NONPARAMETRIC ESTIMATION AND INFERENCE UNDER SHAPE RESTRICTIONS

by

Joel L. Horowitz
Northwestern University

and

Sokbae Lee
Columbia University

INTRODUCTION

- This talk is about nonparametric estimation of the unknown function g in the model

$$Y = g(X) + \varepsilon; \quad E(\varepsilon \mid X) = 0.$$

- X is continuously distributed.
- g is assumed to satisfy a shape restriction
- Examples of shape restrictions:
    - Monotonicity or convexity
    - The Slutsky restriction of consumer theory
    - Increasing or decreasing returns to scale

WHY IMPOSE SHAPE RESTRICTIONS?

- Economic theory does not provide finite-dimensional parametric models but often provides shape restrictions such as monotonicity, convexity or the Slutsky restriction.
  - This motivates nonparametric estimation under shape restrictions.
- g can be estimated consistently and with the optimal rate of convergence without shape restrictions.

WHY SHAPE RESTRICTIONS? (2)

- Nonparametric estimates can be noisy and inconsistent with economic theory due to random sampling errors.
    - Blundell, Horowitz, and Parey (2012, 2016) found that nonparametric estimates of a demand function were wiggly and non-monotonic in the price.
- Shape restrictions can stabilize nonparametric estimates and make them consistent with economic theory without the need for arbitrary parametric or semiparametric restrictions.
- In addition, shape restrictions can improve estimation accuracy and reduce the widths of confidence bands.

FORM OF THE SHAPE RESTRICTION

- Write the restriction as $(Ag)(x) \leq 0$, where A is an operator.
- A is a linear operator if the shape restriction is monotonicity or convexity.
- A is a nonlinear operator if the shape restriction is the Slutsky condition.
- There is a large statistics literature on nonparametric estimation under monotonicity or convexity but not under a nonlinear shape restriction (e.g., Slutsky)

THE ESTIMATION PROBLEM

- Let $\{Y_i, X_i : i = 1,...,n\}$ be an independent random sample of (Y, X).

- In principle, g can be estimated by solving the problem

$$\hat{g}(x) = \arg\min_{f \in \mathcal{F}} \sum_{i=1}^{n} [Y_i - f(x)]^2$$

subject to:

$$(Af)(x) \leq 0 \quad \text{for all } x \in \operatorname{supp}(X).$$

  - $\mathcal{F}$ is a class of estimators, such as Nadaraya-Watson, local polynomial, or series estimators.

WHY INFERENCE IS DIFFICULT

- There are two sources of difficulty
  - The estimation problem has uncountably many constraints
  - The values of x for which the shape constraint is or is not binding are unknown.
    - The asymptotic distribution of $\hat{g}$ depends on where the constraint binds.
- The first problem could be dealt with easily if it were known where the shape constraint is binding.
  - Therefore, the second problem is more fundamental.

OUTLINE OF APPROACH

- To minimize complexity, assume that X is scalar and its density is bounded away from 0 on its support [0,1].
  - Extension to a multi-dimensional X involves mainly notational complications.
- Let $x_j = j/(J+1);\ \ j = 1,...,J$ be a grid of J points in [0,1].
  - Define $g_j = g(x_j)$
- We estimate and impose the shape restriction only on g_j's, thereby obtaining finitely many constraints.
  - $J \to \infty$ as $n \to \infty$, so estimation is consistent uniformly over $x \in [0,1]$.

EXAMPLES OF SHAPE RESTRICTIONS ON THE GRID

- Monotonicity

$$g_j - g_{j+1} \leq 0; \quad j = 1, \ldots, J-1$$

  - Use finite differences for faster convergence.

- A Slutsky-like nonlinear constraint (two-dimensional covariate: X, Z)

$$[g(x_{j+1}, z_k) - g(x_j, z_k)]$$

$$+ g(x_j, z_k)[g(x_j, z_{k+1}) - g(x_j, z_k)] \leq 0$$

REPRESENTING CONSTRAINTS ON THE GRID

- Define the vector $\boldsymbol{g} = (g_1, ..., g_J)'$.
- Write the shape restrictions on the grid as

$$A_k(\boldsymbol{g}) \leq 0; \quad k = 1, ..., \kappa,$$

where A_k is a real-valued function $\mathbb{R}^J$ and there are κ constraints.

