grmpy-semipar

A Local Instrumental Variables Approach to Estimating the Generalized Roy Model

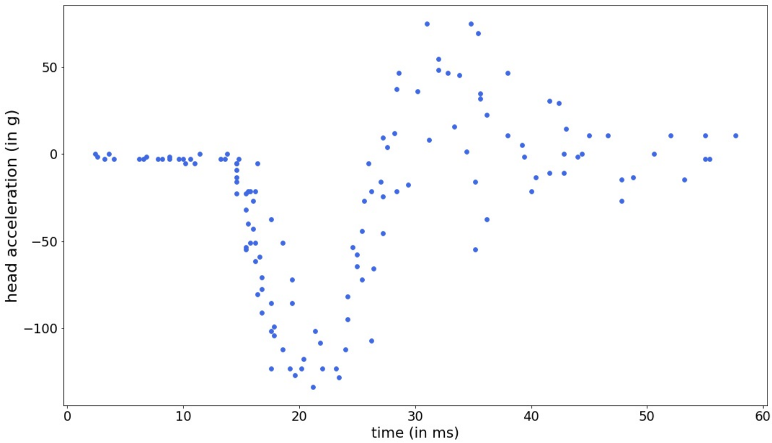
Sebastian Gsell

University of Bonn
sebastian.gsell@uni-bonn.de

Table of contents

locpoly

Heads up!



Preliminaries

1. Specifiy order of the polynomial p to be estimated; e.g.
   $p = 1$ results in a local linear,
   $p = 2$ in a local quadratic fit
2. Choose derivative v of estimator $\widehat{\beta}$ you want to obtain.
   Make sure that $v = p + 1$, where $\widehat{\beta} = (\widehat{\beta_0}, \ldots, \widehat{\beta_p})$
3. Determine length of the grid M over which the local polynomial fit shall be peformed ($M = 400$ is a good benchmark for most cases)
4. Pick a *bandwidth*

For a detailed description of the *locpoly* function , go to:
`https://github.com/segsell/grmpy-semipar/tree/master/semipar/KernReg`

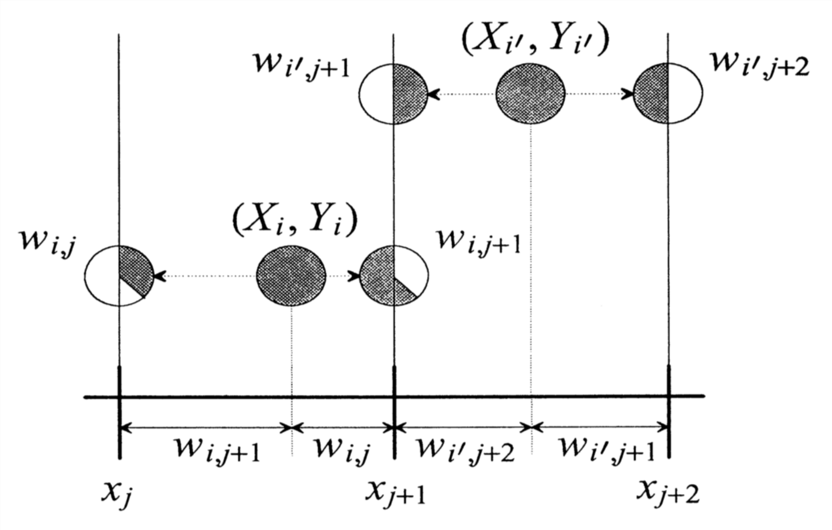
Strategy

1. Bin the data, yields *bin counts* for X and y
2. Compute *kernel weights* (here, Gaussian kernel): W
3. Combine *bin counts* and *kernel weights* yielding $\mathsf{X}'\mathsf{W}\,\mathsf{X}$, $\mathsf{X}'\mathsf{W}\,\mathsf{y}$
4. Solve the locally weighted least-squares regression problem at each point in the grid:

$$\widehat{\beta} = (\mathsf{X}'\mathsf{W}\,\mathsf{X})^{-1}\,\mathsf{X}'\mathsf{W}\,\mathsf{y}$$

   where $\widehat{\beta} = (\widehat{\beta_0}, \ldots, \widehat{\beta_p})$

Linear binning procedure



Source: Fan, J., Marron, S. (1994), "Fast Implementation of Nonparametric Curve Estimators", *J. Comp. Graph. Statist.*, **3**, 35-56.

