
grmpy Documentation

Release 1.0

Development Team

Jul 16, 2018

Contents

1	Economics	3
2	Installation	13
3	Tutorial	15
4	Reliability & Robustness	19
5	Software Engineering	23
6	Contributing	25
7	Contact and Credits	27
8	Changes	29
9	Bibliography	31
	Bibliography	33

[PyPI](#) | [GitHub](#) | [Issues](#)

`grmpy` is an open-source Python 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. Vytlačil. 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. Vytlačil. 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.

Contents:

This section provides a general discussion of the generalized Roy model and selected issues in the econometrics of policy evaluation. The `grmpy` package implements a parametric normal version of the model, please see the next section for details.

1.1 Generalized Roy Model

The generalized Roy model ([24]; [14]) provides a coherent framework to explore the econometrics of policy evaluation. Its parametric version 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

$$D = I[S > 0]$$

$$S = E[Y_1 - Y_0 - C \mid \mathcal{I}]$$

$$C = \mu_C(Z) + U_C$$

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. C denotes the subjective cost of treatment participation. Any subjective benefits, e.g. job amenities, are included (as a negative contribution) in the subjective cost of treatment. Agents take up treatment D if they expect the objective benefit to outweigh the subjective cost. In that case, their subjective evaluation, i.e. the expected surplus from participation S , is positive. 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 ([21], [22]). 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 ([24]).

From the perspective of the econometrician, (X, Z) are observable while (U_1, U_0, U_C) are not. X are the observed determinants of potential outcomes (Y_1, Y_0) , and Z are the observed determinants of the cost of treatment C . Potential outcomes and cost are decomposed into their means $(\mu_1(X), \mu_0(X), \mu_C(Z))$ and their deviations from the mean (U_1, U_0, U_C) . (X, Z) might have common elements. Observables and unobservables jointly determine program participation D .

If their ex ante surplus S from participation is positive, then agents select into treatment. Yet, this does not require their expected objective returns to be positive as well. Subjective cost C might be negative such that agents which expect negative returns still participate. Moreover, in the case of imperfect information, an agent's ex ante evaluation of treatment is potentially different from their ex post assessment.

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 ([17]).

1.2 Selected Issues

We now highlight some selected issues in the econometrics of policy evaluation that can be fruitfully discussed within the framework of the model.

1.2.1 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 ([4]; [8]).

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 I and thus known to the agent at the time of the treatment decision.

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 ([12]).

1.2.2 Essential Heterogeneity

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. 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 ([6]; [23] ; [10]), are not able to identify any parameter of interest ([14]; [12]).

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 ([14], [16], [15]). 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 (ATE), or the average effect on the treated (TT) or untreated (TUT).

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

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, then agents that take up treatment benefit more than those that do not and thus $TT > TUT$. If agents select their treatment status at random, then all parameters are equal. The policy relevance of the conventional treatment effect parameters is limited. They are only informative about extreme policy alternatives. The 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 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 TUT is the policy relevant parameter. To ensure a tight link between the posed policy question and the parameter of interest, Heckman and Vytlacil ([13]) propose the policy-relevant treatment effect (PRTE). They consider policies that do not change potential outcomes, but only affect individual choices. Thus, they account for voluntary program participation. Policy-Relevant Average Treatment Effects The 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

$$\begin{aligned}
 Y_B &= D_B Y_1 + (1 - D_B) Y_0 \\
 Y_A &= D_A Y_1 + (1 - D_A) Y_0.
 \end{aligned}$$

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

$$PRTE = E[D_A] - E[D_B](E[Y_A] - E[Y_B]).$$

In our empirical illustration, in which we consider education policies, the lack of policy relevance of the conventional effect parameters is particularly evident. 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. That is why we estimate the policy-relevant effects of alternative education policies and contrast them with the conventional treatment effect parameters. We also show how the PRTE varies for alternative policy proposals as different agents are induced to change their treatment status.

