Two new approaches in specification tests for the propensity score

Wu Zixuan

January 15, 2020

Content

Introduction

The propensity score, defined as the conditional probability of receiving treatment given covariates, is one of the most widely used tools in causal inference. Several methods exploiting this important tool include matching, e.g. Rosenbaum and Rubin (1985), Heckman et al. (1997) and Abadie and Imbens (2016); inverse probability weighting (IPW), e.g. Rosenbaum (1987), Hirano et al. (2003) and Donald and Hsu (2014); regression methods, e.g. Hahn (1998) and Firpo (2007).

However, the propensity score is usually unknown and has to be estimated. If the covariate space is of high dimensionality, we prefer to construct a parametric model for the propensity score since nonparametric methods suffers from "the curse of dimensionality". Thus we need to pay extra attention to the problem of model misspecification. Frölich (2004), Huber et al. (2013) and Busso et al. (2014) have shown that propensity score misspecifications can lead to misleading treatment effect estimates.

Introduction

Specification test for propensity score:

$$H_0 : \exists \theta_0 \in \Theta, \text{ s.t. } E[D - q(X, \theta_0) | X] = 0 \text{ a.s.}$$

First, consider dimensionality reduction:

$$H_0 : \exists \theta_0 \in \Theta, \text{ s.t. } E[D - q(X, \theta_0) | q(X, \theta_0)] = 0 \text{ a.s.}$$

Next, consider getting rid of tunning parameters:

$$H_0 : \exists \theta_0 \in \Theta, \text{ s.t. } E[\varepsilon(\theta_0) w(q(X, \theta_0), u)] = 0 \text{ for } \forall u \in \Pi$$

We focus on two new approaches proposed by Shaikh et al's (2009) and Sant'Anna et al's (2019). The former one adopts the local smoothing approach, see e.g. Zheng (1996), Fan and Li (1996), while the latter one adopt the intergrated conditional moment approach, see e.g. Stute (1997) and Escanciano (2006).

Introduction

Let D be a binary random variable indicating participation in the program. Define $Y(1)$ and $Y(0)$ as potential outcomes under treatment and control. Then the observed outcome is $Y = DY(1) + (1 - D)Y(0)$. And X is an observable $k \times 1$ pre-treatment covariates vector. We have a random sample $\{(Y_i, D_i, X_i')'\}_{i=1}^{n}$ of size $n \geqslant 1$ from $(Y, D, X')'$.

The main goal in causal inference is to assess the effect of a treatment D on the outcome Y, such as the average treatment effect, $ATE = E[Y(1) - Y(0)]$, and the average treatment effect on the treated, $ATT = E[Y(1) - Y(0)|D = 1]$.

And the propensity score $p(x)$ is defined as

$$p(x) = \mathbb{P}(D = 1|X = x)$$

Assumptions and Tests

Assumption 2.1 $\quad (Y(1), Y(0)) \perp\!\!\!\perp D | X.$

Assumption 2.2 $\quad \forall x \in \mathcal{X}, 0 < p(x) < 1.$

Rosenbaum (1987) has proved that under assumption 2.1-2.2, ATE and ATT can be identified by

$$ATE = \mathbb{E}\left[\left(\frac{D}{p(X)} - \frac{(1-D)}{1-p(X)}\right) Y\right] \text{ and } ATT = \frac{\mathbb{E}\left[\left(D - \frac{p(X)(1-D)}{1-p(X)}\right)\mathrm{Y}\right]}{\mathbb{E}[D]}$$

Assumptions and Tests

Lemma 1.

Let $\alpha = \frac{\mathbb{P}(D=0)}{\mathbb{P}(D=1)}$ and assume that $0 < \mathbb{P}(D = 1) < 1$. If $0 < p(X) < 1$ a.s., then for $\forall u \in [0, 1]$,

$$\mathbb{E}[I\{p(X) \leqslant u\}|D = 1] = \alpha \mathbb{E}\left[\frac{p(X)}{1-p(X)} I\{p(X) \leqslant u\}|D = 0\right] \qquad (1)$$

Furthermore, (1) holds if and only if for $\forall u \in [0, 1]$,

$$\mathbb{E}[(D - p(X))I\{p(X) \leqslant u\}] = 0 \qquad (2)$$

Assumptions and Tests

Motivated from Lemma 1, we can test whether a parametric model for $p(x)$ is correctly specified based on:
For some $\theta_0 \in \Theta$ and all $u \in \Pi$,

$$H_0 : \mathbb{E}\left[\left(D - q\left(X, \theta_0\right)\right) I\left\{q\left(X, \theta_0\right) \leqslant u\right\}\right] = 0 \tag{3}$$

where $\Theta \subset \mathbb{R}^k, \Pi = [0, 1]$, and $q(X, \theta) : \mathcal{X} \times \Theta \mapsto [0, 1]$ is a family of parametric functions known up to the finite dimensional parameter θ.

