

# Package ‘gravity’

December 14, 2017

**Type** Package

**Title** Estimation Methods for Gravity Models

**Version** 0.6

**Date** 2017-12-14

**Author** Anna-

Lena Woelwer <woelwer@uni-trier.de>, Jan Pablo Burgard <burgardj@uni-trier.de>

**Maintainer** Anna-Lena Woelwer <woelwer@uni-trier.de>

**Description** Gravity models are used to explain bilateral flows related to the sizes of bilateral partners, a measure of distance between them and other influences on interaction costs. The underlying idea is rather simple. The greater the masses of two bodies and the smaller the distance between them, the stronger their attraction. This concept is applied to several research topics such as trade, migration or foreign direct investment. Even though their basic idea is rather simple, gravity models can become complex regarding the choice of models or estimation methods. As especially for gravity beginners it is difficult to get an overview of the different methods and implement them in R, the package provides a wrapper of different standard estimation methods for gravity models. By considering the descriptions and codes of the estimation methods, users get a comprehensive and application-oriented overview of the different methods, see which method may be suitable for a certain research question or type of data and extend the code available to their research projects. The package contains two types of estimation methods: those estimating gravity models in their log-log form, such as Ordinary Least Squares (OLS), Fixed Effects, Double Demeaning (DDM), Bonus vetus OLS with simple averages (BVU) and with GDP-weights (BVW), Structural Iterated Least Squares (SILS), Tetrads, and different Tobit models (Tobit, ET-Tobit, EK-Tobit). These methods are partly complex to understand and program and thus a comparison of them is not straightforward, wherefore the package aims at easing an overview of the different methods combined with a direct application. A second type of estimation methods utilizes the gravity model in its multiplicative form with a log-link and different families of distributions. They are relatively easy to compute, added for completeness and contain the methods Poisson Pseudo Maximum Likelihood (PPML), Gamma Pseudo Maximum Likelihood (GPML), Negative Binomial Pseudo Maximum Likelihood (NBPML), and Nonlinear Least Squares (NLS). All functions named estimate gravity models, but they differ in whether they estimate them in their multiplicative or additive form, their requirements and assumptions with respect to data, their handling of Multilateral Resistance terms as well as their possibilities concerning the inclusion of unilateral independent vari-

ables. Therefore, they normally lead to different estimation results. We refer the user to the Gravity Cookbook website (<https://sites.google.com/site/hiegravity/>) for more information on gravity models in general. Head, K. and Mayer, T. (2014) <DOI:10.1016/B978-0-444-54314-1.00003-3> provide a comprehensive and accessible overview of the theoretical and empirical development of the gravity literature as well as the use of gravity models and the various estimation methods, especially their merits and potential problems regarding applicability as well as different datasets. In order to have a straightforward application of all methods, the package comes with two example dataset, one including and the other one excluding zero trade flows. Examples utilizing these dataset are included in the description of the methods. As the range of gravity models and their suitable estimation methods is huge and specific to data and research question at hand, the functions are kept simple such that researchers can utilize them as a starting point of their research and get familiar with them. On the Gravity Cookbook website (<https://sites.google.com/site/hiegravity/>) Keith Head and Thierry Mayer provide Stata code for most methods. Where possible, the functions are tested to lead the same results as the Stata code when choosing the option of robust variance estimation. However, compared to the Stata code available, the functions presented in this package provide users with more flexibility regarding the type of estimation, the number and type of independent variables as well as the possible data. Furthermore, additional functions are added. The functions of the package include distance as an independent variable. If a panel model is used time-invariant effects, such as geographical distance, may have to be excluded; see the respective descriptions. Therefore, it is up to the user to ensure that the functions can be applied to panel data. Depending on the panel dataset and the variables - specifically the type of fixed effects - included in the model, it may easily occur that the model is not computable. Depending on the specific model, the code of the respective function may have to be changed in order to exclude the distance variable from the estimation. For a comprehensive overview of gravity models for panel data see Egger, P., & Pfaffermayr, M. (2003) <DOI:10.1007/s001810200146>, Gomez-Herrera, E. (2013) <DOI:10.1007/s00181-012-0576-2> and Head, K., Mayer, T., & Ries, J. (2010) <DOI:10.1016/j.jinteco.2010.01.002> as well as the references therein (see also the references included in the descriptions of the different functions). At the very least, the user should take special care with respect to the meaning of the estimated coefficients and variances as well as the decision about which effects to include in the estimation. As, to our knowledge at the moment, there is no explicit literature covering the estimation of a gravity equation by Double Demeaning, Structural Iterated Least Squares or Bonus vetus OLS using panel data, we do not recommend to apply these methods in this case. Contributions, extensions and error corrections are very welcome. Please do not hesitate to contact us. We thank Martin Bresslein for valuable comments and advice on early versions of the package.

**Imports** sandwich, multiwayvcov, lmtest, glm2, MASS, stats, censReg, survival

**License** GPL-2

**Encoding** UTF-8

**Depends** R(>= 3.1.1)

**RoxygenNote** 6.0.1

**LazyData** true

**NeedsCompilation** no

Repository CRAN

Date/Publication 2017-12-14 16:25:32 UTC

**R topics documented:**

BVU . . . . .	3
BVW . . . . .	6
DDM . . . . .	8
EK_Tobit . . . . .	10
ET_Tobit . . . . .	13
Fixed_Effects . . . . .	16
GPML . . . . .	18
Gravity_no_zeros . . . . .	21
Gravity_zeros . . . . .	23
NBPML . . . . .	24
NLS . . . . .	27
OLS . . . . .	29
PPML . . . . .	32
SILS . . . . .	35
Tetrads . . . . .	38
Tobit . . . . .	40
<b>Index</b>	<b>44</b>

---

BVU	<i>Bonus vetus OLS, BVU</i>
-----	-----------------------------

---

**Description**

BVU estimates gravity models via Bonus vetus OLS with simple averages.

**Usage**

```
BVU(y, dist, x, inc_o, inc_d, vce_robust = TRUE, data, ...)
```

**Arguments**

y	name (type: character) of the dependent variable in the dataset data, e.g. trade flows. This dependent variable is divided by the product of unilateral incomes (named inc_o and inc_d, e.g. GDPs or GNPs of the countries of interest) and logged afterwards. The transformed variable is then used as the dependent variable in the estimation.
dist	name (type: character) of the distance variable in the dataset data containing a measure of distance between all pairs of bilateral partners. It is logged automatically when the function is executed.

<code>x</code>	vector of names (type: character) of those bilateral variables in the dataset <code>data</code> that should be taken as the independent variables in the estimation. If an independent variable is a dummy variable, it should be of type numeric (0/1) in the dataset. If an independent variable is defined as a ratio, it should be logged. Unilateral metric variables such as GDPs should be inserted via the arguments <code>inc_o</code> for the country of origin and <code>inc_d</code> for the country of destination. As country specific effects are subdued due to demeaning, no further unilateral variables apart from <code>inc_o</code> and <code>inc_d</code> can be added.
<code>inc_o</code>	variable name (type: character) of the income of the country of origin in the dataset <code>data</code> . The dependent variable <code>y</code> is divided by the product of the incomes <code>inc_d</code> and <code>inc_o</code> .
<code>inc_d</code>	variable name (type: character) of the income of the country of destination in the dataset <code>data</code> . The dependent variable <code>y</code> is divided by the product of the incomes <code>inc_d</code> and <code>inc_o</code> .
<code>vce_robust</code>	robust (type: logic) determines whether a robust variance-covariance matrix should be used. The default is set to TRUE. If set TRUE the estimation results are consistent with the Stata code provided at the website <a href="#">Gravity Equations: Workhorse, Toolkit, and Cookbook</a> when choosing robust estimation.
<code>data</code>	name of the dataset to be used (type: character). To estimate gravity equations, a square gravity dataset including bilateral flows defined by the argument <code>y</code> , ISO-codes of type character (called <code>iso_o</code> for the country of origin and <code>iso_d</code> for the destination country), a distance measure defined by the argument <code>dist</code> and other potential influences given as a vector in <code>x</code> are required. All dummy variables should be of type numeric (0/1). Missing trade flows as well as incomplete rows should be excluded from the dataset. Furthermore, flows equal to zero should be excluded as the gravity equation is estimated in its additive form. As, to our knowledge at the moment, there is no explicit literature covering the estimation of a gravity equation by BVU using panel data, cross-sectional data should be used.
<code>...</code>	additional arguments to be passed to BVU.

## Details

Bonus `vetus` OLS is an estimation method for gravity models developed by Baier and Bergstrand (2009, 2010) using simple averages to center a Taylor-series (see the references for more information). To execute the function a square gravity dataset with all pairs of countries, ISO-codes for the country of origin and destination, a measure of distance between the bilateral partners as well as all information that should be considered as dependent and independent variables is needed. Make sure the ISO-codes are of type "character". Missing bilateral flows as well as incomplete rows should be excluded from the dataset. Furthermore, flows equal to zero should be excluded as the gravity equation is estimated in its additive form. The BVU function considers Multilateral Resistance terms and allows to conduct comparative statics. Country specific effects are subdued due to demeaning. Hence, unilateral variables apart from `inc_o` and `inc_d` cannot be included in the estimation. BVU is designed to be consistent with the Stata code provided at the website [Gravity Equations: Workhorse, Toolkit, and Cookbook](#) when choosing robust estimation. As, to our knowledge at the moment, there is no explicit literature covering the estimation of a gravity equation by BVU using panel data, we do not recommend to apply this method in this case.

**Value**

The function returns the summary of the estimated gravity model as an [lm](#)-object.

**References**

For estimating gravity equations via Bonus Vetus OLS see

Baier, S. L. and Bergstrand, J. H. (2009) <DOI:10.1016/j.jinteco.2008.10.004>

Baier, S. L. and Bergstrand, J. H. (2010) in Van Bergeijk, P. A., & Brakman, S. (Eds.) (2010) chapter 4 <DOI:10.1111/j.1467-9396.2011.01000.x>

For more information on gravity models, theoretical foundations and estimation methods in general see

Anderson, J. E. (1979) <DOI:10.12691/wjssh-2-2-5>

Anderson, J. E. (2010) <DOI:10.3386/w16576>

Anderson, J. E. and van Wincoop, E. (2003) <DOI:10.3386/w8079>

Head, K., Mayer, T., & Ries, J. (2010) <DOI:10.1016/j.jinteco.2010.01.002>

Head, K. and Mayer, T. (2014) <DOI:10.1016/B978-0-444-54314-1.00003-3>

Santos-Silva, J. M. C. and Tenreyro, S. (2006) <DOI:10.1162/rest.88.4.641>

and the citations therein.

See [Gravity Equations: Workhorse, Toolkit, and Cookbook](#) for gravity datasets and Stata code for estimating gravity models.

**See Also**

[lm](#), [coefstest](#), [vcovHC](#)

**Examples**

```
## Not run:
data(Gravity_no_zeros)

BVU(y="flow", dist="distw", x=c("rta"),
inc_o="gdp_o", inc_d="gdp_d", vce_robust=TRUE, data=Gravity_no_zeros)

BVU(y="flow", dist="distw", x=c("rta", "contig", "comcur"),
inc_o="gdp_o", inc_d="gdp_d", vce_robust=TRUE, data=Gravity_no_zeros)

## End(Not run)
```

---

 BVW

*Bonus vetus OLS, BVW*


---

### Description

BVW estimates gravity models via Bonus vetus OLS with GDP-weights.