1.3.2 Local Average Treatment Effect

The Local Average Treatment Effect (LATE) was introduced by Imbens and Angrist ([20]). They show that instrumental variable estimator identify LATE, which measures the mean gross return to treatment for individuals induced into treatment by a change in an instrument.

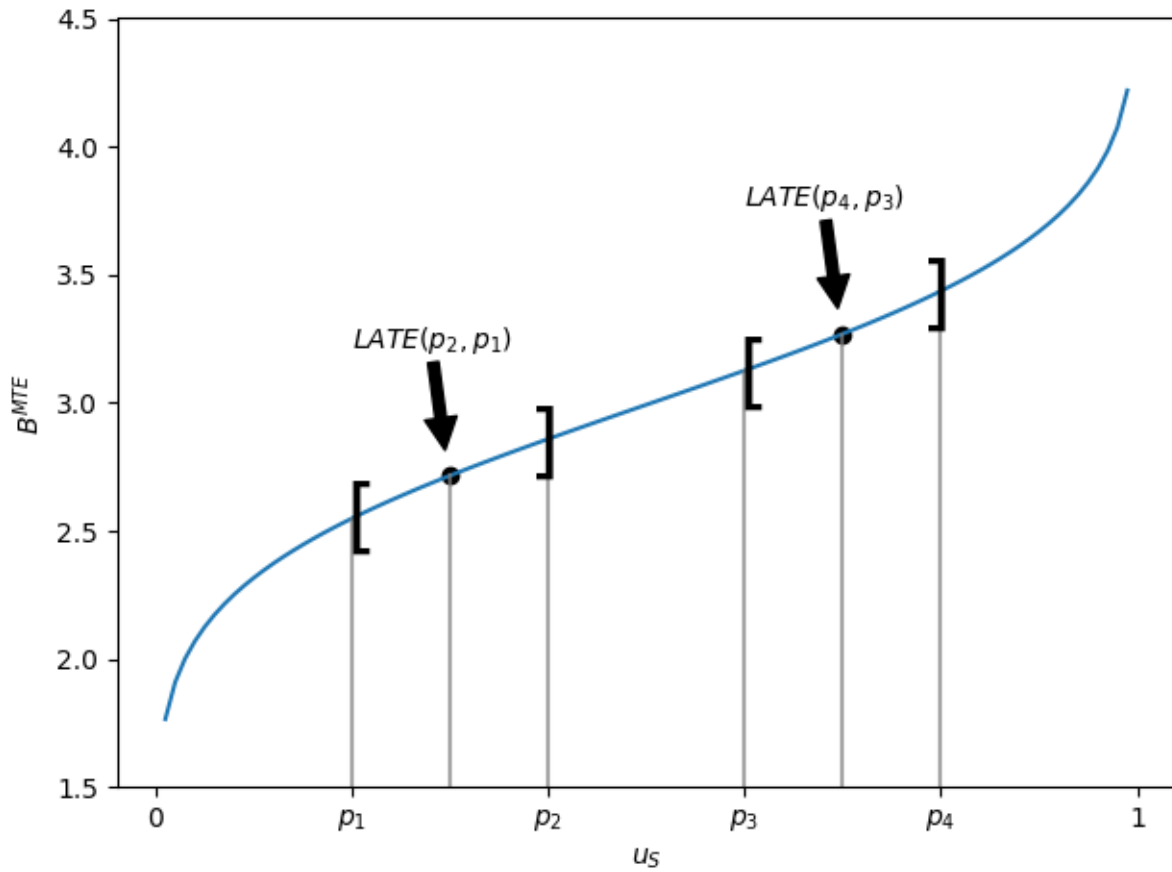


Fig. 1: LATE at different values of u_S

Unfortunately, the people induced to go into state 1 ($D = 1$) by a change in any particular instrument need not to be the same as the people induced to go to state 1 by policy changes other than those corresponding exactly to the variation in the instrument. A desired policy effect may not be directly correspond to the variation in the IV. 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. Heckman develops this argument in detail ([9]).