Notice that according to Stute (1997), (3) can be equivalently written as

$$H_0 : \mathbb{E}\left[D - q\left(X, \theta_0\right) | q\left(X, \theta_0\right)\right] = 0 \text{ a.s. for some } \theta_0 \in \Theta \tag{4}$$

Assumptions and Tests

[Shaikh et al.'s (2009) test]

Following the test proposed by Zheng (1996), Shaikh et al's (2009) test is given by

$$H_0 : \exists \theta_0 \in \Theta \text{ s.t. } \Pr\left\{E\left[D|q\left(X,\theta_0\right)\right] = q\left(X,\theta_0\right)\right\} = 1$$
$$H_1 : \Pr\{E[D|q(X,\theta)] = q(X,\theta)\} < 1 \quad \forall \theta \in \Theta$$

And the test statistic is given by

$$T_n\left(h_n\right) = \sqrt{\frac{n-1}{n}}\frac{nh_n^{1/2}\hat{V}_n(h_n)}{\sqrt{\hat{\Sigma}_n(h_n)}} \tag{5}$$

where

$$\hat{V}_n\left(h_n\right) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\frac{1}{h_n}K\left(\frac{q\left(X_i,\hat{\theta}_n\right)-q\left(X_j,\hat{\theta}_n\right)}{h_n}\right)\varepsilon_i\left(\hat{\theta}_n\right)\varepsilon_j\left(\hat{\theta}_n\right)$$
$$\hat{\Sigma}_n\left(h_n\right) = \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\frac{1}{h_n}K^2\left(\frac{q\left(x_i,\hat{\theta}_n\right)-q\left(x_j,\hat{\theta}_n\right)}{h_n}\right)\varepsilon_i^2\left(\hat{\theta}_n\right)\varepsilon_j^2\left(\hat{\theta}_n\right)$$
$$\varepsilon_i\left(\hat{\theta}_n\right) = D_i - q\left(X_i,\hat{\theta}_n\right)$$

Assumptions and Tests

Denote $\varepsilon = D - q(X, \theta_0)$ and $f_q(\cdot)$ as the density of $q(X, \theta_0)$. Then under H_0,

$$\mathbb{E}[\varepsilon \mathbb{E}[\varepsilon|X] f_q(q(X, \theta_0))] = 0$$

while under H_1 this term is greater than 0.

Apply kernel estiomation method,

$$\hat{\mathbb{E}}(\varepsilon|x) = \frac{1}{n-1} \sum_{j=1, j\neq i}^{n} \frac{1}{h_n} K\left(\frac{q(X_i, \hat{\theta}_n) - q(X_j, \hat{\theta}_n)}{h_n}\right) \varepsilon_j(\hat{\theta}_n) / \hat{f}_q\left(q(X_i, \hat{\theta}_n)\right)$$

Thus the sample analogue is

$$\hat{V}_n(h_n) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \frac{1}{h_n} K\left(\frac{q\left(X_i, \hat{\theta}_n\right) - q\left(X_j, \hat{\theta}_n\right)}{h_n}\right) \varepsilon_i\left(\hat{\theta}_n\right) \varepsilon_j\left(\hat{\theta}_n\right)$$

Assumptions and Tests

[Projection-based specification test]

Recall that $\varepsilon_i(\theta) = D_i - q(X_i, \theta)$. For all $u \in \Pi$, define

$$\mathcal{P}_n I\{q(X,\theta) \leqslant u\} = I\{q(X,\theta) \leqslant u\} - g'(X,\theta)\Delta_n^{-1}(\theta) G_n(u,\theta)$$

where $g(x,\theta) = \partial q(x,\theta)/\partial\theta$ is the score function of $q(X,\theta)$, and

$$G_n(u,\theta) = \tfrac{1}{n}\textstyle\sum_{i=1}^{n} g(X_i,\theta)\, I\{q(X_i,\theta) \leqslant u\}$$

$$\Delta_n(\theta) = \tfrac{1}{n}\textstyle\sum_{i=1}^{n} g(X_i,\theta)\, g'(X_i,\theta)$$

Notice that $\mathbb{E}\left[g(X,\theta_0)\,\mathcal{P} I\{q(X,\theta_0) \leqslant u\}\right] \equiv 0$ almost everywhere in $u \in \Pi$.

