estimation of models featuring moment inequalities (II)

3. **Inference for conditional moment inequalities:**
    - The asymptotic distribution of the test statistic depends on the set of binding conditional moments for each $x \in \mathcal{X}$
    - If conditioning variable X is continuous, the identified set is characterized by an infinite number of inequalities.
    - Conditional moment inequalities are non-parametric objects that are harder to estimate with slower convergence rates and a usually unknown asymptotic distribution.

4. **Implementation:** the estimation of the confidence region is done by inverting a test over a grid $\implies$ curse of dimensionality
    - One needs to repeat the test for each point in a grid of tested points!

5. **Subvector inference**

Outline

1. Unconditional moment inequalities
2. Conditional moment inequalities
3. Monte Carlo simulations
4. Conclusion

Unconditional moment inequalities

General Setup

- We observe an i.i.d. sample $\{W_i\}_{i=1}^n$ where $W_i \in \mathcal{W} \subset \mathbb{R}^{d_W}$ is distributed according to $P \in \mathcal{P}$.
- We consider an economic model where the parameter of interest is characterized by the following p unconditional moment inequalities.

$$\mathbb{E}[m(W_i, \theta)] \geq 0_p,$$

where $m : \mathbb{R}^{d_W} \times \Theta \to \mathbb{R}^p$ is a known measurable function.

General Setup

- We observe an i.i.d. sample $\{W_i\}_{i=1}^n$ where $W_i \in \mathcal{W} \subset \mathbb{R}^{d_W}$ is distributed according to $P \in \mathcal{P}$.
- We consider an economic model where the parameter of interest is characterized by the following p unconditional moment inequalities.

$$\mathbb{E}[m(W_i, \theta)] \geq 0_p,$$

  where $m : \mathbb{R}^{d_W} \times \Theta \to \mathbb{R}^p$ is a known measurable function.

- The identified set Θ_I is defined as follows:

$$\Theta_I = \left\{\theta \in \Theta \mid \mathbb{E}[m(W_i, \theta)] \geq 0_p\right\}.$$

- Notation: $m_\theta \equiv \mathbb{E}[m(W_i, \theta)]$

Inference

- In practice, the econometrician doesn't observe the true moment m_θ but an empirical counterpart $m_{\theta,n}$:

$$m_{\theta,n} = \frac{1}{n}\sum_{i=1}^{n} m(W_i, \theta)$$

- The objective for the econometrician is to construct a confidence region CR_n that satisfies the following two properties:

- **Asymptotic validity**: $$\forall \theta \in \Theta_I, \quad \liminf_{n\to\infty} P(\theta \in CR_n) \geq 1 - \alpha.$$

- **Consistency**: $$\forall \theta \notin \Theta_I, \quad \lim_{n\to\infty} P(\theta \in CR_n) = 0.$$

Inference

- In practice, the econometrician doesn't observe the true moment m_θ but an empirical counterpart $m_{\theta,n}$:

$$m_{\theta,n} = \frac{1}{n}\sum_{i=1}^{n} m(W_i,\theta)$$

- The objective for the econometrician is to construct a confidence region CR_n that satisfies the following two properties:
- **Asymptotic validity**: $$\forall\theta\in\Theta_I,\quad \liminf_{n\to\infty} P(\theta\in CR_n)\geq 1-\alpha.$$
- **Consistency**: $$\forall\theta\notin\Theta_I,\quad \lim_{n\to\infty} P(\theta\in CR_n)=0.$$
- Additional desirable properties: uniform validity over $(\mathcal{P},\Theta_I(P))$ and non-conservativeness

Canonical estimation procedure

- The traditional inference procedure usually relies on a test statistic of the form:

$$\xi_n(\theta) = \min_{j=1,\ldots,p} \sqrt{n} \frac{\frac{1}{n}\sum_{i=1}^{n} m_j(W_i,\theta)}{\sqrt{\widehat{var}(m_j(W_i,\theta))}}$$

  - Other test statistics are possible: MMM, QLR, ...

Canonical estimation procedure

- The traditional inference procedure usually relies on a test statistic of the form:

$$\xi_n(\theta) = \min_{j=1,\dots,p} \sqrt{n} \frac{\frac{1}{n}\sum_{i=1}^n m_j(W_i,\theta)}{\sqrt{\widehat{var}(m_j(W_i,\theta))}}$$

  - Other test statistics are possible: MMM, QLR, ...

- Confidence region: $$CR_n(\theta) = \{\theta \in \Theta \mid \xi_n(\theta) \geq c^*\}$$

  with c^* a critical value chosen to recover asymptotic validity.

Canonical estimation procedure

- The traditional inference procedure usually relies on a test statistic of the form:

$$\xi_n(\theta) = \min_{j=1,\ldots,p} \sqrt{n} \frac{\frac{1}{n}\sum_{i=1}^{n} m_j(W_i,\theta)}{\sqrt{\widehat{var}(m_j(W_i,\theta))}}$$

  - Other test statistics are possible: MMM, QLR, ...

- Confidence region: $CR_n(\theta) = \{\theta \in \Theta \mid \xi_n(\theta) \geq c^*\}$

  with c^* a critical value chosen to recover asymptotic validity.

- **Main challenge in deriving** c^*: the asymptotic distribution of the test statistic depends on the identity of the binding moments, which is unknown to the econometrician.

Computation of the critical value in the literature

There are three strands of methods to compute the critical value:

1. **Simulation-based methods** that seek to approximate the asymptotic distribution: [Chernozhukov et al., 2007], [Andrews and Soares, 2010], [Andrews and Barwick, 2012], [Romano et al., 2014], [Chen et al., 2018]
    - The most established procedure is the generalized moment selection (GMS) in [Andrews and Soares, 2010]: empirical selection of the binding moments.
    - critical values must be simulated for each candidate parameter in the grid.

Computation of the critical value in the literature

There are three strands of methods to compute the critical value:

1. **Simulation-based methods** that seek to approximate the asymptotic distribution: [Chernozhukov et al., 2007], [Andrews and Soares, 2010], [Andrews and Barwick, 2012], [Romano et al., 2014], [Chen et al., 2018]
   - The most established procedure is the generalized moment selection (GMS) in [Andrews and Soares, 2010]: empirical selection of the binding moments.
   - critical values must be simulated for each candidate parameter in the grid.

3. **Upper bounds on the exact or asymptotic distribution**: [Chernozhukov et al., 2018b], [Rosen, 2008]
   - Simpler implementation but can be conservative.

3. **Conditional tests**: [Cox and Shi, 2022]

The smoothed min approach

- For $z = (z_1, z_2, ..., z_p) \in \mathbb{R}^p$, we consider the following smooth approximation of the minimum between the elements of z and 0:

$$g_{\rho}(z) = \frac{\sum_{j=1}^{p} z_j \exp(-\rho z_j)}{1 + \sum_{j=1}^{p} \exp(-\rho z_j)},$$

- ρ is the smoothing parameter: it controls the level of approximation.

The smoothed min approach

- For $z = (z_1, z_2, ..., z_p) \in \mathbb{R}^p$, we consider the following smooth approximation of the minimum between the elements of z and 0:

$$g_\rho(z) = \frac{\sum_{j=1}^{p} z_j \exp(-\rho z_j)}{1 + \sum_{j=1}^{p} \exp(-\rho z_j)},$$

- ρ is the smoothing parameter: it controls the level of approximation.

