Nonparametric Instrumental Variable Estimation
Under Monotonicity

Denis Chetverikov ¹  Daniel Wilhelm ²

¹University of California Los Angeles
²University College London & CeMMAP
Motivation:

NPIV estimators may suffer from very slow convergence rates.
Motivation:
NPIV estimators may suffer from very slow convergence rates

Approach: impose additional restrictions:
(a) monotone regression function
(b) monotone instrument

Results:
1. Ill-posedness is weakened
2. NPIV estimator: imposing monotonicity yields convergence rate that is not worse
3. Fast convergence rate local to constants
4. Non-asymptotic risk bound: tightens as distance to constants shrinks
5. Simulations: dramatic finite sample performance improvements
Motivation:
NPIV estimators may suffer from very slow convergence rates

Approach: impose additional restrictions:
(a) monotone regression function
(b) monotone instrument

Results:
1. **ill-posedness** is weakened
Motivation:
NPIV estimators may suffer from very slow convergence rates

Approach: impose additional restrictions:
(a) monotone regression function
(b) monotone instrument

Results:
1. **ill-posedness** is weakened
2. **NPIV estimator**:
   - imposing monotonicity yields convergence rate that is not worse
   - fast convergence rate local to constants
Motivation:
NPIV estimators may suffer from very slow convergence rates

Approach: impose additional restrictions:
(a) monotone regression function
(b) monotone instrument

Results:
1. ill-posedness is weakened
2. NPIV estimator:
   - imposing monotonicity yields convergence rate that is not worse
   - fast convergence rate local to constants
3. non-asymptotic risk bound: tightens as distance to constants shrinks
Motivation:
NPIV estimators may suffer from very slow convergence rates

Approach: impose additional restrictions:
(a) monotone regression function
(b) monotone instrument

Results:
1 ill-posedness is weakened
2 NPIV estimator:
   ■ imposing monotonicity yields convergence rate that is not worse
   ■ fast convergence rate local to constants
3 non-asymptotic risk bound: tightens as distance to constants shrinks
4 simulations: dramatic finite sample performance improvements
Overview

1. Motivation and Background

2. Bounds on Restricted Measure of Ill-Posedness

3. Non-Asymptotic Risk Bounds

4. Simulations

5. Gasoline Demand in the U.S.
Overview

1 Motivation and Background

2 Bounds on Restricted Measure of Ill-Posedness

3 Non-Asymptotic Risk Bounds

4 Simulations

5 Gasoline Demand in the U.S.
NPIV model:

\[ Y = g(X) + \varepsilon, \quad E[\varepsilon | W] = 0 \]

where

- \( Y \) dependent variable (continuous)
- \( X \) endogenous covariate (continuous)
- \( W \) instrumental variable (continuous)
- all variables are scalar
NPIV vs. Linear IV and Nonparametric Regression

\[ Y = g(X) + \varepsilon \]

1. **Linearity with endogeneity**
   - \( g(x) = \beta_0 + \beta_1 x \) and \( E[\varepsilon | W] = 0 \)
   - TSLS estimator:
     \[ |\hat{\beta} - \beta| = O_p(n^{-1/2}) \]

2. **Nonlinearity without endogeneity**

3. **NPIV: Nonlinearity with endogeneity**
   - \( g(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + ... \) and \( E[\varepsilon | W] = 0 \)

   TSLS regression of \( Y \) on 1, \( X \), \( X^2 \), ..., \( X^K \) using IVs 1, \( W \), \( W^2 \), ..., \( W^J \)

   This problem is much harder than the two above!
NPIV vs. Linear IV and Nonparametric Regression

\[ Y = g(X) + \varepsilon \]

1 **Linearity with endogeneity**
   - \( g(x) = \beta_0 + \beta_1 x \) and \( E[\varepsilon | W] = 0 \)
   - TSLS estimator:
     \[ |\hat{\beta} - \beta| = O_p(n^{-1/2}) \]

2 **Nonlinearity without endogeneity**
   - \( g(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots \) and \( E[\varepsilon | X] = 0 \)
   - Regress \( Y \) on \( 1, X, X^2, \ldots, X^{K_n} \):
     \[ \|\hat{g} - g\| = O_p(n^{-s/(2s+1)}) \]
Y = g(X) + \varepsilon

1. **Linearity with endogeneity**
   - \( g(x) = \beta_0 + \beta_1 x \) and \( E[\varepsilon|W] = 0 \)
   - TSLS estimator:
     \[ |\hat{\beta} - \beta| = O_p(n^{-1/2}) \]

2. **Nonlinearity without endogeneity**:
   - \( g(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots \) and \( E[\varepsilon|X] = 0 \)
   - Regress \( Y \) on \( 1, X, X^2, \ldots, X^{K_n} \):
     \[ \|\hat{g} - g\| = O_p(n^{-s/(2s+1)}) \]

3. **NPIV: Nonlinearity with endogeneity**:
   - \( g(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots \) and \( E[\varepsilon|W] = 0 \)
   - TSLS regression of \( Y \) on \( 1, X, X^2, \ldots, X^{K_n} \) using IVs \( 1, W, W^2, \ldots, W^{J_n} \)
   - This problem is much harder than the two above!
NPIV model:

\[ Y = g(X) + \varepsilon, \quad E[\varepsilon|W] = 0 \]
Comparison to Linear Model

NPIV model:

\[ Y = g(X) + \varepsilon, \quad E[\varepsilon | W] = 0 \]

If \( g(x) = \beta x \):

\[ E[YW] = \beta E[XW] \]
Comparison to Linear Model

NPIV model:

\[ Y = g(X) + \varepsilon, \quad E[\varepsilon|W] = 0 \]

