Consistency of Plug-In Estimators of Upper Contour and Level Sets

Neşe Yıldız

October 2008

Outline

- Propose an estimator for a set of parameters that satisfy finite number of moment inequalities or equalities. Provide conditions for (Hausdorff) consistency of this estimator.
- Contemplate about non-parametric extensions.

Recent Partial Identification Literature:

- Horowitz & Manski (2000): confidence interval for the identified set (univ.),
- Manski & Tamer (2002): consistent set estimator with interval data; extra slackness is introduced into objective function (*),
- Imbens & Manski (2004): confidence interval for set identified parameter (univ.),
- Chernozhukov, Hong & Tamer (2007) (CHT): first to provide general method of constructing confidence intervals for these models, extra slackness required,
- Andrews, Berry & Jia (2004): inference methods for incomplete models, no extra slackness required, but equality constraints are ruled out (*),
- Romano & Shaikh (2006): construct confidence sets by iterating on CHT,
- Rosen (2006): confidence sets of parameters defined by finite number of inequalities,
- Beresteaunu & Molinari (2006): inference method for sets that are (Aumann) expectations of SVRVs.
- Galichon & Henry (2006a,b): dilation bootstrap + mass transportation

Set Up I:

$\Theta \subseteq \mathbb{R}^I$: parameter set,
$\{X_1, ..., X_n\}$: data, random variables defined on probability space $(\Omega, \mathcal{F}, P)$.
$d_H(A, B) = \max\{\sup_{a \in A} \inf_{b \in B} \|a - b\|, \sup_{b \in B} \inf_{a \in A} \|a - b\|\}$.

$$\tilde{g} : \mathcal{X} \times \Theta \to \mathbb{R}^M \qquad \tilde{\varphi} : \mathcal{X} \times \Theta \to \mathbb{R}^S$$
$$g(\theta) = E\tilde{g}(X, \theta) \qquad \varphi(\theta) = E\tilde{\varphi}(X, \theta)$$
$$\hat{g}(\theta) = \frac{1}{n} \sum_i \tilde{g}(X_i, \theta) \qquad \hat{\varphi}(\theta) = \frac{1}{n} \sum_i \tilde{\varphi}(X, \theta)$$

$$\Theta_0 = \{\theta \in \Theta : g(\theta) \geq 0, \varphi(\theta) = 0\},$$
$$\hat{\Theta} = \{\theta \in \Theta : \hat{g}(\theta) \geq 0, \hat{\varphi}(\theta) = 0\}.$$

Goal: Show $d_H(\hat{\Theta}, \Theta_0) \xrightarrow{P} 0$.

Inequalities Only ($S = 0$):

Assumptions:

(A-1) Θ is compact and convex. $\Theta_0 \neq \emptyset$.

(AI-2) $\sup_{\theta \in \Theta} \|\hat{g}(\theta) - g(\theta)\| \xrightarrow{P} 0$.

Define

$$h(\theta) := \min\{g_1(\theta), ..., g_M(\theta)\}, \qquad \Theta^* := \{\theta \in \Theta : h(\theta) = 0\}.$$

Inequalities Only ($S = 0$):

Define:

$$\overline{\Theta}^{\epsilon} := \{\theta \in \Theta : g(\theta) \geq -\epsilon\}, \qquad \underline{\Theta}^{\epsilon} := \{\theta \in \Theta : g(\theta) \geq \epsilon\}.$$

Note for each $\epsilon > 0$,
(i) $\underline{\Theta}^{\epsilon} \subseteq \Theta_0 \subseteq \overline{\Theta}^{\epsilon}$, and
(ii) $\underline{\Theta}^{\epsilon} \subseteq \hat{\Theta} \subseteq \overline{\Theta}^{\epsilon}$ for all sufficiently large n with probability close to 1.

To show Hausdorff consistency of $\hat{\Theta}$ we will show $d_H(\overline{\Theta}^{\epsilon}, \Theta_0) \longrightarrow 0$ and $d_H(\underline{\Theta}^{\epsilon}, \Theta_0) \longrightarrow 0$ as $\epsilon \longrightarrow 0$.