- Example: monotonicity

$$A_k(\boldsymbol{g}) = g_k - g_{k+1}; \; k = 1, ..., J - 1 = K$$

ESTIMATION ON THE GRID

- We use two estimators, one that does not impose shape restrictions, and one that does.
- The unrestricted estimator is used to find the set of "possibly binding" constraints.
- Both estimators are local quadratic with bandwidths $h \propto n^{-1/5}$.
  - Local quadratic estimation with $h \propto n^{-1/5}$ provides an estimator that is free of asymptotic bias.
  - The bandwidth can be chosen by standard cross-validation or plug-in methods for local constant or local linear estimation.

UNRESTRICTED ESTIMATION

- For $j = 1, \ldots, J$, $\tilde{g}(x_j) = \tilde{b}_0$, where

$$\tilde{\boldsymbol{b}} = \arg \min_{b_0, b_1, b_2} \sum_{i=1}^{n} \left[Y_i - b_0 - b_1(X_i - x_j) + b_2(X_i - x_j)^2 \right]^2 K\left(\frac{X_i - x_j}{h} \right)$$

- Let the distance between grid points exceed $2h$.
  - Specifically, $J \ll n^{1/5}(\log n)^{-1/4}$.
  - Then $\tilde{g}(x_j)$ and $\tilde{g}(x_k)$ are statistically independent if $j \neq k$.

UNRESTRICTED ESTIMATION (2)

- Define $\boldsymbol{g} = [g(x_1),...,g(x_J)]'$ and $\tilde{\boldsymbol{g}} = [\tilde{g}(x_1),...,\tilde{g}(x_J)]'$.
- Under standard smoothness conditions,

$$(nh)^{1/2}(\tilde{\boldsymbol{g}} - \boldsymbol{g}) \to^d N(0, \omega)$$

  - ω is a $J \times J$ diagonal matrix whose diagonal elements can be estimated consistently.
- This is a standard result when J is fixed, and a multivariate version of the Berry-Esséen theorem (Bentkus 2003) can be used to show that it remains true as $J \to \infty$

FINDING THE SET OF POSSIBLY BINDING CONSTRAINTS, $\tilde{S}$

- Let $\mathcal{S}$ denote the set of constraints that bind in the population and $\tilde{\mathcal{S}}$ denote the data-based set of possibly binding constraints.

- Write the k'th constraint as $A_k(\boldsymbol{g}) \leq 0$.

- Define $\mathcal{M}_{Cn}$ by

$$\mathcal{M}_{Cn} = \{k : C^{-1}(nh)^{-1/2} \leq -A_k(\boldsymbol{g}) \leq C[(\log n)/nh]^{1/2}\}.$$

- Define $\bar{\mathcal{M}} = \{k : k \notin \mathcal{M}_{Cn}\}$

ASSUMPTIONS ABOUT THE CONSTRAINTS

- Assume:
  - Each A_k is twice differentiable and depends on only q components of $\boldsymbol{g}$ for some fixed q.
  - $|\mathcal{M}_{Cn}| \to 0$ as $n \to \infty$.
- The first assumption is satisfied by shape restrictions important in applications (e.g. monotonicity, Slutsky).
- The second is satisfied by these shape restrictions under the assumed h and grid spacing and permits some constraints to be "nearly binding."

FINDING $\tilde{\mathcal{S}}$

- Define $c_n = (\log n)^{1/2}$. Let κ be the number of constraints. Define $\boldsymbol{A} = (A_1, ..., A_\kappa)'$.

- It follows from asymptotic normality of $(nh)^{1/2}(\tilde{\boldsymbol{g}} - \boldsymbol{g})$ and differentiability of the constraints that

$$(nh)^{1/2}[\boldsymbol{A}(\tilde{\boldsymbol{g}}) - \boldsymbol{A}(\boldsymbol{g})] \to^d N(0, \Upsilon)$$

where Υ is a consistently estimable matrix.

FINDING $\tilde{\mathcal{S}}$ (2)

- Therefore,

$$P[\Upsilon_{kk}^{-1/2}(nh)^{1/2}\,|\,A_k(\tilde{\boldsymbol{g}})-A_k(\boldsymbol{g})\,|>c_n \text{ for any } k\le\kappa]=o(1)$$

- Set $\tilde{\mathcal{S}}=\{k:\ -\Upsilon_{kk}^{-1/2}(nh)^{1/2}A_k(\tilde{\boldsymbol{g}})\le c_n\}$.
  - Then $P(\tilde{\mathcal{S}}=\mathcal{S})\to 1$.
  - Same if $\Upsilon_{kk}^{-1/2}$ is estimated.

