grmpy Documentation

Release 1.0

Development Team

Apr 14, 2021

CONTENTS

PyPI | GitHub | Issues

`grmpy` is an open-source package for the simulation and estimation of the generalized Roy model. It serves as a teaching tool to promote the conceptual framework of the generalized Roy model, illustrate a variety of issues in the econometrics of policy evaluation, and showcase basic software engineering practices.

We build on the following main references:

> James J. Heckman and Edward J. Vytlacil. Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation. In *Handbook of Econometrics*, volume 6B, chapter 70, pages 4779–4874. Elsevier Science, 2007.
>
> James J. Heckman and Edward J. Vytlacil. Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. In *Handbook of Econometrics*, volume 6B, chapter 71, pages 4875–5143. Elsevier Science, 2007.
>
> Jaap H. Abbring and James J. Heckman. Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. *Handbook of Econometrics*, volume 6B, chapter 72, pages 5145-5303. Elsevier Science, 2007.

The remainder of this documentation is structured as follows. We first present the basic economic model and provide installation instructions. We then illustrate the basic use case of the package in a tutorial and showcase some evidence regarding its reliability. In addition we provide some information on the software engineering tools that are used for transparency and dependability purposes. The documentation concludes with further information on contributing, contact details as well as a listing of the latest releases.

The package is used as a teaching tool for a course on the analysis human capital at the University of Bonn. The affiliated lecture material is available on GitHub .

CHAPTER

ONE

ECONOMICS

This section provides a general discussion of the generalized Roy model and selected issues in the econometrics of policy evaluation.

1.1 Generalized Roy Model

The generalized Roy model ([20], [30]) provides a coherent framework to explore the econometrics of policy evaluation. Its is characterized by the following set of equations.

Potential Outcomes

$$Y_1 = \mu_1(X) + U_1$$
$$Y_0 = \mu_0(X) + U_0$$

Choice

$$D^* = \mu_D(Z) - V$$
$$D = I[D^* > 0]$$
$$B = E[Y_1 - Y_0 \mid \mathcal{I}]$$

Observed Outcome

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

(Y_1, Y_0) are objective outcomes associated with each potential treatment state D and realized after the treatment decision. Y_1 refers to the outcome in the treated state and Y_0 in the untreated state. D^* denotes the latent tendency for treatment participation. It includes any subjective benefits, e.g. job amenities, and costs, e.g. tuition costs. Agents take up treatment D if their latent tendency D^* is positive. $\mathcal{I}$ denotes the agent's information set at the time of the participation decision. The observed outcome Y is determined in a switching-regime fashion ([27], [28]). If agents take up treatment, then the observed outcome Y corresponds to the outcome in the presence of treatment Y_1. Otherwise, Y_0 is observed. The unobserved potential outcome is referred to as the counterfactual outcome. If costs are identically zero for all agents, there are no observed regressors, and $(U_1, U_0) \sim N(0, \Sigma)$, then the generalized Roy model corresponds to the original Roy model ([30]).

From the perspective of the econometrician, (X, Z) are observable while (U_1, U_0, V) are not. X are the observed determinants of potential outcomes (Y_1, Y_0), and Z are the observed determinants of the latent indicator variable D^*. The Potential outcomes as well as the latent indicator D^* are decomposed into their means $(\mu_1(X), \mu_0(X), \mu_D(Z))$ and their deviations from the mean (U_1, U_0, V). (X, Z) might have common elements. Observables and unobservables jointly determine program participation D.

If their ex ante latent indicator for participation D^* is positive, then agents select into treatment. Yet, this does not require their expected objective benefit B to be positive as well. Note that the unobservable term V enters D^* with a negative sign. Therefore conditional on Z , high values of V indicate a lower propensity for selecting into treatment and vice versa.

The evaluation problem arises because either Y_1 or Y_0 is observed. Thus, the effect of treatment cannot be determined on an individual level. If the treatment choice D depends on the potential outcomes, then there is also a selection problem. If that is the case, then the treated and untreated differ not only in their treatment status but in other characteristics as well. A naive comparison of the treated and untreated leads to misleading conclusions. Jointly, the evaluation and selection problem are the two fundamental problems of causal inference ([23]). Using the setup of the generalized Roy model, we now highlight several important concepts in the economics and econometrics of policy evaluation. We discuss sources of agent heterogeneity and motivate alternative objects of interest.

1.2 Agent Heterogeneity

What gives rise to variation in choices and outcomes among, from the econometrician's perspective, otherwise observationally identical agents? This is the central question in all econometric policy analyses ([6], [13]).

The individual benefit of treatment is defined as

$$B = Y_1 - Y_0 = (\mu_1(X) - \mu_0(X)) + (U_1 - U_0).$$

From the perspective of the econometrician, differences in benefits are the result of variation in observable X and unobservable characteristics $(U_1 - U_0)$. However, $(U_1 - U_0)$ might be (at least partly) included in the agent's information set $\mathcal{I}$ and thus known to the agent at the time of the treatment decision. Therefore we are able to distinguish between observable and unobservable heterogeneity. Observable heterogeneity is reflected by by the difference $\mu_1(X) - \mu_0(X)$. t denotes the differences between individuals that are based on differences of observable individual specific characteristics captured by X. Since we are able to take observable heterogeneity into account by conditioning on X this kind of heterogeneity is a negligible problem for the evaluation of policy interventions. Therefore all following concepts condition on X.

Consequently the second type of heterogeneity is represented by the differences between individuals captured by $U_1 - U_0$. This differences are unobservable from the perspective of an econo-

metrician. It should be noted that the term *unobservable* does not implicate that U 1 and U 0 are not completely or at least partly included in an individual's information set. As a result, unobservable treatment effect heterogeneity can be distinguished into private information and uncertainty. Private information is only known to the agent but not the econometrician; uncertainty refers to variability that is unpredictable by both.

The information available to the econometrician and the agent determines the set of valid estimation approaches for the evaluation of a policy. The concept of essential heterogeneity emphasizes this point ([18]). If agents select their treatment status based on benefits unobserved by the econometrician (selection on unobservables), then there is no unique effect of a treatment or a policy even after conditioning on observable characteristics. In terms of the Roy model this is characterized by the following condition

$$Y_1, Y_0 \not\perp\!\!\!\perp D$$

Average benefits are different from marginal benefits, and different policies select individuals at different margins. Conventional econometric methods that only account for selection on observables, like matching ([8], [15], [29]), are not able to identify any parameter of interest ([18], [20]). For example, [7] present evidence on agents selecting their level of education based on their unobservable gains and demonstrate the importance of adjusting the estimation strategy to allow for this fact. [16] propose a variety of tests for the the presence of essential heterogeneity.

1.3 Objects of Interest

Treatment effect heterogeneity requires to be precise about the effect being discussed. There is no single effect of neither a policy nor a treatment. For each specific policy question, the object of interest must be carefully defined ([20], [21], [22]). We present several potential objects of interest and discuss what question they are suited to answer. We start with the average effect parameters. However, these neglect possible effect heterogeneity. Therefore, we explore their distributional counterparts as well.

1.3.1 Conventional Average Treatment Effects

It is common to summarize the average benefits of treatment for different subsets of the population. In general, the focus is on the average effect in the whole population, the average treatment effect B^{ATE}, or the average effect on the treated B^{TT} or untreated B^{TUT}.

$$\begin{aligned} B^{ATE} &= E[Y_1 - Y_0] \\ B^{TT} &= E[Y_1 - Y_0 | D = 1] \\ B^{TUT} &= E[Y_1 - Y_0 | D = 0] \end{aligned}$$

All average effect parameter possibly hide considerable treatment effect heterogeneity. The relationship between these parameters depends on the assignment mechanism that matches agents to treatment. If agents select their treatment status based on their own benefits, like in the presence of essential heterogeneity, then agents that take up treatment benefit more than those that do not and thus $B^{TT} > B^{ATE}$. If agents select their treatment status at random, which is equivalent with the absence of essential heterogeneity, then all parameters are equal. Figure 1 illustrates an example for both cases. Both graphs show the distribution of benefits which is characterized by the difference of potential outcomes $Y_1 - Y_0$. Additionally the figures shows the conventional effects for both setups whereupon the selection process on the left side is affected by essential heterogeneity whereas the right side displays the effects in the absence of essential heterogeneity.