Note (i) $\Rightarrow \sup_{\theta_0 \in \Theta_0} \inf_{\theta \in \overline{\Theta}^{\epsilon}} \|\theta - \theta_0\| = 0$ and $\sup_{\theta \in \underline{\Theta}^{\epsilon}} \inf_{\theta_0 \in \Theta_0} \|\theta - \theta_0\| = 0$

Inequalities Only ($S = 0$):

Remark 1 *Note that* $\Theta_0 = \{\theta \in \Theta : h(\theta) \geq 0\}$. *Therefore, when* $\Theta_0 \neq \emptyset$,

$$\Theta_0 = argmin_{\theta \in \Theta} \Big(|h(\theta)| 1\{h(\theta) \leq 0\} \Big).$$

Inequalities Only ($S = 0$):

Proposition 1 *If* Θ *is compact,* g *is continuous, and* $\Theta_0 \neq \emptyset$*, we have*

$$\sup_{\theta \in \overline{\Theta}^{\epsilon}} \inf_{\theta \in \Theta_0} ||\theta - \theta_0|| \to 0 \text{ as } \epsilon \to 0.$$

Remark 2 Θ *does not have to be connected for this proposition.*

Remark 3 *This part is not so novel.*

Inequalities Only ($S = 0$):

(AI-3) Let $\mathcal{O} \subseteq \mathbb{R}^I$, with $\mathcal{O}$ open and containing Θ. Assume that g is continuously differentiable over $\mathcal{O}$, and for each $\theta^* \in \Theta^*$ $Dg(\theta^*)$ has rank M.

Proposition 2 *Suppose (A-1), (AI-2) and (AI-3) hold. In addition suppose that $M \leq I$ and $\Theta^* \subseteq \text{int}(\Theta)$. Then we have*

$$\sup_{\theta \in \Theta_0} \inf_{\theta \in \underline{\Theta}^\epsilon} ||\theta - \theta_0|| \to 0 \text{ as } \epsilon \to 0.$$

Generalized Inverse Function Theorem:

Theorem 1 *(Luenberger)* *Let* x_0 *be a regular point of a transformation* T *mapping the Banach space* X *into the Banach space* Y *(i.e.* $DT(x_0)$ *maps* X *onto* Y*). Then there is a neighborhood* $N(y_0)$ *of the point* $y = T(x_0)$ *and a constant* K *such that the equation* $y = T(x)$ *has a solution for every* $y \in N(y_0)$ *and the solution satisfies* $\|x - x_0\| \leq K\|y - y_0\|$.

Corollary 1 *Suppose every element,* x_0*, of a compact set* F *is a regular point of* T*. Then there is* $r > 0$ *and* $K < \infty$ *(not depending on* x_0*) for each* $x_0 \in F$ *with* $y_0 = T(x_0)$*, the equation* $y = T(x)$ *has a solution for every* $y \in B_r(y_0)$*, and the solution satisfies* $\|x - x_0\| \leq K\|y - y_0\|$.

A few observations:

1. (AI-3) implies $Dg(\theta)Dg(\theta)^T$ has rank M for each $\theta \in \Theta^*$. And all eigenvalues of this matrix are real and positive. Let $\lambda_M(\theta)$ denote the smallest eigenvalue.
2. Using Courant-Fischer Min-Max Theorem we could write each eigenvalue as the value function of an optimization problem. Then Maximum Theorem implies that the eigenvalues vary continuously in θ.
3. Θ^* is compact. Thus $\inf_{\theta^* \in \Theta^*} \lambda_M(\theta^*) =: \underline{\lambda} > 0$. Also there is $\rho > 0$ such that $\theta \in \overline{B_\rho(\theta^*)} \Rightarrow \lambda_M(\theta) \geq \frac{1}{2}\underline{\lambda}$ and $\theta \in \Theta$.
4. Every point in the compact set $\cup_{\theta^* \in \Theta^*} \overline{B_\rho(\theta^*)}$ is regular. We apply the Corollary to this set.
5. $\epsilon_1^* := \frac{1}{2} \inf\{h(\theta) : h(\theta) \geq 0, \theta \in \Theta \setminus B_\rho(\theta^*)\} > 0$.

