"Honest confidence sets in nonparametric IV regression and other ill-posed models"

Andrii Babii

UNC Chapel Hill

December 7, 2018

Preview

Motivation

How to build confidence sets in the NPIV and other ill-posed models?

Preview

Motivation

How to build confidence sets in the NPIV and other ill-posed models?

Approach

1. functional CLT does not work;
2. alternative ideas: data-driven concentration inequality or Gaussian/bootstrap approximation.

Preview

Motivation

How to build confidence sets in the NPIV and other ill-posed models?

Approach

1. functional CLT does not work;
2. alternative ideas: data-driven concentration inequality or Gaussian/bootstrap approximation.

Results

Confidence sets have appealing properties:

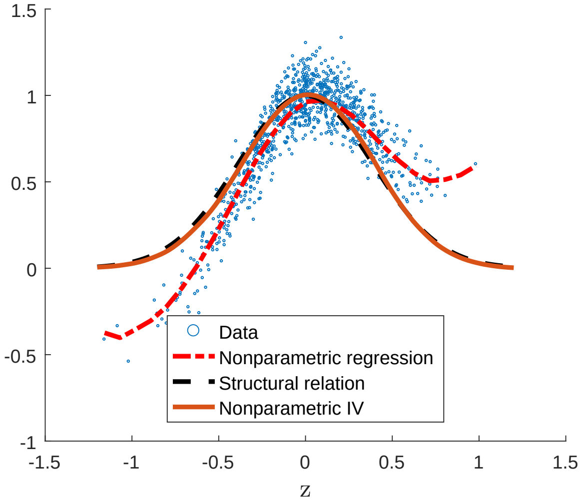
- uniform over the domain of the function;
- honest: uniform over the large DGPs.

Nonparametric IV regression

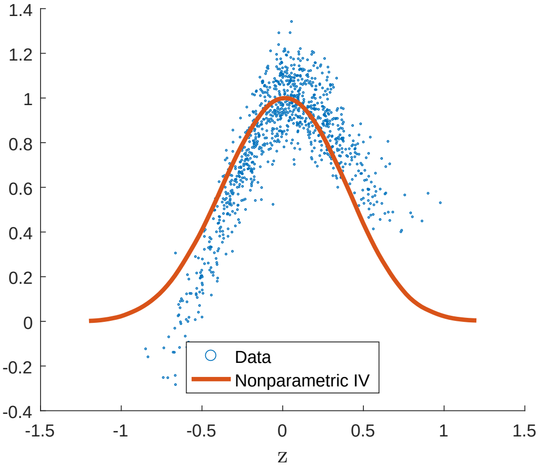
$$Y = \varphi(Z) + U, \qquad \mathbb{E}[U|W] = 0$$

- robust to model mispecification (e.g. nonlinearities) comparing to the linear IV regression;
- robust to endogeneity comparing to traditional non-parametric approaches.

Endogeneity bias

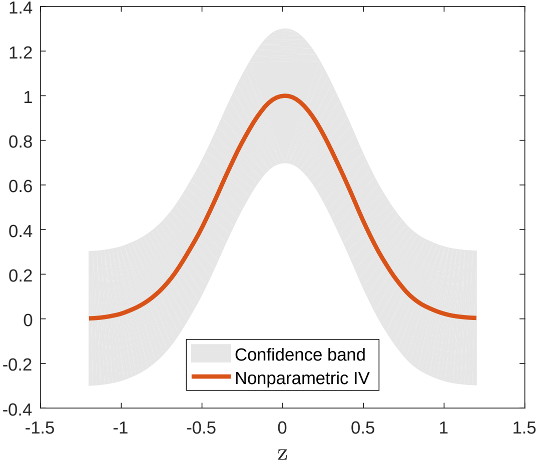


What can we learn about structural function φ?



Significant effects, nonlinearities, monotonicity, concavity/convexity?

Uniform confidence band



There is evidence for nonlinearities, not monotone, not convex.