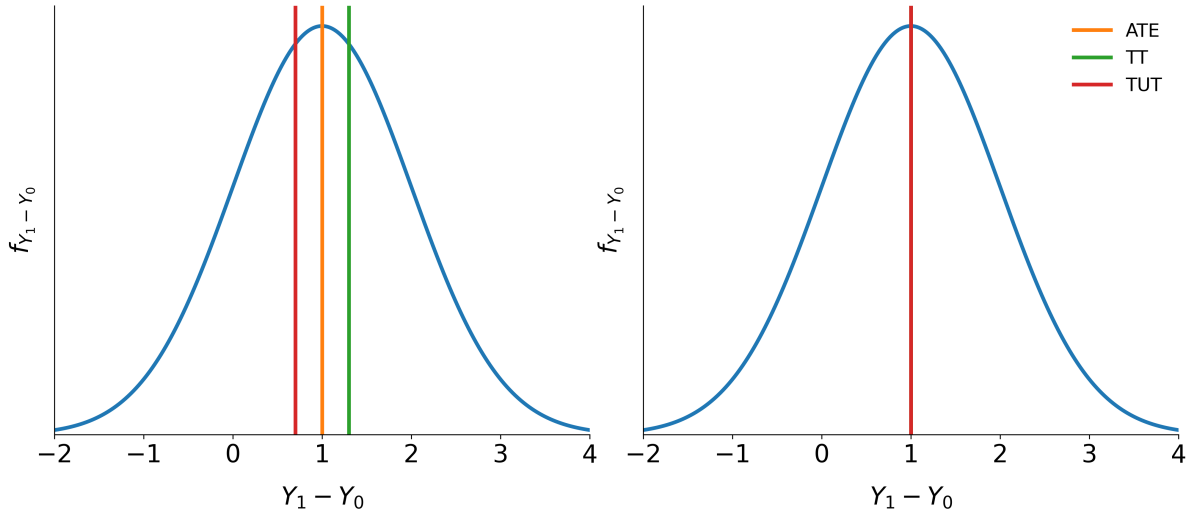


Fig. 1: **Fig. 1:** Conventional treatment effects with and without essential heterogeneity

The policy relevance of the conventional treatment effect parameters is limited in the presence of essential heterogeneity. They are only informative about extreme policy alternatives. The B^{ATE} is of interest to policy makers if they weigh the possibility of moving a full economy from a baseline to an alternative state or are able to assign agents to treatment at random. The B^{TT} is informative if the complete elimination of a program already in place is considered. Conversely, if the same program is examined for compulsory participation, then the B^{TUT} is the policy relevant parameter.

To ensure a tight link between the posed policy question and the parameter of interest, [19] propose the policy-relevant treatment effect B^{PRTE}. They consider policies that do not change potential outcomes, but only affect individual choices. Thus, they account for voluntary program participation.

1.3.2 Policy-Relevant Average Treatment Effect

The B^{PRTE} captures the average change in outcomes per net person shifted by a change from a baseline state B to an alternative policy A. Let D_B and D_A denote the choice taken under the baseline and the alternative policy regime respectively. Then, observed outcomes are determined as

$$Y_B = D_B Y_1 + (1 - D_B) Y_0$$
$$Y_A = D_A Y_1 + (1 - D_A) Y_0.$$

A policy change induces some agents to change their treatment status $(D_B \neq D_A)$, while others are unaffected. More formally, the B^{PRTE} is then defined as

$$B^{PRTE} = \frac{1}{E[D_A] - E[D_B]} (E[Y_A] - E[Y_B]).$$

As an example consider that policy makers want to increase the overall level of education. Rather than directly assigning individuals a certain level of education, policy makers can only indirectly affect schooling choices, e.g. by altering tuition cost through subsidies. The individuals drawn into treatment by such a policy will neither be a random sample of the whole population, nor the whole population of the previously (un-)treated. Therefore the implementation of conventional effects run the risk of being biased, whereas the B^{PRTE} is able to evaluate the average change in outcomes per net individual that is shifted into treated.

1.3.3 Local Average Treatment Effect

The Local Average Treatment Effect B^{LATE} was introduced by [26]. They show that instrumental variable approaches (IV) identify B^{LATE}, which measures the mean gross return to treatment for individuals induced into treatment by a change in an instrument.

Unfortunately, the people induced to go into treatment by a change in any particular instrument need not to be the same as the people induced to to select into treatment by policy changes other than those corresponding exactly to the variation in the instrument. A desired policy effect may be directly correspond to the variation in the instrument. Moreover, if there is a vector of instruments that generates choice and the components of the vector are intercorrelated, IV estimates using the components of Z as the instruments, one at a time, do not, in general, identify the policy effect corresponding to varying that instruments, keeping all other instruments fixed, the ceteris paribus effect of the change in the instrument. [14] develop this argument in detail.

The average effect of a policy and the average effect of a treatment are linked by the marginal treatment effect $\left(B^{MTE}\right)$. The B^{MTE} was introduced into the literature by [4] and extended by [19], [20] and [21].

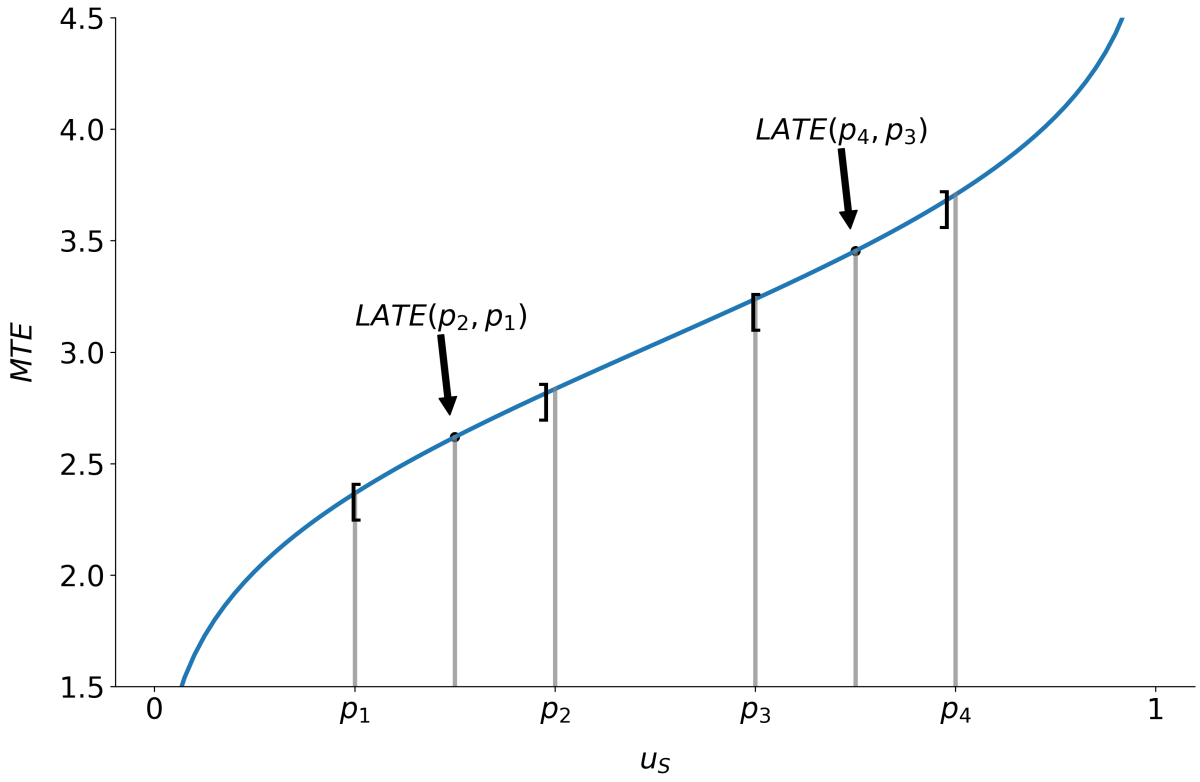


Fig. 2: **Fig. 2:** B^{LATE} at different values of u_S

1.3.4 Marginal Treatment Effect

The B^{MTE} is the treatment effect parameter that conditions on the unobserved desire to select into treatment. Let V be the heterogeneity effect that impacts the propensity for treatment participation and let $U_S = F_V(V)$. Then, the B^{MTE} is defined as

$$B^{MTE}(u_S) = E[Y_1 - Y_0|U_S = u_S].$$

The B^{MTE} is the average benefit for persons with observable characteristics $X = x$ and unobservables $U_S = u_S$. By construction, U_S denotes the different quantiles of V . So, when varying U_S but keeping X fixed, then the B^{MTE} shows how the average benefit varies along the distribution of V . For U_S evaluation points close to zero, the B^{MTE} is the average effect of treatment for individuals with a value of V that makes them most likely to participate. The opposite is true for high values of U_S. The B^{MTE} provides the underlying structure for all average effect parameters previously discussed. These can be derived as weighted averages of the B^{MTE} ([20]).