### Usage

```
BVW(y, dist, x, inc_o, inc_d, vce_robust = TRUE, data, ...)
```

### Arguments

<code>y</code>	name (type: character) of the dependent variable in the dataset <code>data</code> , e.g. trade flows. This dependent variable is divided by the product of unilateral incomes (named <code>inc_o</code> and <code>inc_d</code> , e.g. GDPs or GNPs of the countries of interest) and logged afterwards. The transformed variable is then used as the dependent variable in the estimation.
<code>dist</code>	name (type: character) of the distance variable in the dataset <code>data</code> containing a measure of distance between all pairs of bilateral partners. It is logged automatically when the function is executed.
<code>x</code>	vector of names (type: character) of those bilateral variables in the dataset <code>data</code> that should be taken as the independent variables in the estimation. If an independent variable is a dummy variable, it should be of type numeric (0/1) in the dataset. If an independent variable is defined as a ratio, it should be logged. Unilateral metric variables such as GDPs should be inserted via the arguments <code>inc_o</code> for the country of origin and <code>inc_d</code> for the country of destination. As country specific effects are subdued due to demeaning, no further unilateral variables apart from <code>inc_o</code> and <code>inc_d</code> can be added.
<code>inc_o</code>	variable name (type: character) of the income of the country of origin in the dataset <code>data</code> . The dependent variable <code>y</code> is divided by the product of the incomes <code>inc_d</code> and <code>inc_o</code> .
<code>inc_d</code>	variable name (type: character) of the income of the country of destination in the dataset <code>data</code> . The dependent variable <code>y</code> is divided by the product of the incomes <code>inc_d</code> and <code>inc_o</code> .
<code>vce_robust</code>	robust (type: logic) determines whether a robust variance-covariance matrix should be used. The default is set to TRUE. If set TRUE the estimation results are consistent with the Stata code provided at the website <a href="#">Gravity Equations: Workhorse, Toolkit, and Cookbook</a> when choosing robust estimation.
<code>data</code>	name of the dataset to be used (type: character). To estimate gravity equations, a square gravity dataset including bilateral flows defined by the argument <code>y</code> , ISO-codes of type character (called <code>iso_o</code> for the country of origin and <code>iso_d</code> for the destination country), a distance measure defined by the argument <code>dist</code> and other potential influences given as a vector in <code>x</code> are required. All dummy variables should be of type numeric (0/1). Missing trade flows as well as incomplete rows

should be excluded from the dataset. Furthermore, flows equal to zero should be excluded as the gravity equation is estimated in its additive form. As, to our knowledge at the moment, there is no explicit literature covering the estimation of a gravity equation by BVW using panel data, cross-sectional data should be used.

... additional arguments to be passed to BVW.

## Details

Bonus vetus OLS is an estimation method for gravity models developed by Baier and Bergstrand (2009, 2010) using GDP-weights to center a Taylor-series (see the references for more information). To execute the function a square gravity dataset with all pairs of countries, ISO-codes for the country of origin and destination, a measure of distance between the bilateral partners as well as all information that should be considered as dependent an independent variables is needed. Make sure the ISO-codes are of type "character". Missing bilateral flows as well as incomplete rows should be excluded from the dataset. Furthermore, flows equal to zero should be excluded as the gravity equation is estimated in its additive form. The BVW function considers Multilateral Resistance terms and allows to conduct comparative statics. Country specific effects are subdued due to demeaning. Hence, unilateral variables apart from `inc_o` and `inc_d` cannot be included in the estimation. BVW is designed to be consistent with the Stata code provided at the website [Gravity Equations: Workhorse, Toolkit, and Cookbook](#) when choosing robust estimation. As, to our knowledge at the moment, there is no explicit literature covering the estimation of a gravity equation by BVW using panel data, we do not recommend to apply this method in this case.

## Value

The function returns the summary of the estimated gravity model as an `lm`-object.

## References

For estimating gravity equations via Bonus Vetus OLS see

Baier, S. L. and Bergstrand, J. H. (2009) <DOI:10.1016/j.jinteco.2008.10.004>

Baier, S. L. and Bergstrand, J. H. (2010) in Van Bergeijk, P. A., & Brakman, S. (Eds.) (2010) chapter 4 <DOI:10.1111/j.1467-9396.2011.01000.x>

For more information on gravity models, theoretical foundations and estimation methods in general see

Anderson, J. E. (1979) <DOI:10.12691/wjssh-2-2-5>

Anderson, J. E. (2010) <DOI:10.3386/w16576>

Anderson, J. E. and van Wincoop, E. (2003) <DOI:10.3386/w8079>

Head, K., Mayer, T., & Ries, J. (2010) <DOI:10.1016/j.jinteco.2010.01.002>

Head, K. and Mayer, T. (2014) <DOI:10.1016/B978-0-444-54314-1.00003-3>

Santos-Silva, J. M. C. and Tenreyro, S. (2006) <DOI:10.1162/rest.88.4.641>

and the citations therein.

See [Gravity Equations: Workhorse, Toolkit, and Cookbook](#) for gravity datasets and Stata code for estimating gravity models.

**See Also**

[lm](#), [coefstest](#), [vcovHC](#)

**Examples**

```
## Not run:
data(Gravity_no_zeros)

BVW(y="flow", dist="distw", x=c("rta"),
inc_o="gdp_o", inc_d="gdp_d", vce_robust=TRUE, data=Gravity_no_zeros)

BVW(y="flow", dist="distw", x=c("rta", "comcur", "contig"),
inc_o="gdp_o", inc_d="gdp_d", vce_robust=TRUE, data=Gravity_no_zeros)

## End(Not run)
```

---

DDM

---

*Double Demeaning, DDM*


---

**Description**

DDM estimates gravity models via double demeaning the left hand side and right hand side of the gravity equation.

**Usage**

```
DDM(y, dist, x, vce_robust = TRUE, data, ...)
```

**Arguments**

y	name (type: character) of the dependent variable in the dataset data, e.g. trade flows. This variable is logged and then used as the dependent variable in the estimation.
dist	name (type: character) of the distance variable in the dataset data containing a measure of distance between all pairs of bilateral partners. It is logged automatically when the function is executed.
x	vector of names (type: character) of those bilateral variables in the dataset data that should be taken as the independent variables in the estimation. If an independent variable is a dummy variable, it should be of type numeric (0/1) in the dataset. If an independent variable is defined as a ratio, it should be logged. Unilateral effects drop out due to double demeaning and therefore cannot be estimated.

<code>vce_robust</code>	robust (type: logic) determines whether a robust variance-covariance matrix should be used. The default is set to TRUE. If set TRUE the estimation results are consistent with the Stata code provided at the website <a href="#">Gravity Equations: Workhorse, Toolkit, and Cookbook</a> when choosing robust estimation.
<code>data</code>	name of the dataset to be used (type: character). To estimate gravity equations, a square gravity dataset including bilateral flows defined by the argument <code>y</code> , ISO-codes of type character (called <code>iso_o</code> for the country of origin and <code>iso_d</code> for the destination country), a distance measure defined by the argument <code>dist</code> and other potential influences given as a vector in <code>x</code> are required. All dummy variables should be of type numeric (0/1). Missing trade flows as well as incomplete rows should be excluded from the dataset. Furthermore, flows equal to zero should be excluded as the gravity equation is estimated in its additive form. As, to our knowledge at the moment, there is no explicit literature covering the estimation of a gravity equation by DDM using panel data, cross-sectional data should be used.
<code>...</code>	additional arguments to be passed to DDM.

### Details

DDM is an estimation method for gravity models presented in Head and Mayer (2014) (see the references for more information). To execute the function a square gravity dataset with all pairs of countries, ISO-codes for the country of origin and destination, a measure of distance between the bilateral partners as well as all information that should be considered as dependent an independent variables is needed. Make sure the ISO-codes are of type "character". Missing bilateral flows as well as incomplete rows should be excluded from the dataset. Furthermore, flows equal to zero should be excluded as the gravity equation is estimated in its additive form. Country specific effects are subdued due double demeaning. Hence, unilateral income proxies such as GDP cannot be considered as exogenous variables.

DDM is designed to be consistent with the Stata code provided at the website [Gravity Equations: Workhorse, Toolkit, and Cookbook](#) when choosing robust estimation. As, to our knowledge at the moment, there is no explicit literature covering the estimation of a gravity equation by DDM using panel data, we do not recommend to apply this method in this case.

### Value

The function returns the summary of the estimated gravity model as an `lm`-object.

### References

For more information on Double Demeaning as well as information on gravity models, theoretical foundations and estimation methods in general see

Head, K. and Mayer, T. (2014) <DOI:10.1016/B978-0-444-54314-1.00003-3>

as well as

Anderson, J. E. (1979) <DOI:10.12691/wjssh-2-2-5>

Anderson, J. E. (2010) <DOI:10.3386/w16576>

Anderson, J. E. and van Wincoop, E. (2003) <DOI:10.3386/w8079>

Baier, S. L. and Bergstrand, J. H. (2009) <DOI:10.1016/j.jinteco.2008.10.004>

Baier, S. L. and Bergstrand, J. H. (2010) in Van Bergeijk, P. A., & Brakman, S. (Eds.) (2010) chapter 4 <DOI:10.1111/j.1467-9396.2011.01000.x>

Head, K., Mayer, T., & Ries, J. (2010) <DOI:10.1016/j.jinteco.2010.01.002>

Santos-Silva, J. M. C. and Tenreyro, S. (2006) <DOI:10.1162/rest.88.4.641>

and the citations therein.

See [Gravity Equations: Workhorse, Toolkit, and Cookbook](#) for gravity datasets and Stata code for estimating gravity models.

### See Also

[lm](#), [coefstest](#), [vcovHC](#)

### Examples

```
## Not run:
data(Gravity_no_zeros)

DDM(y="flow", dist="distw", x=c("rta"),
vce_robust=TRUE, data=Gravity_no_zeros)

DDM(y="flow", dist="distw", x=c("rta", "comcur", "contig"),
vce_robust=TRUE, data=Gravity_no_zeros)

## End(Not run)
```

---

EK\_Tobit

*Eaton and Kortum (2001) Tobit model, EK\_Tobit*

---

### Description

EK\_Tobit estimates gravity models in their additive form by conducting a censored regression. It follows the Eaton and Kortum (2001) Tobit model where each country is assigned specific censoring bounds.

### Usage

```
EK_Tobit(y, dist, x, vce_robust = TRUE, data, ...)
```

**Arguments**

<code>y</code>	name (type: character) of the dependent variable in the dataset data, e.g. trade flows. The variable is logged and then taken as the dependent variable in the regression. As the log of zero is not defined, all flows equal to zero are replaced by a left open interval with the logged minimum trade flow of the respective importing country as right border.
<code>dist</code>	name (type: character) of the distance variable in the dataset data containing a measure of distance between all pairs of bilateral partners. It is logged automatically when the function is executed.
<code>x</code>	vector of names (type: character) of those bilateral variables in the dataset data that should be taken as the independent variables in the estimation. If an independent variable is a dummy variable, it should be of type numeric (0/1) in the dataset. If an independent variable is defined as a ratio, it should be logged. Unilateral variables such as country dummies or incomes can be added. If unilateral metric variables such as GDPs should be used as independent variables, those variables have to be logged first and the logged variable can be used in <code>x</code> . Interaction terms can be added.
<code>vce_robust</code>	robust (type: logic) determines whether a robust variance-covariance matrix should be used. The default is set to TRUE. If set TRUE the estimation results are consistent with the Stata code provided at the website <a href="#">Gravity Equations: Workhorse, Toolkit, and Cookbook</a> when choosing robust estimation.
<code>data</code>	name of the dataset to be used (type: character). To estimate gravity equations, a square gravity dataset including bilateral flows defined by the argument <code>y</code> , ISO-codes of type character (called <code>iso_o</code> for the country of origin and <code>iso_d</code> for the destination country), a distance measure defined by the argument <code>dist</code> and other potential influences given as a vector in <code>x</code> are required. All dummy variables should be of type numeric (0/1). Missing trade flows as well as incomplete rows should be excluded from the dataset. Zero trade flows are allowed.
<code>...</code>	additional arguments to be passed to <code>EK_Tobit</code> .

**Details**

`EK_Tobit` represents the Eaton and Kortum (2001) Tobit model. When taking the log of the gravity equation flows equal to zero constitute a problem as their log is not defined. Therefore, in `EK_Tobit` all values of the dependent variable are redefined as intervals. The positive observations have both interval bounds equal to their original value. For zero flows the interval is left open. The right border of the interval is set to the log of the minimum positive trade flow of the respective importing country. The defined data object of class `Surv` is then inserted in `survreg` for the parameter estimation.