CONSTRAINED ESTIMATION

- The constrained local quadratic estimator is $\hat{\boldsymbol{g}} = \hat{\boldsymbol{b}}_0$, where

$$\tilde{\boldsymbol{b}} = \arg \min_{b_0, b_1, b_2}$$

$$\sum_{i=1}^{n} \sum_{j=1}^{J} \left[Y_i - b_{0j} - b_{1j}(X_i - x_j) + b_{2j}(X_i - x_j)^2 \right]^2 K\left(\frac{X_i - x_j}{h} \right)$$

subject to: $A_k(\boldsymbol{g}) = 0; \quad k \in \tilde{\mathcal{S}}$

CONSTRAINED ESTIMATION (2)

- If the A_k's are linear (e.g., monotonicity, convexity), this is a standard constrained least squares problem with an analytic solution.

- The solution gives

$$(\hat{\boldsymbol{g}} - \boldsymbol{g}) = \Xi(\tilde{\boldsymbol{g}} - \boldsymbol{g}) + r_n$$

where Ξ is a matrix that can be estimated and $\|r_n\| = o_p(n^{-1/2})$.

INFERENCE WITH LINEAR CONSTRAINTS

- Asymptotic inference can be carried out as if $\mathcal{S}$ were known, because $P(\tilde{\mathcal{S}} = \mathcal{S}) \to 1$.

- It follows from this and asymptotic normality of $(nh)^{1/2}(\tilde{\boldsymbol{g}} - \boldsymbol{g})$ that

$$\lim_{n\to\infty} \sup_{\boldsymbol{g}\in\bar{\mathcal{M}}_n} \sup_{\boldsymbol{t}} | P[(nh)^{1/2}(\hat{\boldsymbol{g}} - \boldsymbol{g}) \le \boldsymbol{t}] - P(Z \le \boldsymbol{t}) | = 0,$$

  where $Z \sim N(0, \Sigma_n)$ and Σ_n is a matrix whose components can be estimated consistently.

UNIFORMITY

- The foregoing result holds uniformly over a class of functions in which some constraints may be nearly binding
- The class does not include functions $\boldsymbol{g} \in \mathcal{M}_{Cn}$ so there is a "gap" in the class of functions for which uniformity holds.
- A similar gap arises in penalized estimation of high-dimensional models (Bühlmann and van de Geer, 2011; Horowitz and Huang, 2013).
- Results of Leeb and Pötscher indicate that removing the gap for the large class of shape restrictions treated here would be difficult or impossible.

UNIFORMITY (cont.)

- In moment inequalities, it is not known which constraints do or do not bind.
- Uniformity is achieved without a gap by relaxing the constraints slightly so they bind only asymptotically.
- We do not adopt this approach here because:
  - It would allow the shape restrictions to be violated in finite samples, which would have undesirable consequences in some applications.
  - Asymptotically, the constrained confidence bands would be as wide as the unconstrained ones.

POINTWISE CONFIDENCE INTERVALS

- These can be obtained in the usual way.
- Pointwise confidence intervals for the g_j's are

$$\gamma_{j1} \leq (nh)^{1/2}(\hat{g}_j - g_j) \leq \gamma_{n2}$$

- The critical values can be obtained from the $N(0, \Sigma_n)$ distribution.

UNIFORM CONFIDENCE BAND

- As before, let $Z \sim N(0, \Sigma_n)$.
- A symmetrical uniform $(1-\alpha)$ confidence band is

$$\{g(x_1),....,g(x_J): -\gamma_j \le \hat{g}(x_j) - g(x_j) \le \gamma_j;\ \ j = 1,...,J\}$$

- The critical values γ_j solve

$$P\left[\bigcap_{j=1}^{J} [|Z_j| \le (nh)^{1/2}\gamma_j]\right] - (1-\alpha).$$

The γ_j's can be found by simulation by sampling the $N(0, \Sigma_n)$ distribution.

UNIFORM CONFIDENCE BAND (2)

- The grid points become dense in [0,1] as $n \to \infty$.
- Therefore, the confidence band is asymptotically uniform over [0,1] as $n \to \infty$.
- The minimum average width shape restricted confidence band can be found by solving

$$\min_{\gamma_1,\ldots,\gamma_J} \sum_{j=1}^{J} \gamma_j \text{ subj. to } P\left[\bigcap_{j=1}^{J} [|Z_j| \le (nh)^{1/2}\gamma_j]\right] = (1-\alpha)$$

  - Shape restrictions do not affect the rate of convergence of the width but reduce the constant that multiplies the rate.

NONLINEAR CONSTRAINTS

- The unconstrained estimator is uniformly consistent.
    - Therefore, it satisfies the constraints with probability approaching 1 as $n \to \infty$.
- In a finite sample, the random sampling errors may cause the unconstrained estimator to violate the constraints.
    - But the magnitude of the violation is small if n is large.
- Therefore, the constrained estimation problem can be regarded as a perturbation of the unconstrained problem and treated by the theory of sensitivity analysis in nonlinear programming (Fiacco 1983).