The average effect of a policy and the average effect of a treatment are linked by the marginal treatment effect (MTE). The MTE was introduced into the literature by Björklund and Moffitt ([3]) and extended by Heckman and Vytlacil ([13], [14], [15]).

1.3.3 Marginal Treatment Effect

The MTE is the treatment effect parameter that conditions on the unobserved desire to select into treatment. Let $V = E[U_C - (U_1 - U_0)|I]$ summarize the expectations about all unobservables determining treatment choice and let $U_S = F_V(V)$. Then, the MTE is defined as

$$MTE(x, u_S) = E[Y_1 - Y_0 | X = x, U_S = u_S].$$

The 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 MTE shows how the average benefit varies along the distribution of V . For u_S evaluation points close to zero, the 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 MTE provides the underlying structure for all average effect parameters previously discussed. These can be derived as weighted averages of the MTE ([14]).

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

$$\Delta j(x) = \int_0^1 MTE(x, u_S) h_j(x, u_S) du_S,$$

where the weights $h_j(x, u_S)$ are specific to parameter j , integrate to one, and can be constructed from data.

All parameters are identical only in the absence of essential heterogeneity. Then, the $MTE(x, u_S)$ is constant across the whole distribution of V as agents do not select their treatment status based on their unobservable benefits.

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 ([11]).