- Following [Chernozhukov et al., 2015], we have:

$$|min(0, z_1, z_2, \ldots, z_p) - g_\rho(z)| \leq \frac{1}{\rho} \log\left(\frac{p-1}{e}\right), \text{ for} p > 1$$

- The larger ρ, the closer the approximation is to the minimum

The smoothed min approach: A pivotal test statistic

- We define our smooth test statistic as follows:

$$\xi_n(\theta) = \sqrt{n}\frac{g_{\rho_n}(m_{\theta,n})}{\sqrt{\nabla g_{\rho_n}(m_{\theta,n})^T \Sigma_n \nabla g_{\rho_n}(m_{\theta,n})}}$$

with Σ_n a consistent estimator of Σ_0 the variance of the moments and ∇g_{ρ_n} the gradient of g_{ρ_n}.

- Our confidence region of confidence level $1-\alpha$ is defined as follows:

$$\mathrm{CR}_n(1-\alpha) = \{\xi_n(\theta) \geq z_\alpha\}$$

with z_α the α-quantile of the standard normal distribution.

The smoothed min approach: Regularity assumptions

Assumption (Regularity assumption on the moments)

1. $\exists C$ *such that* $\forall \theta \in \Theta$, $\forall j$, $\mathbb{E}\left[m_j(W_i, \theta)^2\right] < C$
2. $\forall \theta \in \Theta, \;\; \Sigma_\theta \equiv \mathbb{E}\left[\left(m(W_i, \theta) - \mathbb{E}[m(W_i, \theta)]\right)\left(m(W_i, \theta) - \mathbb{E}[m(W_i, \theta)]\right)^T\right]$ *is positive definite*

- Regularity conditions that are very common when conducting inference in parametric models
- They allow us to write a CLT for the vector of moments with a positive definite asymptotic variance-covariance matrix.

The smoothed min approach: Asymptotic expansion

Proposition (Asymptotic expansion of the test statistic)

Assumption 1 holds. Let ρ_n a divergent sequence of positive number such that $\rho_n = cn^\alpha + o(1)$, $0 < \alpha < 1/2$, then

$$\sqrt{n}g_{\rho_n}(m_{\theta,n}) = \underbrace{\sqrt{n}g_{\rho_n}(m_\theta)}_{(1)} + \underbrace{\nabla g_{\rho_n}(m_\theta)\sqrt{n}(m_{\theta,n} - m_\theta)}_{(2)} + O_P\left(\frac{\rho_n}{\sqrt{n}}\right),$$

The smoothed min approach: Asymptotic expansion

Proposition (Asymptotic expansion of the test statistic)

Assumption 1 holds. Let ρ_n *a divergent sequence of positive number such that* $\rho_n = cn^{\alpha} + o(1)$, $0 < \alpha < 1/2$, *then*

$$\sqrt{n} g_{\rho_n}(m_{\theta,n}) = \underbrace{\sqrt{n} g_{\rho_n}(m_\theta)}_{(1)} + \underbrace{\nabla g_{\rho_n}(m_\theta)\sqrt{n}(m_{\theta,n} - m_\theta)}_{(2)} + O_P\left(\frac{\rho_n}{\sqrt{n}}\right),$$

- Term (1) converges to 0 if $\theta \in \Theta_I$ and $-\infty$ otherwise.
- Term (2) is asymptotically normal.

The smoothed min approach: Asymptotic expansion

Proposition (Asymptotic expansion of the test statistic)

Assumption 1 holds. Let ρ_n *a divergent sequence of positive number such that* $\rho_n = cn^{\alpha} + o(1)$, $0 < \alpha < 1/2$, *then*

$$\sqrt{n} g_{\rho_n}(m_{\theta,n}) = \underbrace{\sqrt{n} g_{\rho_n}(m_\theta)}_{(1)} + \underbrace{\nabla g_{\rho_n}(m_\theta)\sqrt{n}(m_{\theta,n} - m_\theta)}_{(2)} + O_P\left(\frac{\rho_n}{\sqrt{n}}\right),$$

- Term (1) converges to 0 if $\theta \in \Theta_I$ and $-\infty$ otherwise.
- Term (2) is asymptotically normal.
- The constraint on ρ_n implies that the amount of smoothing cannot decrease faster than the parametric convergence rate.

Let $\mathcal{J}_0(\theta) = \{j \in \{1, \dots, p\} : m_{\theta,j} = 0\}$.

Proposition (Asymptotic properties of the test statistic)

Assumption 1 holds. Let ρ_n a divergent sequence of positive numbers such that $\rho_n = cn^{\alpha} + o(1)$, then the following holds:

1. $\theta \in \Theta_I$ *and* $J_0 = \mathrm{card}(\mathcal{J}_0(\theta)) = 0$*:*

$$\Pr(\xi_n(\theta) > z_\alpha) \underset{n\to\infty}{\longrightarrow} 1.$$

2. $\theta \in \Theta_I$ *and* $J_0 = \mathrm{card}(\mathcal{J}_0(\theta)) > 0$*:*

$$\xi_n(\theta) \overset{d}{\underset{n\to\infty}{\longrightarrow}} \mathcal{N}(0,1).$$

3. $\theta \notin \Theta_I$:

$$\Pr(\xi_n(\theta) > z_\alpha) \underset{n\to\infty}{\longrightarrow} 0.$$

Additional remarks

- Choice of the smoothing parameter is crucial. We propose a method to calibrate ρ_n.

$\Longrightarrow$ Trade-off between (i) the "**bias**" implied by the difference between the min and its smooth approximation and (ii) the **accuracy** of the first-order approximation.

Additional remarks

- Choice of the smoothing parameter is crucial. We propose a method to calibrate ρ_n.

  $\Longrightarrow$ Trade-off between (i) the "**bias**" implied by the difference between the min and its smooth approximation and (ii) the **accuracy** of the first-order approximation.

- A variant of our test statistic is to standardize all the moments beforehand $\Longrightarrow$ yields better results in our simulations

Additional remarks

- Choice of the smoothing parameter is crucial. We propose a method to calibrate ρ_n.

  $\Longrightarrow$ Trade-off between (i) the "**bias**" implied by the difference between the min and its smooth approximation and (ii) the **accuracy** of the first-order approximation.

- A variant of our test statistic is to standardize all the moments beforehand $\Longrightarrow$ yields better results in our simulations

- We show uniform validity over $(\mathcal{P}, \Theta_I(P))$ under mild additional restrictions.

▸ calibration ▸ uniformity

Conditional moment inequalities

General set-up

- We observe an i.i.d. sample $\{Y_i, X_i\}_{i=1}^n$ where $Y_i \in \mathcal{Y} \subset \mathbb{R}^{d_Y}$ and $X_i \in \mathcal{X} \subset \mathbb{R}^{d_X}$ are distributed according to a probability distribution $P \in \mathcal{P}$.
- We consider a model where the parameter of interest θ is characterized by the following p conditional moment inequalities.

$$\mathbb{E}[m(W_i, \theta)|X_i] \geq 0_p \ a.s.$$

where $m : \mathbb{R}^{d_W} \times \Theta \to \mathbb{R}^p$ is a known measurable function and $W_i = (Y_i, X_i)$.