Parameter $j, \Delta j(x)$, can be written as

$$\Delta j = \int_0^1 B^{MTE}(u_S)\omega^j(u_S)du_S,$$

where the weights $\omega^j(u_S)$ are specific to parameter j, integrate to one, and can be constructed from data. For instance figure 3 shows weights for the B^{ATE}, B^{TT} and the B^{TUT} as well as the corresponding B^{MTE}. Contrary, figure 2 emphasizes that the concept of B^{LATE} is closely related to the idea of B^{MTE}. It illustrates that B^{LATE} evaluates B^{MTE} along a particular interval of the distribution of the unobservable Variable V. The specific range depends on the chosen instrument.

All parameters are identical only in the absence of essential heterogeneity. Then, the $B^{MTE}(x, u_S)$ is constant across the whole distribution of V as agents do not select their treatment status based on their unobservable benefits. This can be seen in figure 4 which illustrates B^{MTE} in the absence of essential heterogeneity, represented by the dotted orange line as well as an example for the B^{MTE} in the presence of essential heterogeneity portrayed by the blue graph.

So far, we have only discussed average effect parameters. However, these conceal possible treatment effect heterogeneity, which provides important information about a treatment. Hence, we now present their distributional counterparts ([1]).

1.4 Distribution of Potential Outcomes

Several interesting aspects of policies cannot be evaluated without knowing the joint distribution of potential outcomes ([2], [17]). The joint distribution of (Y_1, Y_0) allows to calculate the whole distribution of benefits. Based on it, the average treatment and policy effects can be constructed just as the median and all other quantiles. In addition, the portion of people that benefit from treatment can be calculated for the overall population $Pr(Y_1 - Y_0 > 0)$ or among any subgroup of particular interest to policy makers $Pr(Y_1 - Y_0 > 0|X)$. This is important as a treatment which

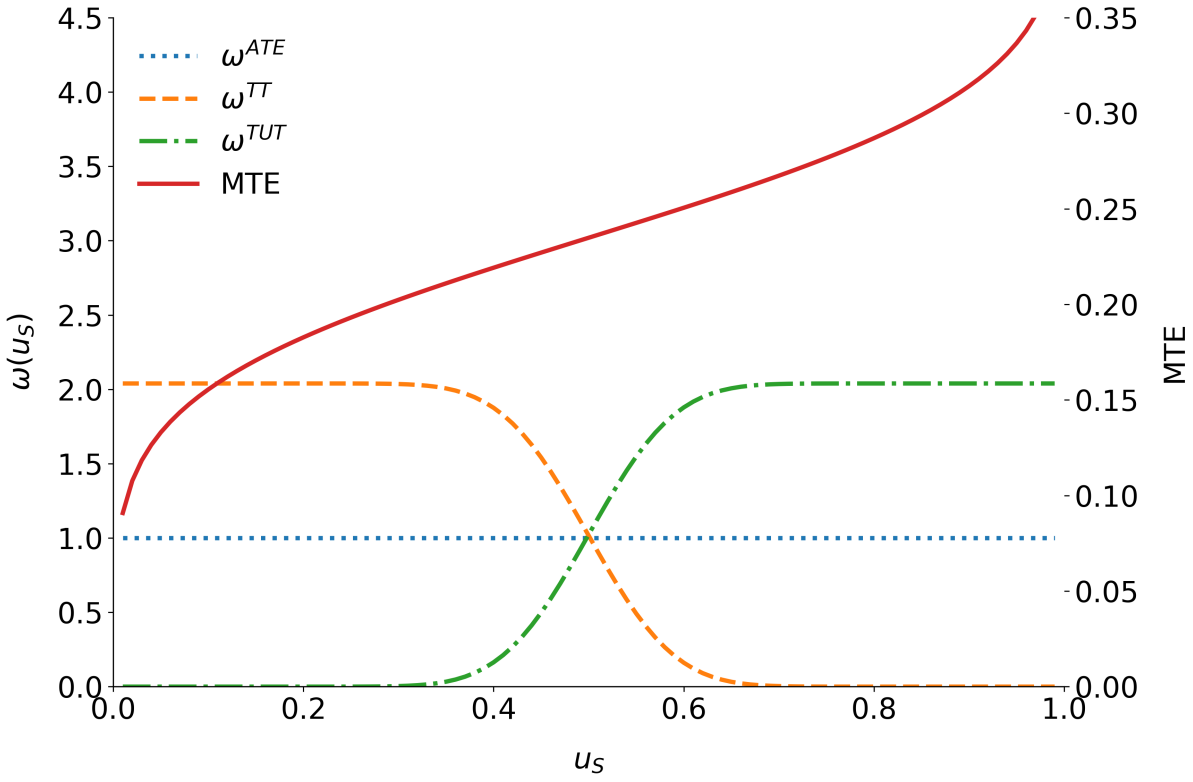


Fig. 3: **Fig. 3:** Weights for the marginal treatment effect for different parameters.

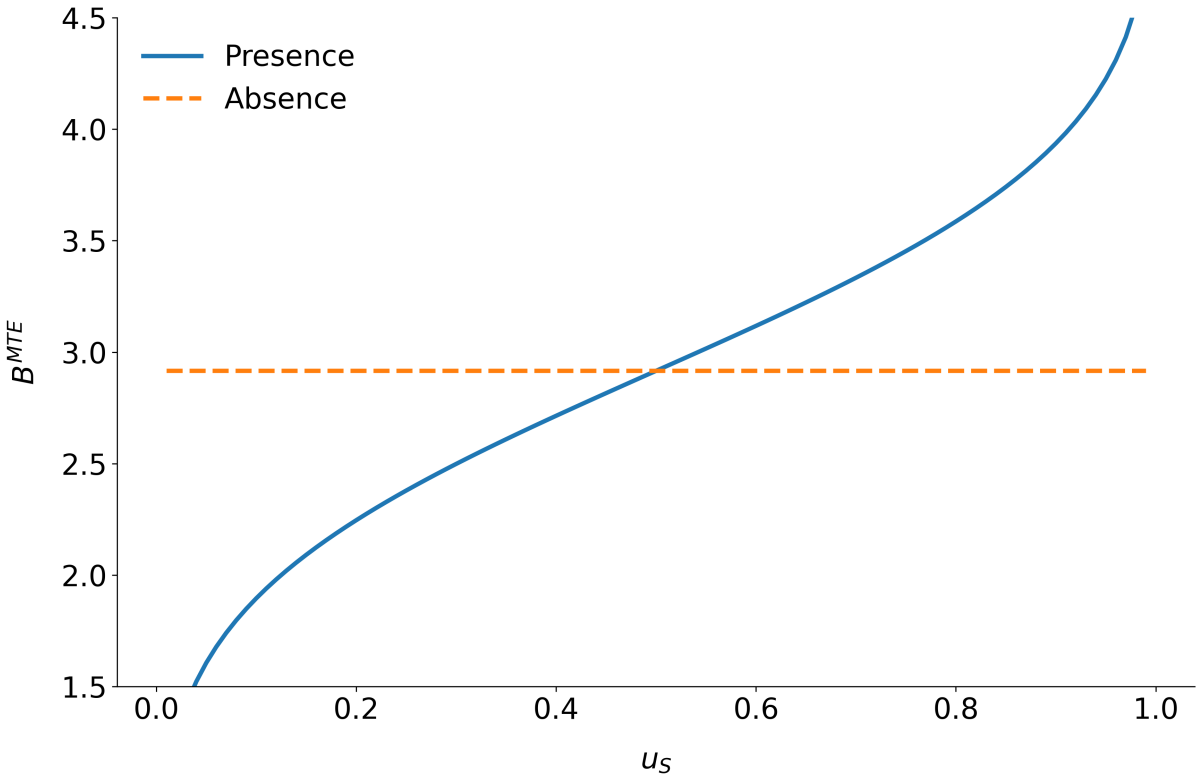


Fig. 4: **Fig 4:** B^{MTE} in the presence and absence of essential heterogeneity.

is beneficial for agents on average can still be harmful for some. For a comprehensive overview on related work see [2] and the work they cite. The survey by [12] provides an overview about the alternative approaches to the construction of conterfactual observed outcome distributions. See [3], [31] and [32] for their studies of quantile treatment effects.