The Generalized Roy Model

Framework

Potential Outcomes

$$Y_1 = \beta_1 X + U_1$$

$$Y_0 = \beta_0 X + U_0$$

Observed Outcome

$$Y = DY_1 + (1 - D)Y_0$$

Choice

$$I = Z\gamma - V$$

$$D_i = \begin{cases} 1 & if\ I > 0 \\ 0 & if\ I \leq 0 \end{cases}$$

The Marginal Treatment Effect (MTE)

$$
\begin{aligned}
MTE(\bar{x}, u_D) &\equiv E(Y_1 - Y_0 | X = \bar{x}, U_D = u_D) \\
&= \underbrace{\bar{x}(\beta_1 - \beta_0)}_{\textit{heterogeneity in observables}} + \underbrace{E(U_1 - U_0 | U_D = u_D)}_{k(u):\ \textit{heterogeneity in unobservables}} \qquad (1)
\end{aligned}
$$

where $u_D = \Phi(V)$

Parametric Estimation

Assumptions

Distributional Characteristics

$$\{U_1, U_0, V\} \sim \mathcal{N}(0, \Sigma) \qquad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{1,0} & \sigma_{1,V} \\ \sigma_{1,0} & \sigma_0^2 & \sigma_{0,V} \\ \sigma_{1,V} & \sigma_{0,V} & \sigma_V^2 \end{bmatrix}$$

$$\begin{aligned} MTE(\bar{x}, u_D) &\equiv E(Y_1 - Y_0 | X = \bar{x}, U_D = u_D) \\ &= \underbrace{\bar{x}(\beta_1 - \beta_0)}_{\text{heterogeneity in observables}} + \underbrace{(\sigma_{1,V} - \sigma_{0,V})\Phi^{-1}(u_D)}_{k(u)\text{: heterogeneity in unobservables}} \end{aligned} \tag{2}$$

Estimate parameters $(\beta_1, \beta_0, \sigma_{1,V}, \sigma_{0,V})$ via Maximum Likelihood and plug into (2).

Semiparametric Design

Assumptions

(X, Z) is independent of $\{U_1, U_0, V\}$:

1) Shape of *MTE* is independent of X,

2) *MTE* is identified over the common support of $P(Z)$, unconditional on X.

The Local Instrumental Variables *(LIV)* Estimator

$$E(Y|X=\bar{x}, P(Z)=p) = \bar{x}\beta_0 + \bar{x}(\beta_1-\beta_0)p + E[U_1-U_0|U_D \leq p]p \quad (3)$$

$$\Delta^{LIV}(\bar{x}, u_D) = \left.\frac{E(Y|X=\bar{x}, P(Z)=p)}{\partial p}\right|_{p=u_D}$$

$$= \bar{x}(\beta_1-\beta_0) + \left.\frac{[E(U_1-U_0|U_D \leq p]p}{\partial p}\right|_{p=u_D} \quad (4)$$

$$= \underbrace{\bar{x}(\beta_1-\beta_0)}_{\substack{\text{observable}\\ \text{component}}} + \underbrace{E(U_1-U_0|U_D=u_D)}_{\substack{k(u):\ \text{unobservable}\\ \text{components}}} = MTE(\bar{x}, u_D)$$

Observable component $\bar{x}(\beta_1 - \beta_0)$

1. Estimate *Treatment Propensity* $P(z)$
2. Define *Common Support* and trim the sample
3. Multiply each regressor in X with $\widehat{P}(z)$
4. Fit local linear regressions of X, $X \times p$ and Y on $\widehat{P}(z)$ and compute residuals
5. Estimate $\beta_0,\ \beta_1 - \beta_0$ by running OLS:

$$e_Y = e_X\beta_0 \ + \ e_{X \times p}(\beta_1 - \beta_0) \ + \ \epsilon$$

6. Multiply matrices $\bar{x}$ and $(\widehat{\beta_1 - \beta_0})$

Unobservable components $k(u)$

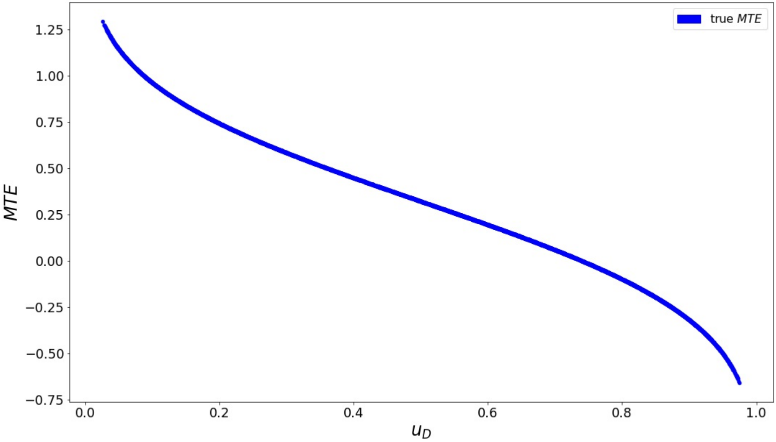
1. Compute the unobserved part of Y

$$\tilde{Y} = Y - X'\widehat{\beta_0} - X'(\widehat{\beta_1 - \beta_0})p$$

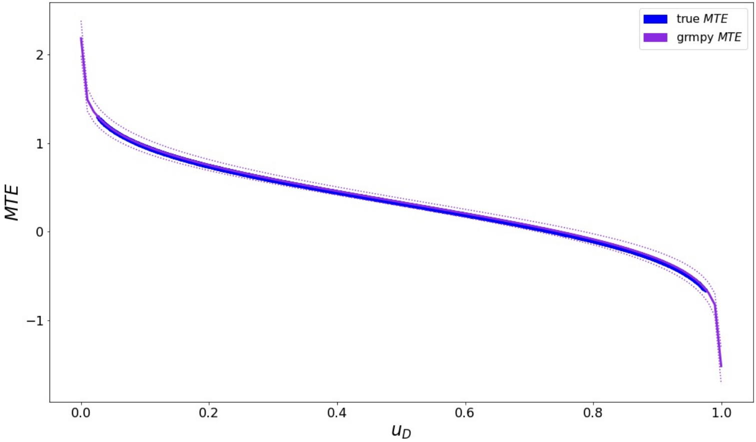
2. Estimate $k(u)$ through a locally quadratic regression of $\tilde{Y}$ on $\widehat{P}(z)$

Comparison

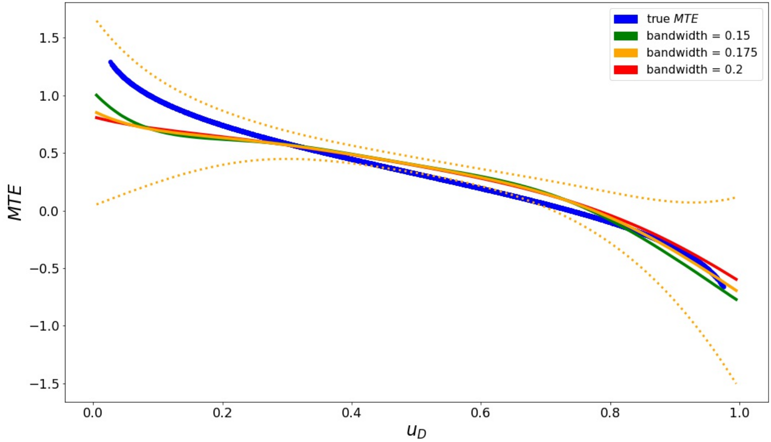
$\{U_1, U_0, V\}$ are normally distributed