Proof of Proposition (2):

Proof 1 *Let* $\delta > 0$*, and consider* $0 < \epsilon < \min\{\frac{r}{\sqrt{M}}, \frac{\delta}{K\sqrt{M}}, \epsilon_1^*, \frac{\eta}{2K\sqrt{M}}\}$*. For any* θ_0 *with* $h(\theta_0) \geq \epsilon$*,* $\theta_0 \in \underline{\Theta}^\epsilon$*, we have* $\inf_{\theta \in \underline{\Theta}^\epsilon} ||\theta - \theta_0|| = 0$*. Thus, consider* θ_0 *such that* $h(\theta_0) \in [0, \epsilon)$*. Note that if there is no such* θ_0 *we have nothing to prove because in that case* $\Theta_0 = \underline{\Theta}^\epsilon$*. Next, let* $t \in \mathbb{R}^K$ *be defined by* $t_m := g_m(\theta_0) + \epsilon - h(\theta_0)$*. Then*

$$\|t - g(\theta_0)\| = \sqrt{\sum_m (g_m(\theta_0) + \epsilon - h(\theta_0) - g_m(\theta_0))^2} = \sqrt{M}(\epsilon - h(\theta_0)) \leq \sqrt{M}\epsilon < r.$$

Therefore, there is a θ' *such that* $g_m(\theta') = g_m(\theta_0) + \epsilon - h(\theta_0)$*, and*

$$\|\theta_0 - \theta'\| \leq K\sqrt{M}\epsilon < \delta.$$

On the other hand, since $h(\theta_0) < \epsilon$*,* $h(\theta') \geq \epsilon$*, i.e.* $\theta' \in \underline{\Theta}^\epsilon$*. Thus,* $\inf_{\theta \in \underline{\Theta}^\epsilon} ||\theta - \theta_0|| \leq ||\theta' - \theta_0|| < \delta$*. Thus,* $\inf_{\theta \in \underline{\Theta}^\epsilon} ||\theta - \theta_0|| \leq \delta$ *for each* $\theta_0 \in \Theta_0$*.*

What if $\hat{\Theta}_n$ is empty?

$\Theta_0 \neq \emptyset \Rightarrow \hat{\Theta}_n \neq \emptyset \; w.p.a.1.$

In small samples $\hat{\Theta}_n$ could be empty.

Recall $\Theta_0 = \{\theta \in \Theta : \theta \text{ minimizes } |h(\theta)| 1\{h(\theta) \leq 0\}\}$.

Alternative estimator: $\hat{\Theta}_n^a = \{\theta \in \Theta : \theta \text{ minimizes } |\hat{h}(\theta)| 1\{\hat{h}(\theta) \leq 0\}\}$.

$\hat{\Theta}_n$ and $\hat{\Theta}_n^a$ are equal asymptotically because they coincide whenever $\hat{\Theta}_n \neq \emptyset$.

Equality Constraints Only ($M = 0$):

(AE-2) $\sup_{\theta \in \Theta} \|\hat{\varphi}(\theta) - \varphi(\theta)\| \xrightarrow{P} 0.$

(AE-3) Let $\mathcal{O} \subseteq \mathbb{R}^I$, with $\mathcal{O}$ open and containing Θ. Assume that φ is continuously differentiable over $\mathcal{O}$, and for each $\theta^* \in \Theta^*$ $D\varphi(\theta^*)$ has rank S.

$\forall \epsilon > 0$, let $\Theta(\epsilon) := \{\theta \in \Theta : \varphi(\theta) \in [-\epsilon, \epsilon]^S\}$. Note, for sufficiently large n, with probability close 1 we have,

$$\sup_{\theta \in \hat{\Theta}} \inf_{\theta' \in \Theta_0} \|\theta - \theta'\| \leq \sup_{\theta \in \Theta(\epsilon)} \inf_{\theta' \in \Theta_0} \|\theta - \theta'\|.$$

Proposition 3 *Suppose (A-1), (AE-2) and (AE3) hold. Then*

$$\sup_{\theta \in \Theta(\epsilon)} \inf_{\theta' \in \Theta_0} \|\theta - \theta'\| \to 0, \text{as } \epsilon \to 0.$$