Assumptions and Tests

Consider projection-based empirical process $\hat{R}_n^p(u)$, $\hat{\theta}_n$ is a $\sqrt{n}$-consistent estimator for θ_0 under H_0:

$$\hat{R}_n^p(u) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i \left(\hat{\theta}_n\right) \mathcal{P}_n I \left\{ q\left(X_i, \hat{\theta}_n\right) \leqslant u \right\} \tag{6}$$

Our test statistics are based on continuous functionals of (6), especially the Cramér–von Mises-type and Kolmogorov–Smirnov-type functionals:

$$CvM_n = \int_{\Pi} \left| \hat{R}_n^p(u) \right|^2 F_n(du) = \frac{1}{n} \sum_{i=1}^{n} \left[\hat{R}_n^p \left(q\left(X_i, \hat{\theta}_n\right)\right)\right]^2 \tag{7}$$

$$KS_n = \sup_{u \in \Pi} \left| \hat{R}_n^p(u) \right| \tag{8}$$

Assumptions and Tests

Consider the unprojection version

$$\hat{R}_n(u) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i \left(\hat{\theta}_n\right) I \left\{ q \left(X_i, \hat{\theta}_n \right) \leqslant u \right\}$$

We can prove that

$$\hat{R}_n(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i \left(\theta_0\right) I \left\{ q \left(X_i, \theta_0 \right) \leqslant u \right\} - \sqrt{n} \left(\hat{\theta}_n - \theta_0\right)' \mathbb{E} \left[g \left(X, \theta_0 \right) I \left\{ q \left(X, \theta_0 \right) \leqslant u \right\} \right] + o_p(1)$$

Asymptotic Theory —— Shaikh et al. (2009)

Assumption 3.1
(1) Θ is a compact subset of $\mathbb{R}^k$;
(2) $q(X, \theta)$ has a continuous density $f(x, \theta)$ w.r.t. Lebesgue measure for all θ in a neighborhood of the true parameter θ_0, and $q(X, \theta)$ is Lipschitz continuous w.r.t θ.

Assumption 3.2
$K : \mathbb{R} \to \mathbb{R}$ is bounded, Lipschitz continuous, symmetric and satisfies: (a) $\int K(u)\mathrm{d}u = 1$; (b) $\int |K(u)|\mathrm{d}u < \infty$;
(c) $\int |uK(u)|\mathrm{d}u < \infty$;(d) $\int \left|u^2 K(u)\right| \mathrm{d}u < \infty$

Assumption 3.3 $\quad \left\|\hat{\theta}_n - \theta_0\right\| = O_p\left(n^{-1/2}\right)$

Assumption 3.4
The bandwidth sequence satisfies $0 < h = h_n \to 0$ and $nh^4 \to \infty$.

Asymptotic Theory —— Shaikh et al. (2009)

Theorem 1.
Suppose Assumptions 3.1–3.4 hold, under H_0, for test statistic (5), we have

$$T_n(h_n) \xrightarrow{d} N(0,1)$$

Asymptotic Theory —— Santa'Anna et al. (2019)

Assumption 4.1

(1) The parameter space Θ is a compact subset of $\mathbb{R}^k$

(2) The true parameter θ_0 belongs to the interior of Θ

(3) $\left\|\hat{\theta}_n - \theta_0\right\| = O_p\left(n^{-1/2}\right)$

Assumption 4.2

The parametric propensity score function $q(x, \theta)$ is twice continuously differentiable in Θ_0 for each $x \in \mathcal{X}$ with its first derivative $g(x, \theta)$ satisfying $\mathbb{E}\left[\sup_{\theta \in \Theta_0} \|g(X, \theta)\|\right] < \infty$ and its second derivative satisfying $\mathbb{E}\left[\sup_{\theta \in \Theta_0} \|\partial g(X, \theta)/\partial \theta\|\right] < \infty$. And the matrix $\Delta(\theta) \equiv \mathbb{E}\left[g(X, \theta) g'(X, \theta)\right]$ is nonsingular in Θ_0.