To execute the function a square gravity dataset with all pairs of countries, ISO-codes for the country of origin and destination, a measure of distance between the bilateral partners as well as all information that should be considered as dependent and independent variables is needed. Missing bilateral flows as well as incomplete rows should be excluded from the dataset. Zero trade flows are allowed.

`EK_Tobit` is designed to be consistent with the Stata code provided at the website [Gravity Equations: Workhorse, Toolkit, and Cookbook](#) when choosing robust estimation.

Up to now, the function is designed for cross-sectional data, but can be extended to panel data using the [survreg](#) function.

For other Tobit functions, see [Tobit](#) for a simple Tobit model where number 1 is added to all observations and [ET\\_Tobit](#) for the Eaton and Tamura (1994) threshold Tobit model where instead of simply adding number 1 to the data the threshold is estimated.

### Value

The function returns the summary of the estimated gravity model as a [survreg](#)-object.

### References

For more information on gravity models, theoretical foundations and estimation methods in general see

Anderson, J. E. (1979) <DOI:10.12691/wjssh-2-2-5>

Anderson, J. E. (2010) <DOI:10.3386/w16576>

Anderson, J. E. and van Wincoop, E. (2003) <DOI:10.3386/w8079>

Baier, S. L. and Bergstrand, J. H. (2009) <DOI:10.1016/j.jinteco.2008.10.004>

Baier, S. L. and Bergstrand, J. H. (2010) in Van Bergeijk, P. A., & Brakman, S. (Eds.) (2010) chapter 4 <DOI:10.1111/j.1467-9396.2011.01000.x>

Head, K., Mayer, T., & Ries, J. (2010) <DOI:10.1016/j.jinteco.2010.01.002>

Head, K. and Mayer, T. (2014) <DOI:10.1016/B978-0-444-54314-1.00003-3>

Santos-Silva, J. M. C. and Tenreyro, S. (2006) <DOI:10.1162/rest.88.4.641>

and the citations therein.

Especially for Tobit models see

Tobin, J. (1958) <DOI:10.2307/1907382>

Eaton, J., & Tamura, A. (1994) <DOI:10.3386/w4758>

Eaton, J., & Kortum, S. (2001) <DOI:10.3386/w8070>.

See Carson, R. T., & Sun, Yixiao (2007) <DOI:10.1111/j.1368-423X.2007.00218.x> for the estimation of the threshold in a Tobit model.

See [Gravity Equations: Workhorse, Toolkit, and Cookbook](#) for gravity datasets and Stata code for estimating gravity models.

### See Also

[Surv](#), [survreg](#)

### Examples

```
## Not run:
# Example for data with zero trade flows
data(Gravity_zeros)

Gravity_zeros$lgdp_o <- log(Gravity_zeros$gdp_o)
Gravity_zeros$lgdp_d <- log(Gravity_zeros$gdp_d)
```

```

EK_Tobit(y="flow", dist="distw", x=c("rta","lgdp_o","lgdp_d"),
vce_robust=TRUE, data=Gravity_zeros)

EK_Tobit(y="flow", dist="distw", x=c("rta","iso_o","iso_d"),
vce_robust=TRUE, data=Gravity_zeros)

## End(Not run)

```

---

ET\_Tobit

*Eaton and Tamura (1994) threshold Tobit model, ET\_Tobit*


---

### Description

ET\_Tobit estimates gravity models in their additive form by conducting a left-censored regression. It follows the Eaton and Tamura (1994) Tobit model, also called threshold Tobit model, where, instead of adding number 1 to the dependent variable as done in [Tobit](#), the constant added to the data is estimated and interpreted as a threshold. For estimating this threshold, we follow Carson and Sun (2007).

### Usage

```
ET_Tobit(y, dist, x, data, ...)
```

### Arguments

- |      |  |
|------|--|
| y    | name (type: character) of the dependent variable in the dataset data, e.g. trade flows. Following Carson and Sun (2007), the smallest positive flow value is used as an estimate of the threshold. It is added to y, the transformed variable is logged and taken as the dependent variable in the Tobit estimation with lower bound equal to the log of the smallest possible flow value.   |
| dist | name (type: character) of the distance variable in the dataset data containing a measure of distance between all pairs of bilateral partners. It is logged automatically when the function is executed.  |
| x    | vector of names (type: character) of those bilateral variables in the dataset data that should be taken as the independent variables in the estimation. If an independent variable is a dummy variable, it should be of type numeric (0/1) in the dataset. If an independent variable is defined as a ratio, it should be logged. Unilateral variables such as country dummies or incomes can be added. If unilateral metric variables such as GDPs should be used as independent variables, those variables have to be logged first and the logged variable can be used in x. Interaction terms can be added. |

`data` name of the dataset to be used (type: character). To estimate gravity equations, a square gravity dataset including bilateral flows defined by the argument `y`, ISO-codes of type character (called `iso_o` for the country of origin and `iso_d` for the destination country), a distance measure defined by the argument `dist` and other potential influences given as a vector in `x` are required. All dummy variables should be of type numeric (0/1). Missing trade flows as well as incomplete rows should be excluded from the dataset. Zero trade flows are allowed.

`...` additional arguments to be passed to `ET_Tobit`.

## Details

`ET_Tobit` represents the Eaton and Tamura (1994) Tobit model which is often used when several gravity models are compared. When taking the log of the gravity equation flows equal to zero constitute a problem as their log is not defined. Therefore, a constant is added to the flows. This constant, opposed to `Tobit`, is estimated. Compared to the usual ET-Tobit approaches, in this package, the estimation of the threshold is done before the other parameters are estimated. We follow Carson and Sun (2007), who show that taking the minimum positive flow value as an estimate of the threshold is super-consistent and that using this threshold estimate ensures that the parameter MLE are asymptotically normal with the asymptotic variance identical to the variance achieved when the threshold is known. Hence, first the threshold is estimated as the minimum positive flow. This threshold is added to the flow variable. It is logged afterwards and taken as the dependent variable. The Tobit estimation is then conducted using the `censReg` function and setting the lower bound equal to the log of the minimum positive flow value which was added to all observations. A Tobit regression represents a combination of a binary and a linear regression. This procedure has to be taken into consideration when interpreting the estimated coefficients. The marginal effects of an explanatory variable on the expected value of the dependent variable equals the product of both the probability of the latent variable exceeding the threshold and the marginal effect of the explanatory variable of the expected value of the latent variable.

To execute the function a square gravity dataset with all pairs of countries, ISO-codes for the country of origin and destination, a measure of distance between the bilateral partners as well as all information that should be considered as dependent an independent variables is needed. Missing bilateral flows as well as incomplete rows should be excluded from the dataset. Zero trade flows are allowed.

Up to now, the function is designed for cross-sectional data, but can be easily extended to panel data using the `censReg` function. A robust estimations is not implemented to the present as the `censReg` function is not compatible with the `vcovHC` function.

For a more elaborate Tobit function, see `EK_Tobit` for the Eaton and Kortum (2001) Tobit model where each zero trade volume is assigned a country specific interval with the upper bound equal to the minimum positive trade level of the respective importing country.

## Value

The function returns the summary of the estimated gravity model as a `censReg`-object.

## References

For more information on gravity models, theoretical foundations and estimation methods in general see

Anderson, J. E. (1979) <DOI:10.12691/wjssh-2-2-5>  
 Anderson, J. E. (2010) <DOI:10.3386/w16576>  
 Anderson, J. E. and van Wincoop, E. (2003) <DOI:10.3386/w8079>  
 Baier, S. L. and Bergstrand, J. H. (2009) <DOI:10.1016/j.jinteco.2008.10.004>  
 Baier, S. L. and Bergstrand, J. H. (2010) in Van Bergeijk, P. A., & Brakman, S. (Eds.) (2010) chapter 4 <DOI:10.1111/j.1467-9396.2011.01000.x>  
 Head, K., Mayer, T., & Ries, J. (2010) <DOI:10.1016/j.jinteco.2010.01.002>  
 Head, K. and Mayer, T. (2014) <DOI:10.1016/B978-0-444-54314-1.00003-3>  
 Santos-Silva, J. M. C. and Tenreyro, S. (2006) <DOI:10.1162/rest.88.4.641>  
 and the citations therein.

Especially for Tobit models see

Tobin, J. (1958) <DOI:10.2307/1907382>  
 Eaton, J., & Tamura, A. (1994) <DOI:10.3386/w4758>  
 Eaton, J., & Kortum, S. (2001) <DOI:10.3386/w8070>.

See Carson, R. T., & Sun, Yixiao (2007) <DOI:10.1111/j.1368-423X.2007.00218.x> for the estimation of the threshold in a Tobit model.

See [Gravity Equations: Workhorse, Toolkit, and Cookbook](#) for gravity datasets and Stata code for estimating gravity models.

## See Also

[censReg](#)

## Examples

```
## Not run:
# Example for data with zero trade flows
data(Gravity_zeros)

Gravity_zeros$lgdp_o <- log(Gravity_zeros$gdp_o)
Gravity_zeros$lgdp_d <- log(Gravity_zeros$gdp_d)

ET_Tobit(y="flow", dist="distw", x=c("rta","lgdp_o","lgdp_d"),
data=Gravity_zeros)

## End(Not run)
```

---

Fixed_Effects	<i>Fixed_Effects</i>
---------------	----------------------

---

### Description

Fixed\_Effects estimates gravity models via OLS and fixed effects for the countries of origin and destination. These effects catch country specific effects.

### Usage

```
Fixed_Effects(y, dist, fe = c("iso_o", "iso_d"), x, vce_robust = TRUE, data,
...)
```

### Arguments

y	name (type: character) of the dependent variable in the dataset data, e.g. trade flows. This variable is logged and taken as the dependent variable in the estimation.
dist	name (type: character) of the distance variable in the dataset data containing a measure of distance between all pairs of bilateral partners. It is logged automatically when the function is executed.
fe	vector of names (type: character) of fixed effects. The default is set to the unilateral identifiers "iso_o" and "iso_d" for cross-sectional data. When using panel data, interaction terms of the iso-codes and time may be added in either fe or x.
x	vector of names (type: character) of those bilateral variables in the dataset data that should be taken as the independent variables in the estimation. If an independent variable is a dummy variable it should be of type numeric (0/1) in the dataset. If an independent variable is defined as a ratio, it should be logged. The fixed effects catch all unilateral effects. Therefore, no other unilateral variables such as GDP can be included as independent variables in the estimation.
vce_robust	robust (type: logic) determines whether a robust variance-covariance matrix should be used. The default is set to TRUE. If set TRUE the estimation results are consistent with the Stata code provided at the website <a href="#">Gravity Equations: Workhorse, Toolkit, and Cookbook</a> when choosing robust estimation.
data	name of the dataset to be used (type: character). To estimate gravity equations, a square gravity dataset including bilateral flows defined by the argument y, ISO-codes of type character (called iso_o for the country of origin and iso_d for the destination country), a distance measure defined by the argument dist and other potential influences given as a vector in x are required. All dummy variables should be of type numeric (0/1). Missing trade flows as well as incomplete rows should be excluded from the dataset. Furthermore, flows equal to zero should be excluded as the gravity equation is estimated in its additive form. When using panel data, a variable for the time may be included in the dataset. Note that the variable for the time dimension should be of type: factor. The time variable can

be used as a single dependent variable or interaction term with other variables such as country identifiers by inserting it into `x` or `fe`. See the references for more information on panel data.

... additional arguments to be passed to `Fixed_Effects`.

## Details

To account for MR terms, Feenstra (2002) and Feenstra (2004) propose to use importer and exporter fixed effects. Due to the use of these effects, all unilateral influences such as GDPs can no longer be estimated. A disadvantage of the use of `Fixed_Effects` is that, when applied to panel data, the number of country-year or country-pair fixed effects can be too high for estimation. In addition, no comparative statistics are possible with `Fixed_Effects` as the MR terms are not estimated explicitly. Nevertheless, Head and Mayer (2014) highlight the importance of the use of fixed effects. To execute the function a square gravity dataset with all pairs of countries, ISO-codes for the country of origin and destination, a measure of distance between the bilateral partners as well as all information that should be considered as dependent an independent variables is needed. Make sure the ISO-codes are of type "character". Missing bilateral flows as well as incomplete rows should be excluded from the dataset. Furthermore, flows equal to zero should be excluded as the gravity equation is estimated in its additive form. Country specific fixed effects are considered by incorporating "iso\_o" and "iso\_d" in `fe`. By including country specific fixed effects, all monadic effects are captured, including Multilateral Resistance terms. Therefore, no other unilateral variables such as GDP can be included as independent variables in the estimation.