Equality Constraints Only ($M = 0$):

(AE-2′) $\sup_{\theta\in\Theta} \|\hat{\varphi}(\theta) - \varphi(\theta)\| \overset{a.s.}{\to} 0.$

Proposition 4 *Suppose (A-1), (AE-2′)-(AE-3) hold. Then*

$$\sup_{\theta\in\Theta_0} \inf_{\theta'\in\hat{\Theta}} \|\theta - \theta'\| \overset{P}{\to} 0.$$

Proof is slightly more involved than that of Proposition 2 because we cannot use a set like $\underline{\Theta}^{\epsilon}$*. We can apply the Inverse Function Theorem for each* $\hat{\varphi}_n$*, but the radius* r *and Lipschitz constant* K *may depend on* n*. To get uniformity over* n *use Egoroff's Theorem.*

Equality Constraints Only ($M = 0$):

$\hat{\Theta}_n$ is non-empty asymptotically, but may be empty in small samples.

Use $\hat{\Theta}_n^b := \{\theta \in \Theta : \theta \text{ minimizes } \hat{\varphi}_n(\theta')^T W \hat{\varphi}_n(\theta')\}$, where W is positive definite symmetric matrix, or

Use $\hat{\Theta}_n^a := \{\theta \in \Theta : \theta \text{ minimizes } |\hat{h}(\theta)| 1\{\hat{h}(\theta) \leq 0\}$,
where $\hat{h}(\theta) = \min\{\hat{\varphi}_1(\theta), ..., \hat{\varphi}_1(\theta), -\hat{\varphi}_1(\theta), ..., -\hat{\varphi}_1(\theta)\}$.

$\hat{\Theta}_n, \hat{\Theta}_n^a$ and $\hat{\Theta}_n^b$ are asymptotically equal.

Equality and Inequality Constraints Together ($M + S \leq I$, $M, S > 0$):

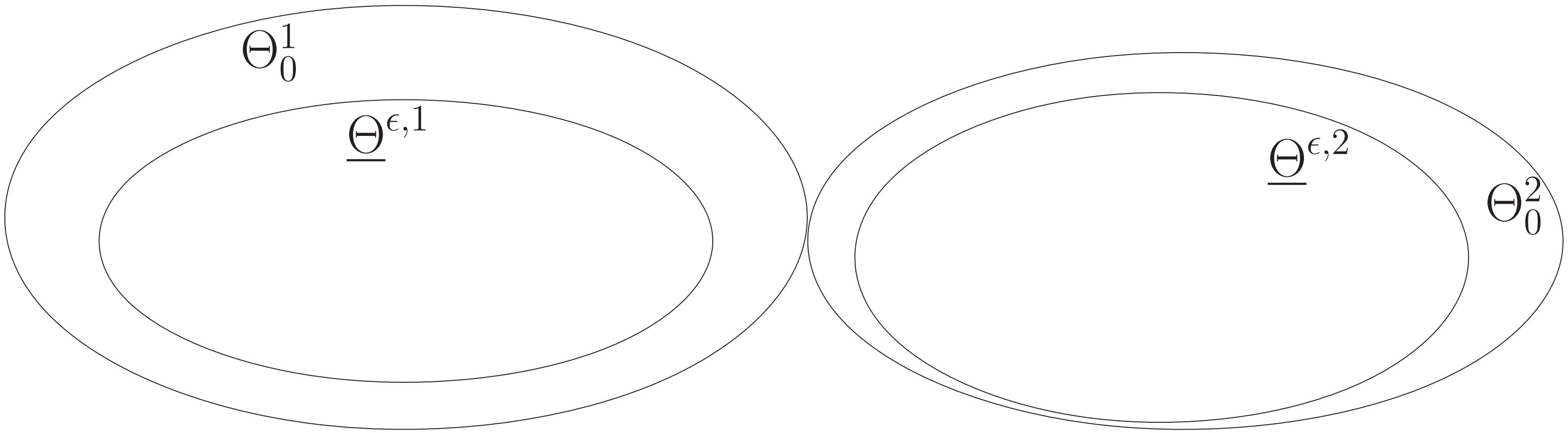
Can show $d_h(\hat{\Theta}_n, \Theta_0) \xrightarrow{P} 0$ under appropriate rank condition.