General set-up

- We observe an i.i.d. sample $\{Y_i, X_i\}_{i=1}^n$ where $Y_i \in \mathcal{Y} \subset \mathbb{R}^{d_Y}$ and $X_i \in \mathcal{X} \subset \mathbb{R}^{d_X}$ are distributed according to a probability distribution $P \in \mathcal{P}$.
- We consider a model where the parameter of interest θ is characterized by the following p conditional moment inequalities.

$$\mathbb{E}[m(W_i, \theta)|X_i] \geq 0_p \ a.s.$$

  where $m : \mathbb{R}^{d_W} \times \Theta \to \mathbb{R}^p$ is a known measurable function and $W_i = (Y_i, X_i)$.

- The identified set Θ_I is defined as follows:

$$\Theta_I = \{\theta \in \Theta \mid \mathbb{E}[m(W_i, \theta)|X_i] \geq 0_p \ a.s.\}$$

General set-up

- We observe an i.i.d. sample $\{Y_i, X_i\}_{i=1}^n$ where $Y_i \in \mathcal{Y} \subset \mathbb{R}^{d_Y}$ and $X_i \in \mathcal{X} \subset \mathbb{R}^{d_X}$ are distributed according to a probability distribution $P \in \mathcal{P}$.
- We consider a model where the parameter of interest θ is characterized by the following p conditional moment inequalities.

$$\mathbb{E}[m(W_i, \theta)|X_i] \geq 0_p \; a.s.$$

where $m : \mathbb{R}^{d_W} \times \Theta \to \mathbb{R}^p$ is a known measurable function and $W_i = (Y_i, X_i)$.

- The identified set Θ_I is defined as follows:

$$\Theta_I = \{\theta \in \Theta \mid \mathbb{E}[m(W_i, \theta)|X_i] \geq 0_p \; a.s.\}$$

- **Objective for the econometrician:** construct a **consistent** and **asymptotically valid** confidence region for Θ_I

The canonical estimation procedure

- Additional Challenges on top of the ones outlined in the unconditional moment inequalities case:
  - X is continuous $\implies$ the identified set is characterized by an infinite number of inequalities.
  - Conditional moments are non-parametric objects that are harder to estimate and display non-standard asymptotic properties (eg: no CLT, curse of dimensionality,...).

The canonical estimation procedure

- Additional Challenges on top of the ones outlined in the unconditional moment inequalities case:
  - X is continuous $\implies$ the identified set is characterized by an infinite number of inequalities.
  - Conditional moments are non-parametric objects that are harder to estimate and display non-standard asymptotic properties (eg: no CLT, curse of dimensionality,...).
- **Preeminent methods:** [Andrews and Shi, 2013], [Chernozhukov et al., 2013]

The canonical estimation procedure

- Additional Challenges on top of the ones outlined in the unconditional moment inequalities case:
  - X is continuous $\implies$ the identified set is characterized by an infinite number of inequalities.
  - Conditional moments are non-parametric objects that are harder to estimate and display non-standard asymptotic properties (eg: no CLT, curse of dimensionality,...).
- **Preeminent methods:** [Andrews and Shi, 2013], [Chernozhukov et al., 2013]
  - The leading method in [Andrews and Shi, 2013] transforms the conditional moment inequalities into a growing number of unconditional ones.
  - They consider a collection $\mathcal{N}$ of non-negative functions of X_i, denoted $\nu(X_i)$:

$$\theta \in \Theta_I \implies \mathbb{E}\left[m_j(W_i, \theta)\nu(X_i)\right] \geq 0, \ \forall j \in \{1, \ldots, p\}, \ \forall \nu \in \mathcal{N}$$

The canonical estimation procedure

- Additional Challenges on top of the ones outlined in the unconditional moment inequalities case:
  - X is continuous $\implies$ the identified set is characterized by an infinite number of inequalities.
  - Conditional moments are non-parametric objects that are harder to estimate and display non-standard asymptotic properties (eg: no CLT, curse of dimensionality,...).
- **Preeminent methods:** [Andrews and Shi, 2013], [Chernozhukov et al., 2013]
  - The leading method in [Andrews and Shi, 2013] transforms the conditional moment inequalities into a growing number of unconditional ones.
  - They consider a collection $\mathcal{N}$ of non-negative functions of X_i, denoted $\nu(X_i)$:

$$\theta \in \Theta_I \implies \mathbb{E}\left[m_j(W_i, \theta)\nu(X_i)\right] \geq 0, \ \forall j \in \{1, \ldots, p\}, \ \forall \nu \in \mathcal{N}$$

  - Under strong conditions on $\mathcal{N}$ (eg: $\mathcal{N}$ contains an infinite number of elements), the implication becomes an equivalence

challenges

An alternative characterization of the identified set

- Our approach relies on the following characterization of the sharp identified set:

$$\theta \in \Theta_I \iff m_{\theta,j}(X_i) \equiv \mathbb{E}\left(m_j(W_i,\theta)|X_i\right) \geq 0, \quad \forall \;\; j = 1, \ldots, p, \;\; a.s$$

$$\iff \min\left\{0, \min_{j=1,\ldots,p} m_{\theta,j}(X_i)\right\} = 0, \;\; a.s.$$

$$\iff \mathbb{E}\left[\min\left\{0, \min_{j=1,\ldots,p} m_{\theta,j}(X_i)\right\}\right] = 0$$

An alternative characterization of the identified set

- Our approach relies on the following characterization of the sharp identified set:

$$\begin{aligned}\theta \in \Theta_I &\iff m_{\theta,j}(X_i) \equiv \mathbb{E}\left(m_j(W_i,\theta)|X_i\right) \geq 0, \quad \forall \;\; j = 1, \ldots, p, \;\; a.s \\ &\iff \min\left\{0, \min_{j=1,\ldots,p} m_{\theta,j}(X_i)\right\} = 0, \;\; a.s. \\ &\iff \mathbb{E}\left[\min\left\{0, \min_{j=1,\ldots,p} m_{\theta,j}(X_i)\right\}\right] = 0\end{aligned}$$

- We transform p conditional moment inequalities into one unconditional moment equality without losing any identification power.

- Characterization akin to the one used in [Lee et al., 2013].

The smoothed-min approach for conditional moments

- **Main idea**: use the smooth approximation of the minimum to recover asymp. normal estimator for $\mathbb{E}\left[\min\{0, \min_{j=1,\dots,p} m_{\theta,j}(X_i)\}\right]$.

The smoothed-min approach for conditional moments

- **Main idea**: use the smooth approximation of the minimum to recover asymp. normal estimator for $\mathbb{E}\left[\min\{0, \min_{j=1,\dots,p} m_{\theta,j}(X_i)\}\right]$.

- **A pivotal test statistic**:

$$\xi_n(\theta) = \sqrt{n}\frac{1}{\sqrt{V_n}}\left(\frac{1}{n}\sum_{i=1}^{n}\underbrace{g_{\rho_n}(m_{\theta,n}(X_i))}_{(1)} + \underbrace{\nabla g_{\rho_n}(m_{\theta,n}(X_i))^\top (m(W_i,\theta) - m_{\theta,n}(X_i))}_{(2)}\right)$$

  where

  - $m_{\theta,n}(X_i)$ is a non-parametric estimator $\mathbb{E}\left(m(W_i,\theta)|X_i\right)$

The smoothed-min approach for conditional moments

- **Main idea**: use the smooth approximation of the minimum to recover asymp. normal estimator for $\mathbb{E}\left[\min\{0, \min_{j=1,\ldots,p} m_{\theta,j}(X_i)\}\right]$.