`Fixed_Effects` estimation can be used for both, cross-sectional as well as panel data. Nonetheless, the function is designed to be consistent with the Stata code for cross-sectional data provided at the website [Gravity Equations: Workhorse, Toolkit, and Cookbook](#) when choosing robust estimation. The function `Fixed_Effects` was therefore tested for For the use with panel data no tests were performed. Therefore, it is up to the user to ensure that the functions can be applied to panel data. Depending on the panel dataset and the variables - specifically the type of fixed effects - included in the model, it may easily occur that the model is not computable. Also, note that by including bilateral fixed effects such as country-pair effects, the coefficients of time-invariant observables such as distance can no longer be estimated. Depending on the specific model, the code of the respective function may have to be changed in order to exclude the distance variable from the estimation. At the very least, the user should take special care with respect to the meaning of the estimated coefficients and variances as well as the decision about which effects to include in the estimation. When using panel data, the parameter and variance estimation of the models may have to be changed accordingly. For a comprehensive overview of gravity models for panel data see Egger and Pfaffermayr (2003), Gomez-Herrera (2013) and Head, Mayer and Ries (2010) as well as the references therein.

## Value

The function returns the summary of the estimated gravity model as an `lm`-object.

## References

For more information on fixed effects as well as information on gravity models, theoretical foundations and suitable estimation methods in general see

Anderson, J. E. (2010) <DOI:10.3386/w16576>

Head, K. and Mayer, T. (2014) <DOI:10.1016/B978-0-444-54314-1.00003-3>

as well as

Anderson, J. E. (1979) <DOI:10.12691/wjssh-2-2-5>

Anderson, J. E. and van Wincoop, E. (2003) <DOI:10.3386/w8079>

Baier, S. L. and Bergstrand, J. H. (2009) <DOI:10.1016/j.jinteco.2008.10.004>

Baier, S. L. and Bergstrand, J. H. (2010) in Van Bergeijk, P. A., & Brakman, S. (Eds.) (2010) chapter 4 <DOI:10.1111/j.1467-9396.2011.01000.x>

Head, K., Mayer, T., & Ries, J. (2010) <DOI:10.1016/j.jinteco.2010.01.002>

Santos-Silva, J. M. C. and Tenreyro, S. (2006) <DOI:10.1162/rest.88.4.641>

and the citations therein.

See [Gravity Equations: Workhorse, Toolkit, and Cookbook](#) for gravity datasets and Stata code for estimating gravity models.

For estimating gravity equations using panel data see

Egger, P., & Pfaffermayr, M. (2003) <DOI:10.1007/s001810200146>

Gomez-Herrera, E. (2013) <DOI:10.1007/s00181-012-0576-2>

and the references therein.

### See Also

[lm](#), [coefstest](#), [vcovHC](#)

### Examples

```
## Not run:
data(Gravity_no_zeros)

Fixed_Effects(y="flow", dist="distw", fe=c("iso_o", "iso_d"),
x=c("rta"), vce_robust=TRUE, data=Gravity_no_zeros)

Fixed_Effects(y="flow", dist="distw", fe=c("iso_o", "iso_d"),
x=c("rta", "comcur", "contig"), vce_robust=TRUE, data=Gravity_no_zeros)

## End(Not run)
```

---

GPML

*Gamma Pseudo Maximum Likelihood, GPML*

---

### Description

GPML estimates gravity models in their multiplicative form via Gamma Pseudo Maximum Likelihood.

**Usage**

```
GPML(y, dist, x, vce_robust = TRUE, data, ...)
```

**Arguments**

<code>y</code>	name (type: character) of the dependent variable in the dataset <code>data</code> , e.g. trade flows.
<code>dist</code>	name (type: character) of the distance variable in the dataset <code>data</code> containing a measure of distance between all pairs of bilateral partners. It is logged automatically when the function is executed.
<code>x</code>	vector of names (type: character) of those bilateral variables in the dataset <code>data</code> that should be taken as the independent variables in the estimation. If an independent variable is a dummy variable, it should be of type numeric (0/1) in the dataset. If an independent variable is defined as a ratio, it should be logged. Unilateral variables such as country dummies or incomes can be added. If unilateral metric variables such as GDPs should be used as independent variables, those variables have to be logged first and the logged variable can be used in <code>x</code> . Interaction terms can be added.
<code>vce_robust</code>	robust (type: logic) determines whether a robust variance-covariance matrix should be used. The default is set to TRUE.
<code>data</code>	name of the dataset to be used (type: character). To estimate gravity equations, a square gravity dataset including bilateral flows defined by the argument <code>y</code> , ISO-codes of type character (called <code>iso_o</code> for the country of origin and <code>iso_d</code> for the destination country), a distance measure defined by the argument <code>dist</code> and other potential influences given as a vector in <code>x</code> are required. All dummy variables should be of type numeric (0/1). Missing trade flows as well as incomplete rows should be excluded from the dataset. Flows equal to zero have to be excluded as the gamma distribution is used. When using panel data, a variable for the time may be included in the dataset. Note that the variable for the time dimension should be of type: factor. See the references for more information on panel data.
<code>...</code>	additional arguments to be passed to functions used by GPML.

**Details**

GPML is an estimation method for gravity models belonging to generalized linear models. It is estimated via [glm2](#) using the gamma distribution and a log-link. To execute the function a square gravity dataset with all pairs of countries, ISO-codes for the country of origin and destination, a measure of distance between the bilateral partners as well as all information that should be considered as dependent and independent variables is needed. Make sure the ISO-codes are of type "character". Missing bilateral flows as well as incomplete rows should be excluded from the dataset. Flows equal to zero have to be excluded as the gamma distribution is used. For similar functions, utilizing the multiplicative form via the log-link, but different distributions, see [PPML](#), [NLS](#), and [NBPML](#).

GPML estimation can be used for both, cross-sectional as well as panel data. It is up to the user to ensure that the functions can be applied to panel data. Depending on the panel dataset and the variables - specifically the type of fixed effects - included in the model, it may easily occur that the model is not computable. Also, note that by including bilateral fixed effects such as country-pair effects, the coefficients of time-invariant observables such as distance can no longer be estimated.

Depending on the specific model, the code of the respective function may have to be changed in order to exclude the distance variable from the estimation. At the very least, the user should take special care with respect to the meaning of the estimated coefficients and variances as well as the decision about which effects to include in the estimation. When using panel data, the parameter and variance estimation of the models may have to be changed accordingly. For a comprehensive overview of gravity models for panel data see Egger and Pfaffermayr (2003), Gomez-Herrera (2013) and Head, Mayer and Ries (2010) as well as the references therein.

### Value

The function returns the summary of the estimated gravity model similar to a `glm`-object.

### References

For more information on the estimation of gravity equations via Gamma Pseudo maximum Likelihood see

Santos-Silva, J. M. C. and Tenreyro, S. (2006) <DOI:10.1162/rest.88.4.641>

For more information on gravity models, theoretical foundations and estimation methods in general see

Anderson, J. E. (1979) <DOI:10.12691/wjssh-2-2-5>

Anderson, J. E. (2010) <DOI:10.3386/w16576>

Anderson, J. E. and van Wincoop, E. (2003) <DOI:10.3386/w8079>

Baier, S. L. and Bergstrand, J. H. (2009) <DOI:10.1016/j.jinteco.2008.10.004>

Baier, S. L. and Bergstrand, J. H. (2010) in Van Bergeijk, P. A., & Brakman, S. (Eds.) (2010) chapter 4 <DOI:10.1111/j.1467-9396.2011.01000.x>

Head, K., Mayer, T., & Ries, J. (2010) <DOI:10.1016/j.jinteco.2010.01.002>

Head, K. and Mayer, T. (2014) <DOI:10.1016/B978-0-444-54314-1.00003-3>

and the citations therein.

See [Gravity Equations: Workhorse, Toolkit, and Cookbook](#) for gravity datasets and Stata code for estimating gravity models.

For estimating gravity equations using panel data see

Egger, P., & Pfaffermayr, M. (2003) <DOI:10.1007/s001810200146>

Gomez-Herrera, E. (2013) <DOI:10.1007/s00181-012-0576-2>

and the references therein.

### See Also

[glm2](#), [coefstest](#), [vcovHC](#)

### Examples

```
## Not run:
data(Gravity_no_zeros)

Gravity_no_zeros$lgdp_o <- log(Gravity_no_zeros$gdp_o)
```

```
Gravity_no_zeros$lgdp_d <- log(Gravity_no_zeros$gdp_d)

GPML(y="flow", dist="distw", x=c("rta", "lgdp_o", "lgdp_d"),
vce_robust=TRUE, data=Gravity_no_zeros)

GPML(y="flow", dist="distw", x=c("rta", "iso_o", "iso_d"),
vce_robust=TRUE, data=Gravity_no_zeros)

## End(Not run)
```

---

Gravity_no_zeros	<i>Gravity dataset without zero trade flows</i>
------------------	---

---

### Description

An edited version of the full gravity dataset: The "square" gravity dataset for all possible pairs of Countries worldwide, 1948-2006, which is used in the article: Head, K., T. Mayer and J. Ries, 2010, "The erosion of colonial linkages after independence". Journal of International Economics, 81(1):1-14 (lead article). Dataset Gravity\_no\_zeros corresponds to the dataset without zero trade flows, Gravity\_zeros, on the other hand, includes zero trade flows.

### Usage

```
data(Gravity_no_zeros)
```

### Format

A data frame with 17088 rows and 10 variables:

**iso\_o** ISO-Code of country of origin  
**iso\_d** ISO-Code of country of destination  
**distw** weighted distance  
**gdp\_o** GDP of country of origin  
**gdp\_d** GDP of country of destination  
**rta** regional trade agreement  
**flow** trade flow  
**contig** contiguity  
**comlang\_off** common official language  
**comcur** common currency

## Details

An edited version of the full gravity dataset: The "square" gravity dataset for all possible pairs of Countries worldwide, 1948-2006, which is used in the article: Head, K., T. Mayer and J. Ries, 2010, "The erosion of colonial linkages after independence". *Journal of International Economics*, 81(1):1-14 (lead article).

In order to have a dataset suited for all functions, a cross-sectional dataset is chosen. All incomplete rows and observations with missing trade flows are therefore excluded from the dataset. As some of the functions in the package are capable of handling zero values in trade flows and some are not, two datasets, Gravity\_zeros and Gravity\_no\_zeros, are provided.

The original dataset downloaded at [http://econ.sciences-po.fr/sites/default/files/file/tmayer/data/col\\_regfile09.zip](http://econ.sciences-po.fr/sites/default/files/file/tmayer/data/col_regfile09.zip) was edited in the following way:

```
# Reading in the dataset
library(foreign)
col_regfile09 <- read.dta("col_regfile09.dta")
# Isolation of one year
data06 <- col_regfile09[col_regfile09$year == 2006,]
# Choosing variables (select columns)
data06 <- data06[,c(2, 3, 6, 8, 12, 27, 34, 4, 5, 29)]
# Transforming data
# Isolation of complete cases
data06 <- data06[complete.cases(data06) == TRUE,]
# Exclusion of trade flows equal to 0
Gravity_no_zeros <- data06[data06$flow != 0,]
row.names(Gravity_no_zeros) <- 1:length(row.names(Gravity_no_zeros))
# Divide GDPs by 1,000,000 for scaling purposes
Gravity_no_zeros$gdp_o <- Gravity_no_zeros$gdp_o / 1000000
Gravity_no_zeros$gdp_d <- Gravity_no_zeros$gdp_d / 1000000
```

## Source

<http://econ.sciences-po.fr/node/131> <https://sites.google.com/site/hiegravity/data-sources>

## References

Head, K. and Mayer, T. (2014). Chapter 3 - gravity equations: Workhorse, toolkit, and cookbook. In Gita Gopinath, E. H. and Rogoff, K., editors, *Handbook of International Economics*, volume 4 of *Handbook of International Economics*, pages 131-195. Elsevier. ([Gravity Equations: Workhorse, Toolkit, and Cookbook](#))

Head, K., T. Mayer and J. Ries, 2010, "The erosion of colonial linkages after independence". *Journal of International Economics*, 81(1):1-14 (lead article).