In small samples may want to use

$$\hat{\Theta}_n^b := \{\theta \in \Theta : \theta \text{ minimizes } |\hat{h}(\theta')| 1\{\hat{h}(\theta') \leq 0\} + \hat{\varphi}(\theta')^T W \hat{\varphi}(\theta')\}.$$

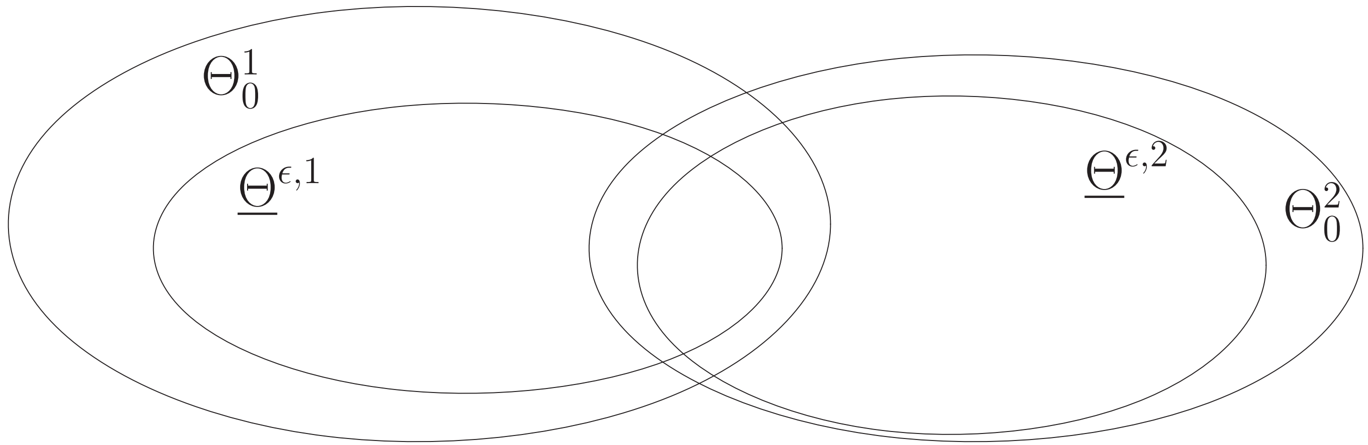
What if $M + S > I$?

Divide constraints into L groups with $M_l + S_l \leq I$ for each group?



Problem: the intersection of Θ_0^l will have empty interior with equality constraints.

What if $M + S > I$?



It's tempting to think non-empty intersection saves the day.

$M > I$, $S = 0$

(AI-4) Suppose either that Θ_0 is singleton or that for each $\theta^* \in \Theta^*$ there is j_0 such that h is strictly monotone in θ_{j_0} at θ^* where $h(\theta) := \min\{h^l(\theta) : l = 1, ..., L\}$ and $\Theta^* = \{\theta \in \Theta : h(\theta) = 0\}$.

Proposition 5 *Suppose that (A1), (AI-2) through (AI-4) hold, and that* $\Theta^* \subseteq \mathrm{int}(\Theta)$. *Then* $d_H(\hat{\Theta}_n^a, \Theta_0) \xrightarrow{P} 0$ *where* $\hat{\Theta}_n^a := \{\theta \in \Theta : \theta \text{ minimizes } |\hat{h}(\theta')|1\{\hat{h}(\theta') \leq 0\}\}$.

Proof 2 *Verify the conditions for Theorem 1 of Andrews, Berry and Jia (2004).*

$$M + S > I, \; S \neq 0$$

Find I constraints that satisfy the rank condition. Use that as the initial estimator for the CHT /Romano-Shaikh procedure.