The zero of an average effect might be the result of part of the population having a positive effect, which is just offset by a negative effect on the rest of the population. This kind of treatment effect heterogeneity is informative as it provides the starting point for an adaptive research strategy that tries to understand the driving force behind these differences ([24], [25]).

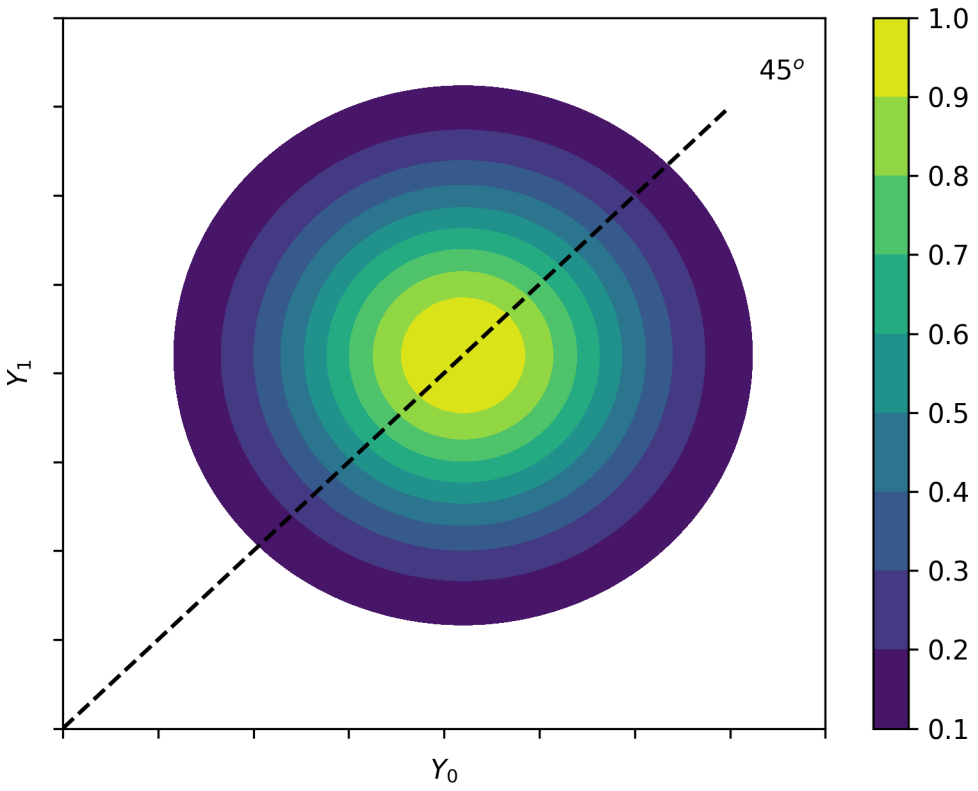


Fig. 5: **Fig 5:** Distribution of potential Outcomes

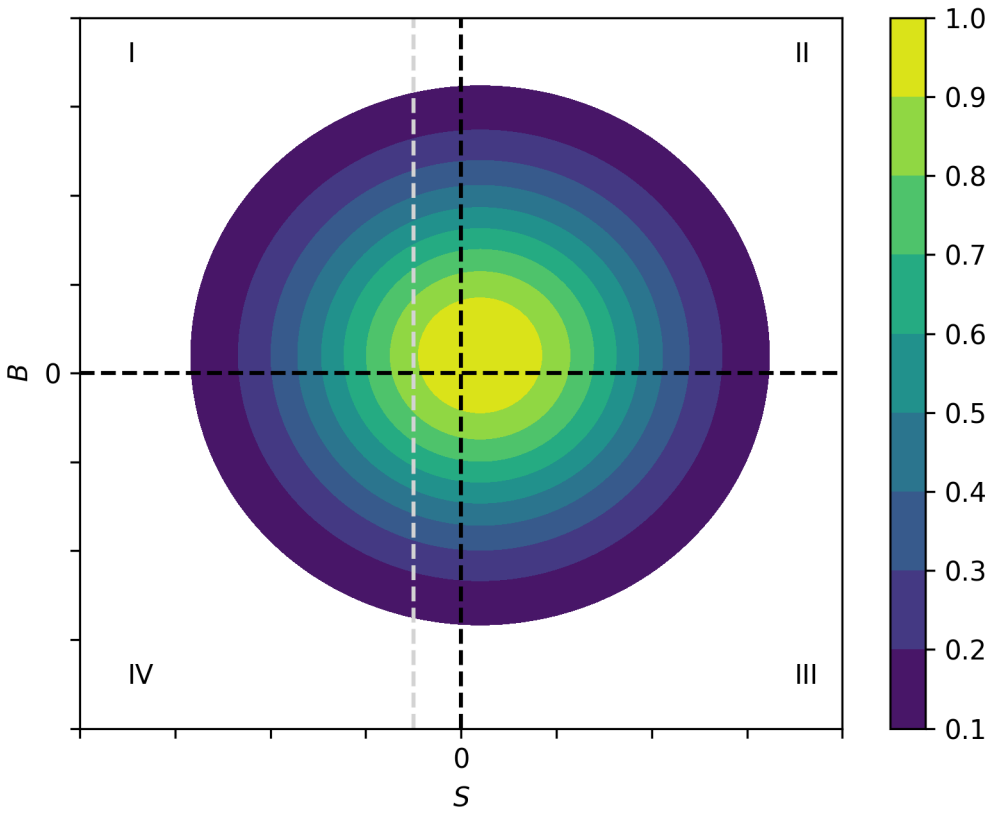


Fig. 6: **Fig. 6:** Distribution of benefits and surplus

CHAPTER

TWO

INSTALLATION

The `grmpy` package can be conveniently installed from the Python Package Index (PyPI) or directly from its source files. We currently support Python 3.6< on Linux systems.

2.1 Python Package Index

You can install the stable version of the package the usual way.

```
$ pip install grmpy
```

2.2 Source Files

You can download the sources directly from our GitHub repository.

```
$ git clone https://github.com/OpenSourceEconomics/grmpy.git
```

Once you obtained a copy of the source files, installing the package in editable model is straightforward.

```
$ pip install -e .
```

2.3 Test Suite

Please make sure that the package is working properly by running our test suite using `pytest`.

```
$ python -c "import grmpy; grmpy.test()"
```

CHAPTER

THREE

TUTORIAL

We now illustrate the basic capabilities of the `grmpy` package. We start by outlining some basic functional form assumptions before introducing to alternative models that can be used to estimate the marginal treatment effect (MTE). We then turn to some simple use cases.

3.1 Assumptions

The `grmpy` package implements the normal linear-in-parameters version of the generalized Roy model. Both potential outcomes and the choice (Y_1, Y_0, D) are a linear function of the individual's observables (X, Z) and random components (U_1, U_0, V).

$$
\begin{aligned}
Y_1 &= X\beta_1 + U_1 \\
Y_0 &= X\beta_0 + U_0 \\
D &= I[D^* > 0] \\
D^* &= Z\gamma - V
\end{aligned}
$$

Individuals decide to select into latent indicator variable D^* is positive. Depending on their decision, we either observe Y_1 or Y_0.

3.1.1 Parametric Normal Model

The parametric model imposes the assumption of joint normality of the unobservables $(U_1, U_0, V) \sim \mathcal{N}(0, \Sigma)$ with mean zero and covariance matrix Σ.

3.1.2 Semiparametric Model

The semiparametric approach invokes no assumption on the distribution of the unobservables. It requires a weaker condition $(X, Z)U_1, U_0, V$

Under this assumption, the MTE is:

- additively separable in X and U_D, which means that the shape of the MTE is independent of X, and
- identified over the common support of $P(Z)$, unconditional on X.

The assumption of common support is crucial for the application of LIV and needs to be carefully evaluated every time. It is defined as the region where the support of $P(Z)$ given $D = 1$ and the support of $P(Z)$ given :math:`D=0 overlap.

3.2 Model Specification

You can specify the details of the model in an initialization file (example). This file contains several blocks:

SIMULATION

The *SIMULATION* block contains some basic information about the simulation request.

Key	Value	Interpretation
agents	int	number of individuals
seed	int	seed for the specific simulation
source	str	specified name for the simulation output files

ESTIMATION

Depending on the model, different input parameters are required.