If \( g(x) = \beta x \):

\[ E[YW] = \beta E[XW] \]

- assume \( E[XW] \) has nonzero eigenvalues
- invert \( E[XW] \) to solve for \( \beta \)
NPIV model:

\[ Y = g(X) + \varepsilon, \quad E[\varepsilon|W] = 0 \]
Operator Representation of the NPIV Model

NPIV model:

\[ Y = g(X) + \varepsilon, \quad E[\varepsilon | W] = 0 \]

If \( g(x) \) is nonparametric:

\[ E[Y | W] = E[g(X) | W] \]
Operator Representation of the NPIV Model

NPIV model:

\[ Y = g(X) + \varepsilon, \quad E[\varepsilon|W] = 0 \]

If \( g(x) \) is nonparametric:

\[ E[Y|W] = E[g(X)|W] \]
\[ \Leftrightarrow \quad m = Tg \]

where

- \( m(w) := E[Y|W = w]f_w(w) \)
- \( Th(w) := \int h(x)f_{X,W}(x, w)dx \)
Ill-posedness

\[ g = T^{-1}m \]
Ill-posedness

\[ g = T^{-1} m \]

- singular values of \( T \) tend to zero
Ill-posedness

\[ g = T^{-1} m \]

- singular values of \( T \) tend to zero
- small changes in \( m \) may translate into large changes in \( g \)
Ill-posedness

\[ g = T^{-1} m \]

- Singular values of \( T \) tend to zero
- Small changes in \( m \) may translate into large changes in \( g \)

Recovering \( g \) from estimates of \( m \) and \( T \) is an **ill-posed inverse problem**!
Ill-posedness means $T^{-1}$ is not continuous:

$$\|T(g_1 - g_2)\| \text{ small} \quad \not\Rightarrow \quad \|g_1 - g_2\| \text{ small}$$
Ill-posedness means $T^{-1}$ is not continuous:

$$\| T(g_1 - g_2) \| \text{ small} \not\Rightarrow \| g_1 - g_2 \| \text{ small}$$

Measure of ill-posedness:

$$\tau_n = \sup_{h \in H_{K_n}} \frac{\|h\|}{\|Th\|}$$

where $\{H_{K_n}\}_{n \geq 1}$ is a sequence of sieve spaces of dim. $K_n$. 
Ill-posedness means $T^{-1}$ is not continuous:

\[ \| T(g_1 - g_2) \| \text{ small} \not\Rightarrow \| g_1 - g_2 \| \text{ small} \]

Measure of ill-posedness:

\[ \tau_n = \sup_{h \in \mathcal{H}_{K_n}} \frac{\| h \|}{\| Th \|} \to \infty \quad K_n \to \infty \]

where $\{ \mathcal{H}_{K_n} \}_{n \geq 1}$ is a sequence of sieve spaces of dim. $K_n$. 

Chetverikov and Wilhelm (UCLA and UCL) Monotone Nonparametric IV
Measure of Ill-Posedness

Ill-posedness means $T^{-1}$ is not continuous:

$$\|T(g_1 - g_2)\| \text{ small} \not\Rightarrow \|g_1 - g_2\| \text{ small}$$

Measure of ill-posedness:

$$\tau_n = \sup_{h \in \mathcal{H}_{K_n}} \frac{\|h\|}{\|Th\|} \to \infty \quad K_n \to \infty$$

where $\{\mathcal{H}_{K_n}\}_{n \geq 1}$ is a sequence of sieve spaces of dim. $K_n$.

Implication:

$\tau_n$ multiplies variance and slows down the convergence rate:

$$\|\hat{g} - g\|_2 = O_p \left( \tau_n(K_n/n)^{1/2} + K_n^{-s} \right)$$
$\tau_n \to \infty$ faster the faster the singular values of $T$ tend to zero:
Types of Ill-posedness

\( \tau_n \to \infty \) faster the faster the singular values of \( T \) tend to zero:

- **mildly ill-posed:** \( \tau_n = O(K_n^r) \)

\[
\| \hat{g} - g \|_2 = O_p(K_n^r(K_n/n)^{1/2} + K_n^{-s}) = O(n^{2r-2s+1})
\]
Types of Ill-posedness

\[ \tau_n \to \infty \] faster the faster the singular values of \( T \) tend to zero:

- **mildly ill-posed:** \( \tau_n = O(K_n^r) \)

\[
\| \hat{g} - g \|_2 = O_p(K_n^r(K_n/n)^{1/2} + K_n^{-s}) = O(n^{2r+2s+1})
\]

- **severely ill-posed:** \( \tau_n = O(e^{cK_n}) \)

\[
\| \hat{g} - g \|_2 = O_p(e^{cK_n}(K_n/n)^{1/2} + K_n^{-s}) = O((\log n)^{-s})
\]
Effect of Imposing Monotonicity

Chetverikov and Wilhelm (UCLA and UCL) Monotone Nonparametric IV
Effect of Imposing Monotonicity

Chetverikov and Wilhelm (UCLA and UCL) Monotone Nonparametric IV
Effect of Imposing Monotonicity

Chetverikov and Wilhelm (UCLA and UCL) Monotone Nonparametric IV
Consider two monotonicity assumptions:

1. $g$ is monotone
2. $W$ is a monotone IV (roughly: $w_2 \geq w_1 \Rightarrow F_{X|W=w_2} \succeq F_{X|W=w_1}$)
Consider two monotonicity assumptions:

1. $g$ is monotone
2. $W$ is a monotone IV (roughly: $w_2 \geq w_1 \Rightarrow F_{X|W=w_2} \geq F_{X|W=w_1}$)

Central result:

$\tau_n$, slightly modified, is bounded on the set of monotone functions
Boundary Effects

Monotonicity constraint:

\[ g'(x) \geq 0 \quad \forall x \]
Boundary Effects

Monotonicity constraint:

\[ g'(x) \geq 0 \quad \forall x \]

1. **Interior**: \( g \) is strictly monotone
   - constrained and unconstrained estimator are equal wpa 1
   - no convergence rate improvements possible

Chetverikov and Wilhelm (UCLA and UCL) Monotone Nonparametric IV
Boundary Effects

Monotonicity constraint:

\[ g'(x) \geq 0 \quad \forall x \]

1. **Interior**: \( g \) is strictly monotone
   - constrained and unconstrained estimator are equal wpa 1
   - no convergence rate improvements possible

2. **Boundary**: \( g \) is constant
   - convergence rate improvements possible

Implications of bounded \( \tau \):

1. fast convergence rate on the boundary
2. fast convergence rate also in slowly shrinking neighborhood of the boundary
3. finite sample performance depends on distance to boundary
Boundary Effects

Monotonicity constraint:

\[ g'(x) \geq 0 \quad \forall x \]

1 Interior: \( g \) is strictly monotone
   - constrained and unconstrained estimator are equal wpa 1
   - no convergence rate improvements possible

2 Boundary: \( g \) is constant
   - convergence rate improvements possible

Implications of bounded \( \tau \):

1 fast convergence rate on the boundary
2 fast convergence rate also in slowly shrinking neighborhood of the boundary
3 finite sample performance depends on distance to boundary
NPIV estimators:

- surveys: Carrasco, Florens, and Renault (2007), Horowitz (2011)

NPIV optimality:

- Hall and Horowitz (2005), Chen and Reiß (2011), Chen and Christensen (2013)

NPIV and shape restrictions:

- Matzkin (1994)
Monotone nonparametric regression:


Risk bounds for monotone regression:

Overview

1 Motivation and Background

2 Bounds on Restricted Measure of Ill-Posedness

3 Non-Asymptotic Risk Bounds

4 Simulations

5 Gasoline Demand in the U.S.
Sieve space of functions with derivative bounded from below:

$$\mathcal{H}_K(a) := \left\{ h \in \mathcal{H}_K : \inf_{x \in [0,1]} h'(x) \geq -a \right\}$$
A Restricted Measure of Ill-posedness

Sieve space of functions with derivative bounded from below:

\[ \mathcal{H}_K(a) := \left\{ h \in \mathcal{H}_K : \inf_{x \in [0,1]} h'(x) \geq -a \right\} \]

restricted measure of ill-posedness:

\[ \tau_{n,t}(a) := \sup_{h \in \mathcal{H}_{K_n}(a)} \frac{\|h\|_{2,t}}{\|Th\|_2} \]

\|h\|_{2,t} \quad \quad \|h\|_{2,t=1} \]
Assumption 1 (Monotone IV)

(i) **Stochastic monotonicity**: For all $x, w', w'' \in (0,1)$,

$$w' \leq w'' \Rightarrow F_{X|W}(x|w') \geq F_{X|W}(x|w'')$$
Assumption 1 (Monotone IV)

(i) **Stochastic monotonicity:** For all $x, w', w'' \in (0, 1)$,

\[ w' \leq w'' \implies F_{X|W}(x|w') \geq F_{X|W}(x|w'') \]

(ii) **Instrument relevance:** inequality is strict for some pair $w', w''$
Monotone IV Assumption

Assumption 1 (Monotone IV)

(i) **Stochastic monotonicity:** For all $x, w', w'' \in (0, 1)$,

$$w' \leq w'' \implies F_{X|W}(x|w') \geq F_{X|W}(x|w'')$$

(ii) **Instrument relevance:** inequality is strict for some pair $w', w''$

- first stage monotone in $W$
- allows for multi-dimensional first-stage error

Examples

Chetverikov and Wilhelm (UCLA and UCL) Monotone Nonparametric IV
Regularity Conditions

Assumption 2 (Density)

1. \((X, W)\) has density \(f_{X,W}(x, w)\) with respect to the Lebesgue measure on \([0, 1]^2\).
2. \(\int_0^1 \int_0^1 f_{X,W}(x, w)^2 \, dx \, dw \leq C_T\) for some finite constant \(C_T\).
3. There exists \(c_f > 0\) such that \(f_{X|W}(x|w) \geq c_f\) for all \(x \in [x_1, x_2]\) and \(w \in \{w_1, w_2\}\).
4. There exist \(0 < c_W \leq C_W < \infty\) such that \(c_W \leq f_W(w) \leq C_W\) for all \(w \in [0, 1]\).
Theorem 1

Under Assumptions 1 and 2, there exist constants $0 < c, C < \infty$ independent of $n$ such that

$$\tau_{n,t}(a) \leq C \quad \text{for all } n \text{ and all } a \leq c$$
Bounded Measure of Ill-Posedness

**Theorem 1**

*Under Assumptions 1 and 2, there exist constants $0 < c, C < \infty$ independent of $n$ such that*

$$\tau_{n,t}(a) \leq C \quad \text{for all } n \text{ and all } a \leq c$$

- $\tau_{n,t}(a)$ does **not** grow with dimension $K_n$ of sieve space
Theorem 1