Example: Returns to Schooling

$$Y = \alpha_0 + \alpha_1 Educ + \alpha_2 Educ^2 + U.$$

B, quarter of birth, is orthogonal to U. Let $\theta := (\alpha_0, \alpha_1, \alpha_2) \in \mathcal{A}$.

$$\varphi(\theta) = \begin{pmatrix} E(Y - \alpha_0 - \alpha_1 Educ - \alpha_2 Educ^2) \\ E[(Y - \alpha_0 - \alpha_1 Educ - \alpha_2 Educ^2)B] \end{pmatrix}, \quad \Theta_0 = \{(\alpha_0, \alpha_1, \alpha_2) \in \mathcal{A} : \varphi(\theta) = 0\}.$$

Note that the Andrews, Berry and Jia (2004) approach is not applicable in this case. The Jacobian condition imposed in this paper on the other hand is that the rank of the following matrix is 2:

$$D\varphi(\theta) = \begin{bmatrix} -1 & -E(Educ) & -E(Educ^2) \\ -E(B) & -E(Educ * B) & -E(Educ^2 * B) \end{bmatrix}.$$

This condition will be satisfied unless $E(Educ * B) = E(Educ) * E(B)$ and $E(Educ^2 * B) = E(Educ^2) * E(B)$.

Example: Returns to Schooling

Chernozhukov & Hansen (2006), Chernozhukov, Hansen & Jansson (2005) estimate τ quantile function of potential income.
Y: potential income, X: covariates including schooling, Z: instruments, $d_X \geq d_Z$.

$$\varphi(\theta) := E[(\tau - 1\{Y \leq X^T\theta\})Z].$$

$$D\varphi(\theta) = -E\left[\; ZX_1 f_Y(X^T\theta|X,Z) \;\; ZX_2 f_Y(X^T\theta|X,Z) \;\; \cdots \;\; ZX_{d_x} f_Y(X^T\theta|X,Z) \;\right]$$

Rank condition: For all θ satisfying $\varphi(\theta) = 0$, the $d_z \times d_x$ matrix $D\varphi(\theta)$ has rank d_z.

Set Up II:

$Y = g_0(X) + U$, $E(U|Z) = 0$, $\pi(Z) := E[Y|Z]$, $g_0 \in \mathcal{L}^2(X)$, $\pi \in \mathcal{L}^2(Z)$.

Data $\{Y_i, X_i, Z_i\}$ implies π is identified. g_0 unknown.

Goal: Estimate the set of functions $\mathcal{G}_0 := \{g \in \mathcal{L}^2(X) : E[g(X)|Z] = \pi(Z) \ a.s.\}$.

Idea: Use the Generalized Inverse Function Theorem to do this.

Estimation of g_0:

$T_F : \mathcal{L}_F^2(W) \to \mathcal{L}_F^2(Z), g \to T_F(g) := E[g(X_0, Z_1)|Z_1, Z_2] - \pi(Z_1, Z_2).$

$\delta T_F(g;h)$: Fréchet differential of T_F at g with increment h is defined by

$$\lim_{\|h\|_{\mathcal{L}^2} \to 0} \frac{\|T_F(g+h) - T_F(g) - \delta T_F(g;h)\|_{\mathcal{L}^2}}{\|h\|_{\mathcal{L}^2}}.$$

g is a regular point of T_F if $\delta T_F(g;h)$ maps $\mathcal{L}_F^2(X)$ onto $\mathcal{L}_F^2(Z)$.

In our case, $\delta T_F(g;h)$ does not depend on g, so each $g \in \mathcal{L}_F^2(X)$ is a regular point of T_F if for each $r \in \mathcal{L}_F^2(Z)$ there exists an $h \in \mathcal{L}_F^2(X)$ such that $\delta T_F(g;h) = E[h(X)|Z] = r(Z)$.

Special Case I:

Image of $\delta T_F(g; h)$ for each g:
$\mathcal{A} := \{E[h(X)|Z] : h \in \mathcal{L}^2(X)\} \subseteq \mathcal{L}^2(Z)$.
It is also the image of T_F.

When is $\mathcal{A}$ a Banach space?
$\mathcal{A}$ is a Banach space when $\mathcal{L}^2(Z) \subseteq \mathcal{L}^2(X)$.

Note: In this case, $\pi \in \mathcal{A}$.