Inference in nonparametric IV model

1. Why uniform inference?
   - global shape properties: nonlinearities, endogeneity bias, monotonicity, concavity/convexity;
   - the range of possible economic effects at all points simultaneously;
   - natural, since the estimator is a function;
   - have appealing geometry and are easy to implement numerically.

Inference in nonparametric IV model

1. Why uniform inference?
   - global shape properties: nonlinearities, endogeneity bias, monotonicity, concavity/convexity;
   - the range of possible economic effects at all points simultaneously;
   - natural, since the estimator is a function;
   - have appealing geometry and are easy to implement numerically.
2. Why honest?
   - Adopted in modern non-parametric statistics to rule-out some contradictory results (e.g., the Epanechnikov kernel is not optimal);
   - Safe against the nature (if the nature gives us the problem withing a given class).

Inference in nonparametric IV model

1. Why uniform inference?
   - global shape properties: nonlinearities, endogeneity bias, monotonicity, concavity/convexity;
   - the range of possible economic effects at all points simultaneously;
   - natural, since the estimator is a function;
   - have appealing geometry and are easy to implement numerically.
2. Why honest?
   - Adopted in modern non-parametric statistics to rule-out some contradictory results (e.g., the Epanechnikov kernel is not optimal);
   - Safe against the nature (if the nature gives us the problem withing a given class).
3. Existing uniform inferential methods are only available for sieve estimator introduced in (Newey and Powell, 2003): (Chen and Christensen, 2018).

Contributions of this paper

1. develop uniform inferential methods for the NPIV estimator based on kernels and Tikhonov regularization (Florens, 2003), (Darolles, Fan, Florens, and Renault, 2011);

Contributions of this paper

1. develop uniform inferential methods for the NPIV estimator based on kernels and Tikhonov regularization (Florens, 2003), (Darolles, Fan, Florens, and Renault, 2011);
2. uniform inferential methods for functional regression models;

Contributions of this paper

1. develop uniform inferential methods for the NPIV estimator based on kernels and Tikhonov regularization (Florens, 2003), (Darolles, Fan, Florens, and Renault, 2011);
2. uniform inferential methods for functional regression models;
3. uniform inferential methods for measurement errors (density deconvolution) models based on Tikhonov regularization (Carraco and Florens, 2011).

Why kernels and Tikhonov?

1. The estimator changes smoothly with respect to tuning parameters: stable finite-sample performance;
2. Do not need to select bases: different bases have different approximating properties depending on the unknown function φ;
3. More non-parametric in finite samples: what if we miss important coefficient with series approach?
4. Do not need to invert random matrices, which may be singular and need to be additionally regularized;
5. Robust to various deviations from point identification: best approximation to the function φ (Babii and Florens, 2018).

Roadmap

Roadmap

1. Honest and uniform confidence sets for NPIV: construction
   - Concentration inequalities
   - Bootstrap approximations
2. Honest and uniform confidence sets for NPIV: results
3. Monte Carlo experiments
4. Engel curves in the US

NPIV (Florens, 2003)

1. Nonparametric IV regression model

$$Y = \varphi(Z) + U, \qquad \mathbb{E}[U|W] = 0.$$

NPIV (Florens, 2003)

1. Nonparametric IV regression model

$$Y = \varphi(Z) + U, \qquad \mathbb{E}[U|W] = 0.$$

2. Conditional moment restriction leads to ill-posed equation

$$\mathbb{E}[Y|W = w] = \mathbb{E}[\varphi(Z)|W = w].$$

NPIV (Florens, 2003)

1. Nonparametric IV regression model

$$Y = \varphi(Z) + U, \qquad \mathbb{E}[U|W] = 0.$$

2. Conditional moment restriction leads to ill-posed equation

$$\mathbb{E}[Y|W = w] = \mathbb{E}[\varphi(Z)|W = w].$$

3. Equivalently φ solves integral equation

$$r(w) := \mathbb{E}[Y|W = w]f_W(w) = \int_{[a,b]^p} \varphi(z)f_{ZW}(z, w)\mathrm{d}z =: (T\varphi)(w),$$

shortly denoted as

$$r = T\varphi.$$

NPIV (Florens, 2003)

