grmpy Tutorial

Sebastian Becker

Introduction

grmpy

`grmpy` is an open-source Python package for the simulation and estimation of generalized Roy model. Its main purpose is to serve as a teaching tool to promote the conceptual framework provided by the generalized Roy model to illustrate a variety of issues in the econometrics of policy evaluation.

grmpy

- **grmpy** is ...
    - ...an open-source Python Package for the simulation and estimation of the generalized Roy model.

    - ...intended as an useful device to support and improve the understanding of the framework by the opportunity to experience the effect of particular specifications directly.

Setup

Setup

- Normal linear-in-parameters version of the generalized Roy model.

Potential Outcomes   Cost

$$Y_1 = \beta_1 X + U_1 \qquad C = \gamma Z + U_C$$

$$Y_0 = \beta_0 X + U_0$$

Observed Outcomes   Choice

$$Y = DY_1 + (1 - D)Y_0 \qquad S = Y_1 - Y_0 - C$$

$$D = \mathsf{I}[S > 0]$$

Features

Features

- *grmpy* is currently capable of the following features:

  - Simulating a dataset based on your own specifications.

  - Providing some useful information about the simulated dataset for instance:
    - Distributional outcome characteristics

    - ATE, TT, TUT

    - MTE by decile

  - Estimating the coefficients of interest given a dataset (of a specific form).

Install the package

- ► OS, Linux : Use the pip install manager (*pip install grmpy*) or download the package via GitHub and install it manually.

- ► Windows: The same procedure as for Linux, OS but you have to verify that the numpy package is already installed on your machine.

Initialization file

- The initialization file provides the user the opportunity to specify all parameters of his/her model, for instance:

    - Simulation parameters (number of observations, name of the output files)

    - Estimation parameters (optimization algorithm, start values)

    - Optimization parameters

    - Coefficients and covariance parameters, dummy variables...

- Example

- for a detailed explanation see: *grmpy*-documentation

Simulation

- *grmpy.simulate():* :

    - Input: path of the initialization file.

    - The function returns a data frame based on your specifications and different output files.

        - The data set as a pickle and a txt file.

        - An Info file that provides the distributional characteristics of the data as well as information about the different treatment effects.

Estimation

- *grmpy.estimate():*

  - Input: path of the initialization file.

  - At the moment the estimation process is only capable of two different optimization algorithms:

    - Broyden Fletcher Goldfarb Shanno (BFGS) algorithm

    - Powell's conjugate direction method

- There are two different options for the start values that could be set in the initialization file:

  - *init:* The estimation process uses the coefficient values specified in the initialization file as the start values for the estimation process.

  - *auto:* The start values are determined via a simple OLS followed by a Probit regression for the choice indicator.

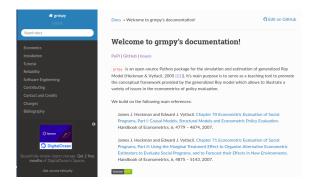- The estimation results are printed to an output file

Test battery

- We also provide a test battery that includes several tests to ensure that the processes perform as intended.

    - Property-based testing

    - Reliability testing

    - Regression testing

Application Example

Additional Information

Online documentation

Additional Information

- ***grmpy***-documentation
- Course material regarding the generalized Roy model
- GitHub Repository

Appendix

References

Heckman, J. J., & Vytlacil, E. J. (2007a). Econometric evaluation of social programs, part i: Causal effects, structural models and econometric policy evaluation. In J. J. Heckman & E. E. Leamer (Eds.), *Handbook of econometrics* (Vol. 6B, pp. 4779–4874). Amsterdam, Netherlands: Elsevier Science.

Heckman, J. J., & Vytlacil, E. J. (2007b). Econometric evaluation of social programs, part ii: Using the marginal treatment effect to organize alternative economic estimators to evaluate social programs and to forecast their effects in new environments. In J. J. Heckman & E. E. Leamer (Eds.), *Handbook of econometrics* (Vol. 6B, pp. 4875–5144). Amsterdam, Netherlands: Elsevier Science.