Microeconometrics I – Problem Set 04

Vinícius de Almeida Nery Ferreira

September 15th, 2024

1 Specification Tests for the Propensity Score

In this question, we will use the data and the methods of Sant'Anna and Song (2019). Very briefly, they develop a specification test for parametric propensity score estimation based on a particular restriction relating the distribution of treated and control groups that must hold if the model is correctly specified. In particular, they test the validity *Overlap* assumption:

$$0 < p(X) < 1 \quad \forall x \in \mathcal{X}, \tag{\textit{Overlap}}$$

where X is the vector of covariates, $\mathcal{X}$ is the covariate space and $p(X) \coloneqq P(D = 1|X = x)$ is the propensity score, with D being the binary treatment indicator.

They argue that the test's main benefits are the immunity to the "curse of dimensionality", the fact that it is fully data-driven, not requiring tuning parameter, and the ability to detect a broad class of local alternative hypothesis.

We reproduce below the main lemma[1] of their paper.

Lemma. *Let* $\alpha = \frac{\mathbb{P}(D=0)}{\mathbb{P}(D=1)}$ *and assume* $0 < \mathbb{P}(D = 1) < 1$*. If* $0 < p(X) < 1$ *almost surely (a.s.), then*

$$\mathbb{E}[1\{p(X) \leq u\} \mid D = 1] = \alpha \mathbb{E}\left[\frac{p(X)}{1 - p(X)} 1\{p(X) \leq u\} \mid D = 0\right], \quad \forall u \in [0, 1].$$

Furthermore, this holds if and only if

$$\mathbb{E}[(D - p(X))1\{p(X) \leq u\}] = 0, \quad \forall u \in [0, 1].$$

Therefore, the test they develop is based on the hypothesis

$$H_0 : \mathbb{E}[(D - q(X, \theta_0))1\{q(X, \theta_0) \leq u\}] = 0 \quad \text{for some } \theta_0 \in \Theta \quad \text{and for all } u \in \Pi$$

or equivalently

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

where $\Theta \subset \mathbb{R}^k$, $\Pi = [0, 1]$ is the unit interval, and $q(X, \theta)$ is a parametric specification for the propensity score, such as the probit or the logit.

To test this hypothesis, they propose test statistics based on continuous functionals, such as the Cramér-von-Mises (CvM) and Kolmogorov-Smirnov (KS) functionals. Critical values are constructed using a multiplier bootstrap process.

1. As they state in Remark 1, this result does not involve outcome data and therefore holds and is well motivated even when CIA does not hold, which may happen in decomposition exercises.

1.1 Reproducing (part of) Table 4

To reproduce the first columns of their Table 4, we use the paper's accompanying **R** package, `pstest`. We model the probability of being a WTO member in 1990 and 1995 as a parametric function of these covariates:

- X_1: real per capita GDP;
- X_2: land area per capita;
- X_3: polity index, which measures the degree of democracy or autocracy in the country.

We run two specifications:

$$\begin{aligned} &\text{Specification 1:} \quad D = \Phi\left(\theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3\right) \\ &\text{Specification 2:} \quad D = \Phi\left(\theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \theta_4 X_1 X_2 + \theta_5 X_1 X_3 + \theta_6 X_2 X_3\right) \end{aligned}$$

where $D = 1$ indicates being a WTO member and Φ is the standard normal CDF. Thus, our link function is a probit, which is the same as theirs.

In each of these specifications, we test the hypothesis

$$H_0 : \exists \theta_0 \in \Theta : \mathbb{E}\left[D - \Phi(\boldsymbol{X}'\theta_0) \mid \Phi(\boldsymbol{X}'\theta_0)\right] = 0 \quad \text{a.s.}$$

where $\boldsymbol{X}$ is the appropriate covariate vector of each specification.

The replication of their Table 1 is shown in our Table 1. For both specifications, the p-values, highlighted in bold, are constructed based on 10,000 bootstrap draws.

Table 1: Propensity Score Specification Test Results

	Specification 1	Specification 2
CvM_n Statistic	0.02	0.01
$\boldsymbol{CvM_n}$ P-Value	**0.01**	**0.29**
KS_n Statistic	0.30	0.28
$\boldsymbol{KS_n}$ P Value	**0.12**	**0.25**

1.2 Interpreting Table 4

When using the CvM statistic, we have evidence that the first specification is not correctly specified – in the sense that the overlap assumption does not hold –, as we have a p-value of 0.01. This evidence is weaker when using the KS statistic, which has a p-value of 0.12.