1. Nonparametric IV regression model

$$Y = \varphi(Z) + U, \qquad \mathbb{E}[U|W] = 0.$$

2. Conditional moment restriction leads to ill-posed equation

$$\mathbb{E}[Y|W = w] = \mathbb{E}[\varphi(Z)|W = w].$$

3. Equivalently φ solves integral equation

$$r(w) := \mathbb{E}[Y|W = w]f_W(w) = \int_{[a,b]^p} \varphi(z)f_{ZW}(z, w)\mathrm{d}z =: (T\varphi)(w),$$

shortly denoted as

$$r = T\varphi.$$

4. Ill-posed problem, since the generalized inverse map $(T^*T)^{-1}$ is not continuous.

NPIV (Florens, 2003)

1. Nonparametric IV regression model

$$Y = \varphi(Z) + U, \qquad \mathbb{E}[U|W] = 0.$$

2. Conditional moment restriction leads to ill-posed equation

$$\mathbb{E}[Y|W = w] = \mathbb{E}[\varphi(Z)|W = w].$$

3. Equivalently φ solves integral equation

$$r(w) := \mathbb{E}[Y|W = w]f_W(w) = \int_{[a,b]^p} \varphi(z)f_{ZW}(z,w)\mathrm{d}z =: (T\varphi)(w),$$

shortly denoted as

$$r = T\varphi.$$

4. Ill-posed problem, since the generalized inverse map $(T^*T)^{-1}$ is not continuous.
5. Tikhonov regularization: smooth out discontinuity of inversion.

NPIV: estimation

1. Joint density f_{ZW}

$$\hat{f}_{ZW}(z,w) = \frac{1}{nh_n^{p+q}} \sum_{i=1}^{n} K_z\left(h_n^{-1}(Z_i - z)\right) K_w(h_n^{-1}(W_i - w)).$$

NPIV: estimation

1. Joint density f_{ZW}

$$\hat{f}_{ZW}(z, w) = \frac{1}{nh_n^{p+q}} \sum_{i=1}^{n} K_z\left(h_n^{-1}(Z_i - z)\right) K_w(h_n^{-1}(W_i - w)).$$

2. Left side r

$$\hat{r}(w) = \frac{1}{nh_n^{q}} \sum_{i=1}^{n} Y_i K_w\left(h_n^{-1}(W_i - w)\right).$$

NPIV: estimation

1. Joint density f_{ZW}

$$\hat{f}_{ZW}(z,w) = \frac{1}{nh_n^{p+q}} \sum_{i=1}^{n} K_z\left(h_n^{-1}(Z_i - z)\right) K_w(h_n^{-1}(W_i - w)).$$

2. Left side r

$$\hat{r}(w) = \frac{1}{nh_n^{q}} \sum_{i=1}^{n} Y_i K_w\left(h_n^{-1}(W_i - w)\right).$$

3. Operator T

$$(\hat{T}\phi)(w) = \int_{[a,b]^p} \phi(z)\hat{f}_{ZW}(z,w)\mathrm{d}z.$$

NPIV: estimation

1. Joint density f_{ZW}

$$\hat{f}_{ZW}(z,w) = \frac{1}{nh_n^{p+q}} \sum_{i=1}^{n} K_z\left(h_n^{-1}(Z_i - z)\right) K_w(h_n^{-1}(W_i - w)).$$

2. Left side r

$$\hat{r}(w) = \frac{1}{nh_n^{q}} \sum_{i=1}^{n} Y_i K_w\left(h_n^{-1}(W_i - w)\right).$$

3. Operator T

$$(\hat{T}\phi)(w) = \int_{[a,b]^p} \phi(z)\hat{f}_{ZW}(z,w)\mathrm{d}z.$$

4. Adjoint operator T^*

$$(\hat{T}^*\psi)(z) = \int_{[a,b]^q} \psi(w)\hat{f}_{ZW}(z,w)\mathrm{d}w.$$

Tikhonov regularization

Tikhonov-regularized estimator

$$\hat{\varphi}_{\alpha_n} = (\alpha_n I + \hat{T}^* \hat{T})^{-1} \hat{T}^* \hat{r},$$

where $\alpha_n \to 0$ is the regularization/ridge parameter.