PARAMETRIC MODEL

Key	Value	Interpretation
semipar	False	choose the parametric normal model
agents	int	number of individuals (for the comparison file)
file	str	name of the estimation specific init file
optimizer	str	optimizer used for the estimation process
start	str	flag for the start values
maxiter	int	maximum numbers of iterations
dependent	str	indicates the dependent variable
indicator	str	label of the treatment indicator variable
output_file	str	name for the estimation output file
comparison	int	flag for enabling the comparison file creation

SEMIPARAMETRIC MODEL

Key	Value	Interpretation
semipar	True	choose the semiparametric model
show_output	bool	If *True*, intermediate outputs of the estimation process are displayed
dependent	str	indicates the dependent variable
indicator	str	label of the treatment indicator variable
file	str	name of the estimation specific init file
logit	bool	If false: probit. Probability model for the choice equation
nbins	int	Number of histogram bins used to determine common support (default is 25)
bandwidth	float	Bandwidth for the locally quadratic regression
gridsize	int	Number of evaluation points for the locally quadratic regression (default is 400)
ps_range	list	Start and end point of the range of $p = u_D$ over which the MTE shall be estimated
rbandwidth	int	Bandwidth for the double residual regression (default is 0.05)
trim_support	bool	Trim the data outside the common support, recommended (default is *True*)
reesti-mate_p	bool	Re-estimate $P(Z)$ after trimming, not recommended (default is *False*)

In most empirical applications, bandwidth choices between 0.2 and 0.4 are appropriate. [11] find that a gridsize of 400 is a good default for graphical analysis. For data sets with less than 400 observations, we recommend a gridsize equivalent to the maximum number of observations that remain after trimming the common support. If the data set of size N is large enough, a gridsize of 400 should be considered as the minimal number of evaluation points. Since *grmpy*'s algorithm is fast enough, gridsize can be easily increased to N evaluation points.

The "rbandwidth", which is 0.05 by default, specifies the bandwidth for the LOESS (Locally Estimated Scatterplot Smoothing) regression of X, $X \times p$, and Y on $\widehat{P}(Z)$. If the sample size is

small (N < 400), the user may need to increase "rbandwidth" to 0.1. Otherwise *grmpy* will throw an error.

Note that the MTE identified by LIV consists of wo components: $\overline{x}(\beta_1 - \beta_0)$ (which does not depend on $P(Z) = p$) and $k(p)$ (which does depend on p). The latter is estimated nonparametrically. The key "p_range" in the initialization file specifies the interval over which $k(p)$ is estimated. After the data outside the overlapping support are trimmed, the locally quadratic kernel estimator uses the remaining data to predict $k(p)$ over the entire "p_range" specified by the user. If "p_range" is larger than the common support, *grmpy* extrapolates the values for the MTE outside this region. Technically speaking, interpretations of the MTE are only valid within the common support. In our empirical applications, we set "p_range" to $[0.005, 0.995]$.

The other parameters ("trim_support" and "reestimate_p") are set by default and do not need to be specified by the user. In rare cases, the user might wish to change these parameters. In general, we do not recommend this.

TREATED

The *TREATED* block specifies the number and order of the covariates determining the potential outcome in the treated state and the values for the coefficients β_1. Note that the length of the list which determines the parameters has to be equal to the number of variables that are included in the order list.

Key	Container	Values	Interpretation
params	list	float	Parameters
order	list	str	Variable labels

UNTREATED

The *UNTREATED* block specifies the covariates that a the potential outcome in the untreated state and the values for the coefficients β_0.

Key	Container	Values	Interpretation
params	list	float	Parameters
order	list	str	Variable labels

CHOICE

The *CHOICE* block specifies the number and order of the covariates determining the selection process and the values for the coefficients γ.

Key	Container	Values	Interpretation
params	list	float	Parameters
order	list	str	Variable labels

3.2.1 Further Specifications for the Parametric Model

DIST

The *DIST* block specifies the distribution of the unobservables.

Key	Container	Values	Interpretation
params	list	float	Upper triangular of the covariance matrix

VARTYPES

The *VARTYPES* section enables users to specify optional characteristics to specific variables in their simulated data. Currently there is only the option to determine binary variables. For this purpose the user have to specify a key which reflects the corresponding variable label and assign a list to this label which contains the type (*binary*) as a string as well as a float (<0.9) that determines the probability for which the variable is one.

Key	Container	Values	Interpretation
Variable label	list	string and float	Type of variable + additional information

SCIPY-BFGS

The *SCIPY-BFGS* block contains the specifications for the *BFGS* minimization algorithm. For more information see: SciPy documentation.

Key	Value	Interpretation
gtol	float	the value that has to be larger as the gradient norm before successful termination
eps	float	value of step size (if *jac* is approximated)

SCIPY-POWELL

The *SCIPY-POWELL* block contains the specifications for the *POWELL* minimization algorithm. For more information see: SciPy documentation.

Key	Value	Interpretation
xtol	float	relative error in solution values *xopt* that is acceptable for convergence
ftol	float	relative error in fun(*xopt*) that is acceptable for convergence

3.3 Examples

3.3.1 Parametric Normal Model

In the following chapter we explore the basic features of the `grmpy` package. The resources for the tutorial are also available online. So far the package provides the features to simulate a sample from the generalized Roy model and to estimate some parameters of interest for a provided sample as specified in your initialization file.

Simulation

First we will take a look on the simulation feature. For simulating a sample from the generalized Roy model you use the `simulate()` function provided by the package. For simulating a sample of your choice you have to provide the path of your initialization file as an input to the function.

```
import grmpy

grmpy.simulate('tutorial.grmpy.yml')
```

This creates a number of output files that contain information about the resulting simulated sample.

- **data.grmpy.info**, basic information about the simulated sample
- **data.grmpy.txt**, simulated sample in a simple text file
- **data.grmpy.pkl**, simulated sample as a pandas data frame

Estimation

The other feature of the package is the estimation of the parameters of interest. By default, the parametric model is chosen, in which case the parameter *semipar* in the *ESTIMATION* section of the initialization file is set to *False*. The start values and optimizer options need to be specified in the *ESTIMATION* section.

```
grmpy.fit('tutorial.grmpy.yml', semipar=False)
```

As in the simulation process this creates an output files that contain information about the estimation results.

3.3.2 Local Instrumental Variables

If the user wishes to estimate the parameters of interest using the semiparametric LIV approach, *semipar* must be changed to *True*.

```
grmpy.fit('tutorial.semipar.yml', semipar=True)
```

If *show_output* is *True*, `grmpy` plots the common support of the propensity score and shows some intermediate outputs of the estimation process.

CHAPTER

FOUR

RELIABILITY

The following section illustrates the reliability of the estimation strategy behind the `grmpy` package when facing agent heterogeneity and shows also that the corresponding results withstand a critical examination. The checks in both subsections are based on a mock data set respectively the estimation results from

> Carneiro, Pedro, James J. Heckman, and Edward J. Vytlacil. Estimating Marginal Returns to Education. *American Economic Review*, 101 (6):2754-81, 2011.