Under Assumptions 1 and 2, there exist constants $0 < c, C < \infty$ independent of $n$ such that

$$\tau_{n,t}(a) \leq C \quad \text{for all } n \text{ and all } a \leq c$$

- $\tau_{n,t}(a)$ does not grow with dimension $K_n$ of sieve space
- bound independent of whether unrestricted problem is mildly or severely ill-posed
Theorem 1

Under Assumptions 1 and 2, there exist constants $0 < c, C < \infty$ independent of $n$ such that

$$\tau_{n,t}(a) \leq C \quad \text{for all } n \text{ and all } a \leq c$$

- $\tau_{n,t}(a)$ does not grow with dimension $K_n$ of sieve space
- bound independent of whether unrestricted problem is mildly or severely ill-posed
- in contrast:
  $$\tau_{n,t}(\infty) = O(K_n^r) \quad \text{or} \quad \tau_{n,t}(\infty) = O(e^{CK_n})$$
  in mildly or severely ill-posed case, respectively
Overview

1 Motivation and Background

2 Bounds on Restricted Measure of Ill-Posedness

3 Non-Asymptotic Risk Bounds

4 Simulations

5 Gasoline Demand in the U.S.
Restricted and Unrestricted Estimators

- iid sample \((Y_i, X_i, W_i), i = 1, \ldots, n\), from the distribution of \((Y, X, W)\)
Restricted and Unrestricted Estimators

- iid sample \((Y_i, X_i, W_i), \ i = 1, \ldots, n\), from the distribution of \((Y, X, W)\)
- \(p(x)\) vector of \(K_n\) basis functions
Restricted and Unrestricted Estimators

- iid sample \((Y_i, X_i, W_i), i = 1, \ldots, n\), from the distribution of \((Y, X, W)\)
- \(p(x)\) vector of \(K_n\) basis functions
- \(q(w)\) vector of \(J_n \geq K_n\) basis functions
- iid sample \( (Y_i, X_i, W_i), \ i = 1, \ldots, n \), from the distribution of \( (Y, X, W) \)
- \( p(x) \) vector of \( K_n \) basis functions
- \( q(w) \) vector of \( J_n \geq K_n \) basis functions
- \( \xi_n := \max(\sup_{x \in [0,1]} \|p(x)\|, \sup_{w \in [0,1]} \|q(w)\|) \)
Restricted and Unrestricted Estimators

- iid sample \((Y_i, X_i, W_i), i = 1, \ldots, n\), from the distribution of \((Y, X, W)\)
- \(p(x)\) vector of \(K_n\) basis functions
- \(q(w)\) vector of \(J_n \geq K_n\) basis functions
- \(\xi_n := \max(\sup_{x \in [0,1]} \|p(x)\|, \sup_{w \in [0,1]} \|q(w)\|)\)
- restricted estimator (imposing monotonicity of \(g\)):
  \[
  \hat{g}^r(x) := p(x)' \hat{\beta}^r
  \]
Assumption 3

\[ g \text{ is monotone.} \]
Theorem 2

Let Assumptions 1–3 and some more regularity conditions hold. Then with probability at least $1 - \alpha - n^{-1}$:

$$\|\hat{g}^r - g\|_2, t \leq C \min \left\{ \|g\|_{\infty} + \left( \frac{K_n}{\alpha n} + \frac{\xi_n^2 \log n}{n} \right)^{1/2}, \tau_n \cdot \left( \frac{K_n}{\alpha n} + \frac{\xi_n^2 \log n}{n} \right)^{1/2} \right\} + CK_n^{-s}$$

where $C$ depends only on constants appearing in the assumptions and $\tau_n := \tau_{n,t}(\infty)$. 
Theorem 2

Let Assumptions 1–3 and some more regularity conditions hold. Then with probability at least $1 - \alpha - n^{-1}$:

$$\|\hat{g}^r - g\|_{2,t} \leq C \min \left\{ \|g'\|_{\infty} + \left( \frac{K_n}{\alpha n} + \frac{\xi_n^2 \log n}{n} \right)^{1/2}, \tau_n \cdot \left( \frac{K_n}{\alpha n} + \frac{\xi_n^2 \log n}{n} \right)^{1/2} \right\} + CK_n^{-s}$$

where $C$ depends only on constants appearing in the assumptions and $\tau_n := \tau_{n,t}(\infty)$.

monotonicity constraint matters if

$$\|g'\|_{\infty} \leq (\tau_n - 1) \left( \frac{K_n}{\alpha n} + \frac{\xi_n^2 \log n}{n} \right)^{1/2}$$
Implications of the Bound

1. No rate improvements in interior of monotonicity constraint:

\[ \| \hat{g}^r - g \|_{2,t} = O_P \left( \tau_{n,t}(\infty)(K_n/n)^{1/2} + K_n^{-s} \right) \]
Implications of the Bound

1. No rate improvements in **interior** of monotonicity constraint:

\[
\| \hat{g}^r - g \|_{2,t} = O_P \left( \tau_{n,t}(\infty)(K_n/n)^{1/2} + K_n^{-s} \right)
\]

2. Fast rate in \( n^{-s/(1+2s)} \sqrt{\log n} \)-neighborhood of the **boundary**:

\[
\| \hat{g}^r - g \|_{2,t} = O_P \left( n^{-s/(1+2s)} \sqrt{\log n} \right).
\]

- Independent of mild or severe ill-posedness
Implications of the Bound

1. No rate improvements in **interior** of monotonicity constraint:

\[ \| \hat{g}^r - g \|_{2,t} = O_P \left( \tau_n, t(\infty)(K_n/n)^{1/2} + K_n^{-s} \right) \]