Goal: honest uniform confidence set

1. Uniform confidence set is described by the pair of functions

$$C_n = \{[C_l(z), C_u(z)] : \ z \in [a,b]^p\}.$$

2. Confidence set, honest to the class $\mathcal{P}$

$$\inf_{P\in\mathcal{P}} \Pr(\varphi \in C_n) \geq \underbrace{1-\gamma}_{\text{coverage level}} - \underbrace{O(\delta_n)}_{\text{coverage error}}, \quad \gamma \in (0,1),$$

where $P \in \mathcal{P}$ is the distribution of (Y, Z, W).

Goal: honest uniform confidence set

1. Uniform confidence set is described by the pair of functions

$$C_n = \{[C_l(z), C_u(z)] : \; z \in [a, b]^p\}.$$

2. Confidence set, honest to the class $\mathcal{P}$

$$\inf_{P \in \mathcal{P}} \Pr(\varphi \in C_n) \geq \underbrace{1-\gamma}_{\text{coverage level}} - \underbrace{O(\delta_n)}_{\text{coverage error}}, \quad \gamma \in (0,1),$$

where $P \in \mathcal{P}$ is the distribution of (Y, Z, W).

3. Expected diameter of C_n should be as "small" as possible and should shrink as sample increases

$$\sup_{P \in \mathcal{P}} \mathbb{E} |C_n|_\infty = O(r_n).$$

Nonparametric IV: "variance"-"bias" decomposition

$$\hat{\varphi} - \varphi = \nu_n + \xi_n + B_n,$$

where

- ν_n is the leading stochastic term, loosely called the "variance";
- ξ_n is the stochastic noise generated by the estimation of the operator T (negligible comparing to the "variance");
- B_n is the regularization bias.

"Variance" term is empirical process

1. Building uniform confidence sets amounts to approximating quantiles of supremum of the "variance"

$$\|\nu_n\|_\infty = \left\| \frac{1}{nh_n^q} \sum_{i=1}^{n} U_i (\alpha_n I + T^* T)^{-1} T^* K_w \left(h_n^{-1}(W_i - .) \right) \right\|_\infty$$

2. Several complications:

"Variance" term is empirical process

1. Building uniform confidence sets amounts to approximating quantiles of supremum of the "variance"

$$\|\nu_n\|_\infty = \left\| \frac{1}{nh_n^q} \sum_{i=1}^{n} U_i(\alpha_n I + T^* T)^{-1} T^* K_w \left(h_n^{-1}(W_i - .)\right) \right\|_\infty$$

2. Several complications:
   - the process is a triangular array, changing with the sample size n;

"Variance" term is empirical process

1. Building uniform confidence sets amounts to approximating quantiles of supremum of the "variance"

$$\|\nu_n\|_\infty = \left\| \frac{1}{nh_n^q} \sum_{i=1}^{n} U_i(\alpha_n I + T^*T)^{-1} T^* K_w \left(h_n^{-1}(W_i - .)\right) \right\|_\infty$$

2. Several complications:
   - the process is a triangular array, changing with the sample size n;
   - as sample size increases, we need $\alpha_n \to 0$ for consistency of the estimator and in the limit we obtain discontinuous $(T^*T)^{-1}$.

Uniform central limit theorem breaks down

Theorem

For arbitrary ill-posed inverse problem, it is not possible to find a sequence s_n such that $s_n \nu_n$ would converge weakly in the space of continuous functions to a non-degenerate random process.

Alternatives to CLT...

1. Data-driven concentration inequality
   - works for any finite sample size n;
   - allows to obtain finite-sample bounds on quantiles;
   - free from coverage errors of Gaussian and/or bootstrap approximations;
   - does not impose any regularity on the process.