---

Gravity\_zeros

*Gravity dataset with zero trade flows*

---

### Description

An edited version of the full gravity dataset: The "square" gravity dataset for all possible pairs of Countries worldwide, 1948-2006, which is used in the article: Head, K., T. Mayer and J. Ries, 2010, "The erosion of colonial linkages after independence". Journal of International Economics, 81(1):1-14 (lead article). Dataset Gravity\_no\_zeros corresponds to the dataset without zero trade flows, Gravity\_zeros, on the other hand, includes zero trade flows.

### Usage

```
data(Gravity_zeros)
```

### Format

A data frame with 17088 rows and 10 variables:

**iso\_o** ISO-Code of country of origin

**iso\_d** ISO-Code of country of destination

**distw** weighted distance

**gdp\_o** GDP of country of origin

**gdp\_d** GDP of country of destination

**rta** regional trade agreement

**flow** trade flow

**contig** contiguity

**comlang\_off** common official language

**comcur** common currency

### Details

An edited version of the full gravity dataset: The "square" gravity dataset for all possible pairs of Countries worldwide, 1948-2006, which is used in the article: Head, K., T. Mayer and J. Ries, 2010, "The erosion of colonial linkages after independence". Journal of International Economics, 81(1):1-14 (lead article).

In order to have a dataset suited for all functions, a cross-sectional dataset is chosen. All incomplete rows and observations with missing trade flows are therefore excluded from the dataset. As some of the functions in the package are capable of handling zero values in trade flows and some are not, two datasets, Gravity\_zeros and Gravity\_no\_zeros, are provided.

The original dataset downloaded at [http://econ.sciences-po.fr/sites/default/files/file/tmayer/data/col\\_regfile09.zip](http://econ.sciences-po.fr/sites/default/files/file/tmayer/data/col_regfile09.zip) was edited in the following way:

```
# Reading in the dataset
```

```

library(foreign)
col_regfile09 <- read.dta("col_regfile09.dta")
# Isolation of one year
data06 <- col_regfile09[col_regfile09$year == 2006,]
# Choosing variables (select columns)
data06 <- data06[,c(2, 3, 6, 8, 12, 27, 34, 4, 5, 29)]
# Transforming data
# Isolation of complete cases
data06 <- data06[complete.cases(data06) == TRUE,]
Gravity_zeros <- data06
row.names(Gravity_zeros) <- 1:length(row.names(Gravity_zeros))
# Divide GDPs by 1,000,000 for scaling purposes
Gravity_zeros$gdp_o <- Gravity_zeros$gdp_o / 1000000
Gravity_zeros$gdp_d <- Gravity_zeros$gdp_d / 1000000

```

### Source

<http://econ.sciences-po.fr/node/131> <https://sites.google.com/site/hiegravity/data-sources>

### References

Head, K. and Mayer, T. (2014). Chapter 3 - gravity equations: Workhorse, toolkit, and cookbook. In Gita Gopinath, E. H. and Rogoff, K., editors, Handbook of International Economics, volume 4 of Handbook of International Economics, pages 131-195. Elsevier. ([Gravity Equations: Workhorse, Toolkit, and Cookbook](#))

Head, K., T. Mayer and J. Ries, 2010, "The erosion of colonial linkages after independence". Journal of International Economics, 81(1):1-14 (lead article).

---

NBPML

*Negative Binomial Pseudo Maximum Likelihood, NBPML*

---

### Description

NBPML estimates gravity models in their multiplicative form via Negative Binomial Pseudo Maximum Likelihood.

### Usage

```
NBPML(y, dist, x, vce_robust = TRUE, data, ...)
```

**Arguments**

<code>y</code>	name (type: character) of the dependent variable in the dataset <code>data</code> , e.g. trade flows.
<code>dist</code>	name (type: character) of the distance variable in the dataset <code>data</code> containing a measure of distance between all pairs of bilateral partners. It is logged automatically when the function is executed.
<code>x</code>	vector of names (type: character) of those bilateral variables in the dataset <code>data</code> that should be taken as the independent variables in the estimation. If an independent variable is a dummy variable, it should be of type numeric (0/1) in the dataset. If an independent variable is defined as a ratio, it should be logged. Unilateral variables such as country dummies or incomes can be added. If unilateral metric variables such as GDPs should be used as independent variables, those variables have to be logged first and the logged variable can be used in <code>x</code> . Interaction terms can be added.
<code>vce_robust</code>	robust (type: logic) determines whether a robust variance-covariance matrix should be used. The default is set to TRUE.
<code>data</code>	name of the dataset to be used (type: character). To estimate gravity equations, a square gravity dataset including bilateral flows defined by the argument <code>y</code> , ISO-codes of type character (called <code>iso_o</code> for the country of origin and <code>iso_d</code> for the destination country), a distance measure defined by the argument <code>dist</code> and other potential influences given as a vector in <code>x</code> are required. All dummy variables should be of type numeric (0/1). Missing trade flows as well as incomplete rows should be excluded from the dataset. Zero trade flows are allowed. When using panel data, a variable for the time may be included in the dataset. Note that the variable for the time dimension should be of type: factor. See the references for more information on panel data.
<code>...</code>	additional arguments to be passed to functions used by NBPML.

**Details**

NBPML is an estimation method for gravity models belonging to generalized linear models. It is estimated via `glm.nb` using the negative binomial distribution and a log-link. To execute the function a square gravity dataset with all pairs of countries, ISO-codes for the country of origin and destination, a measure of distance between the bilateral partners as well as all information that should be considered as dependent and independent variables is needed. Make sure the ISO-codes are of type "character". Missing bilateral flows as well as incomplete rows should be excluded from the dataset. Zero trade flows are allowed. For similar functions, utilizing the multiplicative form via the log-link, but different distributions, see [PPML](#), [GPML](#), and [NLS](#).

NBPML estimation can be used for both, cross-sectional as well as panel data. It is up to the user to ensure that the functions can be applied to panel data. Depending on the panel dataset and the variables - specifically the type of fixed effects - included in the model, it may easily occur that the model is not computable. Also, note that by including bilateral fixed effects such as country-pair effects, the coefficients of time-invariant observables such as distance can no longer be estimated. Depending on the specific model, the code of the respective function may have to be changed in order to exclude the distance variable from the estimation. At the very least, the user should take special care with respect to the meaning of the estimated coefficients and variances as well as the decision

about which effects to include in the estimation. When using panel data, the parameter and variance estimation of the models may have to be changed accordingly. For a comprehensive overview of gravity models for panel data see Egger and Pfaffermayr (2003), Gomez-Herrera (2013) and Head, Mayer and Ries (2010) as well as the references therein.

### Value

The function returns the summary of the estimated gravity model similar to a `glm`-object.

### References

For more information on gravity models, theoretical foundations and estimation methods in general see

Anderson, J. E. (1979) <DOI:10.12691/wjssh-2-2-5>

Anderson, J. E. (2010) <DOI:10.3386/w16576>

Anderson, J. E. and van Wincoop, E. (2003) <DOI:10.3386/w8079>

Baier, S. L. and Bergstrand, J. H. (2009) <DOI:10.1016/j.jinteco.2008.10.004>

Baier, S. L. and Bergstrand, J. H. (2010) in Van Bergeijk, P. A., & Brakman, S. (Eds.) (2010) chapter 4 <DOI:10.1111/j.1467-9396.2011.01000.x>

Head, K., Mayer, T., & Ries, J. (2010) <DOI:10.1016/j.jinteco.2010.01.002>

Head, K. and Mayer, T. (2014) <DOI:10.1016/B978-0-444-54314-1.00003-3>

Santos-Silva, J. M. C. and Tenreyro, S. (2006) <DOI:10.1162/rest.88.4.641>

and the citations therein.

See [Gravity Equations: Workhorse, Toolkit, and Cookbook](#) for gravity datasets and Stata code for estimating gravity models.

For estimating gravity equations using panel data see

Egger, P., & Pfaffermayr, M. (2003) <DOI:10.1007/s001810200146>

Gomez-Herrera, E. (2013) <DOI:10.1007/s00181-012-0576-2>

and the references therein.

### See Also

[glm.nb](#), [coefstest](#), [vcovHC](#)

### Examples

```
## Not run:
# Example for data with zero trade flows
data(Gravity_zeros)

NBPML(y="flow", dist="distw", x=c("rta","iso_o","iso_d"),
vce_robust=TRUE, data=Gravity_zeros)

# Example for data without zero trade flows
data(Gravity_no_zeros)
```

```
Gravity_no_zeros$lgdp_o <- log(Gravity_no_zeros$gdp_o)
Gravity_no_zeros$lgdp_d <- log(Gravity_no_zeros$gdp_d)

NBPML(y="flow", dist="distw", x=c("rta", "lgdp_o", "lgdp_d"),
vce_robust=TRUE, data=Gravity_no_zeros)

## End(Not run)
```

---

NLS

*Nonlinear Least Squares, NLS*


---

### Description

NLS estimates gravity models in their multiplicative form via Nonlinear Least Squares.

### Usage

```
NLS(y, dist, x, vce_robust = TRUE, data, ...)
```

### Arguments

<code>y</code>	name (type: character) of the dependent variable in the dataset <code>data</code> , e.g. trade flows.
<code>dist</code>	name (type: character) of the distance variable in the dataset <code>data</code> containing a measure of distance between all pairs of bilateral partners. It is logged automatically when the function is executed.
<code>x</code>	vector of names (type: character) of those bilateral variables in the dataset <code>data</code> that should be taken as the independent variables in the estimation. If an independent variable is a dummy variable, it should be of type numeric (0/1) in the dataset. If an independent variable is defined as a ratio, it should be logged. Unilateral variables such as country dummies or incomes can be added. If unilateral metric variables such as GDPs should be used as independent variables, those variables have to be logged first and the logged variable can be used in <code>x</code> . Interaction terms can be added.
<code>vce_robust</code>	robust (type: logic) determines whether a robust variance-covariance matrix should be used. The default is set to TRUE.
<code>data</code>	name of the dataset to be used (type: character). To estimate gravity equations, a square gravity dataset including bilateral flows defined by the argument <code>y</code> , ISO-codes of type character (called <code>iso_o</code> for the country of origin and <code>iso_d</code> for the destination country), a distance measure defined by the argument <code>dist</code> and other potential influences given as a vector in <code>x</code> are required. All dummy variables should be of type numeric (0/1). Missing trade flows as well as incomplete rows should be excluded from the dataset. Zero trade flows are allowed. When using panel data, a variable for the time may be included in the dataset. Note that the

variable for the time dimension should be of type: factor. See the references for more information on panel data.

... additional arguments to be passed to functions used by NLS.

## Details

NLS is an estimation method for gravity models belonging to generalized linear models. It is estimated via `glm` using the gaussian distribution and a log-link. As the method may not lead to convergence when poor starting values are used, the linear predictions, fitted values, and estimated coefficients resulting from a PPML estimation are used for the arguments `etastart`, `mustart`, and `start`. To execute the function a square gravity dataset with all pairs of countries, ISO-codes for the country of origin and destination, a measure of distance between the bilateral partners as well as all information that should be considered as dependent an independent variables is needed. Missing bilateral flows as well as incomplete rows should be excluded from the dataset. Zero trade flows are allowed. For similar functions, utilizing the multiplicative form via the log-link, but different distributions, see `PPML`, `GPML`, and `NBPML`.