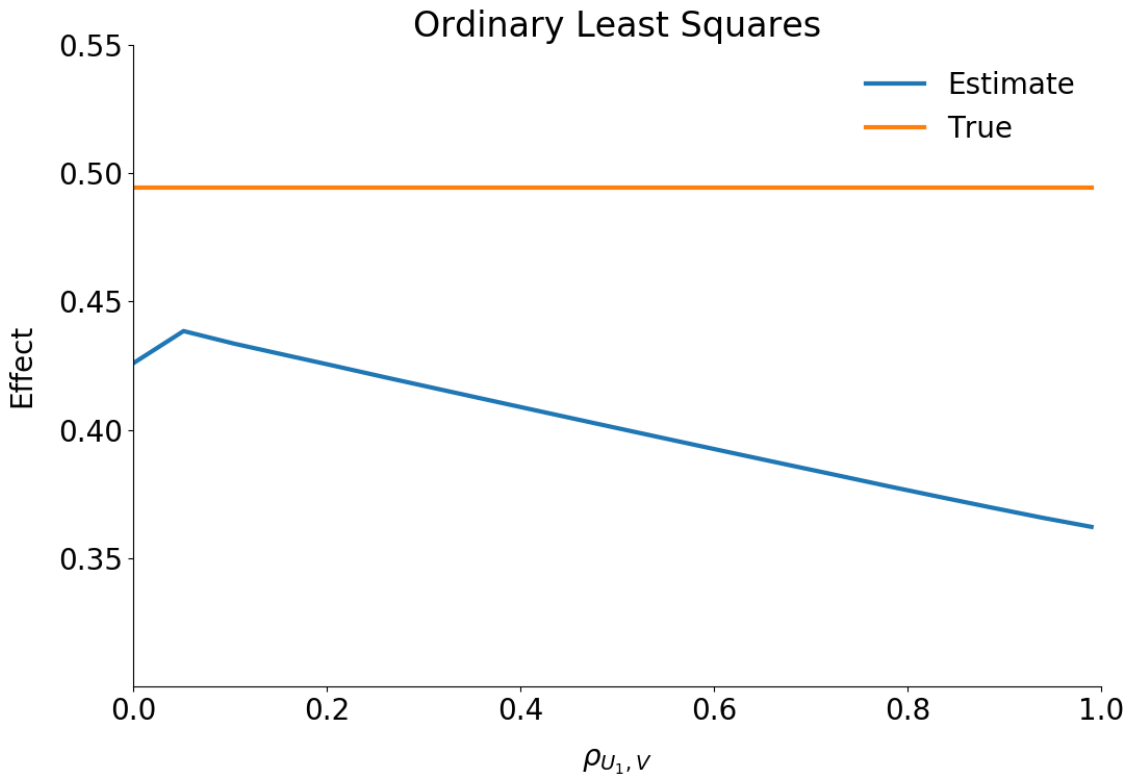
We conduct two different test setups. Firstly we show that `grmpy` is able to provide better results than simple estimation approaches in the presence of essential heterogeneity. For that purpose we rely on a monte carlo simulation setup which enables us to compare the results obtained by the `grmpy` estimation process with such of a standard Ordinary Least Squares (OLS) approach. Secondly we show that `grmpy` is capable of replicating the B^{MTE} results by [7] for the parametric version of the Roy model.

4.1 Reliability of Estimation Results

The estimation results and data from [7] build the basis of the reliability test setup. The data is extended by combining them with simulated unobservables that follow a distribution that is prespecified in the following initialization file. In the next step the potential outcomes and the choice of each individual are calculated by using the estimation results.

This process is iterated a certain amount of times. During each iteration the rate of correlation between the simulated unobservables increases. Translated in the Roy model framework this is equivalent to an increase in the correlation between the unobservable variable U_1 and V, the unobservable that indicates the preference for selecting into treatment. Additionally the specifications of the distributional characteristics are designed so that the expected value of each unobservable is equal to zero. This ensures that the true B^{ATE} is fixed to a value close to 0.5 independent of the correlation structure.

For illustrating the reliability we estimate B^{ATE} during each step with two different methods. The first estimation uses a simple OLS approach.

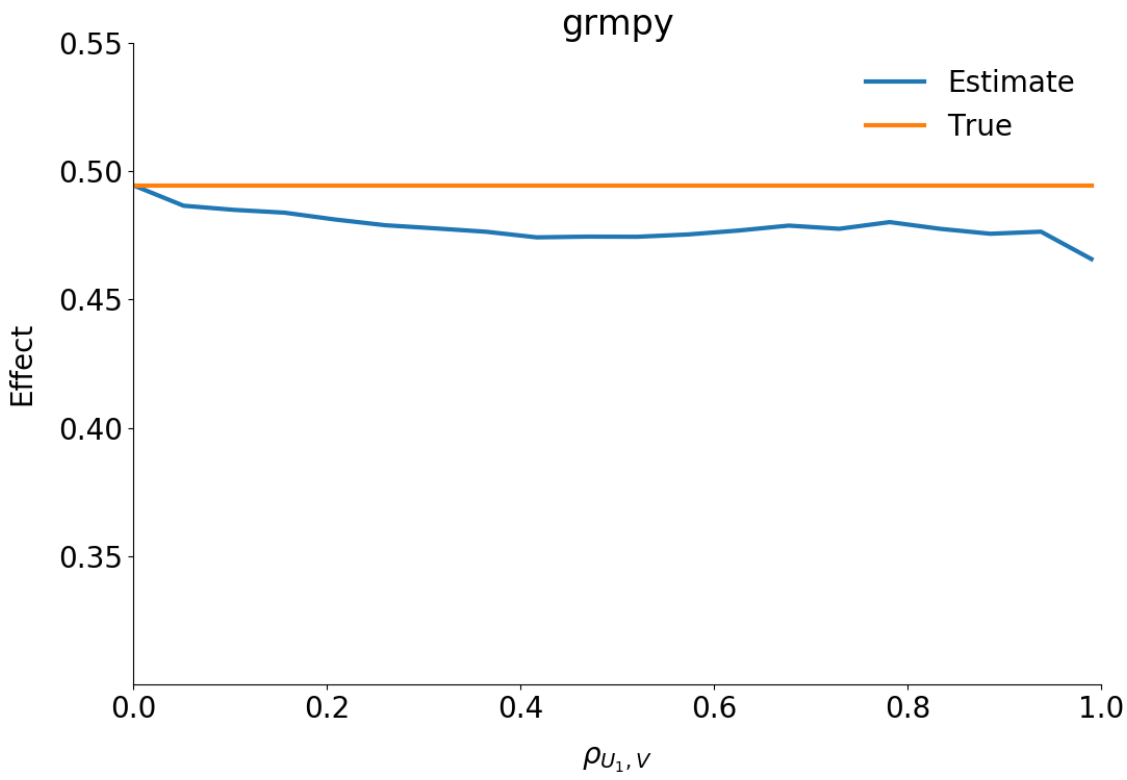


As can be seen from the figure, the OLS estimator underestimates the effect significantly. The stronger the correlation between the unobservable variables the more or less stronger the downwards bias.

The second figure shows the estimated B^{ATE} from the `grmpy` estimation process. Conversely to the OLS results the estimate of the average effect is close to the true value even if the unobservables are almost perfectly correlated.

4.2 Sensitivity to Different Distributions of the Unobservables

The parametric specification makes the strong assumption that the unobservables follow a joint normal distribution. The semiparametric method of local instrumental variables is more flexible, as it does not invoke conditions on the functional form. We test how sensitive the two methods to different distributions of the unobservables. To that end, we use a toy model of the returns to college (based on [5]), where we know the true shape of the B^{MTE}.