Alternatives to CLT...

1. Data-driven concentration inequality
   - works for any finite sample size n;
   - allows to obtain finite-sample bounds on quantiles;
   - free from coverage errors of Gaussian and/or bootstrap approximations;
   - does not impose any regularity on the process.
2. Bootstrap/Gaussian approximations (Chernozhukov, Chetverikov, and Kato, 2014), (Chernozhukov, Chetverikov, and Kato, 2016)
   - for each finite sample-size, approximate the supremum of empirical process by the supremum of the Gaussian process;
   - works even if there does not exist a well-behaved limiting Gaussian process.
   - have coverage errors.
   - relatively mild regularity conditions on the process: VC-type.

Concentration vs Gaussian approximation

Let $(X_i)_{i=1}^n$ be i.i.d.$\sim F$ estimated with EDF $F_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{(-\infty,x]}(X_i)$

1. Dvoretzky–Kiefer-Wolfowitz concentration inequality

$$\Pr\left(\|F_n - F\|_\infty \leq \sqrt{\frac{\log(2/\gamma)}{2n}}\right) \geq 1-\gamma$$

Concentration vs Gaussian approximation

Let $(X_i)_{i=1}^n$ be i.i.d.$\sim F$ estimated with EDF $F_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{(-\infty,x]}(X_i)$

1. Dvoretzky–Kiefer-Wolfowitz concentration inequality

$$\Pr\left(\|F_n - F\|_\infty \leq \sqrt{\frac{\log(2/\gamma)}{2n}}\right) \geq 1-\gamma$$

2. Komlós-Major-Tusnády strong approximation

$$\Pr\left(\|F_n - F\|_\infty \leq \frac{c_{1-\gamma}}{\sqrt{n}}\right) = 1-\gamma + O\left(\frac{\log n}{\sqrt{n}}\right),$$

where $c_{1-\gamma}$ is $1-\gamma$ quantile of supremum of some Gaussian process.

Roadmap

Supremum of empirical process

1. Concentration inequality: "statistics $T_n = f(X_1, \dots, X_n)$ that depends smoothly on many independent variables $(X_i)_{i=1}^n$ concentrates sharply around its mean"

$$\Pr\left(\|\nu_n\|_\infty > \mathbb{E}\|\nu_n\|_\infty + \sqrt{\frac{2x}{n}}\right) \leq e^{-x}, \qquad x > 0,$$

   where $\|\nu_n^\varepsilon\|_\infty$ is bootstrapped statistics.

2. Can replace $\mathbb{E}\|\nu_n\|_\infty$ by some data-driven quantity due to the symmetrization inequality.

Confidence sets based on concentration inequality

Confidence sets with $1-\gamma$, $\gamma \in (0,1)$ coverage can be described as

$$C_n^{\mathrm{ci}} = \left\{ \left[\hat{\varphi}(z) - q_n^{\mathrm{ci}}, \hat{\varphi}(z) + q_n^{\mathrm{ci}}\right] :\ z \in [a,b]^p \right\},$$

where

$$q_n^{\mathrm{ci}} = 2\left\|\hat{\nu}_n^{\varepsilon}\right\|_{\infty} + \frac{3\|\hat{T}^*\|_{2,\infty}\|K\|F\sqrt{2\log(1/\gamma)} + \log^{-1} n}{\alpha_n\sqrt{nh_n^q}},$$

where $\|\hat{\nu}_n^{\varepsilon}\|_{\infty}$ and $\|T^*\|_{2,\infty}$ are some data-dependent quantities.

Bootstrap/Gaussian approximations

Confidence set with $1-\gamma$, $\gamma \in (0,1)$ coverage can be described as

$$C_n^{\mathrm{g}} = \left\{ [\hat{\varphi}_{\alpha_n}(z) - q_n^{\mathrm{g}}, \hat{\varphi}_{\alpha_n}(z) + q_n^{\mathrm{g}}] : \; z \in [a,b]^p \right\},$$

where

$$q_n^{\mathrm{g}} = \frac{c_{1-\gamma}^* \|\hat{T}^*\|_{2,\infty} + \log^{-1} n}{\alpha_n \sqrt{n h_n^q}},$$

and $c_{1-\gamma}^*$ is quantile of bootstrapped supremum of empirical process.