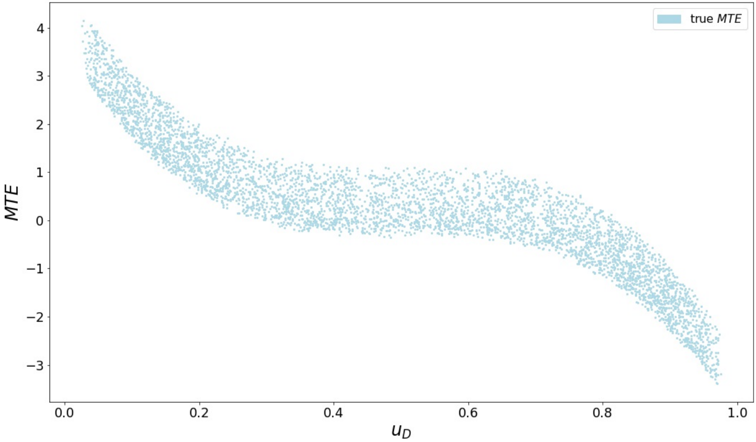
grmpy results



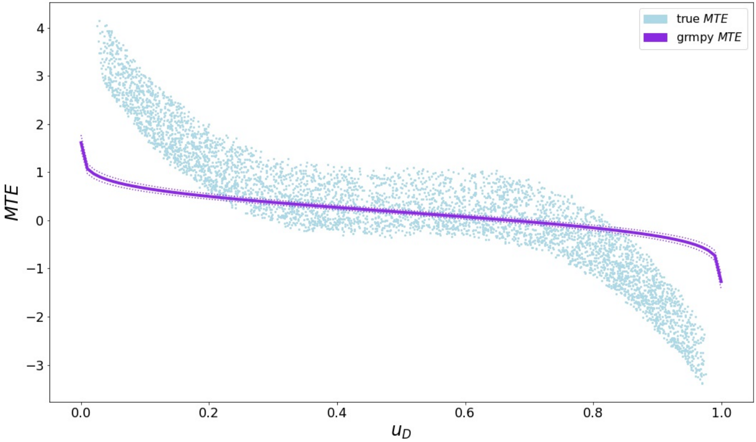
grmpy-semipar results

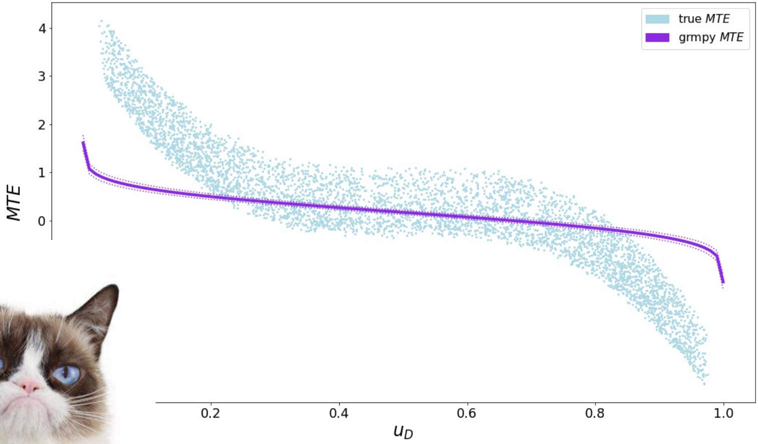


$\{U_1, U_0, V\}$ are non-normally distributed

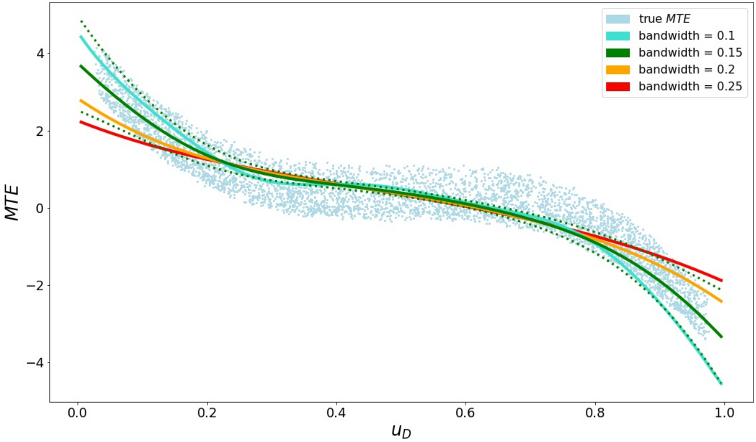


grmpy results




true MTE
grmpy MTE
MTE
4
3
2
1
0
0.2
0.4
0.6
0.8
1.0
u_D

grmpy-semipar results



Thank you!