hregmonkey: Be A Happy Regression Monkey! (v1.0.3)

Weiwei Zheng （郑维伟）

Antai College of Economics and Management
Shanghai Jiao Tong University

2024 年 12 月 27 日

hregmonkey: Be A Happy Regression Monkey!

hregmonkey is a command for automating multiple regression analyses.

- "**h**" means "**hi**", "**hello**", "**happy**", "**high-dimension**", or even "**hate**" multiple regressions.
- Utilizes the **reghdfe** command for high-dimensional fixed effects.
- Print Standard academic style results table.
- Exports results via **esttab** to RTF files.
- Click **here** to visit the package **webpage** for more details.
- Click **here** to view the package **introduction slides**.
- Click **here** to watch the package **demonstration video**.

hregmonkey Stata Syntax (v1.0.3)

```
hregmonkey [if] [in], yvars(varlist) xvars(varlist)
    [cvars(varlist) idvar(varname) timevar(varname)
    absorb(absvars) groupvar(groupvar) indvar(indvar)
    aggregation(string) vcetype(vcetype) betadot(#)
    sedot(#) elsedot(#) quietly timer noestprint
    subfolder(string) save replace]
```

Regression Variable Options

yvars(varlist)

- Specify the list of dependent variables, recommended to use global variables.

xvars(varlist)

- Specify the list of independent variables, recommended to use global variables.

cvars(varlist)

- List of control variables, recommended to use global variables.

Cont': **hregmonkey** Stata Syntax

Fixed Effects and Clustering Options

idvar(varname)

- Individual identifier variable, default is ***id***.

timevar(varname)

- Time identifier variable, default is ***time***.

absorb(absvars)

- Absorption variables for fixed effects, default is ***i.id i.time***.

vcetype(vcetype)

- Specify one of three types of standard error in the report.

Report Format Options

betadot(#)

- Set the number of decimal places for regression coefficients, default is ***4***.

sedot(#)

- Set the number of decimal places for standard errors, default is ***4***.

elsedot(#)

- Set the number of decimal places for other parameters, default is ***4***.

Cont': **hregmonkey** Stata Syntax

Regression Result Reporting Options

quietly

- Run multiple regressions quietly.

timer

- Show **start**, **elapse**, and **finish times** by stage of computation.

noestprint

- **Hide** the standard regression results table. Default is **show**.

Output Configuration Options

subfolder(string)

- Specify the subfolder where output files will be saved.

save

- Export results to RTF files.

replace

- Overwrite existing RTF files.

Syntax & Options Description

Syntax

hregmonkey [*if*] [*in*], **yvars**(*varlist*) **xvars**(*varlist*) [**cvars**(*varlist*) **idvar**(*varname*) **timevar**(*varname*) **absorb**(*absvars*) **groupvar**(*groupvar*) **indvar**(*indvar*) **aggregation**(*string*) **vcetype**(*vcetype*) **betadot**(#) **sedot**(#) **elsedot**(#) **quietly** **timer** **noestprint** **subfolder**(*string*) **save** **replace**]

options	Description
Regression Variables	
yvars(*varlist*)	List of dependent variables. Suggest using syntax "***global yvarlist "y1 y2"***" to specify.
xvars(*varlist*)	List of independent variables. Suggest using syntax "***global xvarlist "x1 x2"***" to specify.
cvars(*varlist*)	List of control variables. Suggest using syntax "***global cvarlist "cv1 cv2"***" to specify.
Fixed Effects and Clustering	
idvar(*varname*)	Name of the individual identifier variable for fixed effects. Default is "***id***", which defined by command **tsset** or **xtset**.
timevar(*varname*)	Name of the time identifier variable for fixed effects. Default is "***time***", which defined by command **tsset** or **xtset**.
absorb(*absvars*)	Absorption variables for fixed effects. Default is "***i.id i.time***". Options include "***i.id***", "***i.time***", "***i.id#c.time***", "***i.id##c.time***", etc.
groupvar(*groupvar*)	Categorical variable representing each ***group***.
indvar(*indvar*)	Categorical variable representing each ***individual***.
aggregation(*string*)	Method of aggregation for the individual components of the group fixed effects. Valid options are ***mean*** (default), and ***sum***.
vcetype(*vcetype*)	Specify one of the following **three types** of standard error in the report. ***Notice***: it is not recommended to run clustered SEs if any of the clustering variables have too few d levels. A frequent rule of thumb is that **each cluster variable must have at least 50 different categories**.
vcetype(*unadjusted|ols*)	Estimate conventional standard errors, valid under the assumptions of **homoscedasticity** and **no correlation between observations** even in small samples.
vcetype(*robust*)	Estimate **heteroscedasticity consistent** standard errors (**Huber/White/Sandwich estimators**), which still assume **independence between observations**.
vcetype(*cluster clustervars*)	Estimate **consistent** standard errors even when the observations are **correlated within groups**, which allows multi-way-clustering. ***1. vcetype(cluster var1 var2)*** allows for intragroup correlation across individuals, time, country, etc. For instance, **vcetype**(*cluster firm year*) estimates SEs with firm and year clustering ({it:{opt two-way clusteri ***2.*** Interactions of the type ***vcetype(cluster var1#var2)*** i.e. where all observations of a given firm and year are clustered together. ***vcetype(cluster firm#year)*** estimates SEs with one-way clustering (***one-way clustering***).
Report Style	
betadot(#)	Set the number of decimal places displayed for regression coefficients. Default is **betadot**(*4*).
sedot(#)	Set the number of decimal places displayed for standard errors. Default is **sedot**(*4*).
elsedot(#)	Set the number of decimal places displayed for other parameters. Default is **elsedot**(*4*).
Print Regression Results	
quietly	Quietly run multiple regressions with command **reghdfe**.
timer	Show **start**, **elapse**, and **finish** times by stage of computation.
noestprint	Hide the standard regression results table. Default is "**show**".
Output Configuration	
subfolder(*string*)	Take effect with option "**save**". Specify where output RTF files will be saved. Default is folder "***result***" in the "***cd***" path. Support multi-level folders syntax (e.g. "***my_folder***", "***result\my_folder***", both "\" and "/" are fine).
save	Take effect after install package **estout**. Support **hregmonkey** to export results to RTF files in "***Chinese***" with the **esttab** command.
replace	Take effect with option "**save**". Overwrite existing RTF files when exporting results repeatedly.