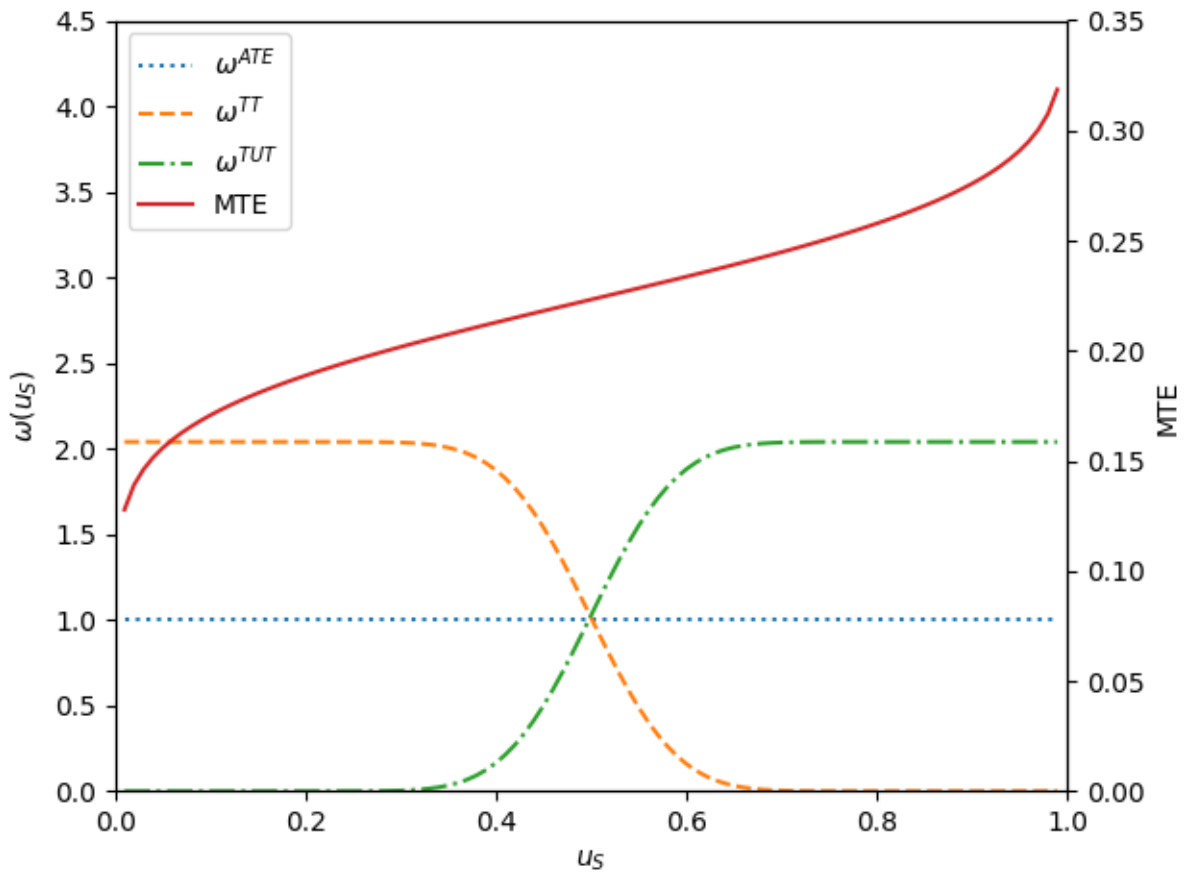


Fig. 2: Weights for the marginal treatment effect for different parameters.

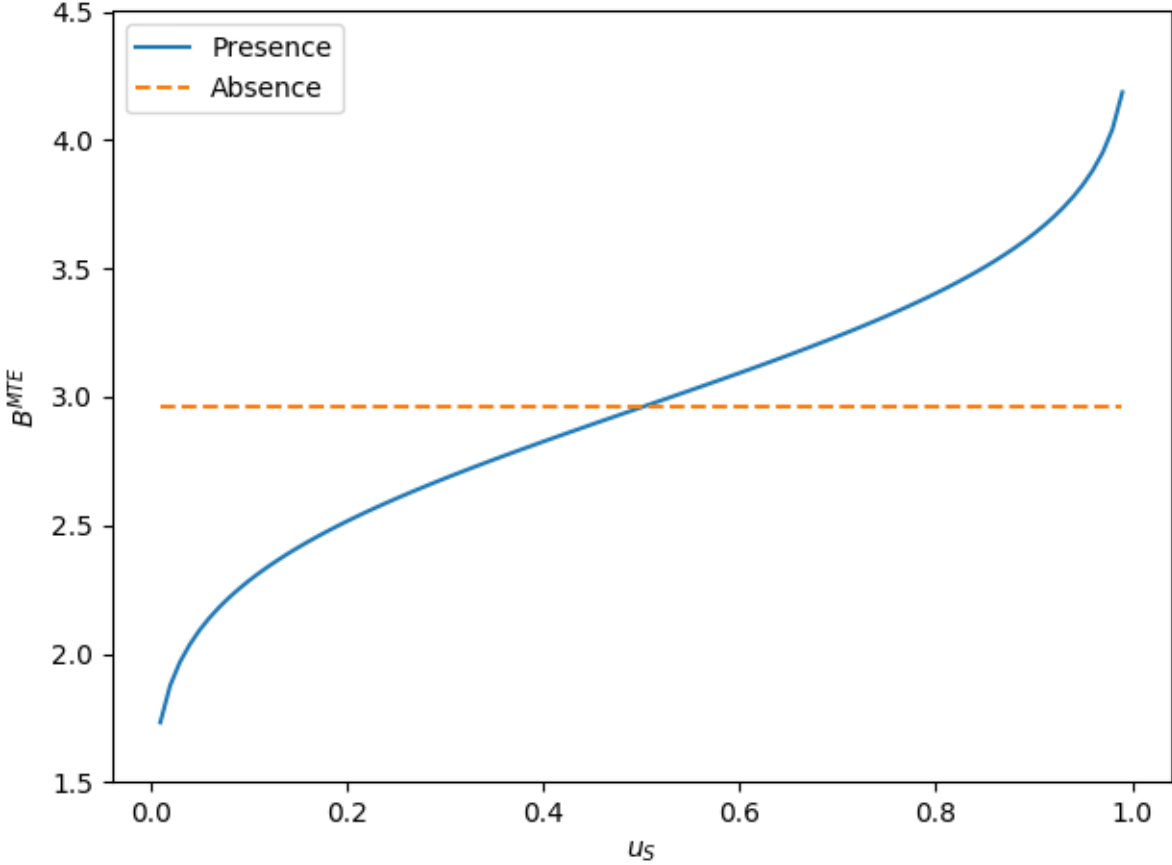


Fig. 3: MTE in the presence and absence of essential heterogeneity.