Assumption 4.3

The function $F_\theta(u) = \mathbb{P}(q(X, \theta) \leqslant u)$ satisfies $\sup_{u \in \Pi} |F_{\theta_1}(u) - F_{\theta_2}(u)| \leqslant C \|\theta_1 - \theta_2\|$, where C is a bounded positive number, not depending on θ_1 and θ_2.

Asymptotic Theory —— Santa'Anna et al. (2019)

Theorem 2. Let Assumptions 4.1-4.3 hold, under H_0, we have that

$$\sup_{u\in\Pi}\left|\hat{R}_n^p(u)-R_{n0}^p(u)\right|=o_p(1)\quad \text{and}\quad \hat{R}_n^p(u)\Rightarrow R_\infty^p$$

where $R_{n0}^p(u)=\frac{1}{\sqrt{n}}\sum_{i=1}^n\varepsilon_i\left(\theta_0\right)\mathcal{P}I\left\{q\left(X_i,\theta_0\right)\leqslant u\right\}$, and R_∞^p denotes a gaussian process with mean zero and covariance structure given by

$$K^p(u_1,u_2)=\mathbb{E}[q(X,\theta_0)(1-q(X,\theta_0))\mathcal{P}I\{q(X,\theta_0)\leqslant u_1\}\mathcal{P}I\{q(X,\theta_0)\leqslant u_2\}]$$

Corollary 1. Under assumptions of Theorem 1 and H_0, for any continuous functional $\Gamma(\cdot)$ from $\ell^\infty(\Pi)$ to $\mathbb{R}$, we have

$$\Gamma(\hat{R}_n^p)\xrightarrow{d}\Gamma(R_\infty^p)$$

Asymptotic Theory —— Santa'Anna et al. (2019)

[Computation of Critical Values]

We use multiplier bootstrap to compute the critical values. More precisely, we approximate the asymptotic behavior of $\hat{R}_n^p(u)$ by

$$\hat{R}_n^{p*}(u) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i \left(\hat{\theta}_n\right) \mathcal{P}_n I \left\{ q \left(X_i, \hat{\theta}_n \right) \leqslant u \right\} V_i$$

where $\{V_i\}_{i=1}^n$ is a sequence of i.i.d. random variables with zero mean, unit variance and bounded support, independent of the original sample $\{(D_i, X_i')'\}_{i=1}^n$. We use i.i.d. Bernoulli variates $\{V_i\}$ proposed by Mammen (1993):

$$V_i = \begin{cases} (1-\sqrt{5})/2, & p = (\sqrt{5}+1)/2\sqrt{5} \\ (\sqrt{5}+1)/2, & p = (\sqrt{5}-1)/2\sqrt{5} \end{cases}$$

Asymptotic Theory —— Santa'Anna et al. (2019)

With $\hat{R}_n^{p*}(u)$, the bootstrapped version of our test statistics (7) and (8) are given by

$$CvM_n^* = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{R}_n^{p*}\left(q\left(X_i,\hat{\theta}_n\right)\right)\right]^2$$

$$KS_n^* = \sup_{u\in\Pi}\left|\hat{R}_n^{p*}(u)\right|$$

The asymptotic critical values are then estimated by

$$c_{n,\alpha}^{\Gamma*} \equiv \inf\left\{c_\alpha \in [0,\infty) : \lim_{n\to\infty}\mathbb{P}_n^*\left\{\Gamma\left(\hat{R}_n^{p*}\right) > c_\alpha\right\} = \alpha\right\}$$

Theorem 2. Assume Assumptions 4.1–4.3 hold, then,

$$\hat{R}_n^{p*} \Rightarrow R_\infty^p \quad \text{a.s.}$$

Monte Carlo Simulation

We conduct a series of Monte Carlo experiments to study the finite sample properties of these two tests. Specifically, we compare Santa'Anna et al.'s (2019) Cramér-von Mises test CvM_n and Kolmogorov- Smirnov test KS_n to the Shaikh et al.'s (2009) test.

For Shaikh et al.'s (2009) test, we use the gaussian kernel $K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$, and the bandwidth h_n is chosen to be $cn^{-1/8}$ where c equal to 0.01, 0.05, 0.10 and 0.15.

We consider sample sizes n equal to 100, 200, 400, 600, 800 and 1000. For each design, we simulate 1000 Monte Carlo experiments.