$\mathcal{A}$ is a Banach Space:

1. $\mathcal{A}$ is a normed vector space because $0 \in \mathcal{L}^2(X)$ and the mapping T_F is linear.
2. Completeness: let $\{\varphi_n\}_{n=1}^{\infty} \subseteq \mathcal{A}$ be a Cauchy sequence.
   $\mathcal{L}^2(Z)$ complete implies there is a $\varphi \in \mathcal{L}^2(Z)$ such that $\varphi_n \xrightarrow{\mathcal{L}^2} \varphi$.
   Since $\mathcal{L}^2(Z) \subseteq \mathcal{L}^2(X)$, $\varphi \in \mathcal{L}^2(X)$ and $E[\varphi(Z)|Z] = \varphi(Z)$ so that $\varphi \in \mathcal{A}$.

   Thus, $\mathcal{A}$ is a Banach space.

Special Case I:

When T_F is viewed as a mapping from $\mathcal{L}^2(X)$ to $\mathcal{A}$ every point of $\mathcal{L}^2(X)$ is a regular point of T_F by construction.

As a result, the Generalized Inverse Function Theorem tells us that there is a ball with radius r centered at π and a constant K (in this case neither r nor K depend on g_0) such that for all θ in that ball the equation $T_F(h) = \theta$ has a solution. Moreover, the solution will satisfy $\|h - g_0\| \leq K\|\pi - \theta\|$.

Special Case I:

Theorem 2 *Suppose we have an estimator,* π_n*, for* π *such that* $\pi_n \in \mathcal{L}^2(Z)$ *and* $\pi_n \xrightarrow{\mathcal{L}^2} \pi$ *as* $n \to \infty$*, then*

$$\max\{\sup_{g_0 \in \mathcal{G}_0} \inf_{g \in \mathcal{G}_n} \|g - g_0\|_{\mathcal{L}^2}, \sup_{g \in \mathcal{G}_n} \inf_{g_0 \in \mathcal{G}_0} \|g - g_0\|_{\mathcal{L}^2}\} \to 0,$$

where

$$\mathcal{G}_0 := \{g \in \mathcal{L}^2(X) : E[g(X)|Z] = \pi(Z) \ a.s.\},$$
$$\mathcal{G}_n := \{g \in \mathcal{L}^2(X) : E[g(X)|Z] = \pi_n(Z) \ a.s.\}.$$

Special Case I:

Proof 3 *We know that there exists* $r > 0$ *and* $K < \infty$ *such that for all* $a \in \mathcal{A}$ *whenever* $b \in B_r^{\mathcal{L}^2}(a)$ *the equation* $b(Z) = E[g(X)|Z]$ *has a solution* $g_b \in \mathcal{L}^2(X)$*, and* $\|g_b - g_a\|_{\mathcal{L}^2} \leq K\|a - b\|_{\mathcal{L}^2}$*. Let* $\varepsilon \in (0, r)$ *and* n *be sufficiently large so that* $\|\pi - \pi_n\|_{\mathcal{L}^2} < \varepsilon$*. Consider* $g_0 \in \mathcal{G}_0$*. By the inverse function theorem there exists* $\tilde{g} \in \mathcal{G}_n$ *such that* $E[\tilde{g}(X)|Z] = \pi_n(Z)$ *and* $\|g_0 - \tilde{g}\|_{\mathcal{L}^2} \leq K\varepsilon$*. Thus,*

$$\sup_{g_0 \in \mathcal{G}_0} \inf_{g \in \mathcal{G}_n} \|g - g_0\|_{\mathcal{L}^2} \to 0.$$

Using almost identical steps, we can also show that

$$\sup_{g \in \mathcal{G}_n} \inf_{g_0 \in \mathcal{G}_0} \|g - g_0\|_{\mathcal{L}^2} \to 0. \qquad \blacksquare$$

Special Case II:

$Y \leq g_0(X) + U$ and $E(U|Z) = 0$.
Note that this model is not distinguishable from the model where $Y = g_0(X) + U$ and $E(U|Z) \geq 0$.

Define the mapping $T_F : \mathcal{L}^2(X) \to \mathcal{L}^2(Z)$ as before.