NLS estimation can be used for both, cross-sectional as well as panel data. It is up to the user to ensure that the functions can be applied to panel data. Depending on the panel dataset and the variables - specifically the type of fixed effects - included in the model, it may easily occur that the model is not computable. Also, note that by including bilateral fixed effects such as country-pair effects, the coefficients of time-invariant observables such as distance can no longer be estimated. Depending on the specific model, the code of the respective function may have to be changed in order to exclude the distance variable from the estimation. At the very least, the user should take special care with respect to the meaning of the estimated coefficients and variances as well as the decision about which effects to include in the estimation. When using panel data, the parameter and variance estimation of the models may have to be changed accordingly. For a comprehensive overview of gravity models for panel data see Egger and Pfaffermayr (2003), Gomez-Herrera (2013) and Head, Mayer and Ries (2010) as well as the references therein.

## Value

The function returns the summary of the estimated gravity model similar to a `glm`-object.

## References

For more information on gravity models, theoretical foundations and estimation methods in general see

Anderson, J. E. (1979) <DOI:10.12691/wjssh-2-2-5>

Anderson, J. E. (2010) <DOI:10.3386/w16576>

Anderson, J. E. and van Wincoop, E. (2003) <DOI:10.3386/w8079>

Baier, S. L. and Bergstrand, J. H. (2009) <DOI:10.1016/j.jinteco.2008.10.004>

Baier, S. L. and Bergstrand, J. H. (2010) in Van Bergeijk, P. A., & Brakman, S. (Eds.) (2010) chapter 4 <DOI:10.1111/j.1467-9396.2011.01000.x>

Head, K., Mayer, T., & Ries, J. (2010) <DOI:10.1016/j.jinteco.2010.01.002>

Head, K. and Mayer, T. (2014) <DOI:10.1016/B978-0-444-54314-1.00003-3>

Santos-Silva, J. M. C. and Tenreyro, S. (2006) <DOI:10.1162/rest.88.4.641>

and the citations therein.

See [Gravity Equations: Workhorse, Toolkit, and Cookbook](#) for gravity datasets and Stata code for estimating gravity models.

For estimating gravity equations using panel data see

Egger, P., & Pfaffermayr, M. (2003) <DOI:10.1007/s001810200146>

Gomez-Herrera, E. (2013) <DOI:10.1007/s00181-012-0576-2>

and the references therein.

### See Also

[glm](#), [coefstest](#), [vcovHC](#)

### Examples

```
## Not run:
# Example for data with zero trade flows
data(Gravity_zeros)

Gravity_zeros$lgdp_o <- log(Gravity_zeros$gdp_o)
Gravity_zeros$lgdp_d <- log(Gravity_zeros$gdp_d)

NLS(y="flow", dist="distw", x=c("rta","lgdp_o","lgdp_d"),
    vce_robust=TRUE, data=Gravity_zeros)

# Example for data without zero trade flows
data(Gravity_no_zeros)

Gravity_no_zeros$lgdp_o <- log(Gravity_no_zeros$gdp_o)
Gravity_no_zeros$lgdp_d <- log(Gravity_no_zeros$gdp_d)

NLS(y="flow", dist="distw", x=c("rta","lgdp_o","lgdp_d"),
    vce_robust=TRUE, data=Gravity_no_zeros)

## End(Not run)
```

---

OLS

*OLS*

---

### Description

OLS estimates gravity models in their traditional form via Ordinary Least Squares (OLS). It does not consider Multilateral Resistance terms.

### Usage

```
OLS(y, dist, x, inc_d, inc_o, uie = FALSE, vce_robust = TRUE, data, ...)
```

**Arguments**

<code>y</code>	name (type: character) of the dependent variable in the dataset data, e.g. trade flows. This variable is logged and taken as the dependent variable in the estimation. If <code>uie=TRUE</code> the dependent variable is divided by the product of unilateral incomes <code>inc_o</code> and <code>inc_d</code> , e.g. GDPs or GNPs of the countries of interest and logged afterwards. If <code>uie=FALSE</code> the dependent variable is logged directly. The transformed variable is then used as the dependent variable and the logged income variables are used as independent variables in the estimation.
<code>dist</code>	name (type: character) of the distance variable in the dataset data containing a measure of distance between all pairs of bilateral partners. It is logged automatically when the function is executed.
<code>x</code>	vector of names (type: character) of those bilateral variables in the dataset data that should be taken as the independent variables in the estimation. If an independent variable is a dummy variable, it should be of type numeric (0/1) in the dataset. If an independent variable is defined as a ratio, it should be logged. Unilateral metric variables such as GDPs should be inserted via the arguments <code>inc_o</code> for the country of origin and <code>inc_d</code> for the country of destination. Interaction terms can be added.
<code>inc_d</code>	variable name (type: character) of the income of the country of destination in the dataset data. If <code>uie=TRUE</code> , the dependent variable <code>y</code> is divided by the product of the incomes <code>inc_d</code> and <code>inc_o</code> . If <code>uie=FALSE</code> , the incomes are logged and taken as independent variables in the estimation. If one wants to use more than one unilateral variable, e.g. GDP and population, those variables have to be merged into one variable, e.g. GDP per capita, which can be inserted into <code>inc_d</code> .
<code>inc_o</code>	variable name (type: character) of the income of the country of origin in the dataset data. If <code>uie=TRUE</code> , the dependent variable <code>y</code> is divided by the product of the incomes <code>inc_d</code> and <code>inc_o</code> . If <code>uie=FALSE</code> , the incomes are logged and taken as independent variables in the estimation. If one wants to use more than one unilateral variable, e.g. GDP and population, those variables have to be merged into one variable, e.g. GDP per capita, which can be inserted into <code>inc_o</code> .
<code>uie</code>	Unitary Income Elasticities (type: logic) determines whether the parameters are to be estimated assuming unitary income elasticities. The default value is set to <code>FALSE</code> . If <code>uie</code> is set <code>TRUE</code> , the flows in the dependent variable <code>y</code> are divided by the product of the country pairs' incomes before the estimation. If <code>uie</code> is set to <code>FALSE</code> , the income variables are logged and taken as independent variables in the estimation. The variable names for the incomes should be inserted into <code>inc_o</code> for the country of origin and into <code>inc_d</code> for destination country.
<code>vce_robust</code>	<code>robust</code> (type: logic) determines whether a robust variance-covariance matrix should be used. The default is set to <code>TRUE</code> . If set <code>TRUE</code> the estimation results are consistent with the Stata code provided at the website <a href="#">Gravity Equations: Workhorse, Toolkit, and Cookbook</a> when choosing robust estimation.
<code>data</code>	name of the dataset to be used (type: character). To estimate gravity equations, a square gravity dataset including bilateral flows defined by the argument <code>y</code> , ISO-codes of type character (called <code>iso_o</code> for the country of origin and <code>iso_d</code> for the destination country), a distance measure defined by the argument <code>dist</code> and other potential influences given as a vector in <code>x</code> are required. All dummy variables

should be of type numeric (0/1). Missing trade flows as well as incomplete rows should be excluded from the dataset. Furthermore, flows equal to zero should be excluded as the gravity equation is estimated in its additive form. When using panel data, a variable for the time may be included in the dataset. Note that the variable for the time dimension should be of type: factor. See the references for more information on panel data.

... additional arguments to be passed to OLS.

## Details

OLS estimates gravity models in their traditional, additive, form via Ordinary Least Squares using the `lm` function. Multilateral Resistance terms are not considered by this function. To execute the function a square gravity dataset with all pairs of countries, ISO-codes for the country of origin and destination, a measure of distance between the bilateral partners as well as all information that should be considered as dependent an independent variables is needed. Make sure the ISO-codes are of type "character". Missing bilateral flows as well as incomplete rows should be excluded from the dataset. Furthermore, flows equal to zero should be excluded as the gravity equation is estimated in its additive form. As the coefficients for the country's incomes were often found to be close to unitary and unitary income elasticities are in line with some theoretical foundations on international trade, it is sometimes assumed that the income elasticities are equal to unity. In order to allow for the estimation with and without the assumption of unitary income elasticities, the option `uie` is built into OLS with the default set to FALSE.

OLS estimation can be used for both, cross-sectional and panel data. Nonetheless, the function is designed to be consistent with the Stata code for cross-sectional data provided at the website [Gravity Equations: Workhorse, Toolkit, and Cookbook](#) when choosing robust estimation. The function OLS was therefore tested for cross-sectional data. For the use with panel data no tests were performed. Therefore, it is up to the user to ensure that the functions can be applied to panel data. Depending on the panel dataset and the variables - specifically the type of fixed effects - included in the model, it may easily occur that the model is not computable. Also, note that by including bilateral fixed effects such as country-pair effects, the coefficients of time-invariant observables such as distance can no longer be estimated. Depending on the specific model, the code of the respective function may have to be changed in order to exclude the distance variable from the estimation. At the very least, the user should take special care with respect to the meaning of the estimated coefficients and variances as well as the decision about which effects to include in the estimation. When using panel data, the parameter and variance estimation of the models may have to be changed accordingly. For a comprehensive overview of gravity models for panel data see Egger and Pfaffermayr (2003), Gomez-Herrera (2013) and Head, Mayer and Ries (2010) as well as the references therein.

## Value

The function returns the summary of the estimated gravity model as an `lm`-object.

## References

For more information on gravity models, theoretical foundations and estimation methods in general see

Anderson, J. E. (1979) <DOI:10.12691/wjssh-2-2-5>

Anderson, J. E. (2010) <DOI:10.3386/w16576>

Anderson, J. E. and van Wincoop, E. (2003) <DOI:10.3386/w8079>  
 Baier, S. L. and Bergstrand, J. H. (2009) <DOI:10.1016/j.jinteco.2008.10.004>  
 Baier, S. L. and Bergstrand, J. H. (2010) in Van Bergeijk, P. A., & Brakman, S. (Eds.) (2010) chapter 4 <DOI:10.1111/j.1467-9396.2011.01000.x>  
 Head, K., Mayer, T., & Ries, J. (2010) <DOI:10.1016/j.jinteco.2010.01.002>  
 Head, K. and Mayer, T. (2014) <DOI:10.1016/B978-0-444-54314-1.00003-3>  
 Santos-Silva, J. M. C. and Tenreyro, S. (2006) <DOI:10.1162/rest.88.4.641>  
 and the citations therein.

See [Gravity Equations: Workhorse, Toolkit, and Cookbook](#) for gravity datasets and Stata code for estimating gravity models.

For estimating gravity equations using panel data see

Egger, P., & Pfaffermayr, M. (2003) <DOI:10.1007/s001810200146>

Gomez-Herrera, E. (2013) <DOI:10.1007/s00181-012-0576-2>

and the references therein.

### See Also

[lm](#), [coefstest](#), [vcovHC](#)

### Examples

```
## Not run:
data(Gravity_no_zeros)

OLS(y="flow", dist="distw", x=c("rta", "contig", "comcur"),
inc_o="gdp_o", inc_d="gdp_d", uie=FALSE,
vce_robust=TRUE, data=Gravity_no_zeros)

OLS(y="flow", dist="distw", x=c("rta", "comcur", "contig"),
inc_o="gdp_o", inc_d="gdp_d", uie=TRUE,
vce_robust=TRUE, data=Gravity_no_zeros)

## End(Not run)
```

### Description

PPML estimates gravity models in their multiplicative form via Poisson Pseudo Maximum Likelihood.