Roadmap

NPIV: results

Theorem (Coverage errors and expected diameters)

Under some assumptions for some $\delta_n \to 0$

$$\inf_{P\in\mathcal{P}} \Pr\left(\varphi \in C_n^{\mathrm{ci}}\right) \geq 1-\gamma - O\left(\delta_n\right)$$

and

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \Pr\left(\varphi \in C_n^{\mathrm{g}}\right) \geq 1-\gamma.$$

Moreover, there exists $r_n \to 0$ *such that for any* $P \in \mathcal{P}$

$$|C_n^{\mathrm{s}}|_\infty = O_p\left(r_n\right), \qquad \mathrm{s} \in \{\mathrm{g}, \mathrm{ci}\}.$$

Roadmap

Numerical implementation

1. All estimators are extremely easy to implement: solving a system of linear equations

$$\hat{\varphi} = \left(\alpha_n \mathbf{I} + \mathbf{K}^\top \mathbf{K}\right)^{-1} \mathbf{K}^\top \mathbf{r}.$$

Numerical implementation

1. All estimators are extremely easy to implement: solving a system of linear equations
$$\hat{\varphi} = \left(\alpha_n \mathbf{I} + \mathbf{K}^\top \mathbf{K}\right)^{-1} \mathbf{K}^\top \mathbf{r}.$$
2. To obtain confidence sets, we just need to subtract and add corresponding q_n^{s} values to the estimator, where $\mathrm{s} \in \{\mathrm{g}, \mathrm{ci}\}$.

Monte Carlo experiments for NPIV

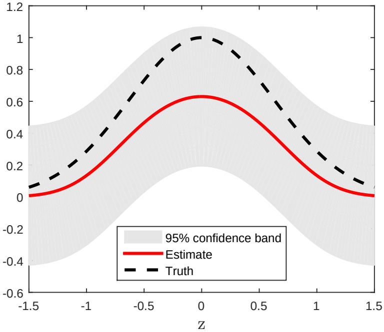
DGP

$$Y = \varphi(Z) + U, \qquad \varphi(z) = e^{-z^2/0.8},$$

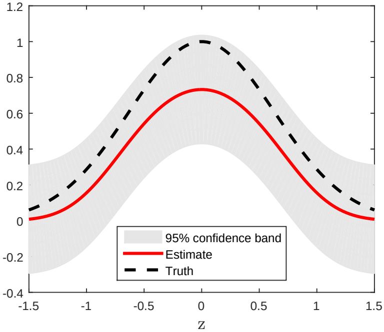
$$\begin{pmatrix} Z \\ W \\ U \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_z^2 & \rho\sigma_z\sigma_w & \sigma_{zu} \\ \rho\sigma_z\sigma_w & \sigma_w^2 & 0 \\ \sigma_{zu} & 0 & \sigma_u^2 \end{pmatrix} \right),$$

where $\sigma_z = \sigma_w = 0.3$, $\sigma_u = \sqrt{0.03}, \sigma_{zu} = 0.04$, and $\rho = 0.3$.

Monte Carlo experiments



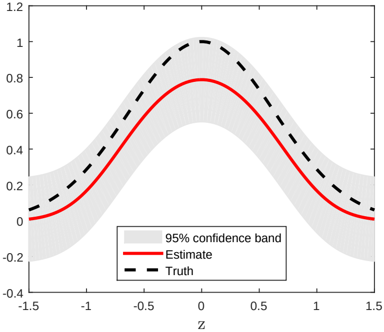
(a) $n = 1000$, $\hat{\gamma} = 0.89$



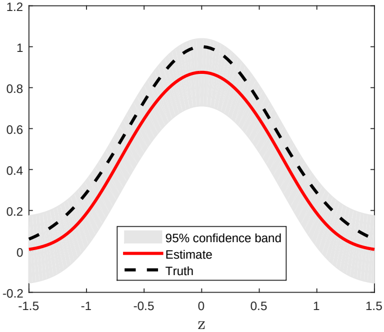
(b) $n = 5000$, $\hat{\gamma} = 0.94$

Figure: Estimates and confidence bands based on concentration inequality, average over 5000 experiments.

