Heterogeneous Treatment Effect Analysis in Stata

Ben Jann

University of Bern, ben.jann@soz.unibe.ch

Heterogeneous Treatment Effects Project Workshop
University of Michigan, Ann Arbor, June 19, 2015

Introduction

- Methods for causal inference from observational data have received much attention in the last two decades or so, especially in econometrics, but also in many other fields.
- Starting point of this literature is the Rubin Causal Model (a.k.a. Potential Outcomes Model a.k.a Counterfactual Causality).
- Assume a binary treatment variable D and let Y^1 and Y^0 be the potential outcomes with and without treatment, respectively. The treatment effect for individual i is then simply the difference between the potential outcomes, that is

$$\delta_i = Y_i^1 - Y_i^0$$

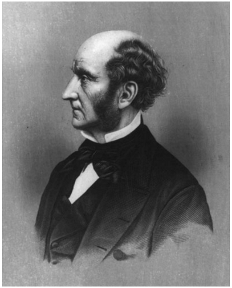
- The fundamental problem of causal inference, however, is that we can only observe Y_i^1 *or* Y_i^0. One of the potential outcomes must be counterfactual because what we observe is

$$Y_i = \begin{cases} Y_i^1 & \text{if } D_i = 1 \\ Y_i^0 & \text{if } D_i = 0 \end{cases}$$

Introduction

- The idea of defining causality in terms of potential outcomes is not new:

> *"Thus, if a person eats of a particular dish, and dies in consequence, that is, would not have died if he had not eaten of it, people would be apt to say that eating of that dish was the cause of his death." (Mill 2002[1843]: 214)*



John Stuart Mill (1806–1873)

Introduction

- A basic paradigm of the literature based on the potential outcomes model is that there can be individual heterogeneity in treatment effects, which stands in contrast to traditional regression modeling assuming constant structural parameters.
- The view that treatment effects can be heterogeneous led to new methods for causal inference and also to new uses and interpretations of existing methods (e.g. LATE interpretation of IV estimators, revival of matching and regression discontinuity designs).
- Surprisingly, however, not much attention is usually paid to the explicit analysis of the heterogeneity of treatment effects in applied studies.
- The basic quantity of interest is the average treatment effect (ATE)

$$ATE = E[\delta_i] = E[Y_i^1 - Y_i^0] = E[Y_i^1] - E[Y_i^0]$$

or sometimes the average treatment effect on the treated ($ATT = E[\delta_i|D_i = 1]$) or the average treatment effect on the untreated ($ATC = E[\delta_i|D_i = 0]$).

Introduction

- Why should we care about analyzing heterogeneous treatment effects?
- The naive estimator of the average treatment effect based on observational data can be decomposed as

$$
\begin{aligned}
NATE &= E[Y_i^1|D_i = 1] - E[Y_i^0|D_i = 0] \\
&= E[\delta_i] + \underbrace{E[Y_i^0|D_i = 1] - E[Y_i^0|D_i = 0]}_{\text{pre-treatment heterogeneity bias}} \\
&\quad + (1 - E[D_i])\underbrace{(E[\delta_i|D_i = 1] - E[\delta_i|D_i = 0])}_{\text{treatment-effect heterogeneity bias}}
\end{aligned}
$$

- The focus of most estimation approaches is to eliminate the first type of bias, but also the second type of bias might threaten the validity of causal inference.

Introduction

- For example, in the literature on economic returns to higher education various theories have been proposed that imply heterogeneous effects depending on the probability to go to college.
  - Human-capital theory in economics predicts *positive selection* into treatment, because people choose to go to college based on the expected economic returns. This is a widely accepted view.
  - More sociologically oriented literature suggests that college attendance is strongly influenced by social origin, which leads to *negative selection* into treatment under certain conditions.
- To evaluate these theories it is therefore crucial to analyze how treatment effects vary with treatment probability.
- Ultimately, believes about the mechanisms at play determine educational policy.