2. Fast rate in \( n^{-s/(1+2s)} \sqrt{\log n} \)-neighborhood of the **boundary**:

\[ \| \hat{g}^r - g \|_{2,t} = O_P \left( n^{-s/(1+2s)} \sqrt{\log n} \right). \]

- independent of mild or severe ill-posedness

3. Finite samples:

\[ \| \hat{g}^r - g \|_{2,t} \leq C \min \left\{ \| g' \|_{\infty} + \left( \frac{K_n}{\alpha n} + \frac{\xi_n^2 \log n}{n} \right)^{1/2}, \tau_n \cdot \left( \frac{K_n}{\alpha n} + \frac{\xi_n^2 \log n}{n} \right)^{1/2} \right\} + CK_n^{-s} \]

- distance to boundary matters
Overview

1 Motivation and Background

2 Bounds on Restricted Measure of Ill-Posedness

3 Non-Asymptotic Risk Bounds

4 Simulations

5 Gasoline Demand in the U.S.
Simulation Setup

- NPIV model:

\[ Y = g(X) + \varepsilon, \quad E[\varepsilon|W] = 0 \]

where

\[ g(x) = 10\kappa \left[ -(x - 0.25)^2 1\{x \in [0, 0.25]\} + (x - 0.75)^2 1\{x \in [0.75, 1]\} \right] \]
Simulation Setup

- NPIV model:
  \[ Y = g(X) + \varepsilon, \quad E[\varepsilon|W] = 0 \]
  
  where
  \[ g(x) = 10\kappa \left[ - (x - 0.25)^2 \mathbf{1}\{x \in [0, 0.25]\} + (x - 0.75)^2 \mathbf{1}\{x \in [0.75, 1]\} \right] \]

- \( X = \Phi(\xi) \) and \( W = \Phi(\zeta) \), where \( \Phi \) is the \( N(0, 1) \) cdf

- \( \xi = \rho \zeta + \sqrt{1 - \rho^2} \varepsilon \)

- \( \varepsilon = \kappa \sigma_\varepsilon (\eta \varepsilon + \sqrt{1 - \eta^2} \nu) \)

- \((\nu, \zeta, \varepsilon) \sim N(0, I)\)

- \( \kappa \) in \( \{1, 0.5, 0.1\} \) governs flatness of \( g \)

- 1,000 MC samples

- normalized B-spline basis for \( p(x) \) and \( q(w) \) of degree 3 and 4, varying number of knots \( k_X, k_W \)
Figure: $N = 500$, $\rho = 0.3$, $\eta = 0.3$, $\sigma_\varepsilon = 0.1$, $k_X = 3$, $k_W = 4$. 
### Performance Summary

<table>
<thead>
<tr>
<th>$k_X$</th>
<th>$k_W$</th>
<th>$\kappa = 1$</th>
<th>$\kappa = 0.5$</th>
<th>$\kappa = 0.1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>unrestr.</td>
<td>restr.</td>
<td>unrestr.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\kappa = 1$</td>
<td>$\kappa = 0.5$</td>
<td>$\kappa = 0.1$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>unrestr.</td>
<td>restr.</td>
<td>unrestr.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\kappa = 1$</td>
<td>$\kappa = 0.5$</td>
<td>$\kappa = 0.1$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>unrestr.</td>
<td>restr.</td>
<td>unrestr.</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>0.002 0.005</td>
<td>0.001 0.001</td>
<td>0.000 0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.102 0.032</td>
<td>0.321 0.008</td>
<td>0.012 0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.104 0.038</td>
<td>0.321 0.009</td>
<td>0.012 0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.034</td>
<td>0.029</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>0.001 0.006</td>
<td>0.000 0.002</td>
<td>0.000 0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.462 0.031</td>
<td>0.103 0.008</td>
<td>0.004 0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.463 0.037</td>
<td>0.104 0.009</td>
<td>0.004 0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.080</td>
<td>0.088</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>0.001 0.004</td>
<td>0.000 0.001</td>
<td>0.000 0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.936 0.036</td>
<td>0.255 0.009</td>
<td>0.012 0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.936 0.040</td>
<td>0.255 0.010</td>
<td>0.012 0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.043</td>
<td>0.039</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>0.001 0.005</td>
<td>0.000 0.001</td>
<td>0.000 0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.387 0.035</td>
<td>0.110 0.009</td>
<td>0.004 0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.388 0.040</td>
<td>0.110 0.010</td>
<td>0.004 0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.103</td>
<td>0.089</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>0.002 0.005</td>
<td>0.000 0.001</td>
<td>0.000 0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.508 0.041</td>
<td>0.144 0.010</td>
<td>0.007 0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.510 0.046</td>
<td>0.144 0.011</td>
<td>0.007 0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.090</td>
<td>0.078</td>
<td></td>
</tr>
</tbody>
</table>

Table: \( N = 500, \sigma_\varepsilon = 0.7, \rho = 0.3, \eta = 0.3. \)
<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\eta$</th>
<th>$\kappa = 1$</th>
<th></th>
<th>$\kappa = 0.5$</th>
<th></th>
<th>$\kappa = 0.1$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>unst.</td>
<td>restr.</td>
<td>unst.</td>
<td>restr.</td>
<td>unst.</td>
<td>restr.</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>bias sq.</td>
<td>var</td>
<td>0.022</td>
<td>0.003</td>
<td>0.006</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>MSE</td>
<td>MSE ratio</td>
<td>0.192</td>
<td>0.176</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>0.3</td>
<td>0.7</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>bias sq.</td>
<td>var</td>
<td>0.020</td>
<td>0.003</td>
<td>0.006</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>MSE</td>
<td>MSE ratio</td>
<td>0.209</td>
<td>0.163</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>0.7</td>
<td>0.3</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>bias sq.</td>
<td>var</td>
<td>0.013</td>
<td>0.000</td>
<td>0.002</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>MSE</td>
<td>MSE ratio</td>
<td>0.040</td>
<td>0.063</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>0.7</td>
<td>0.7</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>bias sq.</td>
<td>var</td>
<td>0.010</td>
<td>0.000</td>
<td>0.002</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>MSE</td>
<td>MSE ratio</td>
<td>0.051</td>
<td>0.060</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**Table:** $N = 500$, $\sigma_\varepsilon = 0.1$, $k_X = 3$, $k_W = 4$. 
Overview