**Usage**

```
PPML(y, dist, x, vce_robust = TRUE, data, ...)
```

**Arguments**

<code>y</code>	name (type: character) of the dependent variable in the dataset <code>data</code> , e.g. trade flows.
<code>dist</code>	name (type: character) of the distance variable in the dataset <code>data</code> containing a measure of distance between all pairs of bilateral partners. It is logged automatically when the function is executed.
<code>x</code>	vector of names (type: character) of those bilateral variables in the dataset <code>data</code> that should be taken as the independent variables in the estimation. If an independent variable is a dummy variable, it should be of type numeric (0/1) in the dataset. If an independent variable is defined as a ratio, it should be logged. Unilateral variables such as country dummies or incomes can be added. If unilateral metric variables such as GDPs should be used as independent variables, those variables have to be logged first and the logged variable can be used in <code>x</code> . Interaction terms can be added.
<code>vce_robust</code>	<code>robust</code> (type: logic) determines whether a robust variance-covariance matrix should be used. The default is set to <code>TRUE</code> . If set <code>TRUE</code> the estimation results are consistent with the Stata code provided at the website <a href="#">Gravity Equations: Workhorse, Toolkit, and Cookbook</a> when choosing robust estimation.
<code>data</code>	name of the dataset to be used (type: character). To estimate gravity equations, a square gravity dataset including bilateral flows defined by the argument <code>y</code> , ISO-codes of type character (called <code>iso_o</code> for the country of origin and <code>iso_d</code> for the destination country), a distance measure defined by the argument <code>dist</code> and other potential influences given as a vector in <code>x</code> are required. All dummy variables should be of type numeric (0/1). Missing trade flows as well as incomplete rows should be excluded from the dataset. Zero trade flows are allowed. When using panel data, a variable for the time may be included in the dataset. Note that the variable for the time dimension should be of type: factor. See the references for more information on panel data.
<code>...</code>	additional arguments to be passed to functions used by PPML.

**Details**

PPML is an estimation method for gravity models belonging to generalized linear models. It is estimated via `glm` using the quasipoisson distribution and a log-link. PPML is presented in Santos-Silva and Tenreyro (2006) (see the references for more information). To execute the function a square gravity dataset with all pairs of countries, ISO-codes for the country of origin and destination, a measure of distance between the bilateral partners as well as all information that should be considered as dependent and independent variables is needed. Make sure the ISO-codes are of type "character". Missing bilateral flows as well as incomplete rows should be excluded from the dataset. Zero trade flows are allowed. For similar functions, utilizing the multiplicative form via the log-link, but different distributions, see [GPML](#), [NLS](#), and [NBPML](#).

PPML estimation can be used for both, cross-sectional as well as panel data. The function is designed to be consistent with the results from the Stata function `ppml` written by J. M. C. Santos-Silva and

S. Tenreyro. The function PPML was therefore tested for cross-sectional data. For the use with panel data no tests were performed. Therefore, it is up to the user to ensure that the functions can be applied to panel data. Depending on the panel dataset and the variables - specifically the type of fixed effects - included in the model, it may easily occur that the model is not computable. Also, note that by including bilateral fixed effects such as country-pair effects, the coefficients of time-invariant observables such as distance can no longer be estimated. Depending on the specific model, the code of the respective function may have to be changed in order to exclude the distance variable from the estimation. At the very least, the user should take special care with respect to the meaning of the estimated coefficients and variances as well as the decision about which effects to include in the estimation. When using panel data, the parameter and variance estimation of the models may have to be changed accordingly. For a comprehensive overview of gravity models for panel data see Egger and Pfaffermayr (2003), Gomez-Herrera (2013) and Head, Mayer and Ries (2010) as well as the references therein.

### Value

The function returns the summary of the estimated gravity model as an `glm`-object.

### References

For more information on the estimation of gravity equations via Poisson Pseudo maximum Likelihood see

Santos-Silva, J. M. C. and Tenreyro, S. (2006) <DOI:10.1162/rest.88.4.641>

For more information on gravity models, theoretical foundations and estimation methods in general see

Anderson, J. E. (1979) <DOI:10.12691/wjssh-2-2-5>

Anderson, J. E. (2010) <DOI:10.3386/w16576>

Anderson, J. E. and van Wincoop, E. (2003) <DOI:10.3386/w8079>

Baier, S. L. and Bergstrand, J. H. (2009) <DOI:10.1016/j.jinteco.2008.10.004>

Baier, S. L. and Bergstrand, J. H. (2010) in Van Bergeijk, P. A., & Brakman, S. (Eds.) (2010) chapter 4 <DOI:10.1111/j.1467-9396.2011.01000.x>

Head, K., Mayer, T., & Ries, J. (2010) <DOI:10.1016/j.jinteco.2010.01.002>

Head, K. and Mayer, T. (2014) <DOI:10.1016/B978-0-444-54314-1.00003-3>

and the citations therein.

See [Gravity Equations: Workhorse, Toolkit, and Cookbook](#) for gravity datasets and Stata code for estimating gravity models.

For estimating gravity equations using panel data see

Egger, P., & Pfaffermayr, M. (2003) <DOI:10.1007/s001810200146>

Gomez-Herrera, E. (2013) <DOI:10.1007/s00181-012-0576-2>

and the references therein.

### See Also

[glm](#), [coeftest](#), [vcovHC](#)

**Examples**

```
## Not run:
# Example for data with zero trade flows
data(Gravity_zeros)

PPML(y="flow", dist="distw", x=c("rta","iso_o","iso_d"),
vce_robust=TRUE, data=Gravity_zeros)

# Example for data without zero trade flows
data(Gravity_no_zeros)

Gravity_no_zeros$lgdp_o <- log(Gravity_no_zeros$gdp_o)
Gravity_no_zeros$lgdp_d <- log(Gravity_no_zeros$gdp_d)

PPML(y="flow", dist="distw", x=c("rta","lgdp_o","lgdp_d"),
vce_robust=TRUE, data=Gravity_no_zeros)

## End(Not run)
```

---

SILS

*Structural Iterated Least Squares, SILS*


---

**Description**

SILS estimates gravity models via Structural Iterated Least Squares and an explicit inclusion of the Multilateral Resistance terms.

**Usage**

```
SILS(y, dist, x, inc_o, inc_d, maxloop = 50, maxloop2 = 50,
dec_places = 4, vce_robust = TRUE, verbose = FALSE, data, ...)
```

**Arguments**

<code>y</code>	name (type: character) of the dependent variable in the dataset <code>data</code> , e.g. trade flows. This dependent variable is divided by the product of unilateral incomes ( <code>inc_o</code> and <code>inc_d</code> , e.g. GDPs or GNPs of the countries of interest) and logged afterwards. The transformed variable is then used as the dependent variable in the estimation.
<code>dist</code>	name (type: character) of the distance variable in the dataset <code>data</code> containing a measure of distance between all pairs of bilateral partners. It is logged automatically when the function is executed.

<code>x</code>	vector of names (type: character) of those bilateral variables in the dataset <code>data</code> that should be taken as the independent variables in the estimation. If an independent variable is a dummy variable, it should be of type numeric (0/1) in the dataset. If an independent variable is defined as a ratio, it should be logged. Unilateral metric variables such as GDPs should be inserted via the arguments <code>inc_o</code> for the country of origin and <code>inc_d</code> for the country of destination.
<code>inc_o</code>	variable name (type: character) of the income of the country of origin in the dataset <code>data</code> . The dependent variable <code>y</code> is divided by the product of the incomes <code>inc_d</code> and <code>inc_o</code> .
<code>inc_d</code>	variable name (type: character) of the income of the country of destination in the dataset <code>data</code> . The dependent variable <code>y</code> is divided by the product of the incomes <code>inc_d</code> and <code>inc_o</code> .
<code>maxloop</code>	maximum number of outer loop iterations. The default is set to 50. There will be a warning if the iterations did not converge.
<code>maxloop2</code>	maximum number of inner loop iterations. The default is set to 50. There will be a warning if the iterations did not converge.
<code>dec_places</code>	number of decimal places that should not change after a new iteration for the estimation to stop. The default is set to 4.
<code>vce_robust</code>	robust (type: logic) determines whether a robust variance-covariance matrix should be used. The default is set to TRUE. If set TRUE the estimation results are consistent with the Stata code provided at the website <a href="#">Gravity Equations: Workhorse, Toolkit, and Cookbook</a> when choosing robust estimation.
<code>verbose</code>	(type: logic) determines whether the estimated coefficients of each iteration should be printed in the console. The default is set to FALSE.
<code>data</code>	name of the dataset to be used (type: character). To estimate gravity equations, a square gravity dataset including bilateral flows defined by the argument <code>y</code> , ISO-codes of type character (called <code>iso_o</code> for the country of origin and <code>iso_d</code> for the destination country), a distance measure defined by the argument <code>dist</code> and other potential influences given as a vector in <code>x</code> are required. All dummy variables should be of type numeric (0/1). Missing trade flows as well as incomplete rows should be excluded from the dataset. Furthermore, flows equal to zero should be excluded as the gravity equation is estimated in its additive form. As, to our knowledge at the moment, there is no explicit literature covering the estimation of a gravity equation by SILS using panel data, cross-sectional data should be used.
<code>...</code>	additional arguments to be passed to functions used by SILS.

## Details

SILS is an estimation method for gravity models developed by Head and Mayer (2014) (see the references for more information). To execute the function a square gravity dataset with all pairs of countries, ISO-codes for the country of origin and destination, a measure of distance between the bilateral partners as well as all information that should be considered as dependent an independent variables is needed. Make sure the ISO-codes are of type "character". Missing bilateral flows as well as incomplete rows should be excluded from the dataset. Furthermore, flows equal to zero should be excluded as the gravity equation is estimated in its additive form. The function SILS

utilizes the relationship between the Multilateral Resistance terms and the transaction costs. The parameters are estimated by an iterative procedure. The function executes loops until the parameters stop changing significantly.

SILS is designed to be consistent with the Stata code provided at the website [Gravity Equations: Workhorse, Toolkit, and Cookbook](#) when choosing robust estimation. As, to our knowledge at the moment, there is no explicit literature covering the estimation of a gravity equation by SILS using panel data, we do not recommend to apply this method in this case.

### Value

The function returns the summary of the estimated gravity model as an `lm`-object. It furthermore returns the resulting coefficients for each iteration.

### References

For information on SILS as well as more information on gravity models, theoretical foundations and suitable estimation methods in general see

Head, K. and Mayer, T. (2014) <DOI:10.1016/B978-0-444-54314-1.00003-3>  
and

Anderson, J. E. and van Wincoop, E. (2003) <DOI:10.3386/w8079>  
as well as

Anderson, J. E. (1979) <DOI:10.12691/wjssh-2-2-5>

Anderson, J. E. (2010) <DOI:10.3386/w16576>

Baier, S. L. and Bergstrand, J. H. (2009) <DOI:10.1016/j.jinteco.2008.10.004>

Baier, S. L. and Bergstrand, J. H. (2010) in Van Bergeijk, P. A., & Brakman, S. (Eds.) (2010) chapter 4 <DOI:10.1111/j.1467-9396.2011.01000.x>

Head, K., Mayer, T., & Ries, J. (2010) <DOI:10.1016/j.jinteco.2010.01.002>

Santos-Silva, J. M. C. and Tenreyro, S. (2006) <DOI:10.1162/rest.88.4.641>  
and the citations therein.

See [Gravity Equations: Workhorse, Toolkit, and Cookbook](#) for gravity datasets and Stata code for estimating gravity models.

### See Also

[lm](#), [coefstest](#), [vcovHC](#)

### Examples

```
## Not run:
data(Gravity_no_zeros)

SILS(y="flow", dist="distw", x=c("rta"), inc_o="gdp_o", inc_d="gdp_d",
maxloop=50, maxloop2=50, dec_places=4, vce_robust=TRUE, verbose=FALSE,
data=Gravity_no_zeros)

SILS(y="flow", dist="distw", x=c("rta", "comcur", "contig"),
```

```
inc_o="gdp_o", inc_d="gdp_d", maxloop=50, maxloop2=50, dec_places=4,
vce_robust=TRUE, verbose=TRUE, data=Gravity_no_zeros)
```

```
## End(Not run)
```

---

Tetrads

*Tetrads*


---

## Description

Tetrads estimates gravity models by taking the ratio of the ratio of flows.