Monte Carlo Simulation

Simulation 1:
We first consider cases where the number of covariates is relatively small.

DGP 1. $D^* = \frac{(X_1+X_2)}{3} - \varepsilon$

DGP 2. $D^* = -1 + \frac{(X_1+X_2+X_1X_2)}{3} - \varepsilon$

DGP 3. $D^* = -0.2 + \frac{\left(X_1^2-X_2^2\right)}{2} - \varepsilon$

DGP 4. $D^* = \frac{(0.1+X_1/3)}{\exp((X_1+X_2)/3)} - \varepsilon$

DGP 5. $D^* = \frac{(-0.8+(X_1+X_2+X_1X_2)/3)}{\exp(0.2+(X_1+X_2)/3)} - \varepsilon$

Monte Carlo Simulation

For each of these five DGPs, $D = I\{D^* > 0\}$, $\varepsilon \perp\!\!\!\perp (X_1, X_2)$, where $X_1 = Z_1$, $X_2 = (Z_1 + Z_2)/\sqrt{2}$, and Z_1, Z_2 and ε are independent standard normal random variables.

Let $X = (1, X_1, X_2)'$, the null hypothesis H_0 is:

$$H_0 : \exists \theta_0 = (\beta_0, \beta_1, \beta_2)' \in \Theta : \mathbb{E}\left[D | \Phi\left(X'\theta_0\right)\right] = \Phi\left(X'\theta_0\right) \text{ a.s.}$$

where $\Phi(\cdot)$ is the cumulative distribution function(CDF) of the standard normal distribution.

We estimate θ_0 by MLE.

Monte Carlo Simulation

dgp	n	CvM	KS	Shaikh(0.01)	Shaikh(0.05)	Shaikh(0.10)	Shaikh(0.15)
1	100	0.057	0.052	0.033	0.023	0.008	0.004
	200	0.054	0.058	0.054	0.02	0.011	0.003
	400	0.061	0.053	0.045	0.021	0.015	0.005
	600	0.061	0.057	0.052	0.022	0.015	0.009
	800	0.06	0.056	0.04	0.027	0.017	0.011
	1000	0.062	0.054	0.052	0.025	0.012	0.007
2	100	0.269	0.246	0.062	0.073	0.069	0.055
	200	0.602	0.526	0.091	0.19	0.211	0.191
	400	0.931	0.862	0.215	0.459	0.519	0.526
	600	0.987	0.961	0.318	0.668	0.751	0.763
	800	1	0.994	0.509	0.844	0.901	0.906
	1000	1	0.999	0.662	0.939	0.963	0.968
3	100	0.35	0.279	0.061	0.059	0.018	0
	200	0.544	0.464	0.165	0.132	0.046	0.011
	400	0.714	0.658	0.417	0.319	0.107	0.018
	600	0.778	0.735	0.579	0.378	0.147	0.038
	800	0.805	0.778	0.65	0.437	0.157	0.034
	1000	0.83	0.797	0.685	0.472	0.154	0.04
4	100	0.166	0.14	0.043	0.03	0.02	0.01
	200	0.353	0.282	0.061	0.065	0.05	0.031
	400	0.639	0.523	0.088	0.192	0.21	0.182
	600	0.853	0.77	0.138	0.348	0.425	0.399
	800	0.935	0.886	0.193	0.508	0.612	0.604
	1000	0.981	0.964	0.302	0.691	0.8	0.808
5	100	0.136	0.111	0.058	0.03	0.013	0.01
	200	0.182	0.16	0.046	0.039	0.036	0.028
	400	0.357	0.315	0.058	0.08	0.084	0.073
	600	0.503	0.456	0.08	0.136	0.155	0.145
	800	0.609	0.578	0.085	0.185	0.234	0.223
	1000	0.691	0.635	0.109	0.249	0.31	0.319

Monte Carlo Simulation

Simulation2:
We next extent to cases with 10 covariates.