- **A pivotal test statistic**:

$$\xi_n(\theta) = \sqrt{n}\frac{1}{\sqrt{V_n}}\left(\frac{1}{n}\sum_{i=1}^{n}\underbrace{g_{\rho_n}(m_{\theta,n}(X_i))}_{(1)} + \underbrace{\nabla g_{\rho_n}(m_{\theta,n}(X_i))^\top (m(W_i,\theta) - m_{\theta,n}(X_i))}_{(2)}\right)$$

  where

  - $m_{\theta,n}(X_i)$ is a non-parametric estimator $\mathbb{E}\left(m(W_i,\theta)|X_i\right)$
  - (1) is the smoothed min operator and (2) is an orthogonalization term that ensures that the test statistic is "locally insensitive" to the fact that $m_\theta(X_i)$ is estimated.

The smoothed-min approach for conditional moments

- **Main idea**: use the smooth approximation of the minimum to recover asymp. normal estimator for $\mathbb{E}\left[\min\{0, \min_{j=1,\dots,p} m_{\theta,j}(X_i)\}\right]$.

- **A pivotal test statistic**:

$$\xi_n(\theta) = \sqrt{n}\frac{1}{\sqrt{V_n}}\left(\frac{1}{n}\sum_{i=1}^{n}\underbrace{g_{\rho_n}(m_{\theta,n}(X_i))}_{(1)} + \underbrace{\nabla g_{\rho_n}(m_{\theta,n}(X_i))^\top (m(W_i,\theta) - m_{\theta,n}(X_i))}_{(2)}\right)$$

  where

  - $m_{\theta,n}(X_i)$ is a non-parametric estimator $\mathbb{E}\left(m(W_i,\theta)|X_i\right)$
  - (1) is the smoothed min operator and (2) is an orthogonalization term that ensures that the test statistic is "locally insensitive" to the fact that $m_\theta(X_i)$ is estimated.
  - V_n a consistent estimator of the variance of: $V_0 = \lim_{n\to\infty} \mathrm{Var}[(1)+(2)]$

Confidence region

Our confidence region is as follows:

$$CR_n(1-\alpha) = \{\theta \in \Theta \mid \xi_n(\theta) \geq z_\alpha\}, \qquad \text{with } z_\alpha \text{ the } \alpha\text{-quantile of } \mathcal{N}(0,1)$$

Confidence region

Our confidence region is as follows:

$$CR_n(1-\alpha) = \{\theta \in \Theta \mid \xi_n(\theta) \geq z_\alpha\}, \qquad \text{with } z_\alpha \text{ the } \alpha\text{-quantile of } \mathcal{N}(0,1)$$

- **Remarks on the estimation of** $\mathbb{E}\left(m(W_i,\theta)|X_i\right)$
  - This step is the most demanding one in our procedure.
  - In most cases, $\mathbb{E}\left(m(W_i,\theta)|X_i\right)$ only needs to be estimated once for all the candidates θ in the grid.

Confidence region

Our confidence region is as follows:

$$CR_n(1-\alpha) = \{\theta \in \Theta \mid \xi_n(\theta) \geq z_\alpha\}, \qquad \text{with } z_\alpha \text{ the } \alpha\text{-quantile of } \mathcal{N}(0,1)$$

- **Remarks on the estimation of $\mathbb{E}\left(m(W_i,\theta)|X_i\right)$**
  - This step is the most demanding one in our procedure.
  - In most cases, $\mathbb{E}\left(m(W_i,\theta)|X_i\right)$ only needs to be estimated once for all the candidates θ in the grid.
- **Sample splitting:** following [Chernozhukov et al., 2018a], we use sample splitting for the estimation of $m_\theta(\cdot)$
  - We split the data into K samples and estimate $m_\theta(X_i)$ using all the sub-samples that don't contain observation i.
  - Sample splitting improves the finite sample performance and allows us to relax some regularity conditions likely violated when $dim(X_i)$ large.

Asymptotic properties of the test statistic

Proposition (Asymptotic properties of the test statistic)

Let ρ_n *a diverging sequence such that* $\rho_n = cn^{\alpha} + o(1)$ *with* $0 < \alpha < 2\gamma - \frac{1}{2}$*, Under mild regularity conditions ,* $\mathrm{CR}_n(1-\alpha)$ *is asymptotically valid and consistent, i.e.,*

1. *Asymptotic validity:*

$$\forall \theta \in \Theta_I, \quad \liminf_{n \to \infty} \Pr(\theta \in \mathrm{CR}_n(1-\alpha)) \geq 1 - \alpha.$$

2. *Consistency:*

$$\forall \theta \notin \Theta_I, \quad \Pr(\theta \in \mathrm{CR}_n(1-\alpha)) \underset{n \to \infty}{\longrightarrow} 0.$$

- Remark: our asymptotic results don't place any restrictions on the methods to be used to estimate $m_\theta(X_i)$

▸ proof

▸ regularity conditions

Monte Carlo simulations

Simulation setup conditional moment inequalities

$$\Theta_I = \{\theta \in \Theta \mid \mathbb{E}\left[m_j(\theta, W_i)|X_i\right] \geq 0 \text{ for } j = 1, ..., 6 \}$$

where the moment functions are defined as follows:

$$m_1(W_i, \theta) = -\theta_2 + (Y_{ij} + 3)$$
$$m_2(W_i, \theta) = \theta_2 + Y_{ij}$$
$$m_3(W_i, \theta) = \theta_2 + 4 - (1 + Y_{ij})\theta_1$$
$$m_4(W_i, \theta) = -\theta_2 + 1 + (1 + Y_{ij})\theta_1$$
$$m_5(W_i, \theta) = \theta_2 - 3 + (1 + Y_{ij})\theta_1$$
$$m_6(W_i, \theta) = -\theta_2 + 6 - (1 + Y_{ij})\theta_1$$

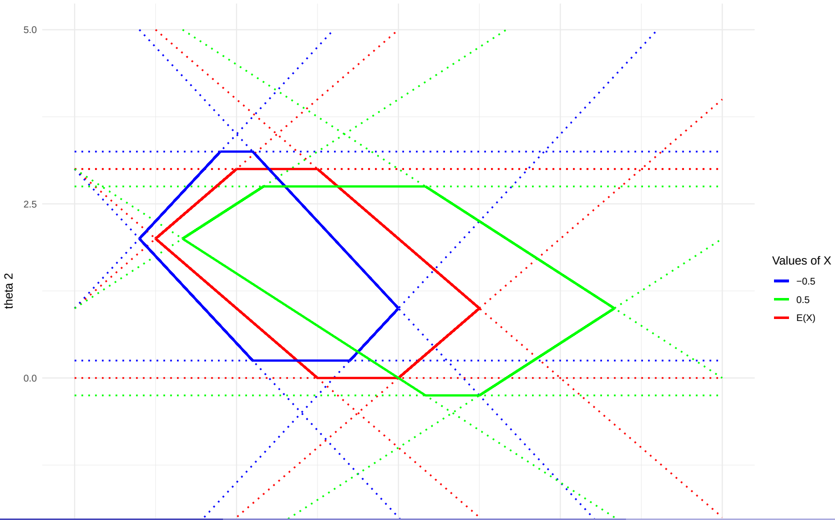
$$Y_{ij} = \frac{1}{2}(-\frac{1}{4} - X_i + X_i^2) + \varepsilon_{ij} \quad a.s.\text{with} \quad \mathbb{E}[\varepsilon_{ij}|X_i] = 0 \;\; a.s.$$

$X_i \sim U[-0.5, 0.5]$ and $\varepsilon_i \sim \mathcal{N}(0, 0.5)$ $\forall j$.