1 Motivation and Background

2 Bounds on Restricted Measure of Ill-Posedness

3 Non-Asymptotic Risk Bounds

4 Simulations

5 Gasoline Demand in the U.S.
Gasoline Demand Curves

\[ Q = g(P, Y) + \gamma'Z + \varepsilon, \quad E[\varepsilon|Y, W, Z] = 0 \]

where

- \( Q \): annual log gasoline consumption
- \( P \): log price of gasoline (endogenous), avg. price per gallon
- \( Y \): log household income (exogenous)
- \( Z \): covariates (exogenous)
- \( W \): distance to major oil platform (instrument)

analysis same as in Blundell, Horowitz, and Parey (2012), but allow for endogeneity of price
Data and Estimators

Data:

- 2001 National Household Travel Survey
- sample size 4,812

Estimators:

- nonparametric kernel regression taking price as exogenous
  - bandwidths as in Blundell, Horowitz, and Parey (2012)
  - estimate $\gamma$ as in Robinson (1988), then remove the covariates from $Y$
- restricted and unrestricted NPIV estimators:
  - normalized B-spline basis for $p(x)$ and $q(w)$, degree: both 3, knots: 3 and 5
  - impose linearity in $Z$
- estimates at three income levels: $42,500$, $57,500$, and $72,500$
Estimates of the Demand Curve

Chetverikov and Wilhelm (UCLA and UCL) Monotone Nonparametric IV
Conclusions

Impose two monotonicity assumptions in NPIV model:

1. monotone regression function
2. monotone IV

Consequences:

1. ill-posedness is weakened
2. obtain non-asymptotic risk bounds
3. simulations show dramatic performance improvements of estimator imposing monotonicity
More in the paper:

1. monotonicity implies non-trivial identification bounds
2. new adaptive test of the monotone IV assumption
Examples of Monotone IV

Example 1 (Joint Normal Distribution)

\[ X = \Phi(\tilde{X}) \text{ and } W = \Phi(\tilde{W}) \text{ where} \]

- \((\tilde{X}, \tilde{W}) \sim N\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)\) with \(\rho > 0\)
- \(\Phi\) is the cdf of \(N(0, 1)\)
Examples of Monotone IV

Example 1 (Joint Normal Distribution)

\[ X = \Phi(\tilde{X}) \text{ and } W = \Phi(\tilde{W}) \text{ where} \]

- \((\tilde{X}, \tilde{W}) \sim N\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)\) with \(\rho > 0\)
- \(\Phi\) is the cdf of \(N(0, 1)\)

Example 2 (Random Coefficients)

\[ X = U_1 + U_2 W \]

where

- \(U_1, U_2, W\) mutually independent
- \(U_1, U_2 \sim U[0, \frac{1}{2}]\) and \(W \sim U[0, 1]\)
Monotonicity Implied by Theory

- firm produces log output $Y$ from labor input $X$
- $W$ price of log output
- $U$ log wage
- production function:
  $$Y = g(X) + \varepsilon$$
  where $\varepsilon$ denotes capital, total factor productivity etc.
- profits: $\pi(X) = e^W e^Y - e^U e^X$
- $g$ increasing, strictly concave
- elasticity of output with respect to labor is strictly less than one
Under the above conditions:

- $\frac{\partial^2 \pi}{\partial X \partial W} \geq 0$
- $\frac{\partial^2 \pi}{\partial X^2} < 0$
Under the above conditions:

- $\frac{\partial^2 \pi}{\partial X \partial W} \geq 0$
- $\frac{\partial^2 \pi}{\partial X^2} < 0$

$W$ is a monotone IV:

$$\Rightarrow \frac{\partial X}{\partial W} = -\frac{\frac{\partial^2 \pi}{\partial X \partial W}}{\frac{\partial^2 \pi}{\partial X^2}} \geq 0$$
A Restricted Measure of Ill-posedness

- NPIV model: \( E[Y|W] = E[g(X)|W] \iff m = Tg \)
- truncated norm: \( \|h\|_{2,t} := \left( \int_{x_1}^{x_2} |h(x)|^2 \, dx \right)^{1/2} \) where \( 0 < x_1 < x_2 < 1 \)
- \( p_1, p_2, \ldots \) basis of \( L^2[0, 1] \)
- sieve space of functions with derivative bounded from below:

\[
\mathcal{H}_K(a) := \left\{ h \in L^2[0, 1] : \exists b_1, \ldots, b_K \in \mathbb{R} \text{ with } h = \sum_{j=1}^{K} b_j p_j \right. \\
\left. \text{ and } \inf_{x \in [0,1]} h'(x) \geq -a \right\}
\]

restricted measure of ill-posedness:

\[
\tau_{n,t}(a) := \sup_{h \in \mathcal{H}_{K_n}(a)} \frac{\|h\|_{2,t}}{\|Th\|_2} \\
\|h\|_{2,t} = 1
\]
Relationship Between Measures of Ill-posedness