Analysis of Heterogeneous Treatment Effects in Stata

- To support the analysis of treatment-effect heterogeneity we developed a Stata command called `hte` (Xie, Brand, and Jann 2012).
- The approach of `hte` is to assume conditional unconfoundedness given a set of covariates and then analyze the treatment effect across the propensity score.
- Such an analysis can be revealing even if unconfoundedness does not hold.
- Three different algorithms are provided by `hte`:
  - The Stratification-Multilevel Method (SM)
  - The Matching-Smoothing Method (MS)
  - The Smoothing-Differencing Method (SD)

The Stratification-Multilevel Method (SM)

- The SM algorithm consists of four steps.
  1. Estimate the propensity score (i.e. the conditional probability to receive treatment) given the covariates (using `probit` or `logit`).
  2. Construct balanced propensity score strata.
     - ★ `hte` calls the `pscore` command for this purpose (Becker and Ichino 2002).
  3. Estimate strata-specific average treatment effects.
     - ★ In each stratum, a regression model on treatment is estimated, optionally including control variables to account for remaining covariate imbalance within strata.
  4. Estimate the trend of treatment effects across propensity score strata.
     - ★ `hte` regresses the strata-specific treatment effects on strata rank using variance weighted least squares (`vwls`; with the variance based on the standard errors of the strata specific treatment effects).

The Matching-Smoothing Method (MS)

- The MS algorithm consists of the following steps.
  1. Estimate the propensity score (i.e. the conditional probability to receive treatment) given the covariates (using `probit` or `logit`).
  2. Match treated units and control units with a matching algorithm based on the propensity score and compute a counterfactual outcome for each observation based on the matched observations from the other group.
  3. Plot the differences between observed and potential outcomes against the propensity score.
  4. Apply a nonparametric model such as local polynomial regression (Fan and Gijbels 1996) or lowess smoothing (Cleveland 1979) to the matched differences to yield a pattern of treatment effect heterogeneity across the propensity score.

The Smoothing-Differencing Method (SD)

- The SD algorithm works as follows.
  1. Estimate the propensity score (i.e. the conditional probability to receive treatment) given the covariates (using `probit` or `logit`).
  2. For each group (the control group and the treatment group) fit separate nonparametric regressions of the dependent variable on the propensity score.
  3. Obtain the pattern of treatment effect heterogeneity as a function of the propensity score by taking the take the difference in the nonparametric regression line between the treated and the untreated at different levels of the propensity score.

Example

- Effect of preschool childcare on the probability to go to high school in Germany (Fritschi and Jann 2009).
- Data: German Socio-Economic Panel (GSOEP)
- Variables:
  - Attended preschool childcare (the treatment)
  - Goes to high school (the outcome)
  - Gender
  - Birth cohort
  - Number of siblings
  - Education of parents
  - Household income
  - Mother's labor force participation
  - Migration background
  - East/west germany

The Stratification-Multilevel Method (SM)

```
. hte sm highschool childcare                                          ///
>     female cohort1991 cohort1992 cohort1993 cohort1994 cohort1995   ///
>     siblings1 siblings2 siblings3                                    ///
>     peduclow peduchigh lnhhinc motherlfp immigrant east              ///
>     , nograph
```