Monte Carlo experiments



(a) $n = 1000, \hat{\gamma} = 0.87$



(b) $n = 5000, \hat{\gamma} = 0.97$

Figure: Estimates and confidence bands based on Gaussian approximation, average over 5000 experiments.

Roadmap

1. Honest and uniform confidence sets for NPIV: construction
   - Concentration inequalities
   - Bootstrap approximations
2. Honest and uniform confidence sets for NPIV: results
3. Monte Carlo experiments
4. Engel curves in the US

Engel curves

1. Engel curves describe how the demand for commodity changes while the household's budget increases:
   - Y - share spent on commodity;
   - Z - total budget allocated to subgroup of commodities.

Engel curves

1. Engel curves describe how the demand for commodity changes while the household's budget increases:
   - Y - share spent on commodity;
   - Z - total budget allocated to subgroup of commodities.

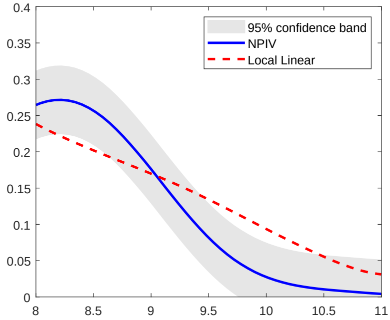
Engel curves

1. Engel curves describe how the demand for commodity changes while the household's budget increases:
   - Y - share spent on commodity;
   - Z - total budget allocated to subgroup of commodities.
2. Endogeneity: both Y and Z are likely to be jointly determined. Gross earnings of the household head is a valid IV, assuming that heterogeneity in earnings is not related to households' preferences over consumption.

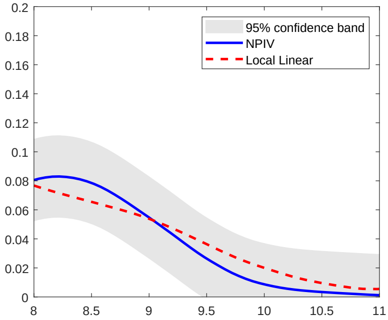
Application: confidence sets for Engel curves

- 2015 Consumer Expenditure Survey data in US;
- 5598 married couples with positive income during the past 12 month;
- Confidence sets for Engel curves for different commodities;
- Y share of total non-durable expenditures spent;
- Z log of total non-durable expenditures;
- W gross earnings of households' head.

Confidence sets for Engel curves



(a) Food



(b) Tobacco

Conclusions

1. Uniform confidence sets give us useful information about global shape properties: nonlinearities, monotonicity, concavity/convexity.

Conclusions

1. Uniform confidence sets give us useful information about global shape properties: nonlinearities, monotonicity, concavity/convexity.
2. Uniform CLT does not work and non-conventional approaches to inference are needed.

Conclusions

1. Uniform confidence sets give us useful information about global shape properties: nonlinearities, monotonicity, concavity/convexity.
2. Uniform CLT does not work and non-conventional approaches to inference are needed.
3. This paper provides two uniform inferential methods for NPIV.

Conclusions

1. Uniform confidence sets give us useful information about global shape properties: nonlinearities, monotonicity, concavity/convexity.
2. Uniform CLT does not work and non-conventional approaches to inference are needed.
3. This paper provides two uniform inferential methods for NPIV.
4. Extensions in the paper: a very general class of ill-posed models with Tikhonov regularization, including functional regressions and density deconvolutions.

Thank you!

email: babii.andrii@gmail.com