Methods: [Andrews and Shi, 2013], smoothed-min, a subset of the methods used for unconditional moment inequalities

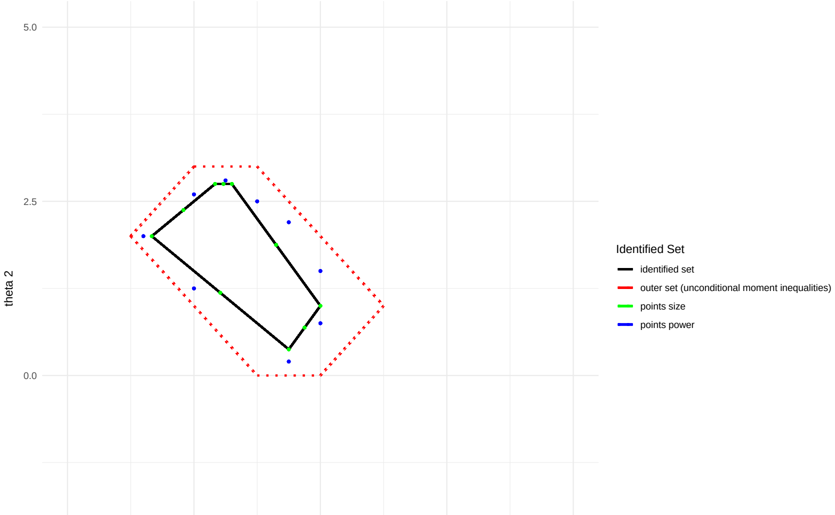
The identified set

Figure 1: Identified set in experimental design 4



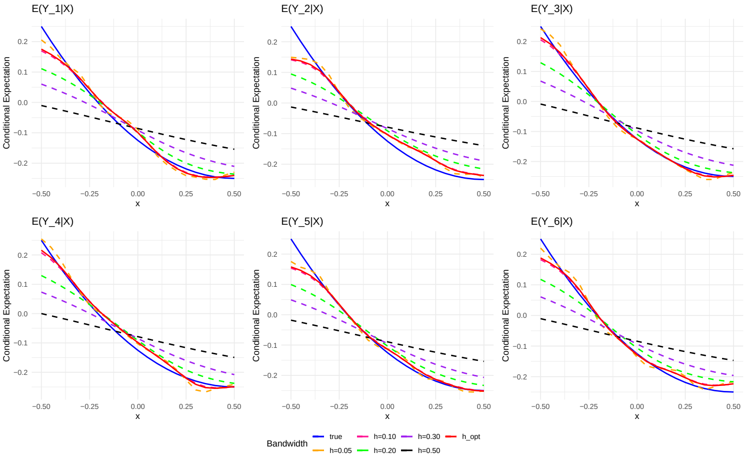
The identified set (II)

Figure 2: Identified set in experimental design 4



First stage estimator

Figure 3: First-stage kernel estimator



The bandwidth is chosen by cross-validation

Empirical size

Table 1: Null Rejection Probability (5000 replications)

Test statistic	critical value	n=250	n=500	n=1000	n=5000
smoothed min ($\rho = 1$)	z_α	0	0	0	0
smoothed min ($\rho = 5$)	z_α	0	0	0	0
smoothed min ($\rho = 10$)	z_α	0.013	0.009	0.004	0
smoothed min ($\rho = 50$)	z_α	0.048	0.039	0.036	0.024
smoothed min ($\rho = 100$)	z_α	0.052	0.042	0.045	0.036
smoothed min ($\rho = 200$)	z_α	0.052	0.042	0.047	0.043
CVM modified method of moments ($r_{1,n}$)	GMS	0.001	0.001	0	0
CVM standardized min ($r_{1,n}$)	GMS	0	0	0	0
CVM modified method of moments ($r_{2,n}$)	GMS	0	0.001	0	0
CVM standardized min ($r_{2,n}$)	GMS	0	0	0	0
CVM modified method of moments ($r_{3,n}$)	GMS	0	0	0	0
CVM standardized min ($r_{3,n}$)	GMS	0	0	0	0

The empirical size is the average of the empirical rejection probability over 10 points on the boundary of the identified set (5000 replications).

Empirical power

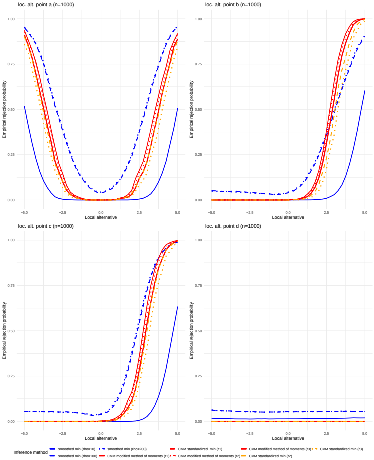
Table 2: Average power against fixed alternatives (5000 replications)

Test statistic	critical value	n=250	n=500	n=1000	n=5000
smoothed min ($\rho = 1$)	z_α	0	0	0	0
smoothed min ($\rho = 5$)	z_α	0.028	0.046	0.079	0.146
smoothed min ($\rho = 10$)	z_α	0.083	0.145	0.228	0.428
smoothed min ($\rho = 50$)	z_α	0.171	0.267	0.406	0.752
smoothed min ($\rho = 100$)	z_α	0.179	0.276	0.422	0.777
smoothed min ($\rho = 200$)	z_α	0.181	0.279	0.427	0.786
CVM modified method of moments ($r_{1,n}$)	GMS	0.076	0.147	0.259	0.608
CVM standardized min ($r_{1,n}$)	GMS	0.063	0.11	0.197	0.569
CVM modified method of moments ($r_{2,n}$)	GMS	0.067	0.139	0.259	0.606
CVM standardized min ($r_{2,n}$)	GMS	0.053	0.102	0.187	0.568
CVM modified method of moments ($r_{3,n}$)	GMS	0.058	0.129	0.247	0.596
CVM standardized min ($r_{3,n}$)	GMS	0.045	0.09	0.171	0.555

The empirical power is the average of the rejection probability over 10 points on the boundary of the identified set (5000 replications).

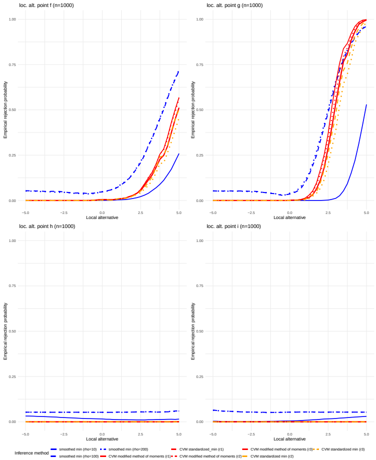
Power against local alternatives on vertices

Figure 4: Power against local alternatives of the form $\theta_2^v + \frac{1}{\sqrt{n}}$



Power against local alternatives on edges

Figure 5: Identified set in experimental design 4



Conclusion

Conclusion

- In this paper, we introduce a novel testing procedure for models characterized by conditional and unconditional moment inequalities.
- We derive a test statistic that is asymptotically normal by considering a smooth approximation of the minimum of the empirical moments.
- We show that our method can be easily adapted to handle conditional moment inequalities and remains consistent and asymptotically valid under weak regularity conditions.