1.3.4 Distribution of Potential Outcomes

Several interesting aspects of policies cannot be evaluated without knowing the joint distribution of potential outcomes (see [2] and [11]). 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 is beneficial for agents on average can still be harmful for some. The absence 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 ([18], [19]).

The `grmpy` package can be conveniently installed from the [Python Package Index \(PyPI\)](#) or directly from its source files. We currently support Python 2.7 and 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()"
```

We now illustrate the basic capabilities of the `grmpy` package. We start with the assumptions about functional form and the distribution of unobservables and 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 cost associated with treatment participations (Y_1, Y_0, C) are a linear function of the individual's observables (X, Z) and random components (U_1, U_0, U_C) .

$$Y_1 = X\beta_1 + U_1$$

$$Y_0 = X\beta_0 + U_0$$

$$C = Z\gamma + U_C$$

We collect all unobservables determining treatment choice in $V = U_C - (U_1 - U_0)$. The unobservables follow a normal distribution $(U_1, U_0, V) \sim \mathcal{N}(0, \Sigma)$ with mean zero and covariance matrix Σ . Individuals decide to select into treatment if their surplus from doing so is positive $S = Y_1 - Y_0 - C$. Depending on their decision, we either observe Y_1 or Y_0 .

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

The *ESTIMATION* block determines the basic information for the estimation process.

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 .

Key	Column	Value	Interpretation
coeff	str	float	intercept coefficient
coeff	str	float	coefficient of the first covariate
coeff	str	float	coefficient of the second covariate

UNTREATED

The *UNTREATED* block specifies the number and order of the covariates determining the potential outcome in the untreated state and the values for the coefficients β_0 . In particular, the integer in the column **Column** specifies the column in the relevant dataset.

COST

The *COST* block specifies the number and order of the covariates determining the cost of treatment and the values for the coefficients γ . In particular, the integer in the column **Column** specifies the column in the relevant dataset.

Key	Column	Value	Interpretation
coeff	str	float	intercept coefficient
coeff	str	float	coefficient of the first covariate
coeff	str	float	coefficient of the second covariate

DIST

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

Key	Value	Interpretation
coeff	float	σ_{U_1}
coeff	float	σ_{U_1, U_0}
coeff	float	$\sigma_{U_1, V}$
coeff	float	σ_{U_0}
coeff	float	$\sigma_{U_0, V}$
coeff	float	σ_V

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 <i>jac</i> 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 <i>xopt</i> that is acceptable for convergence
ftol	float	relative error in $\text{fun}(xopt)$ that is acceptable for convergence

3.3 Examples

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 the parameters of interest (given a data set) 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.ini')
```

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. The specification regarding start values and the optimizer options are determined in the *ESTIMATION* section of the initialization file.

```
grmpy.estimate('tutorial.grmpy.ini')
```

As in the simulation process this creates a number of output file that contains information about the estimation results.

- **est.grmpy.info**, basic information of the estimation process
- **comparison.grmpy.txt**, distributional characteristics of the input sample and the samples simulated from the start and result values of the estimation process