NONLINEAR CONSTRAINTS

- The sensitivity calculations are messy, but the result is simple.
  - Replace the nonlinear constraint functions with the linear approximations obtained from first-order Taylor series expansions.
  - Then proceed as with linear constraints.
- Therefore, the pointwise and uniform confidence bands obtained with linear constraints also apply to nonlinear constraints.

FINITE SAMPLE CORRECTION

- Asymptotically, $\tilde{\mathcal{S}} = \mathcal{S}$ with probability approaching 1.
  - In a finite sample, constraints for which $A_k(\boldsymbol{g}) < 0$ may be included in $\tilde{\mathcal{S}}$ and erroneously treated as binding in constrained estimation.
  - The resulting bias can cause severe under coverage.
- The possibility of a large error in cov. prob. can be reduced at the cost of a wider confidence band by
  - Reducing size of $\tilde{\mathcal{S}}$
  - Including a bias correction term in the asymptotic distribution of $(nh)^{1/2}(\hat{\boldsymbol{g}} - \boldsymbol{g})$.

FINITE SAMPLE PROCEDURE

- Redefine the set of possibly binding constraints as

$$\hat{\mathcal{S}} = \{k : -\Upsilon_{kk}^{-1/2}(nh)^{1/2} A_k(\tilde{\boldsymbol{g}}) \leq c(\log n)^{1/2}\}$$

  for some $c > 0$.

  - Choose upper bound m on acceptable bias due to erroneously treating a non-binding constraint as binding.
    - m is user-selected tuning parameter
  - Choose c to so that the maximum bias of constraints in $\hat{\mathcal{S}}$ does not exceed m.

FINITE SAMPLE PROCEDURE (2)

- Find confidence limits from

$$P\left[\bigcap_{j=1}^{J}\{-\gamma_j + m_j \le Z_j \le \gamma_j + m_j\right] = 1-\alpha$$

- $Z = (Z_1,...,Z_J) \sim N[0,(nh)^{-1}\hat{\Sigma}_n]$
- $m_j = m$ if $k \in \mathcal{S}$; $m_j = 0$ if $k \notin \mathcal{S}$.
- The confidence band based on the bias-corrected estimate can be as wide as the unconstrained band
    - But the bias corrected estimate satisfies the shape restriction. The unconstrained estimator does not.

MONTE CARLO EVIDENCE

- Design mimics empirical application of estimating a production function for the Chinese chemical industry

- Model

$$Y = [(K^{1/2} + L^{1/2})^{2\tau}]e^{U};\ U \sim N(0, 0.01)$$

- Shape constraint: Non-increasing returns to scale
  - $\tau = 1$: Constant returns to scale
  - $\tau = 0.9$: Slightly decreasing returns to scale
  - $\tau = 0.5$: Strongly decreasing returns to scale

- $n = 1000$, $m = 0.005$, and $n = 2000$

RESULTS

Estimator	n	τ	Cov. prob.	Rel. width
Unconstrained	1000	1.0	0.95	1.0
Constrained			0.95	0.82
Oracle			0.96	0.82
Bias Corr.			0.95	0.90
Unconstrained		0.9	0.94	1.0
Constrained			0.74	0.83
Oracle			0.94	1.0
Bias Corr.			0.91	0.97
Unconstrained		0.5	0.94	1.0
Constrained			0.94	1.0
Oracle			0.94	1.0
Bias Corr.			0.94	1.0

RESULTS

Estimator	n	τ	Cov. prob.	Rel. width
Unconstrained	2000	1.0	0.95	1.0
Constrained			0.95	0.82
Oracle			0.96	0.82
Bias Corr.			0.945	0.91
Unconstrained		0.9	0.95	1.0
Constrained			0.66	0.84
Oracle			0.95	1.0
Bias Corr.			0.93	0.98
Unconstrained		0.5	0.95	1.0
Constrained			0.95	1.0
Oracle			0.95	1.0
Bias Corr.			0.96	1.0

CONCLUSIONS

- Talk has been about nonparametric inference of a conditional mean function under a shape restriction
    - The main result is a uniform confidence band for the unknown, shape-restricted conditional mean function.
    - Method accommodates nonlinear constraints such as the Slutsky restriction and non-increasing returns to scale.
- There is a large statistics and econometrics literature on estimation under monotonicity and convexity/concavity.
        - To our knowledge, this is first construction of uniform confidence bands under nonlinear shape restrictions.