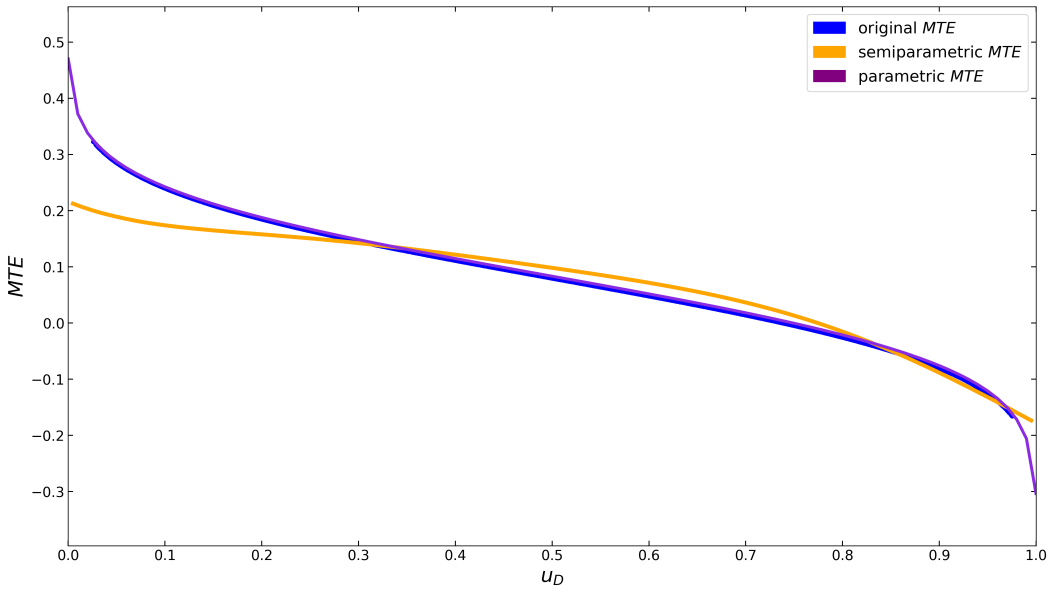
4.2.1 Normal Distribution

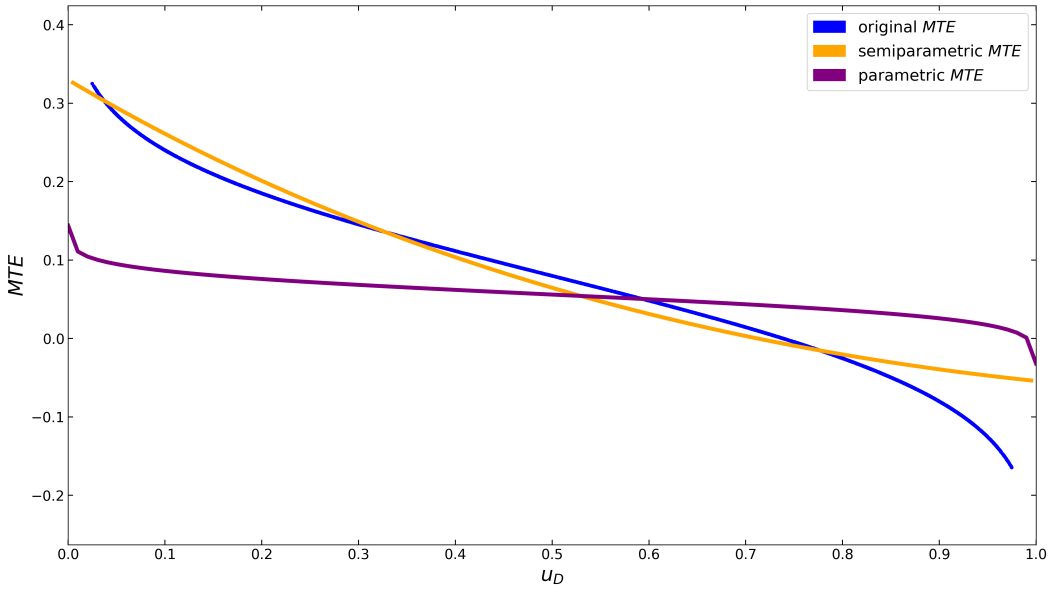
Both specifications come very close to the original curve. The parametric model even gets a perfect fit.

4.2.2 *beta* Distribution

The shape of the *beta* distribution can be flexibly adjusted by the tuning parameters α and β, which we set to 4 and 8, respectively.

The parametric model underestimates the returns to college, whereas the semiparametric B^{MTE} still fits the original curve pretty well. The latter makes no assumption on the functional form of the unobservables and, thus, is more flexible in estimating the parameter of interest when the assumption of joint normality is violated. Which model is superior depends on the context. In empirical applications, we recommend to examine both.


original MTE
semiparametric MTE
parametric MTE
MTE
u_D


original MTE
semiparametric MTE
parametric MTE
MTE
u_D

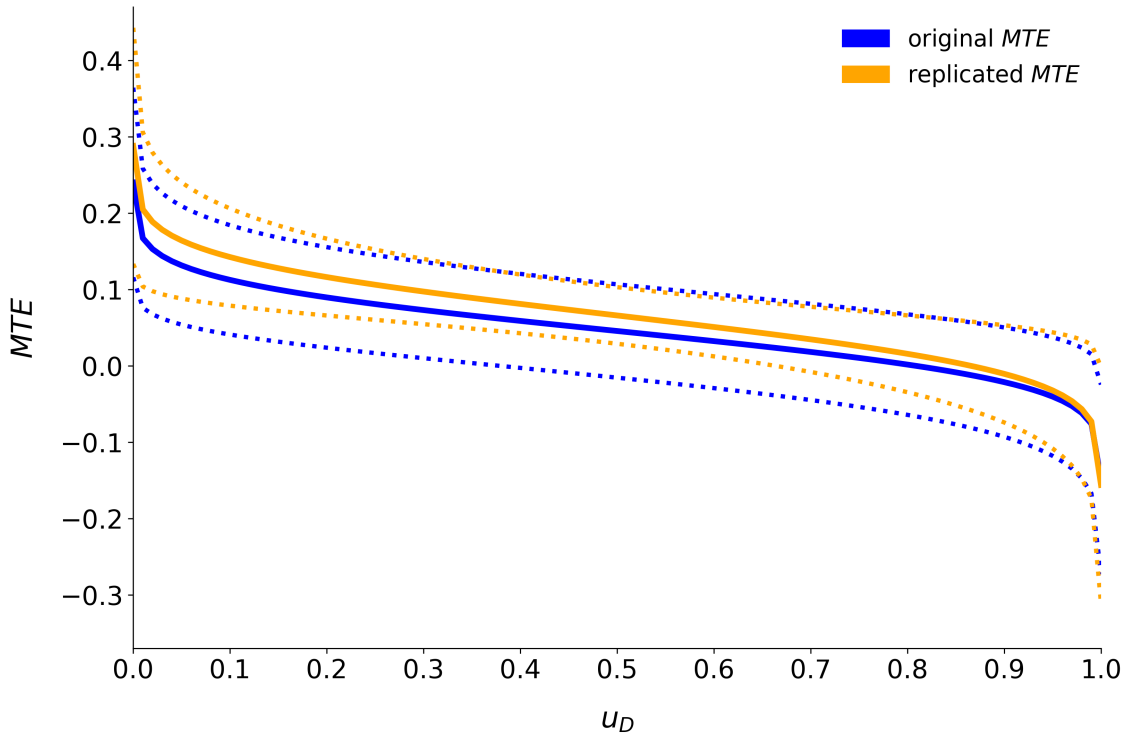
4.3 Reliability

In another check of reliability, we compare the results of our estimation process with already existing results from the literature. For this purpose we replicate the results for both the parametric and semiparametric MTE from Carneiro 2011 ([7]). Note that we make use of a mock data set, as the original data cannot be fully recreated from the replication material.

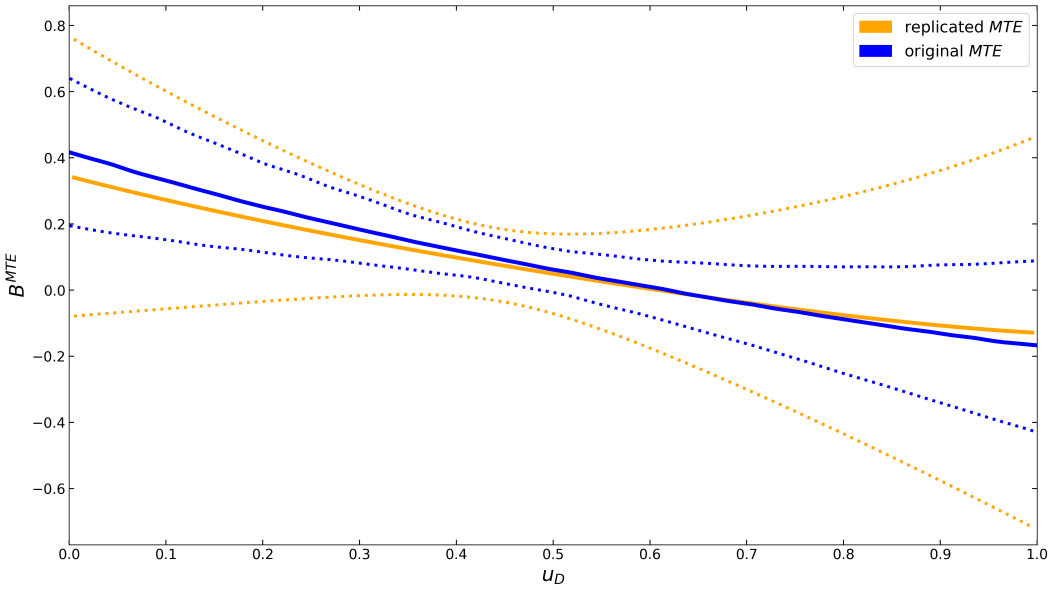
We provide two jupyter notebooks for easy reconstruction of the parametric as well as the semi-parametric setup. The corresponding initialization files can be found here and here.

4.3.1 Parametric Replication

As shown in the figure below, the parametric B^{MTE} is really close to the original results. The deviation seems to be negligible because of the use of a mock dataset.



4.3.2 Semiparametric Replication



The semiparametric B^{MTE} also gets very close to the original curve. However, the 90 percent confidence bands (250 bootstrap replications) are wider. As opposed to the parametric model, where the standard error bands are computed analytically, confidence bands in the semiparametric setup are obtained via the bootstrap method, which is sensitive to the discrepancies in the mock data set.

CHAPTER

FIVE

SOFTWARE ENGINEERING

We now briefly discuss our software engineering practices that help us to ensure the transparency, reliability, scalability, and extensibility of the `grmpy` package. Please visit us at the Software Engineering for Economists Initiative for an accessible introduction on how to integrate these practices in your own research.