Reliability & Robustness

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

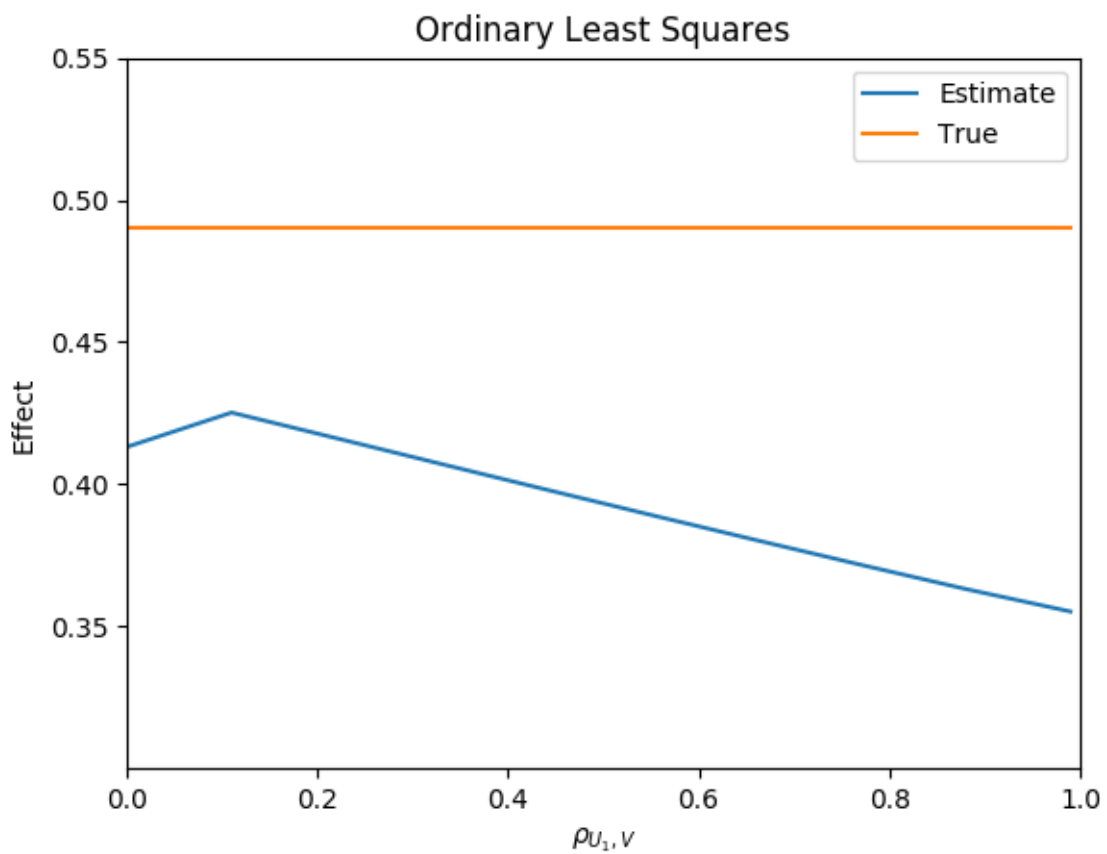
Carneiro, Pedro, James J. Heckman, and Edward J. Vytlačil. [Estimating Marginal Returns to Education](#). *American Economic Review*, 101 (6):2754-81, 2011.