Install Package and Dependencies

Install Package hregmonkey from **Weiwei Zheng**'s Personal Website

Note: More features will be added in the future, please update regularly to ensure a better user experience.

- `net install hregmonkey, replace from("https://weiweizheng.eu.org/uploads/hregmonkey/")`
- `help hregmonkey`

Install Dependencies (Skip if has installed before)

- `ssc install reghdfe, replace`
- `ssc install estout, replace`

Examples Setup & Data Description

Setup Test Environment

- `sysuse auto, clear`
- `gen long city = ceil(_n/20)`
- `gen long village = ceil(_n/10)`
- `gen long id = ceil(_n/5)`
- `gen year = mod(_n-1, 5) + 2001`
- `xtset id year`
- `sort city village id year`

Declare Global Variables

- `global y1 "weight length"`
- `global y2 "price"`
- `global x1 "mpg"`
- `global x2 "trunk turn"`
- `global cv1 "headroom gear_ratio"`
- `global cv2 "rep78"`

Multiple Regressions Demo

Demo 1: Basic Regression

Run regressions for "**y1**" on "**x1**" with controls "**cv1**", default "**id**" and "**year**" fixed-effects, cluster by "**id**".

- `hregmonkey, y($y1) x($x1) cv($cv1)`

Demo 2: Quietly Run Regression with Clustering

Quietly run above regressions, cluster by "**city**", report regression tables.

- `hregmonkey, y($y1) x($x1) cv($cv1) clus(city) q`

Demo 3: Conditional Regression

Run regressions only for observations where "**foreign==0**", specify "**village**" and "**year**" fixed-effects, cluster by "**city**", and do not report regression tables.

- `hregmonkey if foreign==0, y($y1 $y2) x($x1) cv($cv1) i(village) t(year) clus(city) noest`

Cont': Multiple Regressions Demo

Demo 4: Default Subfolder for Output

Run above regressions with the default subfolder ("***result***") for output files.

- `hregmonkey if foreign==0, y($y1 $y2) x($x1) cv($cv1) i(village) t(year) clus(city) noest s r`

Demo 5: User-Specific Subfolder

Run above regressions with a user specific subfolder (e.g. "***my_folder***") or multiple-level subfolder (e.g. "***result/my_folder***", both "\" and "/" are fine) for output files.

- `hregmonkey if foreign==0, y($y1 $y2) x($x1) cv($cv1) i(village) t(year) clus(city) noest sub(my_folder) s r`

Demo 6: Multiple Regressions

Run regressions for "**y1**" and "**y2**" on "**x1**" and "**x2**" with controls "**cv1**" and "**cv2**", specify "**village**" and "**year**" fixed-effects, cluster by "**city**", and report regression tables.

- `hregmonkey, y($y1 $y2) x($x1 $x2) cv($cv1 $cv2) i(village) t(year) clus(city) sub(result/my_folder) save replace`

Author Information

- Author: **Weiwei ZHENG (郑维伟)**
- E-mail: `vivid_zheng@126.com`
- Homepage: `https://weiweizheng.eu.org`
- Research Fields: Regional Economics, Industrial Economics, Spatial Econometrics

Author

Weiwei ZHENG **(郑维伟)**

Homepage: https://weiweizheng.eu.org

I am a Ph.D. candidate of Applied Economics at Antai College of Economics and Management, Shanghai Jiao Tong University.

My research fields include regional economics, industrial economics, and spatial econometrics theory & application. My interests focus on talent agglomeration, knowledge spillover, peer effect, R&D manipulation, firm innovation, total fa ductivity, etc.

I operate the 计量经济理论与应用研究室 "***Econometric Theory & Application Seminar***" (*https://etaseminar.eu.org*).

I designed the best academic navigation webpage, "***ETASeminar***" (*https://vividzheng.eu.org*), to assist your academic research!

Click for my ENG CV or CHN version.

For academic cooperation, contact via E-mail: *vivid_zheng@126.com* or *etaseminar@163.com*.

Thanks

Weiwei ZHENG (郑维伟)

2024 年 12 月 27 日