## Usage

```
Tetrads(y, dist, x, k = "USA", ell = "JPN", multiway_vcov = TRUE, data,
...)
```

## Arguments

<code>y</code>	name (type: character) of the dependent variable in the dataset <code>data</code> , e.g. trade flows. It is logged and taken as the dependent variable in the estimation.
<code>dist</code>	name (type: character) of the distance variable in the dataset <code>data</code> containing a measure of distance between all pairs of bilateral partners. It is logged automatically when the function is executed.
<code>x</code>	vector of names (type: character) of those bilateral variables in the dataset <code>data</code> that should be taken as the independent variables in the estimation. If an independent variable is a dummy variable, it should be of type numeric (0/1) in the dataset. If an independent variable is defined as a ratio, it should be logged. Unilateral effects drop as the ratio of ratios is taken.
<code>k</code>	reference importing country, default is set to "USA".
<code>ell</code>	reference exporting country, default is set to "JPN".
<code>multiway_vcov</code>	(type: logic) optional; determines whether a function implementing Cameron, Gelbach, & Miller (2011) multi-way clustering of variance-covariance matrices in the package <code>multiway_vcov</code> is used for the estimation. In case <code>multiway_vcov=TRUE</code> , the <code>cluster.vcov</code> function is used. The default value is set to <code>TRUE</code> .
<code>data</code>	name of the dataset to be used (type: character). To estimate gravity equations, a square gravity dataset including bilateral flows defined by the argument <code>y</code> , ISO-codes of type character (called <code>iso_o</code> for the country of origin and <code>iso_d</code> for the destination country), a distance measure defined by the argument <code>dist</code> and other potential influences given as a vector in <code>x</code> are required. All dummy variables should be of type numeric (0/1). Missing trade flows as well as incomplete rows should be excluded from the dataset. Furthermore, flows equal to zero should be excluded as the gravity equation is estimated in its additive form. When using

panel data, a variable for the time may be included in the dataset. Note that the variable for the time dimension should be of type: factor. See the references for more information on panel data.

... additional arguments to be passed to functions used by Tetrads.

## Details

Tetrads is an estimation method for gravity models developed by Head et al. (2010) (see the references for more information). To execute the function a square gravity dataset with all pairs of countries, ISO-codes for the country of origin and destination, a measure of distance between the bilateral partners as well as all information that should be considered as dependent and independent variables is needed. Make sure the ISO-codes are of type "character". Missing bilateral flows as well as incomplete rows should be excluded from the dataset. Furthermore, flows equal to zero should be excluded as the gravity equation is estimated in its additive form.

The function Tetrads utilizes the multiplicative form of the gravity equation. After choosing a reference importer `k` and exporter `e11` one can eliminate importer and exporter fixed effects by taking the ratio of ratios. Only those exporters trading with the reference importer and importers trading with the reference exporter will remain for the estimation. Therefore, reference countries should preferably be countries which trade with every other country in the dataset. After restricting the data in this way, Tetrads estimates the gravity equation in its additive form by OLS. As, by taking the ratio of ratios, all monadic effects diminish, no unilateral variables such as GDP can be included as independent variables.

Tetrads estimation can be used for both, cross-sectional as well as panel data. Nonetheless, the function is designed to be consistent with the Stata code for cross-sectional data provided on the website [Gravity Equations: Workhorse, Toolkit, and Cookbook](#) when choosing robust estimation. The function Tetrads was therefore tested for cross-sectional data. If Tetrads is used for panel data, the user may have to drop distance as an independent variable as time-invariant effects drop. For applying Tetrads to panel data see Head, Mayer and Ries (2010).

## Value

The function returns the summary of the estimated gravity model as an `lm`-object.

## References

For information on Tetrads see

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2011) <DOI:10.3386/t0327>

Head, K., Mayer, T., & Ries, J. (2010) <DOI:10.1016/j.jinteco.2010.01.002>

For more information on gravity models, theoretical foundations and estimation methods in general see

Anderson, J. E. (1979) <DOI:10.12691/wjssh-2-2-5>

Anderson, J. E. (2010) <DOI:10.3386/w16576>

Anderson, J. E. and van Wincoop, E. (2003) <DOI:10.3386/w8079>

Baier, S. L. and Bergstrand, J. H. (2009) <DOI:10.1016/j.jinteco.2008.10.004>

Baier, S. L. and Bergstrand, J. H. (2010) in Van Bergeijk, P. A., & Brakman, S. (Eds.) (2010) chapter 4 <DOI:10.1111/j.1467-9396.2011.01000.x>

Head, K. and Mayer, T. (2014) <DOI:10.1016/B978-0-444-54314-1.00003-3>  
 Santos-Silva, J. M. C. and Tenreyro, S. (2006) <DOI:10.1162/rest.88.4.641>  
 and the citations therein.

See [Gravity Equations: Workhorse, Toolkit, and Cookbook](#) for gravity datasets and Stata code for estimating gravity models.

For estimating gravity equations using panel data see

Egger, P., & Pfaffermayr, M. (2003) <DOI:10.1007/s001810200146>

Gomez-Herrera, E. (2013) <DOI:10.1007/s00181-012-0576-2>

and the references therein.

### See Also

[lm](#), [coefstest](#), [cluster.vcov](#)

### Examples

```
## Not run:
data(Gravity_no_zeros)

Tetrads(y="flow", dist="distw", x=c("rta"), k="USA", ell="JPN",
multiway_vcov=TRUE, data=Gravity_no_zeros)

Tetrads(y="flow", dist="distw", x=c("rta", "comcur", "contig"),
k="USA", ell="JPN", multiway_vcov=FALSE, data=Gravity_no_zeros)

## End(Not run)
```

---

Tobit

*Left-censored Tobit model with known threshold*

---

### Description

Tobit estimates gravity models in their additive form by conducting a left-censored regression, which, after adding the constant 1 to the dependent variable, utilizes  $\log(1)=0$  as the censoring value.

### Usage

```
Tobit(y, dist, x, added_constant = 1, data, ...)
```

**Arguments**

<code>y</code>	name (type: character) of the dependent variable in the dataset <code>data</code> , e.g. trade flows. The number 1 is added and the transformed variable is logged and taken as the dependent variable in the Tobit estimation with lower bound equal to $\emptyset$ as $\log(1)=\emptyset$ represents the smallest flows in the transformed variable.
<code>dist</code>	name (type: character) of the distance variable in the dataset <code>data</code> containing a measure of distance between all pairs of bilateral partners. It is logged automatically when the function is executed.
<code>x</code>	vector of names (type: character) of those bilateral variables in the dataset <code>data</code> that should be taken as the independent variables in the estimation. If an independent variable is a dummy variable, it should be of type numeric (0/1) in the dataset. If an independent variable is defined as a ratio, it should be logged. Unilateral variables such as country dummies or incomes can be added. If unilateral metric variables such as GDPs should be used as independent variables, those variables have to be logged first and the logged variable can be used in <code>x</code> . Interaction terms can be added.
<code>added_constant</code>	scalar (type: numeric); represents the constant to be added to the dependent variable. The default value is 1. The minimum of $\log(y+\text{added\_constant})$ is taken as the left boundary in the Tobit model. In the often used case of <code>added_constant=1</code> , the dependent variable is left-censored at value $\emptyset$ as $\log(1)=\emptyset$ .
<code>data</code>	name of the dataset to be used (type: character). To estimate gravity equations, a square gravity dataset including bilateral flows defined by the argument <code>y</code> , ISO-codes of type character (called <code>i so_o</code> for the country of origin and <code>i so_d</code> for the destination country), a distance measure defined by the argument <code>dist</code> and other potential influences given as a vector in <code>x</code> are required. All dummy variables should be of type numeric (0/1). Missing trade flows as well as incomplete rows should be excluded from the dataset. Zero trade flows are allowed.
<code>...</code>	additional arguments to be passed to <code>Tobit</code> .

**Details**

`Tobit` represents the left-censored Tobit (Tobin, 1958) approach utilizing a known censoring threshold which is often used when several gravity models are compared. When taking the log of the gravity equation flows equal to zero constitute a problem as their log is not defined. Therefore, in the execution of the function the number 1 is added to all flows and the  $\log(\text{flows}+1)$  is taken as the dependent variable. The Tobit estimation is conducted using the `censReg` function and setting the lower bound equal to  $\emptyset$  as  $\log(1)=\emptyset$  represents the smallest flows in the transformed variable. A Tobit regression represents a combination of a binary and a linear regression. This procedure has to be taken into consideration when interpreting the estimated coefficients. The marginal effects of an explanatory variable on the expected value of the dependent variable equals the product of both the probability of the latent variable exceeding the threshold and the marginal effect of the explanatory variable of the expected value of the latent variable.

To execute the function a square gravity dataset with all pairs of countries, ISO-codes for the country of origin and destination, a measure of distance between the bilateral partners as well as all information that should be considered as dependent an independent variables is needed. Missing bilateral flows as well as incomplete rows should be excluded from the dataset. Zero trade flows are allowed.

Up to now, the function is designed for cross-sectional data, but can be easily extended to panel data using the [censReg](#) function. A robust estimations is not implemented to the present as the [censReg](#) function is not compatible with the [vcovHC](#) function.

For more elaborate Tobit functions, see [ET\\_Tobit](#) for the Eaton and Tamura (1994) threshold Tobit model where instead of simply adding number 1 to the data the threshold is estimated or [EK\\_Tobit](#) for the Eaton and Kortum (2001) Tobit model where each zero trade volume is assigned a country specific interval with the upper bound equal to the minimum positive trade level of the respective importing country.

### Value

The function returns the summary of the estimated gravity model as a [censReg](#)-object.

### References

For more information on gravity models, theoretical foundations and estimation methods in general see

Anderson, J. E. (1979) <DOI:10.12691/wjssh-2-2-5>

Anderson, J. E. (2010) <DOI:10.3386/w16576>

Anderson, J. E. and van Wincoop, E. (2003) <DOI:10.3386/w8079>

Baier, S. L. and Bergstrand, J. H. (2009) <DOI:10.1016/j.jinteco.2008.10.004>

Baier, S. L. and Bergstrand, J. H. (2010) in Van Bergeijk, P. A., & Brakman, S. (Eds.) (2010) chapter 4 <DOI:10.1111/j.1467-9396.2011.01000.x>

Head, K., Mayer, T., & Ries, J. (2010) <DOI:10.1016/j.jinteco.2010.01.002>

Head, K. and Mayer, T. (2014) <DOI:10.1016/B978-0-444-54314-1.00003-3>

Santos-Silva, J. M. C. and Tenreyro, S. (2006) <DOI:10.1162/rest.88.4.641>

and the citations therein.

Especially for Tobit models see

Tobin, J. (1958) <DOI:10.2307/1907382>

Eaton, J., & Tamura, A. (1994) <DOI:10.3386/w4758>

Eaton, J., & Kortum, S. (2001) <DOI:10.3386/w8070>.

See [Gravity Equations: Workhorse, Toolkit, and Cookbook](#) for gravity datasets and Stata code for estimating gravity models.

### See Also

[censReg](#)

### Examples

```
## Not run:
# Example for data with zero trade flows
data(Gravity_zeros)

Gravity_zeros$lgdp_o <- log(Gravity_zeros$gdp_o)
```

```
Gravity_zeros$lgdp_d <- log(Gravity_zeros$gdp_d)

Tobit(y="flow", dist="distw", x=c("rta", "lgdp_o", "lgdp_d"),
      added_constant = 1, data=Gravity_zeros)

## End(Not run)
```

# Index

## \*Topic **datasets**

Gravity\_no\_zeros, 21

Gravity\_zeros, 23

BVU, 3

BVW, 6

censReg, 14, 15, 41, 42

cluster.vcov, 40

coefstest, 5, 8, 10, 18, 20, 26, 29, 32, 34, 37,  
40

DDM, 8

EK\_Tobit, 10, 14, 42

ET\_Tobit, 12, 13, 42

Fixed\_Effects, 16

glm, 20, 26, 28, 29, 33, 34

glm.nb, 25, 26

glm2, 19, 20

GPML, 18, 25, 28, 33

Gravity\_no\_zeros, 21

Gravity\_zeros, 23

lm, 5, 7–10, 17, 18, 31, 32, 37, 39, 40

NBPML, 19, 24, 28, 33

NLS, 19, 25, 27, 33

OLS, 29

PPML, 19, 25, 28, 32

SILS, 35

Surv, 11, 12

survreg, 11, 12

Tetrads, 38

Tobit, 12–14, 40

vcovHC, 5, 8, 10, 14, 18, 20, 26, 29, 32, 34, 37,  
42