4.1 Reliability

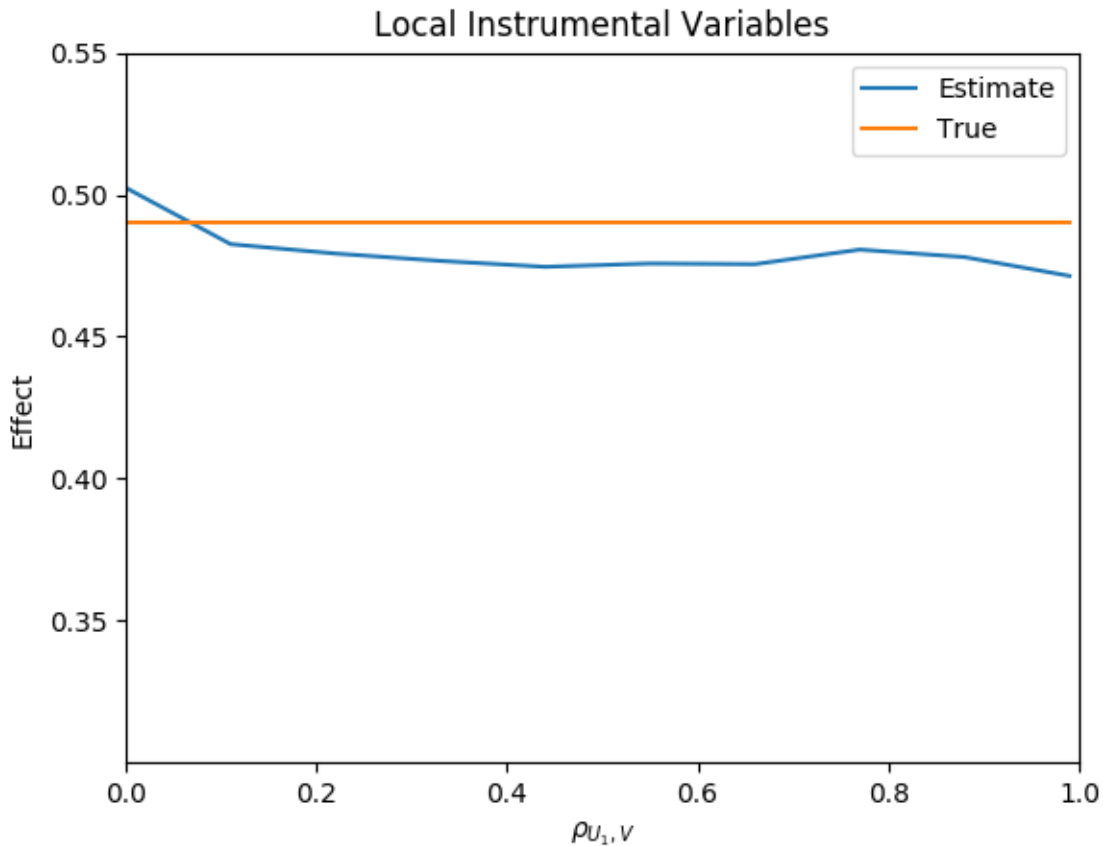
The estimation results and data from Carneiro 2011 build the basis of the reliability test setup. The data is extended by combining them with simulated unobservables that follow a distribution that is pre-specified in the following `initialization file`. In the next step the potential outcomes and the 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 average effect of treatment (ATE) is fixed to a value close to 0.5 independent of the correlation structure.

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



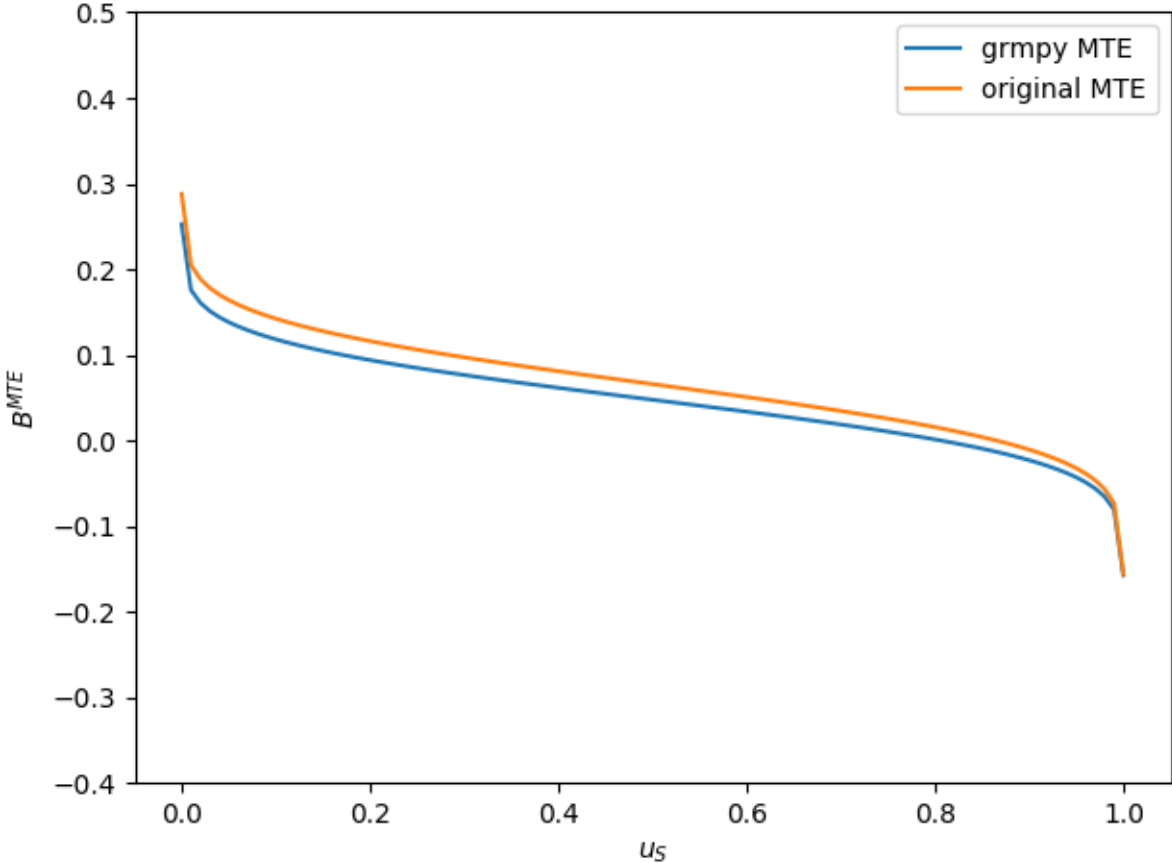
As can be seen from the figure, the OLS estimator underestimates the effect significantly. The stronger the correlation between the unobservable variables the more or less stronger the downwards bias.



