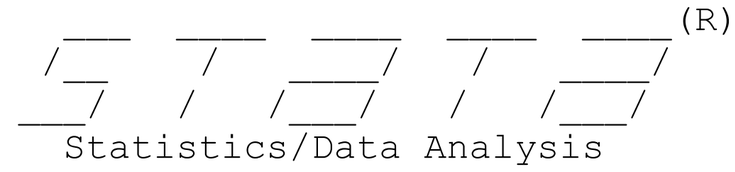

Title

gintreg — Generalized Interval Regression

Syntax

gintreg *depvar1* *depvar2* [*indepvars*] [*if*] [*in*] [, *options*]

depvar1 and *depvar2* should have the following form:

Type of data		*depvar1*	*depvar2*
point data	*a* = [*a*,*a*]	*a*	*a*
interval data	[*a*,*b*]	*a*	*b*
left-censored data	(-inf,*b*]	.	*b*
right-censored data	[*a*,inf)	*a*	.

If using grouped data then the form will be similar:

Type of data		*depvar1*	*depvar2*	*frequency*
point data	*a* = [*a*,*a*]	*a*	*a*	*n*
interval data	[*a*,*b*]	*a*	*b*	*n*
left-censored data	(-inf,*b*]	.	*b*	*n*
right-censored data	[*a*,inf)	*a*	.	*n*

options	Description
Main	
distribution(dist_type)	dist_type may be gb2, gg, lnormal, sgt, sged, or normal; default is normal.
constraints(*numlist***)**	specified linear constraints by number to be applied. Can use this option along with **distribution** to allow for any distribution in the SGT or GB2 family trees.
frequency(*varlist***)**	if using group data specify variable that denotes frequency.
Model	
sigma(*varlist***)**	allow the log of sigma to vary as a function of independent variables; can use with dist_type normal, lnormal, gg, gb2, sgt, or sged.
lambda(*varlist***)**	allow lambda to vary as a function of independent variables; can use with dist_type sgt or sged.
p(*varlist***)**	allow p to vary as a function of independent variables; can use with dist_type gb2, gg, sgt, or sged.
q(*varlist***)**	allow q to vary as a function of independent variables; can use with dist_type gb2 or sgt.
SE/Robust	
vce(*vcetype***)**	*vcetype* may be **oim**,**robust**, **cluster** *clustvar*, **opg**, **bootstrap**, or **jackknife**.
robust	use robust standard errors.
cluster(*varlist***)**	cluster standard errors with respect to sampling unit *varlist*.
Estimation	
initial(*numlist***)**	initial values for p,q and lambda in that order. if the distribution does not have p, q or lambda, key in initial values for mu and lnsigma in that order.
maximize options	control the maximization process

Display

showconstonly	Show the estimated constant only model.
eyx(*stat***)**	Show the expected value of *depvar* conditional on *indepvars* at level of stat; default is mean.

Description

gintreg fits a model of *depvar* on *indepvars* using maximum likelihood where the dependent variable can be point data, interval data, right-censored data, or left-censored data. This is a generalization of the built in STATA command **intreg** and will yield identical estimates if the normal distribution option is used. Unlike **intreg**, **gintreg** allows the underlying variable of interest to be distributed according to a more general distribution including all distributions in the Skewed Generalized T family and Generalized Beta of the Second Kind tree. Finally, **gintreg** allows for grouped data when using the frequency option.

The assumed model for interval regression is y = XB + eps where only the thresholds containtin the latent variable y are observed, X is a vector of explanatory variables with a corresponding coefficient vector B and eps is assumed to be independently and identically distributed random distrubances. The upper and lower thresholds for y can be denoted by U and L respectively.

The conditional probability that y is in the interval (L,U) is: Pr(L <= y <= U) = F(eps = U - XB: theta) - F(eps: L-XB: theta), where F denotes the cdf of the random disturbances and theta denotes a vector of distributional parameters. **gintreg** uses MLE on the corresponding log-likelihood function to estimate beta (displayed as mu or delta in the output) and the distributional parameters theta.

Options

Main

distribution(dist_type) specifies the type of distribution used in the interval regressions. **gintreg** will use a log-likelihood function composed of the pdf and cdf of this distribution (pdf for point data and cdf for intervals and censored observations). dist_type may be gb2, gg, lnormal, sgt, sged, or normal; Default is normal.

constraints(*numlist)* specified linear constraints by number to be applied. Can use this option along with **distribution** to allow for any distribution in the SGT or GB2 family trees. Constraints are defined using the **constraint** command; see **[R] constraint**.

frequency(*varlist)* if using grouped data, specify the variable that denotes the frequency of the observation. Can be in percentage terms or levels as **gintreg** will normalize by summing the value of frequency for all observations.