Conclusion

- In this paper, we introduce a novel testing procedure for models characterized by conditional and unconditional moment inequalities.
- We derive a test statistic that is asymptotically normal by considering a smooth approximation of the minimum of the empirical moments.
- We show that our method can be easily adapted to handle conditional moment inequalities and remains consistent and asymptotically valid under weak regularity conditions.
- What remains to be done: propose a way to calibrate ρ_n in the case with conditional moment inequalities.

References I

Ackerberg, D., Chen, X., Hahn, J., and Liao, Z. (2014).
Asymptotic efficiency of semiparametric two-step gmm.
Review of Economic Studies, 81(3):919–943.

Andrews, D. W. (1994).
Asymptotics for semiparametric econometric models via stochastic equicontinuity.
Econometrica: Journal of the Econometric Society, pages 43–72.

Andrews, D. W. and Barwick, P. J. (2012).
Inference for parameters defined by moment inequalities: A recommended moment selection procedure.
Econometrica, 80(6):2805–2826.

Andrews, D. W. and Guggenberger, P. (2009).
Validity of subsampling and "plug-in asymptotic" inference for parameters defined by moment inequalities.
Econometric Theory, 25(3):669–709.

References II

Andrews, D. W. and Guggenberger, P. (2010).

Asymptotic size and a problem with subsampling and with the m out of n bootstrap.

Econometric Theory, 26(2):426–468.

Andrews, D. W. and Shi, X. (2013).

Inference based on conditional moment inequalities.

Econometrica, 81(2):609–666.

Andrews, D. W. and Soares, G. (2010).

Inference for parameters defined by moment inequalities using generalized moment selection.

Econometrica, 78(1):119–157.

Chen, X., Christensen, T. M., and Tamer, E. (2018).

Monte carlo confidence sets for identified sets.

Econometrica, 86(6):1965–2018.

References III

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018a).
Double/debiased machine learning for treatment and structural parameters.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2015).
Comparison and anti-concentration bounds for maxima of gaussian random vectors.
Probability Theory and Related Fields, 162(1):47–70.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2018b).
Inference on causal and structural parameters using many moment inequalities.
The Review of Economic Studies, 86(5):1867–1900.

Chernozhukov, V., Hong, H., and Tamer, E. (2007).
Estimation and confidence regions for parameter sets in econometric models 1.
Econometrica, 75(5):1243–1284.

References IV

Chernozhukov, V., Lee, S., and Rosen, A. M. (2013).
Intersection bounds: Estimation and inference.
Econometrica, 81(2):667–737.

Cox, G. and Shi, X. (2022).
Simple Adaptive Size-Exact Testing for Full-Vector and Subvector Inference in Moment Inequality Models.
The Review of Economic Studies.
rdac015.

Lee, S., Song, K., and Whang, Y.-J. (2013).
Testing functional inequalities.
Journal of Econometrics, 172(1):14–32.

Manski, C. F. and Tamer, E. (2002).
Inference on regressions with interval data on a regressor or outcome.
Econometrica, 70(2):519–546.

References V

Newey, W. K. (1994).
The asymptotic variance of semiparametric estimators.
Econometrica: Journal of the Econometric Society, pages 1349–1382.

Romano, J. P., Shaikh, A. M., and Wolf, M. (2014).
A practical two-step method for testing moment inequalities.
Econometrica, 82(5):1979–2002.

Rosen, A. M. (2008).
Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities.
Journal of Econometrics, 146(1):107–117.

A regression model where the outcome variable is partially observed

- Example taken from [Manski and Tamer, 2002].
- Assume that a latent outcome variable Y_i^* satisfies the following conditional mean restriction:

$$Y_i^* = \theta_1 + X_i\theta_2 + U_i \text{ where } \mathbb{E}[U_i|X_i] = 0 \ a.s.$$

- The econometrician only observes $[Y_{L,i}; Y_{U,i}]$ that contains Y_i^*. $Y_{L,i} = \lfloor Y_i \rfloor$ and $Y_{U,i} = \lfloor Y_i \rfloor + 1$.

A regression model where the outcome variable is partially observed

- Example taken from [Manski and Tamer, 2002].
- Assume that a latent outcome variable Y_i^* satisfies the following conditional mean restriction:

$$Y_i^* = \theta_1 + X_i\theta_2 + U_i \text{ where } \mathbb{E}[U_i|X_i] = 0 \; a.s.$$

- The econometrician only observes $[Y_{L,i}; Y_{U,i}]$ that contains Y_i^*. $Y_{L,i} = \lfloor Y_i \rfloor$ and $Y_{U,i} = \lfloor Y_i \rfloor + 1$.
- Without additional restrictions, one can show that θ must satisfy the following two conditional moment inequalities.

$$\mathbb{E}[\theta_1 + X_i\theta_2 - Y_{L,i}|X_i] \geq 0 \; a.s.$$
$$\mathbb{E}[Y_{U,i} - \theta_1 - X_i\theta_2|X_i] \geq 0 \; a.s.$$

back

Other challenges

1. **Selection and derivation of the moment inequalities:**
    - **Selection:** in many contexts (eg: games), the number of inequalities implied by the model quickly becomes intractable.
        - How to select a subset of inequalities while limiting the information loss?
    - **Derivation:** inequalities may come from equilibrium conditions that need to be simulated...

Other challenges

1. **Selection and derivation of the moment inequalities:**
   - **Selection:** in many contexts (eg: games), the number of inequalities implied by the model quickly becomes intractable.
     - How to select a subset of inequalities while limiting the information loss?
   - **Derivation:** inequalities may come from equilibrium conditions that need to be simulated...

5. **Subvector inference:** how to do inference efficiently on one of the parameters (not the full vector of parameters) or a known function of the parameters?

back

Our contribution

We provide a novel inference method based on a smooth approximation of the minimum across the empirical moments (and we let the smoothing decrease with n)

Our contribution

We provide a novel inference method based on a smooth approximation of the minimum across the empirical moments (and we let the smoothing decrease with n)

1. **Good statistical properties**.
   - The test statistic behaves asymptotically as the sum of a weighted sum of normals and a deterministic drift
     - Under H_0: the drift converges to 0 $\implies$ **Asymptotic normality**

Our contribution

We provide a novel inference method based on a smooth approximation of the minimum across the empirical moments (and we let the smoothing decrease with n)

1. **Good statistical properties**.
   - The test statistic behaves asymptotically as the sum of a weighted sum of normals and a deterministic drift
     - Under H_0: the drift converges to 0 $\implies$ **Asymptotic normality**
     - Under H_a, the drift diverges to $-\infty$ $\implies$ **consistency**

Our contribution

We provide a novel inference method based on a smooth approximation of the minimum across the empirical moments (and we let the smoothing decrease with n)

1. **Good statistical properties**.
   - The test statistic behaves asymptotically as the sum of a weighted sum of normals and a deterministic drift
     - Under H_0: the drift converges to 0 $\implies$ **Asymptotic normality**
     - Under H_a, the drift diverges to $-\infty$ $\implies$ **consistency**

Our contribution

We provide a novel inference method based on a smooth approximation of the minimum across the empirical moments (and we let the smoothing decrease with n)

1. **Good statistical properties**.
   - The test statistic behaves asymptotically as the sum of a weighted sum of normals and a deterministic drift
     - Under H_0: the drift converges to 0 $\implies$ **Asymptotic normality**
     - Under H_a, the drift diverges to $-\infty$ $\implies$ **consistency**
2. **Ease of implementation**
   - Test statistic and the critical value are straightforward to derive: no minimization, no simulations.
   - One tuning parameter: the smoothing parameter.