The second figure shows the estimated 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 Robustness

For checking the robustness of the package's estimation process we replicate the results for the marginal treatment effect from Carneiro 2011 ([5]). For this purpose we created a jupyter notebook that runs an estimation based on an [initialization file](#). The init file corresponds to the specifications of the authors. As shown in the figure below the results are really close to the original results. The deviation seems to be negligible because of the usage of a mock dataset.



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 corresponding 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.

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 three 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 Eisenhauer et al. ([7])

6.1.2 Estimation Methods

- implementing polynomial and local-instrumental variable estimation as outlined by Heckman et al. ([12])

6.1.3 Numerical Methods

- exploring alternative optimization algorithms to address large estimation tasks

CHAPTER 7

Contact and Credits

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

7.1 Development Lead

Philipp Eisenhauer

7.2 Contributors

Sebastian Becker, Maximilian Blesch, Benedikt Kauf, Tobias Raabe

7.3 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](#).

7.3.1 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 8

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 9

Bibliography

Bibliography

- [1] Arild Aakvik, James J. Heckman, and Edward J. Vytlačil. 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] 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.
- [4] 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.
- [5] Pedro Carneiro, James J. Heckman, and Edward J. Vytlačil. Estimating Marginal Returns to Education. *American Economic Review*, 101(6):2754–81, 2011.
- [6] 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.
- [7] Philipp Eisenhauer, James J. Heckman, and Edward J. Vytlačil. The generalized Roy model and the cost-benefit analysis of social programs. *Journal of Political Economy*, 123(2):413–443, 2015.
- [8] James J. Heckman. Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *Journal of Political Economy*, 109(4):673–748, 2001.
- [9] James J. Heckman. Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic Literature*, 48(2):356–398, 2010.

- [10] 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.
- [11] 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.
- [12] 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.
- [13] James J. Heckman and Edward J. Vytlacil. Policy-relevant treatment effects. *American Economic Review*, 91(2):107–111, 2001.
- [14] James J. Heckman and Edward J. Vytlacil. Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, 73(3):669–739, 2005.
- [15] 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.
- [16] 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.
- [17] Paul W. Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81(386):945–960, 1986.
- [18] 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.
- [19] 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.
- [20] Guido W. Imbens and Joshua D. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475, 1994.
- [21] Richard E. Quandt. The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, 53(284):873–880, 1958.
- [22] Richard E. Quandt. A new approach to estimating switching regressions. *Journal of the American Statistical Association*, 67(338):306–310, 1972.
- [23] 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.
- [24] Andrew D. Roy. Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3(2):135–146, 1951.