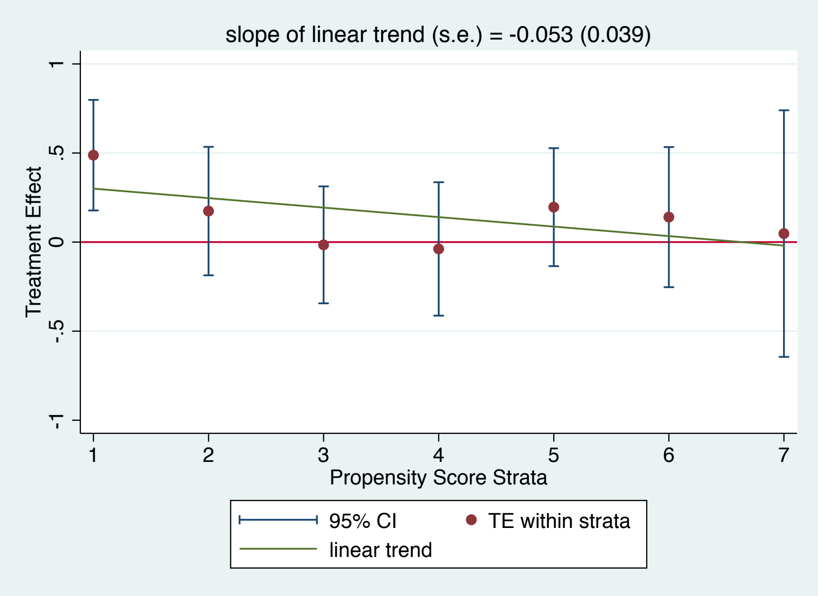
Number of obs = 594

highschool	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
TE by strata						
1	.4878277	.158317	3.08	0.002	.1775321	.7981234
2	.1740196	.1840194	0.95	0.344	-.1866517	.5346909
3	-.0155844	.1674786	-0.09	0.926	-.3438365	.3126677
4	-.0384615	.1910365	-0.20	0.840	-.4128862	.3359632
5	.1960784	.1689491	1.16	0.246	-.1350557	.5272126
6	.1401515	.2007209	0.70	0.485	-.2532543	.5335573
7	.047619	.3531523	0.13	0.893	-.6445468	.7397849
Linear trend						
_slope	-.0532744	.0388194	-1.37	0.170	-.1293591	.0228102
_cons	.3533125	.1521936	2.32	0.020	.0550185	.6516065

TE = treatment effect

```
. hte sm graph, yline(0)
```

The Stratification-Multilevel Method (SM)



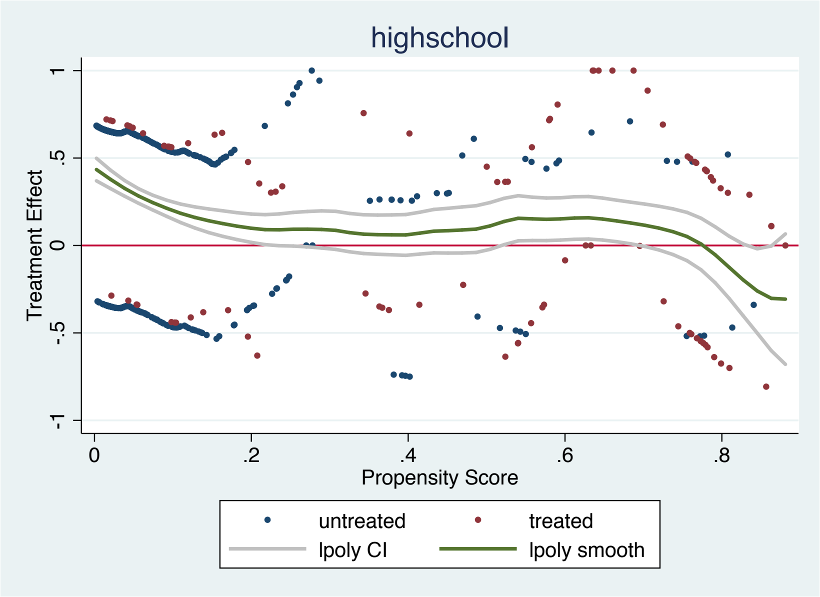
The Matching-Smoothing Method (MS)

```
. hte ms highschool childcare                                            ///
>     female cohort1991 cohort1992 cohort1993 cohort1994 cohort1995     ///
>     siblings1 siblings2 siblings3                                      ///
>     peduclow peduchigh lnhhinc motherlfp immigrant east                ///
>     , kernel lpolyci(degree(1) lw(*2) ciplot(rline)) yline(0)
(running psmatch2 ...)

Probit regression                               Number of obs     =        594
                                                LR chi2(15)       =     184.44
                                                Prob > chi2       =     0.0000
Log likelihood =  -165.5364                     Pseudo R2         =     0.3578
```

childcare	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
female	-.0505153	.1566255	-0.32	0.747	-.3574957	.2564651
cohort1991	-.1728172	.217302	-0.80	0.426	-.5987214	.253087
cohort1992	-.6398071	.2502713	-2.56	0.011	-1.13033	-.1492844
cohort1993	-.5347488	.249634	-2.14	0.032	-1.024023	-.0454751
cohort1994	-.6569049	.2735351	-2.40	0.016	-1.193024	-.120786
cohort1995	-.727918	.35289	-2.06	0.039	-1.41957	-.0362663
siblings1	.0277498	.1834415	0.15	0.880	-.331789	.3872885
siblings2	-.4883101	.2967347	-1.65	0.100	-1.069899	.0932792
siblings3	-.4729711	.4123559	-1.15	0.251	-1.281174	.3352316
peduclow	.0221446	.2546728	0.09	0.931	-.4770049	.5212941
peduchigh	.449848	.1978372	2.27	0.023	.0620942	.8376018
lnhhinc	-.1936814	.1942748	-1.00	0.319	-.5744529	.1870902
motherlfp	.3912866	.1703624	2.30	0.022	.0573825	.7251907
immigrant	-.1015982	.2362028	-0.43	0.667	-.5645473	.3613508
east	1.778842	.191471	9.29	0.000	1.403566	2.154118