Our contribution

We provide a novel inference method based on a smooth approximation of the minimum across the empirical moments (and we let the smoothing decrease with n)

1. **Good statistical properties**.
   - The test statistic behaves asymptotically as the sum of a weighted sum of normals and a deterministic drift
     - Under H_0: the drift converges to 0 $\implies$ **Asymptotic normality**
     - Under H_a, the drift diverges to $-\infty$ $\implies$ **consistency**
2. **Ease of implementation**
   - Test statistic and the critical value are straightforward to derive: no minimization, no simulations.
   - One tuning parameter: the smoothing parameter.
3. Our test statistic can be adapted to handle **conditional** moment inequalities
   - **Asymptotic normality**
   - **Consistency**

example

Calibration of ρ: Bias

- Smoothing creates an identification bias: $g_{\rho_n}(m_\theta) \geq \min\{0, m_{\theta,1}, ... m_{\theta,p}\}$.

- For a fixed ρ, we can define the outer set $\Theta_I^o(\rho)$

$$\Theta_I^o(\rho) = \left\{ \theta \in \mathbb{R}^{\dim(\theta)} | \ g_\rho(m_\theta) = \frac{\sum_{j=1}^p m_{\theta,j} e^{-\rho m_{\theta,j}}}{1 + \sum_{j=1}^p e^{-\rho m_{\theta,j}}} \geq 0 \right\},$$

  - For any $\rho > 0$, $\Theta_I \subset \Theta_I^o(\rho)$
  - $\lim_{\rho \to +\infty} d_H(\Theta_I, \Theta_I^o(\rho)) = 0$, where d_H is the Hausdorff distance.

Calibration of ρ: Bias

- Smoothing creates an identification bias: $g_{\rho_n}(m_\theta) \geq \min\{0, m_{\theta,1}, ...m_{\theta,p}\}$.

- For a fixed ρ, we can define the outer set $\Theta_I^o(\rho)$

$$\Theta_I^o(\rho) = \left\{\theta \in \mathbb{R}^{\dim(\theta)} |\ g_\rho(m_\theta) = \frac{\sum_{j=1}^p m_{\theta,j} e^{-\rho m_{\theta,j}}}{1 + \sum_{j=1}^p e^{-\rho m_{\theta,j}}} \geq 0\right\},$$

  - For any $\rho > 0$, $\Theta_I \subset \Theta_I^o(\rho)$
  - $\lim_{\rho \to +\infty} d_H(\Theta_I, \Theta_I^o(\rho)) = 0$, where d_H is the Hausdorff distance.

- Asymptotically, smoothing has no effect because we let ρ_n diverge

- In finite sample, excessive smoothing negatively affects the power of our test.

Calibration of ρ: Bias

- Smoothing creates an identification bias: $g_{\rho_n}(m_\theta) \geq \min\{0, m_{\theta,1}, ...m_{\theta,p}\}$.

- For a fixed ρ, we can define the outer set $\Theta_I^o(\rho)$

$$\Theta_I^o(\rho) = \left\{\theta \in \mathbb{R}^{\dim(\theta)} |\ g_\rho(m_\theta) = \frac{\sum_{j=1}^p m_{\theta,j} e^{-\rho m_{\theta,j}}}{1 + \sum_{j=1}^p e^{-\rho m_{\theta,j}}} \geq 0\right\},$$

  - For any $\rho > 0$, $\Theta_I \subset \Theta_I^o(\rho)$
  - $\lim_{\rho \to +\infty} d_H(\Theta_I, \Theta_I^o(\rho)) = 0$, where d_H is the Hausdorff distance.

- Asymptotically, smoothing has no effect because we let ρ_n diverge

- In finite sample, excessive smoothing negatively affects the power of our test.

- To quantify this effect, we define a local measure of the distance between Θ_I and $\Theta_I^o(\rho_n)$ at θ.
  - we take the largest deviation $c_n < 0$ s.t. if θ were at the frontier of Θ_I, $(m_\theta, c_n) \in \Theta_I^o(\rho_n)$.

calibration of ρ: size distortion

- We show that the asymptotic expansion at the second order can be expressed as follows:

$$\sqrt{n} g_{\rho_n}(m_{\theta,n}) = \sqrt{n} g_{\rho_n}(m_\theta) + \nabla g_\rho(m_\theta)\sqrt{n}(m_{\theta,n} - m_\theta) + U_n + o_p\left(\frac{\rho_n}{\sqrt{n}}\right),$$

  - with $\mathbb{E}\left[U_n\right] = \frac{\rho_n}{\sqrt{n}} K_0(\theta)$
  - and $K_0(\theta)$ a negative constant that depends on the set of binding moments and the variance-covariance matrix Σ_θ.

- $K_0(\theta)$ can be estimated.

Calibration of ρ: loss function

- To solve the trade-off between the identification bias and the size distortion, we choose ρ_n^*

$$\rho_n = \underset{\rho>0}{argmin} \left\{ \frac{\rho_n}{\sqrt{n}} \lambda_{size}(|K_0(\theta)|) + \frac{1}{\rho_n} \lambda_{power}\left(LD(\theta)\right) \right\}$$

where λ_{size} and λ_{power} are increasing functions chosen by the researchers, $LD(\theta)$ is an upper bound on the local distance.

$$\rho_n^* = n^{1/4} \sqrt{\frac{\lambda_{power}\left(LD(\theta)\right)}{\lambda_{size}(|K_0(\theta)|)}}$$

- The "optimal" choice of ρ_n increases with the number of non-binding moments and decreases with the number of binding moments and the variance of these moments.

- The "optimal" speed of divergence $\alpha^* = \frac{1}{4}$ is also contained in $(0, 1/2)$

back

Asymptotic Uniform validity

- CR_n is asymptotically uniformly valid over the family of distributions $\mathcal{P}$ and over the points $\theta \in \Theta_I$ if:

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \inf_{\theta\in\Theta_I(P)} Pr(\theta \in CR_n) \geq 1 - \alpha.$$

where $\Theta_I(P) = \left\{\theta \in \Theta \mid \mathbb{E}_P[m(W_i, \theta)] \geq 0\right\}$.