Our measure of ill-posedness:

\[ \tau_{n,t}(a) = \sup_{h \in \mathcal{H}_{K_n}(a)} \frac{\|h\|_{2,t}}{\|T_h\|_2} \]  

Measure of ill-posedness by Blundell, Chen, and Kristensen (2007), Horowitz and Lee (2012) etc:

\[ \tau_n = \sup_{h \in \mathcal{H}_{K_n}(\infty)} \frac{\|h\|_2}{\|T_h\|_2} \]
Relationship Between Measures of Ill-posedness

Our measure of ill-posedness:

\[ \tau_{n,t}(a) = \sup_{h \in H_{K_{n}(a)}} \frac{\|h\|_{2,t}}{\|Th\|_{2}} \]

Measure of ill-posedness by Blundell, Chen, and Kristensen (2007), Horowitz and Lee (2012) etc:

\[ \tau_{n} = \sup_{h \in H_{K_{n}(\infty)}} \frac{\|h\|_{2}}{\|Th\|_{2}} \]

Relationship:

- \( \tau_{n,t}(a) \leq \tau_{n,t}(\infty) \leq \tau_{n} \)
- **severe ill-posedness**: \( \tau_{n} = O(e^{cK_{n}}) \) \( \Rightarrow \) \( \tau_{n,t}(\infty) = O(e^{cK_{n}}) \)
- **mild ill-posedness**: paper gives condition under which \( \tau_{n} = O(K_{n}^{r}) \) \( \Rightarrow \) \( \tau_{n,t}(\infty) = O(K_{n}^{r}) \)
Restricted and Unrestricted Estimators

- iid sample $(Y_i, X_i, W_i), i = 1, \ldots, n$, from the distribution of $(Y, X, W)$

unrestricted estimator: $\hat{g}^u(x) := p(x)'\hat{\beta}^u$ with

$$\hat{\beta}^u := \arg\min_{b \in \mathbb{R}^K} (Y - Pb)'Q(Q'Q)^{-1}Q'(Y - Pb)$$

restricted estimator: $\hat{g}^r(x) := p(x)'\hat{\beta}^r$ with

$$\hat{\beta}^r := \arg\min_{b \in \mathbb{R}^K: p(\cdot)'b \in \mathcal{H}_0} (Y - Pb)'Q(Q'Q)^{-1}Q'(Y - Pb)$$
Restricted and Unrestricted Estimators

- iid sample \((Y_i, X_i, W_i), \ i = 1, \ldots, n\), from the distribution of \((Y, X, W)\)
- \(p_1(x), p_2(x), \ldots\) and \(q_1(w), q_2(w), \ldots\) two orthonormal bases in \(L^2[0, 1]\)

**unrestricted estimator:** \(\hat{g}^u(x) := p(x)' \hat{\beta}^u\) with
\[
\hat{\beta}^u := \arg\min_{b \in \mathbb{R}^K} (Y - Pb)' Q(Q'Q)^{-1} Q'(Y - Pb)
\]

**restricted estimator:** \(\hat{g}^r(x) := p(x)' \hat{\beta}^r\) with
\[
\hat{\beta}^r := \arg\min_{b \in \mathbb{R}^K: p(\cdot)'b \in \mathcal{H}_K(0)} (Y - Pb)' Q(Q'Q)^{-1} Q'(Y - Pb)
\]
Restricted and Unrestricted Estimators

- iid sample \((Y_i, X_i, W_i), \ i = 1, \ldots, n\), from the distribution of \((Y, X, W)\)
- \(p_1(x), p_2(x), \ldots\) and \(q_1(w), q_2(w), \ldots\) two orthonormal bases in \(L^2[0, 1]\)
- \(p(x) := (p_1(x), \ldots, p_K(x))'\) and \(q(w) := (q_1(w), \ldots, q_J(w))'\)

unrestricted estimator: \(\hat{g}^u(x) := p(x)' \hat{\beta}^u\) with

\[
\hat{\beta}^u := \text{argmin}_{b \in \mathbb{R}^{K}} (Y - P b)' Q (Q' Q)^{-1} Q' (Y - P b)
\]

restricted estimator: \(\hat{g}^r(x) := p(x)' \hat{\beta}^r\) with

\[
\hat{\beta}^r := \text{argmin}_{b \in \mathbb{R}^{K}: p(\cdot)' b \in \mathcal{H}_K(0)} (Y - P b)' Q (Q' Q)^{-1} Q' (Y - P b)
\]
Restricted and Unrestricted Estimators

- iid sample \((Y_i, X_i, W_i), i = 1, \ldots, n\), from the distribution of \((Y, X, W)\)
- \(p_1(x), p_2(x), \ldots\) and \(q_1(w), q_2(w), \ldots\) two orthonormal bases in \(L^2[0, 1]\)
- \(p(x) := (p_1(x), \ldots, p_K(x))'\) and \(q(w) := (q_1(w), \ldots, q_J(w))'\)
- \(P := (p(X_1), \ldots, p(X_n))', Q := (p(W_1), \ldots, p(W_n))', Y = (Y_1, \ldots, Y_n)'\)

**unrestricted estimator:** \(\hat{g}^u(x) := p(x)' \hat{\beta}^u\) with

\[
\hat{\beta}^u := \arg\min_{b \in \mathbb{R}^K} (Y - P b)' Q (Q' Q)^{-1} Q' (Y - P b)
\]