5.1 Test Battery

We use pytest as our test runner. We broadly group our tests in three categories:

- **property-based testing**

  We create random model parameterizations and estimation requests and test for a valid return of the program.

- **reliability testing**

  We conduct numerous Monte Carlo exercises to ensure that we can recover the true underlying parameterization with an estimation. Also by varying the tuning parameters of the estimation (e.g. random draws for integration) and the optimizers, we learn about their effect on estimation performance.

- **regression testing**

  We provide a regression test. For this purpose we generated random model parameterizations, simulated the coresponding outputs, summed them up and saved both, the parameters and the sums in a json file. The json file is part of the package. Through this the provided test is able to draw parameterizations randomly from the json file. In the next step the test simulates the output variables and compares the sum of the simulated output with the associated json file information. This ensures that the package works accurate even after an update to a new version.

5.2 Documentation

The documentation is created using Sphinx and hosted on Read the Docs.

5.3 Code Review

We use several automatic code review tools to help us improve the readability and maintainability of our code base. For example, we work with Codacy. However, we also conduct regular peer code-reviews using Reviewable.

5.4 Continuous Integration Workflow

We set up a continuous integration workflow around our GitHub Organization. We use the continuous integration services provided by Travis CI. tox ensures that the package installs correctly with different Python versions.

CHAPTER

SIX

CONTRIBUTING

Great, you are interested in contributing to the package.

To get acquainted with the code base, you can check out our issue tracker for some immediate and clearly defined tasks. For more involved contributions, please see our roadmap below. All submissions are required to follow this project-agnostic contribution guide

6.1 Roadmap

We aim for improvements to the `grmpy` package in four domains: Objects of Interest, Estimation Methods, and Numerical Methods.

6.1.1 Objects of Interest

- adding marginal surplus and marginal cost parameters as presented by [10]

6.1.2 Estimation Methods

- implement polynomial and local-instrumental variable estimation as outlined by [18]
- implementing capability to control for unobservables by adding a factor structure assumption as in [9]

6.1.3 Program Structure

- refactoring code to incorporate elements of object-oriented programming

6.1.4 Numerical Methods

- exploring alternative optimization algorithms to address large estimation tasks

CHAPTER

SEVEN

CONTACT AND CREDITS

If you have any questions or comments, please do not hesitate to contact us directly.

7.1 BDFL

Philipp Eisenhauer

7.2 Development Lead

Sebastian Becker, Sebastian Gsell

7.3 Contributors

Maximilian Blesch, Benedikt Kauf, Tobias Raabe

7.4 Acknowledgments

We appreciate the financial support of the AXA Research Fund and the University of Bonn. We are indebted to the open source community as we build on top of numerous open source tools such as the SciPy Stack and statsmodels.

7.4.1 Suggested Citation

We appreciate citations for `grmpy` because it helps us find out how people have been using the package and it motivates further work. Please use our Digital Object Identifier (DOI) and see here for further citation styles.

CHAPTER

EIGHT

CHANGES

This is a record of all past `grmpy` releases and what went into them in reverse chronological order. We follow semantic versioning and all releases are available on PyPI.

8.1 1.0.0 - 2018-XX-XX

This is the initial release of the `grmpy` package.

CHAPTER

NINE

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Arild Aakvik, James J. Heckman, and Edward J. Vytlacil. Estimating treatment effects for discrete outcomes when responses to treatment vary: An application to Norwegian vocational rehabilitation programs. *Journal of Econometrics*, 124(1-2):15–51, 2005.

[2] Jaap H. Abbring and James J. Heckman. Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. In *Handbook of Econometrics*, volume 6B, chapter 72, pages 5145 –5303. Elsevier Science, 2007.

[3] Abadie Alberto, Angrist Joshua, and Imbens Guido. Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings. *Econometrica*, 70(1):91–117, 2002.

[4] Anders M. Björklund and Robert Moffitt. The estimation of wage gains and welfare gains in self-selection models. *Review of Economics and Statistics*, 69(1):42–49, 1987.

[5] S. Brave and T. Walstrum. Estimating marginal treatment effects using parametric and semi-parametric methods. *Stata Journal*, 14(1):191–217, 2014.

[6] Martin Browning, Lars Peter Hansen, and James J. Heckman. Micro data and general equilibrium models. In *Handbook of Macroeconomics*, volume 1A, chapter 8, pages 543–633. Elsevier Science, 1999.

[7] Pedro Carneiro, James J. Heckman, and Edward J. Vytlacil. Estimating Marginal Returns to Education. *American Economic Review*, 101(6):2754–81, 2011.

[8] William G. Cochran and Donald B. Rubin. Controlling bias in observational studies: A review. *Sankhya: The Indian Journal of Statistics Series A*, 35(4):417–446, 1972.

[9] Philipp Eisenhauer. Issues in the econometrics of policy evaluation: explorations using a factor structure model. *Unpublished Manuscript*, 2013.

[10] Philipp Eisenhauer, James J. Heckman, and Edward J. Vytlacil. The generalized Roy model and the cost-benefit analysis of social programs. *Journal of Political Economy*, 123(2):413–443, 2015.

[11] Jianqing Fan and James S. Marron. Fast implementations of nonparametric curve estimators. *Journal of Computational and Graphical Statistics*, 3(1):35–56, 1994.

[12] Nicole Fortin, Thomas Lemieux, and Sergio Firpo. Chapter 1 - Decomposition Methods in Economics. In Orley Ashenfelter and David Card, editors, *Handbook of Labor Economics*, volume 4, pages 1 – 102. Elsevier, 2011.

[13] James J. Heckman. Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *Journal of Political Economy*, 109(4):673–748, 2001.

[14] James J. Heckman. Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic Literature*, 48(2):356–398, 2010.

[15] James J. Heckman, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. Characterizing selection bias using experimental data. *Journal of Political Economy*, 66(5):1017–1098, 1998.

[16] James J. Heckman and Daniel Schmierer. Tests of hypotheses arising in the correlated random coefficient model. *Economic Modelling*, 27(6):1355 – 1367, 2010.

[17] James J. Heckman, Jeffrey Smith, and Nancy Clements. Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *Review of Economic Studies*, 64(4):487–535, 1997.

[18] James J. Heckman, Sergio Urzua, and Edward J. Vytlacil. Understanding instrumental variables in models with essential heterogeneity. *Journal of Political Economy*, 88(3):389–432, 2006.

[19] James J. Heckman and Edward J. Vytlacil. Policy-relevant treatment effects. *American Economic Review*, 91(2):107–111, 2001.

[20] James J. Heckman and Edward J. Vytlacil. Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, 73(3):669–739, 2005.

[21] James J. Heckman and Edward J. Vytlacil. Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. In *Handbook of Econometrics*, volume 6B, chapter 71, pages 4875–5143. Elsevier Science, 2007.

[22] James J. Heckman and Edward J. Vytlacil. Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation. In *Handbook of Econometrics*, volume 6B, chapter 70, pages 4779–4874. Elsevier Science, 2007.

[23] Paul W. Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81(386):945–960, 1986.

[24] Ralph I. Horwitz, Burton H. Singer, Robert. W. Makuch, and Catherine M. Viscoli. Can treatment that is helpful on average be harmful to some patients? A study of conflicting information needs of clinical inquiry and drug regulation. *Journal of Clinical Epidemiology*, 49(4):395–400, 1996.

[25] Ralph I. Horwitz, Burton H. Singer, Robert. W. Makuch, and Catherine M. Viscoli. Reaching the tunnel at the end of the light. *Journal of Clinical Epidemiology*, 50(7):753–755, 1997.

[26] Guido W. Imbens and Joshua D. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475, 1994.

[27] Richard E. Quandt. The estimation of the parameters of a linear regression system obeying two seperate regimes. *Journal of the American Statistical Association*, 53(284):873–880, 1958.

[28] Richard E. Quandt. A new approach to estimating switching regressions. *Journal of the American Statistical Association*, 67(338):306–310, 1972.

[29] Paul R. Rosenbaum and Donald B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[30] Andrew D. Roy. Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3(2):135–146, 1951.

[31] Firpo Sergio. Efficient Semiparametric Estimation of Quantile Treatment Effects. *Econometrica*, 75(1):259–276, 2007.

[32] Chernozhukov Victor and Hansen Christian. An IV Model of Quantile Treatment Effects. *Econometrica*, 73(1):245–261, 2005.