DGP 6. $D^* = -\frac{\sum_{j=1}^{10} X_j}{6} - \varepsilon$

DGP 7. $D^* = -1 - \frac{\sum_{j=1}^{10} X_j}{10} + \frac{X_1 X_2}{2} - \varepsilon$

DGP 8. $D^* = -1 - \frac{\sum_{j=1}^{10} X_j}{10} + \frac{X_1 \sum_{k=2}^{5} X_k}{4} - \varepsilon$

DGP 9. $D^* = -1.5 - \frac{\sum_{j=1}^{10} X_j}{6} + \frac{\sum_{k=1}^{40} X_k^2}{10} - \varepsilon$

DGP 10. $D^* = \frac{-0.1+0.1 \sum_{j=1}^{5} X_j}{\exp\left(-0.2 \sum_{k=1}^{10} X_j\right)} - \varepsilon$

Monte Carlo Simulation

For each of these five DGPs, $X_1, X_2, ..., X_{10}$ are independent standard normal random variables. $D = I\{D^* > 0\}$. Let $X = (1, X_1, X_2, ..., X_{10})'$, $\varepsilon \perp\!\!\!\perp (X)$.

Let $X = (1, X_1, X_2, ..., X_{10})'$, the null hypothesis H_0 is:

$$H_0 : \exists \theta_0 = (\beta_0, \beta_1, \beta_2, ..., \beta_{10})' \in \Theta : \mathbb{E}[D|\Phi(X'\theta_0)] = \Phi(X'\theta_0) \text{ a.s.}$$

where $\Phi(\cdot)$ is the cumulative distribution function(CDF) of the standard normal distribution.

We estimate θ_0 by MLE.

Monte Carlo Simulation

dgp	n	CvM	KS	Shaikh(0.01)	Shaikh(0.05)	Shaikh(0.10)	Shaikh(0.15)
6	100	0.079	0.077	0.048	0.019	0.006	0.001
	200	0.07	0.066	0.043	0.018	0.011	0.007
	400	0.058	0.064	0.051	0.025	0.012	0.008
	600	0.056	0.059	0.046	0.024	0.007	0.004
	800	0.044	0.048	0.039	0.018	0.007	0.005
	1000	0.05	0.058	0.05	0.018	0.005	0.003
7	100	0.079	0.092	0.059	0.018	0.005	0.001
	200	0.102	0.104	0.066	0.014	0.004	0.001
	400	0.132	0.133	0.042	0.017	0.007	0.004
	600	0.163	0.163	0.044	0.025	0.016	0.008
	800	0.247	0.231	0.044	0.046	0.039	0.019
	1000	0.305	0.27	0.054	0.065	0.05	0.031
8	100	0.087	0.1	0.067	0.023	0.006	0.003
	200	0.135	0.131	0.039	0.024	0.008	0.003
	400	0.255	0.234	0.049	0.042	0.034	0.018
	600	0.42	0.372	0.071	0.099	0.088	0.05
	800	0.588	0.54	0.115	0.185	0.172	0.128
	1000	0.661	0.599	0.127	0.245	0.246	0.186
9	100	0.088	0.095	0.045	0.019	0.005	0.001
	200	0.127	0.121	0.042	0.016	0.012	0.005
	400	0.237	0.198	0.039	0.03	0.029	0.021
	600	0.375	0.31	0.059	0.046	0.044	0.034
	800	0.467	0.392	0.051	0.065	0.07	0.066
	1000	0.587	0.486	0.053	0.089	0.108	0.108
10	100	0.08	0.093	0.048	0.016	0.001	0
	200	0.077	0.082	0.037	0.021	0.009	0.003
	400	0.128	0.117	0.042	0.019	0.014	0.01
	600	0.221	0.166	0.061	0.034	0.03	0.014
	800	0.309	0.226	0.054	0.051	0.05	0.033
	1000	0.404	0.31	0.064	0.094	0.09	0.054

Reference

1. Escanciano, J.C., (2006). A consistent diagnostic test for regression models using projections. Econometric Theory 22, 1030–1051.
2. Hirano, K., Imbens, G.W., Ridder, G., 2003. Efficient estimation of average treatment effects using the estimated propensity score. Econometrica 71 (4), 1161–1189.
3. Pedro H.C. Sant'Anna, Xiaojun Song, (2019). Specification tests for the propensity score. J. Econometrics 210, 379–404.
4. Shaikh, A.M., Simonsen, M., Vytlacil, E.J., Yildiz, N., (2009). A specification test for the propensity score using its distribution conditional on participation. J. Econometrics 151 (1), 33–46.
5. Stute, W., (1997). Nonparametric model checks for regression. Ann. Statist. 25 (2), 613–641.
6. Zheng, J.X., (1996). A consistent test of functional form via nonparametric estimation techniques. J. Econometrics 75, 263–289.

The end.

Thanks for listening.