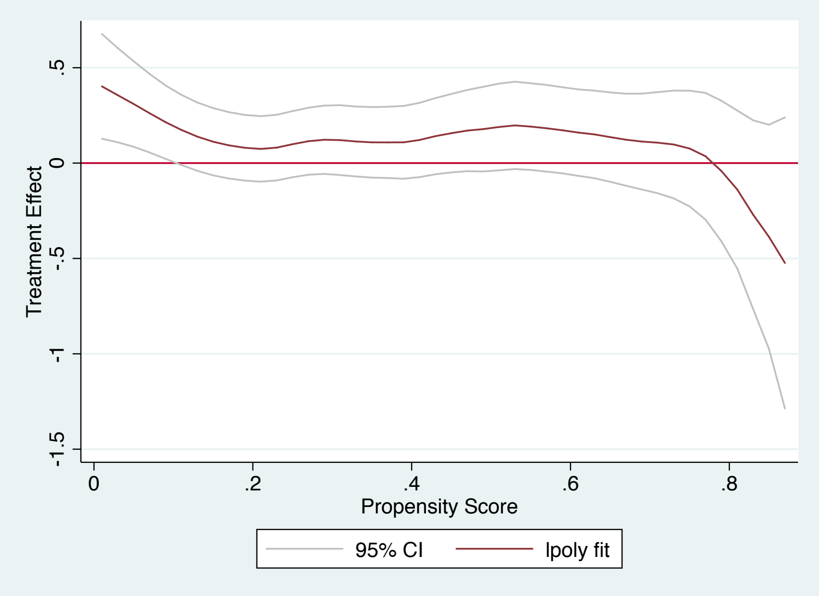
The Matching-Smoothing Method (MS)



The Smoothing-Differencing Method (SD)

```
. hte sd highschool childcare                                                ///
>     female cohort1991 cohort1992 cohort1993 cohort1994 cohort1995   ///
>     siblings1 siblings2 siblings3                                          ///
>     peduclow peduchigh lnhhinc motherlfp immigrant east                   ///
>     , ciopts(recast(rline)) yline(0)
```

The Smoothing-Differencing Method (SD)



Wishes and Grumbles

- `hte` has many limitations and various requests for additions have been made.
  - support for multiple imputation
  - Rosenbaum bounds or similar
  - nonlinear level-2 models in `hte sm`
  - user-provided propensity score in `hte sm` (requires reimplementation of `pscore`)
  - better control over construction PS strata and better returns (requires reimplementation of `pscore`)
  - support for sampling weights/complex surveys; better support for confidence interval estimation
  - better support for matching algorithms in `hte ms` (requires reimplementation of `psmatch2`)
  - using `hte` to generalize results from experimental data to a populatiion
  - ...

References

- Becker, S. O., A. Ichino (2002). Estimation of average treatment effects based on propensity scores. *The Stata Journal* 2:358–377.
- Cleveland, William S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association* 74:829–836.
- Fan, Jianqing and Irène Gijbels (1996). *Local Polynomial Modelling and Its Applications.* London: Chapman & Hall.
- Fritschi, Tobias and Ben Jann (2009). Zum Einfluss vorschulischer Kinderbetreuung auf den Bildungsweg und den erwarteten Erfolg am Arbeitsmarkt [The influence of preschool childcare on educational outcomes and expected labor market success]. *Empirische Pädagogik* 23:500–520.
- Mill, John Stuart (2002). *A System of Logic.* Reprinted from the 1981 edition (first published 1843). Honolulu, Hawaii: University Press of the Pacific.
- Xie, Yu, Jennie E. Brand, and Ben Jann (2012). Estimating Heterogeneous Treatment Effects with Observational Data. *Sociological Methodology* 42:314-347.