**restricted estimator:** \(\hat{g}^r(x) := p(x)' \hat{\beta}^r\) with

\[
\hat{\beta}^r := \arg\min_{b \in \mathbb{R}^K : p(\cdot)' b \in \mathcal{H}_K(0)} (Y - P b)' Q (Q' Q)^{-1} Q' (Y - P b)
\]
Assumption 4 (Monotone IV)

(i) **Stochastic monotonicity:** For all $x, w', w'' \in (0, 1)$,

\[ w' \leq w'' \implies F_{X|W}(x|w') \geq F_{X|W}(x|w'') \]

(ii) **Instrument relevance:** For some constants $C_F > 1$, $0 < x_1 < x_2 < 1$, and $0 \leq w_1 < w_2 \leq 1$,

\[ F_{X|W}(x|w_1) \geq C_F F_{X|W}(x|w_2) \quad \forall x \in (0, x_2), \]

and

\[ C_F(1 - F_{X|W}(x|w_1)) \leq 1 - F_{X|W}(x|w_2) \quad \forall x \in (x_1, 1). \]
Notation

- \( m(w) := E[Y|W = w]f_W(w) \)
- \( Th(w) := \int h(x)f_{X,W}(x, w)dx \)
- NPIV model: \( E[Y|W] = E[g(X)|W] \Leftrightarrow m = Tg \)
- Define \( T_n : L^2[0, 1] \rightarrow L^2[0, 1] \) by
  \[
  (T_n h)(w) := q(w)' E[q(W)p(X)']E[p(U)h(U)]
  \]
  for all \( w \in [0, 1] \) where \( U \sim U[0, 1] \).
- \( f_W(w) \) is the marginal density of \( W \)
- \( \xi_{K,p} := \sup_{x \in [0,1]} \|p(x)\|, \xi_{J,q} := \sup_{w \in [0,1]} \|q(w)\|, \) and \( \xi_n := \max(\xi_{K,p}, \xi_{J,q}) \).
### Assumption 5 (Moments)

*For some constant $C_B < \infty$*

1. $E[\varepsilon^2 | W] \leq C_B$
2. $E[g(X)^2 | W] \leq C_b$

### Assumption 6 (Approximation of $g$)

*There exists $\beta_n \in \mathbb{R}^K$ and a constant $C_g < \infty$ such that the function $g_n(x) = p(x)' \beta_n$ satisfies*

1. $g_n \in \mathcal{H}_n(0)$
2. $\|g - g_n\|_2 \leq C_g K^{-s}$
3. $\|T(g - g_n)\|_2 \leq C_g \tau_n^{-1} K^{-s}$
Assumption 7 (Approximation of $m$)

There exist $\gamma_n \in \mathbb{R}^d$ and a constant $C_m < \infty$ such that the function $m_n(w) := q(w)' \gamma_n$, defined for all $w \in [0, 1]$, satisfies $\|m - m_n\|_2 \leq C_m \tau_n^{-1} K^{-s}$.

Assumption 8 (Operator $T$)

1. $T$ is injective
2. for some constant $C_a < \infty$, $\|(T - T_n)h\|_2 \leq C_a \tau_n^{-1} K^{-s} \|h\|_2$ for all $h \in \mathcal{H}_n(\infty)$
Overview

6 Identification Power of the Monotonicity Assumptions
Lemma 3

If Assumptions 1–3 hold and $g$ is continuously differentiable, then $\text{sign}(g'(x))$ is identified.
Lemma 3

If Assumptions 1–3 hold and $g$ is continuously differentiable, then $\text{sign}(g'(x))$ is identified.

- By monotone IV assumption:
  
  $g$ increasing \implies E[g(X)|W = \cdot] = E[Y|W = \cdot]$ increasing

- estimate $\text{sign}(g'(x))$ simply from regression of $Y$ on $W$
Lemma 4 (Identification bounds)

Suppose Assumptions 1 and 2 hold, \( g', g'' \in L^2[0, 1] \), and \( \bar{C} \) is a constant depending only on observable quantities. If there exists a function \( h \in L^2[0, 1] \) such that

1. \( g' - g'' + h \) is monotone
2. \( \| h \|_{2,t} + \bar{C} \| T \|_2 \| h \|_2 < \| g' - g'' \|_{2,t} \)

then \( g' \) and \( g'' \) are not observationally equivalent.

- identified set \( \Theta := \{ g \in L^2[0, 1] : E[Y|W] = E[g(X)|W], \ g \ monotone \} \)
- \( g' \) and \( g'' \) observationally equivalent if \( E[g'(X) - g''(X)|W] = 0 \)
Identified Set for the Regression Function

- all functions in $\Theta$ have to intersect
Identified Set for the Regression Function

- all functions in $\Theta$ have to intersect
- $\Theta$ does not contain any functions whose difference is monotone
Identified Set for the Regression Function

- all functions in $\Theta$ have to intersect
- $\Theta$ does not contain any functions whose difference is monotone
- $\Theta$ does not contain any functions whose difference is close to monotone
Identified Set for the Regression Function

- All functions in \( \Theta \) have to intersect
- \( \Theta \) does not contain any functions whose difference is monotone
- \( \Theta \) does not contain any functions whose difference is close to monotone
- If \( g \in \Theta \), then all other functions in \( \Theta \) cannot be much steeper than \( g \)
Identified Set for the Regression Function

- all functions in $\Theta$ have to intersect
- $\Theta$ does not contain any functions whose difference is monotone
- $\Theta$ does not contain any functions whose difference is close to monotone
- if $g \in \Theta$, then all other functions in $\Theta$ cannot be much steeper than $g$
- $\Theta$ contains either only increasing or only decreasing functions, not both