Model

The *indepvars* specified will allow the location parameter (mu or delta) to vary as a function of the independent variables. The other parameters in the distribution can also be a function of explanatory variables by using the commands below. If the user specifies a parameter that is not part of dist_type then **gintreg** will indicate an error; e.g. specifying independent variables for q when using the Generalized Gamma distribution.

sigma(*varlist)* allows the log of sigma to be a function of *varlist* and can model heteroskedasticity.

lambda(*varlist)* allows lambda to be a function of *varlist* and can model skewness.

p(*varlist)* allows p to be a function of *varlist*. A shape parameter that impacts the tail thickness and peakedness of the distribution.

q(*varlist)* allows q to be a function of *varlist*. A shape parameter that impacts the tail thickness and peakedness of the distribution.

Standard Errors

vce(*vcetype***)** specifies the type of standard error reported, which includes types that are robust to some kinds of misspecification (robust), that allow for intragroup correlation (cluster clustvar), and that are derived from asymptotic theory (oim, opg); see **[R] vce_option**.

robust use robust standard errors.

cluster(*varlist***)** cluster standard errors with respect to sampling unit *varlist*.

Estimation

initial(*numlist***)** list of numbers that specifies the initial values of the parameters in the constant only model. This must be equal to the number of distributional parameters; i.e. two for the normal and log-normal (mu, sigma), one for the Generalized Gamma (p), two for the GB2 (p, q) and the SGED (p, lambda), and three for the SGT(p, q, lambda).

maximize_options: **difficult**, **technique(***algorithm_spec***)**, **iterate(#)**, [**no**]**log**, **trace**, **gradient**, **showstep**, **hessian**, **showtolerance**, **tolerance(#)**, **ltolerance(#)**, **nrtolerance(#)**; see **[R] maximize**. Allowed techniques include Newton-Raphson (nr), Berndt-Hall-Hausman (bhhh), Davidon -Fletcher-Powell (dfp), and Broyden-Fletcher-Goldfarb-Shanno (bfgs). The default algorithm is Newton-Raphson.

Display

showconstonly : gintreg will always estimate the constant only model first prior to estimating the model with *indepvars*, but this output is suppressed. Use this option to see the estimate of the constant only model.

eyx(*stat***)** This option helps with inference in models with a positive distribution (gb2, gg, lnormal). At the end of the STATA printout, it displays the estimated conditional value of the dependent variable with respect to the independent variables being at the level of stat. This result is returned and is accessible after estimation by e(eyx).

If stat is not specified then the independent variables will be taken at their mean levels:

stat

mean	mean values of independent variables
min	minimum values of independent variables
max	maximum values of independent variables
p1	1st percentile of independent variables
p5	5th percentile of independent variables
p10	10th percentile of independent variables
p25	25th percentile of independent variables
p50	50th percentile of independent variables
p75	75th percentile of independent variables
p90	90th percentile of independent variables
p95	95th percentile of independent variables
p99	99th percentile of independent variables

Remarks

If the optimization is not working, try using the difficult option. You can also use the option **technique(bfgs)**, or the other two **technique** options, which are often more robust than the default **technique(nr)**.

Examples

We have a dataset containing wages, truncated and in categories. Some of the observations on wages appear below

```
    wage1    wage2
    20       25       meaning 20000 <= wages <= 25000
    50        .       meaning 50000 <= wages
```

Setup

```
    . webuse intregxmpl
```

Interval regression with a normal distribution

```
    . gintreg wage1 wage2 age c.age#c.age nev_mar rural school tenure
```

Interval regression with a gb2 distribution (use difficult option)

```
    . gintreg wage1 wage2 age c.age#c.age nev_mar rural school, distribution(gb2)
      difficult
```

Interval regression with a gb2 distribution with the expected value of the dependent variable evaluated when the independent variables are at the 25 percentile (E[Y|X] appears at the end of the printout

```
    . gintreg wage1 wage2 age c.age#c.age, distribution(gb2) eyx(p25) difficult
```

Interval regression with a sgt distribution allowing sigma to vary as a function of independent variables

```
    . gintreg wage1 wage2 age c.age#c.age nev_mar rural school tenure, distribution(sgt)
      sigma(age)
```

Interval regression using the burr3 distribution

```
    . constraint define 1 [q]_cons=1

    . gintreg wage1 wage2 age c.age#c.age nev_mar rural school tenure, distribution(gb2)
      constraints( 1 )
```

Author

Authored by James McDonald and Jacob Orchard at Brigham Young University. For support contact Jacob at orchard.jake@gmail.com.

References

James B., McDonald, Olga Stoddard, and Daniel Walton. 2016. *On using interval response data in experimental economics*, working paper.

James B., McDonald, and Daniel Walton. 2016. *Distributional Assumptions and the Estimation of Contingent Valuation Models.*