When $\sigma(X) \subseteq \sigma(Z)$ the Fréchet differential maps $\mathcal{L}^2(X)$ onto $\mathcal{L}^2(Z)$, and every $g \in \mathcal{L}^2(X)$ is a regular point of T_F.

However, X is exogenous because

$$E(U|X) = E[E(U|Z)|X] = 0.$$

Nonparametric Regression Inequalities:

Assume $X = Z$. Under this assumption T_F is the identity mapping, and finding inverse is trivial.

Suppose that there exists a set $B \in \sigma(X)$ and estimator, π_n, of π such that $P(B) > 0$ and $\sup_{x \in B} |\pi_n(x) - \pi(x)| \stackrel{a.s.}{\rightarrow} 0$. Let $\nu > 0$ and

$$\begin{aligned} \mathcal{G}_0 &:= \{g \in \mathcal{L}^2(X) : g(x) \geq \pi(x) \; for \; a.e. \; x \in B\}, \\ \overline{\mathcal{G}}^{\nu} &:= \{g \in \mathcal{L}^2(X) : g(x) \geq \pi(x) - \nu \; for \; a.e. \; x \in B\}, \\ \underline{\mathcal{G}}^{\nu} &:= \{g \in \mathcal{L}^2(X) : g(x) \geq \pi(x) + \nu \; for \; a.e. \; x \in B\}. \end{aligned}$$

Nonparametric Regression Inequalities:

Theorem 3

$$\max\{\sup_{g_0\in\mathcal{G}_0}\inf_{g\in\underline{\mathcal{G}}^\nu}\|(g-g_0)1_B\|_{\mathcal{L}^2}, \sup_{g\in\overline{\mathcal{G}}^\nu}\inf_{g_0\in\mathcal{G}_0}\|(g-g_0)1_B\|_{\mathcal{L}^2}\}\to 0,$$

as $\nu \to 0$.

Proof 4 *Easy.*

Nonparametric Regression Inequalities:

Proof 5 *Let* $g_0 \in \mathcal{G}_0$. *If* $g_0(x) \geq \pi(x) + \nu$ *for* $a.e.$ $x \in B$, *then* $\inf_{g \in \underline{\mathcal{G}}^\nu} \|(g - g_0)1_B\|_{\mathcal{L}^2} = 0$. *So consider* $g_0 \in \mathcal{G}_0$ *such that* $\pi(x) \leq g_0(x) \leq \pi(x) + \nu$ *on a set* $A \subseteq B$ *such that* $A \in \sigma(X)$ *and* $P(A \cap B) > 0$.[1] *Let* $\tilde{g}(x) = g_0(x)$ *for* $x \in B \setminus A$ *and* $\tilde{g}(x) = g_0(x) + \nu$ *for* $x \in B \cap A$. *Then* $\tilde{g} \in \underline{\mathcal{G}}^\nu$. *Moreover,* $\|(\tilde{g} - g_0)1_B\|_{\mathcal{L}^2} = \nu P(A \cap B)$. *Thus,* $\sup_{g_0 \in \mathcal{G}_0} \inf_{g \in \underline{\mathcal{G}}^\nu} \|(g - g_0)1_B\|_{\mathcal{L}^2} \to 0$ *as* $\nu \to 0$. *Similarly, let* $g \in \overline{\mathcal{G}}^\nu \setminus \mathcal{G}_0$. *Since* g *and* π *are* $\sigma(X)$ *measurable, this means that there exists a set* $E \in \sigma(X)$ *with* $P(E \cap B) > 0$ *such that* $\pi(x) - \nu \leq g(x) \leq \pi(x)$ *for* $x \in E$. *Define* $h(x) := g(x)$ *for* $x \in B \setminus E$ *and* $h(x) := g(x) + \nu$ *for* $x \in B \cap E$. *Then* $h \in \mathcal{G}_0$ *and* $\|(h - g_0)1_B\|_{\mathcal{L}^2} = \nu P(E \cap B)$.

[1] Since g_0 and π are $\sigma(X)$ measurable $A \in \sigma(X)$.

Future:

Try to find other conditions on the relationship between X and Z under which

$$\{E[h(X)|Z] : h \in \mathcal{L}^2(X)\}$$

is a complete subspace of $\mathcal{L}^2(Z)$.