- The uniform validity requirement is motivated by the observation that the asymptotic distributions of test statistics employed in moment inequality models often exhibit discontinuities
  - Confidence sets that are only valid pointwise can be deceptive in finite samples (on this topic, see [Andrews and Guggenberger, 2009] and [Andrews and Guggenberger, 2010])

- We show that if the moments have finite moments of order $2 + \delta$, then our confidence regions are asymptotically uniformly valid

back

Implementation and challenges in [Andrews and Shi, 2013]

- **Implementation:**
  (i) $\forall \nu \in \mathcal{N}$, one compute the test statistic associated with the unconditional moments generated by ν

  (ii) One must integrate the test statistics derived for each ν over a certain measure μ to construct the final CvM or KS test statistic

Implementation and challenges in [Andrews and Shi, 2013]

- **Implementation:**
  (i) $\forall \nu \in \mathcal{N}$, one compute the test statistic associated with the unconditional moments generated by ν

  (ii) One must integrate the test statistics derived for each ν over a certain measure μ to construct the final CvM or KS test statistic

  (iii) Derivation of the critical value

    a) $\forall \nu \in \mathcal{N}$, select the set of binding moments following a form of GMS procedure.

    b) Use this information to simulate the asymptotic distribution of CvM or KS test statistic under the null

Implementation and challenges in [Andrews and Shi, 2013]

- **Implementation:**
  (i) $\forall \nu \in \mathcal{N}$, one compute the test statistic associated with the unconditional moments generated by ν
  (ii) One must integrate the test statistics derived for each ν over a certain measure μ to construct the final CvM or KS test statistic
  (iii) Derivation of the critical value
    a) $\forall \nu \in \mathcal{N}$, select the set of binding moments following a form of GMS procedure.
    b) Use this information to simulate the asymptotic distribution of CvM or KS test statistic under the null

- **Practical and theoretical Challenges:**
  - Curse of dimensionality with $dim(X_i)$: in theory, $card(\nu)$ must increase exponentially with $dim(X_i)$ (increase in computational times, too few observations per ν).
  - Curse of dimensionality with p: total number of moments $\approx p \times card(\mathcal{N})$
  - Many tuning parameters: selection of the binding moments, $\mathcal{N}$, μ,...
  - Repeat the procedure for every θ in the grid!

back

Regularity conditions

Assumption (Regularity conditions on the moments)

1. $\exists\ C,\ \forall \theta \in \Theta, \forall j,\ \mathbb{E}[m_j(W_i, \theta)^2] < C.$

Assumption (Regularity conditions for the non-parametric estimator)

For all $\theta \in \Theta$, The estimator $\hat{m}_{\theta,-k}$ belongs to the class $\mathcal{M}_\theta$ that satisfies:

1. $\mathbb{E}\left[\|\hat{m}_{\theta,-k}(X_i) - m_\theta(X_i)\|_2^2\right]^{1/2} = o(n^{-\gamma})$ *with* $\gamma > 1/4$

2. $Z_{i,n} = \left(n^\gamma \|\hat{m}_{\theta,-k}(X_i) - m_\theta(X_i)\|_2\right)^2 \|m(W_i, \theta)\|_2^2$ *is uniformly integrable. That is:*

$$\forall n,\ \forall \varepsilon,\ \exists K > 0 \ \ \textit{such that:}\ \ \mathbb{E}(Z_{i,n}\mathbf{1}\{Z_{i,n} > K\}) \leq \varepsilon$$

back

Overview of the results: unconditional moment inequalities

- We consider 2 simulation designs: the one in [Andrews and Soares, 2010] and a static entry game.
- We compare our method with [Andrews and Soares, 2010] (min, MMM), [Chernozhukov et al., 2018b] (min), [Romano et al., 2014] (min, MMM)
- **Results:** calibration matters!
  - **Size**: calibrated smoothed-min slightly over-sized in some configurations (majority of binding moments+ small sample, $n = 250, 500$). **Rejection probability converges to the nominal size when n increases**

Overview of the results: unconditional moment inequalities

- We consider 2 simulation designs: the one in [Andrews and Soares, 2010] and a static entry game.
- We compare our method with [Andrews and Soares, 2010] (min, MMM), [Chernozhukov et al., 2018b] (min), [Romano et al., 2014] (min, MMM)
- **Results:** calibration matters!
  - **Size**: calibrated smoothed-min slightly over-sized in some configurations (majority of binding moments+ small sample, $n = 250, 500$). **Rejection probability converges to the nominal size when n increases**
  - **Power**: Smoothed min outperforms [Chernozhukov et al., 2018b] and [Romano et al., 2014], similar to [Andrews and Shi, 2013] (min) in most conf.
    - When negative correlation between the moments, smoothed min beats all the methods.

Overview of the results: unconditional moment inequalities

- We consider 2 simulation designs: the one in [Andrews and Soares, 2010] and a static entry game.
- We compare our method with [Andrews and Soares, 2010] (min, MMM), [Chernozhukov et al., 2018b] (min), [Romano et al., 2014] (min, MMM)
- **Results:** calibration matters!
  - **Size**: calibrated smoothed-min slightly over-sized in some configurations (majority of binding moments+ small sample, $n = 250, 500$). **Rejection probability converges to the nominal size when n increases**
  - **Power**: Smoothed min outperforms [Chernozhukov et al., 2018b] and [Romano et al., 2014], similar to [Andrews and Shi, 2013] (min) in most conf.
    - When negative correlation between the moments, smoothed min beats all the methods.
  - **Speed of implementation**: [Chernozhukov et al., 2018b]$>$ smoothing$>$ [Andrews and Soares, 2010]$>>$ [Romano et al., 2014]

Sketch of the proof

To prove our result, we use the decomposition below:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{g}_{\rho_n}(W_i, m_{\theta,n}) = \underbrace{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{g}_{\rho_n}(W_i, m_{\theta,n}) - \sum_{i=1}^{n}\tilde{g}_{\rho_n}(W_i, m_{\theta})\right)}_{A_n} +$$

$$\underbrace{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{g}_{\rho_n}(W_i, m_{\theta}) - \mathbb{E}\left[g_{\rho_n}(m_\theta(X_i))\right]\right)}_{B_n} + \underbrace{\sqrt{n}\mathbb{E}\left[g_{\rho_n}(m_\theta(X_i))\right]}_{C_n}$$

- First, we leverage key results in the literature on semi-parametric estimation([Newey, 1994], [Andrews, 1994], [Ackerberg et al., 2014], [Chernozhukov et al., 2018a]) to show that $A_n = o_p(1)$.

Sketch of the proof

To prove our result, we use the decomposition below:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{g}_{\rho_n}(W_i, m_{\theta,n}) = \underbrace{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{g}_{\rho_n}(W_i, m_{\theta,n}) - \sum_{i=1}^{n}\tilde{g}_{\rho_n}(W_i, m_{\theta})\right)}_{A_n} +$$

$$\underbrace{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{g}_{\rho_n}(W_i, m_{\theta}) - \mathbb{E}\left[g_{\rho_n}(m_\theta(X_i))\right]\right)}_{B_n} + \underbrace{\sqrt{n}\mathbb{E}\left[g_{\rho_n}(m_\theta(X_i))\right]}_{C_n}$$

- First, we leverage key results in the literature on semi-parametric estimation([Newey, 1994], [Andrews, 1994], [Ackerberg et al., 2014], [Chernozhukov et al., 2018a]) to show that $A_n = o_p(1)$.
- Second, we show that $B_n \xrightarrow{d} \mathcal{N}(0, V_0)$ by proving that the characteristic function of B_n converges to the characteristic function of $\mathcal{N}(0, V_0)$.
- Third, we prove that C_n is almost surely non-negative when $\theta \in \Theta_I$ and $C_n \xrightarrow{P} -\infty$ when $\theta \notin \Theta_I$. back