For both statistics, we don't have evidence to reject the null that Specification 2 is correctly specified, as both p-values are above 0.2. Therefore, it should be preferred over Specification 1.

2 Sensitivity Analysis

In this question, we will use the data and the methods of Cinelli and Hazlett (2020), which develop a sensitivity analysis framework based on an extended version of the usual omitted variable bias.

Cinelli and Hazlett (2020) are interested in the effect of being in an attacked village in Sudan's western region of Darfur on the support of peace, which they measure as an index. They argue that attacks were indiscriminate within a given village, with the exception that women suffered more than men. Thus, controlling for village and sex should be enough for the CIA to hold in this (tragic) context.

Their sensitivity analysis aims to measure how strong an unobservable confounder ought to be to completely erase the treatment effects. For better interpretability, they also display the results in terms of a baseline covariate, which is Female in our case. We defer further details to their (excellent) paper.

2.1 Reproducing Table 1

To reproduce their Table 1, we use the paper's accompanying **R** package, `sensemakr`. The specification we are using is the same as their Equation 1:

$$\text{PeaceIndex} = \hat{\tau}_{\text{res}}\text{DirectHarm} + \hat{\beta}_{f,\text{res}}\text{Female} + \text{Village}\hat{\boldsymbol{\beta}}_{v,\text{res}} + \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\text{res}} + \hat{\varepsilon}_{\text{res}},$$

"where PeaceIndex is an index measuring individual attitudes towards peace, DirectHarm is a dummy variable indicating whether an individual was reportedly injured or maimed during such an attack, Female is a fixed effect for being female and Village is a matrix of village fixed effects. Other pretreatment covariates are included through the matrix $\boldsymbol{X}$, such as age, whether they were a farmer, herder, merchant or trader, their household size and whether or not they voted in the past" (Cinelli and Hazlett 2020, p.42).

Their proposed minimal reporting is in Table 2, which reproduces their Table 1.

Table 2: Sensitivity Results

Treatment	Estimate	Standard Error	T Statistic	$R^2_{Y\sim D\|\boldsymbol{X}}$ (%)	RV (%)	$\text{RV}_{\alpha=.05}$ (%)
DirectHarm	0.097	0.023	4.184	2.2%	13.9%	7.6%

Notes: Standard errors assume homoskedasticity. Degrees of Freedom: 783.
Statistics of a cofounder as strong as Female: $R^2_{Y\sim Z|D,X} = 12.5\%$, $R^2_{D\sim Z|X} = 0.9\%$.

2.2 Interpreting Table 1

Taking their identification argument as credible, having suffered from violence in the village attacks increases support of peace by about 10% on their index. This estimate is significant, with a t-statistic of 4.18 when assuming homoskedasticity.

The Robustness Value (RV) is 13.87%. This means that unobservable confounders would have to explain 13.87% of the residual variance not explained by the other covariates of both the treatment and the outcome in order for the estimate of DirectHarm to go to zero. Note that this implies that any confounder that is "weaker" than this is not capable to completely erase the results.

However, we usually are not interested in the cases where confounders are strong enough to completely erase the effect we care about, but whether they can affect the significance of the result at usual significance levels (in our case, 5%). For that to happen, the unobservable confounders would have to explain about 7.6% of the residual variance of both the treatment and the outcome.

Finally, the partial R^2 of treatment with the outcome is 2.2% – after controlling for the other covariates, treatment explains 2.2% of the peace index variability. They show that this is also the percentage of the residual variance of treatment that an extreme confounder which explains 100% of the residual variance of the outcome would need to be responsible for in order to bring the effect measured by our estimator to zero.

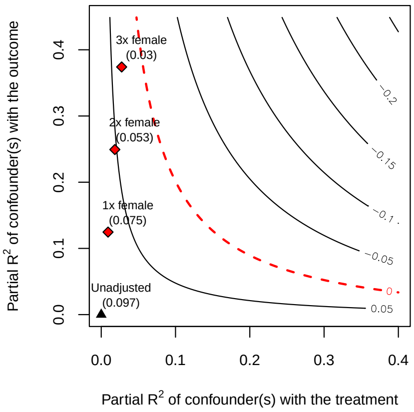
Note that, fundamentally, robustness to misspecification is determined by the share of variation of the outcome that the treatment uniquely explains.

The notes tells us what would be the sensitivity statistics of an unobservable confounder that is as strong as the Female indicator. This confounder would explain about 12.5% of the residual variance of the outcome, while only 0.9% of the residual variance of the treatment. Since the RV is higher than both statistics, it would not be enough to erase all results.

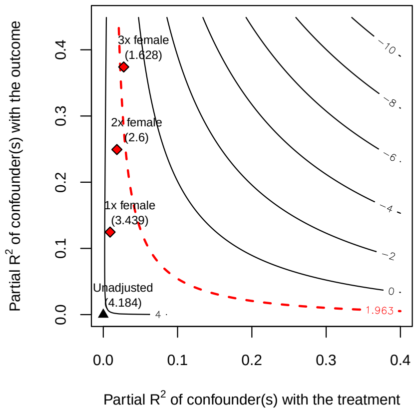
Furthermore, since 2.2% (the partial R^2) is higher than 0.9%, a extreme confounder that explains 100% of the residual variance of the outcome and that is as strong as female in terms of association to treatment would not be able to completely erase the results.

2.3 Reproducing Figure 2

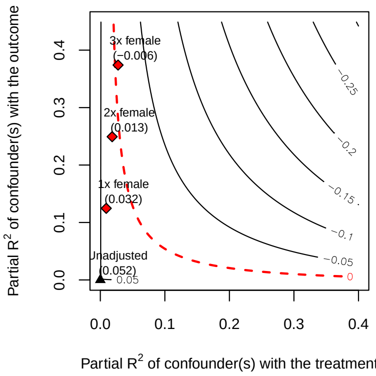
For convenience, we report Figure 2a and 2b in the same panel. We also display the contour plots for the lower and upper bounds of the confidence interval (CI).



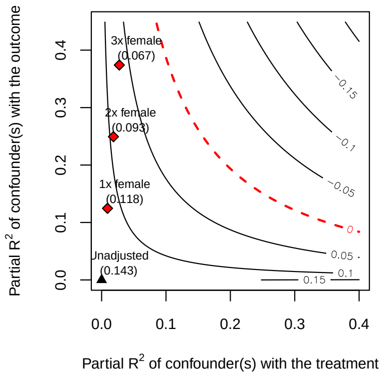
(a) Figure 2a: Contour Plot of the Point Estimate



(b) Figure 2b: Contour Plot of the T-Statistic



(c) Contour Plot of the Lower Bound of the 95% CI



(d) Contour Plot of the Upper Bound of the 95% CI

Figure 1: Contour Plots of Cinelli and Hazlett (2020)

2.4 Interpreting Figure 2

Figure 2a plots the contour lines of the point estimate. That is, what would be the point estimate of DirectHarm if confounders explained a given amount of the residual variance of treatment (horizontal axis) and of the outcome (vertical axis).

We see that not even a confounder three times as strong as Female would be enough to completely erase the estimated effect, as we don't cross the red dashed line that indicates that the estimated coefficient is 0. However, the point estimate could be reduced to as low as 0.03 from the original value of 0.097.

Figure 2b shows the contour lines of the t-statistic. That is, what would be the t-statistic of the coefficient of DirectHarm if confounders explained a given amount of the residual variance of treatment (horizontal axis) and of the outcome (vertical axis).

We see that not even a confounder two times as strong as female would be enough to overturn the significance of our result at the usual 5% level. However, a confounder three times as strong would reduce the point estimate and inflate the variance of the estimator in such a way that the estimate would no longer be significant at the 5% level, being marginally significant at the 10% level (which has a t-statistic of 1.64). This becomes more apparent when looking at the contour plot for the lower bound of the confidence interval, which crosses zero for a confounder three times as strong as Female.

Overall, results are relatively robust to misspecification, as we would need extremely strong confounders that are not observable in order to overturn the point estimate and the significance of the results. Of course, the existence of such a variable is impossible to verify, but the tests give some reassurance and steers the discussion in a more disciplined, quantitative direction.

References

Cinelli, C., and C. Hazlett. 2020. "Making sense of sensitivity: Extending omitted variable bias." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82 (1): 39–67. https://academic.oup.com/jrsssb/article/82/1/39/7056023.

Sant'Anna, P. H., and X. Song. 2019. "Specification tests for the propensity score." *Journal of Econometrics* 210 (2): 379–404. https://www.sciencedirect.com/science/article/pii/S0304407619300272.